Computational Modeling of Catalytic Mechanisms of Glycyl Radical Enzymes

Dissertation zur Erlangung der Doktorwürde
der Fakultät für Biologie, Chemie und Geowissenschaften
der Universität Bayreuth

vorgelegt von

Mikolaj Feliks

UNIVERSITÄT
BAYREUTH

January 29, 2014

This doctoral thesis was prepared at the University of Bayreuth between March 2009 until June 2013, supervised by Prof. Dr. G. Matthias Ullmann.

This is a full reprint of the dissertation submitted to obtain the academic degree of Doctor of Natural Sciences (Dr. rer. nat.) and approved by the Faculty of Biology, Chemistry and Geosciences of the University of Bayreuth.

Acting dean:	Prof. Dr. Rhett Kempe
Date of submission:	03.07.2013
Date of defence (disputation):	09.01.2014

Doctoral committee:

Prof. Dr. G. Matthias Ullmann	1st reviewer:	Prof. Dr. G. Matthias Ullmann
Prof. Dr. Franz Xavier Schmid	2nd reviewer:	Prof. Dr. Holger Dobbek
Prof. Dr. Jürgen Senker	Chairman:	Prof. Dr. Rainer Schobert
Prof. Dr. Rainer Schobert		

I would like to express my gratitude to people without whom this thesis would not have been possible:

- Prof. Dr. G. Matthias Ullmann (supervisor) for guidance, constructive advice and constant support
- Dr. Berta Martins for disclosing unpublished experimental results and many valuable discussions
- Dr. Martin J. Field for the pDynamo software library
- Prof. Dr. W. Andrzej Sokalski for introducing me into the quite exciting field of computational chemistry
- Colleagues from the Computational Biochemistry Group, in particular Lars Müller and Timm Essigke, for creating a friendly and supportive work environment
- Friends from the Welcome Center of the Univeristy of Bayreuth for the great time spent together exploring Oberfranken
- My parents Maria and Jerzy Feliks for their continuous support throughout my student years

Contents

Abstract

In this thesis, different computational methods have been used to study the catalytic mechanisms of three glycyl radical enzymes.

The mechanism of the B_{12}-independent glycerol dehydratase has been elucidated by combining continuum electrostatic and density functional theory calculations. A mechanism for the dehydration of glycerol is proposed that does not involve a complex intramolecular 1→2 shift of the middle hydroxyl group, as previously suggested. Instead, the enzyme uses a pair of residues in the active site, glutamate and histidine, to facilate direct release of a water molecule from glycerol. The mechanism of 4-hydroxyphenylacetate decarboxylase has been explored based on continuum electrostatic and hybrid quantum chemical/molecular mechanical calculations. The calculations suggest that the substrate is activated to a radical form by two simultaneous transfers, one of an electron to the thiyl radical and second of a proton to the active site glutamate. This activation mode has not been reported for any known radical enzyme. The mechanism of pyruvate formate-lyase has been investigated based for the first time on the complete enzyme model. The key new finding is that quenching of the formyl radical is performed by one of the active site cysteines and not by coenzyme A, as previously suggested.

Zusammenfassung

In der vorliegenden Arbeit werden unterschiedliche Berechnungsmethoden angewendet, um die katalytischen Mechanismen von drei Glycyl-Radikalenzyme zu untersuchen. Der Mechanismus der B_{12}-unabhängigen Glyceroldehydratase wurde mit Hilfe von Kontinuumselektrostatik- und Quantenchemierechnungen untersucht. Der vorgeschlagene Dehydratisierungsmechanismus benötigt nicht die komplexe 1→2 Migration der mittleren Hydroxyl-Gruppe des Glycerins. Die Dehydratisierung des Glycerins involvierte einige Reste im aktiven Zentrum des Enzyms, im speziellen ein Glutamat und ein Histidin. Der Mechanismus der 4-Hydroxyphenylacetat-Decarboxylase wurde mit Hilfe von Kontinuumselektrostatik und Quantenchemischen/Molekülmechanischen Rechnungen untersucht. Die Berechnungen zeigen, dass das Substrat zu einer radikalen Form durch zwei gleichzeitige Transfers aktiviert wird. Zum einen wird ein Elektron zum Thiylradikal transferiert und zum anderen ein Proton zu einem Glutamat im aktiven Zentrum. Dieser Aktivierungs-

mechanismus wurde bisher in keinem anderen Radikalenzym beobachtet. Der Mechanismus der Pyruvat-Formiat-Lyase wurde zum ersten Mal mit einem kompletten Enzymsmodell untersucht. Die wichtigste neue Erkenntnis ist, dass die Deaktivierung des Formyl-Intermediats nicht durch Koenzym A durchgeführt wird, sondern durch ein Cystein im aktiven Zentrum.

List of Abbreviations

4Hpad ..... 4-Hydroxyphenylacetate decarboxylase
ARNR ..... Anaerobic (class III) ribonucleotide reductase
BDE ..... Bond dissociation energy
BSS ..... Benzylsuccinate synthase
CTL ..... Choline trimethylamine-lyase
DFT ..... Density functional theory
EPR ..... Electron paramagnetic resonance spectroscopy
GDH ..... B_{12}-dependent glycerol dehydratase
iGDH ..... B_{12}-independent glycerol dehydratase
GRE ..... Glycyl radical enzyme
NMR ..... Nuclear magnetic resonance
NEB ..... Nudged elastic band method
PBE ..... Poisson-Boltzmann electrostatics
PES ..... Potential energy surface
PFL ..... Pyruvate formate-lyase
QC/MM ..... Quantum chemistry/molecular mechanics
SAM ..... *S*-adenosylmethionine
TMA ..... Trimethylamine
TST ..... Transition state theory

1 Introduction

Radical enzymes have been recognized as extremely robust catalysts in many biological systems.[1–4] These enzymes use the high reactivity of radicals to initiate reactions of practically nonreactive compounds. The ability of radical enzymes to catalyze chemically demanding reactions is particularly interesting for industry due to their potential use in organic synthesis.[5,6] Over the years, considerable efforts have been made to understand at the molecular level the intriguing chemistry of radical enzymes.[7,8] However, many aspects of the radical-based catalysis still remain unknown because the reactions catalyzed by radical enzymes involve unstable, short-lived intermediates that are difficult to study experimentally. On the other hand, the methods of computational chemistry for studying enzymatic reactions have in the past decade evolved to a degree that they can compete with experiments offering reasonable accuracy and reliability.[9–12] These methods provide insights into the intrinsic chemistry of the catalyzed reaction that are beyond the scope of most experimental techniques, for example the determination of transition states on the reaction path.

It is in this context that I employ electrostatic,[13–15] quantum chemical,[16–18] and combined quantum chemical/molecular mechanical calculations[19–22] to study the catalytic mechanisms of three glycyl radical enzymes,[23,24] namely the B_{12}-independent glycerol dehydratase (iGDH), 4-hydroxyphenylacetate decarboxylase (4Hpad) and pyruvate formate-lyase (PFL). The mechanisms of the former two enzymes have not been studied before experimentally nor computationally. My calculations show that both iGDH and 4Hpad adopt catalytic mechanisms that are without precedence among the radical enzymes discovered to date. Although the mechanism of PFL is thought to be well understood, the performed calculations are able to explain some of the still confusing aspects of the catalyzed reaction.

The B_{12}-independent glycerol dehydratase is a novel glycyl radical enzyme catalyzing the conversion of glycerol into 3-hydroxylpropionaldehyde.[25,26] The previously suggested reaction mechanism assumed that the dehydration of glycerol is accomplished by an intramolecular 1→2 shift of the middle hydroxyl group, as seen in the catalysis by B_{12}-dependent enzymes. The shift was postulated to involve a cyclic transition state and result in a 1,2-diol intermediate that spontaneously releases a water molecule.[26–28] However, the electrostatic and density functional theory calculations suggest that the mechanistically

complicated migration of the middle hydroxyl group can be avoided[29] (Manuscript A). One of the histidine residues in the active site of iGDH donates a proton to the leaving hydroxyl group, which leads to the immediate formation of a water molecule. The release of water is coupled to a proton transfer from one of the terminal hydroxyl groups of glycerol to glutamate in the active site, which generates a C=O double bond of the future product. Despite many structural similarities between the active sites of iGDH and its B_{12}-dependent counterpart,[30,31] GDH, these enzymes seem to adopt totally different catalytic mechanisms for the dehydration of glycerol.[32–34] The key histidine residue, whose protonation state seems to determine the reaction route taken by the enzyme, is doubly-protonated in iGDH and singly-protonated in GDH. The unexpected catalytic mechanism of iGDH derived from the calculations is one of the highlights of the present thesis.

4-hydroxyphenylacetate decarboxylase catalyzes the production of *p*-cresol from its substrate.[35,36] The electrostatic and hybrid quantum chemical/molecular mechanical calculations performed on the full-enzyme model of 4Hpad reveal a catalytic mechanism that involves an unusual activation mode of the substrate, which has not been described for any of the known radical enzymes (Manuscripts B and C). Usually, the substrate in radical enzymes is activated to a radical form by abstraction of a hydrogen atom by transient radical species in the active site, such as the thiyl radical. The calculations indicate, however, that the activation of 4-hydroxyphenylacetate is accomplished by two simultaneous transfers from the substrate, first of an electron to the radical cysteine and second of the phenolic proton to the active site glutamate. The substrate is therefore activated by the netto abstraction of a hydrogen atom. The exceptional separation of proton and electron during the substrate activation by 4Hpad as suggested by the calculations is another highlight of this thesis.

Pyruvate formate-lyase was the first glycyl radical enzyme to be discovered and its mechanism of action has been extensively studied since then.[37,38] The calculations done for the first time on the complete enzyme model attempt to answer the questions regarding some of the unrevealed mechanistic aspects of the PFL-catalyzed reaction (Manuscript D). One of the new findings is that the reaction proceeds through a stable protein-bound tetrahedral intermediate formed by pyruvate and the radical cysteine in the active site. Moreover, the calculations indicate that another active site cysteine is responsible for quenching of the formyl radical. It is therefore quite likely that the release of formate

precedes the binding of coenzyme A to the active site.

In summary, the present thesis discusses computational studies on the catalytic mechanisms of three glycyl radical enzymes. The B_{12}-independent glycerol dehydratase and 4-hydroxyphenylacetate decarboxylase have been shown to perform catalysis via unusual and previously unreported reaction mechanisms. The mechanism of pyruvate formate-lyase has been updated with new information from the calculations on the full-enzyme model. The performed calculations provide novel insights into the molecular basis of radical enzymatic catalysis and prepare the ground for future experiments.

2 Computer simulations of enzyme catalysis

Enzymes are at the heart of most, if not all, life processes. Therefore, understanding how they work is one of the most important goals of modern biochemistry.[39] Enzymes are complex molecular systems of a size frequently exceeding thousands of atoms and their mechanism of action is often obscure and difficult to predict. Many experimental techniques have been established for studying reactions catalyzed by enzymes. These techniques include, for example, protein crystallography, NMR spectroscopy, site-directed mutagenesis and isotopic labeling.[40,41]

In the last years, computer-assisted molecular modeling has become an important complement to experimental studies on enzyme catalysis.[9–12,39,42–49] Thanks to the growing computing power and the development of new theoretical methods and software, molecular modeling can now greatly contribute to our understanding of how enzymes work. The methods of computational chemistry have matured to the point when they can offer accuracy comparable to that of experiments. For example, the development of density functional theory[50] enables studying molecular systems of >100 atoms, such as enzyme active sites. These systems are of a size that is usually prohibitive for standard *ab inito* methods of quantum chemistry. Some aspects of enzymatic reactions are difficult to tackle experimentally, for example the characterization of transition states and intermediates on the reaction path. This applies especially to radical enzymes, since they catalyze reactions involving transient, short-lived radical species. However, these species can be computationally studied in the same way as the stable, long-lived ones. Calculations can provide detailed insights into the catalytic mechanism, pinpoint catalytically important interactions inside the enzyme active site and help explain the origins of catalytic activity.

The goal of computational modeling of an enzyme-catalyzed reaction is to understand how the enzymatic machinery works at the molecular level. This involves identification of all intermediates and transition states along the catalytic cycle and calculation of their relative energies. The energy barrier for a particular reaction step can be calculated taking the energy difference between the reactants and the transition state. Studying a catalytic mechanism also means identifying which groups in the protein or cofactors are involved in the reaction. Once they are known, the overall catalytic effect of an enzyme can be decomposed into different contributions.

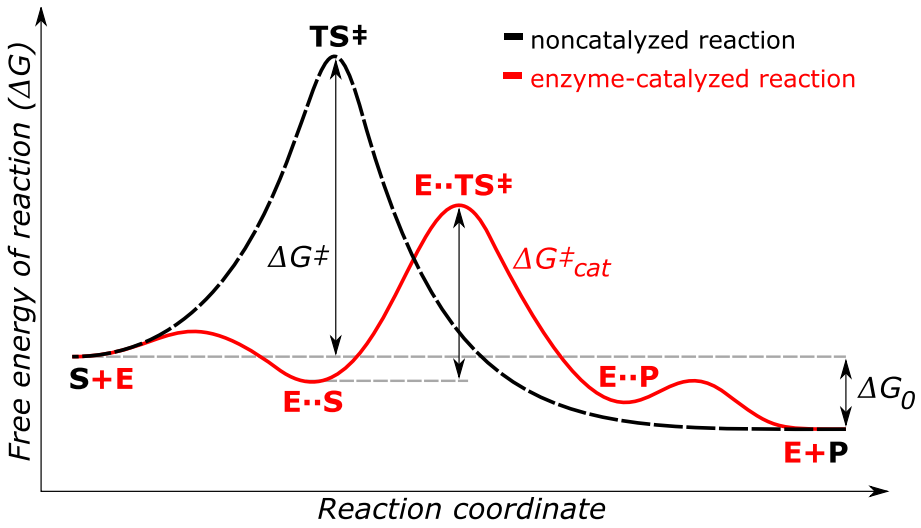

Figure 1: Free energy profile for a hypothetical one-step reaction in solution and catalyzed by an enzyme.

2.1 The principles of enzymatic catalysis

The concept of enzymatic catalysis is depicted on Fig. 1. During the reaction, the substrate (S) is converted into the product (P), passing the transition state ($TS^{\ddagger}$). Without the enzyme catalyst (black dashed line), the energy barrier for substrate conversion is $\Delta G^{\ddagger}$. Alternatively, the substrate first binds to the enzyme active site (red line), forming an enzyme-substrate complex ($E\cdots S$). The energy barrier is now reduced by $\Delta\Delta G^{\ddagger} = \Delta G^{\ddagger} - \Delta G^{\ddagger}_{cat}$. The value of $\Delta\Delta G^{\ddagger}$ corresponds to the catalytic efficiency of the enzyme. Upon conversion, an enzyme-product complex is generated ($E\cdots P$). The product can dissociate from the active site. The enzyme works by lowering the energy barrier ($\Delta G^{\ddagger}$) but does not change the reaction energy (ΔG_0).

There are many factors that contribute to the catalytic efficiency of enzymes. The initial "lock and key" model proposed by Fisher assumed that binding of a substrate to the enzyme activates the substrate to its reactive conformation.[51] This hypothesis was later replaced by the induced fit model ("ground state destabilization"). With the introduction of transition state theory, it became obvious that enzymes provide a pre-organized environment that can stabilize transition states.[52,53] Electrostatic effects due to the polar enzymatic surrounding are the major contributing factor to the stabilization of the transition state in many enzymatic reactions.[54,55] It is now accepted that enzymes can also work by entropic guidance, near attack conformations, desolvation effects, low-barrier

hydrogen bond, effects of protein dynamics, covalent catalysis or tunneling.[56–60] Radical enzymes work by stabilizing radical species that are by themselves catalytically active ("negative catalysis"[61]). To perform the catalysis, many enzymes combine different types of catalytic effects.

The rate constant of a chemical reaction, defined within the framework of transition state theory (TST), takes the form:

$$k(T) = \frac{k_B T}{h} exp\left(\frac{-\Delta G^{\ddagger}}{RT}\right) \tag{1}$$

where k is the rate constant, T is the temperature, $\Delta G^{\ddagger}$ is the free energy of activation and k_B, h and R are the Bolzmann, Planck and universal gas constants, respectively. The rate constant defines in principle how quickly reactants change into products. Eq. 1 links the rate constant, which can be determined experimentally, with the free energy of reaction obtained computationally. Given the Gibbs free energy of activation, expressed by $\Delta G^{\ddagger} = \Delta H^{\ddagger} - T\Delta S^{\ddagger}$, the rate constant becomes:

$$k(T) = \frac{k_B T}{h} exp\left(\frac{-\Delta S^{\ddagger}}{R}\right) exp\left(\frac{-\Delta H^{\ddagger}}{RT}\right) \tag{2}$$

As seen in Eq. 1, the rate constant depends exponentially on the free energy of activation. For the enzyme models discussed in this thesis, the methods of density functional theory were used throughout, in particular B3LYP. The B3LYP method provides the accuracy of ± 3 kcal/mol in predictions of reaction energetics, which leads to error in the rate of two orders of magnitude. On the other hand, this level of accuracy is fully sufficient for the determination of reaction mechanisms and catalytic effects. For the enzymatic reactions studied in this thesis, the calculated reaction energetics correspond only to the enthalpy part of Eq. 2. Entropic effects were not included in the models.

2.2 Computational methods

Theoretical modeling of chemical reactions requires the use of quantum chemical methods for proper description of bond-breaking and bond-forming processes. Classical methods, based on the force fields such as Amber or CHARMM,[62,63] can be used for studying the protein structure and dynamics but they are not applicable to chemical reactions.

Density functional theory

Density functional theory (DFT) of electronic structure has extended the application of quantum chemistry from small molecules to interesting and challenging molecular systems of biological relevance, in particular enzymes. Considering the trade-off between accuracy and computational cost, DFT is currently the best approach for studying enzymatic reactions. In *ab inito* quantum chemical methods, for example Hartree-Fock or MP2, an electronic wavefunction is used as the basic variable that depends on $4N$ coordinates ($3N$ spatial and N spin) for an N-electron system. This function is used to calculate properties of the system, such as geometry, energy and frequencies. In DFT, however, the electron density function $\rho(\vec{r})$ is used that depends on only three spatial coordinates. Employing electron density rather than wavefunction can considerably accelerate calculations due to its much less coordinate dependence. The use of the electron density function is justified, since there exist a unique relationship between $\rho(\vec{r})$ and the properties of the system, as demonstrated by Hohenberg and Kohn.[64] Electron density determines everything about the system in its ground state. For modeling of reactions that involve excited state chemistry, more sophisticated DFT methods are required, for example TD-DFT.[65]

DFT is based on the two Hohenberg-Kohn theorems, which state that 1) the ground state electron density of a system uniquely determines the external potential and 2) the density-dependent functional obeys the variational principle. The total energy of a system can be expressed as a functional of the electron density:

$$E_{tot}[\rho] = T[\rho] + V_{ee}[\rho] + V_{ne}[\rho] \tag{3}$$

where ρ is electron density, T is the kinetic energy of the electrons, V_{ee} is the electron-electron repulsion and V_{ne} is the nuclei-electron attraction. The first two terms are independent of the nuclear positions and represent the density functional. However, in the Hohenberg-Kohn formalism, the exact form of the density functional remains unknown. This problem has been to some extent overcome by introducing the Kohn-Sham orbitals.

The most widely used functionals are the hybrid functionals, which incorporate parts of the Hartree-Fock exchange, for example Becke's three parameter functional[66–68] (B3LYP). These functionals depend on parameters (a, b and c) that are fitted to reproduce molecular properties of interest. B3LYP can be expressed as follows:

$$E_{\mathrm{XC}}^{\mathrm{B3LYP}} = (1-a)E_x^{\mathrm{Slater}} + aE_x^{\mathrm{HF}} + bE_x^{\mathrm{B88}} + (1-c)E_c^{\mathrm{VWN}} + cE_c^{\mathrm{LYP}} \quad (4)$$

where E_x^{Slater} is the Dirac-Slater exchange, E_x^{HF} is the Hartree-Fock exchange, E_x^{B88} is the gradient correction to exchange, E_c^{VWN} and E_c^{LYP} are the correlation functionals of Vosko, Wilkand, Nusair and Lee, Yang, Parr, respectively.

The B3LYP functional performs well for the prediction of molecular geometries and energies of organic molecules and complexes, including the ones that carry a radical.[7] These two properties are crucial for the study of reaction mechanisms. The known deficiencies of approximate functionals include self-interaction error, near-degeneracy error and problems with description of dispersion interactions. Because of this, some interactions like for example π-stacking between DNA bases cannot be treated with DFT.[69] On the other hand, hybrid functionals, such as B3LYP, exhibit better performance due to some error cancellation.

Electrostatic calculations

Many protein residues, ligands and cofactors bind and release protons depending on the current pH and interactions with neighboring sites. Examples of such residues include aspartates, glutamates, histidines and others whose side-chains contain titratable sites. The knowledge of the protonation behavior of these residues and ligands is key for the modeling of enzymatic catalysis.[13–15,54,70] Charged groups can be either directly involved in the enzymatic reaction or influence the chemistry at the reacting region by longer-range electrostatic interactions. The available experimental methods for studying the titration of proteins, such as calorimetry, are unable to assign the protonation states to individual titratable sites. Nuclear magnetic resonance spectroscopy can in principle detect the positions of hydrogen atoms but it is limited to rather small proteins. However, the protonation probabilities of individual sites in the protein can be calculated based on the Poisson-Boltzmann electrostatic model (PBE) combined with a Monte Carlo sampling.[14] The computational studies presented in this thesis highlight the crucial role of electrostatic calculations in the elucidation of enzymatic mechanisms.[15]

There are several methods for the calculation of protein electrostatic potentials and interaction energies.[13,71] The most frequently used is the Poisson-Boltzmann method based on the Poisson equation relating the spatial variation of the protein electrostatic potential

to the charge density and dielectric constant. The system of interest is described starting from the available crystal structure. A set of partial atomic charges, usually taken from the force field, is used to calculate the charge distribution $\rho_{protein}(\vec{r})$. Mobile ions outside the protein are assumed to adopt a Boltzmann distribution. Different dielectric constants are assigned to the protein (usually ϵ=4) and the outer medium (ϵ=80 for aqueous solution) to account for the lower polarizability of protein environment. The electrostatic potential $\phi(\vec{r})$ can be calculated from the linearised Poisson-Boltzmann equation:

$$\vec{\nabla}\left[\epsilon(\vec{r})\vec{\nabla}\phi(\vec{r})\right] = \frac{1}{\epsilon_0}\left[\rho_{protein}(\vec{r}) + \sum_i^I \left(\frac{c_i z_i^2 e^2}{RT}\phi(\vec{r})\right)\right] \tag{5}$$

where c_i is the concentration of ions of type i, z_i is the formal charge of an ion, e is the elementary charge, R is the universal gas constant and T is the temperature. The summation is performed for I different types of ions.[72,73] Because of its complexity, the Poisson-Boltzmann equation is usually solved by using numerical methods that map the protein model onto a cubic lattice, for example the finite difference method.

For a protein with N titratable sites, there can be 2^N different protonation states. On top of the previously calculated electrostatic potential, the energy $G^{(n)}$ of state n in the function of pH can be evaluated as:

$$\begin{aligned} G^{(n)}(\mathrm{pH}) &= \sum_i^N (x_i^{(n)} - x_i^{(0)})(\mathrm{pH} - \mathrm{p}K_i^{\mathrm{intr}}) \\ &\quad + \frac{1}{2}\sum_i^N \sum_j^N (x_i^{(n)} - x_i^{(0)})(x_j^{(n)} - x_j^{(0)})W_{i,j} \end{aligned} \tag{6}$$

where $x_i^{(n)}$ and $x_i^{(0)}$ are the state vectors representing the present and the reference macroscopic protonation state, respectively, collecting the microscopic protonation states of individual sites, $\mathrm{p}K_i^{\mathrm{intr}}$ is the intrinsic pK of site i and $W_{i,j}$ is the interaction energy between two sites. For most proteins, the calculation of all protonation states is not possible because 2^N can be a very large number. Instead, the protonation state energies can be sampled by using the Metropolis Monte Carlo method.[70,74]

2.3 Three approaches to the modeling of enzymatic catalysis

Theoretical modeling of an enzyme-catalyzed reaction mechanism can in general be accomplished at one of the three levels of approximation, i.e. increasing complexity.

Small-molecule approach

The first option is to focus on a small-molecule model of isolated reactants in the gas-phase, neglecting the remaining parts of the enzyme.[42] Obviously, a gas-phase model cannot account for the actual enzymatic mechanism, because it does not include the catalytic effects that arise from the protein environment. These effects can be steric, electrostatic and of different nature. However, calculations on small models allow to investigate the intrinsic chemistry of the catalyzed reaction with accurate methods and at relatively low computational cost. These calculations can serve as a reference for the studies on larger and more realistic models. Because of the limited computing power in the past, calculations on gas-phase models were the first available method for studying enzymatic systems. Some of the radical enzymes have been investigated computationally by using the small-molecule approach, for example pyruvate formate-lyase[75,76] or the enzymes from the B_{12}-dependent family.[8]

Cluster model approach

To better reproduce the behavior of the actual enzymatic system, the model can be extended by inclusion of additional, nonreactive parts of the active site. In the cluster model approach,[16–18,77,78] a model is constructed based on the crystal structure of the enzyme by selecting a discreet number of atoms from the residues that make up the active site. If only a structure of the substrate-free form of the enzyme is available, molecular docking can be used to position the substrate in the active site.[79] In addition to the intrinsic chemistry of the catalyzed reaction, a cluster model also tries to capture the effect of surrounding residues on the reacting region. Since a truncated enzyme model is used, two procedures are employed to account for the missing outer parts of the enzyme. To compensate for the lack of steric effects, selected atoms are kept constrained at their crystallographic positions during the geometry optimizations. The selection of fixed atoms often requires a large number of calculations with different sets of constraints until a

model is found that can keep the integrity of the initial active site during the geometry optimizations in vacuum. Errors that arise due to the coordinate-locking procedure, for example an increased strain of the model, can be avoided by shifting the constrained atoms further away from the reacting center. To simulate the electrostatic polarization effects exerted by the protein surrounding on the cluster, dielectric cavity techniques can be used, for example the polarizable-like continuum model[80,81] (PCM). The missing outer parts of the enzyme are mimicked by a homogeneous medium with a dielectric constant usually set to ϵ=4. For large cluster models, the solvation effects usually saturate, i.e. with an increasing size of the model the choice of the dielectric constant becomes less critical.[82]

The cluster model approach has been demonstrated very successful in studies on different enzymes. Cluster models are easiest to setup for systems like metalloenzymes where the chemistry at the active site is dominated by strong electrostatic interactions due to the presence of the central metal ion. In such cases, further located parts of the enzyme are quite unimportant for the catalyzed reaction and the use of a small cluster model is usually sufficient for proper description of the enzymatic system.

Full-enzyme (QC/MM) approach

Although the performance of the cluster model approach is sufficient for the identification of key features of the catalytic mechanism, it may sometimes be necessary to include additional parts of the enzyme into the model. However, the use of a full-enzyme model requires that the system is partitioned into two parts, the reacting one described quantum chemically and the non-reacting one that can be handled at a simpler molecular mechanical level. That is because the size of a complete enzyme model, which usually exceeds thousands of atoms, would be prohibitive for pure quantum chemical treatment. Given the present computer speed and the available theoretical methods, a model of ∼300 atoms is the maximum that can be studied in a reasonable time using density functional theory. For more sophisticated *ab inito* quantum chemical methods, the maximum available size of the model is usually far less than 100 atoms. On the other hand, force fields such as Amber or CHARMM have been well-parametrized for studying proteins.[62,63] These force fields are known to reproduce ground-state geometries of proteins more accurately than semi-empirical methods of quantum chemistry.[83]

In the hybrid quantum chemical/molecular mechanical approach,[19–22,84] a model of

the enzyme is constructed, usually on top of the available crystal structure, by dividing the system (S) into two parts. The inner part (I) consists of the reacting region and is treated at a quantum chemical level. The outer part (O), which encompasses the rest of the enzyme, is treated at a molecular mechanics level. The QC/MM concept was first introduced in 1976 by Warshel and Levitt[85] but the method has only become popular in the last years. Fig. 2 shows the principle of the QC/MM method.

Due to the strong interactions between the inner (QC) and the outer (MM) regions, the total energy of the system cannot be written as a simple sum of the energies of the subsystems. To account for these interactions, coupling terms are introduced. Two different energy partitioning schemes have been developed within the QC/MM framework, namely the subtractive[86] (ONIOM) and additive scheme. For the calculations described in this thesis, the additive scheme was always used. The additive scheme takes the following form:

$$E^{\mathrm{add}}_{\mathrm{QC/MM}}(\mathrm{S}) = E_{\mathrm{MM}}(\mathrm{O}) + E_{\mathrm{QC}}(\mathrm{I}+\mathrm{L}) + E_{\mathrm{QC-MM}}(\mathrm{I},\mathrm{O}) \tag{7}$$

where $E^{add}_{QC/MM}(S)$ is the total energy of the system, $E_{MM}(O)$ is the energy of the outer region treated at the MM-level, $E_{QC}(I+L)$ is the energy of the inner region including link-atoms treated at the QC-level and $E_{QC-MM}(I,O)$ is the coupling term collecting the interactions between the two regions. For a given reaction step, the difference in electronic energy ($\Delta E^{\mathrm{add}}_{\mathrm{QC/MM}}(\mathrm{S})$) between the intermediate and the transition state gives the potential energy barrier.

The QC/MM method can in principle accommodate any combination of QC- and MM-potentials. From a technical point of view, the applied QC-method must be able to account for the external point charges that represent the outer region. The wavefunction at the reacting region is solved in the presence of point charges. Usually, semi-empirical or DFT methods are used as QC-potentials. The energy of the outer region is calculated from the potential energy function of a force field, for example Amber or CHARMM.[62,63] The energy of these force fields is calculated from the general equation:

$$E_{MM}(O) = E_{\mathrm{bond}} + E_{\mathrm{angle}} + E_{\mathrm{dihedral}} + E_{\mathrm{torsion}} + E_{\mathrm{elec}} + E_{\mathrm{vdW}} \tag{8}$$

The consecutive energy terms correspond to the extension of bonds, bending of angles,

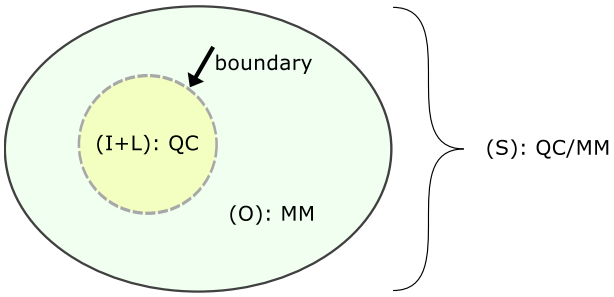

Figure 2: Conceptual drawing showing the principle of the QC/MM approach.

rotation of dihedral angles, electrostatic interactions and Van der Waals interactions, respectively.

The key problem of the QC/MM method is the proper treatment of the boundary between the inner and outer regions.[21] If the boundary cuts through covalent bonds, the valencies of these bonds must be saturated. This is usually done by introducing link-atoms similar to hydrogen atoms that cap the QC-region at the boundary. Other approaches include boundary atoms or frozen localized orbitals.[87]

The exploration of the potential energy surface (PES) of an enzyme model is the central issue in the modeling of enzymatic reactions. Stable structures on the reaction path, i.e. reactants and intermediates, are represented by minima on the PES, whereas transition states are represented by saddle points. Finding the transition state geometry is probably the most challenging aspect in studies on enzymatic reaction paths. A variety of geometry optimization methods have been developed for this purpose.[88,89] To be able to properly characterize the transition state, it is required to calculate a matrix of second-order derivatives of energy (Hessian). However, for large QC/MM models, obtaining and manipulating this matrix becomes computationally too expensive and approximate techniques for localizing transition states have to be used. One approach is the so-called linear transit, where a relaxed PES scan is performed in steps along the assumed major component of the reaction coordinate, for example a distance between two atoms. More sophisticated methods involve generation of a chain of frames interpolated between two energy minima that are bound with a special merit function and optimized together to converge on the minimum energy path, for example the nudged elastic band method[90,91] (NEB). The NEB method has been used in the present work to study the reaction paths in 4Hpad and PFL.

3 The chemistry of radical enzymes from anaerobic bacteria

Radicals are highly reactive atoms, molecules or ions that have an open electron shell, i.e. a single valence electron. This single electron seeks to pair with another electron, which can be derived from the second radical or a σ- or π-bond. To date, many enzymes have been identified that employ a radical-mediated catalytic mechanism.[1–4,24,27,92,93] The protein environment protects reactive radical species from quenching agents, dimerizations and other side reactions and allows the radicals to persist.[92] On the other hand, reactions involving small ligands such as molecular oxygen are difficult to avoid, since these ligands can easily penetrate the protein matrix. For example, glycyl radical enzymes lose their activity due to the cleavage of the protein backbone at the glycyl radical site after exposure to molecular oxygen. Therefore, radical enzymes are usually found in anaerobic microorganisms where the risk of oxygen-induced side reactions is minimal. Whenever possible, nature has evolved enzymes to catalyze biochemical transformations via simple acid-base mechanisms. Since radical chemistry is more demanding in terms of generation, storage and controlled decomposition of open-shell species, radical enzymes are used only when there is no alternative.[4]

Radicals can be introduced into proteins either by homolysis of weak σ-bonds or by electron transfer. There are two radical generators in proteins, namely the vitamin B_{12} (adenosylcobalamin) and *S*-adenosylmethionine (SAM). Alternatively, the radical can sometimes be introduced into the protein by one-electron transfer to coenzyme A or thymidine, leading to the formation of a ketyl radical anion. In such cases, the energy for generating the radical is usually provided by ATP or light. For some enzymes the mechanism by which they obtain the radical is still uncertain, for example acyl-CoA dehydrogenase and 4-hydroxybutyryl-CoA.

The concept of enzymatic radical catalysis is depicted on Fig. 3. Although the conversion of a substrate is thermodynamically feasible, the reaction cannot proceed due to the high energy barrier (A). The situation changes after a catalytic radical ($R^{\bullet}$), for example the thiyl radical, has been introduced into the active site (B). The substrate is now activated to a reactive radical form by abstraction of a hydrogen atom by the catalytic radical. The energy barrier for the conversion of the substrate-derived radical intermediate

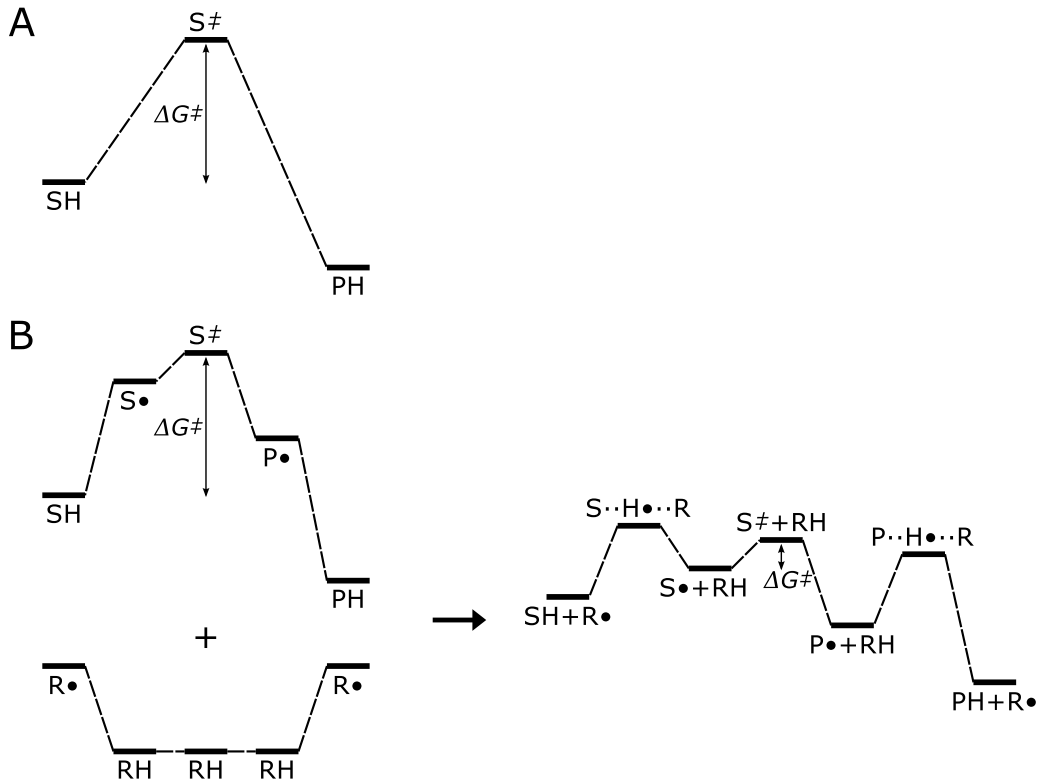

Figure 3: A simplified potential energy diagram showing the principles of enzymatic radical catalysis.[94] (A) Reaction without a catalytic radical. (B) Reaction with a catalytic radical. SH, substrate; PH, product; $S^{‡}$, transition state; $R^{•}$, catalytic radical; $\Delta G^{‡}$, energy barrier for substrate conversion.

($S^{•}$) is significantly lower than that for the substrate. The conversion $S^{•}\rightarrow P^{•}$ gives a product-related radical intermediate. In the last step, $P^{•}$ abstracts a hydrogen atom from the enzyme, which gives the final product (P) and regenerates the initial catalytic radical. In summary, a radical-mediated reaction mechanism usually involves three reaction steps, namely the activation of a substrate by radical species in the active site, conversion of a substrate-derived radical intermediate and deactivation of a product-related intermediate into the final product.

In the following sections, three different classes of radical enzymes from anaerobic bacteria are briefly outlined. Although the present thesis deals specifically with modeling of catalysis by glycyl radical enzymes, there are structural and functional connections between the members of all three classes. For example, SAM-dependent enzymes are essential activators of glycyl radical enzymes. The knowledge of these connections is key for understanding the chemistry of glycyl radical enzymes.

3.1 Vitamin B_{12}-dependent enzymes

Vitamin B_{12} (adenosylcobalamin) is a complex organometallic compound that assists many enzymatic reactions. The octahedral structure of vitamin B_{12} shows a Co^{3+} ion embedded in a corrin ring and two axial ligands interacting with the central ion (Fig. 4). The binding of the cobalt ion is provided by four in-plane nitrogen atoms of corrin. The first ligand is a 5'-deoxyadenosyl residue bound to the cobalt ion via a weak σ-bond. This unusual metal-carbon bond has a dissociation energy of only $\sim$31 kcal/mol in solution and $\sim$15 kcal/mol in protein environment. The nature of the Co$-$C bond in vitamin B_{12} has been extensively studied[8,95,96] but remains poorly understood. One of the key questions still to be addressed is why the bond dissociation energy (BDE) of this bond is significantly different between solution and protein environment. The second ligand can be either 5,6-dimethylbenzimidazole ("base on"), which connects with the corrin ring D, or the imidazole ring of histidine ("base off, his on").

All reactions catalyzed by B_{12}-dependent enzymes start from the homolytic cleavage of the weak Co$-$C bond of the coenzyme. The cleavage gives Co^{+2} and 5'-deoxyadenosyl radical (Ado-$CH_2^{\bullet}$), which abstracts a hydrogen atom from the substrate, generating a substrate-derived radical intermediate. This intermediate rearranges into a product-related radical intermediate and abstracts a hydrogen atom back from 5'-deoxyadenosine, which leads to the final product. The rearrangement usually involves an intramolecular 1$\rightarrow$2 shift of the functional group, for example the middle hydroxyl group of glycerol as in the reaction catalyzed by the B_{12}-dependent glycerol dehydratase.[32,33] Finally, Ado-$CH_2^{\bullet}$ and the Co^{+2} ion can recombine to complete the catalytic cycle. The reactions catalyzed by the B_{12}-dependent enzymes have been extensively studied, both experimentally and computationally.[8]

There are two groups of B_{12}-dependent enzymes. One group comprises of "base on" irreversible eliminases, for example ethanolamine ammonia-lyase, class II ribonucleotide reductase and glycerol dehydratase. Class II RNR is somewhat exceptional, since it uses a cysteine residue in the active site to relay the radical between 5'-deoxyadenosine and the substrate. In all other enzymes from this family, the 5'-deoxyadenosyl radical directly attacks the substrate. Another interesting feature of class II RNR is that the nucleotide reduction by this enzyme does not seem to be accompanied by a 1$\rightarrow$2 rearrangement reaction that is typical for other B_{12}-dependent enzymes.

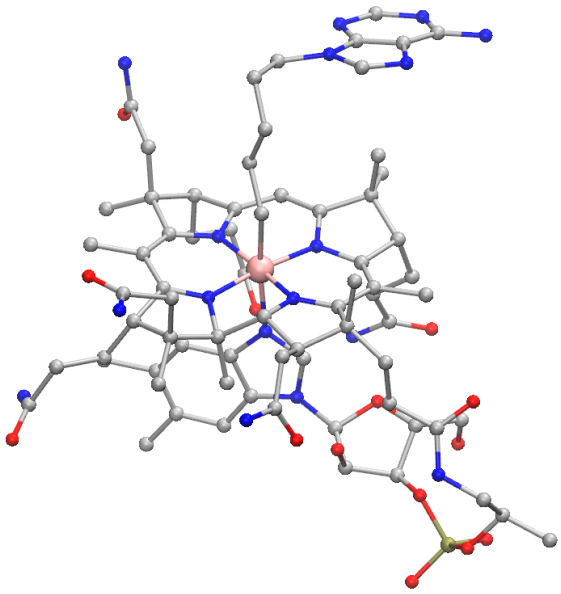

Figure 4: Vitamin B_{12} analog as seen in the crystal structure of diol dehydratase (PDB code 1EEX[30]). The cobalt ion is depicted in pink color.

The second group is represented by "base off, his on" reversible mutases catalyzing rearrangements of the carbon skeleton and migrations of the amino group. Examples of mutases include methylmalonyl CoA mutase and glutamate mutase. Interestingly, all known eliminases have counterparts in the B_{12}-independent enzyme family. These counterparts are either SAM-dependent or glycyl radical enzymes, as in the case of glycerol dehydratase. On the other hand, counterparts of mutases have not been reported and may not exist at all.

A unique feature of the enzymes from the B_{12}-dependent family is that the radical disappears by reformation of the Co−C bond in the coenzyme. Other enzymes can only carry a permanent radical. However, some of the B_{12}-dependent enzymes become inactive after a few thousands of turnovers. For example, diol dehydratase can survive only about 10^4 turnovers. The advantage of these enzymes is that they are much less sensitive to molecular oxygen, unlike glycyl radical and [4Fe−4S] cluster/SAM-containing enzymes.

3.2 *S*-adenosylmethionine radical enzymes

The *S*-adenosylmethionine cofactor (SAM) undergoes transient cleavage to methionine and 5'-deoxyadenosyl radical, which can further propagate the radical by abstracting

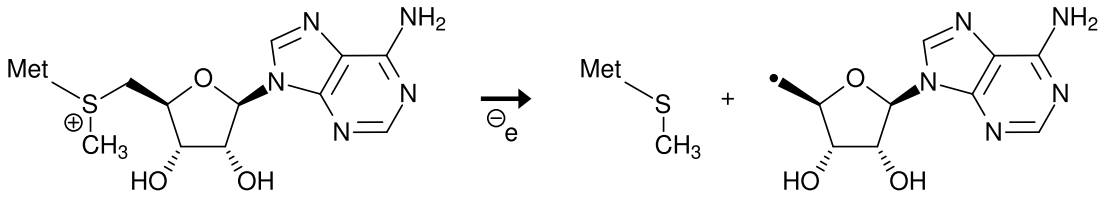

Figure 5: Activation of SAM by addition of an extra electron, leading to methionine and 5'-deoxyadenosyl radical.

hydrogen atoms from substrate molecules or precursor proteins of glycyl radical enzymes. Cleavage of the rather strong sulfur–adenosyl bond in SAM (BDE ∼60 kcal/mol) is only possible if assisted by a low-potential one-electron donor, such as ferredoxin or flavodoxin. These donors contain iron-sulfur clusters ([4Fe−4S]) that reduce SAM by supplying an extra electron to the cofactor, as depicted on Fig. 5. After the cleavage of the S−C bond, the resulting 5'-deoxyadenosyl radical can be irreversibly released or recycled after each turnover to regenerate the cofactor.

More than 3000 potential SAM-dependent enzymes have been detected in genomes of anaerobic and aerobic bacteria, fungi, plants and animals. However, only a few of these enzymes have been studied in detail. These include the reversible lysine-2,3-aminomutase (the first discovered SAM-dependent enzyme) and spore photoproduct-lyase[97–99] (a DNA repair enzyme) as well as irreversible biotin synthase, oxygen-independent coproporphyrinogen III oxidase and Mo-cofactor biosynthesis protein A (MoaA). The latter three enzymes together with MoaC are involved in the synthesis of molybdopterin, which is a crucial cofactor for many metalloenzymes. An important group of irreversible SAM-dependent enzymes are the activases of glycyl radical enzymes.

3.3 Glycyl radical enzymes

As of this writing, six glycyl radical enzymes (GRE) of known function have been identified,[24,100–102] namely pyruvate formate-lyase (PFL), anaerobic ribonucleotide reductase (ARNR), benzylsuccinate synthase (BSS), B_{12}-independent glycerol dehydratase (iGDH), 4-hydroxyphenylacetate decarboxylase (4Hpad) and choline trimethylamine-lyase (CTL). These enzymes are able to perform chemically difficult transformations, such as cleavage of C−C bonds (PFL, 4Hpad), cleavage of C−O bonds (iGDH, ARNR), addition to double bonds (BSS) or cleavage of C−N bonds (CTL). Fig. 6 shows reactions catalyzed by these

enzymes.

GREs carry a stable radical localized on the protein backbone in the vicinity of the active site. Electron paramagnetic resonance experiments established that the radical in activated GREs resides on a glycine residue.[103,104] In a cell, GREs are synthesised as catalytically inactive precursors and require post-translational activation to the reactive radical form by dedicated enzymes from the *S*-adenosylmethionine family. These SAM-dependent activating enzymes are specific to every GRE enzyme. The activation of a GRE is performed by the 5'-deoxyadenosyl radical produced by the SAM-dependent enzyme. This radical abstracts hydrogen from a conserved glycine residue inside the catalytic unit of the glycyl radical enzyme, generating a stable glycyl radical.

It has been established that the exceptional stability of the glycyl radical results primarily from the so-called captodative effect.[7,105,106] In this effect, the unpaired electron of glycine is delocalized between the adjacent groups of the protein backbone, namely the amino group (electron donor) and the carbonyl group (electron acceptor). The captodative effect derives from the summation of the resonance electron withdrawal by the carbonyl group and the electron donation by the amino group. The glycyl radical is most stable for the planar conformation of protein backbone.[105,106] For steric reasons, the planar conformation is easier to adopt for protein residues with small side-chains and ideally without a side-chain. Therefore, the lack of a side-chain in glycine is another factor that is believed to contribute to the increased stability of the glycyl radical. Interestingly, in the available crystal structures of GREs, the key glycine is visible in a distorted, non-planar conformation. Calculations performed on different models based on the crystal structure of pyruvate formate-lyase indicate that the stability of this non-planar glycyl radical is noticeably lower than that of the active site thiyl radical.[107] It has been postulated that GRE can tune the stability of the glycyl radical through conformational control at the spin-carrying site.[106]

Molecular oxygen has been shown to irreversibly inactivate all GREs by attacking the glycyl radical site, which eventually leads to the cleavage of the protein backbone. The high sensitivity of GREs to molecular oxygen limits the distribution of these enzymes to bacteria living in strictly anaerobic environments. Oxygen-induced inactivation of GREs has been studied by means of DFT calculations.[108] In the first step, addition of an oxygen molecule to the glycyl radical gives a peroxyl radical ($ROO^{\bullet}$). The peroxyl radical abstracts

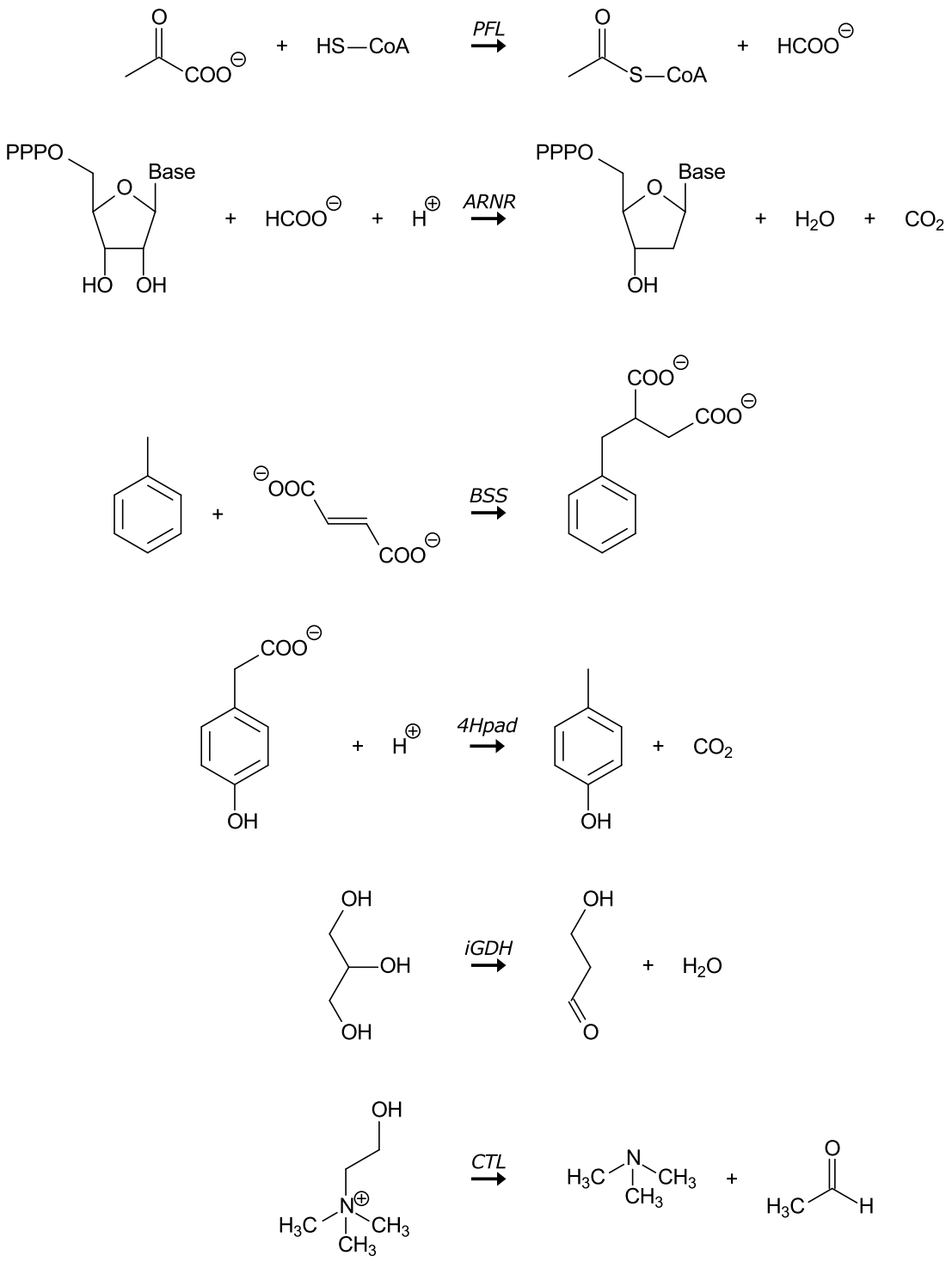

Figure 6: Reactions catalyzed by glycyl radical enzymes of known function: pyruvate formate-lyase (PFL), anaerobic ribonucleotide reductase (ARNR), benzylsuccinate synthase (BSS), 4-hydroxyphenylacetate decarboxylase (4Hpad), B_{12}-independent glycerol dehydratase (iGDH) and choline TMA-lyase (CTL).

a hydrogen atom from the neighboring cysteine, generating a thiyl radical. Transfer of the hydroxyl group between the glycine and the thiyl radical gives a glycyl-alkoxyl radical (Gly-O$^{\bullet}$), which later attacks back the cysteine. The resulting sulfinyl radical (RSO$^{\bullet}$) can be observed in EPR experiments. Fragmentation of the protein backbone is done through the cleavage of the hydroxyglycine moiety or the glycyl-alkoxyl radical.

Glycine in GRE serves only as radical storage. After binding of the substrate to the enzyme active site, the radical shifts from the glycine to the cysteine residue, generating a thiyl radical, which in the next reaction step attacks the substrate. The corresponding cysteine is always located in between the radical storage on glycine and the ligand in the active site. All studied GREs possess one such cysteine with the exception of PFL that uses two cysteines in the catalysis. The substrate is usually activated to a radical form by abstraction of a hydrogen atom by the thiyl radical. However, pyruvate formate-lyase and the recently studied 4-hydroxyphenylacetate decarboxylase are exceptional GREs that use different mechanisms for activating their substrates. Namely, in PFL the thiyl radical attacks pyruvate bound in the active site to generate a protein-bound tetrahedral intermediate, which in an entry point to further transformations. In 4Hpad, electron and proton are abstracted from the substrate separately by the thiyl radical and the active site glutamate, respectively. Both activation modes are discussed in detail in Manuscripts B and C.

Although only a few GRE have been characterized, genome-sequencing experiments predict that many more of these enzymes are present in different anaerobic bacteria. Some of these novel enzymes have been detected but the catalyzed reactions remain unknown. For example, the misannotated PFL2 enzyme shows the active site similar to that of iGDH.[109,110] It is likely that this enzyme is involved in the dehydration of polyols. In the following, the glycyl radical enzymes of known function are briefly reviewed.

Benzylsuccinate synthase

BSS[111,112] is involved in the anaerobic metabolism of toluene in denitrifying bacteria, such as *Thauera aromatica*. The enzyme catalyzes the addition of the methyl carbon of toluene to fumarate, which gives benzylsuccinate. BSS is a complex enzyme composed of three units, each in two copies ($\alpha_2\beta_2\gamma_2$). Unit α contains a sequence motif that is characteristic for the glycyl radical site. The crystal structure of BSS has not been solved so far. On the

basis of biochemical and spectroscopic data, a catalytic mechanism has been proposed for BSS that was later examined by means of DFT calculations on small-molecule models.[113] In the first reaction step, which is common for all GREs, the radical is transferred from the glycyl radical storage to the cysteine in the active site, generating a thiyl radical. Next, the thiyl radical abstracts a hydrogen atom from the methyl group of toluene, which gives a benzyl radical. This radical stereospecifically attacks the double bond of the second substrate, fumarate, which leads to the formation of the 2-benzylsuccinate-3-yl radical. Finally, the hydrogen atom is re-added from the cysteine, generating (*R*)-benzylsuccinate as the final product. The elucidation of the crystal structure of BSS will provide additional structural information and help verify the proposed catalytic mechanism.

4-Hydroxyphenylacetate decarboxylase

4Hpad catalyzes the production of *p*-cresol from its substrate[35,36] . Decarboxylations are key reactions in many biological systems.[114] The phenolic product of 4Hpad is a virulence factor that is used by clostridia against competitive organisms in the human intestine. Although the 4Hpad activity has long been known in several bacteria,[115] for example *C. difficile* and *C. scatologenes*, the crystal structure of the enzyme was solved only recently[116] (see Manuscript B). The crystal structure shows a $(\beta\gamma)_4$ tetramer of heterodimers. Each heterodimer is composed of a larger catalytic β-subunit and a smaller [4Fe−4S] cluster-containing γ-subunit. Among GRE, only 4Hpad and BSS have been reported to contain additional subunits. The exact role of these smaller subunits is unknown but they have been shown to be important for enzyme activation.[117] The larger subunit harbors a characteristic glycyl radical site. The initially proposed mechanism of catalysis by 4Hpad assumed that the catalytic cycle starts from the abstraction of the phenolic hydrogen atom of 4-hydroxyphenylacetate by the thiyl radical. However, in the crystal structure it can be seen that the substrate binds to the enzyme with its carboxyl group close to the thiyl radical, while the hydroxyl group is hydrogen-bound to the glutamate at the opposite end of the active site. A new Kolbe-type[118] catalytic mechanism has been put forward that is now supported by QC/MM calculations[119] (see Manuscript C). In this mechanism, 4-hydroxyphenylacetate is activated to a radical form by two simultaneous transfers, first of an electron from the substrate to the Cys503 thiyl radical and second of a proton from the substrate's hydroxyl group to Glu637. The decarboxylation is done

by cleavage of the C−C bond in the substrate's acetate moiety, generating free carbon dioxide and a product-related radical intermediate. The release of CO_2 is coupled to the proton back-transfer from Glu637 to the ligand. Finally, Glu505 protonates Cys503, which in turn quenches the radical intermediate by hydrogen atom transfer, yielding *p*-cresol as the final product. The calculated mechanism is in line with experiments suggesting that both Cys503 and Glu637 are crucial for the catalysis by 4Hpad.

B_{12}-independent glycerol dehydratase

iGDH from clostridia, for example *C. glycolicum* and *C. butyricum*, catalyzes the fermentative conversion of glycerol into 3-hydroxypropionaldehyde.[25] This reaction enables anaerobic bacteria to grow on glycerol as the main nutrient.[120] Moreover, microbial conversion of glycerol has recently become interesting for biofuel industry.[5,6,121] The crystal structure of the iGDH enzyme shows two monomers in an asymmetric unit.[26] Glycerol binding in the active site of iGDH is provided by an extensive network of hydrogen bonds. The other enzyme known for converting glycerol is the B_{12}-dependent glycerol dehydratase. Interestingly, the specific activity of iGDH is considerably grater than that of its B_{12}-dependent counterpart, GDH, which is probably related to the inactivation of the former after a limited number of turnovers.[26,31] The catalytic mechanism of GDH is believed to involve a 1→2 transfer of the middle hydroxyl group of glycerol to yield an unstable geminal diol. The same mechanism was initially put forward for the B_{12}-independent enzyme. However, recent electrostatic and DFT calculations[29] (see Manuscript A) reveal that the mechanism of catalysis by iGDH most likely does not involve the complicated 1→2 migration step, as previously suggested. Instead, iGDH employs an interesting proton donating/accepting system that consists of His164 and Glu435. This system facilitates direct release of a water molecule from the substrate, without intermediacy of geminal diol species. In the light of these calculations, the catalytic mechanisms of iGDH is rather similar to the one of another glycyl radical enzyme, namely class III RNR, which also performs the cleavage of a C−O bond at some point of its catalytic cycle.

Ribonucleotide reductase

RNR catalyzes the reduction of ribonucleotide triphosphates to deoxyribonucleotides.[23,122,123] This reaction provides building blocks for the synthesis of DNA. The RNR activity is

therefore present in all living organisms. The chemically complicated replacement of the ribose 2'-hydroxyl group by a hydrogen atom is made possible through the use of a radical-mediated mechanism. Three classes of RNR have been identified based on their primary structure, reactivity with molecular oxygen and the way the enzyme obtains the radical. Class I uses the tyrosyl radical and is present in aerobic bacteria and eukaryotes. The tyrosyl radical is generated by self-processing that involves a non-heme [$Fe^{3+}-O-Fe^{3+}$] metal site and molecular oxygen. Class II from *Thermatoga maritima* is a vitamin B_{12}-dependent enzyme that functions independently of oxygen. Class III (ARNR) was first observed in *E. coli* growing under strictly anaerobic conditions. The use of formate as the reducing substrate links ribonucleotide reduction by ARNR to the anaerobic metabolism of pyruvate, which is controlled by another glycyl radical enzyme, PFL. ARNR is a two-enzyme complex of a quaternary ($\alpha_2\beta_2$) structure. The larger unit α performs the catalysis. The smaller unit β is equivalent to a SAM-dependent activating enzyme, since it harbors a redox-active [4Fe−4S] center, *S*-adenosylmethionine and reduced flavodoxin that generate a stable glycyl radical in the larger unit. The catalytic mechanism of ARNR has been a subject of extensive computational studies.[7,124,125] In the first step of the generally accepted reaction mechanism, a thiyl radical abstracts hydrogen at the 3'-position of the ribonucleotide ring. The hydroxyl group at the 2'-position splits off from ribonucleotide. Next, formate donates a proton to the leaving 2'-hydroxyl group to yield a water molecule and subsequently accepts another proton from the 3'-hydroxyl group, generating a 3'-carbonyl group. The active site cysteine completes the reaction by hydrogen atom transfer to the 2'-position of ribonucleotide. An alternative mechanism has been proposed for ARNR that includes two formates participating in the reaction.[125] Given the recent calculations on iGDH, the catalytic role of formate in ARNR seems to be reminiscent to that of His164/Glu435 in iGDH.

Pyruvate formate-lyase

PFL catalyzes the reaction of pyruvate and coenzyme A (CoA) to generate formate and acetyl-CoA, which is a key component of the anaerobic carbon metabolism in many prokaryotes. PFL was the first enzyme for which a radical was detected during the catalysis.[37,103] This radical was later assigned to the Gly734 residue.[104] The crystal structure of PFL from *E. coli* shows a dimer of two identical units.[38,126] The PFL-catalyzed reaction

has been examined by means of experimental and theoretical methods.[75,76,107,127,128] An interesting feature of PFL is that the enzyme employs two active site cysteines that can relay the radical. Other GRE enzymes use only one cysteine. The currently accepted catalytic mechanism starts from the radical transfer between the glycyl radical storage and one of the cysteines. The radical is further propagated to the second cysteine, which attacks the carbonyl moiety of pyruvate, forming a protein-bound radical intermediate (see Manuscript D). The subsequent cleavage of the C−C bond in the intermediate yields a free formyl radical. To complete the reaction, the first cysteine donates a hydrogen atom to the formyl radical, generating formate. Since there is no structural information available on binding of CoA to the active site of PFL, less is known about the second stage of the reaction. Most likely, once formate has left the active site, a molecule of CoA binds in the vicinity of the acylated cysteine. Transacylation between these two results in acetyl-CoA and regenerates the thiyl radical. Unlike for ARNR, iGDH and 4Hpad, the catalytic machinery of PFL is rather simple, since it does not involve residues in the active site other than the two radical cysteines.

Choline trimethylamine-lyase

CTL from the anaerobic bacterium *Desulfovibrio desulfuricans* is a novel glycyl radical enzyme catalyzing the production of trimethylamine (TMA).[101,102] Choline is an important compound for functioning of cell membranes, methyl transfer reactions and neurotransmission. On the other hand, TMA can be used by the bacterium as a source of carbon. The CTL-catalyzed reaction involves cleavage of the C−N bond in choline. This type of reaction has not been described for any of the known GREs. Initially, the conversion of choline was postulated to be catalyzed by a hypothetical B_{12}-dependent enzyme, since the breakdown of a cognate compound, ethanolamine, is catalyzed by the B_{12}-dependent ethanolamine ammonia-lyase. However, the EPR experiments as well as the analysis of the bacterial genome clearly indicate the involvement of a glycyl radical enzyme in this reaction.[101]

4 Synopsis of published and submitted manuscripts

Glycyl radical enzymes have evolved to perform a variety of chemically difficult transformations. The focus of this work was on the computational investigation of enzymes catalyzing the cleavage of C−O and C−C bonds. Because of the evident structural and functional similarities between the enzymes from the glycyl radical family, one of the important issues of the present work was to study whether the catalytic mechanisms of these enzymes share some common features. For example, the crystal structures of the B_{12}-independent glycerol dehydratase and 4-hydroxyphenylacetate decarboxylase show the presence of aspartic and glutamic acid residues in the active site. These residues have been computationally shown to be crucial for the catalysis by both enzymes, acting as proton donors and acceptors.

Manuscript A covers a computational study (electrostatic and cluster model DFT calculations) on the catalytic mechanism of the B_{12}-independent glycerol dehydratase. Manuscript B discusses a crystallographic study, with some support from electrostatic calculations, on the structure of 4-hydroxyphenylacetate decarboxylase. Manuscript C extends this discussion by presenting a combined electrostatic and QC/MM study on the full-enzyme model. In manuscript D, QC/MM calculations are discussed that provide updates to the previously studied mechanism of pyruvate formate-lyase.

4.1 Cleavage of the C−O bond in glycerol

The growing biofuel industry is interested in methods for biochemical degradation of glycerol. This has motivated a study (Manuscript A) on the mechanism of the B_{12}-independent glycerol dehydratase (iGDH), a novel glycyl radical enzyme catalyzing the dehydration of glycerol. The other enzyme capable of converting glycerol is the B_{12}-dependent glycerol dehydratase (GDH). Both enzymes employ radical-mediated mechanisms. For radical storage and activation of the substrate, however, GDH uses the adenosylcobalamin cofactor and iGDH uses a glycyl/thiyl radical diad. Although GDH and iGDH belong to different enzyme families, their active sites are somewhat similar. For example, Glu435, Asp447 and His164 in iGDH are equivalent to Glu170, Asp335 and His143 in GDH. The key structural difference between the two active sites is that the additional histidine in iGDH, His281, is replaced by a potassium ion in GDH.

Enzymes from the B_{12}-dependent family are believed to adopt a common catalytic mechanism that involves an intramolecular 1→2 shift of the functional group of the substrate, for example the middle hydroxyl group of glycerol as in the reaction catalyzed by GDH. It has been established that two effects facilitate the GDH-catalyzed reaction by lowering the energy barrier for the critical 1→2 shift. These effects are the so-called 1) partial protonation of the migrating hydroxyl group by acidic species in the active site, such as histidine, and 2) partial deprotonation of one the terminal groups of glycerol by basic species, such as aspartate or glutamate. However, the migrating/spectator hydroxyl groups never get fully protonated/deprotonated during the 1→2 shift because the deprotonation of neutral histidine would be energetically too unfavorable. The hydroxyl group migration proceeds through a cyclic transition state and gives an unstable 1,2-diol intermediate, which readily collapses into the corresponding aldehyde and water.

To explore the catalytic mechanism of iGDH, the protonation states of titratable residues in the protein were in the first step studied by electrostatic and Monte Carlo calculations on the crystal structure of the substrate-bound form of the enzyme. The calculations indicate that in the lowest energy state at pH=7, both Glu435 and Asp447 are deprotonated and His164 is doubly protonated. The positive charge on His164 is counterbalanced by the negative charges on Glu435 and Asp447. A network of hydrogen bonds inside the active site provides contacts between these residues, the ligand and the rest of the enzyme. The DFT calculations performed on the cluster model of the enzyme show that the dehydration of glycerol by iGDH does not require a complicated intramolecular shift of the middle hydroxyl group, as previously suggested by analogy to the B_{12}-dependent enzyme. Instead, the iGDH-catalyzed reaction seems to involve three elementary steps: 1) activation of glycerol to a radical form by Cys433, 2) dehydration of the substrate-derived radical intermediate, 3) deactivation of the product-related radical intermediate by Cys433. During the second step, a water molecule is released directly from glycerol by full protonation of the leaving hydroxyl group by His164. The dehydration is coupled to a proton transfer from the C1 hydroxyl group of glycerol to Glu435, which generates a C=O double bond of the future aldehyde product. The energetics of all three reaction steps were calculated to be very feasible, i.e. the energy barriers were found to be very low. The barrier for the direct release of water does not exceed 6 kcal/mol. The calculations also suggest the possibility of an alternative reaction mechanism, in which

first His164 and Asp447 exchange a proton and second Asp447 donates the proton to the leaving hydroxyl group of glycerol.

On the basis of the performed calculations, isotope labeling experiments are proposed that could support the theoretical mechanism. The most interesting finding is that the B_{12}-dependent and -independent enzymes adopt fundamentally different mechanisms for the dehydration of glycerol, which is mostly due to a different protonation state of a single histidine in the active site. Moreover, the calculations are able to pinpoint the origins of catalytic efficiency of iGDH. These are: 1) the Cys433 radical that activates the substrate to a reactive radical form, 2) the proton donating/accepting system of His164 (Asp447) and Glu435 that facilitates the release of water from the activated substrate, 3) the hydrogen bond network made up by Asn156, Ser232 and His281 that stabilizes the transition state for the dehydration.

4.2 Cleavage of the C−C bond in 4-hydroxyphenylacetate

Decarboxylation processes are of high significance in biological systems. 4-hydroxyphenylacetate decarboxylase (4Hpad) is a glycyl radical enzyme involved in the fermentative metabolism of tyrosine in clostridia. The enzyme catalyzes the production of *p*-cresol from its substrate. The previously postulated decarboxylation mechanism assumed that the reaction starts from the hydrogen atom transfer between the hydroxyl group of the substrate and the radical cysteine Cys503. However, the recently solved crystal structure of the substrate-bound form of 4Hpad shows an unexpected binding mode of 4-hydroxyphenylacetate in the active site (Manuscript B). Namely, the substrate binds to the enzyme with its carboxyl group close to the radical cysteine Cys503, while the substrate's hydroxyl group is hydrogen-bound to Glu637 at the opposite end of the active site. This finding suggests that the enzyme performs the decarboxylation of 4-hydroxyphenylacetate via a Kolbe-type mechanism. On the other hand, the analysis of the crystal structure alone is insufficient for the determination of the exact enzymatic mechanism. Therefore, electrostatic and quantum chemical/molecular mechanical calculations were used to study the 4Hpad-catalyzed reaction in detail (Manuscript C).

The electrostic calculations indicate that the carboxyl group of the bound substrate is deprotonated and the substrate's hydroxyl group is protonated. The negative charge on the carboxyl group is stabilized by a network of hydrogen bonds from the protonated Glu505,

the protein backbone and Ser344. Based on the QC/MM calculations, the activation of the substrate was found to follow a nonstandard reaction scheme, in which the radical cysteine Cys503 abstracts an electron from the substrate and the active site glutamate Glu637 abstracts a proton from the substrate's hydroxyl group. The activation is hence accomplished by the netto abstraction of a hydrogen atom from the substrate and requires that two active site residues are simultaneously involved in this processes. In the next step, the activated substrate readily decarboxylates. The remaining steps of the reaction involve rotation of the Cys503 side-chain, a proton transfer from Glu505 to Cys503 and a hydrogen atom transfer from Cys503 to the product-related intermediate. All reaction steps were found to be energetically very feasible. The highest energy barrier of about 7 kcal/mol was calculated for the decarboxylation step.

4.3 Cleavage of the C−C bond in pyruvate

Pyruvate formate-lyase catalyzes the conversion of pyruvate and coenzyme A (CoA) into formate and acetyl-CoA, which is a key reaction in the carbon metabolism of various anaerobic bacteria. The reaction catalyzed by PFL has been computationally studied before but only using small-molecule or cluster models of the active site. To better understand the role of protein environment in the catalyzed reaction, not accounted for by the previously used computational methods, I have performed quantum chemical/molecular mechanical calculations on the full-enzyme model (Manuscript D). In general, the present QC/MM calculations support the mechanism derived from the earlier computational studies but introduce some novelties.

The crucial step in the catalysis by PFL is a thiyl attack of the active site cysteine Cys418 on the carbonyl moiety of pyruvate. This step leads to the formation of a protein-bound intermediate. However, the exact nature of this intermediate remained elusive. The study on small-molecule models using deprotonated pyruvate suggests that the thiyl attack of Cys418 on pyruvate's carbonyl moiety and release of the formyl radical are coupled. The studies involving pyruvate in a protonated form or stabilized by two active site arginines, Arg176 and Arg435, show the occurrence of an unstable oxy-radical tetrahedral intermediate of pyruvate bound to Cys418. On the other hand, the present QC/MM calculations indicate that first the energy barrier for the formation of the tetrahedral intermediate is as low as 5 kcal/mol (twice lower than for the earlier models) and second

the intermediate is a relatively stable species on the potential energy surface.

The other new observation is that the collapse of the tetrahedral intermediate induces a conformational change of the Cys419 side-chain in the active site. With the extension of the C−C bond of pyruvate, the Cys419 side-chain rotates towards the leaving formyl radical. After the collapse of the tetrahedral intermediate, Cys419 is perfectly positioned for the hydrogen atom abstraction by the formyl radical. Thus, Cys419 rather than CoA is responsible for quenching of the formyl radical. Moreover, this may suggest that the release of formate precedes binding of the CoA molecule to the active site.

4.4 List of published and submitted manuscripts

A. Mikolaj Feliks, G. Matthias Ullmann **"Glycerol Dehydratation by the B_{12}-independent Enzyme May Not Involve the Migration of a Hydroxyl Group – A Computational Study"** *J. Phys. Chem. B 116*, 7076–7087 (2012)

The project was initiated by myself. All electrostatic and quantum chemical (DFT) calculations for modeling of the enzymatic reaction path described in the paper were designed and performed by myself. The results from the calculations were analyzed and interpreted by myself with support from Matthias Ullmann. The manuscript was written by myself and corrected to the final version by Matthias Ullmann.

B. Berta M. Martins, Martin Blaser, Mikolaj Feliks, G. Matthias Ullmann, Wolfgang Buckel, Thorsten Selmer **"Structural Basis for a Kolbe-Type Decarboxylation Catalyzed by a Glycyl Radical Enzyme"** *J. Am. Chem. Soc. 133*, 14666–14674 (2011)

The experimental part of the paper involving the elucidation of the protein crystal structure was done by Berta Martins and Martin Blaser. The electrostatic calculations and the analysis of protonation states of titratable residues were performed on top of the crystal structure by myself with support from Matthias Ullmann. These calculations were used to formulate the putative catalytic mechanism for the enzyme. The manuscript was written by Berta Martins. All authors were involved in discussions and also contributed to the final version of the manuscript.

C. Mikolaj Feliks, Berta M. Martins, G. Matthias Ullmann **"Catalytic Mechanism of the Glycyl Radical Enzyme 4-Hydroxyphenylacetate Decarboxylase from Continuum Electrostatic and QC/MM Calculations"** *J. Am. Chem. Soc. 135*, 14574–14585 (2013)

All electrostatic and quantum chemical/molecular mechanical (DFT) calculations discussed in the paper were performed by myself. The calculations were designed by myself and Matthias Ullmann. The software for studying the enzymatic mechanism was written by myself using the Python programming language and the framework of the pDynamo library by Martin Field. The results from the calculations (reaction paths, catalytic effects) were analyzed by myself together with Matthias Ullmann. The manuscript was written by myself. Matthias Ullmann and Berta Martins contributed to discussions and helped prepare the final version of the manuscript.

D. Mikolaj Feliks, G. Matthias Ullmann **"New Insights into the Catalytic Mechanism of Pyruvate Formate-Lyase from QC/MM Calculations"**
(Manuscript to be submitted)

The project was initiated by myself. All QC(DFT)/MM calculations described in the paper were carried out by myself. The calculations were peformed with custom Python scripts written using the pDynamo software library. The results were analyzed by myself together with Matthias Ullmann. The manuscript was written by myself and modified by Matthias Ullmann.

Bibliography

[1] J. Stubbe and W. A. van der Donk (1998): Protein Radicals in Enzyme Catalysis. *Chem. Rev. 98*, 705–76

[2] P. Frey (2001): Radical mechanisms of enzymatic catalysis. *Annu. Rev. Biochem. 70*, 121–148

[3] P. A. Frey, A. D. Hegeman and G. H. Reed (2006): Free Radical Mechanisms in Enzymology. *Chem. Rev. 106*, 3302–3316

[4] W. Buckel and B. T. Golding (2006): Radical enzymes in anaerobes. *Annu. Rev. Microbiol. 60*, 27–49

[5] S. S. Yazdani and R. Gonzalez (2007): Anaerobic fermentation of glycerol: a path to economic viability for the biofuels industry. *Curr. Opin. Biotechnol. 18*, 213–219

[6] J. M. Clomburg and R. Gonzalez (2013): Anaerobic fermentation of glycerol: a platform for renewable fuels and chemicals. *Trends Biotechnol. 31*, 20–28

[7] F. Himo and P. Siegbahn (2003): Quantum chemical studies of radical-containing enzymes. *Chem. Rev. 103*, 2421–2456

[8] G. M. Sandala, D. M. Smith and L. Radow (2010): Modeling the Reactions Catalyzed by Coenzyme B_{12}-Dependent Enzymes. *Accounts Chem. Res. 43*, 642–651

[9] M. Field (2002): Simulating enzyme reactions: Challenges and perspectives. *J. Comput. Chem. 23*, 48–58

[10] A. Warshel (2003): Computer simulations of enzyme catalysis: Methods, progress, and insights. *Annu. Rev. Biophys. Biomolec. Struct. 32*, 425–443

[11] M. J. Ramos and P. A. Fernandes (2008): Computational enzymatic catalysis. *Accounts Chem. Res. 41*, 689–698

[12] S. F. Sousa, P. A. Fernandes and M. J. Ramos (2012): Computational enzymatic catalysis – clarifying enzymatic mechanisms with the help of computers. *Phys. Chem. Chem. Phys. 14*, 12431–12441

[13] G. M. Ullmann and E. W. Knapp (1999): Electrostatic models for computing protonation and redox equilibria in proteins. *Eur. Biophys. J. 28*, 533–551

[14] E. Bombarda, T. Becker. and G. M. Ullmann (2006): The Influence of the Membrane Potential on the Protonation of Bacteriorhodopsin: Insights from Electrostatic Calculations into the Regulation of Proton Pumping. *128*, 12129–12139

[15] E. Bombarda and G. M. Ullmann (2010): pH-Dependent pK(a) Values in Proteins-A Theoretical Analysis of Protonation Energies with Practical Consequences for Enzymatic Reactions. *J. Phys. Chem. B 114*, 1994–2003

[16] F. Himo (2006): Quantum chemical modeling of enzyme active sites and reaction mechanisms. *Theor. Chem. Acc. 116*, 232–240

[17] P. E. M. Siegbahn and F. Himo (2009): Recent developments of the quantum chemical cluster approach for modeling enzyme reactions. *J. Biol. Inorg. Chem. 14*, 643–651

[18] P. E. M. Siegbahn and F. Himo (2011): The quantum chemical cluster approach for modeling enzyme reactions. *Wiley Interdiscip. Rev.-Comput. Mol. Sci. 1*, 323–336

[19] H. M. Senn and W. Thiel (2007): QM/MM methods for biological systems. *Top. Curr. Chem. 268*, 173–290

[20] H. M. Senn and W. Thiel (2007): QM/MM studies of enzymes. *Curr. Opin. Chem. Biol. 11*, 182–187

[21] H. M. Senn and W. Thiel (2009): QM/MM Methods for Biomolecular Systems. *Angew. Chem.-Int. Edit. 48*, 1198–1229

[22] O. Acevedo and W. L. Jorgensen (2010): Advances in Quantum and Molecular Mechanical (QM/MM) Simulations for Organic and Enzymatic Reactions. *Accounts Chem. Res. 43*, 142–151

[23] M. Fontecave (1998): Ribonucleotide reductases and radical reactions. *Cell. Mol. Life Sci. 54*, 684–695

[24] H. Eklund and M. Fontecave (1999): Glycyl radical enzymes: a conservative structural basis for radicals. *Struct. Fold. Des. 7*, R257–R262

[25] C. Raynaud, P. Sarcabal, I. Meynial-Salles, C. Croux and P. Soucaille (2003): Molecular characterization of the 1,3-propanediol (1,3-PD) operon of Clostridium butyricum. *Proc. Natl. Acad. Sci. USA 100*, 5010–5015

[26] J. O'Brien, C. Raynaud, C. Croux, L. Girbal, P. Soucaille and W. Lanzilotta (2004): Insight into the mechanism of the B_{12}-independent glycerol dehydratase from Clostridium butyricum: Preliminary biochemical and structural characterization. *Biochemistry 43*, 4635–4645

[27] W. Buckel (2009): Radical and Electron Recycling in Catalysis. *Angew. Chem.-Int. Edit. 48*, 6779–6787

[28] Y. Liu, A. A. Gallo, J. Florian, Y.-S. Liu, S. Mora and W. Xu (2010): QM/MM (ONIOM) Study of Glycerol Binding and Hydrogen Abstraction by the Coenzyme B_{12}-Independent Dehydratase. *J. Phys. Chem. B 114*, 5497–5502

[29] M. Feliks and G. M. Ullmann (2012): Glycerol Dehydratation by the B_{12}-Independent Enzyme May Not Involve the Migration of a Hydroxyl Group: A Computational Study. *J. Phys. Chem. B 116*, 7076–7087

[30] J. Masuda, N. Shibata, Y. Morimoto, T. Toraya and N. Yasuoka (2000): How a protein generates a catalytic radical from coenzyme B_{12}: X-ray structure of a diol-dehydratase-adeninylpentylcobalamin complex. *Struct. Fold. Des. 8*, 775–788

[31] M. Yamanishi, M. Yunoki, T. Tobimatsu, H. Sato, J. Matsui, A. Dokiya, Y. Iuchi, K. Oe, K. Suto, N. Shibata, Y. Morimoto, N. Yasuoka and T. Toraya (2002): The crystal structure of coenzyme B_{12}-dependent glycerol dehydratase in complex with cobalamin and propane-1,2-diol. *Eur. J. Biochem. 269*, 4484–4494

[32] T. Kamachi, T. Toraya and K. Yoshizawa (2007): Computational mutation analysis of hydrogen abstraction and radical rearrangement steps in the catalysis of coenzyme B_{12}-dependent diol dehydratase. *Chem.-Eur. J. 13*, 7864–7873

[33] K. Kinoshita, M. Kawata, K.-i. Ogura, A. Yamasaki, T. Watanabe, N. Komoto, N. Hieda, M. Yamanishi, T. Tobimatsu and T. Toraya (2008): Histidine-alpha 143 assists 1,2-hydroxyl group migration and protects radical intermediates in coenzyme B_{12}-dependent diol dehydratase. *Biochemistry 47*, 3162–3173

[34] G. M. Sandala, B. Kovacevic, D. Baric, D. M. Smith and L. Radom (2009): On the Reaction of Glycerol Dehydratase with But-3-ene-1,2-diol. *Chem.-Eur. J. 15*, 4865–4873

[35] T. Selmer and P. Andrei (2001): p-Hydroxyphenylacetate decarboxylase from Clostridium difficile – A novel glycyl radical enzyme catalysing the formation of p-cresol. *Eur. J. Biochem. 268*, 1363–1372

[36] L. Yu, M. Blaser, P. I. Andrei, A. J. Pierik and T. Selmer (2006): 4-hydroxyphenylacetate decarboxylases: Properties of a novel subclass of glycyl radical enzyme systems. *Biochemistry 45*, 9584–9592

[37] J. Knappe, F. Neugebauer, H. Blaschkowski and M. Ganzler (1984): Post-translational activation introduces a free radical into pyruvate formate-lyase. *Proc. Nat. Acad. Sci. USA 81*, 1332–1335

[38] A. Becker, K. Wolf, W. Kabsch, J. Knappe and W. A. Schultz, S (1999): Structure and mechanism of the glycyl radical enzyme pyruvate formate-lyase. *Nat. Struc. Biol. 6*, 969–975

[39] M. Garcia-Viloca, J. Gao, M. Karplus and D. G. Truhlar (2004): How Enzymes Work: Analysis by Modern Rate Theory and Computer Simulations. *Science 303*, 186–195

[40] I. H. Williams (1993): Interplay of theory and experiment in the determination of transition-state structure. *Chem. Soc. Rev. 22*, 277–283

[41] G. Naray-Szabo (2000): Enzyme mechanisms: interplay of theory and experiment. *J. Mol. Struct. (THEOCHEM) 500*, 157–167

[42] A. Mulholland, G. Grant and W. Richards (1993): Computer modeling of enzyme catalyzed reaction mechanisms. *Protein Eng. 6*, 133–147

[43] A. Warshel and J. Florian (1998): Computer simulations of enzyme catalysis: Finding out what has been optimized by evolution *95*, 5950–5955

[44] T. C. Bruice and K. Kahn (2000): Computational enzymology. *Curr. Opin. Chem. Biol. 4*, 540–544

[45] S. Marti, J. A. Maite Roca, V. Moliner, E. Silla, I. Tunon and J. Bertran (2004): Theoretical insights in enzyme catalysis. *Chem. Soc. Rev. 33*, 98–107

[46] A. Mulholland (2005): Modelling enzyme reaction mechanisms, specificity and catalysis. *Drug Discov. Today 10*, 1393–1402

[47] M. W. van der Kamp and A. J. Mulholland (2008): Computational enzymology: insight into biological catalysts from modelling. *Nat. Prod. Rep. 25*, 1001–1014

[48] R. Lonsdale, K. E. Ranaghan and A. J. Mulholland (2010): Computational enzymology. *Chem. Commun. 46*, 2354–2372

[49] R. Lonsdale, J. N. Harvey and A. J. Mulholland (2012): A practical guide to modelling enzyme-catalysed reactions. *Chem. Soc. Rev. 41*, 3025–3038

[50] A. J. Cohen, P. Mori-Sanchez and W. Yang (2012): Challenges for Density Functional Theory. *Chem. Rev. 112*, 289–320

[51] E. Fisher (1894): Einfluss der Configuration auf die Wirkung der Enzyme. *Ber. Deutsch. Chem. Ges. 27*, 2985–2993

[52] L. Pauling (1946): Molecular architecture and biological reactions. *Chem. Eng. News 24*, 1375–1377

[53] L. Pauling (1948): Nature of forces between large molecules of biological interest. *Nature 161*, 707–709

[54] A. Warshel (1998): Electrostatic Origin of the Catalytic Power of Enzymes and the Role of Preorganized Active Sites. *J. Biol. Chem. 273*, 27035–27038

[55] A. Warshel, P. K. Sharma, M. Kato, Y. Xiang, H. Liu and M. H. M. Olsson (2006): Electrostatic Basis for Enzyme Catalysis. *Chem. Rev. 106*, 3210–3235

[56] W. Cleland and M. Kreevoy (1994): Low-barrier hydrogen bonds and enzymatic catalysis. *Science 264*, 1887–1890

[57] P. Kollman, B. Kuhn, O. Donini, M. Perakyla, R. Stanton and D. Bakowies (2001): Elucidating the nature of enzyme catalysis utilizing a new twist on an old methodology: Quantum mechanical – Free energy calculations on chemical reactions in enzymes and in aqueous solution. *Accounts Chem. Res. 34*, 72–79

[58] T. C. Bruice (2002): A view at the millennium: The efficiency of enzymatic catalysis. *Acc. Chem. Rev. 35*, 139–148

[59] X. Zhang and K. Houk (2005): Why enzymes are proficient catalysts: Beyond the Pauling paradigm. *Accounts Chem. Res. 38*, 379–385

[60] D. Antoniou, J. Basner, S. Nunez and S. D. Schwartz (2006): Computational and theoretical methods to explore the relation between enzyme dynamics and catalysis. *Chem. Rev. 106*, 3170–3187

[61] J. Retey (1990): Enzymatic-reaction selectivity by negative catalysis or how do enzymes deal with highly reactive intermediates? *Angew. Chem.-Int. Edit. Engl. 29*, 355–361

[62] B. R. Brooks, R. D. Bruccoleri, B. O. Olafson, D. J. States, S. Swaminathan and M. Karplus (1983): CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem. 4*, 187–217

[63] W. Cornell, P. Cieplak, C. Bayly, I. Gould, K. Merz, D. Ferguson, D. Spellmeyer, T. Fox, J. Caldwell and P. Kollman (1995): A 2nd generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc. 117*, 5179–5197

[64] P. Hohenberg and W. Kohn (1964): Inhomogeneous electron gas. *Phys. Rev. B 136*, B864

[65] E. Runge and E. Gross (1984): Density functional theory for time-dependent systems.

Phys. Rev. Lett. 52, 997–1000

[66] A. Becke (1988): Correlation energy of an inhomogeneous electron gas – a coordinate space model. *J. Chem. Phys. 88*, 1053–1062

[67] S. Vosko, L. Wilk and M. Nusair (1980): Accurate spin-dependent electron liquid correlation energies for local spin-density calculations – a critical analysis. *Can. J. Phys. 58*, 1200–1211

[68] C. Lee, W. Yang and R. Parr (1988): Development of the Colle-Salvetti correlation energy formula into a functional of the electron density. *Phys. Rev. B 37*, 785–789

[69] Y. Zhao and D. G. Truhlar (2008): Density functionals with broad applicability in chemistry. *Accounts Chem. Res. 41*, 157–167

[70] D. Bashford and M. Karplus (1991): Multiple-site titration curves of proteins – an analysis of exact and approximate methods for their calculation. *J. Phys. Chem. 95*, 9556–9561

[71] F. Fogolari, A. Brigo and H. Molinari (2002): The Poisson-Boltzmann equation for biomolecular electrostatics: a tool for structural biology. *J. Mol. Recognit. 15*, 377–392

[72] A. Klingen (2006): Structure-based theoretical characterization of the redox-dependent titration behaviour of cytochrome bc_1. Ph.D. thesis, University of Bayreuth

[73] T. Essigke (2008): A Continuum Electrostatic Approach for Calculating the Binding Energetics of Multiple Ligands. Ph.D. thesis, University of Bayreuth

[74] R. T. Ullmann and G. M. Ullmann (2012): GMCT : A Monte Carlo Simulation package for macromolecular receptors. *J. Comp. Chem 33*, 887–900

[75] K. Condic-Jurkic, V. T. Perchyonok, H. Zipse and D. M. Smith (2008): On the modeling of arginine-bound carboxylates: A case study with Pyruvate Formate-Lyase. *J. Comput. Chem. 29*, 2425–2433

[76] K. Condic-Jurkic, H. Zipse and D. M. Smith (2010): A Compound QM/MM Procedure: Comparative Performance on a Pyruvate Formate-Lyase Model System. *J. Comput. Chem. 31*, 1024–1035

[77] L. Noodleman, T. Lovell, W. Han, J. Li and F. Himo (2004): Quantum chemical studies of intermediates and reaction pathways in selected enzymes and catalytic synthetic systems. *Chem. Rev. 104*, 459–508

[78] S.-L. Chen, W.-H. Fang and F. Himo (2008): Technical aspects of quantum chemical

modeling of enzymatic reactions: the case of phosphotriesterase. *Theor. Chem. Acc. 120*, 515–522

[79] T. Lengauer and M. Rarey (1996): Computational methods for biomolecular docking. *Curr. Opin. Struct. Biol. 6*, 402–406

[80] M. Cossi, N. Rega, G. Scalmani and V. Barone (2003): Energies, structures, and electronic properties of molecules in solution with the CPCM solvation model. *J. Comput. Chem. 24*, 669–681

[81] J. Tomasi, B. Mennucci and R. Cammi (2005): Quantum Mechanical Continuum Solvation Models. *Chem. Rev. 105*, 2999–3094

[82] R.-Z. Liao, J.-G. Yu and F. Himo (2011): Quantum Chemical Modeling of Enzymatic Reactions: The Case of Decarboxylation. *J. Chem. Theory Comput. 7*, 1494–1501

[83] G. d. M. Seabra, R. C. Walker and A. E. Roitberg (2009): Are Current Semiempirical Methods Better Than Force Fields? A Study from the Thermodynamics Perspective. *J. Phys. Chem. A 113*, 11938–11948

[84] D. Bakowies and W. Thiel (1996): Hybrid models for combined quantum mechanical and molecular mechanical approaches. *J. Phys. Chem. 100*, 10580–10594

[85] A. Warshel and M. Levitt (1976): Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol. 103*, 227–249

[86] F. Maseras and K. Morokuma (1995): IMOMM: A new integrated ab initio + molecular mechanics geometry optimization scheme of equilibrium structures and transition states. *J. Comput. Chem. 16*, 1170–1179

[87] V. Kairys and J. Jensen (2000): QM/MM boundaries across covalent bonds: A frozen localized molecular orbital-based approach for the effective fragment potential method. *J. Phys. Chem. A 104*, 6656–6665

[88] D. Sheppard, R. Terrell and G. Henkelman (2008): Optimization methods for finding minimum energy paths. *J. Chem. Phys. 128*

[89] H. B. Schlegel (2011): Geometry optimization. *Wiley Interdiscip. Rev.-Comput. Mol. Sci. 1*, 790–809

[90] I. F. Galvan and M. J. Field (2008): Improving the efficiency of the NEB reaction path finding algorithm. *J. Comput. Chem. 29*, 139–143

[91] A. Aleksandrov and M. Field (2012): A hybrid elastic band string algorithm for

studies of enzymatic reactions. *Phys. Chem. Chem. Phys. 14*, 12544–12553

[92] W. Buckel and B. Golding (1998): Radical species in the catalytic pathways of enzymes from anaerobes. *Fems Microbiol. Rev. 22*, 523–541

[93] W. Buckel, J. Zhang, P. Friedrich, A. Parthasarathy, H. Li, I. Djurdjevic, H. Dobbek and B. M. Martins (2012): Enzyme catalyzed radical dehydrations of hydroxy acids. *Biochem. Biophys. Acta 1824*, 1278–1290

[94] T. Toraya (2003): Radical catalysis in coenzyme B_{12}-dependent isomerization (eliminating) reactions. *Chem. Rev. 103*, 2095–2127

[95] K. Jensen and U. Ryde (2005): How the Co−C bond is cleaved in coenzyme B_{12} enzymes: A theoretical study. *J. Am. Chem. Soc. 127*, 9117–9128

[96] P. Kozlowski, T. Andruniow, A. Jarzecki, M. Zgierski and T. Spiro (2006): DFT analysis of Co-alkyl and Co-adenosyl vibrational modes in B_{12}-cofactors. *Inorg. Chem. 45*, 5585–5590

[97] J. Guo, Y. Luo and F. Himo (2003): DNA repair by spore photoproduct lyase: A density functional theory study. *J. Phys. Chem. B 107*, 11188–11192

[98] C. Desnous, D. Guillaume and P. Clivio (2010): Spore Photoproduct: A Key to Bacterial Eternal Life. *Chem. Rev. 110*, 1213–1232

[99] A. Benjdia, K. Heil, T. R. M. Barends, T. Carell and I. Schlichting (2012): Structural insights into recognition and repair of UV-DNA damage by Spore Photoproduct Lyase, a radical SAM enzyme. *Nucleic Acids Res. 40*, 9308–9318

[100] T. Selmer, A. Pierik and J. Heider (2005): New glycyl radical enzymes catalysing key metabolic steps in anaerobic bacteria. *Biol. Chem. 386*, 981–988

[101] S. Craciun and E. P. Balskus (2012): Microbial conversion of choline to trimethylamine requires a glycyl radical enzyme. *Proc. Nat. Acad. Sci. USA 109*, 21307–21312

[102] C. Thibodeaux and W. A. van der Donk (2012): Converging on a mechanism for choline degradation. *Proc. Nat. Acad. Sci. USA 109*, 21184–21185

[103] V. Unkrig, F. Neugebauer and J. Knappe (1989): The free-radical of pyruvate formate-lyase – characterization by EPR spectroscopy and involvement in catalysis as studied with the substrate-analog hypophosphite. *Eur. J. Biochem. 184*, 723–728

[104] A. Wagner, M. Frey, F. Neugebauer, W. Schafer and J. Knappe (1992): The free-radical in pyruvate formate-lyase is located on glycine-734. *Proc. Natl. Acad. Sci. USA 89*, 996–1000

[105] F. Himo (2000): Stability of protein-bound glycyl radical: a density functional theory study. *Chem. Phys. Lett. 328*, 270–276

[106] J. Hioe, G. Savasci, H. Brand and H. Zipse (2011): The Stability of C-alpha Peptide Radicals: Why Glycyl Radical Enzymes? *Chem.-Eur. J. 17*, 3781–3789

[107] J. Guo and F. Himo (2004): Catalytic mechanism of pyruvate formate-lyase revisited. *J. Phys. Chem. B 108*, 15347–15354

[108] J. Gauld and L. Eriksson (2000): Oxidative degradation of pyruvate formate-lyase. *J. Am. Chem. Soc. 122*, 2035–2040

[109] L. Lehtio and A. Goldman (2004): The pyruvate formate lyase family: sequences, structures and activation. *Protein Eng. Des. Sel. 17*, 545–552

[110] L. Lehtio, J. Grossmann, B. Kokona, R. Fairman and A. Goldman (2006): Crystal structure of a glycyl radical enzyme from Archaeoglobus fulgidus. *J. Mol. Biol. 357*, 221–235

[111] C. Krieger, W. Roseboom, S. Albracht and A. Spormann (2001): A stable organic free radical in anaerobic benzylsuccinate synthase of Azoarcus sp strain T. *J. Biol. Chem. 276*, 12924–12927

[112] M. Hilberg, A. J. Pierik, E. Bill, T. Friedrich, M.-L. Lippert and J. Heider (2012): Identification of FeS clusters in the glycyl-radical enzyme benzylsuccinate synthase via EPR and Mossbauer spectroscopy. *J. Biol. Inorg. Chem. 17*, 49–56

[113] F. Himo (2002): Catalytic mechanism of benzylsuccinate synthase, a theoretical study. *J. Phys. Chem. B 106*, 7688–7692

[114] T. Li, L. Huo, C. Pulley and A. Liu (2012): Decarboxylation mechanisms in biological systems. *Bioorganic Chem. 43*, 2–14

[115] L. Dari and H. Barker (1985): P-cresol formation by cell-free extracts of Clostridium difficile. *Arch. Microbiol. 143*, 311–312

[116] B. M. Martins, M. Blaser, M. Feliks, G. M. Ullmann, W. Buckel and T. Selmer (2011): Structural Basis for a Kolbe-Type Decarboxylation Catalyzed by a Glycyl Radical Enzyme. *J. Am. Chem. Soc. 133*, 14666–14674

[117] P. Andrei, A. Pierik, S. Zauner, L. Andrei-Selmer and T. Selmer (2004): Subunit composition of the glycyl radical enzyme p-hydroxyphenylacetate decarboxylase – A small subunit, HpdC, is essential for catalytic activity. *Eur. J. Biochem. 271*, 2225–2230

[118] A. Vijh and B. Conway (1967): Electrode kinetic aspects of Kolbe reaction. *Chem. Rev. 67*, 623–664

[119] M. Feliks, B. Martins and G. M. Ullmann (2013): Catalytic Mechanism of the Glycyl Radical Enzyme 4-Hydroxyphenylacetate Decarboxylase from Continuum Electrostatic and QC/MM Calculations. *J. Am. Chem. Soc. 135*, 14574–14585

[120] W. Deckwer (1995): Microbial conversion of glycerol to 1,3-propanediol. *Fems Microbiol. Rev. 16*, 143–149

[121] M. Gonzalez-Pajuelo, I. Meynial-Salles, F. Mendes, P. Soucaille and I. Vasconcelos (2006): Microbial conversion of glycerol to 1,3-propanediol: Physiological comparison of a natural producer, Clostridium butyricum VPI 3266, and an engineered strain, Clostridium acetobutylicum DG1(pSPD5). *Appl. Environ. Microbiol. 72*, 96–101

[122] D. Logan, J. Andersson, B. Sjoberg and P. Nordlund (1999): A glycyl radical site in the crystal structure of a class III ribonucleotide reductase. *Science 283*, 1499–1504

[123] P. Nordlund and P. Reichard (2006): Ribonucleotide reductases. *Annu. Rev. Biochem. 75*, 681–706

[124] K. Cho, F. Himo, A. Graslund and P. Siegbahn (2001): The substrate reaction mechanism of class III anaerobic ribonucleotide reductase. *J. Phys. Chem. B 105*, 6445–6452

[125] K. Cho, V. Pelmenschikov, A. Graslund and P. Siegbahn (2004): Density functional calculations on class III ribonucleotide reductase: Substrate reaction mechanism with two formates. *J. Phys. Chem. B 108*, 2056–2065

[126] A. Becker and W. Kabsch (2002): X-ray structure of pyruvate formate-lyase in complex with pyruvate and CoA – How the enzyme uses the Cys-418 thiyl radical for pyruvate cleavage. *J. Biol. Chem. 277*, 40036–40042

[127] F. Himo and L. Eriksson (1998): Catalytic mechanism of pyruvate formate-lyase (PFL). A theoretical study. *J. Am. Chem. Soc. 120*, 11449–11455

[128] M. Lucas, P. Fernandes, L. Eriksson and M. Ramos (2003): Pyruvate formate-lyase: A new perspective. *J. Phys. Chem. B 107*, 5751–5757

[129] M. J. Field (2008): The pDynamo program for molecular simulations using hybrid quantum chemical and molecular mechanical potentials. *J. Chem. Theory Comput. 4*, 1151–1161

Manuscript A

"Glycerol Dehydratation by the B_{12}-Independent Enzyme May Not Involve the Migration of a Hydroxyl Group: A Computational Study"

Mikolaj Feliks, G. Matthias Ullmann

J. Phys. Chem. B 116, 7076–7087 (2012)

Permanent link: http://pubs.acs.org/doi/abs/10.1021/jp301165b

Abstract

Continuum electrostatic and density functional calculations have been combined to explore the catalytic mechanism of the B_{12}-independent glycerol dehydratase. In sharp contrast to the previously suggested mechanism, the calculations indicate that the release of water from glycerol is accomplished without the intermediacy of a geminal diol species. Instead, the enzyme employs two active site residues, histidine and glutamate, as a proton donating/accepting system. The glutamate accepts a proton from the terminal hydroxyl group of glycerol, whereas the histidine donates a proton to the leaving middle hydroxyl group of glycerol, forming a water molecule. The calculations also show a key role of the active site residues in stabilization of the transition state of the water release step.

Manuscript B

"Structural Basis for a Kolbe-Type Decarboxylation Catalyzed by a Glycyl Radical Enzyme"

Berta M. Martins, Martin Blaser, Mikolaj Feliks,
G. Matthias Ullmann, Wolfgang Buckel, Thorsten Selmer

J. Am. Chem. Soc. 133, 14666–14674 (2011)

Permanent link: http://pubs.acs.org/doi/abs/10.1021/ja203344x

Abstract

In the present study, we report on the high-resolution crystal structure of 4-hydroxyphenylacetate decarboxylase, a novel glycyl radical enzyme proposed to catalyze the last step of tyrosine fermentation in clostridia. The structure shows an unexpected binding mode of 4-hydroxyphenylacetate to the enzyme active site. In this mode, the carboxyl group of the substrate is in close contact to the thiyl radical and the phenolic group is hydrogen-bound to the glutamate on the opposite end of the active site. A Kolbe-type decarboxylation mechanism is suggested for the formation of the *p*-cresol product. This mechanism is supported by our continuum electrostatic calculations.

Manuscript C

"Catalytic Mechanism of the Glycyl Radical Enzyme 4-Hydroxyphenylacetate Decarboxylase from Continuum Electrostatic and QC/MM Calculations"

Mikolaj Feliks, Berta M. Martins, G. Matthias Ullmann

J. Am. Chem. Soc. 135, 14574–14585 (2013)

Permanent link: http://pubs.acs.org/doi/abs/10.1021/ja402379q

Abstract

The catalytic mechanism of 4-hydroxyphenylacetate decarboxylase has been studied based on continuum electrostatic and QC/MM calculations. The calculations suggest an unconventional activating mode of 4-hydroxyphenylacetate. The substrate is activated to a radical form by two simultaneous transfers, first of an electron and second of a proton. The electron is transferred from the substrate to the thiyl radical and the proton is abstracted from the phenolic group of the substrate by the active site glutamate. The activation generates a radical anion intermediate. The release of CO_2 from the intermediate is coupled to a proton back-transfer from the glutamate to the phenolic group. The mechanism based on the calculations corroborates previous experiments showing that the hydroxyl group in the *p*-position of the substrate is crucial for the catalysis.

New Insights into the Catalytic Mechanism of Pyruvate Formate-Lyase from QC/MM Calculations

Mikolaj Feliks,[1] G. Matthias Ullmann[1,†]

[1)] Computational Biochemistry Group, University of Bayreuth, Universitätsstr. 30, BGI, 95447 Bayreuth, Germany

[†)] to whom correspondence should be addressed;

e-mail: Matthias.Ullmann@uni-bayreuth.de

January 29, 2014

Abstract

We have revisited, by performing quantum chemical/molecular mechanical calculations on the full-enzyme model, the first part of the enzymatic reaction catalyzed by the glycyl radical enzyme puruvate formate-lyase. Based on our model, we propose several modifications to the previously studied mechanism, which now agrees well with experimental data. First, we present an energy profile which can better explain the reversibilty of the PFL-catalyzed reaction. Second, the calculations suggest that the thiyl attack on pyruvate results in the formation of the protein-bound tetrahedral radical intermediate. The occurance of this intermediate on the reaction path was often questioned in the previous studies on PFL. In our study, the barrier for the formation step was calculated to be as low as 5 kcal/mol. Moreover, the intermediate appears to be quite stable thanks to the stabilizing interactions from the enzyme. Third, the barrier for the release of the formyl radical was found to be 9 kcal/mol. Fourth, from the calculated reaction path, we provide evidence that not CoA but rather Cys419 is responsible for the quenching of the formyl radical. That is, the function of Cys419 is not limited to the hydrogen atom relay between Gly734 and Cys418, as thought before.

Keywords: glycyl radical enzyme, enzymatic reaction mechanism, quantum chemical calculation, continuum electrostatics, proton transfer

Introduction

Radical enzymes are interesting because of their ability to catalyze transformations of generally unreactive compounds.[1–4] There exist many enzymes that employ a radical-based reaction mechanism but only a few glycyl radical enzymes (GRE) have been identified to date.[5–8] The first glycyl radical enzyme to be discovered was pyruvate formate-lyase[9] (PFL), which catalyzes the reversible conversion of pyruvate and CoA into formate and acetyl-CoA. This essential reaction is part of the anaerobic glucose metabolism in various bacteria. Like the other enzymes from the glycyl radical family, PFL is activated to a reactive radical form by a dedicated *S*-adenosylmethionine-dependent activating enzyme. During the activation, the radical is introduced into PFL by the homolytic cleavage of the C–H bond at Gly734. Upon binding of pyruvate, the radical moves from Gly734 to Cys419 in the active site, where it can initiate substrate reactions. This activation mechanism is thought to be a common step in the catalysis by all members of the GRE family. However, the consecutive reaction steps may be very different depending on the particular enzyme as, for example, shown in the recent mechanistic studies on the B_{12}-independent glycerol dehydratase[10–12] (iGDH) and 4-hydroxyphenylacetate decarboxylase[13] (4-Hpad) as well as the earlier studies on benzylsuccinate synthase[14] (BSS) and anaerobic ribonucleotide reductase[15] (ARNR).

Since the discovery of PFL in 1985, the enzyme has been extensively studied, both experimentally and computationally.[15–18] Although the catalytic mechanism of PFL is now believed to be well understood, questions remain concerning some of the mechanistic details of the catalyzed reaction. For example, it is not clear which of CoA or Cys419 is responsible for quenching of the formyl radical, i.e. whether the binding of CoA to the active site preceeds the release of formate. Revisiting the mechanism of PFL in the light of the recent studies on iGDH and 4-Hpad may in general provide better understanding of chemistry involved in the catalysis by glycyl radical enzymes.

The currently discussed catalytic mechanism of PFL, which was originally proposed

by Knappe and co-workers,[19] involves four reaction steps. After the substrate binding, the radical is transferred from Cys419 to Cys418. The presence of two cysteine residues in the active site of PFL that actively participate in the catalysis is unique compared to the other members of the GRE family. In the next step, the thiyl radical on Cys418 attacks the carbonyl carbon atom of pyruvate, which leads to the formation of a S–C bond. The so-generated enzyme-bound radical intermediate is unstable and collapses into the acylated Cys418 and a formyl radical. The formyl radical abstracts a hydrogen atom from Cys419, which gives formate and regenerates the radical on Cys419. Alternatively, the formyl radical may abstract a hydrogen atom from CoA that binds in the active site during the second stage of the reaction. The reaction is completed by the transfer of the acetate moiety from Cys418 to CoA.

The PFL-catalyzed reaction has been computationally studied three times before. In their work on PFL, Himo and co-workers calculated the mechanism for a simple gas-phase model of the active site.[16] The X-ray structure of the enzyme was not available at the time of their study. They used a model of protonated pyruvate. The rationale for using neutral pyruvate was that charge separation in the protein environment is usually very small. Following similar methodology, Lucas and co-workers recalculated the mechanism of PFL using a deprotonated model of pyruvate.[17] They found that the reaction steps involving the attack of the Cys418 thiyl radical on pyruvate and the release of the formyl radical anion are concerted.

Clearly, studying isolated reactants in vacuum is the simplest approach to the modeling of enzymatic catalysis. Although this method can provide some insights into the intrinsic chemistry of the reaction, it cannot account for the precise catalytic mechanism, because the protein environment is not included in the model. The lack of protein surrounding allows reactants to move freely during the reaction, which is not possible inside the enzyme active site. In their second study on PFL,[18] Himo and co-workers employed the so-called cluster model approach[20–22] to better represent the actual conditions of protein

interior. The cluster models consisted of up to 75 atoms and were based on the crystal structure of PFL by Becker and co-workers.[23] To compensate for the missing parts of the enzyme, selected atoms were kept fixed at their X-ray positions during the geometry optimizations. However, the cluster models used by Himo still suffered from too much flexibility. Because of this, the exothermic energies of some of the reaction steps were found to be overestimated, i.e. they were not in line with experimental data. Although the cluster model approach works reasonably well for the identification of key features of the mechanism, it cannot provide full insight into long-range interactions between the active site and the rest of the enzyme. Do these interactions influence the chemistry at the active site? If so, to what extent can they modify the mechanism derived from the cluster model calculations? To address these questions and to overcome the limitations of the previously used methods, one has to go beyond small molecule models of the enzymatic system. Therefore, we have employed hybrid quantum chemical/molecular mechanical (QC/MM) calculations[24–26] to study the mechanism of PFL for the first time in a full-protein model. The QC/MM (ONIOM) method was previously used by Condic-Jurkic and co-workers to study some aspects of the PFL-catalyzed reaction but they still used only small gas-phase models of the active site.[27,28]

The key problem in modeling of enzymatic reactions is the treatment of titratable residues.[29] These residues can adopt different protonation states depending on the pH, interactions with local environment etc. The presence of charged groups close to the active site, even if they do not participate directly in the catalysis, may have a considerable influence on the reaction mechanism.[30] Therefore, prior to the QC/MM exploration of the reaction path, we performed Poisson-Boltzmann electrostatic calculations combined with a Monte Carlo titration to study the protonation behavior of titratable residues in PFL.

We explored the catalytic mechanism of PFL by systematic build-up of the reaction path from potential energy surface scans and geometry optimizations. Reaction profiles

connecting the optimized minima of intermediates were studied by the nudged elastic band method (NEB).[31,32] Transition states were taken as points of the highest energy on the NEB-calculated reaction profiles. Based on our calculations, we are able to present a mechanism that is in general consistent with the previously studied ones but introduces several corrections and novel observations.

Methods

Preparation of the full-enzyme model. The crystal structure of the pyruvate-bound form of PFL by Becker and co-workers[23] (PDB code 1H16) was used to build the complete enzyme model. At the resolution of 1.53 Å, this is the most accurate structure of PFL availabe to date. In addition to puruvate, the crystal structure shows a molecule of CoA bound at the protein surface. The binding mode of CoA allows to study only the first stage of the PFL-catalyzed reaction (see Fig. 1). The initial model of the reactant state was prepared in CHARMM.[33,34] D-treitol, tetraethylene glycol, sodium and magnesium ions as well as CoA were removed from the model. The CHARMM27 force field[35] was used to describe the protein and the waters. The MM-parameters for pyruvate were taken from the force field based on analogy to similar structures.

Setup of the continuum electrostatic model. The protonation states of all titratable residues were set to their standard values at pH 7. CHARMM program was used to add missing hydrogens atoms and to optimize their positions. Crystallographic water molecules were removed from the model. A Poisson-Boltzmann continuum electrostatic model combined with a Monte Carlo titration was used to calculate the protonation probabilities of titratable residues in the protein. The electrostatic calculations were carried out in MEAD[36] with the following parameters. Dielectric constants of $\epsilon_p = 4$ and $\epsilon_s = 80$ were assigned to the interior of the protein and to the solvent, respectively. The ionic strength of solvent was set to $I = 100$ mM and the temperature was set to $T = 300$ K. An ion exclusion layer of 2.0 Å and a solvent probe radius of 1.4 Å were

used to define the volume of the protein. The electrostatic potential was calculated using four grids of 121^3 nodes with focussing steps at resolutions from 2.0 Å to 0.25 Å. The protonation probabilities of titratable residues were calculated as a function of pH using the program GMCT.[37] The pH was varied from 0 to 14 in steps of 0.2 pH-units. For every pH-step, the MC calculation consisted of 100 equilibration scans and 3000 production scans at $T = 300\ K$.

Setup of the QC/MM model. Starting from the crystal structure, missing hydrogen atoms were added to the protein according to the previous electrostatic calculations. Their positions were subsequently geometry optimized in CHARMM. Since the available X-ray structure is of the inactive enzyme, the radical was introduced into the model by deleting the H_γ hydrogen atom of Cys419. The MM-charge of the removed hydrogen was added to that of the remaining S_γ atom. Crystallographic water molecules were preserved in the QC/MM model. The full-enzyme model consisted of 15712 atoms (759 residues, 1267 water molecules, one pyruvate ligand).

The QC/MM calculations were carried out within the framework of the pDynamo software library.[38] ORCA program[39] was employed to handle the quantum chemical part of calculations. In-house Python scripts were used to setup and control the calculations and to analyze the results. The B3LYP[40–43] density functional theory method was used as a QC-potential and the CHARMM27 force field was used as a MM-potential. The B3LYP method has been demonstrated reliable for studying radical enzymes.[15] Geometry optimizations were performed with a medium-sized 6-31G(d) basis set. A conjugate gradients energy minimization algorithm was employed for all geometry optimizations with the convergance criterion of the root mean square gradient of energy <0.01 kcal/mol Å. Final energies and atomic properties (Mulliken spin densities) were obtained by performing single-point energy evaluations with the 6-311++G(2d,2p) basis set on top of the geometries optimized with the smaller basis set. Unless explicitly stated, the energies discussed in the paper correspond to the calculations with the larger basis set. All QC/MM calcula-

tions, including the geometry optimizations, were performed with electronic embedding.

The calculations were carried out for a QC/MM model with 60 QC-atoms (64 with link-atoms included). The other atoms were treated at the MM-level. The QC-region is made up by the pyruvate ligand, the side-chains of Cys418, Cys419, Arg176 and Arg435, the protein backbone link between the cysteines and two water molecules Wat2859 and Wat3252 (see Fig. 2). Thus, the total charge of the QC-region is +1 and the multiplicity is 2.

No extra water molecules were added to the system except the ones present in the crystal structure. The outer parts of the QC/MM model were kept restrained during the geometry optimizations. All QC-atoms were allowed to move freely as well as the MM-atoms within the radius of 8 Å from every QC-atom. At the distance of 8–16 Å from every QC-atom, harmonic positional restraints with an increasing force constant were applied to the MM-atoms. Outside the radius of 16 Å, all atoms were restrained with the maximum force constant of 12 kcal/mol.

Exploration of the reaction path. The reaction path was gradually constructed starting from the optimized geometry of the crystal structure (the substrate state). Potential energy surface (PES) scans followed by geometry optimizations were performed to find the geometries of intermediates and transition states. The QC/MM model was in the first step preoptimized using only the force field. All subsequent geometry optimizations were performed using the QC/MM-potential. A PES scan was done by extending or shortening, usually in steps of ±0.1 Å, a selected distance between two atoms that best approximates the actual reaction coordinate for the particular reaction step. For example, the attack of the Cys418 radical on pyruvate was simulated by shortening the $S_{\gamma,Cys418}\cdots C2_{pyr}$ distance. At each point of the scan, a restrained geometry optimization was performed. The geometries of energy minima resulting from the scans were reoptimized and used as starting points for consecutive scans. This procedure was repeated until the product state was found. Since the reaction coordinate chosen for a PES only approximates the actual

one, the reaction paths connecting the minima were recalculated using a variant of the nudged elastic band method (NEB),[31,32] as implemented in the pDynamo library. In this method a chain of so-called frames is generated between two geometries of the previously calculated minima. The frames kept separated by a special merit function and optimized simultaneously to the minimum energy path (MEP). Usually a set of thirteen frames was used to represent the MEP. The same convergance criterion was used for the NEB calculations as for the geometry optimizations. The highest energy points on the NEB-derived reaction paths were taken as transition state geometries. In the last step, the reaction profiles were refined by performing single-point energy evaluations with the larger basis set on top of the NEB-calculated geometries. The key benefit of using the NEB method for searching transition states is that it only uses the first derivatives of energy, unlike the other methods that also require the second derivatives. For a large enzymatic system the calculation of second derivatives would be computationally too expensive.

Results and Discussions

The crystal structure of the substrate-bound form of PFL shows pyruvate in close contact with one of the two active site cysteines, namely Cys418. This cysteine is perfectly positioned for the attack of the thiyl radical on the carbonyl carbon atom of pyruvate (C2; see Fig. 2 for atom names). The distance from the S_γ atom of Cys418 to the C2 atom of pyruvate is only 2.6 Å. The second cysteine, Cys419, is located further from the substrate and it was suggested to act as a hydrogen atom relay between Gly734 and Cys418. We started the exploration of the reaction path from the point when the radical is localized on Cys419. Although the primary radical storage in PFL is Gly734, our QC/MM calculations on the present crystal structure indicate that the radical is far more stable on Cys419 than on Gly734. This issue will be discussed in the last section of the paper.

Pyruvate in the active site of PFL is hydrogen-bound to two arginine residues, Arg176 and Arg435. The electrostatic calculations predict a typical protonation behavior for the

arginines, i.e. they are both positively charged. The arginines form hydrogen bonds to the carboxylic group of pyruvate, thus stabilizing the negative charge of the substrate. The shortest of these bonds is 1.7 Å (crystal structure with CHARMM-optimized hydrogens). The active sites of enzymes similar to PFL, namely iGDH and 4-Hpad, show negatively charged aspartic or glutamic acid residues in a direct vicinity of the substrate that have recently been shown to be crucial for the catalysis by these enzymes, since they act as proton acceptors/donors.[12,13] Interestingly, there are no such residues in the reacting region of PFL. The closest to the active site is Asp661, which is involved in the stabilizing interactions with Arg176.

Substrate state. Depending on the orientation of the Cys418 side-chain, we found two possible substrate states for the radical localized on Cys419 (see Fig. 3). In the first substrate state (Sub), the H_γ hydrogen atom of Cys418 points towards the carboxylic group of pyruvate. The optimized geometry of Sub shows the atoms $H_{\gamma,Cys418}$ and $O1_{Pyr}$ at the distance of 2.2,Å from each other. The H_γ hydrogen is rotated by 180° in the second substrate state (Sub'), where it points towards the indole ring of Trp333. Both substrate states are geometrically very similar and differ mainly in the position of the H_γ atom. The second substrate state was calculated to be 6.9 kcal/mol more stable than the first one. However, the barrier for the radical transfer Cys419→Cys418 is significantly lower when the transfer starts from Sub (7.6 kcal/mol) than from Sub' (13.3 kcal/mol). The two substrate states are separated by a barrier of 2.9 kcal/mol, which corresponds to the rotational transition state Sub→Sub' between the states. The radical in Sub/Sub' is located primarily on the S_γ atom of Cys419. The Mulliken atomic spin density was calculated to be 0.82 for this atom (see Fig. 3). Thiyl radicals are known to be highly localized.[44] Some traces of the radical are also visible on the S_γ atom of Cys418 (0.08).

Radical transfer Cys419→Cys418. During the first of the studied reaction steps, the H_γ hydrogen of Cys418 is transferred from Cys418 to Cys419 (see Fig. 3). The transfer of hydrogen will move the radical from Cys419 to Cys418. Interestingly, PFL is

the only glycyl radical known to date, in which two active site cysteines participate in the catalysis; the other enzymes from this family possess only one cysteine. To model the abstraction of hydrogen by Cys419, a PES scan was performed by shortening the $H_{\gamma,Cys418} \cdots S_{\gamma,Cys419}$ distance in steps of 0.1 Å. The reaction path was in the next step recalculated with the NEB method. Two scans and NEB calculations were performed starting from the geometries of Sub and Sub'. Regardless of the used starting point, the distance for the hydrogen atom to overcome to bind at Cys419 is nearly the same (3.2 Å and 3.4 Å for Sub and Sub', respectively). The transition state TS1/TS1' estimated by the NEB method shows the $H_{\gamma,Cys418}$ atom at the distance of around 1.9 kcal/mol from the $S_{\gamma,Cys419}$ atom. The spin density in the transition state is shared by two sulfur atoms, namely $S_{\gamma,Cys418}$ (0.75) and $S_{\gamma,Cys419}$ (0.20). For TS1, the calculated activation energy is 7.6 kcal/mol (see Fig. 5 for the energy profile). However, the transfer of hydrogen starting from Sub' requires 13.3 kcal/mol, which is the highest barrier calculated in the forward direction of the catalyzed reaction. Once TS1/TS1' is passed, the energy decreases by about 10 kcal/mol and the system arrives at the first intermediate (In1). Both transition states are energetically equivalent compared to the energy of In1. The transition states also do not display noticeable differences in geometry.

Attack of the Cys418 thiyl radical on pyruvate. The optimized geometry of In1 shows the radical localized on the sulfur atom of Cys418. The distance between the S_γ atom of Cys418 and the C2 atom of pyruvate is slightly shorter in In1 (2.9 Å) than in Sub or Sub' (3.2 and 3.0 Å, respectively). To simulate the attack of Cys418 on pyruvate, a PES scan was performed by further shortening the $S_{\gamma,Cys418} \cdots C2_{Pyr}$ distance by -0.1 Å in each step. The reaction path was subsequently refined using the NEB method. The barrier for the addition of the thiyl radical to pyruvate was calculated to be as low as 5.0 kcal/mol. The obtained barrier is considerably lower than the one calculated before, which was 12.3 kcal/mol and 11.8 kcal/mol for the gas-phase model of neutral[16] and anionic pyruvate,[17] respectively, and 12.9 kcal/mol for the more recent cluster model.[18]

Interestingly, the energetics of this reaction step were found to be quite insensitive on the used level of theory, i.e. nearly the same barriers were obtained using the smaller and the larger basis set (see Fig. 6 for details). As a result of thiyl addition to pyruvate, a complex protein-bound tetrahedral radical intermediate is formed (In2). In the previous computational studies on PFL, the occurance of this intermediate on the reaction path was often questioned, because its energy was calculated to be similar to that of TS2, that is, the intermediate was localized in a very shallow energy minimum on the PES. However, for the present QC/MM model, In2 is around 2.7 kcal/mol lower in energy than TS2. The stability of In2 with respect to In1 has also improved; the tetrahedral intermediate is now only 2.3 kcal/mol less stable than the starting intermediate In1. In the study by Lucas on the isolated gas-phase model with anionic pyruvate,[17] the S–C bond formation and the C–C bond cleavage were found to be coupled in one reaction step. In our full-enzyme model we observe that the $C1_{Pyr}$—$C2_{Pyr}$ bond has been only extended from 1.6 Å (In1) to 1.8 Å (In2). In the tetrahedral intermediate, the S_{Cys418}—$C2_{Pyr}$ bond length is 1.9 Å. From the analysis of the Mulliken spin density it is visible that the radical is now shared between the sulfur atom of Cys418 (0.25), the carbonyl oxygen of pyruvate (0.32) and the carboxylic group of pyruvate (0.41; see Fig. 4 for the values on particular atoms). Given the relatively short length of the $C1_{Pyr}$—$C2_{Pyr}$ bond in In2 and the calculated delocalized spin distribution, it is clear that the full C–C bond cleavage in pyruvate has not taken place yet. Another interesting aspect of the formation of the tetrahedral intermediate is that the $C2_{Pyr}$—$O2_{Pyr}$ bond length does not change significantly. This bond is 1.2 Å long in In1 and 1.3 Å long after the tetrahedral intermediate has been formed, i.e. it remains as a C=O double bond.

The usual entry point to the catalysis by radical enzymes is the abstraction of a hydrogen atom from the substrate. To perform the abstraction, glycyl radical enzymes employ a transient radical cysteine in the active site. Following the activation to a radical form, the substrate enters a series of complex transformations that are not feasible under

normal conditions. In the case of PFL, however, Cys418 attacks the carbonyl group of pyruvate, forming a radical intermediate covalently bound to the protein. Another example of a glycyl radical enzyme that performs unusual activation of its substrate is the recently studied 4-hydroxyphenylacetate decarboxylase.[13]

Collapse of the tetrahedral radical intermediate. After the addition of the Cys418 thiyl radical to pyruvate, the reaction can proceed to the next stage, which is the cleavage of the C–C bond in pyruvate that gives the formyl radical and acylated Cys418. To model the collapse of the tetrahedral intermediate, a PES was first performed along the $C1_{Pyr}$—$C2_{Pyr}$ bond; the distance between the two atoms was gradually extended by 0.1 kcal/mol. The scan was followed by the NEB calculation. Mechanistically, this reaction step was found to involve several simultaneous events. First, the distance between the atoms $C1_{Pyr}$ and $C2_{Pyr}$ extends from 1.8 Å (In2) to 2.9 Å (In3). Second, the S_{Cys418}—$C2_{Pyr}$ bond is shortened from 1.9 Å to 1.8 Å. Third, the side-chain of Cys419 that pointed towards Cys418 rotates in the direction of the newly formed formyl radical. From the NEB-calculated reaction path it can be seen that the H_γ atom of Cys419 follows the moving carboxylic moiety of pyruvate. The distance between the atoms $H_{\gamma,Cys418}$ and $C1_{Pyr}$ is shortened from 3.8 Å (In2) to 3.0 Å (In3). That is, after the splitting of pyruvate, the formyl radical is perfectly positioned for the hydrogen abstraction from Cys419 (see In3 on Fig. 4). This observation suggests that not CoA but rather Cys419 is responsible for quenching of the formyl radical. In the small-molecule studies on the mechanism of PFL, such movement of the Cys419 side-chain was not visible.

Unlike for the In1→In2 addition step, the energetics of the C–C bond cleavage were found to be considerably dependent on the applied level of theory. Namely, the barrier calculated for the smaller basis set (5.7 kcal/mol) is twice as high as the one obtained using the larger basis set (2.6 kcal/mol; see Fig. 6 for detailed energy profiles). The resulting cleaved intermediate (In3) is virtually isoenergetic with the preceeding transition state (TS3), i.e. In3 lies in a very shallow energy minimum on the PES. Namely, the energy

difference between TS3 and In3 does not exceed 0.5 kcal/mol. Taking the small energy difference between TS3 and In3, the similarity between the two geometries is consistent with the Hammond's postulate.[45]

With the extension of the $C1_{Pyr}$—$C2_{Pyr}$ bond, the radical shifts from the enzyme-bound part towards the leaving carboxylic moiety of pyruvate. Since the transition state TS3 is geometrically and energetically very close to the intermediate In3, the spin distributions of the two are nearly the same. The radical in TS3/In3 is localized entirely on the newly created formyl moiety and is shared between the atoms $C1_{Pyr}$ (0.70), $O12_{Pyr}$ (0.14) and $O11_{Pyr}$ (0.16; see In2→In3 on Fig. 4). The acetyl-Cys418 moiety is free of the radical.

Quenching of the formyl radical. In the last reaction step of the first part of the PFL-catalyzed reaction, the formyl radical has only to abstract a hydrogen atom from the enzyme. The nearest accessible hydrogen atom is $H_{\gamma,Cys419}$, as mentioned before. The abstraction of hydrogen from Cys419 was in the first step modeled with a PES scan by shortening the $C2_{Pyr} \cdots H_{\gamma,Cys419}$ distance by 0.1 Å in each step and subsequently refined by using the NEB method. For the last reaction step, the calculated barrier is 10.0 kcal/mol (or 8.8 kcal/mol if only the smaller basis set is used). The relatively high barrier obtained for our QC/MM model in comparison to the previous isolated models can be related to the rather long distance of 3.0 Å that the hydrogen atom has to overcome to transfer between Cys419 and the formyl radical. Moreover, pyruvate inside the active site of PFL is involved in a network of hydrogen bonds with Arg176 and Arg435. These bonds have to be streched during the radical transfer formyl→Cys418, as visible from the NEB-calculated reaction path. Stretching of the hydrogen bond network requires some energy and increases the activation energy for the last reaction step. After the abstraction of hydrogen is complete, the radical is localized back on Cys419. The Mulliken spin density was calculated to be 0.97 for the $S_{\gamma,Cys419}$ atom. Geometrically, the difference between In3 and Pro is only the new position of the hydrogen atom from Cys419. The first part

of the PFL catalytic reaction ends with acylated Cys418, free formate and the Cys419 thiyl radical. On the basis of our observations for the QC/MM model, we propose that the reaction continues with the release of formate from the active site, binding of CoA and abstraction of hydrogen from CoA by the Cys419 thiyl radical.

Conclusions

In the present work, we have investigated the first part of the PFL catalytic cycle by performing QC/MM calculations on the full-enzyme model. The use of a complete enzyme eliminates some of the problems of small-molecule models, since the protein environment is taken into account during the calculations. For example, the orientation of reactants throughout the catalytic cycle is more realistic because of the spatial restraints imposed by the enzyme.

There are two highlights of this work. First, our calculations indicate that the tetrahedral radical intermediate that results from the thiyl attack of Cys418 on pyruvate is a relatively stable species. The barrier required for the formation of this intermediate was calculated to be as low as 5 kcal/mol, which translates into a twice lower barrier in comparison to the previously studied small-molecule models. Second, we propose that quenching of the formyl radical is performed by Cys419 and a molecule of CoA binds to the active site only after the release of formate. The calculated reaction path shows that the cleavage of the C–C bond in pyruvate is coupled to the movement of the Cys419 side-chain towards the leaving formyl radical. After the cleavage, Cys419 is perfectly positioned for the formyl→Cys419 radical transfer.

Acknowledgments. This work was supported by the DFG grant UL 174/8-2 and the BioMedTec International Graduate School of the Elitenetwork Bavaria.

Supplementary materials. Cartesian coordinates and Mulliken atomic spin densities and charges for the optimized structures (only the active site region). This material is available free of charge via the Internet at http://pubs.acs.org.

References

[1] Buckel, W.; Golding, B. *Fems Microbiol. Rev.*, **1998**, *22*(5), 523–541.

[2] Frey, P. *Annu. Rev. Biochem.*, **2001**, *70*, 121–148.

[3] Buckel, W.; Golding, B. T. *Annu. Rev. Microbiol.*, **2006**, *60*, 27–49.

[4] Buckel, W. *Angew. Chem.-Int. Edit.*, **2009**, *48*(37), 6779–6787.

[5] Eklund, H.; Fontecave, M. *Struct. Fold. Des.*, **1999**, *7*(11), R257–R262.

[6] Selmer, T.; Pierik, A.; Heider, J. *Biol. Chem.*, **2005**, *386*(10), 981–988.

[7] Craciun, S.; Balskus, E. P. *Proc. Nat. Acad. Sci. USA*, **2012**, *109*, 21307–21312.

[8] Thibodeaux, C.; van der Donk, W. A. *Proc. Nat. Acad. Sci. USA*, **2012**, *109*, 21184–21185.

[9] Knappe, J.; Neugebauer, F.; Blaschkowski, H.; Ganzler, M. *Proc. Nat. Acad. Sci. USA*, **1984**, *81*, 1332.

[10] Liu, Y.; Gallo, A. A.; Florian, J.; Liu, Y.-S.; Mora, S.; Xu, W. *J. Phys. Chem. B*, **2010**, *114*(16), 5497–5502.

[11] Liu, Y.; Gallo, A. A.; Xu, W.; Bajpai, R.; Florian, J. *J. Phys. Chem. A*, **2011**, *115*(41), 11162–11166.

[12] Feliks, M.; Ullmann, G. M. *J. Phys. Chem. B*, **2012**, *116*(24), 7076–7087.

[13] Feliks, M.; Martins, B.; Ullmann, G. M. *J. Am. Chem. Soc.*, **2013**, *135*(39), 14574–14585.

[14] Himo, F. *J. Phys. Chem. B*, **2002**, *106*(31), 7688–7692.

[15] Himo, F.; Siegbahn, P. *Chem. Rev.*, **2003**, *103*(6), 2421–2456.

[16] Himo, F.; Eriksson, L. *J. Am. Chem. Soc.*, **1998**, *120*(44), 11449–11455.

[17] Lucas, M.; Fernandes, P.; Eriksson, L.; Ramos, M. *J. Phys. Chem. B*, **2003**, *107*(24), 5751–5757.

[18] Guo, J.; Himo, F. *J. Phys. Chem. B*, **2004**, *108*(39), 15347–15354.

[19] Becker, A.; Wolf, K.; Kabsch, W.; Knappe, J.; Schultz, W. A. S. *Nat. Struc. Biol.*, **1999**, *6*, 969.

[20] Himo, F. *Theor. Chem. Acc.*, **2006**, *116*(1-3), 232–240.

[21] Chen, S.-L.; Fang, W.-H.; Himo, F. *Theor. Chem. Acc.*, **2008**, *120*(4-6), 515–522.

[22] Siegbahn, P. E. M.; Himo, F. *J. Biol. Inorg. Chem.*, **2009**, *14*(5), 643–651.

[23] Becker, A.; Kabsch, W. *J. Biol. Chem.*, **2002**, *277*(42), 40036–40042.

[24] Field, M. *J. Comput. Chem.*, **2002**, *23*(1), 48–58.

[25] Senn, H. M.; Thiel, W. *Angew. Chem.-Int. Edit.*, **2009**, *48*(7), 1198–1229.

[26] Acevedo, O.; Jorgensen, W. L. *Accounts Chem. Res.*, **2010**, *43*(1), 142–151.

[27] Condic-Jurkic, K.; Perchyonok, V. T.; Zipse, H.; Smith, D. M. *J. Comput. Chem.*, **2008**, *29*(14), 2425–2433.

[28] Condic-Jurkic, K.; Zipse, H.; Smith, D. M. *J. Comput. Chem.*, **2010**, *31*(5), 1024–1035.

[29] Ullmann, G. M.; Knapp, E. W. *Eur. Biophys. J.*, **1999**, *28*(7), 533–551.

[30] Bombarda, E.; Ullmann, G. M. *J. Phys. Chem. B*, **2010**, *114*(5), 1994–2003.

[31] Galvan, I. F.; Field, M. J. *J. Comput. Chem.*, **2008**, *29*(1), 139–143.

[32] Aleksandrov, A.; Field, M. *Phys. Chem. Chem. Phys.*, **2012**, *14*(36), 12544–12553.

[33] Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.*, **1983**, *4*, 187–217.

[34] Brooks, B. R.; Brooks, C. L. III; Mackerell, A. D. Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.*, **2009**, *30*(10, SI), 1545–1614.

[35] MacKerell, A. D. et al. *J. Phys. Chem. B*, **1998**, *102*, 3586–3616.

[36] Bashford, D.; Gerwert, K. *J. Mol. Biol.*, **1992**, *224*, 473–486.

[37] Ullmann, R. T.; Ullmann, G. M. *J. Comp. Chem*, **2012**, *33*, 887–900.

[38] Field, M. J. *J. Chem. Theory Comput.*, **2008**, *4*(7), 1151–1161.

[39] Neese, F. *Wiley Interdiscip. Rev.-Comput. Mol. Sci.*, **2012**, *2*(1), 73–78.

[40] Vosko, S.; Wilk, L.; Nusair, M. *Can. J. Phys.*, **1980**, *58*(8), 1200–1211.

[41] Lee, C.; Yang, W.; Parr, R. *Phys. Rev. B*, **1988**, *37*(2), 785–789.

[42] Becke, A. *J. Chem. Phys.*, **1993**, *98*(7), 5648–5652.

[43] Stephens, P.; Devlin, F.; Chabalowski, C.; Frisch, M. *J. Phys. Chem.*, **1994**, *98*(45), 11623–11627.

[44] Himo, F. *Biochim Biophys Acta*, **2005**, *1707*(1), 24–33.

[45] Hammond, G. *J. Am. Chem. Soc.*, **1955**, *77*(2), 334–338.

[46] Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.*, **1996**, *14*(1), 33–38.

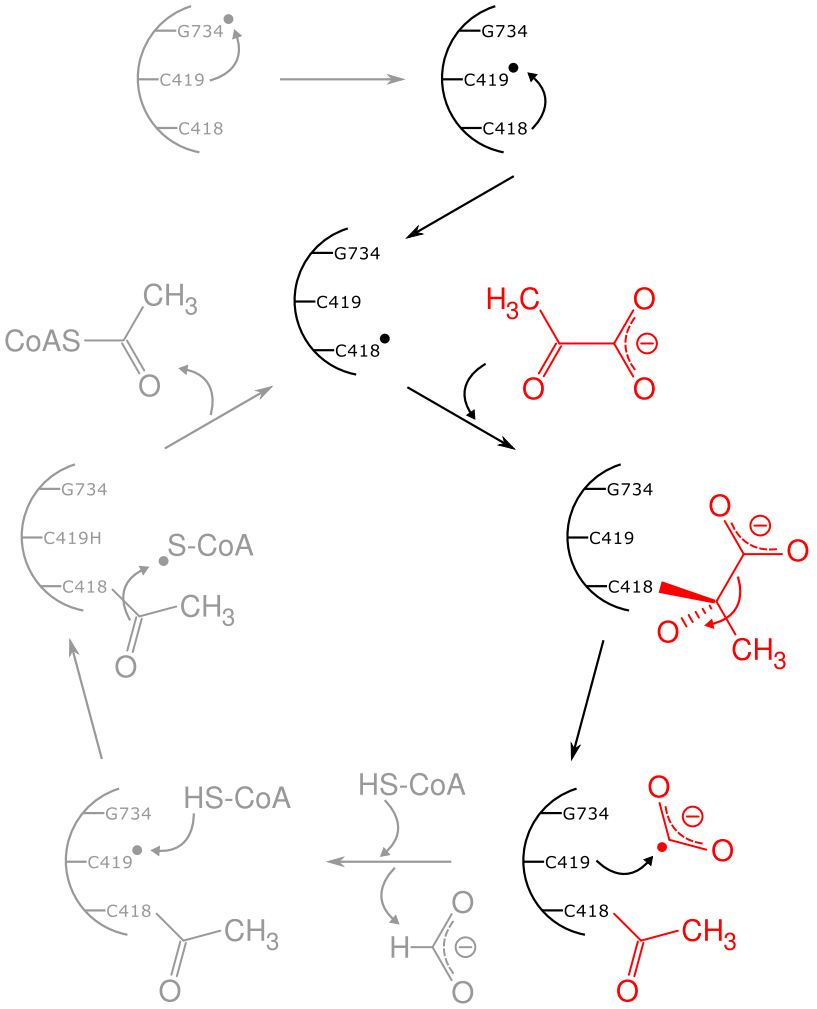

Figure 1: The PFL-catalyzed reaction cycle. The first part of the cycle studied in this paper is depicted in black and red color. The second part not included in this study is shown in grey color. The first part involves four elementary steps. In the first step, a hydrogen atom is transferred from Cys418 to Cys419. In the second step, Cys418 attacks the carbonyl carbon atom of pyruvate, which leads to the formation of a radical tetrahedral intermediate. Next, the intermediate collapses into the formyl radical and acylated Cys418. In the last step, the formyl radical abstracts a hydrogen atom from Cys419, which gives formate and regenerates the radical at Cys419.

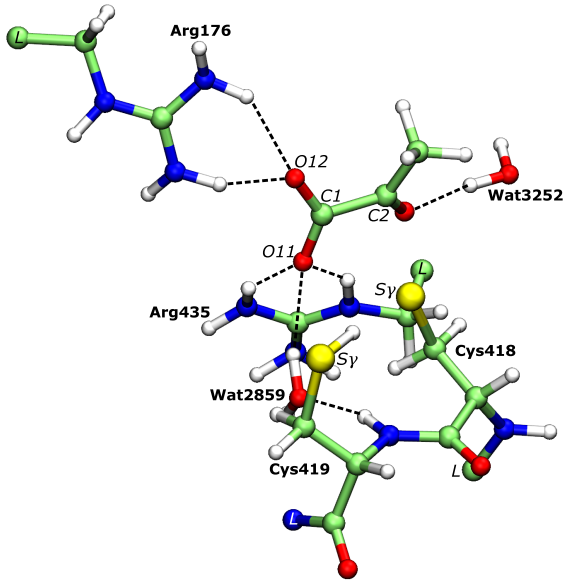

Figure 2: QC-part of the QC/MM model used in the calculations (optimized geometry of In1). The rest of the protein was omitted for clarity. The QC-part encompasses pyruvate, Cys418 and Cys419, the protein backbone link between the cysteines, the side-chains of Arg176 and Arg435 and two water molecules Wat2859 and Wat3252 (64 QC-atoms atoms in total including link-atoms). Dashed lines depict hydrogen bonds. Atoms discussed in the text are labeled. The label *L* indicates atoms replaced by hydrogen-type link-atoms during the QC/MM calculations.

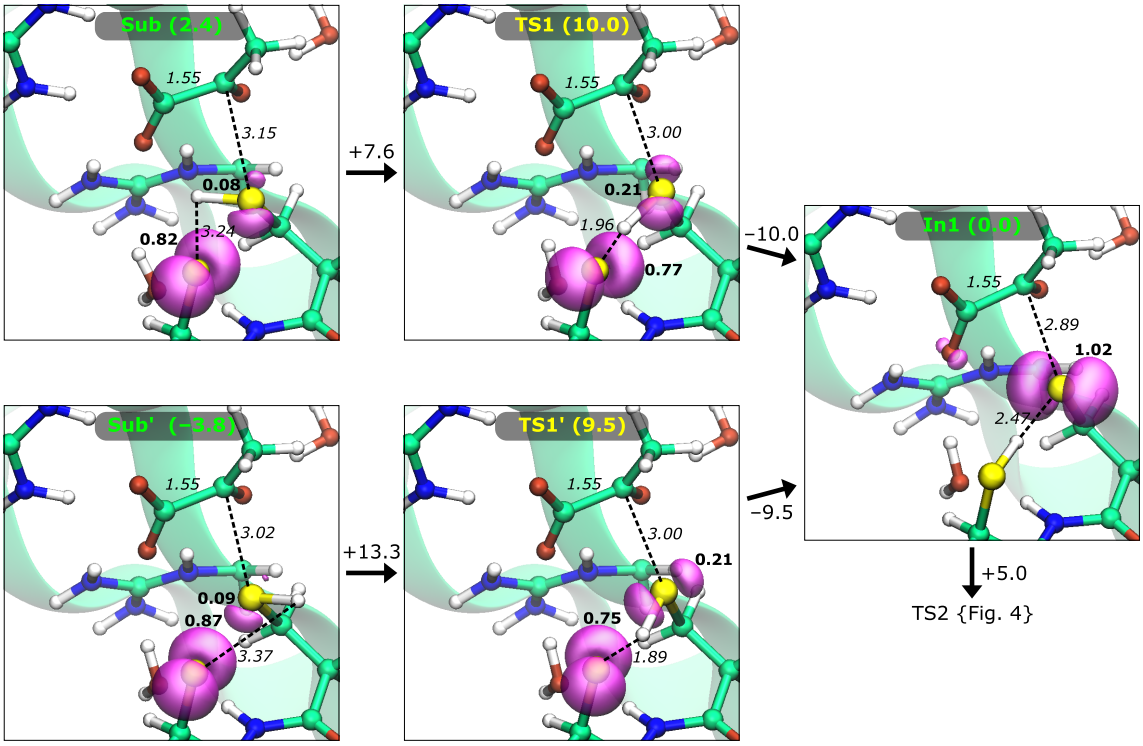

Figure 3: QC/MM-optimized geometries of intermediates and transition states for the radical transfer Cys419→Cys418. Since there are two possible substrate states, the energies (in kcal/mol) are given with respect to the first intermediate (In1; radical localized on Cys418). The transition states were taken from the NEB-derived reaction profiles as the highest energy points. Numbers at the arrows indicate changes of the energy on the reaction path. Relevant inter- and intramolecular distances (in Å) are depicted in italics. The Mulliken atomic spin densities (for clarity, only the ones of the absolute value $\geq$0.1) are depicted in bold. The 3D spin denisity is shown in magenta color at the isovalue of 0.01 a.u. The model was visualized in VMD.[46]

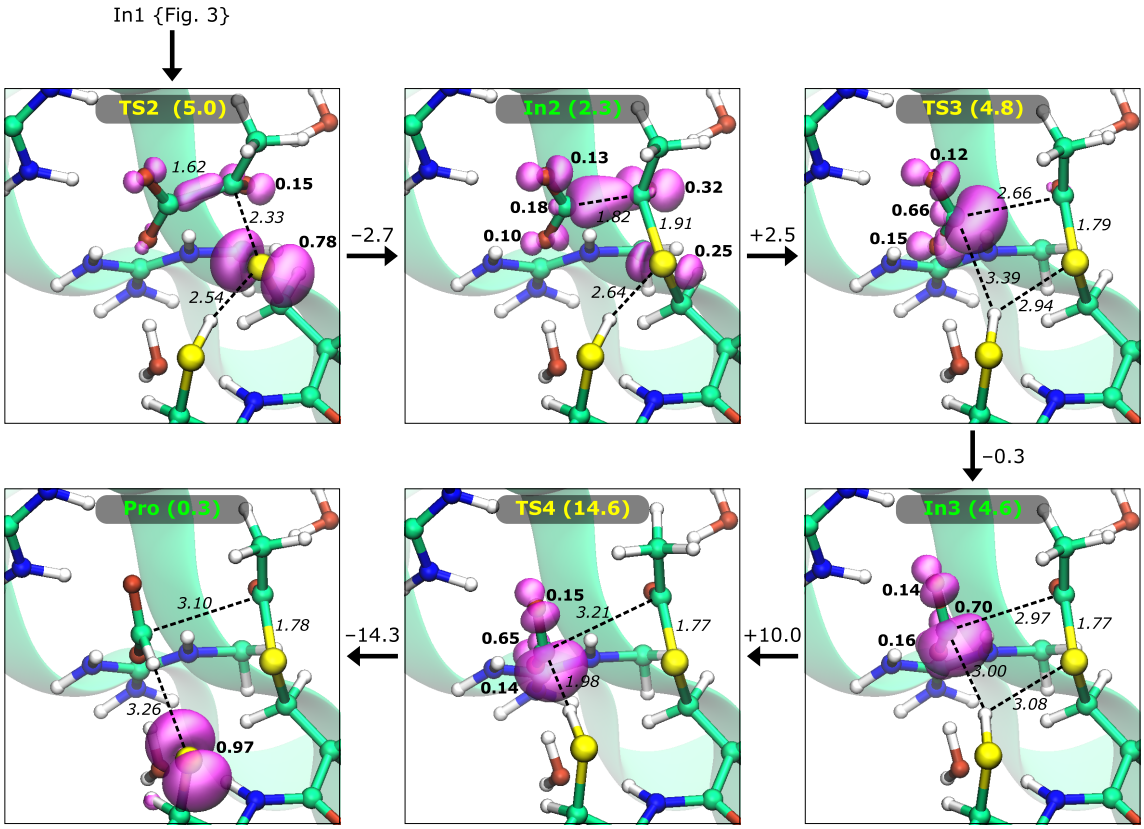

Figure 4: QC/MM-optimized geometries of intermediates and transition states for the formation of the S–C bond between pyruvate and Cys418 (In1→In2), cleavage of the C–C bond of pyruvate (In2→In3) and formation of formate and regeneration of the radical on Cys419 (In3→Pro). See description under Fig. 3 for details.

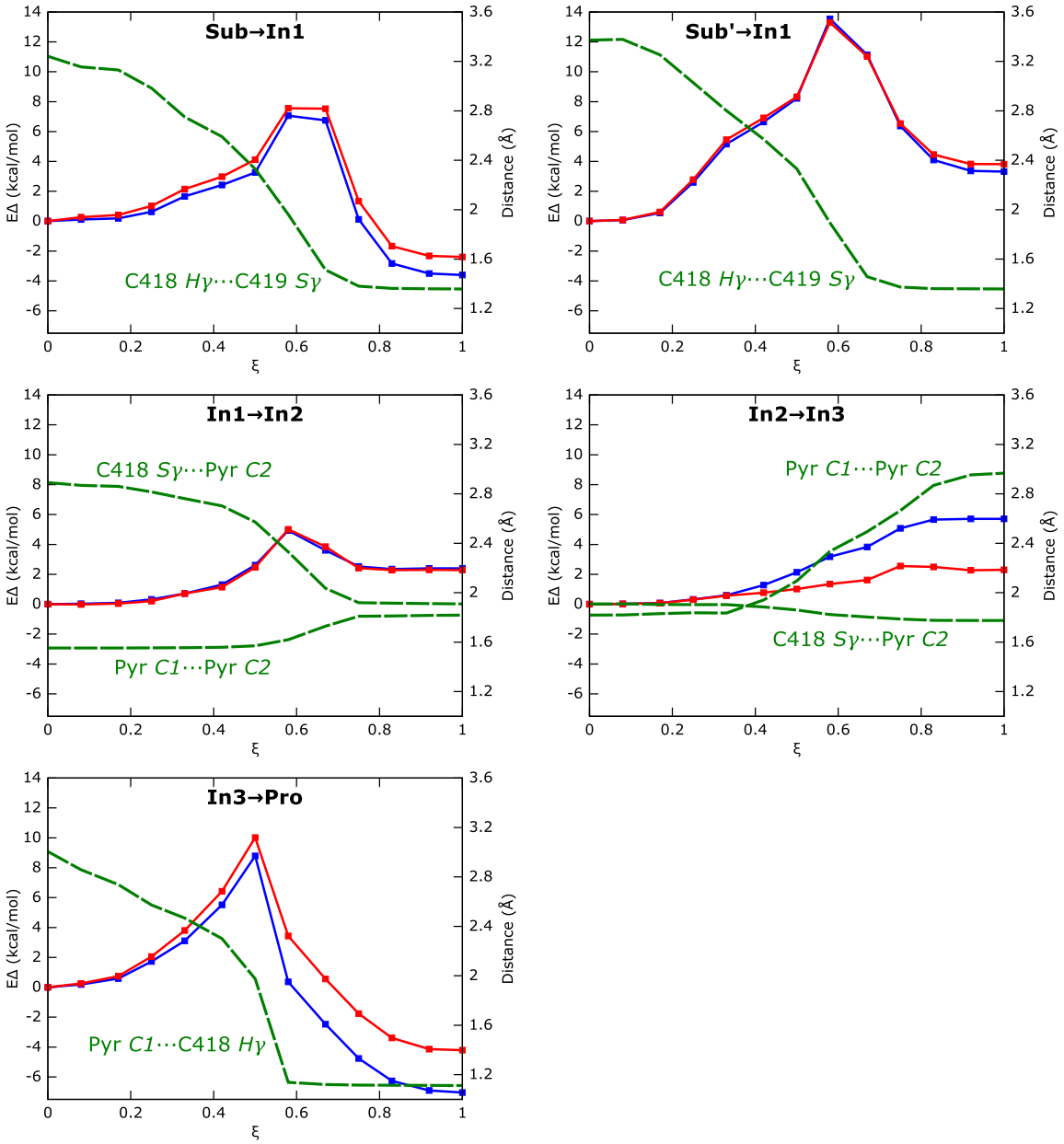

Figure 5: Reaction paths for individual reaction steps calculated using the NEB method. For all paths, 13 NEB frames were used. Blue and red lines represent energy profiles calculated with the smaller 6-31G(d) and the larger 6-311++G(2d,2p) basis set, respectively. Green dashed lines represent changes (in Å) of important distances on the reaction path. ξ is the normalized reaction coordinate ($\xi = 0$ for substrate state; $\xi = 1$ for product state).

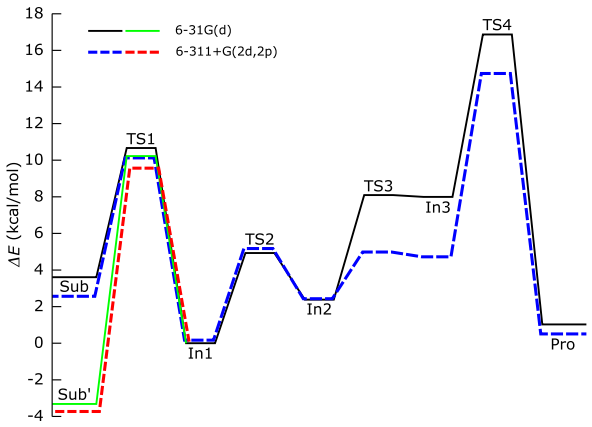

Figure 6: Energy profiles of the PFL-catalyzed reaction for the QC/MM model used in this study. Solid lines are the profiles calculated for the geometries optimized with the smaller 6-31G(d) basis set. Dashed lines are the profiles calculated by performing single-point energy evaluations with the larger 6-311++G(2d,2p) basis set on top of the geometries optimized with the smaller basis set. The profiles are relative to the energy of the first intermediate (In1). Sub/Sub' and In1 are the radical localized on Cys419 and Cys418, respectively. In2 is the tetrahedral pyruvate-enzyme radical intermediate. In3 is the intermediate that have collapsed into the formyl radical and Cys418-acylated enzyme. Pro is formate, Cys418-acylated enzyme and the radical localized back on Cys419.

Python scripts for running QC/MM calculations

The QC/MM calculations for modeling of the reaction mechanisms of 4Hpad and PFL have been carried out by custom Python scripts written within the framework of the pDynamo library (1.7.2) by Martin Field.[129] pDynamo is an open source program for the simulation of molecular systems using quantum chemical, molecular mechanical and hybrid quantum chemical/molecular mechanical methods. In this section, the functionality of the most important scripts is briefly outlined.

The script **setup.py** is used to build a QC/MM model based on the topology (PSF) and coordinate (CRD) files generated by CHARMM. Necessary functions and objects are loaded from pDynamo modules in lines 1-5. The parameter files as well as the PSF and CRD files are loaded in lines 9-11. In the next step, the system object called *mol* is constructed (lines 13-15). The following lines contain a tuple defining the QC-region of the QC model by specifying the chain ID, residue name, residue number and a list of atoms that are to be included into the QC-region. Link atoms are not specified since they are generated automatically by pDynamo. The tuple is translated into a list called *qc* containing the indices of QC-atoms (lines 27-36). Based on this list, positional restraints are calculated that are applied to MM-atoms outside the distance of 8 Å from every QC-atom. The restrains increase linearly to the distance of 16 Å, outside of which the maximum value is applied, in this case 50 kJ/mol. The calculation of restraints is done in lines 38-54. In lines 56-61, two dictionaries are generated containing the indices of restrained and fully moveable atoms. These dictionaries are stored on disk as PKL files for further use by the other scripts. For visualization purposes, XYZ files are generated as well as plain text files containing the atomic indices. In lines 76-77, a PDB file is written containing the geometry of the model. The column with beta factors is substituted by the calculated restraints (lines 79-88). The restraints are scaled to the range of $< 0; 1 >$. The so-modified PDB file can be used for visual inspection of the restrained atoms. Finally, the system object is stored on disk as a compressed PKL file (lines 90-91).

setup.py

```
1 from pBabel           import CHARMMParameterFiles_ToParameters,
    CHARMMPSFFile_ToSystem, CHARMMCRDFile_ToCoordinates3,
    PDBFile_FromSystem, XYZFile_FromSystem
2 from pCore            import Selection, Pickle, GzipPickle, logFile
```

```
from pMoleculeScripts import PruneByAtom
from tempfile         import mktemp
from os               import uname

logFile.Header ("Starting job on %s..." % uname ()[1])

ptab = ("../toppar/par_all27_prot_na.inp", )
psf  = "../charmm/monomer.psf"
crd  = "../charmm/monomer.crd"

mol  = CHARMMPSFFile_ToSystem (psf, isXPLOR = True,
  parameters = CHARMMParameterFiles_ToParameters (ptab))
mol.coordinates3 = CHARMMCRDFile_ToCoordinates3 (crd)

actsite = (
("PRTA", "GLY", "502", "C", "O"),
("PRTA", "CYS", "503", "N", "H", "CA", "HA", "CB", "HB1", "HB2", "SG",
   "HG1", "C", "O"),
("PRTA", "LEU", "504", "N", "H", "CA", "HA", "C", "O"),
("LIGA", "4HP", "001", "C7", "H71", "H72", "C8", "O1", "O2", "C1",
   "C6", "H6", "C5", "H5", "C4", "O4", "HO4", "C2", "H2", "C3", "H3",
   "HO2"),
("PRTA", "SER", "344", "CB", "HB1", "HB2", "OG", "HG1"),
("PRTA", "HIS", "536", "CB", "HB1", "HB2", "CD2", "HD2", "CG", "NE2",
   "HE2", "ND1", "HD1", "CE1", "HE1"),
("PRTA", "GLU", "637", "CG", "HG1", "HG2", "CD", "OE1", "OE2"),
("PRTA", "GLU", "505", "CG", "HG1", "HG2", "CD", "OE1", "OE2", "HE2"), )

qc = []
for res in actsite:
  seg, rname, rnum = res[:3]
  atoms = res[3:]
  for a in atoms:
    pat = "%s:%s.%s:%s" % (seg, rname, rnum, a)
    try:
      qc.append (mol.sequence.AtomIndex (pat))
    except:
      logFile.Text ("WARNING: atom %s not found\n" % pat)

Cmax   = 50.
da, db = 8., 16.
Restr  = []
for i in range (0, mol.sequence.NumberOfAtoms ()):
  c = 0.
  if not i in qc:
    d = 10000.
    for j in qc:
      cd = mol.coordinates3.Distance (i, j)
      if cd < d:
        d = cd
    if d > da:
      if d < db:
```

```
51          c = Cmax / (db - da) * (d - da)
52        else:
53          c = Cmax
54    Restr.append (c)

56  co, fl = {}, {}
57  for i, r in enumerate (Restr):
58    if r > 0.:
59      co[i] = r
60    else:
61      fl[i] = r

63  for (tup, pkl, lst, xyz, label) in (
64    (co, "co.pkl", "co", "co.xyz", "Restrained atoms"),
65    (fl, "fl.pkl", "fl", "fl.xyz", "Flexible atoms"),
66    (qc, "qc.pkl", "qc", "qc.xyz", "QC-region"),
67  ):
68    Pickle (pkl, tup)
69    o = open (lst, "w")
70    o.write (" ".join (map (str, tup)))
71    o.close ()

73    sel = PruneByAtom (mol, Selection (tup))
74    XYZFile_FromSystem (xyz, sel, label = label)

76  tf = mktemp ()
77  PDBFile_FromSystem (tf, mol)

79  fo = open ("see_restr.pdb", "w")
80  i  = 0
81  for p in open (tf).readlines ():
82    if p[:6] in ("ATOM  ", "HETATM"):
83      c  = Restr[i] / Cmax
84      i += 1
85      fo.write ("%s%4.2f%s" % (p[:56], c, p[60:]))
86    else:
87      fo.write (p)
88  fo.close ()

90  GzipPickle ("mol.pklz", mol)
91  logFile.Footer ()
```

The script **qcmmopt.py** is used for geometry optimizations using the QC/MM potential. The preamble of the script contains imports of pDynamo and Python functions and objects. In line 12, a geometry convergence criterion is defined, which is the root mean square of energy lower than 0.04 kJ/mol. Next, the model is loaded from the compressed PKL file and a summary table is printed out (lines 17-18). In lines 20-21, the initial geometry of the model is read from the file *after.xyz*. For the reactant state, the initial geometry

comes from the preoptimization using only the MM-potential. The preoptimization script is similar to the **qcmmopt.py** script but does not contain the setup of the QC-potential and uses the NBModelABFS model instead of the NBModelORCA model for handling of the non-bonding interactions. For other intermediates on the reaction path, the initial geometry is taken as the final geometry generated by the previous QC/MM-optimization. In lines 25-31, a container with restraints is created and attached to the system object. Lines 33-43 initialize the scratch space for the QC-calculation. The actual setup of the QC-potential is done in lines 45-55. An ORCA object is created and attached to the system object. Next, a trajectory object is initialized for storing the intermediate geometries as PKL files. The geometry optimization is done in lines 61-65 by calling the conjugate gradients function in a loop. In the case of a convergence failure, the optimization is automatically restarted. Finally, the QC/MM-optimized geometry is written to disk as an XYZ file (line 67).

qcmmopt.py

```
1  from pBabel           import XYZFile_FromSystem,
       XYZFile_ToCoordinates3, SystemGeometryTrajectory
2  from pCore            import Unpickle, GzipUnpickle, logFile, Selection
3  from pMolecule        import SoftConstraintContainer,
       SoftConstraintEnergyModelHarmonic, SoftConstraintTether,
       NBModelORCA, QCModelORCA, ElectronicState
4  from pMoleculeScripts import ConjugateGradientMinimize_SystemGeometry

6  from sys        import argv
7  from os         import makedirs, uname, getpid
8  from os.path    import basename
9  from time       import strftime
10 from getpass    import getuser

12 TolCrit = 0.04
13 SrcDir  = ".."

15 logFile.Header ("Job started on %s..." % uname ()[1])

17 mol = GzipUnpickle ("%s/mol.pklz" % SrcDir)
18 mol.Summary ()

20 qc  = Unpickle ("%s/qc.pkl" % SrcDir)
21 mol.coordinates3 = XYZFile_ToCoordinates3 ("%s/after.xyz" % SrcDir)

23 ref = XYZFile_ToCoordinates3 ("%s/before.xyz" % SrcDir)

25 co  = Unpickle ("%s/co.pkl" % SrcDir)
26 SCC = SoftConstraintContainer ()
```

```
27 for a in co.keys ():
28   c  = co[a]
29   cm = SoftConstraintEnergyModelHarmonic (0., c)
30   SCC["%s" % a] = SoftConstraintTether (a, ref.GetRow (a), cm)
31 mol.DefineSoftConstraints (SCC)

33 op   = "/home/41/bt260941/local/orca_2_9_0_linux_x86-64/orca"
34 sp   = "/tmp"
35 PID  = getpid ()
36 time = strftime ("%y%m%d%H%M%S")
37 f    = argv[0]
38 fl   = basename (f[:f.rfind (".")])
39 user = getuser ()
40 sp   = "%s/%s_pdynamo/%s-%s-%s" % (sp, user, fl, time, PID)

42 makedirs (sp)
43 logFile.Text ("Scratch directory: %s\n" % sp)

45 mol.electronicState = ElectronicState (charge = -1, multiplicity = 2)

47 qcm = QCModelORCA (
48   "B3LYP:6-31G*", "PAL8", "SCFCONV10",
49   command        = op,
50   scratch        = sp,
51   job            = "job",
52   deleteJobFiles = False )
53 mol.DefineQCModel (qcm, qcSelection = Selection (qc))

55 mol.DefineNBModel (NBModelORCA ())

57 tdir    = "%s/qcmmopt_traj" % sp
58 traj    = SystemGeometryTrajectory (tdir, mol, mode = "w")
59 Conv, a = False, 1

61 while not Conv:
62   logFile.Text ("Optimizing the system (attempt %d)...\n" % a)
63   r = ConjugateGradientMinimize_SystemGeometry (mol, logFrequency = 1,
       maximumIterations = 9999, rmsGradientTolerance = TolCrit,
       trajectories = [(traj, 1)])
64   Conv = r["Converged"]
65   a += 1

67 XYZFile_FromSystem ("after_qcmm.xyz", mol, label = "QC/MM-optimized
     geometry")
68 logFile.Footer ()
```

The script **scan.py** performs a relaxed potential energy surface scan along the distance between two specified atoms. The first part of the script involves the QC/MM model setup and is the same as in the **qcmmopt.py** script. However, Vector3 and SoftConstraintDistance have to be additionally imported from the modules pCore and pMolecule,

respectively. Lines 1-2 define a helper function converting atomic patterns into indices. Lines 4-8 contain the definitions of two atoms along which the scan is performed, step in Å and the final distance between the atoms at which the scan is terminated. Line 10 starts the main loop. First, the current distance between the atoms is calculated. This distance is used as an additional restraint that is added to the restraint container (lines 12-14). A restrained geometry optimization is performed for every step along the scan (lines 19-24). The resulting intermediate geometry is written out as an XYZ file (line 27). In lines 29-30, a check is performed whether the scan has been completed. If not, a vector is calculated between the two atoms (lines 32-37). Based on the vector, the positions of the atoms are updated. The atoms are moved further or closer depending on the variable *step*.

scan.py

```
1  def I (m, t):
2    return m.sequence.AtomIndex ("%s:%s.%s:%s" % t)

4  a     =  I (mol, ("PRTA", "GLY", "734", "HA1"))
5  b     =  I (mol, ("PRTA", "CYS", "419", "SG"))
6  step  = -0.10
7  D_fin =  1.35
8  i     =  0

10 while 1:
11   D  = mol.coordinates3.Distance (a, b)
12   cm = SoftConstraintEnergyModelHarmonic (D, 10000.0)
13   SCC["distance"] = SoftConstraintDistance (a, b, cm)
14   mol.DefineSoftConstraints (SCC)

16   logFile.Text ("*** Optimizing step %d ***\n" % i)
17   XYZFile_FromSystem ("scan%03d.xyz" % i, mol)

19   while 1:
20     r = ConjugateGradientMinimize_SystemGeometry (mol, logFrequency =
         1, maximumIterations = 9999, rmsGradientTolerance = TolCrit)
21     if r["Converged"]:
22       break
23     else:
24       logFile.Text ("*** Restarting the optimization... ***\n")

26   logFile.Text ("*** Done with step %d ***\n" % i)
27   XYZFile_FromSystem ("scan%03d_optimized.xyz" % i, mol)

29   N = abs ((D_fin - D) / step)
30   if N < 1: break

32   A = mol.coordinates3.GetRow (a)
```

```
33    B = mol.coordinates3.GetRow (b)
34    V = Vector3 ()
35    B.CopyTo (V)
36    V.AddScaledVector3 (-1., A)
37    V.Normalize ()

39    f = step * .5
40    mol.coordinates3.AddScaledVector3 (a, -f, V)
41    mol.coordinates3.AddScaledVector3 (b,  f, V)
42    i += 1
```

The nudged elastic band method[90,91] provides a better approximation of the reaction path than a potential energy surface (PES) scan but requires that the geometries of both the initial and the final states are known. The geometry of the latter can be generated by performing a PES scan. The script **neb.py** provides access to the NEB method in pDynamo. The first part of the script is the same as in the **qcmmopt.py** or **scan.py** scripts. Two functions, GrowingStringInitialPath and NudgedElasticBandSplineOptimize_SystemGeometry, have to be imported from the module pMoleculeScripts. Line 1 defines a directory where the NEB-calculated trajectory will be stored. The following lines specify the number of frames and the previously optimized geometries of two energy minima. In line 5, an initial chain of frames is generated by linearly interpolating between the two geometries. Next, the trajectory object is initialized (line 7). Finally, a function is called that performs the calculation of the minimum energy path by using the NEB method.

neb.py

```
1 traj_path = "neb_traj"
2 N         = 11
3 reac      = XYZFile_ToCoordinates3 ("sub_after_qcmm.xyz")
4 prod      = XYZFile_ToCoordinates3 ("in1_after_qcmm.xyz")
5 GrowingStringInitialPath (mol, N, reac, prod, traj_path, log = logFile)

7 traj      = SystemGeometryTrajectory (traj_path, mol, mode = "a+")

9 NudgedElasticBandSplineOptimize_SystemGeometry(mol, traj, log =
     logFile, maximumIterations = 5000, rmsGradientTolerance = TolCrit)
```

The script **cube.py** can be used for the generation of volumetric data containing spin density. The initialization of the QC/MM model follows the same procedure as in the previous scripts. After the scratch space has been set up, a check is performed (lines 3-5)

whether there exists a file containing the parameters of the cube. If so, the parameters in the form of Python commands are loaded. Otherwise, the geometrical center of the QC-region is calculated (lines 7-13). In lines 15-23, a distance is calculated from the center to the furthermost atom. This distance is used to determine the dimensions of the cube (lines 25-31). The calculated cube is 25% larger than the one bonding the QC-region. Lines 33-35 define the resolution (in Å) and evaluate the number of nodes in each direction. In the following lines, an XYZ file is generated for visualization purposes containing the cube vertices. The dimensions are converted into Bohrs (lines 52-58) and a DEF file is written (lines 60-63). Finally, the parameters are passed to the ORCA object (lines 65-70). A single point energy calculation (line 75) generates a cube file in the scratch directory.

cube.py

```
1  cubef = "cube.def"

3  if exists (cubef):
4    f = open (cubef).readlines ()
5    for c in f: exec (c)

7  else:
8    C = Vector3 ()
9    for q in qc:
10     Q = mol.coordinates3.GetRow (q)
11     C.AddScaledVector3 (1., Q)
12   C.Scale (1. / len (qc))
13   cx, cy, cz = C

15   R = 0.
16   V = Vector3 ()
17   for q in qc:
18     Q = mol.coordinates3.GetRow (q)
19     C.CopyTo (V)
20     V.AddScaledVector3 (-1., Q)
21     r = sqrt (V.Dot (V))
22     if r > R:
23       R = r

25   Scale = 1.25
26   minx  = -R * Scale + cx
27   maxx  =  R * Scale + cx
28   miny  = -R * Scale + cy
29   maxy  =  R * Scale + cy
30   minz  = -R * Scale + cz
31   maxz  =  R * Scale + cz

33   Res   = 0.10
```

```
34   t     = (maxx - minx) / Res
35   np    = floor (t)

37   O = open ("cube.xyz", "w")
38   O.write ("9\nCube in Angstroems\n")
39   for a in (
40      (cx,   cy,   cz),
41      (minx, miny, minz),
42      (minx, miny, maxz),
43      (minx, maxy, minz),
44      (minx, maxy, maxz),
45      (maxx, miny, minz),
46      (maxx, miny, maxz),
47      (maxx, maxy, minz),
48      (maxx, maxy, maxz)):
49     O.write ("Xx %6.2f %6.2f %6.2f\n" % a)
50   O.close ()

52   toBohr = 1. / 0.5291772
53   minx   = minx * toBohr
54   maxx   = maxx * toBohr
55   miny   = miny * toBohr
56   maxy   = maxy * toBohr
57   minz   = minz * toBohr
58   maxz   = maxz * toBohr

60   f = open ("cube.def", "w")
61   f.write ("# Cube parameters in a.u.\n")
62   f.write ("minx = %6.2f\nmaxx = %6.2f\nminy = %6.2f\nmaxy =
        %6.2f\nminz = %6.2f\nmaxz = %6.2f\nnp   = %6d\n" %  (minx, maxx,
        miny, maxy, minz, maxz, np))
63   f.close ()

65 qcm = QCModelORCA ("B3LYP:6-311++G (2d,2p)", "PAL4", "SCFCONV10",
      "\n%%plots\n dim1 %d\n dim2 %d\n dim3 %d\n min1 %f\n max1 %f\n min2
      %f\n max2 %f\n min3 %f\n max3 %f\n Format Gaussian_Cube\n SpinDens
      (\"spin_dens.cube\");\n end" % (np, np, np, minx, maxx, miny, maxy,
      minz, maxz),
66   command        = op,
67   scratch        = sp,
68   job            = "job",
69   deleteJobFiles = False
70 )
71 mol.DefineQCModel (qcm, qcSelection = Selection (qc))
72 mol.DefineNBModel (NBModelORCA ())

74 mol.Energy ()

76 logFile.Footer ()
```

Apart from the main research topic that is modeling of catalysis by radical enzymes, I have also contributed to the following papers that in general deal with the problem of enzymatic catalysis:

1. Wiktor Beker, Karol M. Langner, Edyta Dyguda-Kazimierowicz, Mikolaj Feliks, W. Andrzej Sokalski **"Predicting relative stabilities of biomolecular complexes from electrostatic multipole interactions at shortened intermolecular distances"** *J. Comput. Chem. 34*, 1797–1799 (2013)

2. Renata Grzywa, Edyta Dyguda-Kazimierowicz, Michal Sienczyk, Mikolaj Feliks, W. Andrzej Sokalski, Jozef Oleksyszyn **"The molecular basis of urokinase inhibition: from the nonempirical analysis of intermolecular interactions to the prediction of binding affinity"** *J. Mol. Model. 13*, 677–683 (2007)

3. Jing Wang, Juo Gu, Jerzy Leszczynski, Mikolaj Feliks, W. Andrzej Sokalski **"Oxime-induced reactivation of sarin-inhibited AChE: A theoretical mechanism study"** *J. Phys. Chem. B 111*, 2404–2408 (2007)

I declare that this thesis titled, "Computational Modeling of Catalytic Mechanisms of Glycyl Radical Enzymes" and the work presented in it are my own. I confirm that:

- This thesis was prepared entirely while in candidature for a Ph. D. degree at the University of Bayreuth.
- This thesis has not been submitted previously for a Ph. D. degree at any other institution.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Bayreuth, January 2014