Atomistic simulations of the reactive processes in the heme-containing proteins

Inauguraldissertation

zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von
Maksym Soloviov
aus der Ukraine

Basel, Juni 2017

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von:

PROF. DR. MARKUS MEUWLY

PROF. DR. O. ANATOLE VON LILIENFELD

Basel, den 23. Juni 2015

Prof. Dr. Jörg Schibler
Dekan

Having a place to go - is a home.
Having someone to love - is a family.
Having both - is a blessing.

— D. Hedges

Dedicated to my family and friends.

ABSTRACT

Heme proteins have a great impact in the protein research. Due to the unique electronic properties of heme these proteins are abundant in nature and have a wide range of biological functions in most of the organisms from archea to eukaryotes. The ability of heme proteins to bind and release small molecules like CO, NO, O_2 defines the variety of physiological functions and is related to the structural dynamic properties of the protein matrix surrounding heme.

Cytochrome c oxidase (CcO) is a heme-containing protein, which performs oxygen reduction to water as a part of the membrane complex. Cytochrome c oxidase forms stable complex with CO in the binuclear heme $a_3 - Cu_B$ active site and is a model system to study ligand binding and release. The pump-probe experiments performed for the CcO-CO system reported the ultrafast dynamics of the CO transfer from the heme Fe to Cu_B site. Molecular dynamics simulations are used to provide the dynamic structural information during the transfer with atomic resolution. The kinetics of the process determined from the MD simulations is a qualitative agreement with the timescales reported in the experimental studies. The simulations show that the transfer dynamics is ballistic. The doming of the heme Fe observed after the photoexcitation significantly affects the probability of the heme Fe rebinding.

Myglobin (Mb) is an oxygen storage protein, the active site of which contains heme. It allows to study the impact of the structural changes on binding and release of small molecules. The Mb complex with NO was studied using MD simulations. The heme doming effect observed after photodissociation makes the heme Fe less accessible to NO and slows down the rebinding. The DFT parametrized ^2A state predicts the existence of the Fe-ON minimum, which is not observable in the experiments, this might be explained by the effect of the ^4A PES, that activated during the photoexcitation and that has lower energy for the configurations corresponding to the Fe-ON minimum.

ACKNOWLEDGEMENTS

First I would like to thank my supervisor Prof. Markus Meuwly for giving me the opportunity to work in his group. I appreciate his support and patience, and discussions during all years of my PhD study.

I am very grateful to Juvenal Yosa Reyes, Tibor Nagy, Prashant Gupta, Pierre-Andre Cazade, Juan Carlos Castro Palacio and Jaroslaw Szymczak for the great scientific support and productive discussions. I would like to give special thanks to my "godmother" and my officemate for three years in the group, Franzi Hofmann for taking care of me when I have arrived. I thank Prashant Gupta, Juan Carlos Castro and Juvenal Yosa Reyes for reading parts of dissertation - your comments were extremely helpful. I enjoyed working with the members of the Meuwly group: Tristan Bereau, Ana Patricia Gamiz-Hernandez, Lixian Zhang, Christian Kramer, Manuela Utzinger, Myung Won Lee, Tobias Schmidt, Jing Huang, Vijay Solomon, Florent Hedin, Nuria Plattner, Zhen-Hao Xu, Akshaya Kumar Das, Livia Glanzmann, Pierre-Andre Cazade, Florent Hedin, Mike Devereux, Yonggang Yang, Stephan Lutz, Jaroslav Padevet, Maurus Schmid, Ganna Berezovska, Krystel El Hage, Shampa Raghunathan, Sebastian Brickel, Oliver Unke, Polidefkis Diamantis and Marco Pezella.

I would also like to thank Prof. Juerg Osterwalder, Michael Greif, Matthias Hengsberger, Luca Castiglioni and other members of the Osterwalder group at University of Zürich for fruitful collaboration within NCCR MUST projects.

I want also to thank Prof. O. Anatole von Lilienfeld for being my co-referee for my PhD defense.

I would like to thank my parents Tatyana and Andriy Soloviov, and my wonderful wife Viktoria Safonova for all their love and support, for being with me during this adventure. I want to thank Christoph Schweizer for "das produktive Tandemstudium". And last but not least, I want to thank all my family and friends.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1

INTRODUCTION

1.1 A BRIEF HISTORY OF HEME

Heme-containing proteins are one of the most throughly and intensively studied for a long time. In 1850s the hemoglobins of vertebrates were crystalized, with identification of the substance called "hemin" associated with the protein in 1853, by treating blood with glacial acetic acid and heating [59]. The brownish prismatic crystals were found to contain iron. Soon it was figured out that hemin can be removed from hemoglobin making the original protein colorless. It was found that this colorless protein did not contain iron. The later works in 1880s have shown that the mysterious substance obtained from blood contains around 0.335% of Fe (for protein), that leaded to the estimate of the molecular weight of hemoglobin, which had to be ≈ 16670. For the nineteenth century the molecule was huge.

The first spectroscopic information about heme was reported in 1862, Felix Hoppe-Seyler saw the absorption spectrum of blood in visible region. In the presence of oxygen the spectrum contained two absorption bands at $\approx$ 535 and 560 nm. Some years later George G. Stokes removed the oxygen by reduction using a mixture of $FeSO_4$ and tartaric acid. The previously observed two bands in the absorption spectrum were gone, instead he saw a single broad band, that was roughly in between the two, the original red-brown solution was changing its color to purple. By exposing the solution to the oxidation he was able to observe the return of the original color. This reduction/oxidation could be done multiple times, identifying the oxidized and the reduced form [240].

It was evident to the scientists that hemoglobin of blood can bind oxygen, and that it was a carrier of oxygen in blood. The analysis performed on hemoglobin have showed that it binds a single oxygen molecule per one atom of Fe of hemoglobin, if the pressure of oxygen was high enough [79]. Later the similar relationships were found for the carbon dioxide, by its release using the tartaric acid.[12]

In 1929 the structure of the "hemin" (heme chloride) was revealed, Hans Fischer has reported the synthesis of the molecule [64]. In 1930 for this and other discoveries he got a Nobel Prize in chemistry. The final step to reveal the place of the heme in proteins was an atomisticly resolved models based on the X-ray diffraction experiments. These were done first for myoglobin (1958) [117] and hemoglobin (1960) [201]. The Nobel Prize in chemistry was given to M. F. Perutz and J. C. Kendrew "for their studies of the structures of globular proteins". Since then the research on the heme proteins has progressed, providing new information on the impact of the heme on biological functions and properties of the proteins.

1.2 IMPORTANCE OF THE HEME PROTEINS

The heme-containing proteins is a large subclass of the proteins. The Protein Data Bank contain around 1400 heme protein structure, or 268 heme protein entities [19, 20]. These contain 738 hemes, that take part in thousands biochemical reactions [214].The unique electronic structure of the heme leads to the complex behavior of the corresponding proteins. The heme proteins have evolved along different protein families [159], which have different functions such as electron transfer [77, 169, 231], catalysis (kinases [72]), oxidation (peroxidases, oxidases) [57, 171], ion carriers [82], sensors and messaging [33, 168, 219], storage [62, 275] and transport [5, 227]. Heme proteins a parts of large protein complexes, which perform functions in respiratory chain [42, 196], steroid synthesis [41, 212] or apoptosis (programmed cell death) [74, 78]. The diversity of the processes and the critical importance of these processes in the living organisms, in which heme proteins play a key role, show the importance of the heme protein research.

The experimental methods like high-resolution X-ray diffraction and NMR (e.g. based on nuclear Overhauser effect) have assisted the understanding the properties of heme proteins, by providing structural information at atom resolved level [8, 9, 202]. The structural information revealed the effect of the protein, surrounding heme, on the function, reactivity and selectivity of the active site: coordination geometry, electron donor-acceptor of ligand, coordination saturation explicitly effect the reactivity via axial ligands of the heme; the exterior of the protein affects heme via presence and size of the cavities in the protein, hydrogen-bonding, hydrophilic/hydrophobic

properties, electrostatic potential etc. All these factors contribute the function of the heme proteins.

The importance of the axial ligand can be shown in many studies, for example, by replacing the axial histidine by thiolate of cysteine (found in oxidases) in myoglobin the function of the protein is changed from the reversible oxygen binding to the molecular oxygen activation [1]. In cytochrome c, when replacing axial methionine by histidine, the hemoprotein (cytochrome b_5) is obtained [213], whereas if the methionine is replaced by alanine, the protein would have the oxygen-binding properties, similar to the ones of myoglobin [25]. Catalases have tyrosine as an axial ligand of heme, cytochrome P-450 and chloroperoxidase have cysteine, cytochrome c has methionine. This finding leaded to the advances in the bioinorganic chemistry, leading to its explosive growth in last two decades [13].

The protein environment around the heme active site regulates such properties as $Fe(III)/Fe(II)$ redox potential [245]. The $Fe(II)$ state of the heme Fe in the protein can stabilized by many positively charged aminoacid residues surrounding the active site, therefore increasing the affinity of the heme to bind oxygen [262]. Moving from negative charge distribution in the protein exterior in cytochrome b to the positive one in cytochrome c leads to the increase of the corresponding redox potential. This property is important for the mediation of the electron transfer.

Heme proteins are as one of the most well studied metal containing proteins. The variety of the functions they demonstrate makes them one of the most valuable targets in bioinorganic research for modeling the coordination sites [103]. Synthesis of the new changed types of heme proteins based on the known ones, which occur in nature leads to the further understanding of the reactivity of the heme active site [116, 148]. As heme is synthetically available molecule there were many biomimetic inorganic coordination compounds investigated the effect of the coordination on spectroscopic, electrochemical, electronic and magnetic properties without protein [21, 22, 63, 114]. Coleman *et al* have designed model compounds, which mimic the oxygen binding properties as myoglobin and hemoglobin [43, 46]. These molecules demonstrate the affinities similar to the ones in proteins. The further extension of the synthetic analogues of the heme proteins was done by building a sandwich-like structure, in which the heme between short synthetic peptide chains [16]. These allowed to construct the models for complex active sites like the one of the cytochrome c oxidase

[44, 45].

The advancements in genetics (gene analysis and gene manipulations) [107] and molecular biology (polymerase chain reaction) [135, 223] allowed to introduce point mutations to the genes encoding many of the heme proteins. This allowed to study the effects of the protein environment on the reactivity of heme inside the protein. These techniques helped also understanding the evolution of heme proteins and the adaptation-via-mutation mechanisms. For example, Varadarajan *et al* have shown the effect of the electrostatics in the protein by introducing the charged aminoacids to the active site of myoglobin [256]. Such manipulations assisted the protein redesign, enhancing the reactivity. That even made it possible to construct a active site of cytochrome c oxidase in the myoglobin [121, 234]. The mutants with enhanced reactivity can be also obtained by means of directed evolution [39, 61]. Nowadays heme proteins can be designed, resulting in the specimen able to self-assemble at the interfaces, building oriented materials, material, that respond to the external factors like electric fields, changes in pH, ion strength, concentration of specific molecules and can be used in non-linear optics[233], nanoengineering [215].

The unique properties of heme proteins is a result of flexibility of the heme group, modulated by protein environment. These properties lead to the wide spectrum of the functions which heme proteins have. Structural studies supported with many synthetic experiments allowed to see important features of heme protein active sites. The amount of the structural data makes heme proteins a good candidate to study dynamic processes in protein environment.

1.3 DYNAMICS: DIATOMIC MOLECULES AND HEME PROTEINS.

It was shown that heme is bound to the protein covalently via axial bonding between the heme Fe and one or two aminoacid residues of the protein. One of the ligands is often histidine []. The last sixth position in the coordination sphere of the heme Fe might be occupied by external molecules, often these are small diatomic molecules like O_2, NO and CO [128]. In the biochemical processes, in which these small molecules often take part, the binding or release or exchange of the external or internal ligands might happen. The dissociation of the bond between the heme Fe and

the ligand can be carried out by thermal or photoactivation [145, 181]. The further dynamics of the ligand is largely affected by the dynamics of protein matrix.

Recent advancements in spectroscopy have made it possible to investigate the processes on the femto- and picosecond timescale. This allows to study the dynamics of heme, ligand and protein matrix, as well as the effect of each of them on the overall function of heme proteins, because the processes of reorganization and conformational change in response to the ligand binding or release take usually from tenths femtoseconds to nanoseconds. The bonds between a heme and a ligand are mostly covalent (coordination) in nature and quite highly energetic (> 10 kcal/mol), it means that the probability of breaking such a bond on a picosecond timescale simply by thermal activation is low (it happens on the millisecond to second timescale). Therefore the methods to induce the heme-ligand bond dissociation are required in order to study ligand dynamics. It can be done using photoexcitation of the heme system. The photoexcitation of heme does not occur at physiological conditions, but it mimics thermal activation and can be done synchronously for the whole sample, if the photoexcitation reaction has high enough quantum yield. On the other hand the small molecules like CO and NO are good probes to sample the protein environment, because their absorption spectra are affected by the binding/release events, they absorb in the part of the spectrum, where they are not likely to interfere with the signals of the protein matrix and medium of solution/crystal etc (e.g. CO absorption is $\approx 2050 - 2160$ cm^{-}1 in vacuum and ≈ 1900 cm^{-}1 when bond to heme [145, 181]). This way the pump-probe experiments are constructed: a short pulse (pump, usually in visible light range), that excites heme system and causes the heme-ligand bond breaking, is followed by the nonperturbing pulse (probe), which determines the state of the ligand (infrared, visible, Raman scattering or even X-ray diffraction). By varying the delay between the pump and pulse a time-resolved (dynamic) information can be obtained! Now the pump-pulse experiments are routinely used to study ultrafast dynamics in proteins[264].

This dissertation is about revealing the structural features of the reactions in heme proteins using MD simulations. It is an attempt to give a atomistic resolved interpretation of the spectroscopic observation obtained in the pump-probe experiments from structural dynamics point of view. The two case studies are discussed in the following chapters: cytochrome c oxidase - CO (Chapter 3) and myoglobin - NO (Chapters 4 and 5).

1.3.1 *Myoglobin*

Myoglobin was the first protein, the structure of which was identified using the X-ray diffraction method [117]. It is a oxygen storage heme protein in mammals [70]. It is one of the most intensively studied heme protein, which is used as a model system for new experimental and theoretical methods and techniques. The most commonly used form of myoglobin is its complex with CO [40, 128]. The CO in this complex is almost perpendicular to the heme plane [40, 113]. In the pump-probe experiments the myoglobin-CO (MbCO) complex has a large photodissociation quantum yield. After the CO release on photoexcitation in several hundreds femtoseconds CO can be found in two similar configurations and is located close to the heme [86, 230]. These configurations were identified and described by combining a selective point mutations of the aminoacid residues in the distal pocket and IR experiments [145, 181]. MbCO was the first protein system studied using time-resolved X-ray crystallography [237, 238]. The docking site described above was found to be located 2 Åaway from heme, which is in agreement with the previous spectroscopic studies, in wild type myoglobin this site is occupied by CO for $\approx$ 200 ns [6]. CO may also transfer to the so-called Xe_1 pocket (Xe pockets are the cavities in the pocket occupied by Xe atoms, when the crystals of Mb are pressurized in Xe P $\approx$ 150 atm) located next to the distal pocket of heme. It is occupied on the microsecond timescale [6] and CO may leave the protein from the Xe_1 pocket. The time resolution of these X-ray experiments was 150 ps. The very long timescales (hundreds of nanoseconds) make modeling of the pump-probe experiments very hard using MD methods.

NO is much more reactive referring to the heme Fe, and was found to rebind the heme Fe in several hundred picoseconds after photoexcitation in pump-probe experiments. This gives an opportunity to investigate the myoglobin-NO (MbNO) system using MD simulations, providing structural interpretation to the experiments. The kinetics of NO rebinding is largely affected by the heme dynamics and protein matrix was found to be multiexponential. Kim and Lim have identified three distinct states using time-resolved infrared spectroscopy techniques [173]. The NO rebinding in their study is 5-90 ps. The NO rebinding to heme is considered to be barrierless process.

The X-ray structure of the MbNO system has shown that NO binds to the heme Fe via nitrogen. But the DFT calculations for a truncated model system predict the possibility of existence of the inverse configuration (Fe-ON) [187], which can be

found in the synthetic analogs of the MbNO system [30, 35, 277]. Although so far this configuration was not experimentally observed in MbNO [182].

One of the important structural changes to the heme after photodissociation is a so-called heme doming [126], a motion in which the heme Fe atom moves out of the heme plane. The recent experiment report that doming of Fe occurs within 0.5-1 ps after the excitation pulse [127]. Doming largely affects the ability of the heme Fe to rebind NO.

In Chapters 4 and 5 the structural aspects of the NO rebinding kinetics in MbNO after photoexcitation (impact of doming, localization of NO, absence of Fe-ON state) are discussed in detail.

1.3.2 *Cytochrome c oxidase*

Cytochrome c oxidase (CcO) is an example of heme proteins, which are parts of complex enzymatic structures. It is a membrane protein, that performs a terminal stage of respiratory chain for aerobic organisms: four electron oxygen reduction to water is carried out in the heme a_3-Cu_B active site of the protein. In the first stage a dioxygen molecule binds the heme a_3 Fe, CO and NO molecules act as competitors to O_2. NO binds reversibly [28, 47], whereas CO forms a stable complex inhibiting the function of the enzyme. CcO is one of the most well-studied and well-characterized proteins among the other heme copper oxidases [147]. It contains several active sites, the most interesting of which is heme a_3-Cu_B active site, that contains heme a_3 and Cu_B site, which is located only 4.5-5.5 Å. The other active sites are active in electron transfer processes and during proton pumping.

Both heme a_3 and Cu_B can bind external ligands (such as CO), but heme a_3 forms more thermodynamically stable complex. The structure, having two closely located reactive binding sites provides a good opportunity to study ligand transfer processes and binding sites interplay. The complex of the protein with O_2 has a very short lifetime as reduction of oxygen is the main purpose of the protein. Therefore studying the direct transfer of O_2 in the active site involving heme a_3 and Cu_B is non-trivial task, although possible. CO, on the other hand, forms a stable complex with CcO and is well-studied. In case of CcO-CO complex, CO molecule is bound to the heme a_3.

When the heme a_3 is photoexcited, CO is released in less than 100 fs [147, 252] and can bind to the Cu_B. It was experimentally measured that the $Cu_B - CO$ complex is formed after ≈ 1 ps. So fast process suggests that the protein matrix structure assists the transfer reaction, making the process rather ballistic than stochastic (limited by diffusion). J. Treuffet *et al* have carried out a mid-IR pump-probe experiments with 100 fs resolution and were able to observe $Cu_B - CO$ formation in maximum 1.5 ps [252]. Inspired by this study, we have performed the all-atom MD investigation dynamics of this system, that discussed in detail in Chapter 3 of this dissertation.

2

CONCEPTS AND METHODS

2.1 MOLECULAR DYNAMICS

2.1.1 *Theoretical basis for molecular dynamics*

Molecular dynamics is an approach to simulate the dynamics of the particles at the atomistic level, based on the explicit equations of motion. Classical molecular dynamics does it using classical Hamiltonian, which is a function of the coordinates $\mathbf{r}$ of the nuclei and the corresponding momenta $\mathbf{p}$:

$$\mathcal{H}(\mathbf{p},\mathbf{r}) = \mathcal{K}(\mathbf{p}) + \mathcal{U}(\mathbf{r}) = \sum_i \frac{\mathbf{p_i}}{2m_i} + \mathcal{U}(\mathbf{r}) \tag{1}$$

where $\mathcal{K}(\mathbf{p})$ is the kinetic energy, $\mathcal{U}(\mathbf{r})$ is the potential energy, that is independent of time and velocity, p_i is the momentum of particle i, and m_i is the mass of the i-th particle. The microscopic state of a system is defined by a set of $\{\mathbf{r},\mathbf{p}\}$, which can also be considered as coordinates in a phase space.

In order to meet the experimental conditions, the simulation should have the way to connect the microscopic simulations to the measurable macroscopic properties (pressure, temperature). It is done using the statistical mechanics, where the experimental observables are defined based on the averages of the ensemble $\{N, V, T\}$ (N is the number of particles, V is the volume of the system and T is the temperature), via the probability of every state of the phase space (probability distribution $\rho(\mathbf{r},\mathbf{p})$). An ensemble average is calculated over a very large number of states considered simultaneously:

$$\langle A \rangle_{ensemble} = \int d\mathbf{r} \int d\mathbf{p}\, \rho(\mathbf{r},\mathbf{p})\, A(\mathbf{r},\mathbf{p}) \tag{2}$$

where $A(\mathbf{r},\mathbf{p})$ is the observable of interest, $\rho(\mathbf{r},\mathbf{p})$ is the probability distribution (3), which is defined as follows:

$$\rho(\mathbf{r},\mathbf{p}) = \frac{e^{-\mathcal{H}(\mathbf{r},\mathbf{p})/k_B T}}{Q} \tag{3}$$

where Q is a partition function which is defined as an sum over all microscopic states i of Boltzmann factors e^{-E_i/k_BT}, where k_B is the Boltzmann constant and T is the temperature:

$$Q = \int d\mathbf{r} \int d\mathbf{p}\, e^{-\mathcal{H}(\mathbf{r},\mathbf{p})/k_BT} \tag{4}$$

Therefore, although it is possible to calculate the averages of macroscopic observable, one needs to know the Boltzmann probability of each microscopic state $\{\mathbf{r}, \mathbf{p}\}$. The integral Q is computationally hard to calculate, as it is required to know the probability of all possible states within given thermodynamic constraints. Molecular dynamics is an approach to solve the issue by evaluating the motion of a system in the phase space, so that the macroscopic observable is calculated based on the states sampled during the dynamics. In order to obtain a good estimate of the macroscopic property the dynamics must go through all the states of the system, which correspond to the given thermodynamic conditions. It is assumed that if the time which is sufficiently long, the average property A does not depend on the initial conditions for the given thermodynamic constraints and the value of the observable is obtained by integrating the equation of motion of the system. Given the initial set of the positions and momenta $\{\mathbf{r}(0), \mathbf{p}(0)\}$, the evolution of the system in time results in the sets of the instantaneous set of $\{\mathbf{r}(t), \mathbf{p}(t)\}$ for each time t, and the average value of the property A can be evaluated from the trajectory as a time average:

$$\langle A \rangle_{time} = \lim_{\tau\to\infty} \int_{t=0}^{\tau} A(\mathbf{r}(t), \mathbf{p}(t))dt \approx \frac{1}{N} \sum_{i=0}^{N} A(\mathbf{r}, \mathbf{p}) \tag{5}$$

τ is the total time of the trajectory, t is the instantaneous time, N is the number of the instantaneous values of the property A at time each t. In an ideal case the trajectory covers the whole phase space. This assumption is the basis for the ergodic hypothesis which states that the time average is equal to the ensemble average:

$$\langle A \rangle_{time} = \langle A \rangle_{ensemble} \tag{6}$$

So if the system evolve infinitely starting from the initial configuration, all the phase space will be visited. In case of molecular dynamics, the aim is to generate a long enough trajectory in order to satisfy the equality 6. Practically this will hold if the system has been equilibrated and there no sufficient constraints which will prevent

system from exploring the whole phase space in the infinitely long trajectory. The ergodic hypothesis and the conditions are the basis for using molecular dynamics simulations to obtain the estimate of the macroscopic observables.

2.1.2 *Classical (Newtonian) molecular dynamics simulations*

The molecular dynamics approach is based on the Newton's second law of motion.

$$\mathbf{F_i} = m_i \mathbf{a_i} = m_i \frac{d^2 \mathbf{r_i}}{dt^2} \tag{7}$$

where F_i is a force which is applied to the i-th particle, m_i is a mass of the i-th particle, a_i and r_i are an acceleration and a position of the i-th particle respectively. Therefore it is possible to calculate the acceleration of each of the particles of the system, if the force applied to the each of the particles is known. The time evolution of the system result in a trajectory which is obtained by integrating the equations of motion with the positions, velocities and accelerations, and then the trajectory can be used to calculate the time averaged properties.

The force acting on each of particle can be calculated from the potential energy $U(\mathbf{r})$:

$$\mathbf{F_i} = -\nabla_i U(\mathbf{r}) \tag{8}$$

The equations 7 and 8 combined result in:

$$m_i \frac{d^2 \mathbf{r_i}}{dt^2} = -\nabla_i U(\mathbf{r}) \tag{9}$$

Considering the classical Hamiltonian $\mathcal{H}$ (1), the Newtonian equations of motions can be rewritten as:

$$\frac{d\mathbf{r_i}}{dt} = \frac{d\,\mathcal{H}(\mathbf{r},\mathbf{p})}{d\mathbf{p_i}} = \frac{\mathbf{p_i}}{m} \tag{10}$$

$$\frac{d\mathbf{p_i}}{dt} = -\frac{d\,\mathcal{H}(\mathbf{r},\mathbf{p})}{d\mathbf{r_i}} = \mathbf{F_i} \tag{11}$$

The Equation 10 is a first-order differential equation and defines a momentum of the system. The Equation 11 is the Newton's second law. These Newtonian equations of motion have properties, such as: conservation of energy, conservation of the momenta and time-reversibility.

2.1.2.1 *Conservation of energy*

In order to conserve energy in Newtonian mechanics the forces, which are applied to the particles, have to be conservative, that means that the $\partial \mathbf{F}/\partial t = 0$ or in other words the integral of $\mathbf{F}$ is within the interval is independent from the path and depends only on the limits of the integral.

2.1.2.2 *Conservation of momenta*

If we consider the system of two particles A and B in a conservative force, the internal forces defined by the potential change corresponding to the change in momenta during time t in such a way that $\delta p_{A\to B} = F_{A\to B} t$ and $\delta p_{B\to A} = F_{B\to A} t$, where $F_{A\to B}$ is the force which act on B from A and $F_{B\to A}$ is the force which act on A from B. According to the Newtonś third law the following equality should hold: $F_{A\to B} = F_{B\to A}$, this leads to the $\delta p_{A\to B} = -\delta p_{B\to A}$. So in the conservative force field the momentum is conserved.

2.1.2.3 *Time reversibility*

This property states that by only changing the velocities the system should follow the reverse trajectory [140]. The Newtonian equations of motions are time-reversible, however in practice this property holds for a finite number of steps of the numerical trajectory due to the error introduced by the integration schemes.
In order to solve these equations of motion the integration of the Equations 10 and 11 is required. Using finite differences method is a standard approach to solve this type of a problem. In the molecular dynamics it means that the current positions and velocities (at time t) are used to evaluate the next positions and velocities (at time $t + \delta t$). The trajectory is obtained by the iterative solutions of the corresponding equations of motion. The choice of the time interval of the step (time step) of the simulation δt is adjusted according to the studied processes. It is important to note that this interval should be significantly smaller than the period of the fastest motion or normal mod in the system, otherwise a special treatment has to be applied accordingly. The smaller is the time step, the more accurate is the integration of the equations of motion. Normally in all-atom molecular dynamics the fastest normal modes correspond to the bonds with a hydrogen. For example, O-H stretch has frequency of approximately 3600 cm^{-1}, this corresponds to $\approx 1.08 \times 10^{14}$ Hz, therefore a period of this vibration is 9.26×10^{-15} s, or 9.26 fs. Most of the all-atom MD simulations have a time step of the numerical integration of 1 fs. The bigger time steps will affect the trajectory. The time step sets the limits to the overall simulations, a 1 ns simulation will require 10^6 evaluations,

whereas many processes of interest, in particular protein folding, protein-protein complex formation, diffusion through membrane, transport through ion channels, happen in a much longer timescales: micro- or milliseconds, or even seconds. In most cases the timescales of millisecond and second are not feasible for the recent computer architectures, however microsecond simulations can be obtained using the graphical cards [206] or specialized hardware [232].
The solution to the problem of the numerical integration is solved using the Taylor expansion of position $\mathbf{r}$:

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \frac{d\mathbf{r}(t)}{dt}\delta t + \frac{1}{2}\frac{d^2\mathbf{r}(t)}{dt^2}\delta t^2 + O(\delta t^3) \tag{12}$$

where $O(\delta t^3)$ includes all the term of the order of t^3 or greater. Considering that the second term of the expansion correspond to the velocities $\mathbf{v} = \frac{\partial \mathbf{r}(t)}{\partial t}$ and the third is an acceleration, then the Equation 12 can be rewritten as

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \mathbf{v}(t)\delta t + \frac{1}{2}\mathbf{a}(t)\delta t^2 + O(\delta t^3) \tag{13}$$

As the acceleration can obtained from the Newton's second law (Equation 7) then we get

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \mathbf{v}(t)\delta t + \frac{1}{2}\frac{\mathbf{F}(t)}{m}\delta t^2 + O(\delta t^3) \tag{14}$$

The same applies to the velocities

$$\mathbf{v}(t+\delta t) = \mathbf{v}(t) + \frac{\mathbf{F}(t)}{m}\delta t + O(\delta t^2) \tag{15}$$

Therefore given the positions $r(t)$ and the initial velocities $v(t)$ one can approximate the positions $r(t+\delta t)$ and the velocities $v(t+\delta t)$ using the Equations 14 and 15. The following contains the information on the particular implementations of the integration algorithms.

2.1.2.4 *Verlet integrator*

The Verlet integrator algorithm [259] is derived from two Taylor expansions for a forward $(t+\delta t)$ and a backward $(t-\delta t)$ steps. The core idea behind it is that the equations of motion are time-reversible:

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \mathbf{v}(t)\delta t + \frac{1}{2}\frac{\mathbf{F}}{m}\delta t^2 + O(\delta t^)} \tag{16}$$

$$\mathbf{r}(t-\delta t) = \mathbf{r}(t) - \mathbf{v}(t)\delta t + \frac{1}{2}\frac{\mathbf{F}}{m}\delta t^2 - \frac{1}{3!}\mathbf{b}(t)\delta t^3 + O(\delta t^4) \tag{17}$$

The combination of the two leads to:

$$\mathbf{r}(t+\delta t) = 2\mathbf{r}(t) + \frac{\mathbf{F}}{m}\delta t^2 + O(\delta t^4) \approx 2\mathbf{r}(t) + \frac{\mathbf{F}}{m}\delta t^2 \tag{18}$$

Therefore the algorithm contains the error of δt^4.
In the Verlet algorithm, first, the current position $\mathbf{r}(t)$ is used to calculate the forces $\mathbf{F}(t)$, then the current ($\mathbf{r}(t)$) and the previous ($\mathbf{r}(t-\delta t)$) positions together with the forces $\mathbf{F}(t)$ are used to evaluate the next positions (Equation. 18). Also the velocities are not be computed explicitly and are found from first order central difference:

$$\mathbf{v}(t) = \frac{\mathbf{r}(t+\delta t) - \mathbf{r}(t-\delta t)}{2\delta t} \tag{19}$$

The Verlet integrator is straightforward and has good stability.

2.1.2.5 *Leap Frog integrator*

The leap frog algorithm [91] solves the problem of the evaluation of the velocities by first calculating them at time $t + \frac{1}{2}\delta t$, and then use them to compute the positions:

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \mathbf{v}(t+\frac{1}{2}\delta t)\delta t \tag{20}$$

$$\mathbf{v}(t+\frac{1}{2}\delta t) = \mathbf{v}(t-\frac{1}{2}\delta t) + \frac{\mathbf{F}(t)}{m}\delta t \tag{21}$$

In this procedure, first, the velocities are computed half-step ($\frac{1}{2}\delta t$) ahead the positions, so the velocities "leap" over the positions, then the positions are computed half-step ahead. The current velocity $\mathbf{v}(t)$ is approximated by the relationship:

$$\mathbf{v}(t) = \frac{1}{2}\left(\mathbf{v}(t-\frac{1}{2}\delta t) + \mathbf{v}(t+\frac{1}{2}\delta t)\right) \tag{22}$$

If the velocities are removed from Equation 20 and Equation 21, then the leap frog scheme becomes identical to the Verlet algorithm. The advantage of the leap frog algorithm is that the velocities are explicitly calculated, however, they are not calculated at the same time as the positions.

2.1.2.6 *Velocity Verlet integrator*

The Velocity Verlet algorithm [244] is an improvement to the original Verlet algorithm which is intended for handling errors caused by evaluation of the velocities, and, in contrast to the leap frog both velocities and positions are calculated at the same point of time. So positions and velocities are derived using the following equations,

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \mathbf{v}(t)\delta t + \frac{1}{2}\frac{\mathbf{F}(t)}{m}\delta t^2 \tag{23}$$

$$\mathbf{v}(t+\delta t) = \mathbf{v}(t) + \frac{1}{2}\left(\frac{\mathbf{F}(t)}{m} + \frac{\mathbf{F}(t+\delta t)}{m}\right)\delta t \tag{24}$$

First the current force is computed, after that the velocities at mid-step are calculated:

$$\mathbf{v}(t+\frac{1}{2}\delta t) = \mathbf{v}(t) + \frac{1}{2}\left(\frac{\mathbf{F}(t))}{m}\right)\delta t \tag{25}$$

Then the new positions are evaluated:

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \mathbf{v}(t)\delta t + \frac{1}{2}\left(\frac{\mathbf{F}(t))}{m}\right)\delta t^2 \tag{26}$$

And finally the velocity is updated:

$$\mathbf{v}(t+\delta t) = \mathbf{v}(t+\frac{1}{2}\delta t) + \frac{1}{2}\left(\frac{\mathbf{F}(t+\frac{1}{2}\delta t))}{m}\right)\delta t \tag{27}$$

Both Velocity Verlet and Leap Frog algorithms are similar up to the half-timestep shift present in the Leap Frog.

2.1.3 *Thermostats and barostats*

Based on the equations of motions shown above the total energy of the system as well as the time averages obtained in such MD setup correspond to the ones of microcanonical ensemble (NVE), which describes a microstate. In order to go beyond the limitations of the NVE ensembles, the notions of the thermostat and/or barostat have to be introduced, which describe the NPT and NVT ensembles. The thermostats

are needed to match the experimental conditions, study the temperature/pressure-dependent processes, to simulate non-equilibrium systems or to explore phase space by means of temperature (simulated annealing, metadynamics).

Thermostats include the modifications of the Newtonian scheme, which are applied in order to mimic the canonical ensemble. Technically they have to modulate the energy flow through the boundaries of the system. The constant temperature simulations introduce the thermostat based on the temperature of the particle relation to the time average of the velocity of the particles:

$$\langle \sum_{i=1}^{N} \frac{1}{2} m_i v_i^2 \rangle = \frac{3}{2} N k_B T \tag{28}$$

where m_i, v_i are individual masses and velocities, N is number of particles, k_B is a Bolzmann constant and T is temperature. When the MD simulation is started, the initial velocities are often assigned (if not available) randomly based on the Maxwell-Boltzmann distribution at the temperature T. During the simulation the mechanism of the thermostat, should control the temperature via manipulating the velocities. The simplest approach to introduce the thermostat action would be by scaling the velocities by a factor λ (so called strong coupling [276]):

$$\lambda = \sqrt{\frac{T_{required}}{T_{current}}} \tag{29}$$

where $T_{required}$ is the given desired temperature which has to be kept and $T_{current}$ is current temperature of the system (28). The problem of this approach is that it does not represent the statistical ensemble: the temperature fluctuations are not properly reproduced and there is no localized correlation as all the velocities of the particles are scaled by the same value.
The another approach to mimic the thermostat, which was published by Berendsen and co-workers [17], is to introduce a parameter τ, that couples the velocity scaling to an external bath:

$$\lambda = \sqrt{1 + \frac{\Delta t}{\tau}\left(\frac{T_{required}}{T_{current}} - 1\right)} \tag{30}$$

where Δt is a time step. The basic idea is that the temperature of the system is adjusted, the deviation decays exponentially. The algorithm is therefore deterministic,

although it is not canonical, it shows minor deviations from canonical distribution (for large systems). As for $0 < \tau < \infty$ it underestimates temperature fluctuations, the approach is also called weak-coupling.

In 1980 Andersen has proposed the stochastic-coupling approach [3]. In this algorithm the equations of motion (12) is integrated every step with the modification such that the velocities of all atoms in the system are reassigned based on the Maxwell-Boltzmann distribution. The criterion for the selection of the particles is that the time between consecutive reassignments of the particle velocities τ are randomly chosen based on the probability p, $p(\tau) = \alpha e^{-\alpha\tau}$, the parameter α is a reassignment frequency. This thermostat mimics the stochastic collisions of the random atoms with the virtual particles of the bath of the thermostat with the temperature T. It might be interpreted as the number of the short microcanonical simulations with small energy fluctuations due to the collisions with these virtual particles. This leads to the canonical distributions of such microstates. The pros of this approach is that it samples from the canonical ensemble, as each particle is coupled to a local heat bath, the heat trapped in localized modes can be efficiently removed, and it allows larger time steps. The cons of the method are due to the loss of the momentum because of the randomness (e.g. the diffusion coefficients can not be calculated).

The most popular set of the methods are the extended Lagrangian methods, which introduces the additional degrees of freedom, introduced by Nosè [184, 185] and then extended by Hoover [95]. The bath is introduced into the system by means of the additional degree of freedom s, which has a fictitious mass Q and potential energy $gk_BT_{required}\log(s)$, where g is a number of degrees of freedom (effectively for a single thermostat is $3N + 1$, $3N$ comes from the original system and one is the additional coordinate of the thermostat), N is a number of particles, k_B is a Boltzmann constant. The Hamiltonian in this case can be written as:

$$H = \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i s^2} + \sum_{i=1}^{N}\sum_{i=1}^{N} U(\mathbf{r}_i) + \frac{\mathbf{p}_s^2}{2Q} + gk_BT_{required}\log(s) \tag{31}$$

where $\mathbf{p}_i$ and $\mathbf{r}_i$ are the momenta and the positions of the particles, $U(\mathbf{r}_i)$ is the potential energy (the other variables are discussed above). The effective mass Q is associated with s and determines thermostat strength and, therefore, the coupling and the energy flow balance between the bath and the simulated system. Also the

parameter Q may be considered as coupling frequency, the too little overlap between this coupling frequency Q and the frequencies in the original system may lead to bad energy transfer time.

The Nosè-Hoover method has shown to produce results that follow the canonical distributions [95, 185, 186]. Generally the original method by Nose [185] considers that the dynamics itself is ergodic, i.e. that trajectory average is a phase space average. However it was shown that for the small or rigid systems this assumption does not apply and the canonical distributions are not generated [95, 250]. In this case a number of the Nose-Hoover chain can be introduced, which are in fact a number of the "thermostated" thermostats immersed one into another [34, 161].

In order to meet the experimental conditions in macroscopic system the canonical (NVT) ensemble does not fit as good as the isothermal-isobaric (NPT) ensemble. In nature the macroscopic systems maintain the constant pressure by changing the volume. However, the calculation of the pressure in the microscopic system of the particles is not as straightforward as in case of temperature, which was discussed above (28). The pressure is related to the kinetic energy of the system and the virial, which is a tensor that is defined as the direct product of the position and force vectors between atoms.

$$P = \frac{2}{3V}\left\langle \sum_{i=1}^{N} \frac{1}{2} m_i v_i^2 \right\rangle - \left\langle \frac{dV}{d\mathbf{r}} \mathbf{r} \right\rangle \tag{32}$$

where m_i, $\mathbf{v_i}$, $\mathbf{r}$ are the masses, velocities and positions of the particles correspondingly, and V is volume of the system. The volume fluctuation $\frac{dV}{d\mathbf{r}}$ is related to the isothermal compressibility β:

$$\beta = \frac{1}{k_B T} \frac{\langle V^2 \rangle - \langle V \rangle^2}{\langle V \rangle} \tag{33}$$

where V and T are the volume and temperature of the system, k_B is Boltzmann constant. The methods of the controlling pressure are similar to the ones used to control temperature. The Berendsen weak coupling barostat uses the scaling of the dimensions [17]:

$$\mu = \left(1 - \frac{\beta \Delta t}{\pi}(P - P_{required})\right)^{\frac{1}{3}} \tag{34}$$

where Δt is an integrator time step, π is a relaxation constant (similar to τ in 30), P and $P_{required}$ are current (32) and required pressure, β is isothermal compressibility (34).

The Nosè-Hoover algorithm may be used by adding an extra degree of freedom, which can be related to the external piston with particular mass, and as well as in the Nose-Hoover thermostat the mass of the piston corresponds to the coupling frequency: the bigger is the mass, the faster is the relaxation towards the required pressure $P_{required}$ [95, 285].

2.2 FORCE FIELDS

In the Newtonian dynamics the potential energy $U(\mathbf{R})$ of the system of molecules or atoms is a function of the positions of the individual atoms and force fields (FFs) are the mathematical models which describe the dependence of the potential energy of the system based on position. A FF combines the analytical expressions of the potential energy and the parameters or coefficients for those expressions. The classical MD does not properly describe the situations, where the features of the system are strongly affected by the quantum phenomena, *e.g.* electron density or spin, and therefore may be applied only to those systems, where the features (even the ones of quantum nature) could be approximated to the classical model. Due to this, the electrons are not present in the FFs and the electronic degrees of freedom are treated implicitly. A FF is typically parametrized based on the QM calculations (*ab initio*, DFT or semi-empirical calculations) or experimental data (various diffraction techniques (X-ray, electron diffraction), NMR, IR).

In the simplest representation the molecule can be approximated to the balls (atoms) connected by springs (bonds). The potential energy of the system is determined by the stretching or bending of the springs-bonds, twisting of the groups of balls-atoms and the interatomic interactions. By default, the force field methods do not have any reference to electrons and therefore the electronic degrees of freedom are neglected, unless the special treatment is used.

The MM which was initially used to get the quantitative information about the chemical systems, when the using of quantum mechanical calculations was unfeasible

due to the technical limitations. The idea of applying FFs to MM was proposed by D. H. Andrews in 1930 [119]: "Mechanical models have been constructed to represent the dynamical systems believed to exist in the molecule". The approach was based on determining the forces between the atoms from the knowledge of the structure and the spectroscopic data like vibrational frequencies, as the vibrational energies and the corresponding normal modes and force constants are related to the properties similar to electronic energies and wavefunction.

The attempts to reveal the structural properties of the condensed systems have been done for the first time in 1940s. Three research groups have created the versions of what is now known as the force field method or simply MM, which is used to calculate molecular structures. In 1946 Hill T.L. has proposed that stretching and bending deformations (quadratic terms) combined with van der Waals interactions (Lennard-Jones 6-12) and Newtonian mechanics can be used to minimize the energies of the system and that this would lead to information regarding structure and energy [88]. Independently in 1946 I. Dostrovsky, E. D. Hughes and C. K. Ingold [56] have used similar approach to explain the steric and polar effects on the rates of the SN_2 reactions of halides. The problem was extremely complex that time. The used method provided the result with very limited accuracy, but it has opened the field for deeper investigation. The same year F.H. Westheimer and J.E. Mayer were investigating the racemization of the optically active substituted diphenyls [218, 273, 274]. This research showed that the MM can be used to explain the relationships between the structure and energies of molecules. In 1976 N.L. Allinger has published the first MM "standard" force field (MM1). This leaded to the development of CHARMM (Chemistry at HARvard Molecular Mechanics) [26, 66, 125, 154–156], AMBER (Assisted Model Building and Energy Refinement) [267, 268], GROMOS (GROningen MOlecular Simulation package) [193], OPLS (Optimized Potential for Liquid Simulations) [112], MMFF (Merck Molecular Force Field) [80] etc.

The aim of the MM is to relate the structural changes (e.g. conformations) to the differences in energy, that also allows to predict the equilibrium geometries and the geometries of the transition states (first-order saddle points). The force field approach uses the Taylor type expansions, extended with the additional terms, that are empirically parametrized. The force fields offer the method of the energy evaluation, which has the lowest computational cost. In the non-spectroscopic MM force fields the energy

of the system is defined as the function of nuclear coordinates (Born-Oppenheimer approximation) and is defined as:

$$U(\mathbf{R}) = E_{bond}(\mathbf{R}) + E_{nonbond}(\mathbf{R}) \tag{35}$$

where $E_{bond}(\mathbf{R})$ describes the system of the covalent bonds and $E_{nonbond}(\mathbf{R})$ is the contribution of the long-range interactions (electrostatic, van der Waals, polarization etc). For example, for a diatomic molecule in vacuum the energy in the ground state ($R \approx R_e$) can be written:

$$\begin{aligned} U(\mathbf{R}) =& U(\mathbf{R}_e) + \frac{dU(\mathbf{R})}{d\mathbf{R}}(\mathbf{R}-\mathbf{R}_e) + \frac{1}{2}\frac{d^2U(\mathbf{R})}{d\mathbf{R}^2}(\mathbf{R}-\mathbf{R}_e)^2 \\ &+ \frac{1}{3!}\frac{d^3U(\mathbf{R})}{d\mathbf{R}^3}(\mathbf{R}-\mathbf{R}_e)^3 + ... \end{aligned} \tag{36}$$

where the $dU(\mathbf{R})/d\mathbf{R} \approx 0$, because in the ground state ($\mathbf{R} \approx \mathbf{R}_e$) and as the FF aim to reproduce the relative energy changes $U(\mathbf{R}_e) = 0$ can be been set to zero. The second derivative term corresponds to the harmonic oscillator and the third and the following derivatives is an anharmonicity term. This results in:

$$U(\mathbf{R}) = k_2(\mathbf{R}-\mathbf{R}_e)^2 + k_3(\mathbf{R}-\mathbf{R}_e)^3 + ... \tag{37}$$

Suppose one is interested analyzing the effects of halide substitution in biphenyls (e.g. chlorine replaced by bromine) using the force field approach and MM, most of the bond lengths and vibrational modes will not be affected by such a replacement. Therefore the force field parameters of the unaffected bonds can be kept and are transferable. Other example are the hydrocarbons. Spectroscopically pentane and hexane are similar and all the C-H bond of the sp^3 carbons in similar position share same parameters when the MM force field is used.

For a complex molecule the multiple terms should be considered and can be written as follows:

$$\begin{aligned} U(\mathbf{R}) =& \underbrace{\sum_{\text{bonds}} U_{\text{stretch}} + \sum_{\text{angles}} U_{\text{bend}} + \sum_{\text{dihedrals}} U_{\text{torsion}}}_{\text{bonded}} \\ &+ \underbrace{\sum_{\text{pairs}} (U_{\text{electrostatic}} + U_{\text{van der Waals}} + ...)}_{\text{nonbonded}} + ... \end{aligned} \tag{38}$$

where the terms come from the bond stretching, angle bending, torsion of dihedral angles and the long range interactions (like) between the pairs of the atoms etc. Hereinafter, the CHARMM force field will be considered. It is widely used for the biologically relevant molecules, like proteins, nucleic acids, peptides, lipids. The CHARMM force field is based on the harmonic approximation, the nonbonded interactions are limited to the pairwise interactions. The energy function is defined as:

$$\begin{aligned} U = & \sum_{\text{bonds}} K_b(b-b_0)^2 + \sum_{\text{angles}} K_\theta(\theta-\theta_0)^2 + \sum_{\text{Urey-Bradley}} K_{UB}(S-S_0)^2 \\ & \sum_{\text{dihedral}} K_\phi(1+\cos(n\phi-\delta)) + \sum_{\text{impropers}} K_\omega(\omega-\omega_0)^2 + \\ & \sum_{\text{nonbonded}} \left\{ \epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 2\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right\} + \\ & \sum_{\text{residues}} U_{CMAP}(\phi\psi) \end{aligned} \tag{39}$$

$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j} \tag{40}$$

$$\sigma_{ij} = \frac{1}{2}\left(\sigma_i + \sigma_j\right) \tag{41}$$

where the first term corresponds to the stretches of the bond, K_b is the force constant and $(b - b_0)$ is the deviation of the given bond length from its equilibrium value b_0. The second term is the treatment of the angular terms: K_θ is a force constant and $(\theta - \theta_0)$ is the deviation of the angle (defined by the 3 atoms, which form 2 bonds between them) from the equilibrium value θ_0. The third term is the Urey-Bradley component, which corresponds to the cross-term, that takes into account the angle bending using the 1-3 nonbonded interactions. The forth term is how the torsion angles are treated in CHARMM: K_ϕ is the force constant, n is the multiplicity (it defines the number the minima), ϕ is the value of the dihedral angle and δ is the phase shift. The fifth term accounts for the out-of-plane bending (e.g. in benzene the H-Ar motion out of plane of the benzene ring can be described by this term), the plane is defined by the triplet of the atoms and the fourth atom is the one, out-of-plane motion of which should be treated: K_ω is the force constant, $(\omega - \omega_0)$ is the deviation from the equilibrium value ω_0 ($\omega_0 = 0$ is a common value, as any non-zero value leads to the out-of-plane positions similar to angular term ($K_\theta(\theta - \theta_0)^2$)). The sixth

term is a pairwise (i, j) nonbonded interactions, which includes a 12-6 Lennard-Jones and Coulomb potentials: r_{ij} is the distance between atoms, ϵ_{ij} and σ_{ij} are the potential well depth and the van der Waals radius values correspondingly, these parameters are pair specific and are obtained by combining the individual values ϵ_i and σ_i of the atoms (40 and 41). The last seventh term is the numerical correction term [26].

2.3 FITTING OF THE MULTIDIMENSIONAL PES

2.3.1 *Force field parametrization*

The accuracy of the MD simulations is largely affected by the accuracy of the FF. Many popular standard FFs, like CHARMM [66, 125, 154–156], AMBER [267, 268], OPLS [112] or GROMOS [193], were optimized to reproduce the structure and the particular properties of biologically relevant molecules, such as aminoacids, proteins, nucleic acids and lipids. Although these FFs contain parameters for major small molecules and monomers, the set offered by the standard library is often limited to few hundreds of entities. Therefore, there is a need to parametrize new chemical species.

In general the PES of the molecule defined, for example in CHARMM FF, by set of the equations 39 and the corresponding parameters. The sources of the experimental data, to which a FF may be fitted, include crystal structures and NMR spectra (e.g. NOESY) (equilibrium bond lengths, valence angles), vibrational, Raman and microwave spectroscopy (force constants from frequencies). In some cases the presence of the experimental data does not assist for parameterization of the FF, as the property is an ensemble average and can be determined only from running the dynamics (e.g. diffusion coefficients) or might be hard to cross-reference the experimental data with the terms of FFs; the bond lengths and angles can be determined from crystal structures and NMR, the van der Waals parameters can be estimated from crystallographic data, but the energy of the torsion terms and the electrostatic potentials are hard to get from the experimental data [183]. In this case these parameters need to be fitted to the proper reference electronic structure calculations. For example, Merck Molecular Force Field (MMFF) is fully parametrized based on the electronic structure

calculations [80].

The accuracy of the estimated parameters of the FF should correspond the FF application. Some of the applications target reproducing of the molecular geometries rather the dynamic (like vibrational spectra) or thermodynamical properties (e.g. heat capacities or enthalpies of formation). When parameterized, the FF should be able to reach the accuracy in reproducing property as the original method, used as a reference in parametrization process. Universality of the FF is often a trade off between the loss of accuracy in ability to reproduce a certain property. The FF is considered to be good for chemistry application, when the root-square error is ≈ 1 kcal/mol. The reproducing of the property is another issue, which has to be dealt with, for example, one of the most widely used water models, TIP3P has melting temperature of 146 K [258], but it is still good to reproduce the solvent environment for protein dynamics.

The non-bonded terms of FFs (van der Waals and electrostatics) determine inter-molecular interactions. As the electronic structure calculations are often done for gas phase, the non-bonded parameters obtained from such a data may not contain the information involving the multi-body interactions, which are observed in condensed phase, where these type of interaction can have a significant contributions to overall energy of the system, it leads then to large errors in quantitative determination of the corresponding properties.

The algorithm of FF parameter fitting can be generalized as follows:

1. Defining the training set (reference) and the objective function: the data may come from both experiment and electronic structure calculations. The objective function is usually the sum of the squares of deviations between the reference dataset and the results from the FF
2. Definition of the functional form of the FF
3. Preparation of the initial guess of the parameters of the FF
4. Optimization of the parameters (minimization of the objective function)
5. Testing the obtained parameters on the validation set

The optimization of the FF may be done in three "dimension": functional form (adding extra terms to the FF, might lead to the improvement of the reproducing the reference data, e.g. multipoles [200], polarization [216]), reference data (increasing quality and the amount of data can improve the general quality of the properties reproducing) and optimization algorithm. The optimization algorithms applied to solve the minimization problem can be gradient-based [183] and non-gradient-based [175, 281], or their combination [183]. Two representative groups of optimization methods: downhill simplex (non-gradient) and gradient descent and simple Newton (gradient-based), are discussed in detail below.

2.3.2 *Downhill simplex algorithm*

The Downhill Simplex (DS) method is a non-gradient-based single-objective optimization approach for searching the space of N-dimensional real vectors. It was developed in 1965 by Nelder and Mead [175]. The DS uses a simplex, a geometric object, that contains $N + 1$ vertex, where N is the number of independent variables. Although the DS method is not gradient-based, it constructs the pseudo-gradient by evaluating enough points in N-dimensional space.

The $N + 1$ vertices form a polytope, which is in fact a polygon in the N-dimensional space: it is a triangle in $\mathbb{R}^2$ (3-points), a tetrahedron in $\mathbb{R}^3$ (4-points) etc. The given initial configuration $N + 1$ defines a simplex. In the method the worst point, the one that has the highest value of the optimized function, gets updated iteratively, this way the simplex can reflect, expand, contract or shrink. The DS algorithm is sometimes called the "amoeba" method, as the these iterative changes of the polytope is similar to the amoeba crawling-like type of movement. All the possible types of updates are shown in Figure 1.

The reflection motion starts with moving the vertex with the highest value of the objective function to a new position, that is a reflection of this point through the other N points. Then the objective function is computed and if its value is lower then previous lower value, the method attempts to expand along the reflection vector more. This defines the expansion motion.

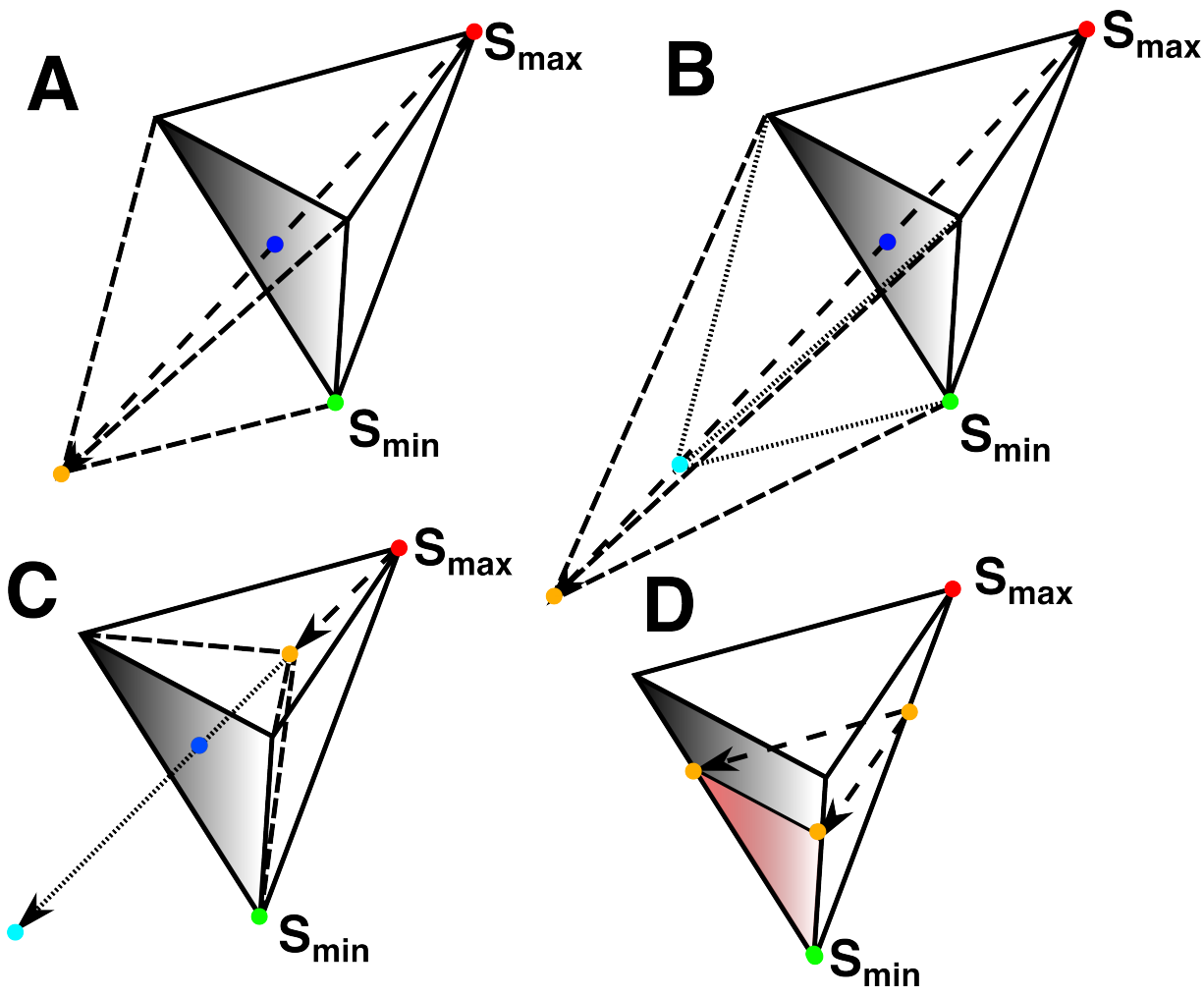


Figure 1: Types of simplex moves: (A) reflection, (B) expansion, (C) contraction, (D) shrinkage

If during the reflection the new point does not lead to the reduction of the objective function the simplex performs a contraction along one dimension from the point which has the lowest value of the objective function. And if in this case the resulting point is worse then the previous, a simplex shrinks along all dimensions towards the point which has the lowest value of the objective function. Iteratively the method finds the local minimum.

The advantages of the algorithm are that it perform 1 or 2 objective function evaluations per iteration, that it does not need a derivative ans that it can work with the objective function which are not smooth. The disadvantage is that it may converge to a non-stationary points.

2.3.3 *Gradient-based methods*

Unlike DS method, in case of gradient-based methods the optimization is performed based on the evaluation of the gradients of the objective function. The objective function can be approximated using the Taylor series around x:

$$f(x+\delta x) = f(x) + \nabla f(x)\delta x + \frac{1}{2}\nabla^2 f(x)\delta x^2 + ... \tag{42}$$

As well as in the DS method the gradient-based optimization algorithms are iterative: given an initial point $\mathbf{x}$, they follow the negative of the gradient of the objective function $-\nabla f(\mathbf{x})$ in order to move from the current position toward a point, which might be a local minimum. The simplest first-order algorithm is steepest descent (gradient descent) method. Given the initial point $\mathbf{x_0}$, the first order approximation of $f(\mathbf{x})$ around $\mathbf{x_0}$ (from Equation 42) is limited to:

$$f(\mathbf{x}) = f(\mathbf{x_0}) + \nabla f(\mathbf{x_0})(\mathbf{x}-\mathbf{x_0}) + \mathcal{O}(\|\mathbf{x}-\mathbf{x_0}\|)^2 ... \tag{43}$$

Suppose it is needed to move from x_0 in a unit direction $\mathbf{u}$ by a step λ. In order to find the proper direction $\mathbf{u}$, we can use the gradient from the Taylor series (43):

$$f(\mathbf{x_0}+\lambda\mathbf{u}) - f(\mathbf{x_0}) = -\lambda\nabla f(\mathbf{x_0})\mathbf{u} + \lambda^2\mathcal{O}(\mathbf{x_0}^2) \tag{44}$$

If the second-order term is omitted by a very small step size λ, the algorithm will perform minimization the fastest at $\nabla f(\mathbf{x_0})\mathbf{u}$. The desired unit direction is found $\mathbf{u} = -\frac{\nabla f(\mathbf{x_0})}{\|\nabla f(\mathbf{x_0})\|}$.

The algorithm is initialized with an initial guess, a maximum number of iterations N_{max}, a tolerance of the gradient absolute value ϵ_g, this is a threshold of the critical point being reached, and a tolerance of value fluctuation convergence ϵ_c in order to determine the change between the iterations is small enough. The algorithm goes as follows:

For $i = 1...N_{max}$

1. set $\mathbf{u}_i = \frac{\nabla f(\mathbf{x}_{i-1})}{\|\nabla f(\mathbf{x}_{i-1})\|}$
2. set $\mathbf{x}_i = \mathbf{x}_{i-1} + \lambda_i \mathbf{u}_i$

3. Is $\|\nabla f(\mathbf{x_i})\| < \epsilon_g$? $\implies$ algorithm converged in stationary point
4. Is $\mathbf{x_i} - \mathbf{x}_{i-1} < \epsilon_c$? $\implies$ algorithm converged in a local minimum
5. Is $f(\mathbf{x_i}) > f(\mathbf{x}_{i-1})$? $\implies$ the algorithm diverges
6. $i == N_{max}$? maximum number of iterations reached

The variable λ should be chosen to balance between the speed of convergence and the risk of divergence. In particular applications it is possible to make a step size λ_i a function of steps or other criteria.

The second-order approach make use of the quadratic term of Taylor series in Equation 42. The quadratic term $\nabla^2 f(\mathbf{x})$ is a Hessian H, therefore by replacing the term we get:

$$f(\mathbf{x}) = f(\mathbf{x_0}) + \nabla f(\mathbf{x_0})(\mathbf{x} - \mathbf{x_0}) + \frac{1}{2}(\mathbf{x} - \mathbf{x_0})^T H_{x_0}(\mathbf{x} - \mathbf{x_0}) \tag{45}$$

The aim is the same as for the SD method, the minimization of the right part of the equation. This way $\nabla f(\mathbf{x}) + H_x(\mathbf{x} - \mathbf{x}) = 0$, if H_x is positive definite, then $\mathbf{x}_{i+1} = \mathbf{x}_i - H_{x_i}^{-1}\nabla f(\mathbf{x_i})$ holds. In order to reduce the risk of the divergence a step size λ_i is added, resulting in:

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \lambda_i H_{x_i}^{-1}\nabla f(\mathbf{x_i}) \tag{46}$$

The Newton method is quadratically convergent near solution. But unfortunately it some issues: it is not always convergent, even if the x_i is close to the stationary point x_s and requires the calculation of the Hessian matrix every iteration. The non-convergence may be illustrated based on 1D case: the equation $f'(x) = 0$, then the iterative algorithm is:

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \frac{f'(\mathbf{x}_i)}{f''(\mathbf{x}_i)} = g(\mathbf{x}_i) \tag{47}$$

In some cases the convergence will not occur, if the initial point $\mathbf{x}_0$ was not chosen to be exact $\mathbf{x}_s$,and will depend on the behavior of the iterative algorithm in the proximity of $\mathbf{x}_s$: if $g'(\mathbf{x}_i) > 1$, the algorithm diverges, $g'(\mathbf{x}_i) < 1$ it converges.

In case of the Hessian calculation, in order to evaluate the step direction $\mathbf{u} = H_{x_i}^{-1}\nabla f(\mathbf{x_i}$ the inverse of the Hessian has to be calculated. The Cholesky factorization is good

for the medium size problems and conjugated gradient approach is applied to large problems. Using the Cholesky factorization one can check if the Hessian H is positive definite, if the contrary is true, then the unit direction switches the sign and the search will continue in the opposite direction.

Another problem is that the inverse of the Hessian matrix has to computed. The inverse will be unstable if the Hessian H is not invertible. In this case an extra factor αI is added, where α is a parameter large enough to guarantee that the resulting form $H + \alpha I$ is positive definite. If α is too large, then the step direction will follow the gradient. For the large datasets the calculation of the Hessian matrix becomes a major factor. In practice this problem is solved by an approximation of the Hessian matrix.

The results of gradient based methods depend on the initial guess provided to the optimization algorithm. One of the strategies to overcome this issue would be running the optimization several times with different initial guesses, especially if there is no a priori knowledge about the behavior of the function (e.g. in simulations).

2.4 KERNEL METHODS

Before now we were considering the parameterization of FFs, that have a known functional form, in which the internal degrees of freedom were explicitly decoupled by means of the superposition. But often the reactive PESs have the coupling of the reactive coordinates and the functional form is not known and the dependence of the coordinates on one another might me highly non-linear and hard to find, especially for the high-dimensional reactive potentials. The kernel-based methods of machine learning become very useful.

The range tasks, in which machine learning methods are applied, may be written as: suppose, there is a given set of data of the size N (x_1, y_1),(x_2, x_2), . . . ,(x_N , y_N), where $x_i \in \mathbb{R}^D$ are inputs and $y_i \in \mathbb{R}$ are outcomes. The task of supervised learning is to find the relation between the inputs and outcomes in order to predict a new outcome y, given a new input x.

2.4.1 *Feature mapping*

Unfortunately in real-world cases there exists numerous problems in which the relationships between the inputs and outcomes is non-linear: it means that in general case the correct mapping might be performed by constructing the feature space $F \in \mathbb{R}^L$, that has extended dimensionality (comparing to the dimensionality of original input space $X \in \mathbb{R}^D, D < L$), in which there exists linear relations and the mapping Φ is such that as $X \xrightarrow{\Phi} F$, e.g. $(x, y) \xrightarrow{\Phi} (x^2, y^2, \sqrt{xy})$. Finding and computing of good feature space is difficult in high dimensional space.

With the proper mapping, one could perform the computation in the good feature space implicitly while working in the original input space without explicit modifications. The feature space should incorporate various nonlinear information of the original data.

The mapping Φ also defines a function space. In this space the norm $\|\cdot\|$ is defined so for the $\psi, \phi \in \Phi$:

$$\|\phi\| \geqslant 0, \|\phi\| = 0 \Leftrightarrow \phi = 0 \tag{48}$$

$$\|\phi + \psi\| \leqslant \|\phi\| + \|\phi\| \tag{49}$$

$$\|\alpha\phi\| = \|\alpha\| + \|\phi\| \tag{50}$$

The norm $\|\phi\|$ can be defined by means of the inner product $\|\phi\| = \sqrt{\langle f, f \rangle}$. The resulting feature space F is a Hilbert space. So if the problem is reduced to linear in the feature space, then the ridge regression problem can be applied to the feature space. The ridge regression is formulated as:

$$\min_w \sum_{i=1}^{N} (y_i - w^T x_i)^2 + \lambda \|w\|^2 \tag{51}$$

where w are the coefficients of solving linear equation, x_i and y_i are input data and corresponding property values and λ is a regularization parameter. The optimal solution $\hat{w}$ of this least squares problem is defined as follows:

$$\hat{w} = X^T(X^TX + \lambda I)^{-1}y \tag{52}$$

where X is a matrix formed by the all inputs, y is a corresponding vector of property values, I is identity matrix. As it was shown the function $\Phi(X)$ performs the mapping of the initial inputs to the extended feature space, therefore the equation 52 applied to the feature space can be rewritten as:

$$\hat{w} = \Phi(X)^T(\Phi(X)^T\Phi(X) + \lambda I)^{-1}y \tag{53}$$

The important here is that $\Phi(X)^T\Phi(X)$ defines the inner product in the feature space, or the measure of similarity, a "distance", between the individual features in the feature space.

2.4.2 *Kernels*

We can avoid computation of the inner products $\langle \phi(x_i), \phi(x_j) \rangle$ explicitly by defining a special function $K(x_i, x_j)$, a kernel, such that:

$$K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle \tag{54}$$

This way a kernel function $K(x_i, x_j)$ computes the inner products in $\mathbb{R}^L$, while operating in $\mathbb{R}^D$, where $L > D$, we can still work in the original input space, implicitly mapping it to the Hilbert space. The main advantage of having a kernel functions is that we limit the optimization problem to the dimensionality of the original input space and not the one of feature space! Hypothetically we can work in the infinite-dimensional feature space.

One might think of a kernel as a similarity measure between x_i and x_j. The main question is how to choose a proper kernel function $K(x_i, x_j)$. It is easy to see that the properties of the kernels are closely related to the ones of the inner product. A kernel should be symmetric:

$$K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle = \langle \phi(x_j), \phi(x_i) \rangle = K(x_j, x_i) \tag{55}$$

Type of kernel	Definition of kernel $K(x_i, x_j)$	Parameters of kernel
Linear	$x_i^T x_j + c$	$c \geqslant 0$ is a free parameter
Polynomial	$(\alpha x_i^T x_j + c)^p$	$c \geqslant 0$ is a free parameters and corresponds to the trade off between the higher and lower polynomial terms, p is a polynomial order, α is a slope
Gaussian	$e^{-\frac{1}{2\sigma^2}\|x_i - x_j\|^2}$	σ is a parameter, that has to be carefully adjusted
Exponential	$e^{-\frac{1}{\sigma^2}\|x_i - x_j\|}$	same as for Gaussian kernel
Laplacian	$e^{-\frac{1}{\sigma}\|x_i - x_j\|}$	kernel is less sensitive to the changes of σ
Sigmoid	$\tanh(\alpha x_i^T x_j + c)$	the kernel is conditionally positive definite, α is commonly $1/N$, where N is a dimensionality of the input data, c is a threshold
Rational Quadratic	$1 - \frac{\|x_i - x_j\|^2}{\|x_i - x_j\|^2 + c}$	c is a threshold. Used instead of the Gaussian, if using Gaussian is too expensive

Table 1: Examples of Kernels

and positive semi-definite:

$$\forall x_i, w_i \in \mathbb{R}, \sum_i^N \sum_j^N w_i w_j K(x_i, x_j) \geqslant 0 \tag{56}$$

The examples of the kernels are shown in Table 1. There is no seemingly magical universal kernel [50, 157], although there are works on the automatic kernel selection [96]. Although some of the kernels are very targeted by structure, in general they do not outperform the other. In practice the proper kernel choice can be done by testing the different kernels and adjusting their parameters, so the error on a test set is minimized. Gaussian and polynomial kernels are good initial starting point [228]. For example, a polynomial kernel can model feature conjunctions via high order of the polynomial. Gaussian, exponential and Laplacian kernels are also known as radial basis functions and allow to model the hyperspheres, linear kernels are good if the feature space is more like a hyperplane.

The kernel is related to the inner product in the feature space and it should be symmetric positive semi-definite function, so technically the kernels might be constructed from the other kernels as following (for kernels K_1 and K_2) [50]:

$$K(x_i, x_j) = cK_1(x_i, x_j) \tag{57a}$$

$$K(x_i, x_j) = c + K_1(x_i, x_j) \tag{57b}$$

$$K(x_i, x_j) = K_1(x_i, x_j) + K_2(x_i, x_j) \tag{57c}$$

$$K(x_i, x_j) = K_1(x_i, x_j)K_2(x_i, x_j) \tag{57d}$$

Kernel method is an approach to solve the problem of the non-linear relations between the input and the outcomes, by solving a linear problem in the extended feature space. This is done by means of the inner products in the feature space while the operations are done in the original space. The kernels perform such mapping between the original input space and the extended feature space at low computational cost.

2.5 REACTIVE FF AND PES

2.5.1 *Reactive Force fields*

Challenges in chemistry are all about chemical changes with molecules, atoms and ions via bonds breaking and formation. The key point of understanding the chemical processes is a revelation of the reaction mechanism and the dynamics of that mechanism when the reaction proceeds. In the equilibrium thermodynamics the molecule samples the configuration space which is close to the local minimum. Temperature, pressure, environment or external factors can perturb the system enough to drive the transition to another minimum and therefore perform a chemical reaction. This process can be investigated using numerous experimental and computational methods, revealing the mechanism and kinetics. The time scales of the most chemical reactions are from picoseconds to hundreds of seconds. Using of some of the computational methods, like DFT and *ab initio* is unfeasible due to the computational costs. The

force fields and molecular dynamics can sample the timescales from femtoseconds to hundreds of milliseconds, making it possible to simulate the processes, which cannot be investigated using other computational methods (especially the kinetics of the processes).

Most of the conventional FFs are based on the harmonic approximation, which means the system is assumed to sample only the configurations, that are close to the minima and cannot describe properly the describe the chemical reaction. Therefore a special treatment has to applied in order to simulate the bond breaking and formation – reactive force fields (RFFs). The RFF have to be parametrized in a different way, comparing to the conventional FFs.

2.5.2 *Empirical Valence Bond Theory*

The first approach to the RFF problem was done by A. Warshel and co-workers, the Empirical Valence Bond method (EVB) [94, 260, 271, 272]. In the EVB method the wavefunction of the system is obtained from a linear combination of the relevant states and the energy is obtained from the the corresponding secular equation. The effective Hamiltonian is constructed based on the parameters, obtained from the empirical data or *ab initio* results. As well as the FFs EVB does not consider electrons or electron density explicitly. In principle EVB method starts with a $N \times N$ potential energy matrix. The N diabatic states are diagonal elements of the matrix, the rest $N(N-1)$ off-diagonal elements are the couplings between the states. The Hamiltonian for two states is defined as:

$$\mathbf{H} = \begin{pmatrix} V_{1,1} & V_{1,2} \\ V_{2,1} & V_{2,2} \end{pmatrix} \tag{58}$$

where $V_{1,1}$ and $V_{2,2}$ are the diabatic states, which are defined based on the non-reactive force field, $V_{1,2}$ and $V_{2,1}$ are coupling elements, which describe the interactions between each diabatic states. After the diagonalization the minimal value is the ground state.

The important point concerning the EVB method is that the off-diagonal, coupling terms are to be parametrized all together in order to describe the reaction path correctly. In the transition state the interactions are controlled by those coupling

terms, these are actually become the reaction coordinate. The great advantage of the EVB method is that at low computational cost, which is lower than the DFT or semi-empirical methods, it produces good quantitative results, when the empirical parameters are properly chosen. This fact was illustrated in the studies on proton transport [226, 266], acid-base reactions in aqueous solutions [257], enzymatic catalysis [269, 270], nucleophilic substitution reactions [15, 176].

The main disadvantage of the EVB method is the due to the definition of the valence bond, the path of the chemical reaction has to be specified by construction. Therefore the EVB method cannot handle the situation, in which it is desired to search for the chemical reaction pathways, which is not in the parametrization.

2.5.3 *ReaxFF*

Another approach to the FF-based reactive processes treatment is ReaxFF. It was developed by Adri van Duin, William A. Goddard, III and co-workers at the California Institute of Technology [254]. As it was shown in the previous cases the bonds, angles, torsions are explicitly defined, when the simulation starts, allowing for only specific reaction sites, the ReaxFF instead uses the definition of bond orders, which leads to the continuous bond formation or breaking.

$$
\begin{aligned}
U(\mathbf{R}) = & \underbrace{\sum_{\text{bonds}} U_{\text{stretch}} + \sum_{\text{pairs}} (U_{\text{electrostatic}} + U_{\text{van der Waals}}}_{\text{2-body interactions}} \\
& + \underbrace{\sum_{\text{angles}} U_{\text{bend}}}_{\text{3-body interactions}} + \underbrace{\sum_{\text{dihedrals}} U_{\text{torsion}}}_{\text{4-body interactions}} \\
& + \underbrace{\sum_{\text{H-bonds}} U_{\text{H-bond}} + \sum_{\text{atoms}} (U_{\text{over}} + U_{\text{under}})}_{\text{additional terms}} \\
& + \underbrace{\sum_{\text{atoms}} U_{\text{lone-pair}} + U_{\text{pen}} + U_{\text{conj}} + \ldots}_{\text{additional terms}}
\end{aligned}
\tag{59}
$$

where U_{stretch} is a bond energy, $U_{\text{electrostatic}}$ are Coulomb interactions are calculated between every atom and the polarizable charges are updated every iteration,

$U_{\text{van der Waals}}$ are the van der Waals terms, which are calculated between every atom, U_{bend} is the angular term, which also depends on the central atom bond order, U_{torsion} corresponds to the energy in the dihedral angles, as well as the angular term, it depends on the bond order of the atoms, $U_{\text{H-bond}}$ is the energy of hydrogen bonds, U_{over} and U_{under} are the overcoordination penalty and the stabilization of the undercoordinated configuration correspondingly (undercoordination stabilization is optional), $U_{\text{lone-pair}}$ penalize the breaking up lone pairs (e.g. of oxygen or nitrogen atoms), U_{pen} is a penalty for allene-type molecules (it is optional), U_{conj} is an stabilization energy gained by the system of conjugated bonds (it is optional). There are many more energy terms, which are specific to the particular cases, e.g. metallic systems, combustion of the hydrocarbons etc.

The significant concept behind the ReaxFF approach is a bond order, which is defined based on the instantaneous configuration: in order to perform a smooth, continuous transition during single, double or triple bond formation the bond length - bond order relationships are introduced for the reacting group of the atoms. Therefore there are no pre-defined reactive sites or reaction pathways, the instantaneous configuration will determine automatically the changes due to the reactions. And the bond orders are updated every integration step of the equations of motion.

The significant difference from the other RFF approaches is that the non-bonded interactions are evaluated for every single pair of the atoms and does not depend on the connectivity between them. At short distances, when the atoms would repel each other due to the Coulomb and van der Waals short-range repulsion, the shielding is applied. All the other "bonded" terms for the multi-body interactions (angles, torsions) are also made bond-order dependent, this way they will change on the changes of the chemical environment. The polarization effects are also explicitly treated via the charge calculation schema, that depends on the configuration.

Another significant difference is that the parametrization is made elementwise, i.e. there is only a single atom type, which corresponds to an element. And the differences between the entities of the element are defined by the chemical environment.

The ReaxFF method has show good results for the reactions in condensed phase [73, 177, 255] (metals included [178]), thermal decomposition [36], combustion

reactions [37, 38, 242].

2.5.4 *Adiabatic Reactive Molecular Dynamics (ARMD)*

FFs, like the one of CHARMM, have the functional form similar to the one in the equation 39, and are the superposition of the individual terms like bond, angles, torsions, non-bonded interactions etc. What happens in the chemical reaction, is the change of the chemical environment, bonds are formed or broken. In ARMD, a number of parametrized PESs, in the simplest case, two (educt and product) are combined. For example, letś consider the case of diatomic with the atoms A and B. When the atoms A and B form a molecule $A-B$ the interaction between them is described using PES V_{prod}, which is a simple Morse potential (similar to the harmonic bonded term in Equation 39):

$$V_{prod}(R) = D_e(1 - e^{-\beta(R-R_e)})^2 \tag{60}$$

where R is the distance between the atoms A and B, D_e is potential well depth, β is the parameter, which characterizes the width of the potential well and R_e is the equilibrium distance. When the molecule $A-B$ dissociates, in the standard FF the atoms A and B will interact with each other by means of PES V_{educt}, a combination of the van der Waals and Coulomb terms (same way as in 39(non-bonded)):

$$V_{educt}(R) = \epsilon\left(\left(\frac{\sigma}{R}\right)^{12} - 2\left(\frac{\sigma}{R}\right)^{6}\right) + \frac{q_A q_B}{kR} \tag{61}$$

Combining the two PESs V_{prod} and V_{educt} should lead to the effective PES. The ARMD solves the problem as follows [52]: suppose the system is initiated in the unbound V_2 state and therefore for a given initial configuration the V_{educt} state is preferable, then each step both of the PESs get evaluated, and, suppose, the two atoms start approach each other and at some point the $V_{prod} < V_{educt}$ condition will hold and V_1 state gets more stable. In this case, the ARMD module stops the simulation and rewinds the configuration of the system τ steps before the crossing. And the simulation restarts from that point, but the PES is mutated from V_{educt} to V_{prod} using a time-dependent switching function:

$$V_{mix} = \frac{1}{2}\left(V_{educt}\left(\tanh[\alpha(t-t_0)]+1\right) + V_{prod}\left(1-\tanh[\alpha(t-t_0)]+1\right)\right) \tag{62}$$

where t is a current time step, t_0 is the time at which the rewind of the trajectory happened and α is a switching parameter, that is related to the time τ. The switching function guarantees the smooth switching and a proper transition from V_2 state to V_1 within $\tau \propto \alpha$ time steps. The functional form of V_{mix} describes a smooth transition. Then the simulation is continued till the conditions of the switching will be met.

The ARMD method introduces an additional variable, a shift Δ_{trans}, which compensates the missing reference energy of transition. This Δ_{trans} parameter is an energy shift of the V_1 PES. The Δ_{trans} may be fitted to the transition enthalpy H_{trans} or free energies of activation $\Delta G^{\ddagger}$, that can be derived from the experimentally measured transition rates [253].

This method allows smooth transition between states, however the general time-dependence of the switching function violates the conservation of energy. Though this problem does not effect the large thermostated systems [52, 191], but may lead to the problems when the multiple transitions might be observed (e.g. hydrogen bonds) and the small systems (especially NVE simulations).

2.5.5 *Multi-Surface Adiabatic Reactive Molecular Dynamics (MS-ARMD)*

The MS-ARMD method solves the issues, which exist in the ARMD approach, precisely, the energy conservation problem [172]. As it was shown above the switching function between a pair of states is time-dependent, this violates the requirement of the Newtonian mechanics regarding the forces to be conservative. The MS-ARMD method states that both the PESs $V_i(\mathbf{R})$ and switching function $w_i(\mathbf{R})$ are the defined for a given configuration $\mathbf{R}$. This way the energy is guaranteed to be conserved during the transition process. The effective PES is the lowest one ($V_{eff}(\mathbf{R}) = \min V_i(\mathbf{R})$), unless the system approaches the crossing. When the configuration of the PESs get close, then the effective PES is handled by the switching function. The parameter ΔV defines the measure of the sensitivity of the switching (similar to the parameter τ in the previous

section 2.5.4). In the MS-ARMD algorithm the effective PES V_{eff} (for N PESs) is a linear combination of the PESs $V_i(\mathbf{R})$:

$$V_{eff}(\mathbf{R}) = \sum_{i=1}^{N} w_i(\mathbf{R}) V_i(\mathbf{R}) \tag{63}$$

where the weights $w_i(\mathbf{R})$ are, in fact, switching functions and defined as:

$$w_i(\mathbf{R}) = \frac{w_{i,0}(\mathbf{R})}{\sum_{i=1}^{n} w_{i,0}(\mathbf{R})} \tag{64}$$

where $w_{i,0}(\mathbf{R})$ are non-normalized weights. The raw non-normalized weights are defined as follows:

$$w_{i,0}(\mathbf{R}) = \exp\left(-\frac{V_i(\mathbf{R}) - V_{min}(\mathbf{R})}{\Delta V}\right) \tag{65}$$

where $V_{min}(\mathbf{R})$ is a current minimum energy PES, and ΔV is a switching parameter, that was discussed above. The information about non-minimum PES is exponentially decaThe value of ΔV is determined during the parametrization.

In the Equation 65, the non-normalized weight $w_{i,0}(\mathbf{R})$ depends on the instantaneous lowest-energy PES, this has a significant impact on the evaluation of the forces as all the weights of the PESs for the non-lowest PESs contains the term, which involves the lowest PES $V_{min}(\mathbf{R})$. The first derivative is defined as follows:

$$\nabla V_{eff}(\mathbf{R}) = \sum_{i=1}^{N} w_i(\mathbf{R}) \cdot \nabla V_i(\mathbf{R}) + \nabla w_i(\mathbf{R}) \cdot V_i(\mathbf{R}) \tag{66}$$

Theoretically, this algorithm can be used for numerous PES. It can also handle the processes, which might have different pathways (e.g. water-elimination in H_2SO_4 [172, 281]).

3

CO TRANSFER IN THE ACTIVE SITE OF CYTOCHROME C OXIDASE

3.1 INTRODUCTION

The dynamics of small ligands in proteins strongly depends on their environment. One of the routinely employed probe molecules is carbon monoxide (CO) which is often used in ligand binding-rebinding experiments [139, 199, 235]. In such experiments the ligand is dissociated from its binding partner, which is typically the iron atom of a heme group, by using a UV/visible laser pulse that promotes the system to an electronically excited and repulsive state. Subsequently, the time-resolved data of the probe molecule allows to follow the time-varying environmental changes by spectroscopic means [4, 282]. The structural interpretation of the data, including the nature of the electrostatic environment [198], the ligand migration pathways and reactivities [180, 194], or the ligand-protein coupling [210], is, however, far from straightforward. Usually, the conclusions that can be drawn from experiment alone about the protein environment are indirect except for circumstances under which spectroscopic and structural signatures can be recorded simultaneously which is only possible in rare cases [229]. Alternatively, validated molecular dynamics (MD) simulations provide the required temporal and spatial resolution to complement experimental investigations and allow to interpret the dynamics at atomic resolution [189, 211].

The ultrafast dynamics of cytochrome c oxidase (CcO) has previously been investigated using a number of experimental techniques [142, 192] which provide information about the protein environment and ligand binding-rebinding reactions [83, 207]. The physiological function of CcO is to reduce molecular oxygen to water. The protein contains two $a-$type heme groups: heme a (low spin) is hexacoordinated and mediates electron transfer from exogenous cytochrome c towards the active site whereas heme a_3 (high-spin), near the copper atom $Cu_B \approx 5$ Å away from the iron atom of heme a_3, acts as the binding site for molecular oxygen and its reaction intermediates during four-electron reduction. The electron equivalents are

provided by cytochrome c in the outside of the mitochondrial inner membrane (the intermembrane space) via a copper site (Cu_A) and a low-spin heme (heme a) to the O_2 reduction site. The protons used for water formation from O_2 are transferred from the inside of the mitochondrial inner membrane (the matrix space) through two hydrogen-bonded networks known as the K and D pathways. In addition to O_2, heme a_3 also binds other diatomic ligands, including nitric oxide (NO) or carbon monoxide (CO) which are physiologically generated and act as biological messengers [92]. Because heme binds CO tightly, its accessibility to the heme a_3 in CcO must be well controlled.

The active site of interest for the current study is the one where the four electron reduction takes place. It contains heme a_3, bound to the protein via histidine His_{376}, and a three-histidine-coordinated Cu center (Cu_B) located 4.5 to 5.0 Å away from the Fe of heme a_3. Together they form a bimetallic active site. Before binding to heme a_3, ligands such as O_2 and CO bind to Cu_B intermediately. The reverse reaction, involving ligand transfer from heme a_3 out of the protein via Cu_B, can be studied using flash photolysis. CO is known to permanently inhibit CcO by forming a thermodynamically stable complex with the heme a_3-Fe. As in other protein environments, the frequency of the CO stretch is sensitive to the changes in the chemical environment [179, 198, 261]. This property has been extensively used in numerous spectroscopic studies of enzymatic active sites as a probe. In CcO, the CO molecule is known to bind transiently to the Cu_B site before forming a thermodynamically more stable complex with the heme a_3–Fe [139]. The reverse reaction can be induced by photolysis of the Fe-CO bond. Subsequent reactions involve the transfer of CO to the Cu_B site which occurs on the 1 to 2 picosecond time scale [252].

In order to better understand the atomistic details underlying CO transfer dynamics in the bimetallic center of CcO, molecular dynamics (MD) simulations together with a parametrized force field are employed. Such simulations are capable of characterizing the ligand transfer trajectories and relate structural and temporal aspects of the process. One of the open questions is whether free CO is formed at all and on what time scales it is expected to exist under such circumstances. One suitable observable to compare with is the infrared spectroscopy of CO which changes characteristically between metal-bound and free CO. Furthermore, it is of interest to explicitly follow the ligand unbinding and rebinding dynamics and to determine the typical time scales for this process which is expected to be on the sub-picosecond to picosecond

time scale. This makes the present system particularly relevant for direct comparison between experiment and simulation because extensive averaging is possible.

3.2 METHODS

3.2.1 *Molecular Dynamics Simulations*

All simulations of Cytochrome C oxidase with bound and unbound CO were carried out using CHARMM [26] with the CHARMM22 force field [156], For CcO the X-ray structure with the Protein Data Bank reference 1AR1 [195] was used. The computational model of the protein contains subunits A (529 amino acids) and B (252 amino acids), 3 copper atoms, 2 heme groups (heme a and heme a_3), calcium and magnesium ions, 9 molecules of lauryldimethylamine oxide, the carbon monoxide molecule and 42 water molecules, which have been found in the X-ray structure. For simulations with Fe-bound CO of heme a_3 the initial Fe–C(CO) bond length was set to 1.90 Å. The crosslink between His240 and Tyr280 (N_ϵ–C_ϵ) known for bovine CcO [115, 195] has been introduced via a CHARMM patch. The two copper atoms, which are not part of the active site of interest are treated by applying harmonic constraints to the bond lengths and the angles between these copper atoms and the protein.

Hydrogen atoms were added using CHARMM and the protein was solvated in a pre-equilibrated water box ($80 \times 117 \times 74$ Å^3) which leads to a total system size of 58092 atoms. We used periodic boundary conditions and the TIP3 model for water [111]. The total charge of the system is $-4.0e$ and was neutralized by replacing random TIP3 waters by potassium ions. Some water molecules were replaced by potassium and chloride ions so that the concentration of KCl in the water box was $\approx$ 0.15 M. The SHAKE algorithm [222, 279]. with a tolerance of 10^{-6} was applied to all bonds which included hydrogen atoms. A time step of 1 fs was used, and the non-bonded interactions were cutoff at 14 Å. First, the system was minimized with the steepest descent algorithm, then heated to 300 K during 300 ps and equilibrated for 700 ps (NVT). Production runs (NVT) were carried out for 5 ns, during which the initial configurations for the excitation were collected every 10 ps.

3.2.2 *Intermolecular interactions*

The global PES $V(\vec{X})$ is a function of the coordinates $\vec{X}$ of all atoms involved. For the particular case of CO transfer between the bimetallic site in CcO we write the total energy as

$$V_{tot}(\vec{X}) = V_{FF}(\vec{Q}) + V(r, R, \theta) \tag{67}$$

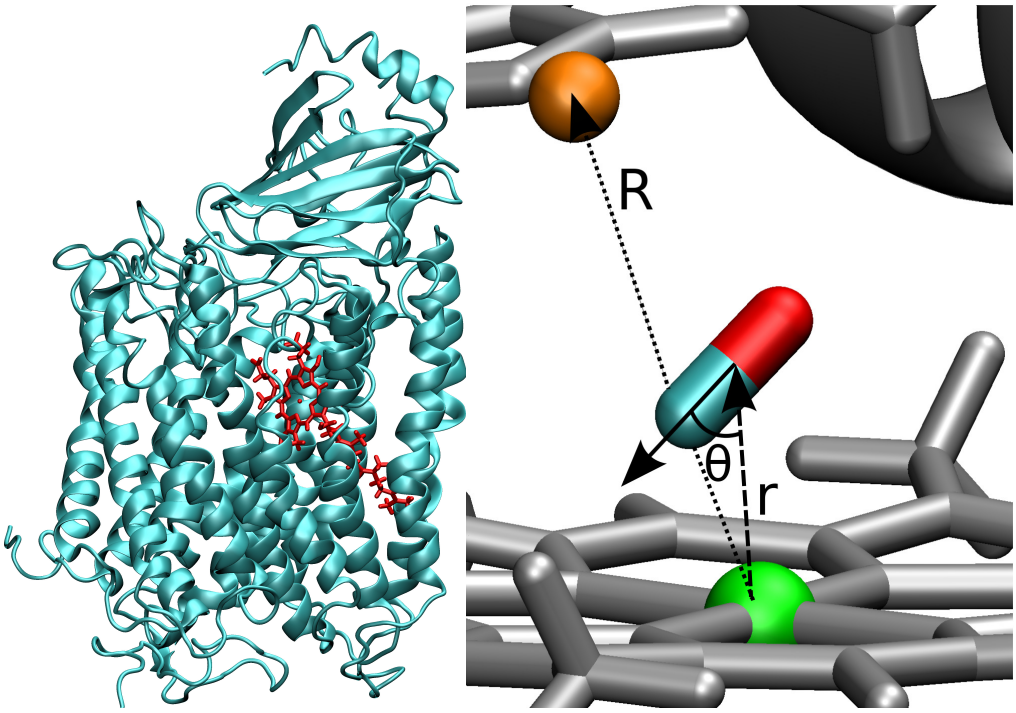
The force field $V_{FF}(\vec{Q})$ is the standard CHARMM22 force field [155] and $\vec{Q}$ contains all configurational coordinates of the system except R, ρ and θ (see below) which describe the coordinate dependence of the PES involving the binuclear site and the ligand in the electronic ground state, see Figure 13. The coordinates of the subsystem are $R = |\vec{R}|$ the distance between the Fe atom of heme a_3 (labeled Fe_{a_3} in the following) and the Cu_B site, the vector $r = |\vec{r}|$ between Fe_{a_3} and the center of mass of CO, and θ which is the angle between $\vec{r}$ and the vector along the CO-bond. Thus θ corresponds to the CO-rotation. In the ligand-bound state the CO stretching coordinate was parametrized as a harmonic potential with a force constant of 1333.1 kcal/mol and an equilibrium separation of 1.128 Å, the charges are $+0.021$ and -0.021 for C and O, respectively [155, 205].

For the 3-dimensional PES $V(r, R, \theta)$ more than 3300 reference energies were calculated using density functional theory (DFT) calculations with the B3LYP [14, 136, 239, 265] functional and the 6-31G(d,p) basis set [203, 204]. All electronic structure calculations were carried out with Gaussian09 [Frisch et al.]. We found it convenient to introduce the dimensionless coordinate ρ defined as $\rho = \frac{r-\sigma}{R-2\sigma}$ where $\sigma = 1.5$ Å is the minimal Fe–C and Cu–C distance encountered in DFT scans, high up on the repulsive wall. This transformation maps $\rho \in [0, 1]$ for all separations R which simplifies and stabilizes the fitting [99, 130]. For an analytical representation of the interaction energies the following parametrization is used

$$V(R, \rho, \theta) = \sum_{\lambda=0}^{10} V_\lambda(R, \rho) P_\lambda(\cos(\theta)) \tag{68}$$

where P_λ are Legendre polynomials and the V_λs are double Morse-like potentials:

$$\begin{aligned} V_\lambda(R, \rho) = {} & D_{1,\lambda}(R) \cdot (1 - e^{-\beta_{1,\lambda}(R)(\rho - \rho_{e_1,\lambda}(R))})^2 \\ & + D_{2,\lambda}(R)(1 - e^{-\beta_{2,\lambda}(R)\cdot(\rho_{e_2,\lambda}(R) - \rho)})^2 \\ & - \alpha(R) \end{aligned} \tag{69}$$



(a) (b)

Figure 2: (a) The model of the system (red - active site of interest, cyan - protein backbone. (b) Diagram showing variables of the potential.

Atoms	A, e	B, e/Å	C, e/Å^2	D, e/Å^3
C	−11.011	20.205	−13.149	3.0461
O	−11.254	22.309	−15.981	3.9834

Table 2: Parameters used for 3-point charge model.

Each parameter X_λ, including $D_i(R)$, $\beta_i(R)$, and $\rho_{e_i}(R)$, is fitted to an expression $X_\lambda(R) = \frac{1}{2}p_0 \tanh(p_1(R - p_2)) + p_3$. This parametrization scheme was already found useful for investigating proton transfer between a donor and an acceptor and allows to work with analytical derivatives required for MD simulations [99, 130].

The parameters p_i were determined from least-squares fitting by using a Nelder-Mead simplex [175, Johnson] followed by a non-linear least squares refinement with the I-Nolls software [55, 133].

For the unligated state, the dissociated CO molecule was described by a three-point fluctuating charge model [188]

$$q_i(r_{CO}) = A_i + B_i \cdot r_{CO} + C_i \cdot r_{CO}^2 + D_i \cdot r_{CO}^3 \tag{70}$$

where the q_i are the charges on the C and O atoms, respectively and an additional charge was placed at the center of mass (COM) of CO with

$$q_{COM} = q_C + q_O \tag{71}$$

Such a model was preferred in the present case over a more rigorous multipolar model for computational efficiency [209] The C-O bond potential is an anharmonic and spectroscopically accurate rotational Rydberg-Klein-Rees (RRKR) potential [100, 101, 152]. This model has performed well in previous studies [54, 188].

3.2.3 *Photodissociation and Ligand Transfer*

Photodissociation was simulated based on the sudden approximation [165]. Starting from the bound state (FeCO), the ground state PES was instantaneously switched to the parameters of the unbound state and a repulsive term between the Fe and the C_{CO} atoms was added:

$$V_{repulsive}(r) = C \cdot \left(\frac{r}{\sigma}\right)^{-12} \tag{72}$$

where σ is the equilibrium distance of the Fe-C bond ($\sigma = 1.9$ Å) and $C = 10$ kcal/mol is an energy parameter. This potential mimics the repulsive part of the excited state PES to which the Fe-C bond is pumped by the photodissociating laser pulse [4, 85]. We note that such a preparation of the system does not necessarily redistribute the excess energy correctly into the available degrees of freedom of the leaving CO-fragment. However, as we are primarily interested in preparing a representative ensemble of photodissociated initial structures and because the dissociative potential is typically only active for a few vibrational periods (a few tenths of femtoseconds) in the present case, a more rigorous treatment was not deemed necessary. More appropriate treatments of such processes have been recently discussed in the context of photodissociation of small molecules together with final state analysis [151]. After applying the photodissociating pulse, the system is switched back to the ground state surface in order to follow the reactive dynamics. The fitted, 3-dimensional PES is reactive and allows to describe both, the Fe-CO and Cu-CO states. However, alternative

procedures to follow chemical reactions in condensed-phase simulations are available, including reactive molecular dynamics [52, 165, 191, 284], reformulations of it [53], or the empirical valence bond (EVB) formalism [271].

3.3 RESULTS AND DISCUSSION

3.3.1 *PES fitting*

The PES was parametrized by fitting to DFT-reference energies. Figure 3 shows a model of the active site that was used for generating the 3-dimensional PES corresponding to the singlet ground state. The model includes the truncated Heme a_3 in which the isoprenoid chain was replaced by a methyl group, the CO ligand, the Cu^+ cation, and the His_{240}, His_{276}, His_{290} and His_{291} side chains which were truncated to the methyl-imidazole core. In order to estimate how strongly the structure deforms when CO is bound to either Cu_B or the Heme a_3, partial optimizations were carried out with the positions of the His-C_β carbons and the carbon atoms of the propionate-carboxylic groups being fixed (see Figure 4). As it is expected the CO-Cu_B bound structure ($\theta = 180°$) is $\approx$ 16 kcal/mol higher in energy than the CO-heme a_3 one ($\theta = 0$). In the partially optimized structure, the most significant structural differences are found for the Cu_B atom (its displacement relative to the reference structure is 0.38 Å) and the heme-a_3-Fe (displacement 0.15 Å). However, the overall structure is well preserved which establishes that using the X-ray structure is a meaningful reference for scanning the PES.

For the electronic structure calculations, a grid was defined which includes Fe–Cu distances between 4.3 and 5.5 Å with a spacing of 0.1 Å, Fe-CO(CoM) distances from 1.5 to 3.5 Å and angles θ corresponding to an 11-point Gauss-Legendre grid. Such a definition for the angular grid was found to enhance the stability of the fit and allow for a convenient representation of the angular degree of freedom in fitting the 3-dimensional PES [130, 166, 187]. The minimum barrier of the transfer reaction on the ground state PES is 38.1 kcal/mol and the root mean square error between the parametrized PES and the DFT reference energies is 1.2 kcal/mol. A direct comparison between the reference data and the fit is reported in Figure 4(c) for the

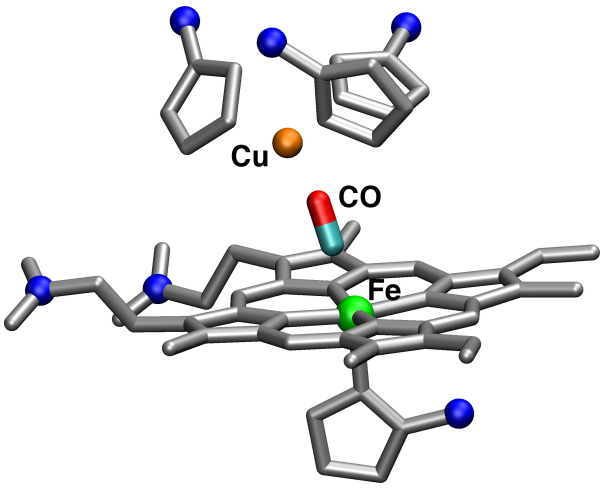
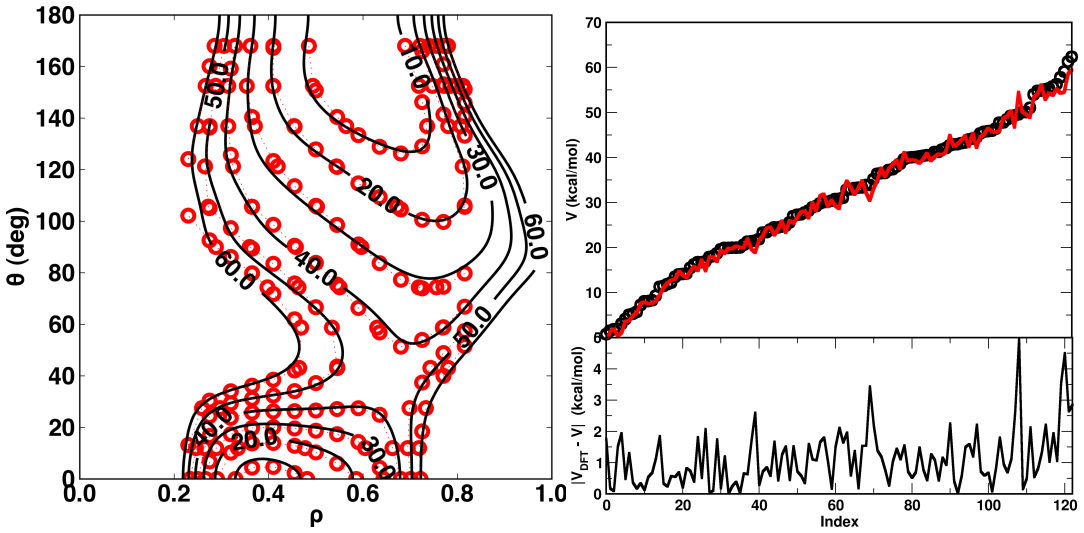


Figure 3: Optimized structure of the active site at the B3LYP/6-31G(d,p) level for CO bound to heme-Fe. The histidine C_β carbons and the carbon atoms of the propionate-carboxylic groups, marked blue, are fixed at their X-ray positions in the partial optimizations.

cut with an Fe–Cu distance of 5.2 Å and demonstrates the high quality of the PES.



(a) (b)

Figure 4: (a) PES based DFT points (red) and fitted potential (black) for R = 5.2 Å. (b) The DFT energies (black circles) and the corresponding values of the fitted PES (red) (upper frame), and the absolute error of the potential for each of the points (lower frame) for R = 5.2 Å.

Reference	Fe–C [Å]	Fe–C–O [°]	Cu–O [Å]
CcO (1AR1) (MD, current work)	1.93 ± 0.067	176 ± 2.04	2.39 ± 0.538
CcO of *Thermus thermophilus*[147]	1.95 ± 0.04	126 ± 8	2.42 ± 0.11
Bovine heart CcO [282]	1.90	152	2.4

Table 3: Characteristic bond lengths and angles in the active site from X-ray structures and the present simulations.

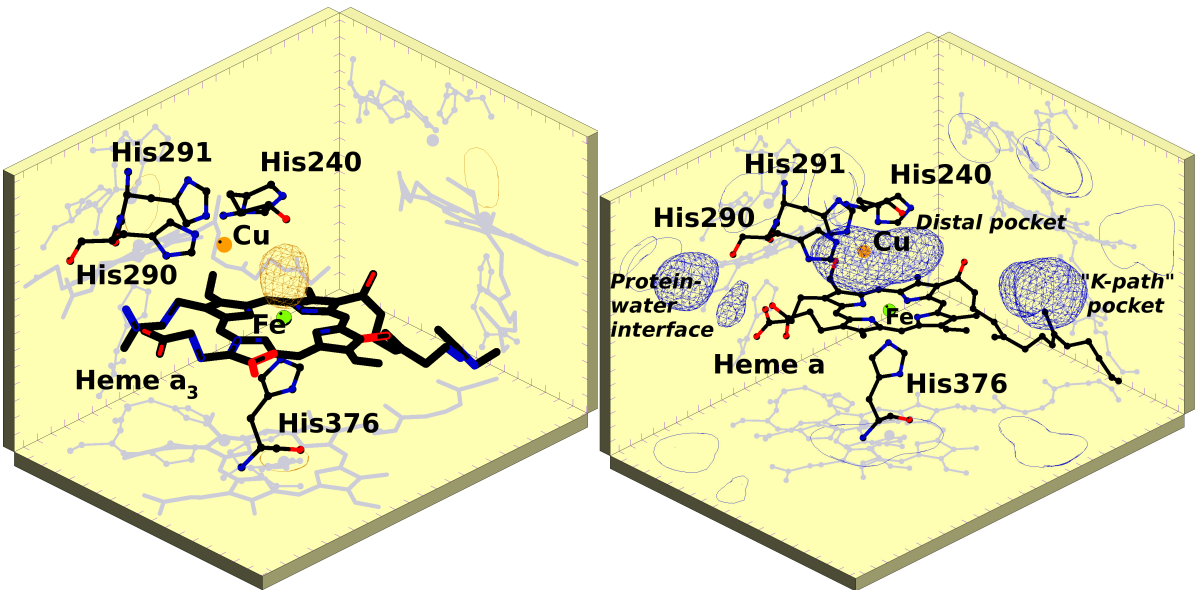
3.3.2 *Dynamics of CO bound to heme a_3 Fe*

As a first validation of the parametrized PES, NVT simulations were carried out for Fe-bound CO. The average Fe–C bond length is 1.93 ± 0.067 Å, the average Fe-C-O valence angle is $176 \pm 2.04°$ – close to linearity – and the average Cu_B–O distance is 2.39 ± 0.538 Å. The structural data reported for the bovine heart CcO [282] and the one of *Thermus thermophilus* [147] are summarized in Table 3. Significant differences are, however, observed for the CO-tilt (Fe–C–O angle). The canonical Fe–CO conformation is linear [261] which is also found in the present MD simulations but is very different in *Thermus thermophilus*. Part of the difference could be related to the rather low resolution of the X-ray diffraction experiments of 2.8–2.9 Å and 2.8–3.2 Å and the low temperature of 100 K at which these experiments were carried out [147, 282]. The tilt observed in the X-ray structures has been proposed to result from electrostatic repulsion between the Cu_B atom and the O of CO [147]. However, given the experimental data on related systems and the fact that the partial optimizations at the B3LYP/6-31G(d,p) level (see above) based on the X-ray structure yield an almost linear Fe-CO arrangement with an angle of 178°, this appears to be unlikely. The current results are also supported by previous computational studies of model heme-CO systems [165, 188–190, 243].

The Fe–Cu distance during the equilibrium dynamics is 4.98 ± 0.08 Å (see Figure 28). This compares with an average distance of 4.4 ± 0.34 Å from simulations for the unligated protein (see below) and 4.5 Å from the X-ray structure [195]. Hence, the presence of the ligand in the active site pushes the two metal centers apart and photodissociation of the ligand allows the Fe- and Cu-atoms to move closer. Also, the comparison between simulations and experiment for the unligated system are an additional validation for the force field employed in the present work.

3.3.3 *Free CO Dynamics*

In order to better characterize the active site once the CO molecule can move in the bimetallic site, simulations were also carried out for photodissociated CO. For this the bond between Fe and CO was removed and replaced by equation 72 which was active for 100 fs after which the Fe–CO nonbonded interactions were reintroduced. This included in particular the van der Waals and electrostatic interactions between the CO ligand and its environment. For the electrostatic part the fluctuating three-point charge model (see Methods) was employed. This model describes the interaction of CO with the protein better than a conventional point charge model [188, 189]. In particular, the molecular dipole and quadrupole moments are correctly captured and thus fluctuate with CO bond length. This also allows to determine the CO-infrared spectrum from the time-varying dipole-moment autocorrelation function $C(t) =< \mu(t)\mu(0) >$.



(a) (b)

Figure 5: Probability distributions for a) Fe-bound CO and b) photodissociated CO in the active site of CcO. Important residues in the active site are labelled. The small green and brown spheres are the heme-Fe and Cu_B atoms, respectively. In b) several docking sites, including the "K-path" and distal site pockets can be distinguished.

The total simulation time for free CO in the bimetallic site was 11 ns. This allowed to map out the shape and the effective volume of the cavity around the active site to which free CO has access, as reported in Figure 5. In the free dynamics simulations

we find three possible locations to which CO escapes after photodissociation: the distal pocket, the water-protein interface and the "K-path" shown in Figure 5(b). Most of the time CO remains in the distal pocket. In one simulation CO moved towards the protein-water interface and in one it diffused into the "K-path" pocket. In cases where CO left the distal pocket, the escape occurred on a timescale of $\approx$ 0.5 ns.

During these simulations the heme a_3 Fe changed from a hexacoordinated to a pentacoordinated state. Hence, the Fe moves below the heme a_3–plane. This coordination change can be described by the distance between the plane defined by the four heme-nitrogen atoms and the Fe atom. In the original crystal structure the heme-Fe separation is $\approx$ 0.42 Å below the plane (in the direction away from the distal pocket) which compares with 0.50 Å from the present simulations and 0.35 Å from previous work on myoglobin [165]. After excitation, CO moves in the direction normal to the heme plane. Hence, the CO dynamics right after photodissociation is rather translational than rotational. After 100 to 250 fs the CO hits the protein and its dynamics is affected by the structure of the pocket of the active site.

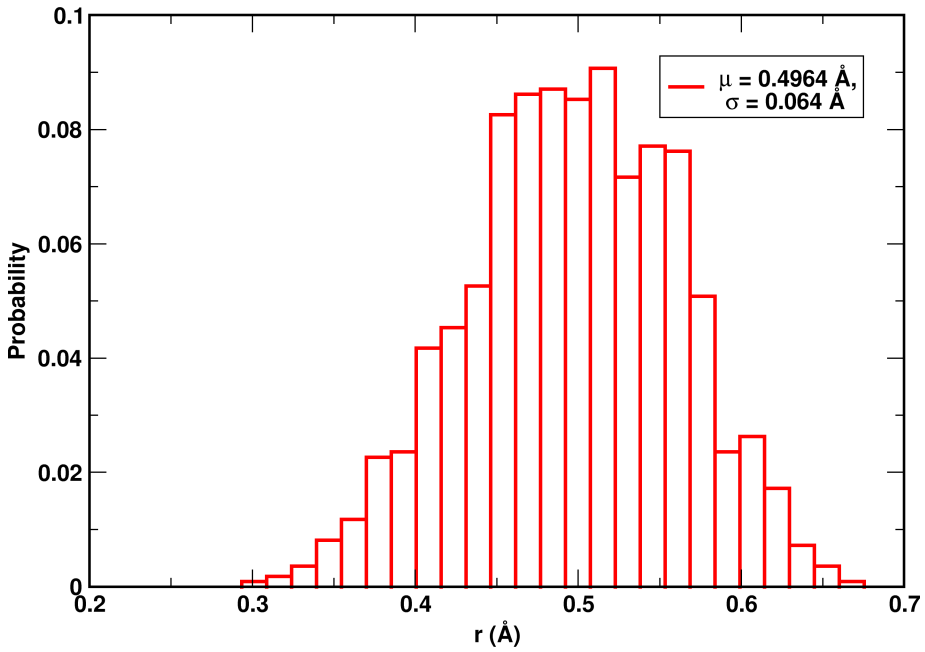


Figure 6: Out of plane motion of Fe in heme a_3 during free CO dynamics. μ - expectation, σ - standard deviation

As the major function of CcOs is proton transfer, the proton channels such as the K-path often contain cavities sufficiently large to accept water molecules. The K-path

is a proton uptake pathway, in which protons are transferred towards heme a_3 and heme a sites, and therefore essential for the function of the protein. In crystal structures of CcOs of other bacteria, e.g. *Thermus thermophilus* [246], up to 3 water molecules can be found in the area close to the "K-path" pocket [29]. Hence this pocket may be similar in character to the well-known Xe-binding sites in CcO [149] or myoglobin [68, 249] which are also accessible to external ligands, including CO and NO. It was therefore of particular interest to investigate the dynamics of the free CO molecule inside the active site of CcO, as experiments [147, 252] do not give definitive answers about the existence of free CO after photodissociation.

Infrared spectroscopy is suitable to address this point but may lack the sensitivity and resolution for unambiguous detection. For a computational characterization, ten individual 500 ps simulations of photodissociated CO in the active site were run. This time scale is appropriate to obtain sufficiently resolved spectra as recent work on NO in Mb has shown [138]. The simulations were started from Fe-bound CO, the repulsive potential to induce photodissociation was active for 100 fs after which the system was switched to the ^{3}A state. The average IR spectrum of the photodissociated, free CO molecule was calculated with a resolution better than 1 cm^{-1} and is reported in Figure 7. For the IR spectra, the real-time dipole-dipole autocorrelation function $C(t) =< \mu(t)\mu(0) >$ is calculated and its Fourier transform $C(\omega)$ yields the IR spectrum according to

$$A(\omega) \propto \omega(1 - \exp(-\hbar\omega/(k_B T)))C(\omega) \tag{73}$$

where k_B is the Boltzmann constant, $\hbar$ is the Planck constant, and T is the temperature. The individual spectra differ as they reflect different environments sampled by the ligand but all of them are broad and centered around 2180 cm^{-1} which is the value for a gas-phase simulation of isolated CO [188, 209]. The difference to the 2143 cm^{-1} band known experimentally originates from the finite step size used in the MD simulations and the classical treatment of the anharmonic oscillator [18, 51, 209]. Consequently, an insignificant shift for photodissociated CO in the active site of CcO is predicted relative to gas-phase CO which contrasts with the situation encountered for free CO in the active sites of Myoglobin or Neuroglobin [138, 144, 152, 179, 180, 188–190, 209, 229].

As anticipated from previous experiments [252] the expected spectrum of photodissociated CO in the active site is broad which is consistent for the dynamics of a free diatomic molecule in a highly inhomogeneous electric field and a largely open chemi-

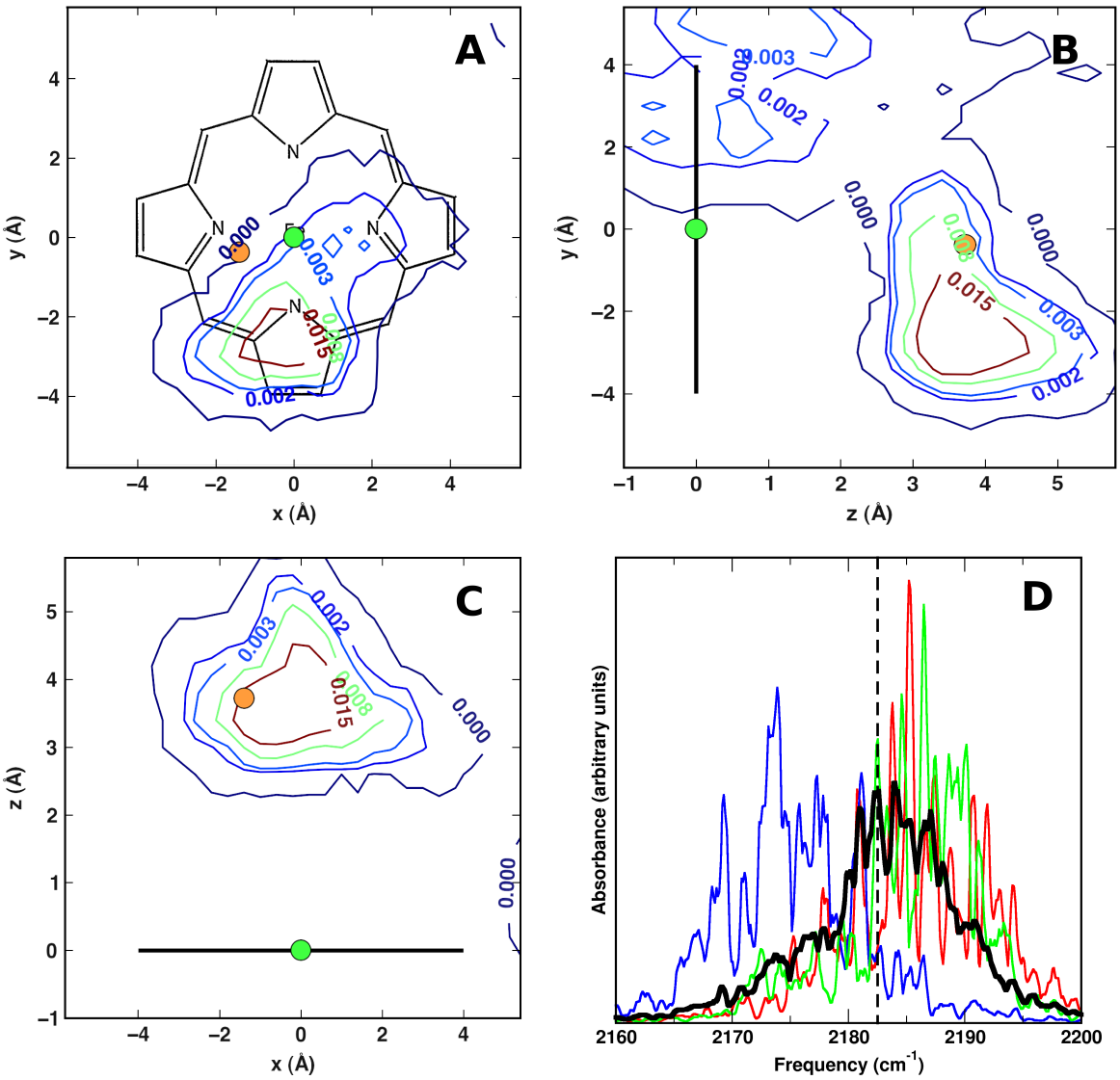


Figure 7: (a,b,c) Localization of the free CO in the active site from 5 ns (10×500 ps) simulations (**a** - XY plane, **b** - YZ plane, **c** - XZ plane). Black bold solid line in YZ and XZ planes shows the approximate linear dimensions of the heme a_3. The blue circle show the position of the heme a_3 Fe, the green circle shows the position of the Cu_B. (d) Infrared spectra of the free CO in the active site of the CcO, calculated from ten 500-ps trajectories at 300 K using the three-point fluctuating charge model: the colored lines are from the individual trajectories and the black bold line is their average and the dashed line corresponds to the position of the IR absorbance peak of the free CO in vacuum, calculated using the three-point fluctuating charge model.

cal environment: the pocket of the active site is formed by both polar (the tri-histidine (His_{276}, His_{325}, His_{326}) coordinated Cu_B, a Tyr_{280} with a hydrogen bond to the hydroxyl of the heme a_3) and apolar groups (Val_{279}, Ile_{315}, Ile_{347}). The pocket volume was analyzed using SURFNET [132]. This program finds cavities by covering space lined by a

given collection of residues using spheres of different radii ($R_{min} = 1.0$Å, $R_{max} = 3.0$Å), which define the volume of the pocket. For estimating the pocket volume relevant in the present work, 400 frames along a trajectory with photodissociated CO were analyzed. For this, the CO molecule was removed from the structure. The average size of the pocket is 105.06 ± 31.8 Å, which is approximately three times larger than the cavity of the distal pocket in myoglobin [23]. This, together with the low signal-to-noise ratio in the experiments could partially explain the absence of the free CO signal in the experiments [252].

3.3.4 *CO transfer following excitation*

In a next step, the ligand transfer reaction was investigated. To this end, 1000 independent excitation trajectories for each excitation time τ_e were run and analyzed. Initially, the excitation was simulated by a repulsive term (Eq. 72) which was already used in the free dynamics. At defined times after the excitation - including $\tau_e = 20$, 30, 50, 100, 150, and 200 fs (for each initial configuration) - the interactions were switched to the interaction potential Eq. 68 describing the ground state PES supporting both, Fe-CO and Cu-CO-bound states. For the heme a_3 separate sets of simulations were carried out with the penta- and the hexacoordinated force fields. In the following, only the results with the pentacoordinated force field are described and comparisons with results from simulations with the hexacoordinated force field are made where it is appropriate. The supplementary material reports the results from simulations with the hexacoordinated force field.

The analysis of excitation trajectories shows that the probability for ligand transfer to yield Cu-CO is only insignificantly affected by increasing the quench delay time (see Figure 8). For all times τ_e the final state population is prominently peaked at the Cu-CO state which suggests that following photodissociation from the heme-Fe the CO ligand is efficiently transferred to the copper atom. After photodissociation, the Fe atom moves out of the heme plane defined by the four heme nitrogens to a distance of $\approx$ 0.5 Å below the plane within the first $\approx$ 100 fs of the simulation. This has also been found in previous work on Myoglobin [165, 263] and the time for breaking the Fe-C bond has been estimated to be on the order of 100 fs [241]. The Fe-out-of-plane displacement is close to that in the experimental 1AR1 X-ray structure

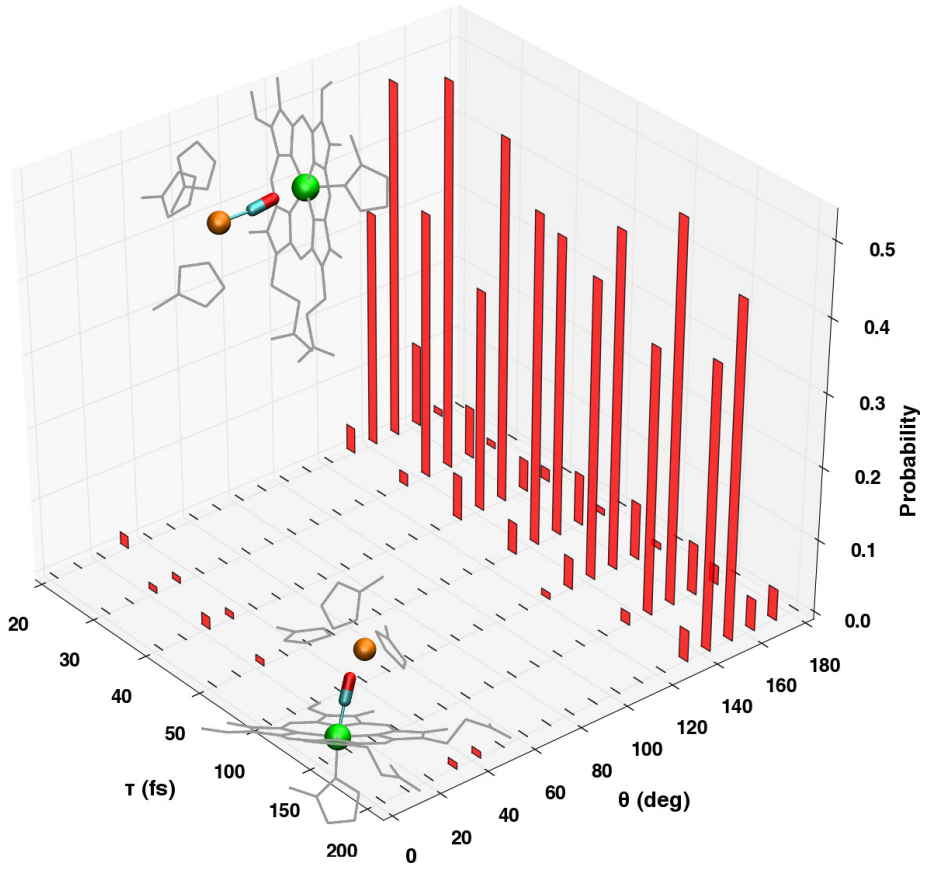


Figure 8: Probability of final states (after 5 ps) as a function of the quench delay time (based on 1000 trajectories). $\theta = 0$ corresponds to Fe–CO, $\theta > 150°$ to Cu–CO.

for pentacoordinated heme a_3, where it is 0.42 Å.

In equilibrium simulations for Fe-bound CO the Fe–Cu distance is typically around 5.0 Å (see Figure 28) which increases to 5.5 Å during the transfer within a picosecond. This displacement is caused partially by the out-of-plane motion of the heme Fe. Therefore the transfer leads to moderate but functionally important changes to the structure of the active site. In particular, the pentacoordinated heme-Fe is "hidden" below the plane which makes it less available for reactions with external ligands. This is in contrast to the tetrahedrally coordinated Cu-ion which always presents an available ligand-binding site to free CO.

The analysis of the trajectories suggests that θ is a meaningful reaction coordinate to describe the transition between the Fe–CO ($\theta \approx 0$) and Cu–CO ($\theta \approx 150°$) states (as θ is the angle between heme a_3, the center of the CO and the C of the CO, the minimum corresponding to the Cu–CO has a value of 150° in the coordinate system used). The

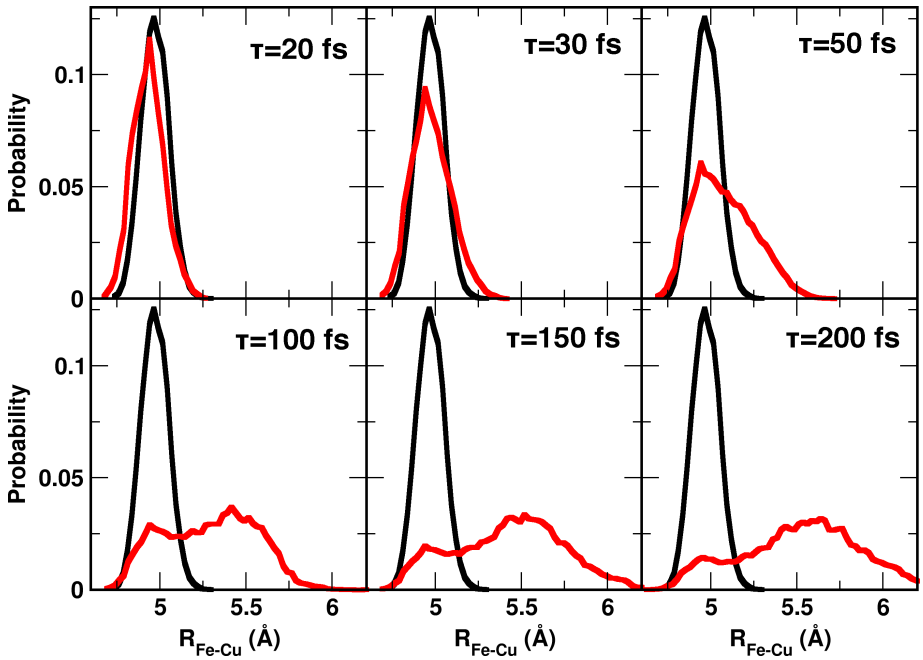


Figure 9: Probability of Fe-Cu distances for simulations (5 ns) of Fe-bound CO (Fe-CO) (black; μ=4.98 Å, σ=0.08 Å) and for 1000 excitation simulations showing the structural changes of the active site (red). The quench delay times τ_e are reported in the panels.

time series $\theta(t)$ obtained from the trajectories (Figure 32) shows that CO arrives at the Cu_B site typically within 1 ps. The transfer itself consists of two phases - a ballistic and a diffusive one (Figure 11).

The kinetic curves show that there is a lag time for the transfer reaction, which corresponds to the minimal time required for the CO molecule to move towards the Cu_B site, followed by the rotation (Figure 11). In the simulations this lag time is the sum of two components: the quench delay time after which the dynamics is driven by the repulsive potential (Eq. 72) and the true ballistic time, when the potential describing the ground PES is active (Eq. 68). The estimated lag time is not greater than 300 fs and the true ballistic time is $\approx$ 100 fs (Figure 34).

To give a quantitative interpretation of the ligand transfer reaction we further analyze the reactive trajectories. Similar to proton transfer reactions, it is useful to consider geometrical criteria with which to distinguish educt from product states and to subsequently determine the sensitivity of the conclusions on the specific criteria chosen [84, 167]. The geometrical criterion for ligand transfer is $\rho \geqslant 0.64$ (CO is closer to Cu_B site) and $\theta \geqslant 150°$, which are the average values of the corresponding variables

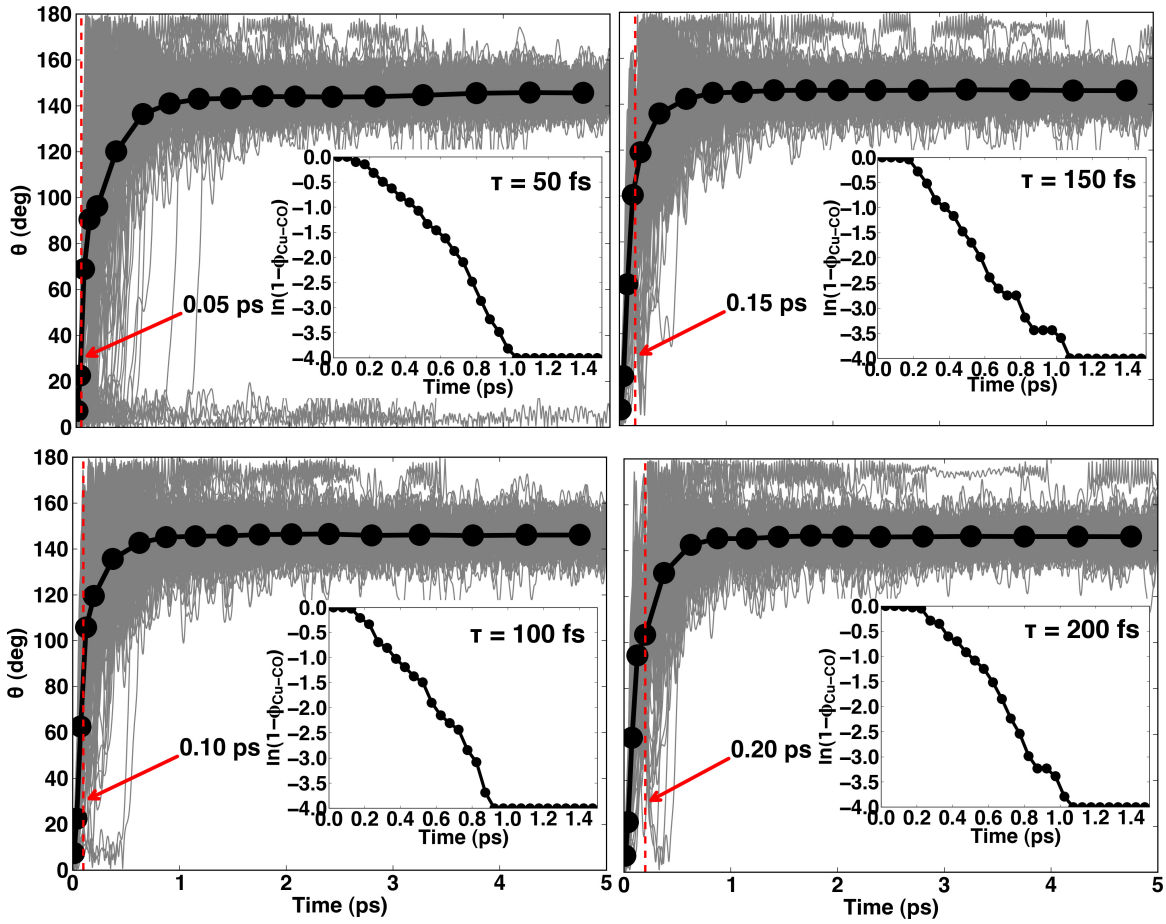


Figure 10: CO transfer kinetics for different quench time delays - see labels in panels from analysis of 1000 trajectories. Main panels: black points are averages $< \theta >$; grey overlapping lines are the individual trajectories, red dashed line shows the point where the quench has been performed. Insets: Φ_{Cu-CO} is the conversion fraction to the Cu–CO state. At $t = 0$, $\Phi_{Cu-CO}(t = 0) = 0$. The slope of $\ln(1 - \Phi)$ is the rate of forming Cu-CO, see text.

obtained from the final state analysis ($\rho_{Cu_B-CO} \geqslant 0.64 \pm 0.05$, $\theta_{Cu_B-CO} = 150 \pm 5°$). Hence, if a trajectory matches this criterion, the transfer is considered to be complete. In the following, the process of CO transfer is treated as a first-order chemical reaction. For determining the rate of conversion from Fe–CO to Cu–CO we use the conversion fraction (Φ) as a concentration term where Φ is the fraction of the trajectories for which CO is transferred to the Cu_B site at a particular time. Fitting the data (see insets in Figure 32) to a first order kinetics yields an estimated rate constant of 3.83 ± 0.7 ps^{-1} and a characteristic time for the conversion to Cu-CO of $\approx$ 260 fs, which is somewhat more rapid than the value (450 fs) obtained from experiment which is based on an exponential fit of the Cu–CO (2065 cm^{-1}) differential transmission

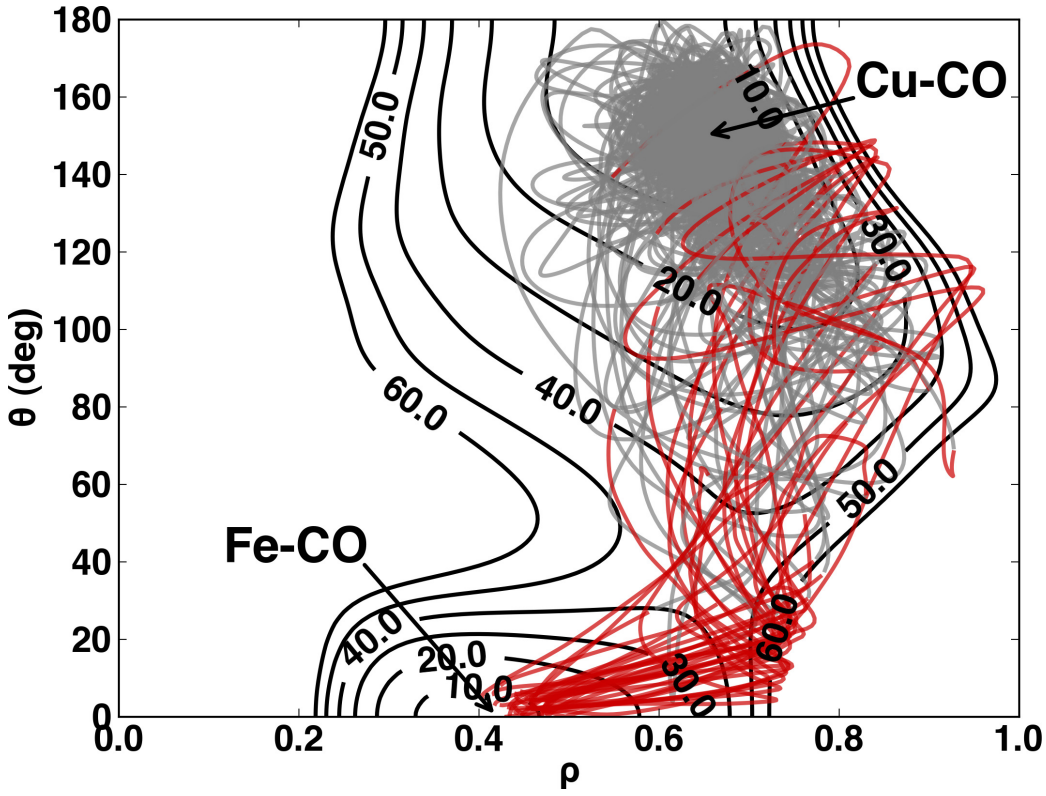


Figure 11: 40 different reactive trajectories which lead to transfer for different quench time delays. The PES for a specific Fe–Cu separation of 5.2 Å is also reported for reference. Red lines - dynamics on the excited state surface (CO unbound), grey - dynamics on the ground state surface (CO bound to Cu). Green and orange areas show the approximate contour of the termination criteria used to validate the approach for Fe–CO and Cu–CO states correspondingly, so that whenever the CO molecule is found within the area the state is considered to be defined.

spectrum [252].

For the hexacoordinated state in which the heme-Fe remains closer to the heme plane, the Fe is still available for rebinding to the CO molecule. Hence, at very short quench delay times ($\tau_e \leqslant 100$ fs) the probability of rebinding to the heme a_3-Fe is high (see Figures 30 to and 31 in the supplementary information). The probability for ligand transfer to form Cu–CO increases with quench delay time and reaches a maximum of close to 100% (see Figure 29) for $\tau_e > 100$ fs and does not significantly change even up to 1 ps. After photodissociation, the Fe atom slightly moves out of the heme plane in the similar fashion as in the pentacoordinated state but only to a distance of ≈ 0.10 Å below the heme plane within the first 20–100 fs of the simulation. Although the equilibrium structures of the penta- and hexacoordinated heme-group are quite

different, the rates of conversion from Fe–CO to Cu–CO are similar. The rate increases from 260 fs to 300 fs when simulations are carried out with the hexacoordinated heme force field. Hence, the conclusions about the time scale and nature of the rebinding reactions are robust with respect to force field parametrizations which characterize two extremes (penta- and hexacoordinated heme-Fe) of the physically relevant situation.

3.3.5 *Validation of the CO transfer protocol*

The approach used so far assumes a certain timescale τ_e for the excited state dynamics. Beyond 100 to 200 fs - depending on the coordination of the heme-Fe - the final state was found to be insensitive to the particular time spent in the unbound state. In order to independently validate this, additional simulations in the unbound state were carried out. Specifically, 600 trajectories, each with a maximum simulation time of 100 ps in length, were started from the bound state. The Fe–CO bond was broken and the repulsive potential was activated, after the 50, 100 and 200 fs delay τ_e the repulsive potential was switched to the CHARMM standard force field with a pentacoordinated heme a_3 and the CO treated with the 3-point fluctuating charge model (analogous to the free CO simulations).

The time of arrival at the Cu_B site or at the heme a_3 was again determined when certain geometrical criteria were met. They included a metal–C_{CO} distance of less than 2.45 Å and a metal–C-O angle greater than 100°. For comparison, the equilibrium Fe–C separation from the X-ray structure is 1.9 Å and the Cu–C distance from DFT calculations is 1.87 Å. With these geometrical criteria, within 1.5 ps after photodissociation 85% (vs. 90% in the hexacoordinated state) of the simulations have a defined state which is either Cu_B–CO or Fe_{a_3}–CO. In the remaining cases, CO is still unbound and samples the active site regions characterized earlier, see Figure 5. Specifically, within 1.5 ps of breaking the Fe-C bond in 512 (85.3%) simulations (out of a total of 600) the CO is bound to Cu_B, in 10 cases (1.7%) it rebinds to the Fe_{a_3} and for 78 (13%) simulations CO remains in a photodissociated state. The kinetics analysis for the validation simulations based on the first order kinetics along the same lines as for the reaction using the explicit reactive 3d-PES (see above) yield rate constants of 2.54 ps^{-1} (characteristic time 393 fs) and 3.11 ps^{-1} (characteristic time 321 fs) for penta-

and hexacoordinated heme-Fe respectively. Within the limitations of the validation simulations this compares favorably with the simulations on the reactive surface.

Similarly to the case when simulations were run on the 3-dimensional PES suitable to explicitly describe both bound states, Fe-CO and Cu-CO, additional simulations were carried out in which the heme was treated with a hexacoordinate force field. Within the first 1.5 ps after the excitation 90% of the trajectories terminate in a metal-bound state (Figure 31). The probability of rebinding to the Fe_{a_3} is higher in this case as the iron atom is readily accessible to the CO ligand due to the smaller out-of plane distortion (0.1 Å vs. 0.5 Å). Hence, after 1.5 ps 484 (81%) of the simulations find the Cu-CO state, 59 (10%) terminate in Fe–CO state and in 57 (9%) case CO remains in a photodissociated state or rebind the metal sites on the longer timescale.

3.4 CONCLUSIONS

In the present work the dynamics of CO between the heme a_3 and the Cu_B binding sites in CcO were studied based on molecular dynamics simulations using an accurate 3-dimensional reactive PES. Independent on the details of the simulation protocols, the time scale for ligand transfer after photodissociation from the heme-Fe is on the sub-picosecond time scale. Following photodissociation from the heme a_3-Fe, approximately 90 % of the ligands rebind on such time scales to either Cu_B (majority) or the heme a_3 (minority). The data obtained from MD simulations and experimental data suggest a ballistic contribution to the transfer process. A time delay of $\approx$ 100 fs before rebinding starts is found in the simulations. This qualitatively agrees with experiment where the time scales were determined by analyzing spectroscopic signatures [252]. The characteristic time scale for rebinding is $\approx$ 300 fs compared with 450 fs from experiment. The remaining population (10 %) consists of unbound CO sampling the active site which, however, are difficult to directly observe experimentally. Computational infrared spectroscopy suggests that the absorption features of unbound CO are broad and centered around the gas-phase value. All these observations agree with experiments which were interpreted based on spectroscopic data [252].

4

REPRODUCING KERNEL POTENTIAL ENERGY SURFACES IN BIOMOLECULAR SIMULATIONS: NITRIC OXIDE BINDING TO MYOGLOBIN

4.1 INTRODUCTION

Following the dynamics of small ligands in (globular) proteins provides information about the internal structure of the macromolecule and about the way how ligand and protein motions are coupled. By combining experimentation and simulation a detailed picture about these processes can be obtained. One of the paradigmatic systems in this context is myoglobin (Mb) and its ligands O_2, CO or NO, see Figure 21. Nitric oxide (NO) is a key messenger for many biological processes in vertebrates. It can bind to myoglobin (Mb), and together with O_2 can be catalyzed by Mb to form nitrate [68]. The rebinding kinetics of NO to Mb has been studied intensively over the past three decades using experimental and computational methods [49, 52, 71, 104, 106, 123, 124, 127, 141, 165, 191, 205, 224]. Compared to rebinding of CO, no substantial rebinding barrier has been found for geminate recombination of NO to the heme-iron [205] and ab initio calculations have even suggested that the recombination reaction may be barrierless for NO in specific conformations [67]. Because of the short time scale involved (picoseconds), this process is ideally suited to be investigated computationally at an atomistic level [52, 141, 165, 191, 224].

Over the past few years, computational investigations of ligand migration in globular proteins has provided considerable insights into their structure, energetics and dynamics [23, 60, 208, 221]. All these efforts require adequate representations of the intermolecular interactions. Methods for this range from purely empirical force fields to mixed quantum mechanical/molecular mechanics (QM/MM) treatments. They all have their merits and disadvantages in terms of computational efficiency and realism in capturing the essentials of the interactions. While empirical force fields are fast in terms of evaluating the energies and forces of a given configuration, they need to be fitted to a parametrized form, their accuracy is limited and they do not allow to describe bond-breaking and bond-forming processes. On the other hand, mixed

QM/MM simulations can be more accurate depending on the level of theory used in the quantum part, but their application is usually limited to single (or few) trajectories of short duration [134, 170]. In particular for systems containing metal atoms - as is the case in Mb where the active site is the Fe-heme unit - even density functional theory (DFT)-based approaches become computationally challenging. Furthermore, technical difficulties can arise, e.g. converging the Hartree Fock wavefunction or maintaining the correct electronic state, if sufficiently distorted conformations are sampled in the MD simulations.

An alternative to QM/MM is to explicitly parametrize the total energy (instead of representing the total energy as sum over individual force field terms) for a subsystem based on rigorous quantum chemistry and combine this with an empirical force field. Again, the parametrization step is critical and often time-consuming. Such an approach has, for example, been used in combining accurate representations for the proton transfer energetics between a donor and an acceptor atom together with empirical force field parameters for the environment [130]. The approach - molecular mechanics with proton transfer (MMPT) - has been successful to realistically follow the proton transfer dynamics in protonated water dimer, diglyme, malonaldehyde or even in a metal-catalyst used for hydroformylation [97, 98, 131]. However, because the quantum-based energies can not always be represented by standard harmonic terms in order to take full advantage of their higher accuracy, tedious non-linear fits are often involved in the parametrization. In the present work kernel interpolation will be used which is based on solving a linear system instead of carrying out a (non-)linear least squares fit.

Recent experiments have followed the interplay between the Fe-out-of-plane (Fe-oop) and the NO-ligand motion in a time resolved fashion [127]. This work points towards a direct coupling between these two degrees of freedom on the 10 to 100 ps time scale. Running a statistically significant number of QM/MM trajectories from which to analyze and atomistically resolve the interplay of the motions involved is beyond current computational methods. In order to retain the accuracy needed, the total interaction in the protein-ligand system is decomposed in a similar fashion and the subsystem - ligand and heme-oop coordinate - is treated at the density functional theory (DFT) level. Subsequently, a suitable representation that preserves the accuracy of the interpolated potential energy surface (PES) is chosen and used in the simulations. An alternative to parametrized representations of PESs are parameter-free,

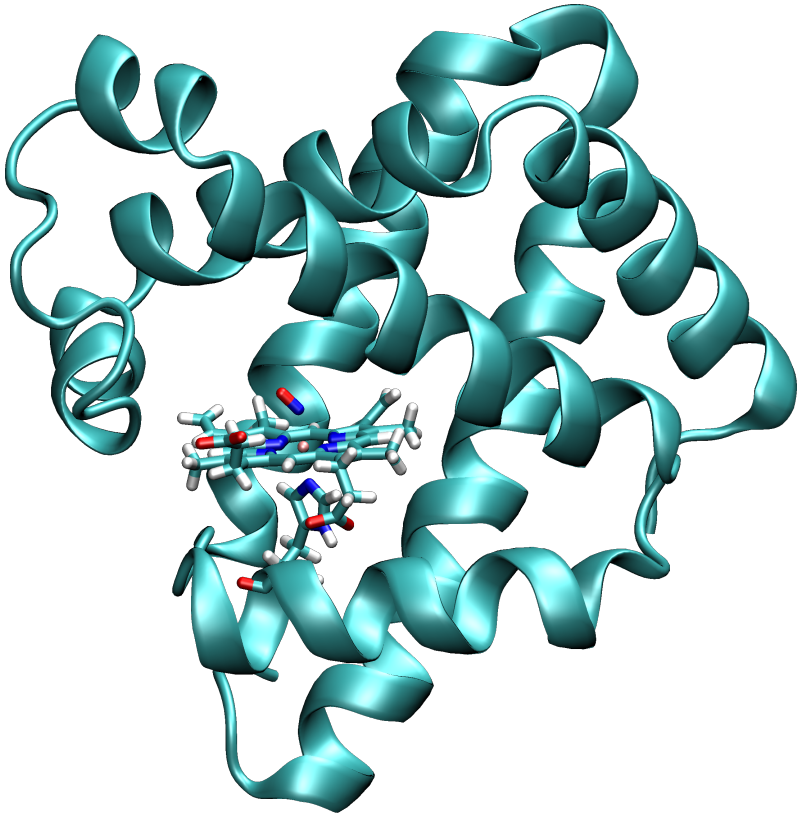


Figure 12: NO-bound Myoglobin with the heme, His93 and NO ligand in licorice and the protein in secondary structure representation.

reproducing kernel Hilbert space (RKHS)-based representations of reference data from quantum chemical calculations [7, 93]. They are based on smoothness criteria of the interpolant and have been successfully applied to study van der Waals complexes [90, 166] and reactions [31, 32]. One of the hallmarks of an RKHS-interpolated PES is that it reproduces the reference energies *exactly* (i.e. by construction) which is not the case for parametrized force fields.

In the present work, RKHS-based representations for a subsystem for which accurate energetics is required are combined with a force field description of the protein environment. The subsystem contains the ligand (nitric oxide) degrees of freedom and the heme-iron out-of-plane motion which has been shown to be decisive for the two limiting states (bound and unbound) and the dynamics between them. The interpolated PESs are then used to explore the ligand dynamics in the ^{2}A and ^{4}A states. The approach used in the present work also allows to following the isomerization dynamics of the ligand which involves bond-breaking and bond-formation between the heme-iron and the ligand. First, the computational models are presented. Next,

the RKHS-PES is discussed and its performance in atomistic simulations is assessed. Finally, the dynamics on the two electronic states of interest is characterized and compared with experiment and previous work.

4.2 METHODS

4.2.1 *Intermolecular interactions*

The total potential energy of the system $V_{tot}(\vec{X})$, where $\vec{X}$ is the position vector for all N atoms, is written as

$$V_{tot}(\vec{X}) = V_{FF}(\vec{Q}) + V(R, \theta, \phi). \tag{74}$$

Here, $V_{FF}(\vec{Q})$ is the standard CHARMM22 force field [155], which is employed for all the coordinates of the configuration space $\vec{Q}$ except (R, θ, ϕ), which are the coordinates describing the interactions between the heme-Fe and NO, and the doming coordinate of the heme-Fe. The doming coordinate describes the transition between an in- and and out-of-plane iron atom upon changing its ligation state from 6- to 5-coordinated (ligand bound to ligand unbound, see Figure 13). The nitric oxide ligand is described by a harmonic bond with force constants of 824.93 and 1101.46 kcal/mol to reflect the changing strength in the NO bond upon binding and unbinding [165]. The R−coordinate is the distance between Fe and the geometrical center of NO, θ is the angle between R and the NO-molecular axis, and ϕ is the average angle between each of the four heme-nitrogen atoms, the heme-Fe and the N_ϵ of the axial His93. Therefore θ describes the NO-rotation and ϕ corresponds to the heme-doming coordinate.

The reference energies for the 3-dimensional PES $V(R, \theta, \phi)$ were obtained from density functional theory (DFT) calculations based on the B3LYP [14, 136, 239, 265] functional together with the 6-31G(d,p) basis set [203, 204] Gaussian09 [Frisch et al.] was used to perform all DFT calculations. The model system contains a truncated heme group, the NO ligand and a methyl-imidazole as a core of the axial histidine (His93), see Figure 13. The reference points were calculated on a 3-dimensional grid: The R−coordinate was scanned between 1.7 and 3.9 Å with step size $\Delta R = 0.1$ Å. The distance d between the heme-Fe and the heme-plane included positions d = 0.2 and 0.1 Å (for the ^{4}A state only) above the plane, the in-plane d = 0.0 Å position, and d = −0.1, –0.3 and –0.5 Å below the plane. For θ an 11-point Gauss-Legendre

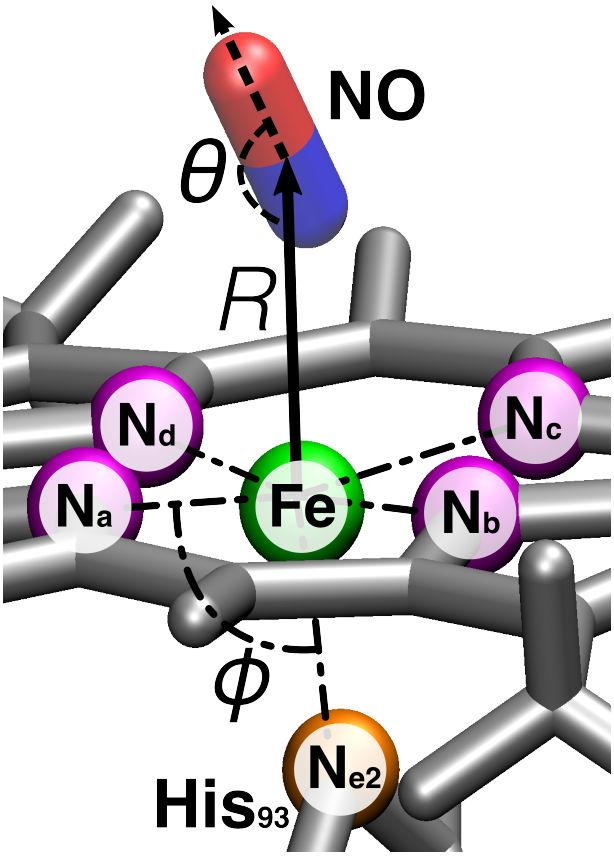


Figure 13: Relevant coordinates for the energetics of bound and unbound NO. Heme-nitrogens N_a to N_d (purple) heme-Fe (green), N_ϵ of His93 (orange), the N and O atoms of NO molecule are blue and red. The distance between the heme Fe and the center of the NO molecule is R, θ is the angle between R and the NO-molecular axis, and ϕ is the average angle between each of the four heme-nitrogen atoms, the heme Fe and the N_ϵ of the axial His93.

grid including angles of θ [°] = 11.815, 27.452, 43.089, 58.726, 74.363, 90.000, 105.637, 121.274, 136.911, 152,548, and 168.185 was employed. The out-of-plane distance used in the DFT grid was translated into an average angle ϕ between the heme-nitrogen, heme-Fe and the N_ϵ of the axial histidine ligand via standard trigonometry.

The calculations were carried out for the ^2A and ^4A states using a 3-dimensional grid resulting in 1210 and 1452 reference points for the two states, respectively. In addition, 100 validation points for both ^2A and ^4A states was computed at off-grid points for different values of r and $\theta = 30, 60, 120°$ and $\phi = 87, 90, 92, 95°$. The RMSE for the validation set as determined from the fitted PES was 1.0 kcal/mol and 1.1 kcal/mol for the ^2A and ^4A PESs, respectively. This is close to chemical accuracy except for

errors in the quantum chemical methods used.

Initially, a parametrized fit of the PES was attempted. However, it was found that in particular representing the Fe-oop coordinate in parametrized form was difficult. Hence a different approach was considered. Kernel-based representations are a powerful means to *exactly* reproduce the target data while maintaining desirable smoothness and asymptotic behavior of the inter- and extrapolant. The explicit form of the RKHS PES is as follows:

$$V(R,\theta,\phi) = \sum_{\lambda=0}^{10} V_\lambda(R,\phi) P_\lambda(\cos(\theta)) + V_c(\phi) \tag{75}$$

where P_λ are Legendre polynomials, and V_c is a harmonic potential. By construction the kernel-based part of the potential $\sum_{\lambda=0}^{10} V_\lambda(R,\phi)P_\lambda(\cos(\theta))$ decays asymptotically $(R \to \infty)$ to zero. The harmonic potential V_c

$$V_c(\phi) = \frac{1}{2} k(\phi - \phi_e)^2 \tag{76}$$

captures the asymptotic energies of the different $V(R,\theta)$ potential energy surfaces depending on the Fe-oop position where k is the force constant, and ϕ_e is the equilibrium position. Both k and ϕ_e are parametrized during the optimization of the potential $V(R,\theta,\phi)$. The final parameters in the ^2A state are $k = 0.165$ kcal/mol and $\phi_e = 95.6°$, and in the ^4A state $k = 0.116$ kcal/mol and $\phi_e = 93.8°$.

The radial strength functions $V_\lambda(R,\phi)$ are represented as a reproducing kernel [7],

$$V_\lambda(R,\phi) = \sum_{i,j} \beta_{\lambda,i,j} \cdot \mathcal{K}(R,R_i) \cdot \mathcal{G}(\phi,\phi_j) \tag{77}$$

where $\mathcal{K}(R,R_i)$ is a radial reproducing kernel [90, 93] and $\mathcal{G}(\phi,\phi_j)$ is a Gaussian reproducing kernel [81]. The explicit form of the radial reproducing kernels used here is

$$\mathcal{K}(R_i,R_j) = \frac{1}{14} R_>^{-7} \left(1 - \frac{7}{9}\frac{R_<}{R_>}\right) \tag{78}$$

where $R_>$ and $R_<$ are the greater and smaller distance, respectively, for any pair of $R-$values. The Gaussian kernels are parametrized as follows:

$$\mathcal{G}(\phi_i,\phi_j) = \exp(-\frac{(\phi_i - \phi_j)^2}{2\sigma}) \tag{79}$$

where σ is the width of the kernel and its value of $\sigma = 5°$ was chosen such that the kernels overlap and the optimization result in smallest error on the reference

data set. Therefore kernels are symmetric positive definitive and describe the correct asymptotic behavior at large R as $\mathcal{K}$ decays to zero.

The $\beta_{\lambda,i,j}$ are determined from a singular value decomposition [58, 76, 87] for each value of λ and hence are not adjustable fitting parameters as would be the case in a typical non-linear least squares fit [166]. In order to prevent overfitting, Tikhonov regularization (parameter $\alpha = 10^{-7}$) is used [247, 248] to penalize solutions of large norm and to stabilize the procedure. The coefficients β_λ are found from solving the least-square minimization problem:

$$\begin{aligned}\hat{\beta_\lambda} &= \arg\min_{\beta_\lambda}\{\|K\beta_\lambda - E_\lambda\|^2 + \alpha\|\beta_\lambda\|^2\} \\ &= (K^T K + \alpha I)^{-1} K^T E_\lambda\end{aligned} \tag{80}$$

Here, K is the kernel matrix and E_λ are the DFT energies. The kernel matrix K is constructed based on the product kernel $k_{i,j}$:

$$k_{i,j} = \mathcal{K}(R_i, R_j)\mathcal{G}(\phi_i, \phi_j) \tag{81}$$

where $\mathcal{K}(R_i, R_j)$ and $\mathcal{G}(\phi_i, \phi_j)$ are the radial and Gaussian kernels, respectively, which are calculated for every single pair of values (R_i, ϕ_i) and (R_j, ϕ_j) of the grid.

In addition, for molecular dynamics simulations forces are required. The necessary derivatives of the RKHS PES can be obtained analytically which yields good energy conservation in NVE simulations.

4.2.2 *Molecular Dynamics Simulations*

The kernel-based interpolation was implemented into the CHARMM program [26] which was used for all molecular dynamics (MD) simulations together with the CHARMM22 force field [156]. The protein was set up as described previously [165, 187]. Mb contains 153 amino acid residues, a heme group and a nitrogen oxide molecule. Simulations were carried out for both, $His_\delta 64$ and $His_\epsilon 64$ protonation, with the latter being the more likely state [108, 163, 164, 181, 220]. Contrary to the previous work which employed stochastic boundaries, in the present work the protein was solvated in a periodic, pre equilibrated waterbox $62.0864 \times 62.0864 \times 62.0864$ Å^3. The final system contains 23711 atoms. All bonds involving hydrogens were treated

using SHAKE [222, 279] with a tolerance of 10^{-6}. For the non-bonded interactions a cutoff of 14 Å was used. First, the system was minimized using steepest-descent and adopted basis Newton-Raphson algorithms. Then it was heated from 100 to 300 K during 60 ps and equilibrated for 500 ps. This was followed by production runs in the NVT ensemble as described further below. The time step in all simulations was $\Delta t = 1$ fs and the velocity Verlet integrator was used to propagate the equations of motion.

4.3 RESULTS AND DISCUSSION

4.3.1 *The Fitted PES*

The results show that the fitted ^2A PES has two well-defined minima, which correspond to the Fe-NO and Fe-ON states. These two states have already been found in earlier work [30, 187] but are absent in other investigations [243]. Previous work found an energy difference of 15.3 kcal/mol between the two states whereas the present B3LYP/6-31G** calculations yield 18.1 kcal/mol. In order to confirm their existence, additional electronic structure calculations were carried out. For this the Fe-NO and Fe-ON structures were separately optimized at the UB3PW91/6-31G** level of theory. The converged energies differ by 18.5 kcal/mol and establish that both conformational isomers also exist with this functional. Experimentally, the existence of an Fe-ON state has been found for model systems, [30, 35, 277] but not for NO in Mb [182].

The NO-binding energy on the ^2A PES is 23.7 kcal/mol, see Figures 14 and 15. Figure 14 illustrates that the RKHS-PES exactly matches the reference DFT-points. Also, it can be seen that the binding energy of the ligand depends on the Fe-oop position and that upon NO-dissociation from an in-plane position the asymptotic energy is higher than for dissociation from an out-of-plane position. At infinite separation of the NO from the heme-Fe (see Figure 13), the Fe atom moves 0.2 Å below the porphyrin plane ($d = -0.2$ Å) and correctly predicts the doming effect. The experimentally observed out-of-plane distance in the high-resolution X-ray structure of Mb [225] is –0.27 Å and the one found in the previous work is –0.35 Å [165]. The out-of-plane positioning of the heme-Fe in the model system and in the protein differ because the protein environment exerts forces which affect the equilibrium structure. This can be taken into account by the harmonic angular potential (N_{heme} - Fe - N_{His}) of the force field.

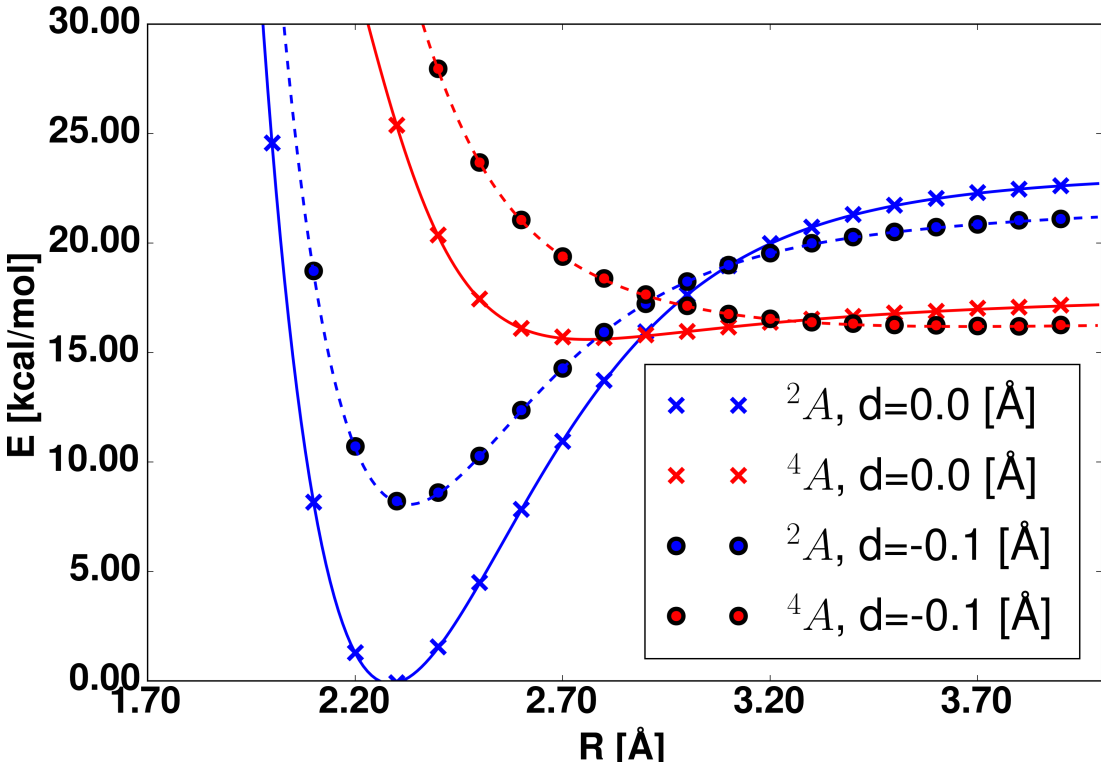


Figure 14: Radial cuts through the PES for $\theta = 152.06°$ and different out-of-plane displacements (shown in the legend). Symbols represent the reference DFT points, solid and dashed lines are the RKHS PES. When the NO molecule dissociates from an in-plane position (crosses) the asymptotic energy is higher than for dissociation from an out- of-plane position (circles). The binding energies of the ligand also depend on the Fe-oop position.

The ^{4}A state does not have a clear minimum energy structure, see Figure 15, and mostly represents a van der Waals complex with an equilibrium Fe-CoM$_{NO}$ distance of 3.5 Å.

In order to validate the implementation of the total potential energy function - and the RKHS-interpolation in particular - and the corresponding derivatives, several NVE simulations 500 ps in length were carried out. The average fluctuation of $p(E - \langle E \rangle)$ is 1 kcal/mol for simulations with the kernel-based PESs for the ^{2}A and ^{4}A states. As a comparison, simulations for the ^{2}A state were run using the force field parameters from previous work [165] and showed identical results. Also, no drift in the total energy is found. This validates the present implementation of the kernel-based PES.

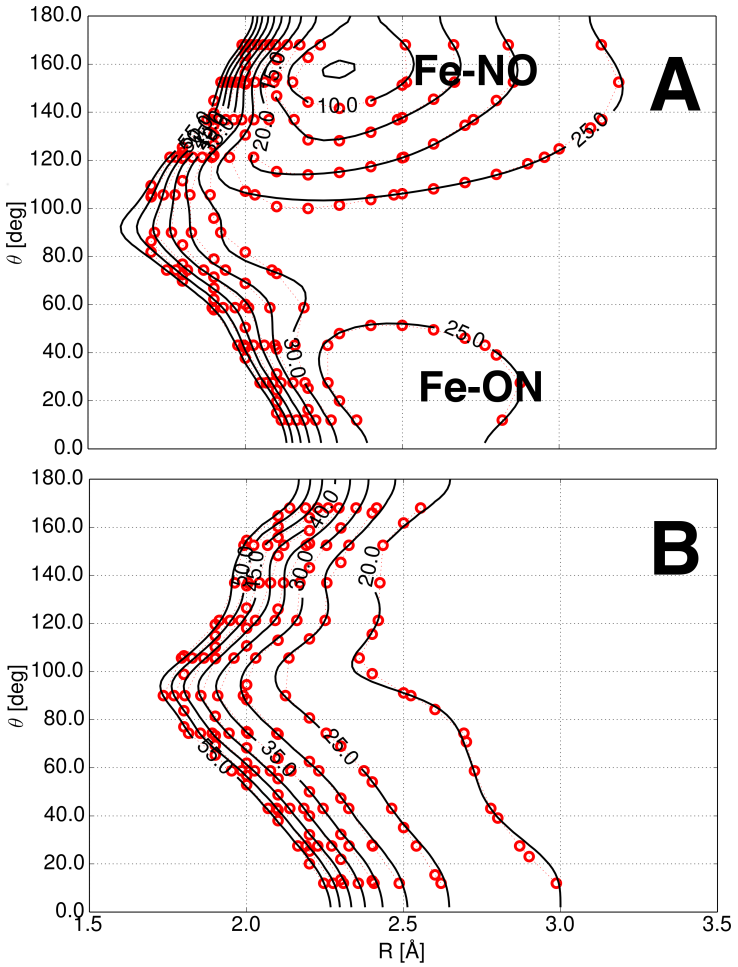


Figure 15: PESs for the two states considered here. Black isocontours are the RKHS-PES and red circles are the DFT reference points for the Fe-in-plane conformation ($\phi = 90°$). (A) The Fe-NO bound state, ^{2}A, with the global minimum in a bent Fe-NO conformation, the secondary minimum (Fe-ON) 20.2 kcal/mol and the transition state separating the two states at 23.7 kcal/mol above the global minimum. (B) The unbound, ^{4}A, state. The energies of this state are close to those of the ^{2}A state around the Fe-ON conformation which may wash out this secondary minimum.

4.3.2 *Equilibrium dynamics of the ^{2}A State*

The well depth of the bound Fe–NO minimum is 23.7 kcal/mol. In the ^{2}A state the conformationally averaged Fe–N distance is 1.79 ± 0.04 Å and the average Fe-N-O angle is $147.5 \pm 4.6°$. These values are in good agreement with the experimentally determined structures with nitric oxide bound to the heme-Fe, including myoglobin [27, 105, 283], Indoleamine 2,3-dioxygenase [2], cytochrome c oxidase [75] and others (see Table 4). The differences might be explained by the effects of the crystal field and

Reference	Fe–N [Å]	Fe–N–O [°]	Structure resolution [Å]
Mb (MD, current work)	1.79 ± 0.04	147.46 ± 4.60	-
Horse heart Mb [48]	1.87	144	1.3
Sperm whale myoglobin (1HJT) [27]	1.89	112	1.7
Horse heart (MS XAFS) [217]	1.75	150	-

Table 4: Characteristic bond lengths and angles for the active site of MbNO obtained from X-ray structures of NO-bound myoglobins and related heme-containing proteins and the present simulations.

the specific structure and electrostatic environment of the distal pocket.

The motion of the NO molecule is significantly affected by the structure of the active site, especially by the His64 residue. One important determinant in the photodissociation process is the local structure surrounding the bound ligand at the moment of excitation. This is largely determined by the positioning of the His64 side chain. The significant lowering in the probability of the dihedral angle ϕ_1 =(N_a-Fe-N-O) in Figure is 16 (A) caused by excluded volume not easily accessible to the bound ligand due to the proximity of His64 and differs for the two protonation states $His_\delta 64$ and $His_\epsilon 64$. For $His_\delta 64$ one low-probability state exists (at $\approx 60°$) whereas for $His_\epsilon 64$ there are two, see Figure 16 (A). The minima at $\phi_1 \approx 60°$ concide whereas that at $\phi_1 \approx -90°$ only occurs for $His_\epsilon 64$. Further analysis of the trajectories suggests that for $His_\epsilon 64$ the side chain can occupy two distinct states (see Figure 16 (B)) whereas there is only one conformation for $His_\delta 64$.

Although the Fe–ON state has not been characterized experimentally so far in MbNO, considering the ground state ^2A dynamics is still valuable for several reasons. Firstly, it allows to assess the coupling between the ligand motion and the Fe-out-of-plane dynamics and secondly, direct comparison with previous simulations at a considerably lower level of theory is possible. Finally, when rebinding from the ^4A state the NO-ligand crosses to the ^2A state in a large range of geometries and hence exploration and characterization of the full ^2A PES is meaningful and relevant.

Previously, the motion of bound NO has been investigated on a fitted, parametrized two-dimensional PES based on B3LYP//VDZ/3-21G calculations [187]. This PES was

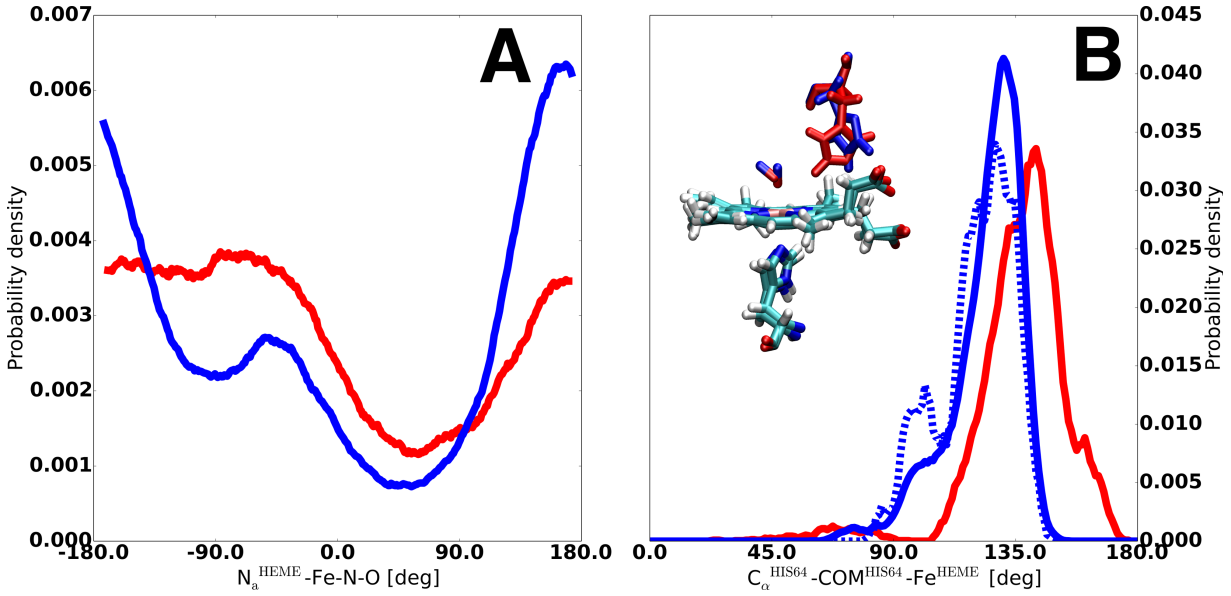


Figure 16: (A) The probability distribution for the N_a-Fe-N-O dihedral angle ϕ_1 (see Figure 21) characterizing the orientation of heme-bound NO with $His_\delta 64$ (red) and $His_\epsilon 64$ (blue). For $His_\delta 64$ one prominent orientation is observed whereas for $His_\epsilon 64$ there are two. (B) The averaged ϕ_2 $=C_\alpha^{His64}$-CoM^{His64}-Fe angle distribution for the two protonation states. The shoulder at $\phi_2 = 100°$ for $His_\epsilon 64$ corresponds to the second state ($\phi_1 = -90°$) in panel (A) (and the blue orientation in the inset) and is typically occupied for tens of picoseconds during the simulations. The blue dashed line is from a single trajectory in which the two states are more clearly separated.

parametrized for the in-plane position of Fe, therefore $\phi = 90°$ throughout. As a first step the performance of the present PES, which relies on a RKHS representation, is compared to simulations on the earlier Legendre expansions with Morse functions for the radial coordinate. As it was already mentioned the ^2A state has two well defined minima (Fe-NO and Fe-ON). The isomerisation barrier for Fe-NO $\rightarrow$ Fe-ON is 23.7 (21.4 previous work) kcal/mol, whereas the barrier for the reverse reaction is around 3.5 kcal/mol (8.4 kcal/mol in previous work) [187]. Part of these differences are due to the considerably larger basis set used in the present work.

To compare results from simulations using the current PES with existing data [187], the Fe–ON$\rightarrow$Fe–NO isomerization dynamics was first studied on the 2-dimensional surface, i.e. for $\phi = 90°$. For this, 250 trajectories were initiated from the Fe–ON local minimum energy structure. The kinetic curves (see Figure 17(A)) were determined from trajectories which lead to isomerization within 250 ps. The ensuing isomerization kinetics are the blue and magenta traces in Figure 17(A) for $His_\delta 64$ and $His_\epsilon 64$

protonation, respectively. For the kinetic curves, isomerization was considered to occur if the NO coordinates fulfill the criterion ($R = 2.55$ Å and $\theta = 130°$). The typical isomerization time is on the 100 ps time scale. This is in qualitative agreement with previous crossing times which ranged from 50 ps to 600 ps [187] using about 10 trajectories, from which, however, no rigorous distribution of the crossing times could be determined. The 100 ps found here qualitatively agree with the earlier work and the shorter isomerization times found here are explained by the lower barrier in the present PES. It is concluded that the two different PES parametrizations (Morse plus Legendre versus RKHS-based Legendre here) yield qualitatively similar isomerization times with the present simulations exhibiting more rapid dynamics due to the lower isomerization barrier. Also, the difference between the two protonation states of His64 are small; the yield within 250 ps differs by about 10 %.

Next, the dynamics on the 3D-PES explicitly including the Fe-oop motion was investigated. It was found that within the maximum simulation time of 250 ps a considerably smaller number of trajectories completed the conformational transition, see red and green traces in Figure 17(A). Compared to ≈ 80 % ($His_\delta 64$) and 95 % ($His_\epsilon 64$) on the 2-dimensional PES only ≈ 25 % ($His_\delta 64$) and 50 % ($His_\epsilon 64$) isomerize when running the simulations on the 3-dimensional PES including the Fe-oop coordinate. This suggests that once the NO ligand rebinds to the ^2A state coming from the ^4A state its dynamics is strongly coupled to the Fe-oop motion. It is evident that not only a smaller fraction of trajectories leads to isomerization but the kinetics is also slowed down compared to the 2D-simulations in which the Fe-oop coordinate is not explicitly included and the iron is always in an in-plane position, available to bond-formation.

Figure 17(B) shows representative isomerization trajectories for $His_\delta 64$ leading to Fe–NO using the 3D potential. First, the Fe-O bond is broken and the NO molecule leaves in the configuration, which is similar to the initial Fe-ON minimum. Then, being affected by the protein active site, the ligand rotates and rebinds to the heme-Fe in the thermodynamically favoured Fe-NO configuration. Depending on the detailed dynamics such a transition lasts up to 10 ps, see Figure 18. When leaving the Fe-ON minimum the heme-Fe moves below the heme-plane, making it more difficult to be accessed by the NO molecule when forming the Fe-NO bond compared to the situation when the Fe-oop coordinate is not explicitly included in the PES.

The isomerization rates using the 3D-PES for $His_\delta 64$ in Figure 17(A) (red trace) can be fit to a single exponential $1 + a_1(e^{(-(t-\tau_1)/t_1)} - 1)$ where $\tau_1 = 12.7$ ps is the lag time, $t_1 = 215$ ps is the characteristic time and $a_1 = 0.25$ is the amplitude. The lag time corresponds to the shortest transition time in the data set. For $His_\epsilon 64$ a fit involving two exponential processes is required which results in $\tau_1 = 2.0$ ps, $t_1 = 131$ ps and $a_1 = 0.38$ for the short component and $\tau_2 = 13.1$ ps, $t_2 = 163$ ps and $a_2 = 0.28$ for the long component. It is noted that the long component (dotted green line) for $His_\epsilon 64$ is almost a fit to the raw data for the isomerization with $His_\delta 64$ (red dots). This suggests that they probe the same conformational substate ($\phi_1 = 60^\circ$ in Figure 16). This is further supported by $\tau_1^{His_\epsilon} \approx \tau_2^{His_\delta}$. Hence the rapid process, which only appears for $His_\epsilon 64$, corresponds to the alternative orientation $\phi_1 = -90^\circ$ in Figure 16 (A).

As already mentioned, the existence of the Fe-ON metastable state has only been established for model systems but not for NO in Mb. Evidently, the PES does support such a state but whether or not it can be observed experimentally also depends on the energetic ordering of the ^2A and ^4A state. Asymptotically, the two states are separated by $\approx$ 5 kcal/mol (see Figure 14) which is close to the stabilization of the Fe-ON minimum relative to the transition state separating it from the Fe-NO state. Furthermore, the protein environment also modulates the energetics of the two electronic states. Taken together, although the DFT calculations establish that the Fe-ON state is a local minimum, observing it in MbNO depends on a delicate balance between the energetics and response to environmental effects of the ^2A and ^4A PESs. The present data suggest that a Fe-NO state has not been found in experiments on MbNO because the ^4A state is lower in energy in this region of phase space, see also Figure 15. Nevertheless, depending on the active-site electrostatics this may be different for mutated myoglobins or other proteins of the globin family because the energy differences between the electronic states involved are small.

4.3.3 *The ^4A state*

The ^4A photodissociated state is characterized from 50 trajectories, each 500 ps in length. The initial configuration of the system was taken from the ^2A state simulations with an instantaneous excitation by switching the effective PES to the one of the ^4A state. This leads to rapid motion of the NO molecule away from the heme-Fe

and consequently diffuse to neighboring protein cavities. The regions accessed by the free NO ligand are illustrated in Figure 19 and include the B state and the Xe4 pocket. These sites match well with the Xe ligand binding pockets found in the X-ray experiments and in previous simulations [10, 146, 249].

After breaking the Fe-NO bond the Fe immediately starts to move below the heme-plane on a ~ 100 fs time scale. This has already been found from optical experiments and previous MD simulations [129, 143]. The slower phases are analyzed in the same fashion as in previous work [129], i.e. the distance of the Fe to the average heme plane and a sliding window of 2 ps was used. However, the data from 10 instead of 4 trajectories, each 200 ps in length, was averaged. The present work (see Figure 20) also finds two additional time scales and a characteristic beating pattern which suggests complex structural changes in the protein. Contrary to the previous analysis, the current work leads to an Fe-oop distance of ≈ -0.28 Å (instead of –0.60 Å) [129] which reproduces the experimentally measured result (–0.27 Å) [225]. However, this number had not been known at the time of the earlier two studies [129, 143]. The time constants were determined from fitting to $ae^{-t/\tau_1} + be^{-t/\tau_2} + c$ and yields $\tau_1 = 3.5$ ps and $\tau_2 = 64.4$ ps, as shown in Figure 20. This compares favourably with the analysis of the optical experiments which find 3.5 ps and 83.0 ps for the two processes on the picosecond time scale, respectively [143].

4.4 SUMMARY AND OUTLOOK

The present work introduces kernel-based PESs into atomistic simulations of biomolecular systems. Specifically, it is possible to carry out simulations at DFT-quality for chemically challenging systems, such as metal centers in proteins, at the speed of regular force fields. For this, the energies of the embedded model system (here heme-His-NO) are precalculated with quantum chemistry and the PES of the relevant coordinates is represented as a RKHS. The implementation yields energy conservation comparable to that of a usual force field. Generalization to more degrees of freedom will require additional technical developments [89, 158] also because the number of reference energies scales exponentially when more degrees of freedom are included. However, extensions to 4 or 5 active degrees of freedom are readily possible.

For NO in Mb the present work clearly shows that including the Fe-out-of-plane coordinate has a profound influence on the ligand's active-site dynamics. This will be of particular interest in better characterizing the rebinding dynamics of photolyzed NO for which recent experimental work suggested the existence of a NO-bound-like, Fe-out-of-plane metastable state [127]. Reactive molecular dynamics [51, 172] involving the present ^{2}A and ^{4}A kernel-based PESs will provide the necessary detail for an atomistically resolved picture linking experiment and molecular dynamics.

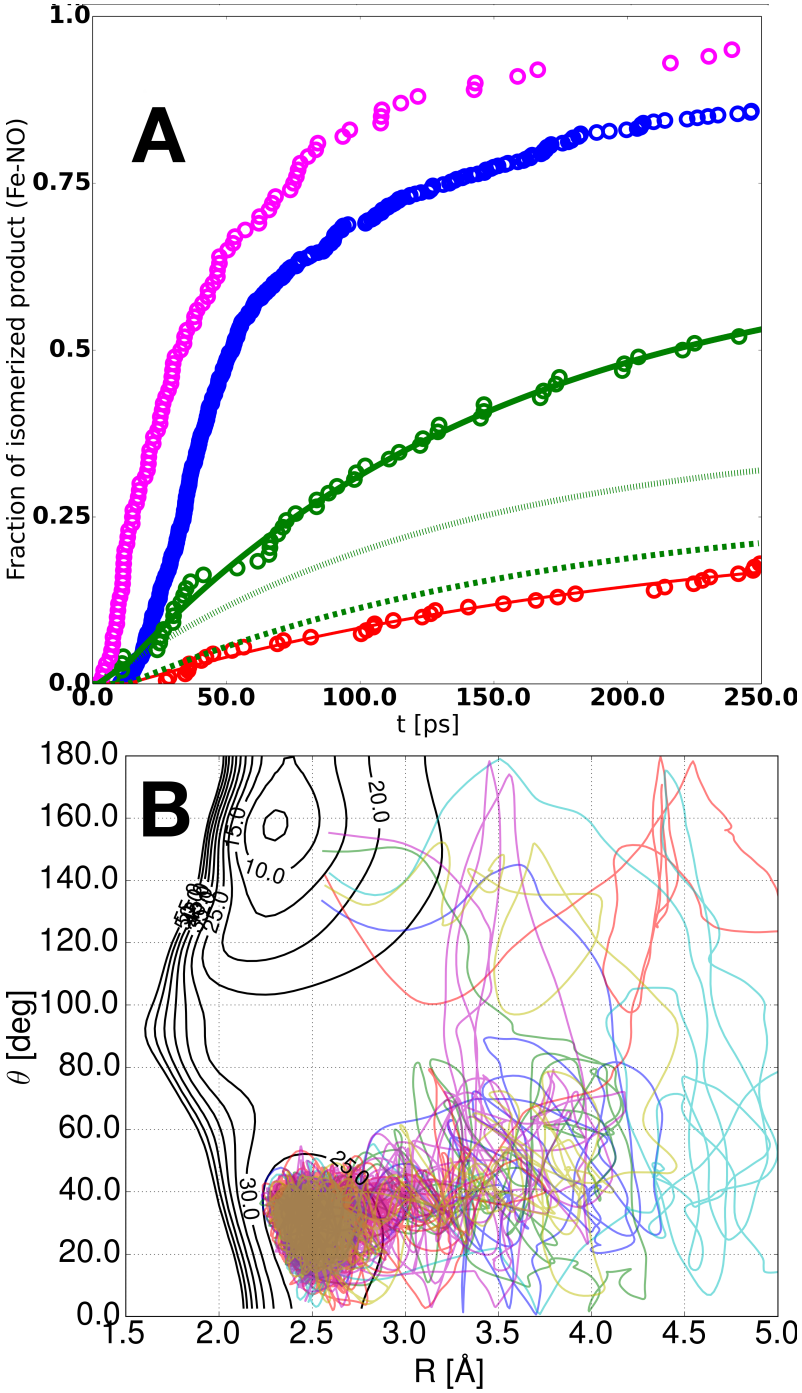


Figure 17: (A) NO isomerization kinetics (fraction of product formed as a function of time) on the ^{2}A PES with (3D - red and green for His$_\delta$64 and His$_\epsilon$64 protonation, respectively) and without (2D - blue and magenta for His$_\delta$64 and His$_\epsilon$64 protonation, respectively) explicitly taking into account the Fe-oop coordinate. Fits to single and double exponentials (see text) are in solid lines. The short (dashed green) and long (dotted green) components describe the two processes found for His$_\epsilon$64 protonation. (B) Projections of several representative trajectories for the Fe-ON→Fe-NO isomerization dynamics on the 3-dimensional PES including the Fe-oop motion, superimposed on the ^{2}A PES. The Fe-ON state has $\theta = 30°$ whereas the Fe-NO state is characterized by $\theta = 150°$.

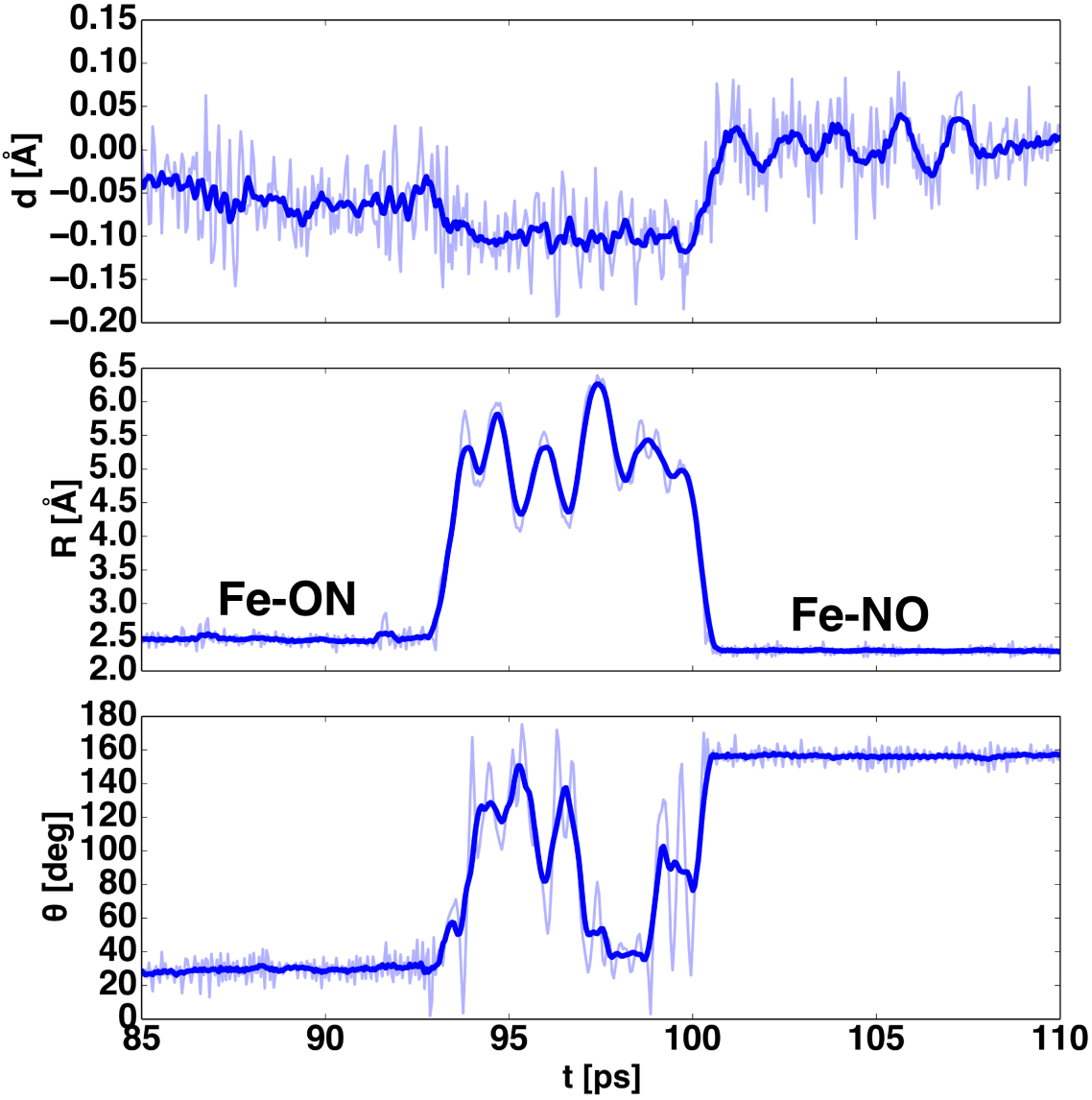


Figure 18: One Fe-ON→Fe-NO isomerization trajectory highlighting the coupling between ligand and Fe-oop motion. During the time the ligand samples regions away from the heme-Fe the iron atom moves below the heme plane (between 95 and 100 ps).

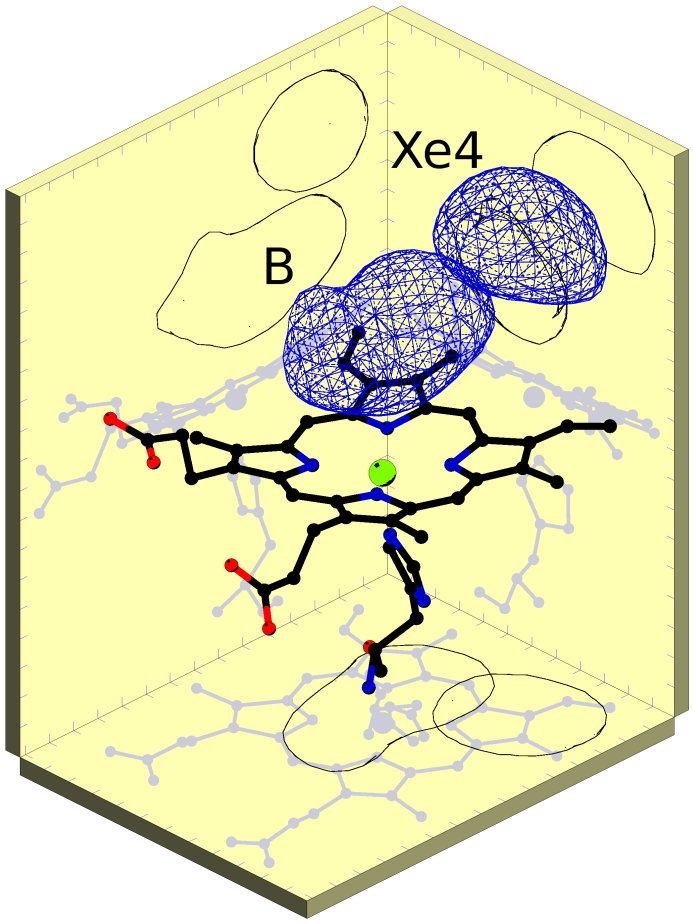


Figure 19: Effective volume accessed by the NO molecule in the ^{4}A state. It includes the B and Xe4 states.

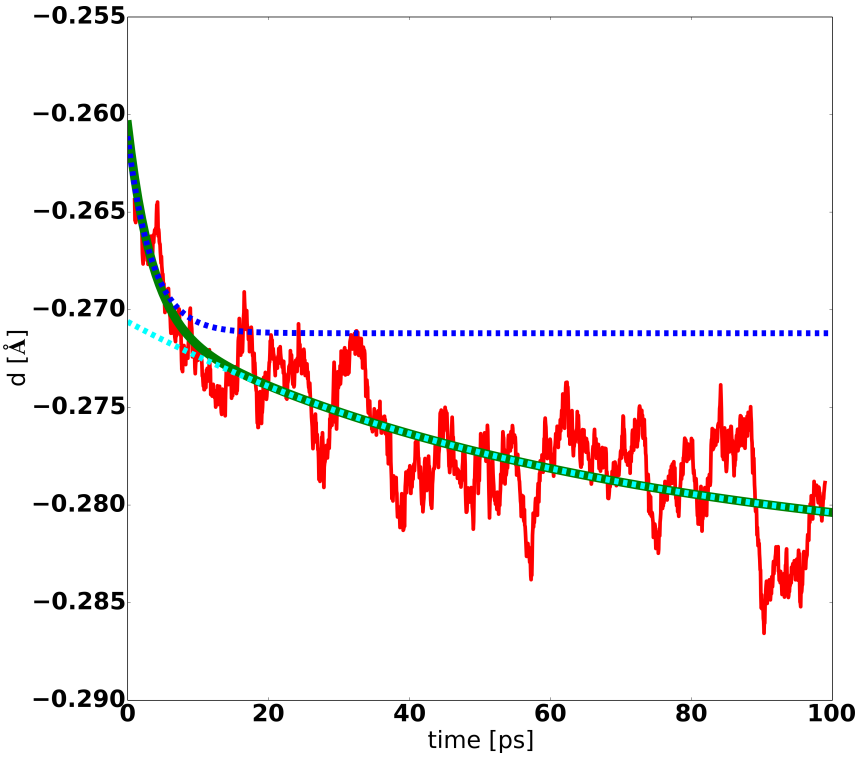


Figure 20: Averaged time dependence over a 2 ps sliding window of the Fe-oop motion after photodissociation of the ligand and subsequent dynamics on the ^{4}A PES. The ultrafast component is averaged out and not visible (see text) and only the two subsequent, longer time scales (blue and cyan dotted curves) are shown together with the overall fit (green solid line).

5

STRUCTURAL INTERPRETATION OF METASTABLE STATES IN MBNO

5.1 INTRODUCTION

Characterizing structure, motions and functional dynamics in biomolecules is essential for understanding their function. The motions usually involve stable states at their endpoints, separated by one or several metastable states (intermediates) in between.[65, 162] Experimentally, it is possible to directly and structurally characterize states with sufficiently long lifetimes.[225, 229] However, when the states live for times too short to stabilize only indirect means to infer their existence are available. Optical spectroscopy can reveal the existence of short lived states through spectral shifts. More direct interrogation of the chemical environment of the iron atom is possible with X-ray absorption spectroscopy (XAS)[236] or multidimensional spectroscopy.[24, 163] However, associating time scales with particular geometrical arrangements of the atoms is usually indirect, except for favourable cases.[229] Under such circumstances computational investigations including the dynamics of the system of interest becomes a meaningful complement.

Following the motion of photodissociated ligands in globular proteins has a long history. Small molecules which can reversibly bind to the protein active center are ideal and sensitive probes of the interior of such complex systems. Nitric oxide (NO) is a physiologically relevant ligand[150, 197, 251], involved in modulating blood flow, thrombosis, and neural activity. Experimentally, the binding kinetics of NO to the heme-group in myoglobin (Mb) has been studied extensively with time resolved spectroscopies, ranging from UV/visible to the mid-IR[49, 106, 122, 123, 127, 205, 280] and resonance Raman[127] techniques. In all cases the rebinding kinetics is multi-exponential with time constants ranging from sub-picoseconds to several hundred picoseconds. However, they can be grouped into two broad classes: processes on the 10 ps time scale and those on the 100 ps time scale.[52, 122, 127, 191, 205, 280]

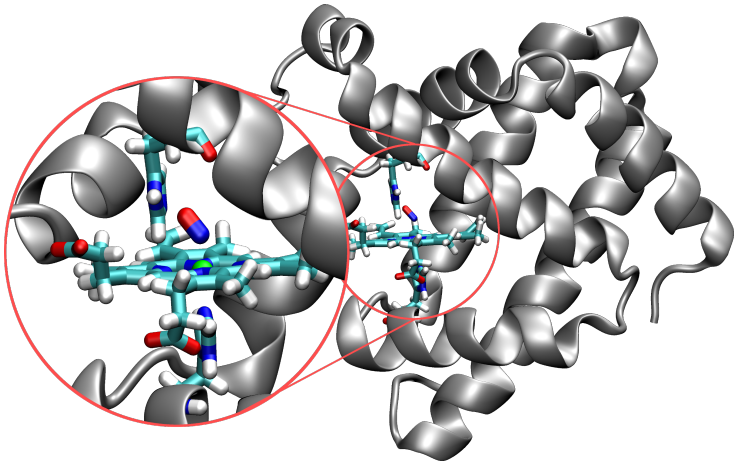


Figure 21: NO-bound Myoglobin with the heme, His93 and NO ligand in licorice and the protein in secondary structure representation.

Recent experiments have followed the interplay between the Fe-out-of-plane (Fe-oop) and the NO-ligand motion in a time resolved fashion.[127, 236, 280] This work points towards a direct coupling between these two degrees of freedom on the 10 to 100 ps time scale. What remains unclear is the structural characterization of this state. Furthermore, earlier studies have attempted to characterize the metastable Fe-ON state[182] and the question remains why this commonly observed motif in model compounds[277] is absent in the protein environment despite a stabilization of 5 kcal/mol or more.[187]

The atomistic dynamics is essential for protein function and provides the basis to interpret the time scales involved at a structural level. Molecular dynamics (MD) simulations are ideal for addressing such questions. Together with validated energy functions such simulations can provide the "missing link" between time resolved experiments and the underlying atomic motions.[137, 189, 229] For NO interacting with heme in Mb such potential energy surfaces (PESs) for the bound ^{2}A and ligand-unbound ^{4}A state are available (see Chapter 4.2.1)

Running a statistically significant number of QM/MM trajectories from which to analyze and resolve the interplay of the motions involved is beyond current computational methods for systems such as MbNO. For quantitative and meaningful computational investigations, suitable representations that preserve the accuracy of the interpolated PES are used in the simulations. Here, a parameter-free, reproducing kernel Hilbert space (RKHS)-based representation which exactly reproduces the refer-

ence data from quantum chemical calculations is employed (for details on the PESs, see the Chapter 4.2.1). [7, 93] RKHS is based on smoothness criteria of the interpolant and has been successfully applied to study van der Waals complexes.[90, 166]

The present work investigates the reactive dynamics between the ^{4}A and ^{2}A PESs following photodissociation of the NO ligand from the heme-iron in Mb. Due to the close energetic ordering of the two states and the existence of two substates on the ^{2}A PES, interesting dynamics can be expected. From comparison with experimentally determined rebinding time scales and the spectroscopies, structural questions underlying the molecular dynamics can be addressed.

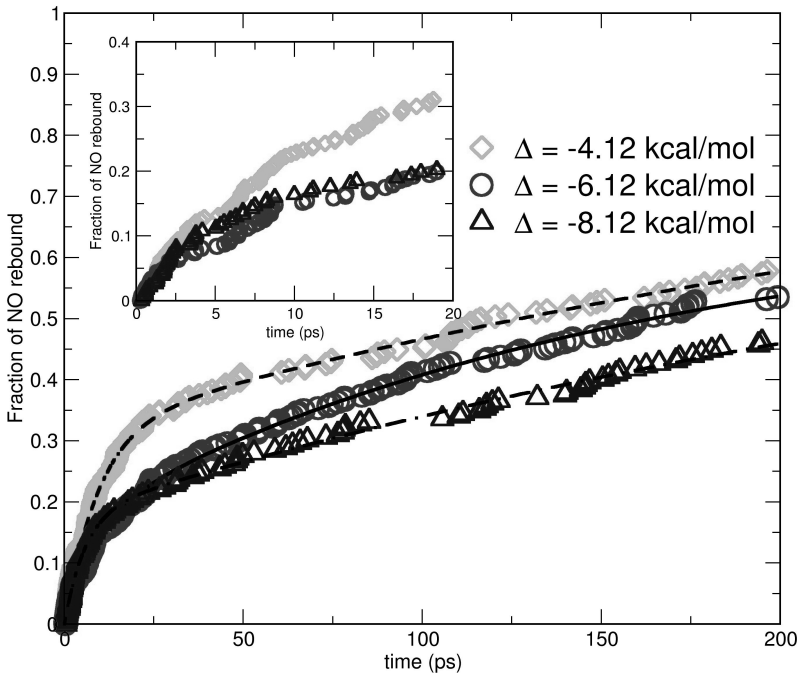


Figure 22: The kinetics and the corresponding exponential fits for NO rebinding to the heme-Fe after photoexcitation. Simulations were run for three different values of the asymptotic shift Δ.

For following the reactive dynamics between the ^{2}A and ^{4}A states, 300 independent simulations were run for a maximum time of 200 ps or until the bound Fe-NO state was formed. Photodissociation was induced by instantaneously switching the force field to the ^{4}A state[165] which introduces around 50 kcal/mol of energy (see Chapter 4), comparable with the photon energies used in experiments (49 to 81 kcal/mol).[143] The rebinding kinetics is reported in Figure 22 and shows the fraction of rebound NO as a function of time. The rebinding kinetics on the

sub-nanosecond time scale follows a multiexponential decay with two time constants $\tau_1 \approx 10$ ps and $\tau_2 \approx 150$ ps, also depending on the asymptotic separation Δ of the two states (see Table 6). The shift Δ relates the asymptotic energies of the ^2A and the ^4A PESs for NO at infinite separation from the heme-Fe (see Chapter 4.2.1).

Typical ligand trajectories are shown in Figure 23. Panel A reports rebinding on the 1 ps time scale whereas panels B and C are representative of rebinding time scales τ_1 and τ_2. For τ_1 the ligand only samples the immediate neighborhood of the heme-iron before rebinding. On the 100 ps time scale rebinding occurs from regions further away. For $\Delta = -6.1$ kcal/mol the total number of rebound trajectories is 17 (rebinding within 2 ps), 28 ($2 \leqslant t < 10$), 115 ($10 \leqslant t < 200$), and 140 rebind on time scales longer than 200 ps, i.e. the rebinding efficiency within 200 ps is ~ 55 %. This compares with a rebound fraction of 75 % on the 200 ps time scale from recent XAS measurements.[236]

The simulations described so far used the PESs fitted to the DFT-data and including environmental effects (see Chapter 4.2.1). Corresponding simulations with 200 independent runs on the DFT-only PESs yield rebinding of 195 trajectories within 200 ps with rebinding times of 2 and 15 ps, i.e. one order of magnitude more rapid than with the refined PES treating the Fe-oop coordinate more accurately and explicitly. This finding is also consistent with previous simulations based on DFT-only PESs.[52, 191] Hence, explicitly including the Fe-oop motion is essential for quantitative results.

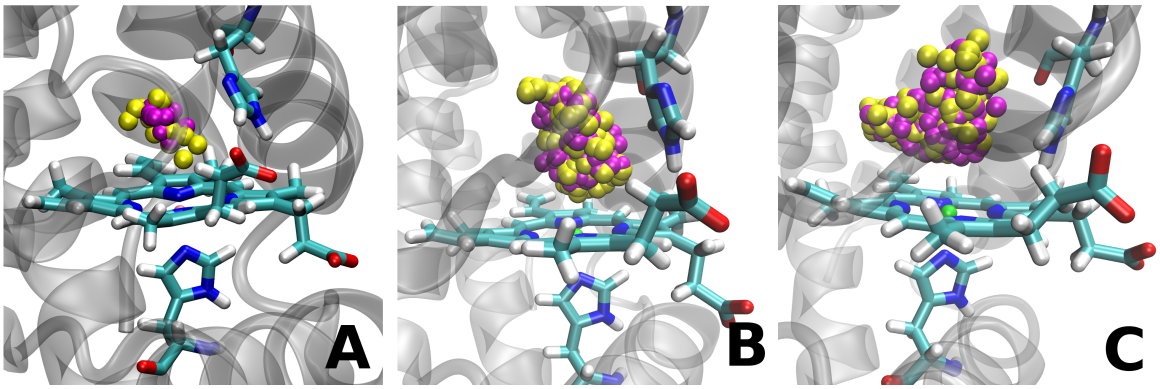


Figure 23: Typical trajectories for different rebinding time scales. NO positions are shown in yellow (N) and magenta (O). (**A**) - the picosecond process ($\tau = 1.6$ ps), (**B**) - the 10 picosecond process ($\tau = 42.1$ ps), (**C**) - the 100 picosecond process ($\tau = 160.2$ ps).

A complementary view of the different rebinding time scales can be gained from analyzing the maximal distance between the Fe and the ligand sampled during a

rebinding simulation. The 300 trajectories were grouped in those that rebind on very short time scales ($\tau < 2$ ps), within 10 ps and longer than 10 ps. The individually normalized probability distribution functions (pdfs) are shown in Figure 38. No NO migration to neighboring xenon-binding sites is found on the 200 ps time scale because, e.g., escape to the closest Xe-site (Xe4) occurs within 1 to 10 ns.[23, 210]

Previous experimental and computational investigations[52, 120, 123, 165, 191, 205, 278] found that the recombination kinetics of MbNO is non-exponential, involving two to three time scales, depending on the model used. The earliest works report time scales of 27.6 ps and 279.3 ps in a double exponential fit.[205] Later, optical and infrared experiments found short time scales ranging from 5 to 30 ps and longer time scales between 100 and 200 ps.[120, 123, 278] The experiments all agree on the existence of two sub-nanosecond time scales which differ by about one order of magnitude. Previous computations with reactive force fields find typically somewhat shorter rebinding times, between 5 ps and 20 ps,[52, 191] depending on the asymptotic separation Δ used.[11] The current simulations find multiple time scales in the 10 to 100 ps range in agreement with optical and infrared experiments. Also, a more rapid component on the 1 to 2 ps time scale is present. The influence of Fe-doming is clearly visible in the rebinding kinetics, see Figure 22. This validates the force fields used here and allows to address a number of hypotheses put forward recently.

5.2 THE FE-OOP, NO-BOUND STATE

First, direct contact with a recently proposed transient structure in the ^{2}A state can be made. Picosecond time resolved[127] and XAS[236] experiments have suggested that the Fe-out-of-plane (Fe-oop) and the NO-ligand motion are closely coupled. The interpretation of the experiments provided evidence for an Fe-oop, ligand bound structure with a lifetime of 30 ± 10 ps. The present simulations on the ^{2}A PES alone indeed show transient stabilization of such a state. The maximal lifetime found was 27 ps and 41.7 % of the trajectories showed such a state (see Figure S3). Despite underestimating the experimentally reported lifetime of 30 ± 10 ps the present simulations support the existence of such a state.

5.3 THE FE–ON STATE

Another structurally elusive state which has been found for heme-model compounds is the Fe–ON isomer.[30, 35, 277] Such a structure is sufficiently stabilized (by several kcal/mol) in NO-bound Mb which should make it observable in infrared experiments.[187] However, no such state was found in experiments.[182] Simulations on the ^2A state alone show that the ligand is stabilized for tens of picoseconds in the Fe–ON state. However, when using the reactive $^2A +^4 A$ PES with a shift of $\Delta = -6.1$ kcal/mol (which best reproduces the DFT energies) the Fe–ON configuration can not be stabilized because in the region of the Fe–ON minimum the repulsive ^4A state is lower in energy than the ^2A state. Hence, the ^2A minimum of the Fe–ON state is masked by the ^4A state and can not be stabilized in wt Mb.

The asymptotic separation Δ depends on the chemical environment around the heme-group. Hence, Δ changes if the environment is modified, e.g. through mutation of amino acids or embedding the heme-group into a different protein. The effect of this can be quantified with simulations on the reactive $^2A +^4 A$ PES with different values for Δ. Reducing the asymptotic separation by 5 kcal/mol ($\Delta = -11.1$ kcal/mol) further destabilizes the Fe–ON state (Figure 39A). For $\Delta = -6.1$ kcal/mol (Figure 39B) the Fe-ON state is insignificantly sampled and the system rebinds efficiently into the Fe-NO state. Contrary to that, reducing the asymptotic separation to $\Delta = -1.1$ kcal/mol (Figure 39C) the Fe-ON state is populated for extended periods during which spectroscopic characterization of this state should be possible. As an illustration of this, recent spectroscopic work found a 2.5 times slower dynamics of NO rebinding in Mb compared to cytochrome c due to different active site architecture.[102] Based on transition state theory, a factor of 2.5 corresponds to ≈ 0.5 kcal/mol in energy. The thermodynamic stability of Mb upon mutation has been found to change by -2 to $+6$ kcal/mol[118] which suggests that modifications in the active site can potentially stabilize the Fe–ON conformation through differential stabilization of the bound state relative to the unbound state.

5.4 STRUCTURAL INTERPRETATION OF REBINDING TIME SCALES

With the coordinates from the reactive MD simulations it is possible to provide a structural interpretation of the $\approx$ 10 ps and $\approx$ 100 ps time scale found in the rebinding simulations. It is of particular interest to investigate in what respect(s) the dynamics between these two states differ. Probability distribution functions of the a) NO-ligand, b) all active-site residues, and c) all His64 side chain atoms projected onto the heme plane are shown in Figure 24.

The pdfs show that the NO-ligand explores a much larger space for long rebinding times (Figure 24 upper right hand panel) than for short rebinding times (Figure 24 upper left hand panel). Furthermore, the NE2 atom of residue His64 occupies two clearly distinguishable states (A and B) in trajectories with long rebinding times which are absent for rebinding on $\tau_1 = 10$ ps (Figure 24 middle row). The middle row emphasizes that for time scales τ_1 the His64 NE2 atom occupies space away from the iron (located at (0,0)) whereas for longer time scales the atom pushes in towards the heme-Fe by almost 2 Å which hinders ligand rebinding. The estimated forward $\Delta G^{NE2}_{A\to B} = 2.5$ kcal/mol and reverse $\Delta G^{NE2}_{B\to A} = 4.0$ kcal/mol barriers suggest that on the 100 ps time scale state A is destabilized relative to state B. Corresponding barrier heights for the other His64-side chain atoms range from 0.5 to 1.5 kcal/mol. Hence, on average state B is separated from state A by a barrier of $\approx$ 2 kcal/mol with state B lower in energy than state A.

The two states (A: green; B: red) are shown in Figure 25 together with the X-ray structure 1HJT[27] (gold). State A ("His64 out") is associated with the short rebinding time scale (10 ps) whereas state B ("His64 in") corresponds to the slow (100 ps) component. The barrier height corresponds to an interconversion time A$\leftrightarrow$B on the sub-nanosecond time scale which is supported by the explicit simulations and compares well with previous findings for CO-bound Mb with interconversion times of a few hundred picoseconds.[153, 163]

For direct contact with the XAS experiments, XANES spectra were computed (see Appendix B) for randomly selected MD-sampled structures with NO-bound (40 structures), NO-unbound (NO within 3.5 Å of the heme-Fe; 20 structures) and NO-unbound (NO within 5.0 Å; 10 structures). For the bound state the computed spectrum ($I_b(E)$) agrees well up to 7.15 keV. For higher energies the absorption signal

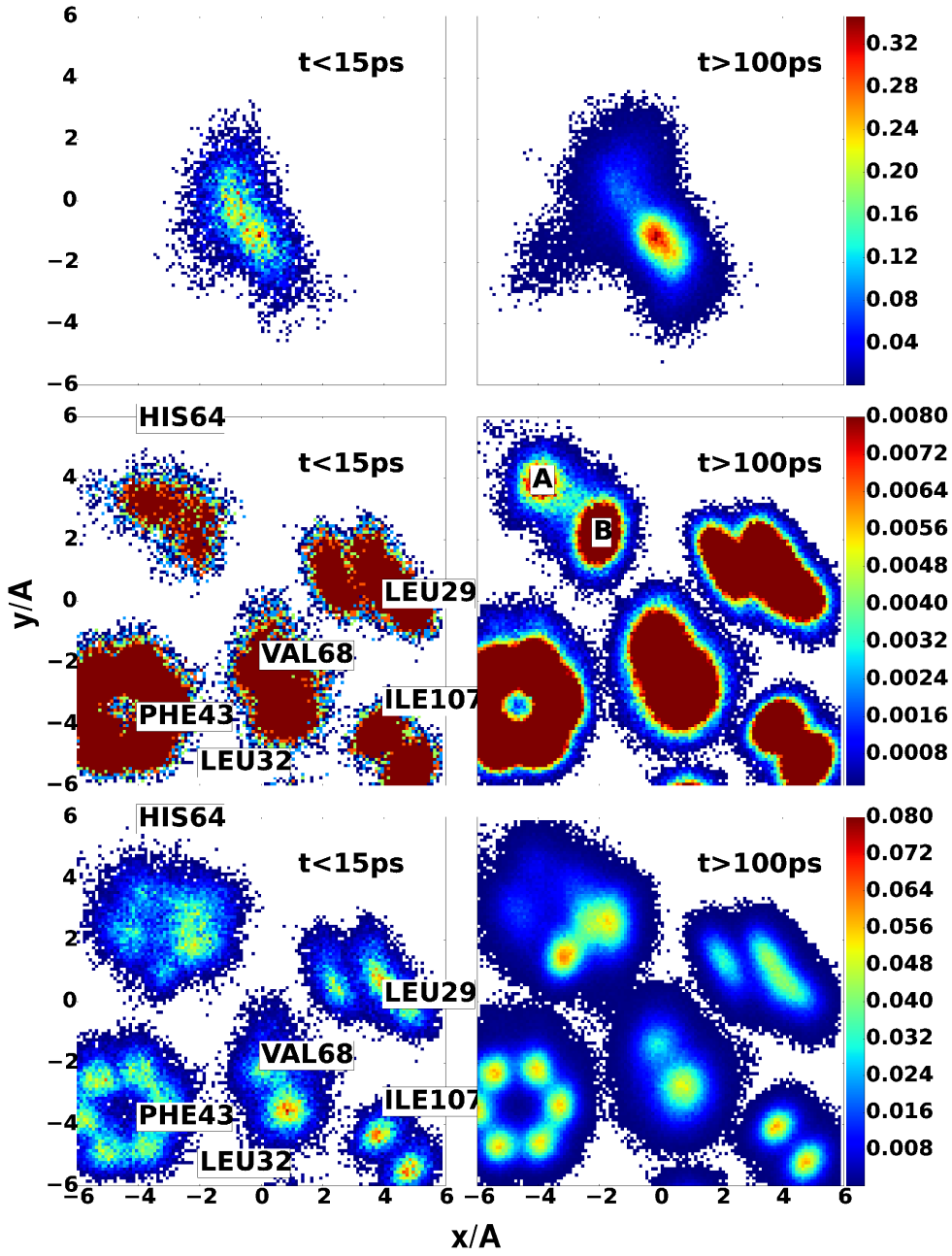


Figure 24: Top Panel: (x, y)—probability distribution function for the free NO ligand for rebinding on the $t < 15$ ps (left) and $t > 100$ ps (right) time scale. Middle Panel: (x, y)—probability distribution function for the NE2 atom of His64 on the two time scales. The two states for NE2 are labeled (A, B) and clearly distinguishable. Also shown are pdfs for all side chain atoms of Phe43, Val68, Leu29, and Ile107. For Phe43 the phenyl ring is always parallel to the heme-plane and all 6 carbon atoms are clearly distinguishable. Bottom Panel: as in the middle panel but for the entire side chain of His64 and with different maximum height of the pdfs.

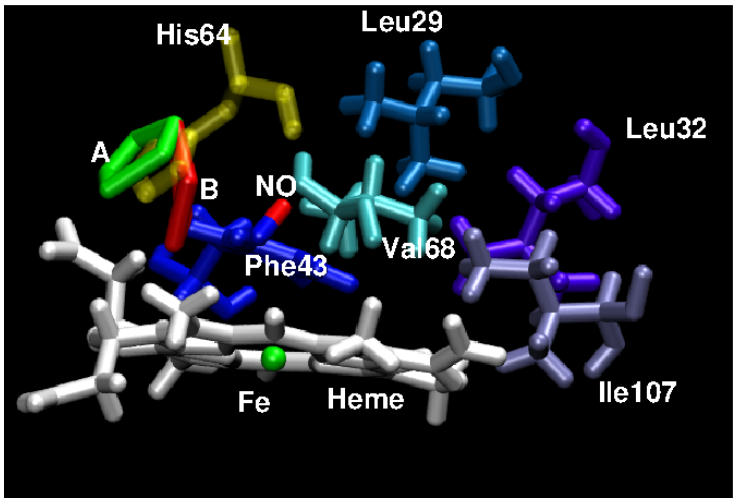


Figure 25: States A (green) and B (red) as found from the per-atom pdfs of the His64-side chain atoms. In gold the X-ray reference structure.

is correctly described but the computed intensity is too high (see Figure 26 top panel). The unbound structures yield $I_{ub,s}$ and $I_{ub,l}$ for the short and long Fe–NO separations, respectively, and nearly superimpose (inset Figure 26) despite the different NO separations. Upon ligand photodissociation the peak at 7.12 keV shifts to lower energy, the intensity between 7.12 and 7.15 keV decreases and between 7.15 and 7.18 keV increases compared to the spectrum for the bound state. All these features are consistent with previous experiments on MbNO and deoxy-Mb.[127, 236]

The experimentally observed transient at 50 ps is compared with the averaged absorbance differences $\Delta I_{ub,s} = I_b - I_{ub,s}$ and $\Delta I_{ub,l} = I_b - I_{ub,l}$ in Figure 26. The two computed difference spectra are quite similar - except for a slightly reduced amplitude around 7.13 keV and 7.15 keV and an enhancement around 7.16 keV for $\Delta I_{ub,s}$ compared to $\Delta I_{ub,l}$ - despite the two very different ligand positions. In both cases the Fe-oop position ranges from 0.3 to 0.15 Å below the plane although for $\Delta I_{ub,s}$ positions closer to in-plane ($d = 0$) also occur. Compared to the experimental transient, $\Delta I_{ub,s}$ and $\Delta I_{ub,l}$ trace the major features but differ from it in the width of the 7.12 keV peak and the behaviour between 7.13 and 7.15 keV. Figure 40 suggests that presence or absence of photodissociated NO affects the XANES spectrum over the entire energy range from 7.1 to 7.15 keV and not just around 7.15 keV as previously assumed[236] because the signal also depends on the motion of the Fe-atom relative to the heme-plane. Within the signal-to-noise of the experiment it can not be distinguished whether the ligand is close to the heme-Fe or further away from it. It

should be recalled that experimentally, a mixture of NO-bound and NO-unbound structures is measured because the photolysis yield is not 100 %.

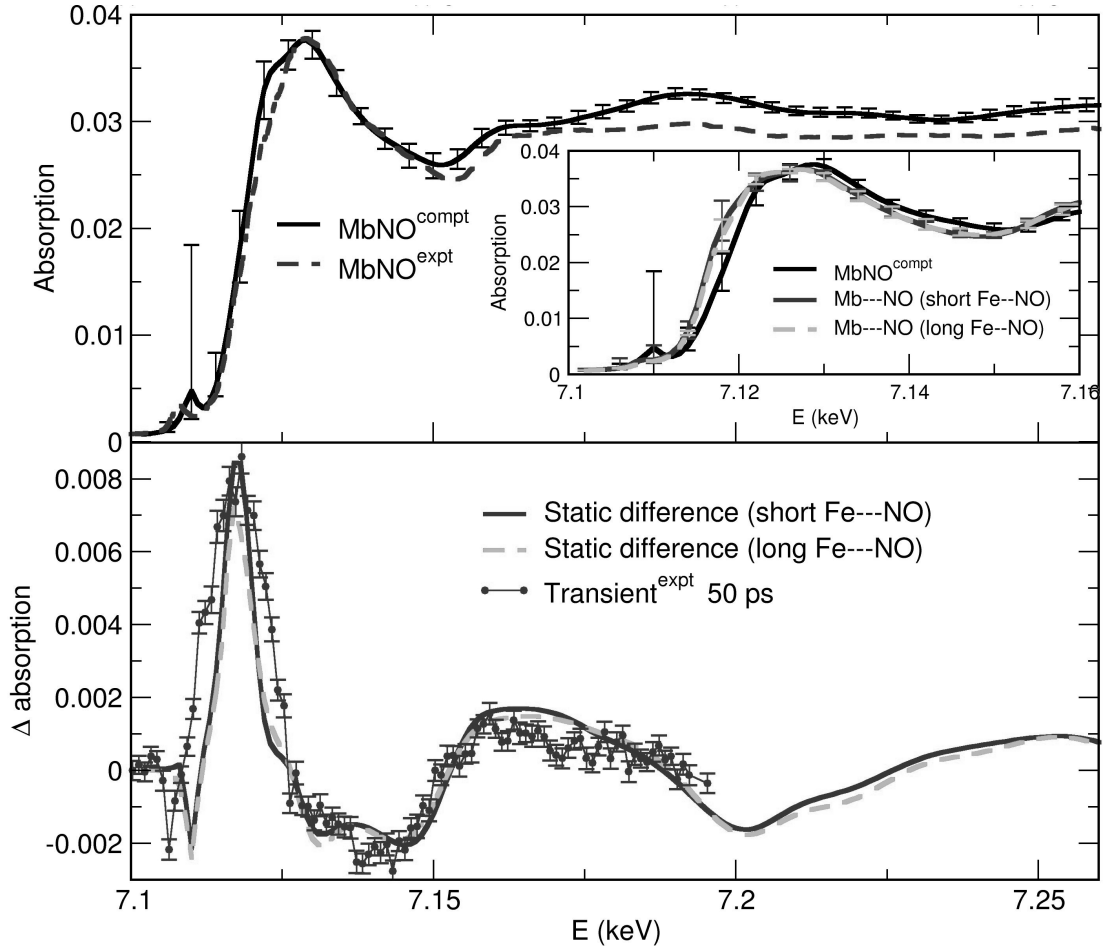


Figure 26: Top panel: Computed XANES spectra for MbNO (black) compared to experiment (dashed). Vertical bars indicate minimum and maximum absorption for all snapshots and are caused by conformational sampling. The inset compares computed spectra for MbNO and 2 sets of photodissociated systems (solid and broken grey, see text). Bottom panel: Static difference spectra (solid and broken grey) compared with the experimental transient at 50 ps (red).

Reactive MD simulations using MS-ARMD yield nonexponential kinetics for ligand rebinding. The time scales (10 and 100 ps) confirm those from optical and infrared experiments. The influence of the Fe-oop coordinate on the rebinding reaction has been directly established. The two time scales are associated with two structurally different states of the His64 side chain - one "out" (state A) and one "in" (state B) - which control ligand access and rebinding dynamics. Although energetically feasible, the ^{2}A Fe–ON metastable state is likely to be unobservable in wild type Mb because

in this region of configurational space the repulsive ^4A state is lower in energy and prevents stabilization. The present work supports a recently proposed, transient Fe–oop/NO-bound structure with lifetimes of up to 30 ps. The computed XAS spectra are compatible with experimentally measured ones but are unable to distinguish between structures with photodissociated NO "close to" or "far away" from the heme-Fe in the active site. The present work provides an atomistically refined picture and structural explanations and assignments for a range of experimental observations, all sensitive to the Fe-oop dynamics of NO after photodissociation in native Mb.

6

GENERAL CONCLUSION AND OUTLOOK

The research of heme protein continues for already 150 years. After the discovery of heme as a cofactor in hemoglobin, it was found in many other enzymes and protein complexes, which carry out very basic functions essential for the survival of living organisms. Such an importance of heme proteins leaded to the intensive studies of their structures and functions, their role in physiological mechanisms and metabolism chains, these include energetic functions (like cytochrome c complexes in mitochondria), oxygen transportation (hemoglobin), oxygen storage (myoglobin) and many other.

The structural identification of hemoglobin and myoglobin using X-ray diffraction method started the revolution in the field: new experimental methods, such as protein X-ray crystallography, elastic neutron scattering, new spectroscopic techniques like 2D IR and Raman spectroscopy, were developed, structural data provided the means to interpret results of spectroscopic experiments. Then new questions regarding the relevance of the protein dynamics in enzymatic reactions have arisen: does dynamics of the protein structure matter? The physiological conditions, at which proteins function, are different from the ones in a crystal. Moreover the crystallography could not effectively capture the conformational states or substates relevant to the function of the protein. It was found that proteins are nanoscale mechanisms and protein matrix has largely affects the the flow of the enzymatic, and often determines the specificity, activity and function of the protein.

Often the protein structure determined in the crystallographic experiments was found to be inactive. UV/vis, IR, Raman and microwave spectroscopy can provide information about the changes to the structure in proteins in time or during enzymatic reaction, but the information, they provide, is not directly structural and requires interpretation based on the structural model. NMR techniques (like NOESY) can provide structural information about the protein in conditions, close to the physiological, but the time resolution is limited to milliseconds-seconds and the method can be applied only to rather small proteins. The overall picture can be described only

by combining diverse experimental and theoretical/ computational techniques. MD simulations were originally used to describe the dynamics of disordered systems. The timescales accessible by MD methods are from femtoseconds to milliseconds, which is the time range of many reactions of physiological interest.

In this work MD simulations were used to describe the kinetics and protein dynamics during the processes involving a small external ligand in heme active site of two proteins: myoglobin and cytochrome c oxidase. The systems involving these proteins are experimentally well studied, providing a good reference for our computational models.

Cytochrome c oxidase is an important part of the protein complex, that carries out oxygen reduction to water, this reaction takes place in the binuclear heme $a_3 - Cu_B$ active site of the protein. The ultrafast dynamics of CcO-CO system was previously investigated using experimental methods, providing time-resolved spectroscopic information. After photoexcitation of heme-CO, the CO moves away from the heme a_3 and with a high probability ($\approx$ 90 %) binds to Cu_B site. The characteristic time of the transfer of $\approx$ 300 fs found from the present study agrees qualitatively with that from the previous experiment investigations (450 fs). It was also shown that the process of the transfer is mainly ballistic with negligible stochastic (diffusion) component. This study has also shown that the heme doming largely affects the reactivity of the heme active site by reducing the probability of the consecutive rebinding.

Myoglobin does not contain the second reactive center in the active site, which might interfere small ligand dynamics. It enables the investigation of the impact of the Fe out-of-plane motion on the NO rebinding following photodissociation. The MD simulations, aimed to reproduce the conditions of the pump-probe experiments, resulted in non-exponential rebinding kinetics, similar to the reported experimental results. The effect of the heme doming was explicitly described, making heme Fe less accessible to the external small ligand and leading to the slower rebinding times. The Fe-ON state, predicted based on the DFT calculations for the ^2A state and non-observable in the experimental studies, might be vanished by the effect of the ^4A PES, which is lower in energy for the configurations, that correspond to the Fe-ON minimum. It may prevent the Fe-ON state from being stabilized.

This work describes some structural aspects of heme protein dynamics in the processes of ligand binding. This knowledge may assist the deeper understanding of the possible mechanisms behind the heme protein functions, in particular, the ones, which involve small ligands (signaling, oxygen storage and transport, electron transfer etc). The combination of experimental and theoretical methods may lead to the advancements in the investigation of the protein structure-to-function relationships.

A

SUPPLEMENTARY INFORMATION: CO-DYNAMICS IN THE ACTIVE SITE OF CYTOCHROME C OXIDASE

The following figures summarize the simulations performed for the hexacoordinated state. The protocol used for the transfer reaction for the hexacoordinated state is similar to that for the pentacoordinated state with the only difference that for the heme unit the force field parameters are such that the heme a_3 Fe remains closer to the nitrogen plane. Hence, after quenching to the ground state PES the CO molecule can more readily bind to the Fe whereas in the pentacoordinated state heme a_3 Fe moves towards the histidine (His_{376}) on the femtosecond scale. This mainly affects the probabilities of the rebinding Fe and the structural changes of the active site via increased Fe–Cu distance.

The kinetics of the CO transfer to the Cu_B site is not significantly affected, resulting only in the minor reduction of the reaction rate constant (3.48 ± 0.44 (hexacoordinated) vs 3.83 ± 0.7 ps^{-1} (pentacoordinated)).

Figures and report rebinding probabilities and their kinetics for the validation simulations.

The parameters for the fitted 3d PES are reported in Table 5.

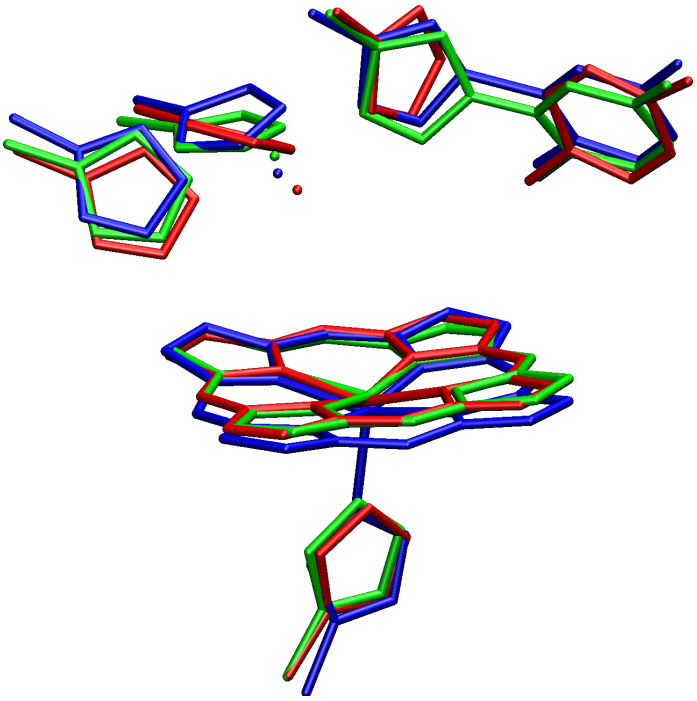


Figure 27: The X-ray structures of the active site of reduced unligated form of the CcO of *P. denitrificans* (1AR1 (red)), *Th. Thermophilus (3EH5 (blue))* and *Bos Taurus* (2EIJ (green)). The RMSD of 1AR1 and 3EH5 is 0.685 Å, the RMSD of 1AR1 and 2EIJ is 0.402 Å

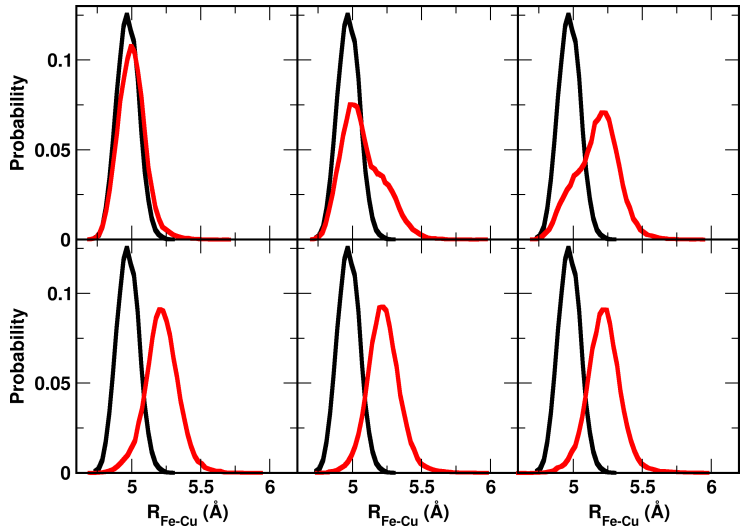
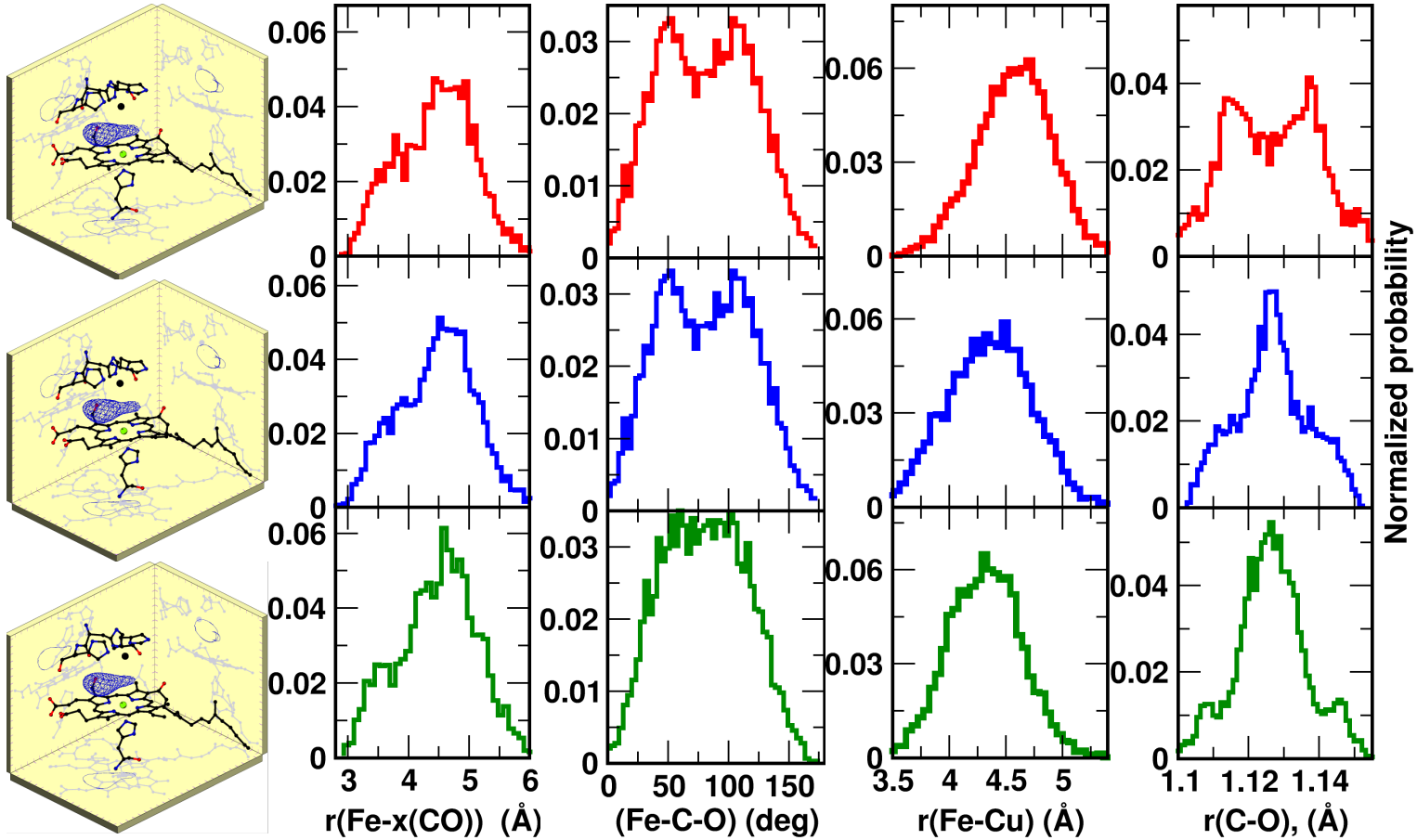


Figure 28: Probability of Fe-Cu distances for production simulations (5 ns) where CO is bound to Fe (black; μ=4.98 Å, σ=0.08 Å) and for 1000 excitation simulations showing the structural changes of the active site (red histograms) for the hexacoordinated state. The quench delay times are shown in the graphs.



(a) (b) (c) (d) (e)

Figure 29: The following illustrates the behavior of the different structural measures of the active site of CcO. Every row corresponds to the individual trajectory, the three (out of 10) of which are shown in the IR spectrum (first row – red spectrum, second row – blue spectrum, third row – green spectrum). (a) Probability distributions for the photodissociated CO in the active site of CcO. (b) Probabilities of the distances between a_3 Fe and the center of mass of CO. (c) Probabilities of the angle between heme a_3 Fe and C and O of CO. (d) Probabilities of the distances between a_3 Fe and Cu_B. (e) Probabilities of C-O distance of the free CO molecule.

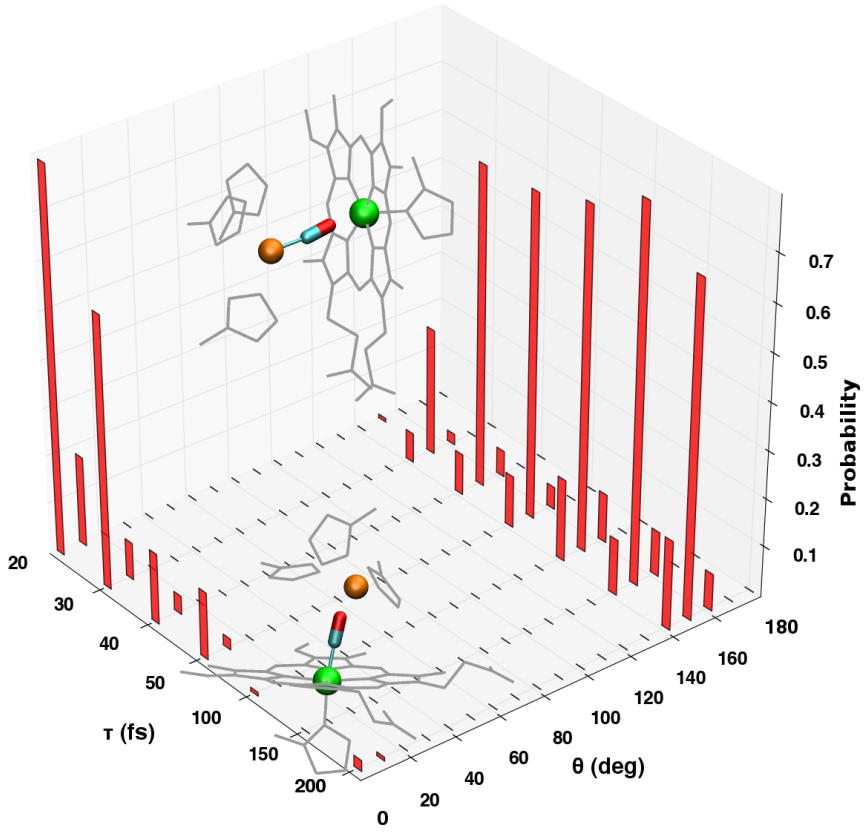


Figure 30: Probability of final states (after 5 ps) as a function of the quench delay time (based on 1000 trajectories). $\theta = 0$ corresponds to Fe–CO, $\theta > 150°$ to Cu–CO for the hexacoordinated state.

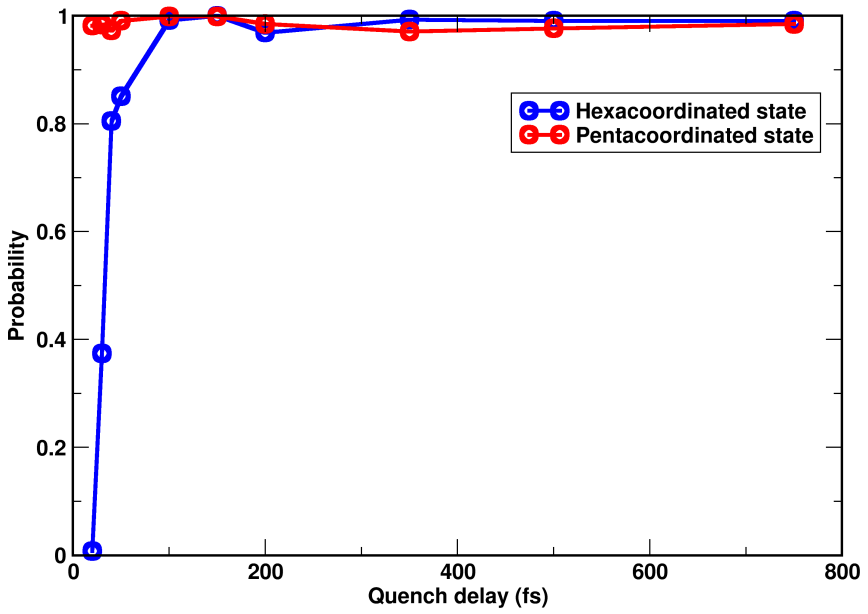


Figure 31: Probability to form Cu_B–CO depending on the quench delay time.

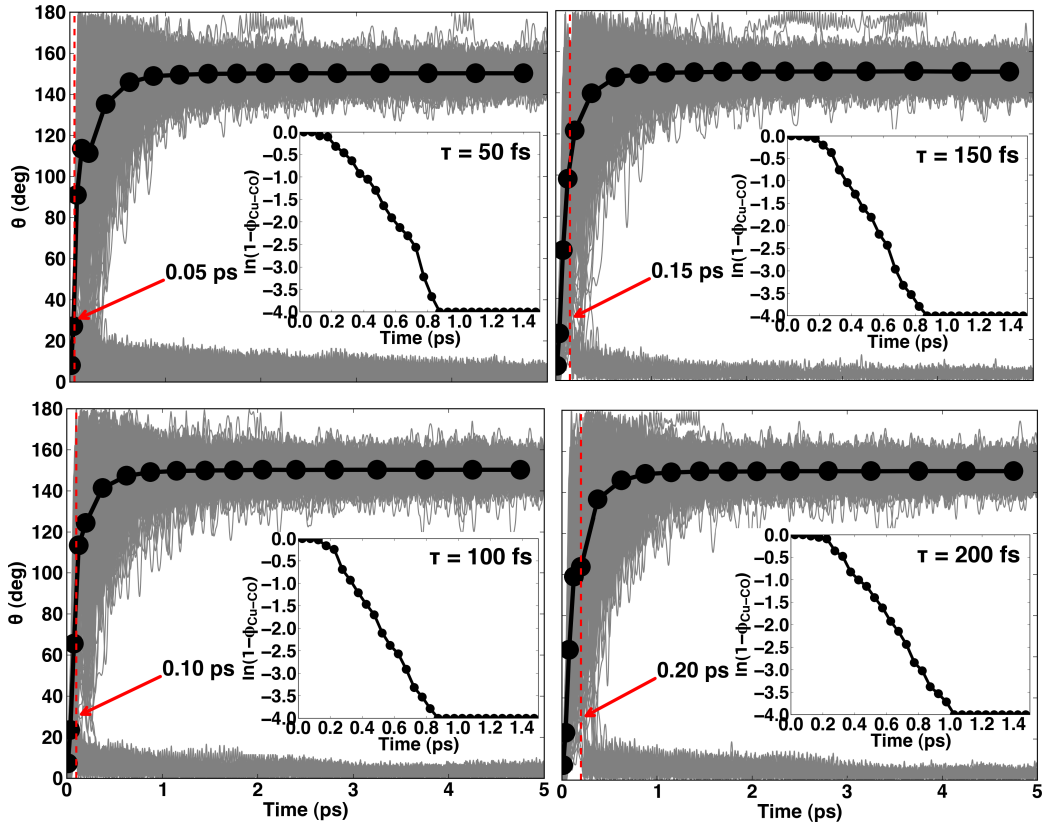
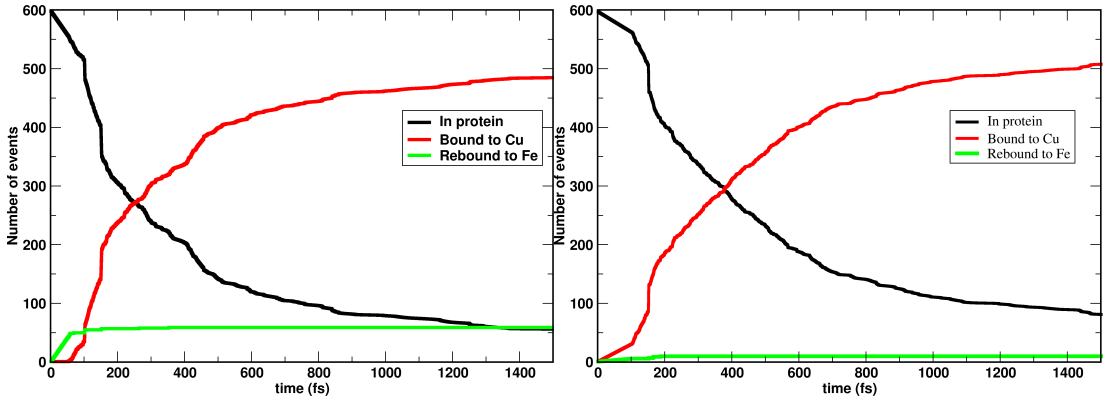
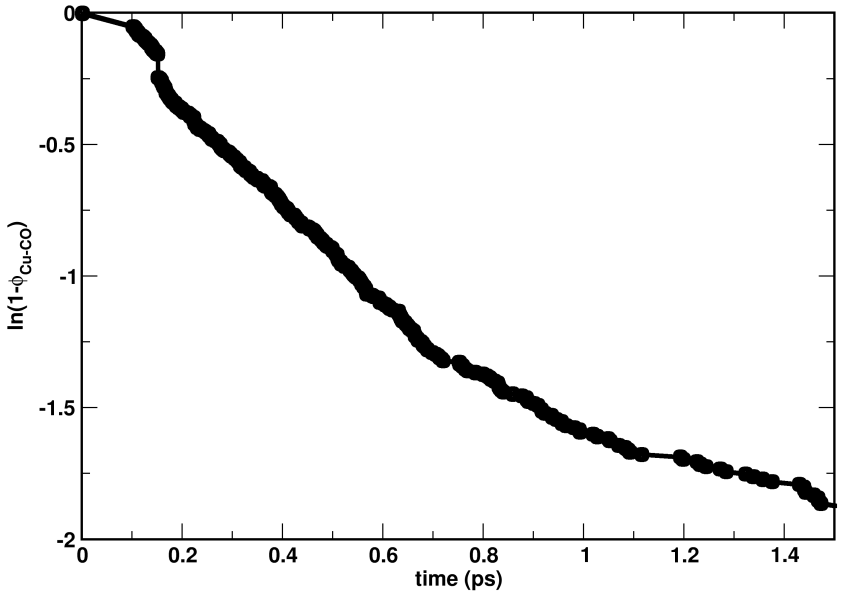


Figure 32: CO transfer kinetics for different quench time delays (shown in the plots) (for 1000 trajectories for the hexacoordinated state). In large graphs: black points are average values $< \theta >$; grey overlaping lines are the individual trajectories; the red dashed line shows the point where the quench has been performed. Insets: ϕ_{Cu-CO} is the conversion fraction to the Cu–CO state. At $t = 0$, $\phi_{Cu-CO}(t = 0) = 0$.

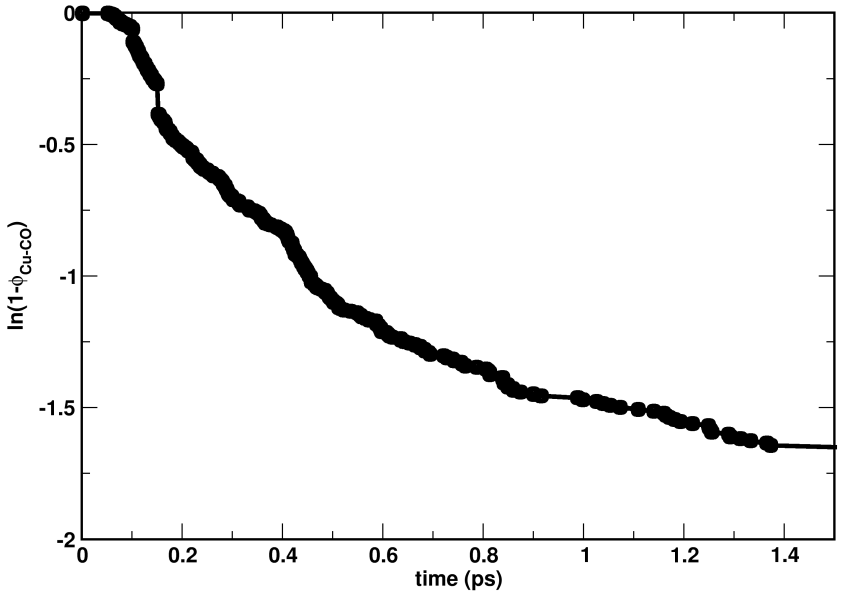


(a) (b)

Figure 33: Number of events for 600 simulations (maximum 100 ps) with the cutoff criteria (Met=Cu,Fe): r(Met-C)=2.45 Å, θ(Met-C-O)=100°; (a) hexacoordinated state, (b) pentacoordinated state



(a)



(b)

Figure 34: CO transfer kinetics for the cutoff criteria (Met=Cu,Fe): r(Met-C)=2.45 Å, θ(Met-C-O)=100°; Φ_{Cu-CO} is the conversion fraction to the Cu–CO state. At $t = 0$, $\Phi_{Cu-CO}(t = 0) = 0$. The slope of $\ln(1 - \Phi)$ is the rate of forming Cu-CO based on the cutoff criteria; (a) hexacoordinated state, (b) pentacoordinated state

Parameter	P_1	P_2	P_3	P_4	Parameter	P_1	P_2	P_3	P_4
$D_{1,\lambda=0}$	87.802	2.653	6.166	108.449	$\beta_{1,\lambda=0}$	1.281	1.036	6.188	3.622
$\rho_{e_1,\lambda=0}$	-0.086	1.027	4.675	0.395	$D_{2,\lambda=0}$	-117.819	1.139	6.281	42.641
$\beta_{2,\lambda=0}$	3.718	0.820	6.131	4.758	$\rho_{e_2,\lambda=0}$	0.302	0.804	4.310	0.637
$\alpha_{\lambda=0}$	-53.093	1.669	4.494	19.943	$D_{1,\lambda=1}$	-9.062	1.652	6.147	11.240
$\beta_{1,\lambda=1}$	14.989	3.085	6.256	14.294	$\rho_{e_1,\lambda=1}$	-0.159	0.890	4.080	0.178
$D_{2,\lambda=1}$	58.026	1.289	5.646	162.878	$\beta_{2,\lambda=1}$	0.518	1.445	5.396	0.433
$\rho_{e_2,\lambda=1}$	-0.030	8.979	4.483	0.505	$\alpha_{\lambda=1}$	-26.328	1.159	5.052	10.785
$D_{1,\lambda=2}$	-75.091	2.122	4.934	57.967	$\beta_{1,\lambda=2}$	5.655	1.715	5.438	5.396
$\rho_{e_1,\lambda=2}$	-0.237	0.986	5.029	0.319	$D_{2,\lambda=2}$	-64.183	3.600	4.168	65.460
$\beta_{2,\lambda=2}$	2.401	1.264	4.749	3.648	$\rho_{e_2,\lambda=2}$	0.212	1.323	4.972	0.588
$\alpha_{\lambda=2}$	-119.905	1.955	3.908	28.987	$D_{1,\lambda=3}$	-10.300	3.296	5.190	68.256
$\beta_{1,\lambda=3}$	14.967	0.803	6.333	10.017	$\rho_{e_1,\lambda=3}$	0.008	3.782	4.117	0.078
$D_{2,\lambda=3}$	21.338	1.686	4.383	-6.907	$\beta_{2,\lambda=3}$	0.696	4.056	6.400	9.579
$\rho_{e_2,\lambda=3}$	-0.206	3.317	5.052	0.849	$\alpha_{\lambda=3}$	2.531	3.986	5.198	-63.280
$D_{1,\lambda=4}$	-22.694	0.986	5.661	33.142	$\beta_{1,\lambda=4}$	5.408	0.837	6.374	5.583
$\rho_{e_1,\lambda=4}$	-0.302	0.884	5.274	0.290	$D_{2,\lambda=4}$	-119.143	1.934	4.481	63.718
$\beta_{2,\lambda=4}$	4.766	2.103	5.294	4.488	$\rho_{e_2,\lambda=4}$	0.220	1.947	5.454	0.644
$\alpha_{\lambda=4}$	-104.127	2.492	3.817	29.988	$D_{1,\lambda=5}$	-68.525	1.427	5.202	127.472
$\beta_{1,\lambda=5}$	1.494	0.816	5.905	1.722	$\rho_{e_1,\lambda=5}$	-0.286	0.810	6.396	0.236
$D_{2,\lambda=5}$	-57.750	2.935	4.553	54.017	$\beta_{2,\lambda=5}$	3.289	3.116	4.907	3.527
$\rho_{e_2,\lambda=5}$	-0.036	5.702	4.051	0.872	$\alpha_{\lambda=5}$	-17.832	0.820	3.937	-21.816
$D_{1,\lambda=6}$	-4.509	1.019	6.249	35.075	$\beta_{1,\lambda=6}$	-0.648	1.696	6.165	2.509
$\rho_{e_1,\lambda=6}$	0.258	3.556	4.138	0.266	$D_{2,\lambda=6}$	-66.102	2.211	4.824	45.577
$\beta_{2,\lambda=6}$	4.792	2.024	5.476	4.512	$\rho_{e_2,\lambda=6}$	0.071	4.447	5.002	0.804
$\alpha_{\lambda=6}$	-9.334	1.143	4.040	-8.903	$D_{1,\lambda=7}$	42.814	8.999	5.861	36.018
$\beta_{1,\lambda=7}$	-10.671	6.754	4.063	7.277	$\rho_{e_1,\lambda=7}$	0.185	8.607	5.032	0.391
$D_{2,\lambda=7}$	15.678	8.715	6.398	15.476	$\beta_{2,\lambda=7}$	-1.625	1.855	5.320	4.772
$\rho_{e_2,\lambda=7}$	0.144	3.305	5.380	0.817	$\alpha_{\lambda=7}$	-1.697	1.742	4.375	-4.588
$D_{1,\lambda=8}$	4.771	2.895	5.175	6.662	$\beta_{1,\lambda=8}$	4.398	0.924	5.062	13.281
$\rho_{e_1,\lambda=8}$	0.118	7.427	6.219	0.193	$D_{2,\lambda=8}$	115.937	5.052	6.164	62.741
$\beta_{2,\lambda=8}$	-14.842	7.519	4.090	7.468	$\rho_{e_2,\lambda=8}$	-0.373	3.178	3.835	0.800
$\alpha_{\lambda=8}$	-14.862	1.237	5.432	-8.442	$D_{1,\lambda=9}$	32.861	2.576	6.398	36.668
$\beta_{1,\lambda=9}$	14.895	1.045	4.244	8.363	$\rho_{e_1,\lambda=9}$	-0.071	0.997	5.614	0.109
$D_{2,\lambda=9}$	92.550	1.144	3.930	18.442	$\beta_{2,\lambda=9}$	5.582	8.956	6.256	3.290
$\rho_{e_2,\lambda=9}$	-0.355	2.117	5.286	0.633	$\alpha_{\lambda=9}$	-9.431	8.995	6.194	-25.686
$D_{1,\lambda=10}$	8.896	1.274	5.774	14.629	$\beta_{1,\lambda=10}$	11.102	1.002	4.527	10.819
$\rho_{e_1,\lambda=10}$	0.019	4.239	5.422	0.080	$D_{2,\lambda=10}$	115.247	5.712	5.005	57.695
$\beta_{2,\lambda=10}$	-14.930	3.153	4.353	7.912	$\rho_{e_2,\lambda=10}$	-0.390	5.821	4.742	0.775
$\alpha_{\lambda=10}$	-12.408	1.058	6.070	-16.642					

Table 5: Parameters of the fitted potential used as the model for the ground PES.

B

SUPPLEMENTARY INFORMATION: STRUCTURAL INTERPRETATION OF METASTABLE STATES IN MBNO

B.1 MOLECULAR DYNAMICS SIMULATIONS

All molecular dynamics (MD) simulations were carried out using CHARMM[26] with the CHARMM22 force field.[155] The protein was set up as described previously.[165, 187] Mb contains 153 amino acid residues, a heme group and a nitrogen oxide molecule. Simulations were carried out for the $His_\epsilon 64$ protonation state, which is the more likely state [108, 163, 164, 181, 220]. The protein was solvated in a periodic, preequilibrated water box $62.0864 \times 62.0864 \times 62.0864$ $Å^3$ and the final system contains 23711 atoms. All bonds involving hydrogens were treated using SHAKE[222, 279] with a tolerance of 10^{-6}. For the non-bonded interactions a cutoff of 14 Å was used. First, the system was minimized using steepest-descent and adopted basis Newton-Raphson algorithms. Then it was heated from 100 to 300 K during 60 ps and equilibrated for 500 ps. This was followed by production runs, each 200 ps in length, in the NVT ensemble. The time step in all simulations was $\Delta t = 1$ fs.

B.2 MULTISTATE PES

Up to this point the ^2A and ^4A PESs are independent. However, for reactive MD simulations they need to be related to each other such that asymptotically (for $R \to \infty$) the relative energetics between the two states reflects their true energetic ordering. This is described by a scalar constant Δ which is added to the energy of the ^4A state. In situations where dynamics on multiple, potentially crossing PESs can occur, smooth dynamics around the crossing region needs to be ensured. Here, we use multi state-ARMD (MS-ARMD) which is based on mixing PESs depending on their relative energies.[172] The method relies on mixing all PESs for a given configuration $\vec{x}$ according to exponentially distributed and energy-dependent weights w_i whereby PESs

with lower total energy have larger weights. The total PES on which the dynamics takes place is thus

$$V_{\text{effective}}(R, \phi, \theta) = \sum_{i=1}^{N} w_i \cdot V_i(R, \theta, \phi) \tag{82}$$

where the $w_i(x)$ are the normalized weights of the *i*-th PES

$$w_i = \frac{w_{0,i}}{\sum_{i=1}^{N} w_{0,i}}. \tag{83}$$

The $w_{0,i}$ are raw weights, defined as

$$w_{0,i} = \exp\left(-\frac{V_i(R, \phi, \theta) - V_{min}(R, \phi, \theta)}{\sigma}\right) \tag{84}$$

where $V_{min}(R, \phi, \theta)$ is a minimal energy PES, and $\sigma = 5.0$ kcal/mol controls the width over which the PESs are mixed. The advantage of such a procedure is that it allows to combine several PESs exhibiting different bonding patterns as is required in empirical force fields, leads to a smooth overall PES and analytical gradients can be computed. It was shown that the total energy along individual trajectories is conserved which also allows simulations in the NVE ensemble.[172] On the other hand, compared to time-based switching the implementation and the fitting of the PESs are somewhat more involved.

Because the Fe-oop motion for the ^{4}A PES (corresponding to the N_{porph}–Fe–$N_{\epsilon_{His93}}$ force field term) is only approximately parametrized in the CHARMM22 force field, it needs to be corrected when the ^{2}A and ^{4}A PESs are mixed in the reactive simulations. For this, a correction potential $V_c(\phi) = A(1 - e^{B(\phi - C)})^2 + D$ was used for ϕ−values between 90° and 95 ° to best reproduce the reference DFT energies in the crossing region. The parameters for $V_c(\phi)$ are $A = 12.0$ kcal/mol, $B = 0.37$ deg^{-1}, $C = 89.0$ deg, and $D = -19.0$ kcal/mol.

The evolution after photodissociation was probed by running 50 independent simulations. They started from the bound state (^{2}A) and the Fe-NO bond was broken. The excitation process was induced by instantaneously switching to the ^{4}A state.[165] The propagation time τ on the excited state ranged from $\tau = 100$ to 750 fs to assess the sensitivity of the final distribution of the ligands on this parameter (see Figure S3). During the excitation the NO molecule moves away from the heme-Fe. When the system is switched to the full reactive ^{2}A + ^{4}A PES (see below) the NO molecule can rebind to the heme-Fe. For comparing the effect of different evolution times

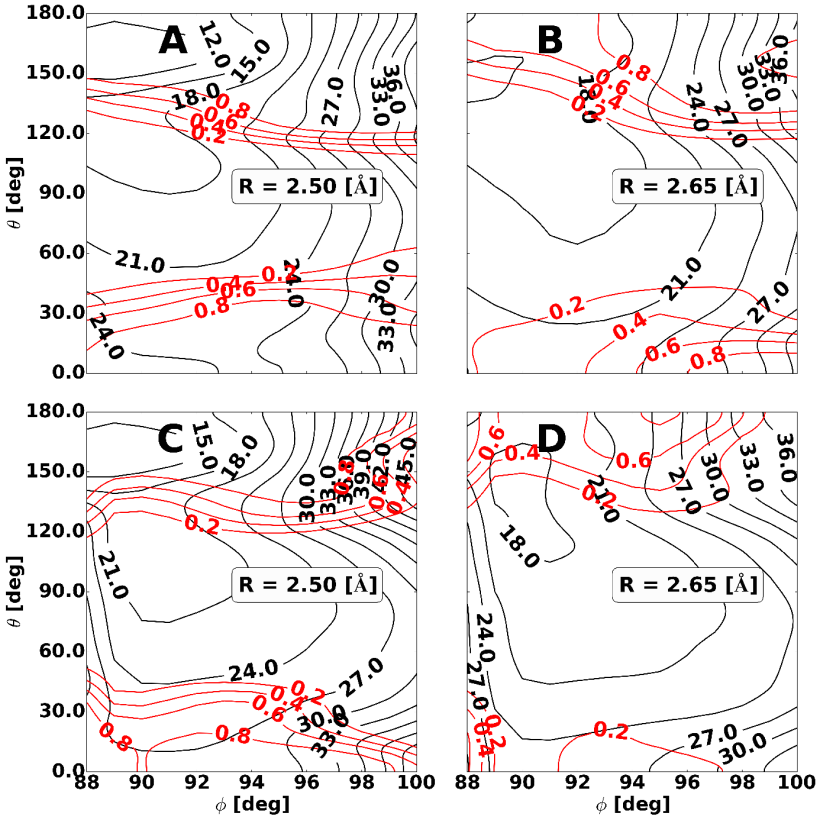


Figure 35: Effective PES for the different R, θ and ϕ. (A,B) DFT only based PES, (C,D) DFT based PES combined with the CHARMM FF. Black contours show the effective PES (in kcal/mol). Red contours show the weight of the ^{2}A PES in the effective PES (the weight ranges from 0.0 to 1.0, and a weight of 1.0 means that only the ^{2}A PES is contributes to the effective PES).

τ the distribution of Fe-CoM(NO) distances and the rebinding times to form the thermodynamically stable Fe-NO state on the ^{2}A PES were determined for each value of τ. It is found that for $\tau \geqslant 250$ fs no major changes in either of the two properties occur and for the subsequent work $\tau_{^4A} = 500$ fs was used. Such a value is also representative in view of the 300 fs derived from earlier experiments on MbCO.[4, 160] The 50 trajectories initially considered all end up in the Fe-NO bound state within 150 ps.

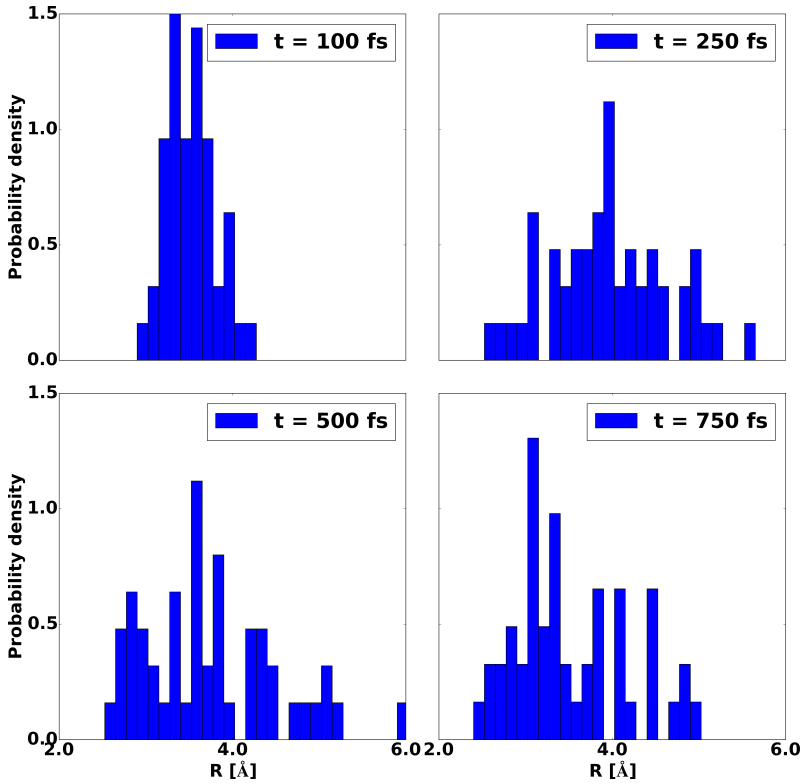


Figure 36: **A-D**) Probability densities of final states after evolving the 50 independent simulations on the excited PES for (**A** - 100 fs, **B** - 250 fs, **C** - 500 fs, **D** - 750 fs).

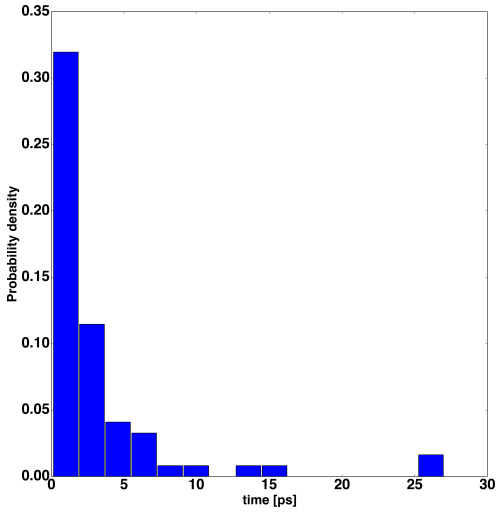


Figure 37: Lifetime of the proposed[127] Fe–oop, ligand-bound structure from simulations on the ^{2}A PES. The longest lifetime observed was 27.8 ps.

B.3 LIGAND DYNAMICS IN THE ACTIVE SITE

The maximal distance between the Fe and the ligand sampled during the rebinding simulation. The probability distribution function for the maximum distance of the

ligand to the heme-Fe atom from trajectories rebinding on very short time scales ($\tau < 2$ ps), within 10 ps and longer than 10 ps is shown in Figure 38.

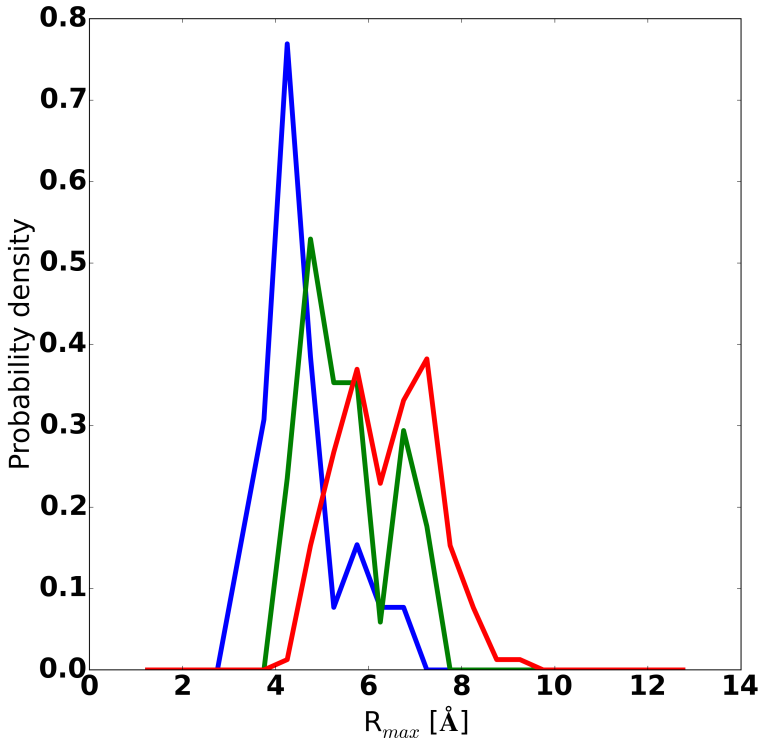


Figure 38: Distribution of the maximum heme Fe–NO distance for the rebinding trajectories with. The blue curve corresponds to trajectories that rebind within $\tau_{reb} < 2.0$ ps, the green curve to those with $2.0 < \tau_{reb} < 10.0$ ps and the red curve to $\tau_{reb} > 10.0$ ps.

B.4 REBINDING KINETICS

Depending on the asymptotic separation Δ of the two states (see SI) the time scales vary. The rapid component has an amplitude of 15-30 % compared with 50 to 60% for the slow process.

B.5 COMPUTATION OF THE XANES SPECTRA

For each of the 20 structures the transient XAS spectrum was determined using the FDMNES package.[110] In the calculations a self-consistent potential of radius 6 Å around the Fe-atom was used and the interaction with the X-ray field was described

Δ [kcal/mol]	a_1	τ_1 [ps]	a_2	τ_2 [ps]
-2 kcal/mol	0.17	5.71	0.63	339.6
0 kcal/mol	0.16	8.52	0.52	157.0
$+2$ kcal/mol	0.31	9.49	0.59	350.0

Table 6: Amplitudes a_i and rebinding times τ_i of the double exponential fits depending on the asymptotic separation Δ.

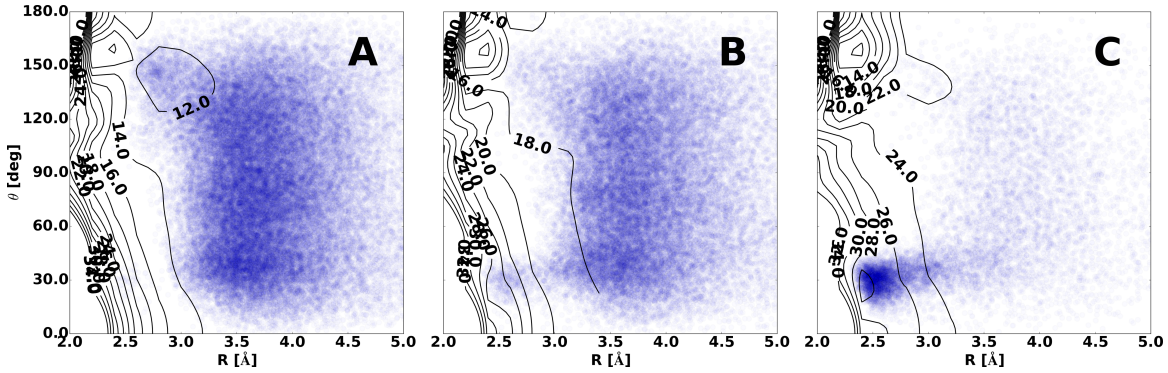


Figure 39: Isomerization trajectories from Fe–ON $\rightarrow$ Fe–NO process on the effective PES with different shifts Δ between the ^2A and ^4A PESs. (A) $\Delta = -11.1$ kcal/mol, (B) $\Delta = -6.1$ kcal/mol, (C) $\Delta = -1.1$ kcal/mol. Shifting the ^2A and ^4A closer to each other (panel C) leads to exposing the Fe–ON minimum.

using the electric quadrupole approximation. Many-body effects and the core hole lifetime broadening were included by convoluting the signal according to

$$\Gamma_f(E - E_f) = \Gamma_{hole} + \Gamma_{max}\left(\frac{1}{2} + \frac{1}{\pi}\arctan\left[\frac{\pi}{3}\frac{\Gamma_{max}}{E_l}\left(\epsilon - \frac{1}{\epsilon^2}\right)\right]\right) \tag{85}$$

with $\epsilon = (E - E_f)/E_{ctr}$, E_f is the Fermi energy, and the parameters are the total height $\Gamma_{max} = 15$ eV, the inflection point $E_{ctr} - E_f = 30$ eV, and the inclination $E_l = 25$ eV. $\Gamma_{hole} = 1.81$ eV is the core hole lifetime. This procedures follows the one used in earlier work.[236]

The influence of the presence and absence of the photodissociated NO ligand for the XANES spectrum is shown in Figure 40. It is found that for NO closer to the heme-Fe the influence is more pronounced but does not vanish for locations 5 Å away from the iron atom. Also, the sensitivity of the difference signal as a function of energy changes. The maximum influence is found around the edge.

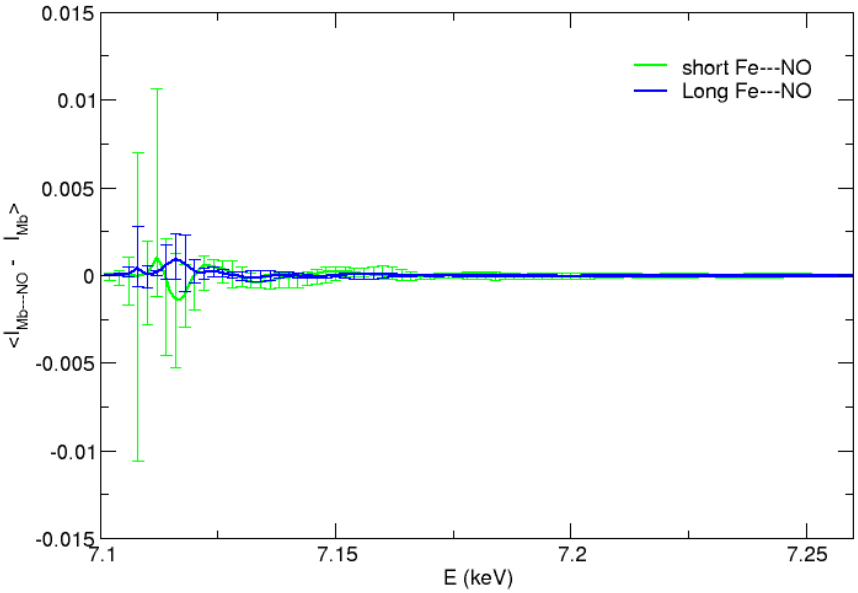


Figure 40: Influence of the presence and absence of the NO-ligand for the two sets of photodissociated snapshots (see main text): NO within 3.5 Å of the heme-Fe (20 structures) and NO-unbound (NO within 5.0 Å (10 structures). For each structure the XANES spectrum was calculated with and without the NO ligand presence and the averaged difference is reported together with the minimum and maximum difference (fluctuation bars).

BIBLIOGRAPHY

[1] Adachi, S., Nagano, S., Ishimori, K., Watanabe, Y., Morishima, I., Egawa, T., Kitagawa, T., and Makino, R. (1993). Roles of proximal ligand in heme proteins: replacement of proximal histidine of human myoglobin with cysteine and tyrosine by site-directed mutagenesis as models for p-450, chloroperoxidase, and catalase. *Biochem.*, 32(1):241–252.

[2] Aitken, J. B., Thomas, S. E., Stocker, R., Thomas, S. R., Takikawa, O., Armstrong, R. S., and Lay, P. A. (2004). Determination of the nature of the heme environment in nitrosyl indoleamine 2,3-dioxygenase using multiple-scattering analyses of x-ray absorption fine structure. *Biochem.*, 43:4892.

[3] Andersen, H. C. (1980). Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.*, 72(4):2384.

[4] Anfinrud, P. A., Han, C., and Hochstrasser, R. M. (1989). Direct observations of ligand dynamics in hemoglobin by subpicosecond infrared spectroscopy. *Proc. Natl. Acad. Sci.*, 86(21):8387.

[5] Antonini, E. and Brunori, M. (1971). *Hemoglobin and Myoglobin in their Reactions with Ligands*, volume 21. North-Holland.

[6] Aranda, R., Levin, E. J., Schotte, F., Anfinrud, P. A., and Phillips, G. N. (2006). Time-dependent atomic coordinates for the dissociation of carbon monoxide from myoglobin. *Acta Crystal. Sec. D: Biol. Crystal.*, 62(7):776–783.

[7] Aronszajn, N. (1950). Theory of reproducing kernels. *Trans. Amer. Math. Soc.*, 68:337.

[8] Babini, E., Bertini, I., Borsari, M., Capozzi, F., Luchinat, C., Zhang, X., Moura, G. L., Kurnikov, I. V., Beratan, D. N., Ponce, A., et al. (2000). Bond-mediated electron tunneling in ruthenium-modified high-potential iron-sulfur protein. *J. Am. Chem. Soc.*, 122(18):4532–4533.

[9] Banci, L. (1997). Structural properties of peroxidases. *J. Biotech.*, 53(2):253–263.

[10] Banushkina, P. and Meuwly, M. (2005). Free-energy barriers in mbco rebinding. *J. Phys. Chem. B*, 109(35):16911.

[11] Banushkina, P. and Meuwly, M. (2007). Diffusive dynamics on multidimensional rough free energy surfaces. *J. Chem. Phys.*, 127(13).

[12] Barcroft, J. and Haldane, J. S. (1902). A method of estimating the oxygen and carbonic acid in small quantities of blood. *J Physiol*, 28(3):232–240. 16992619[pmid].

[13] Barrick, D. (1994). Replacement of the proximal ligand of sperm whale myoglobin with free imidazole in the mutant his-93. fwdarw. gly. *Biochem.*, 33(21):6546–6554.

[14] Becke, A. D. (1993). Density-functional thermochemistry. iii. the role of exact exchange. *J. Chem. Phys.*, 98(7):5648–5652.

[15] Benjamin, I. (2008). Empirical valence bond model of an sn2 reaction in polar and nonpolar solvents. *J. Chem. Phys.*, 129(7):074508.

[16] Benson, D. R., Hart, B. R., Zhu, X., and Doughty, M. B. (1995). Design, synthesis, and circular dichroism investigation of a peptide-sandwiched mesoheme. *J. Am. Chem. Soc.*, 117(33):8502–8510.

[17] Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81(8):3684.

[18] Berens, P. H. and Wilson, K. R. (1981). Molecular dynamics and spectra. i. diatomic rotation and vibration. *J. Chem. Phys.*, 74(9):4872–4882.

[19] Berman, H., Henrick, K., and Nakamura, H. (2003). Announcing the worldwide protein data bank. *Nature Struc. Mol. Biol.*, 10(12):980–980.

[20] Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000). The protein data bank. *Nucl. Acids Res.*, 28(1):235–242.

[21] Bertini, I., Luchinat, C., and Parigi, G. (2000). Hyperfine shifts in low-spin iron (iii) hemes: A ligand field analysis. *Eur. J. Inorg. Chem.*, 2000(12):2473–2480.

[22] Bertini, I., Luchinat, C., Parigi, G., and Walker, F. A. (1999). Heme methyl 1h chemical shifts as structural parameters in some low-spin ferriheme proteins. *JBIC J. Biol. Inorg. Chem.*, 4(4):515–519.

[23] Bossa, C., Anselmi, M., Roccatano, D., Amadei, A., Vallone, B., Brunori, M., and Di Nola, A. (2004). Extended molecular dynamics simulation of the carbon monoxide migration in sperm whale myoglobin. *Biophysical journal*, 86(6):3855.

[24] Bredenbeck, J., Helbing, J., Nienhaus, K., Nienhaus, G. U., and Hamm, P. (2007). Protein ligand migration mapped by nonequilibrium 2D-IR exchange spectroscopy. *Proc. Natl. Acad. Sci.*, 104(36):14243–14248.

[25] Bren, K. L. and Gray, H. B. (1993). Structurally engineered cytochromes with novel ligand-binding sites: oxy and carbon monoxy derivatives of semisynthetic horse heart ala80 cytochrome c. *J. Am. Chem. Soc.*, 115(22):10382–10383.

[26] Brooks, B. R., Brooks III, C. L., Mackerell, A. D. J., Nilsson, L., and et al (2009). Charmm: the biomolecular simulation program. *J. Comp. Chem.*, 30(10):1545.

[27] Brucker, E. A., Olson, J. S., Ikeda-Saito, M., and Phillips, G. N. (1998). Nitric oxide myoglobin: Crystal structure and analysis of ligand geometry. *Prot. Struct. Funct. Genet.*, 30:352.

[28] Brunori, M., Giuffrè, A., Forte, E., Mastronicola, D., Barone, M. C., and Sarti, P. (2004). Control of cytochrome c oxidase activity by nitric oxide. *Bioch. Bioph. Acta*, 1655:365–371.

[29] Buhrow, L., Ferguson-Miller, S., and Kuhn, L. (2012). From static structure to living protein: Computational analysis of cytochrome c oxidase main-chain flexibility. *Biophys. J.*, 102(9):2158.

[30] Carducci, M., Pressprich, M., and Coppens, P. (1997). Diffraction studies of photoexcited crystals: Metastable nitrosyl-linkage isomers of sodium nitroprusside. *J. Am. Chem. Soc.*, 119(11):2669–2678.

[31] Castro-Palacio, J., Bemish, R., and Meuwly, M. (2015). Equilibrium rate coefficients from atomistic simulations: The $o(^3p) + no(^2\pi) \rightarrow o_2(\sigma_g^-) + n(^4s)$ reaction at temperatures relevant to the hypersonic flight regime. *J. Chem. Phys.*, 142:091104.

[32] Castro-Palacio, J., Nagy, T., Bemish, R., and Meuwly, M. (2014). Computational study of collisions between $o(^3p)$ and $no(^2\pi)$ at temperatures relevant to the hypersonic flight regime. *J. Chem. Phys.*, 141:164319.

[33] Chan, M. K. (2001). Recent advances in heme-protein sensors. *Curr. op. chem. biol.*, 5(2):216–222.

[34] Cheng, A. and Merz, K. M. (1996). Application of the nosé-hoover chain algorithm to the study of protein dynamics. *J. Phys. Chem.*, 100(5):1927–1937.

[35] Cheng, L., Novozhilova, I., Kim, C., Kovalevsky, A., Bagley, K. A., Coppens, P., and Richter-Addo, G. B. (2000). First observation of photoinduced nitrosyl linkage isomers of iron nitrosyl porphyrins. *J. Am. Chem. Soc.*, 122(29):7142–7143.

[36] Chenoweth, K., Cheung, S., van Duin, A. C. T., Goddard, W. A., and Kober, E. M. (2005). Simulations on the thermal decomposition of a poly(dimethylsiloxane) polymer using the reaxff reactive force field. *J. Am. Chem. Soc.*, 127(19):7192. PMID: 15884961.

[37] Chenoweth, K., van Duin, A. C. T., Dasgupta, S., and Goddard III, W. A. (2009). Initiation mechanisms and kinetics of pyrolysis and combustion of jp-10 hydrocarbon jet fuel. *J. Phys. Chem. A*, 113(9):1740. PMID: 19209880.

[38] Chenoweth, K., van Duin, A. C. T., and Goddard, W. A. (2008). Reaxff reactive force field for molecular dynamics simulations of hydrocarbon oxidation. *J. Phys. Chem. A*, 112(5):1040. PMID: 18197648.

[39] Cherry, J. R., Lamsa, M. H., Schneider, P., Vind, J., Svendsen, A., Jones, A., and Pedersen, A. H. (1999). Directed evolution of a fungal peroxidase. *Nature biotech.*, 17(4):379–384.

[40] Chu, K., Vojtchovskỳ, J., McMahon, B. H., Sweet, R. M., Berendzen, J., and Schlichting, I. (2000). Structure of a ligand-binding intermediate in wild-type carbonmonoxy myoglobin. *Nature*, 403(6772):921–923.

[41] Clark, B. J. and Stocco, D. M. (1995). Expression of the steroidogenic acute regulatory (star) protein: a novel lh-induced mitochondrial protein required for the acute regulation of steroidogenesis in mouse leydig tumor cells. *Endocr. res.*, 21(1-2):243–257.

[42] Cleeter, M., Cooper, J., Darley-Usmar, V., Moncada, S., , and Schapira, A. (1994). Reversible inhibition of cytochrome c oxidase, the terminal enzyme of the mitochondrial respiratory chain, by nitric oxide: implications for neurodegenerative diseases. *FEBS lett.*, 345(1):50–54.

[43] Collman, J. P. (1977). Synthetic models for the oxygen-binding hemoproteins. *Acc.Chem. Res.*, 10(7):265–272.

[44] Collman, J. P., Fudickar, W., and Shiryaeva, I. (2003). Water-soluble polymer-bound biomimetic analogues of cytochrome c oxidase catalyze 4e-reduction of o2 to water. *Inorg. Chem.*, 42(11):3384–3386.

[45] Collman, J. P., Rapta, M., Bröring, M., Raptova, L., Schwenninger, R., Boitrel, B., Fu, L., and L'Her, M. (1999). Close structural analogues of the cytochrome c oxidase fea3/cub center show clean 4e-electroreduction of o2 to h2o at physiological ph. *J. Am. Chem. Soc.*, 121(6):1387–1388.

[46] Collman, J. P., Zhang, X., Herrmann, P. C., Uffelman, E. S., Boitrel, B., Straumanis, A., and Brauman, J. I. (1994). Congruent multiple michael addition for the synthesis of biomimetic heme analogs. *J. Am. Chem. Soc.*, 116(6):2681–2682.

[47] Cooper, C. E. (2002). Nitric oxide and cytochrome oxidase: substrate, inhibitor or effector? *Trends biochem. sc.*, 27(1):33–39.

[48] Copeland, D. M., Soares, A. S., West, A. H., and Richter-Addo, G. B. (2006). Crystal structures of the nitrite and nitric oxide complexes of horse heart myoglobin. *J. of Inorg. Biochem.*, 100:1413.

[49] Cornelius, P., Hochstrasser, R., and Steele, A. (1983). Ultrafast relaxatio nin picosecond photolysis of nitrosylhemoglobin. *J. Mol. Biol.*, 163(1):119–128.

[50] Cristianini, N. and Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines: And Other Kernel-based Learning Methods*. Cambridge University Press, New York, NY, USA.

[51] Danielsson, J. and Meuwly, M. (2007). Energetics and dynamics in mbcn: Cn- vibrational relaxation from molecular dynamics simulations. *J. Phys. Chem. B*, 111:218.

[52] Danielsson, J. and Meuwly, M. (2008). Atomistic simulation of adiabatic reactive processes based on multi-state potential energy surfaces. *J. Chem. Theo. Comp.*, 4(7):1083.

[53] Dayal, P., Weyand, S. A., McNeish, J., and Mosey, N. J. (2011). Temporal quantum mechanics/molecular mechanics: Extending the time scales accessible in molecular dynamics simulations of reactions. *Chem. Phys. Lett.*, 516(4-6):263.

[54] Devereux, M. and Meuwly, M. (2009). Structural assignment of spectra by characterization of conformational substates in bound mbco. *Biophys. J.*, 96(11):4363.

[55] Devereux, M. and Meuwly, M. (2010). Force field optimization using dynamics and ensemble averaged data: Vibrational spectra and relaxation in bound mbco. *J. Chem. Inf. Model.*, 50:349.

[56] Dostrovsky, I., Hughes, E., and Ingold, C. (1946). Mechanism of substitution as a saturated carbon atom. xxxii. the role of hindrance. g. magnitude of steric effects, range of occurrence of steric and polar effects, and place of wagner rearrangement in nucleophilic substitution and elimination. *J. Chem. Soc.*, 1:173–194.

[57] Dunford, H. B. (2000). Peroxidase-catalyzed halide ion oxidation. *Redox Rep.*, 5(4):169–171.

[58] Eckart, C. and Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, 1:211.

[59] Edsall, J. T. (1972). Blood and hemoglobin: The evolution of knowledge of functional adaptation in a biochemical system. part i: The adaptation of chemical structure to function in hemoglobin. *Journal of the History of Biology*, 5(2):205–257.

[60] Elber, R. and Karplus, M. (1990). Enhanced sampling in molecular dynamics: Use of the time-dependent hart ree approximation for a simulation of carbon monoxide diffusion through Myoglobin. *J. Am. Chem. Soc.*, 112:9161–9175.

[61] Farinas, E. T., Schwaneberg, U., Glieder, A., and Arnold, F. H. (2001). Directed evolution of a cytochrome p450 monooxygenase for alkane oxidation. *Adv. Syn. Cat.*, 343(6-7):601–606.

[62] Fermi, G. and Perutz, M. F. (1981). *Haemoglobin and myoglobin*, volume 2. Clarendon press.

[63] Fiammengo, R., Wojciechowski, K., Crego-Calama, M., Timmerman, P., Figoli, A., Wessling, M., and Reinhoudt, D. N. (2003). Heme-protein active site models via self-assembly in water. *Org. lett.*, 5(19):3367–3370.

[64] Fischer, H. and Zeile, K. (1929). Synthese des hämatoporphyrins, protoporphyrins und hämins. *Justus Liebigs Annalen der Chemie*, 468(1):98–116.

[65] Fischer, S., Olsen, K. W., Nam, K., and Karplus, M. (2011). Unsuspected pathway of the allosteric transition in hemoglobin. *Proc. Natl. Acad. Sci.*, 108(14):5608–5613.

[66] Foloppe, N. and MacKerell, Jr., A. D. (2000). All-atom empirical force field for nucleic acids: I. parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comp. Chem.*, 21(2):86–104.

[67] Franzen, S. (2002). Spin-dependent mechanism for diatomic ligand binding to heme. *Proc. Natl. Acad. Sci.*, 99(26):16754–16759.

[68] Frauenfelder, H., McMahon, B. H., Austin, R. H., Chu, K., and Groves, J. T. (2001). From static structure to living protein: Computational analysis of cytochrome c oxidase main-chain flexibility. *Proc. Natl. Acad. Sci.*, 98(5):2370.

[Frisch et al.] Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, Jr., J. A., Peralta, J. E., Ogliaro, F., Bearpark, M., Heyd, J. J., Brothers, E., Kudin, K. N., Staroverov, V. N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, J. M., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, Ö., Foresman, J. B., Ortiz, J. V., Cioslowski, J., and Fox, D. J. Gaussian 09 Revision A.1. Gaussian Inc. Wallingford CT 2009.

[70] Garry, D. J., Ordway, G. A., Lorenz, J. N., Radford, N. B., Chin, E. R., Grange, R. W., Bassel-Duby, R., and Williams, R. S. (1998). Mice without myoglobin. *Nature*, 395(6705):905–908.

[71] Gibson, Q., Regan, R., Elber, R., Olson, J., and Carver, T. (1992). Distal pocket residues affect picosecond ligand recombination in myoglobin - An experimental and molecular-dynamics study of position 29 mutants. *J. Biol. Chem.*, 267(31):22022–22034.

[72] Gilles-Gonzalez, M. A., Gonzalez, G., Perutz, M. F., Kiger, L., Marden, M. C., and Poyart, C. (1994). Heme-based sensors, exemplified by the kinase fixl, are a new class of heme protein with distinctive ligand binding and autoxidation. *Biochem.*, 33(26):8067–8073. PMID: 8025112.

[73] Ginhoven, R. V. and Hjalmarson, H. (2007). Atomistic simulation of si/sio2 interfaces. *Nucl. Instr. Methods Phys. Res, Sec. B*, 255(1):183.

[74] Giorgio, M., Migliaccio, E., Orsini, F., Paolucci, D., Moroni, M., Contursi, C., Pelliccia, G., Luzi, L., Minucci, S., Marcaccio, M., et al. (2005). Electron transfer between cytochrome c and p66 shc generates reactive oxygen species that trigger mitochondrial apoptosis. *Cell*, 122(2):221–233.

[75] Giuffre, A., Forte, E., Brunori, M., and Sarti, P. (2005). Nitric oxide, cytochrome c oxidase and myoglobin: Competition and reaction pathways. *FEBS Lett.*, 579:2528.

[76] Golub, G. H. and Kahan, W. (1965). Calculating the singular values and pseudo-inverse of a matrix. *J. Soc. Ind. Appl. Math.: Ser. B, Num. Analysis*, 2:205.

[77] Gray, H. B. and Winkler, J. R. (1996). Electron transfer in proteins. *Ann. rev. biochem.*, 65:537–561.

[78] Green, D. R. and Reed, J. C. (1998). Mitochondria and apoptosis. *Science*, 281(5381):1309–1312.

[79] Haldane, J. (1898). A contribution to the chemistry of hæmoglobin and its immediate derivatives. *J. Physiol.*, 22(4):298–306. 16992410[pmid].

[80] Halgren, T. A. (1996). Merck molecular force field. i. basis, form, scope, parameterization, and performance of mmff94. *J. Comp. Chem.*, 17(5-6):490.

[81] Hangelbroek, T. and Ron, A. (2010). Nonlinear approximation using gaussian kernels. *J. Func. Analysis*, 259:203.

[82] Harrison, P. and Huehns, E. (1979). Proteins of iron metabolism. *Nature*, 279:476–477·

[83] Heitbrink, D., Sigurdson, H., Bolwien, C., Brzezinski, P., and Heberle, J. (2002). Transient binding of CO to cu(b) in cytochrome c oxidase is dynamically linked to structural changes around a carboxyl group: a time-resolved step-scan fourier transform infrared investigation. *Biophys. J.*, 82(1 Pt 1):1.

[84] Helfand, E. (1978). Brownian dynamics study of transitions in a polymer chain of bistable oscillators. *J. Chem. Phys.*, 69(3):1010–1018.

[85] Henry, E. R., Levitt, M., and Eaton, W. A. (1985). Molecular dynamics simulation of photodissociation of carbon monoxide from hemoglobin. *Proc. Natl. Acad. Sci.*, 82(7):2034.

[86] Henry, E. R., Sommer, J. H., Hofrichter, J., Eaton, W. A., and Gellert, M. (1983). Geminate recombination of carbon monoxide to myoglobin. *J. Mol. Biol.*, 166(3):443–451.

[87] Hestenes, M. R. (1958). Inversion of matrices by biorthogonalization and related results. *J. Soc. Ind. Appl. Math.*, 6:52.

[88] Hill, T. L. (1946). On steric effects. *The Journal of Chemical Physics*, 14(7):465.

[89] Ho, T. and Rabitz, H. (2003). Reproducing kernel Hilbert space interpolation methods as a paradigm of high dimensional model representations: Application to multidimensional potential energy surface construction. *J. Chem. Phys.*, 119(13):6433–6442.

[90] Ho, T. S. and Rabitz, H. (1996). A general method for constructing multidimensional molecular potential energy surfaces from ab initio calculations. *J. Chem. Phys.*, 104:2584.

[91] Hockney, R. W. (1970). The potential calculation and some applications. *Meth. Comp. Phys.*, 9:136–211.

[92] Hofacker, I. and Schulten, K. (1998). Oxygen and proton pathways in cytochrome c oxidase. *Proteins*, 30(1):100.

[93] Hollebeek, T., Ho, T. S., and Rabitz, H. (1999). Constructing multidimensional molecular potential energy surfaces from ab initio data. *Ann. Rev. Phys. Chem.*, 50:537.

[94] Hong, G., Rosta, E., and Warshel, A. (2006). Using the constrained DFT approach in generating diabatic surf aces and off diagonal empirical valence bond terms for modeling reactions in con densed phases. *J. Phys. Chem. B*, 110:19570.

[95] Hoover, W. G. (1985). Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, 31:1695.

[96] Howley, T. and Madden, M. (2006). *An Evolutionary Approach to Automatic Kernel Construction*, volume 4132 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg.

[97] Huang, J., Buchowiecki, M., Nagy, T., Vanicek, J., and Meuwly, M. (2014). Kinetic isotope effect in malonaldehyde from path integral Monte Carlo simulations. *PCCP*, 16:204–211.

[98] Huang, J., Haeussinger, D., Gellrich, U., Seiche, W., Breit, B., and Meuwly, M. (2012). Hydrogen-Bond and Solvent Dynamics in Transition Metal Complexes: A Combined Simulation and NMR-Investigation. *J. Phys. Chem. B*, 116(49):14406–14415.

[99] Huang, J. and Meuwly, M. (2010). Explicit hydrogen-bond potentials and their application to nmr scalar couplings in proteins. *JCTC*, 6(2):467.

[100] Huffaker, J. N. (1976a). Diatomic molecules as perturbed morse oscillators. i. extension to higher order parameters. *J. Chem. Phys.*, 64:3175.

[101] Huffaker, J. N. (1976b). Diatomic molecules as perturbed morse oscillators. ii. extension to higher order parameters. *J. Chem. Phys.*, 64:4564.

[102] Hunt, N. T., Greetham, G. M., Towrie, M., Parker, A. W., and Tucker, N. P. (2011). Relationship between protein structural fluctuations and rebinding dynamics in ferric haem nitrosyls. *Biochem. J.*, 433(3):459–468.

[103] Ibers, J. A. and Holm, R. H. (1980). Modeling coordination sites in metallobiomolecules. *Science*, 209:223–235.

[104] Ikeda-Saito, M., Dou, Y., Yonetani, T., Olson, J., Li, T., Regan, R., and Gibson, Q. (1993). Ligand diffusion in the distal heme pocket of myoglobin - A primary determinant of geminate rebinding. *J. Biol. Chem.*, 268(10):6855–6857.

[105] Immoos, C. E., Sulc, F., Farmer, P. J., Czarnecki, K., Bocian, D. F., Levina, A., Aitken, J. B., Armstrong, R. S., and Lay, P. A. (2005). Bonding in hno-myoglobin as characterized by x-ray absorption and resonance raman spectroscopies. *J. Am. Chem. Soc.*, 127:814.

[106] Ionascu, D., Gruia, F., Ye, X., Yu, A., Rosca, F., Demidov, C. B. A., Olson, J. S., and Champion, P. M. (2005). Temperature-dependent studies of no recombination to heme and heme proteins. *J. Am. Chem. Soc.*, 127:16921–16934.

[107] Jackson, D. A., Symons, R. H., and Berg, P. (1972). Biochemical method for inserting new genetic information into dna of simian virus 40: circular sv40 dna

molecules containing lambda phage genes and the galactose operon of escherichia coli. *Proc. Natl. Acad. Sci.*, 69(10):2904–2909.

[108] Johnson, J. B., Lamb, D. C., Frauenfelder, H., Müller, J., McMahon, B., Nienhaus, G. U., and Young, R. D. (1996). Ligand binding to heme proteins. vi. interconversion of taxonomic substates in carbonmonoxymyoglobin. *Biophys. J.*, 71(3):1563.

[Johnson] Johnson, S. G. The nlopt nonlinear-optimization package.

[110] Joly, Y. (2001). X-ray absorption near-edge structure calculations beyond the muffin-tin approximation. *Phys. Rev. B*, 63(12):125120.

[111] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79:926.

[112] Jorgensen, W. L., Maxwell, D. S., and Tirado-Rives, J. (1996). Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.*, 118(45):11225.

[113] Kachalova, G. S., Popov, A. N., and Bartunik, H. D. (1999). A steric mechanism for inhibition of CO binding to heme proteins. *Science*, 284(5413):473–476.

[114] Kadish, K. and Bottomley, L. (1980). Influence of substituted pyridines on the redox reactions of iron porphyrins. *Inorg. Chem.*, 19(4):832–836.

[115] Kaila, V. R., Johansson, M. P., Sundholm, D., Laakkonen, L., and Wikström, M. (2009). The chemistry of the cub site in cytochrome c oxidase and the importance of its unique his-tyr bond. *Bioch. Bioph. Acta*, 1787(4):221.

[116] Karlin, K. D. (1993). Metalloenzymes, structural motifs, and inorganic models. *Science*, 261(5122):701–708.

[117] Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., and Philips, D. C. (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, 181(4610):662–666.

[118] Kepp, K. P. (2015). Towards a "Golden Standard" for computing globin stability: Stability and structure sensitivity of myoglobin mutants. *Bioch. Bioph. Acta*, 1854(10, A):1239–1248.

[119] Kettering, C. F., Shutts, L. W., and Andrews, D. H. (1930). A representation of the dynamic properties of molecules by mechanical models. *Phys. Rev.*, 36:531.

[120] Kholodenko, Y., Gooding, E., Dou, Y., Ikeda-Saito, M., and Hochstrasser, R. (1999). Heme protein dynamics revealed by geminate nitric oxide recombination in mutants of iron and cobalt myoglobin. *Biochem.*, 38(18):5918–5924.

[121] Kim, E., Chufán, E. E., Kamaraj, K., and Karlin, K. D. (2004a). Synthetic models for heme-copper oxidases. *Chem. rev.*, 104(2):1077–1134.

[122] Kim, J., Park, J., Lee, T., and Lim, M. (2012). Dynamics of Geminate Rebinding of NO with Cytochrome c in Aqueous Solution Using Femtosecond Vibrational Spectroscopy. *J. Phys. Chem. B*, 116(46):13663–13671.

[123] Kim, S., Jin, G., and Lim, M. (2004b). Dynamics of geminate recombination of no with myoglobin in aqueous solution probed by femtosecond mid-ir spectroscopy. *The Journal of Physical Chemistry B*, 108(52):20366.

[124] Kim, S. and Lim, M. (2005). Protein conformation-induced modulation of ligand binding kinetics: A femtosecond mid-ir study of nitric oxide binding trajectories in myoglobin. *J. Am. Chem. Soc.*, 127:8908–8909.

[125] Klauda, J. B., Venable, R. M., Freites, J. A., OĆonnor, J. W., Tobias, D. J., Mondragon-Ramirez, C., Vorobyov, I., MacKerell, A. D., and Pastor, R. W. (2010). Update of the charmm all-atom additive force field for lipids: Validation on six lipid types. *J. Phys. Chem. B*, 114(23):7830–7843. PMID: 20496934.

[126] Klug, D. D., Zgierski, M. Z., John, S. T., Liu, Z., Kincaid, J. R., Czarnecki, K., and Hemley, R. J. (2002). Doming modes and dynamics of model heme compounds. *Proc. Natl. Acad. Sci.*, 99(20):12526–12530.

[127] Kruglik, S. G., Yoo, B.-K., Franzen, S., Vos, M. H., Martin, J.-L., and Negrerie, M. (2010). Picosecond primary structural transition of the heme is retarded after nitric oxide binding to heme proteins. *Proc. Natl. Acad. Sci.*, 107:13678.

[128] Krzywda, S., Murshudov, G. N., Brzozowski, A. M., Jaskolski, M., Scott, E. E., Klizas, S. A., Gibson, Q. H., Olson, J. S., and Wilkinson, A. J. (1998). Stabilizing bound o2 in myoglobin by valine68 (e11) to asparagine substitution. *Biochem.*, 37(45):15896–15907.

[129] Kuczera, K., Lambry, J., Martin, J., and Karplus, M. (1993). Nonexponential relaxation after ligand dissociation from myoglobin: a molecular dynamics simulation. *Proc. Natl. Acad. Sci.*, 90(12):5805–5807.

[130] Lammers, S., Lutz, S., and Meuwly, M. (2007). Reactive force fields for proton transfer dynamics. *J. Comp. Chem.*, 29(April):1048.

[131] Lammers, S. and Meuwly, M. (2007). On the relationship between infrared spectra of shared protons in diffe rent chemical environments: A comparison of protonated diglyme and protonated wa ter dimer. *J. Phys. Chem. A*, 111:1638.

[132] Laskowski, R. A. (1995). Surfnet: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. *J. Mol. Graph.*, 13(5):323.

[133] Law, M. M. and Hutson, J. M. (1997). I-nolls: a program for interactive nonlinear least-squares fitting of the parameters of physical models. *Comp. phys. comm.*, 102:252.

[134] Lawson Daku, L. M. and Hauser, A. (2010). Ab initio molecular dynamics study of an aqueous solution of [fe(bpy)$_3$]cl$_2$ in the low-spin and in the high-spin states. *J. Phys. Chem. Lett.*, 1(12):1830–1835.

[135] Lawyer, F. C., Stoffel, S., Saiki, R. K., Myambo, K., Drummond, R., and Gelfand, D. H. (1989). Isolation, characterization, and expression in escherichia coli of the dna polymerase gene from thermus aquaticus. *J. Biol. Chem.*, 264(11):6427–6437.

[136] Lee, C., Yang, W., and Parr, R. G. (1988). Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Phys. rev. B*, 37(2):785.

[137] Lee, M. W., Carr, J. K., Göllner, M., Hamm, P., and Meuwly, M. (2013). 2d ir spectra of cyanide in water investigated by molecular dynamics simulations. *J. Chem. Phys.*, 139:054506.

[138] Lee, M. W. and Meuwly, M. (2012). Molecular dynamics simulation of nitric oxide in myoglobin. *J. Phys. Chem. B*, 116(14):4154.

[139] Lemon, D., Calhoun, M., Gennis, R., and Woodruff, W. (1993). The gateway to the active site of heme-copper oxidases. *Biochemistry*, 32(45):11953.

[140] Levesque, D. and Verlet, L. (1993). Molecular dynamics and time reversibility. *Journal of Statistical Physics*, 72(3-4):519–537.

[141] Li, H., Elber, R., and Straub, J. E. (1993). Molecular-dynamics simulation of no recombination to myoglobin mutants. *J. Biol. Chem.*, 268:17908–17916.

[142] Liebl, U., Lipowski, G., Négrerie, M., Lambry, J. C., Martin, J. L., and Vos, M. H. (1999). Coherent reaction dynamics in a bacterial cytochrome c oxidase. *Nature*, 401(6749):181.

[143] Lim, M., Jackson, T., and Anfinrud, P. (1993). Nonexponential relaxation after ligand dissociation from myoglobin: a molecular dynamics simulation. *Proc. Natl. Acad. Sci.*, 90(12):5801–5804.

[144] Lim, M., Jackson, T. A., and Anfinrud, P. A. (1995). Mid-infrared vibrational spectrum of CO after photodissociation from heme: Evidence for a ligand docking site in the heme pocket of hemoglobin and myoglobin. *J. Chem. Phys.*, 102(11):4355.

[145] Lim, M., Jackson, T. A., and Anfinrud, P. A. (1997). Ultrafast rotation and trapping of carbon monoxide dissociated from myoglobin. *Nature Struc. Mol. Biol.*, 4(3):209–214.

[146] Liong, E. C. (1999). *Structural and functional analysis of proximal pocket mutants of sperm whale myoglobin.* PhD thesis, Rice University.

[147] Liu, B., Zhang, Y., Sage, J. T., Soltis, S. M., Doukov, T., Chen, Y., Stout, C. D., and Fee, J. A. (2012). Structural changes that occur upon photolysis of the fe (ii) a3–co complex in the cytochrome ba 3-oxidase of thermus thermophilus: A combined x-ray crystallographic and infrared spectral study demonstrates CO binding to cu b. *Bioch. Bioph. Acta*, 1817(4):658–665.

[148] Lombardi, A., Nastri, F., and Pavone, V. (2001). Peptide-based heme-protein models. *Chem. rev.*, 101(10):3165–3190.

[149] Luna, V. M., Fee, J. A., Deniz, A. A., and Stout, C. D. (2012). Mobility of xe atoms within the oxygen diffusion channel of cytochrome ba3 oxidase. *Biochem.*, 51(23):4669.

[150] Lundberg, J. O., Gladwin, M. T., Ahluwalia, A., Benjamin, N., Bryan, N. S., Butler, A., Cabrales, P., Fago, A., Feelisch, M., Ford, P. C., Freeman, B. A., Frenneaux, M., Friedman, J., Kelm, M., Kevil, C. G., Kim-Shapiro, D. B., Kozlov, A. V., Lancaster, Jr., J. R., Lefer, D. J., McColl, K., McCurry, K., Patel, R. P., Petersson, J., Rassaf, T., Reutov, V. P., Richter-Addo, G. B., Schechter, A., Shiva, S., Tsuchiya, K., van Faassen,

E. E., Webb, A. J., Zuckerbraun, B. S., Zweier, J. L., and Weitzberg, E. (2009). Nitrate and nitrite in biology, nutrition and therapeutics. *Nat. Chem. Biol.*, 5(12):865–869.

[151] Lutz, S. and Meuwly, M. (2012). Photodissociation dynamics of clcn at different wavelengths. *Chem. Phys. Chem.*, 13(1):305.

[152] Lutz, S., Nienhaus, K., Nienhaus, G. U., and Meuwly, M. (2009). Ligand migration between internal docking sites in photodissociated carbonmonoxy neuroglobin. *J. Phys. Chem. B*, 113(46):15334.

[153] Ma, J., Huo, S., and Straub, J. (1997). Molecular dynamics simulation study of the B-states of solvated carbon monoxymyoglobin. *J. Am. Chem. Soc.*, 119(10):2541–2551.

[154] MacKerell, A. D. and Banavali, N. K. (2000). All-atom empirical force field for nucleic acids: Ii. application to molecular dynamics simulations of dna and rna in solution. *J. Comp. Chem.*, 21(2):105–120.

[155] MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, 102(18):3586–3616. PMID: 24889800.

[156] Mackerell, A. D., Feig, M., and Brooks, C. L. (2004). Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comp. Chem.*, 25(11):1400–1415.

[157] Manning, C. D., Raghavan, P., and Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press.

[158] Manzhos, S., Wang, X., Dawes, R., and Carrington, T. (2006). A nested molecule-independent neural network approach for high-quality potential fits. *J. Phys. Chem. A*, 110(16):5295–5304.

[159] Martin, A. C., Orengo, C. A., Hutchinson, E. G., Jones, S., Karmirantzou, M., Laskowski, R. A., Mitchell, J. B., Taroni, C., and Thornton, J. M. (1998). Protein folds and functions. *Structure*, 6(7):875–884.

[160] Martin, J., Migus, A., Poyart, C., Lecarpentier, Y., Astier, R., and Antonetti, A. (1983). *Proc. Natl. Acad. Sci.*, 80:173.

[161] Martyna, G. J., Klein, M. L., and Tuckerman, M. (1992). Nosé-hoover chains: The canonical ensemble via continuous dynamics. *J. Chem. Phys.*, 97(4):2635.

[162] Menting, J. G., Yang, Y., Chan, S. J., Phillips, N. B., Smith, B. J., Whittaker, J., Wickramasinghe, N. P., Whittaker, L. J., Pandyarajan, V., Wan, Z.-l., Yadav, S. P., Carroll, J. M., Strokes, N., Roberts, Jr., C. T., Ismail-Beigi, F., Milewski, W., Steiner, D. F., Chauhan, V. S., Ward, C. W., Weiss, M. A., and Lawrence, M. C. (2014). Protective hinge in insulin opens to enable its receptor engagement. *Proc. Natl. Acad. Sci.*, 111(33):E3395–E3404.

[163] Merchant, K. A., Noid, W. G., Thompson, D. E., Akiyama, R., Loring, R. F., and Fayer, M. D. (2003). Structural assignments and dynamics of the a substates of mbco: spectrally resolved vibrational echo experiments and molecular dynamics simulations. *J. Phys. Chem. B*, 107:4–7.

[164] Meuwly, M. (2006). On the influence of the local environment on the CO stretching freque ncies in native Myoglobin: Assignment of the B-states in MbCO. *Chem. Phys. Chem.*, 7:2061–2063.

[165] Meuwly, M., Becker, O. M., Stote, R., and Karplus, M. (2002). No rebinding to myoglobin: a reactive molecular dynamics study. *Biophys. Chem.*, 98(1-2):183–207.

[166] Meuwly, M. and Hutson, J. M. (1999). Morphing ab initio potentials: A systematic study of ne–hf. *The Journal of chemical physics*, 110(17):8338–8347.

[167] Meuwly, M. and Karplus, M. (2002). Simulation of proton transfer along ammonia wires: An "ab initio" and semiempirical density funct ional comparison of potentials and classical molecular dynamics. *J. Chem. Phys.*, 116:2572.

[168] Min, L., Strushkevich, N. V., Harnastai, I. N., Iwamoto, H., Gilep, A. A., Takemori, H., Usanov, S. A., Nonaka, Y., Hori, H., Vinson, G. P., et al. (2005). Molecular identification of adrenal inner zone antigen as a heme-binding protein. *FEBS J.*, 272(22):5832–5843.

[169] Mines, G. A., Pascher, T., Lee, S. C., Winkler, J. R., and Gray, H. B. (1996). Cytochrome c folding triggered by electron transfer. *Chem. Biol.*, 3(6):491 – 497.

[170] Moret, M.-E., Tavernelli, I., and Rothlisberger, U. (2009). Combined qm/mm and classical molecular dynamics study of [ru(bpy)(3)](2+) in water. *J. Phys. Chem. B*, 113:7737–7744.

[171] Mueller, E., Loida, P., Sligar, S., and De Montellano, P. O. (1995). *Cytochrome P450: Structure, Mechanism, and Biochemistry*. Plenum Press, New York.

[172] Nagy, T., Yosa Reyes, J., and Meuwly, M. (2014). Multisurface adiabatic reactive molecular dynamics. *J. Chem. Theo. Comp.*, 10(4):1366.

[173] Negrerie, M., Cianetti, S., Vos, M. H., Martin, J.-L., and Kruglik, S. G. (2006a). Ultrafast heme dynamics in ferrous versus ferric cytochrome c studied by time-resolved resonance raman and transient absorption spectroscopy. *J. Phys. Chem. B*, 110(25):12766–12781.

[174] Negrerie, M., Kruglik, S., Lambry, J., Vos, M., Martin, J., and Franzen, S. (2006b). Role of heme iron coordination and protein structure in the dynamics and geminate rebinding of nitric oxide to the H93G myoglobin mutant - Implications for nitric oxide sensors. *J. Biol. Chem.*, 281(15):10389–10398.

[175] Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *The computer journal*, 7:308.

[176] Nelson, K. V. and Benjamin, I. (2010). A molecular dynamics-empirical valence bond study of an sn2 reaction at the water/chloroform interface. *J. Phys. Chem. C*, 114(2):1154–1163.

[177] Newsome, D. A., Sengupta, D., Foroutan, H., Russo, M. F., and van Duin, A. C. T. (2012). Oxidation of silicon carbide by o2 and h2o: A reaxff reactive molecular dynamics study, part i. *The Journal of Physical Chemistry C*, 116(30):16111.

[178] Nielson, K. D., van Duin, A. C. T., Oxgaard, J., Deng, W.-Q., and Goddard, W. A. (2005). Development of the reaxff reactive force field for describing transition metal catalyzed reactions, with application to the initial stages of the catalytic formation of carbon nanotubes. *J. Phys. Chem. A*, 109(3):493. PMID: 16833370.

[179] Nienhaus, K., Lutz, S., Meuwly, M., and Nienhaus, G. U. (2010). Structural identification of spectroscopic substates in neuroglobin. *Chem. Phys. Chem.*, 11(1):119.

[180] Nienhaus, K., Lutz, S., Meuwly, M., and Nienhaus, G. U. (2013). Reaction-pathway selection in the structural dynamics of a heme protein. *Chem. Eur. J.*, 19(11):3558.

[181] Nienhaus, K., Olson, J. S., Franzen, S., and Nienhaus, G. U. (2005). The origin of stark splitting in the initial photoproduct state of mbco. *J. Am. Chem. Soc.*, 127(1):40–41.

[182] Nienhaus, K., Palladino, P., and Nienhaus, G. U. (2008). Structural dynamics of myoglobin: FTIR-TDS study of NO migration and binding. *Biochem.*, 47(3):935.

[183] Norrby, P.-O. and Liljefors, T. (1998). Automated molecular mechanics parameterization with simultaneous utilization of experimental and quantum mechanical data. *J. Comp. Chem.*, 19(10):1146–1166.

[184] Nosè, S. (1984a). A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.*, 52(2):255.

[185] Nosè, S. (1984b). A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.*, 81(1):511.

[186] Nosè, S. (1991). Constant temperature molecular dynamics methods. *Progress of Theoretical Physics Supplement*, 103:1.

[187] Nutt, D. R., Karplus, M., and Meuwly, M. (2005). Potential energy surface and molecular dynamics of MbNO: Existence of an unsuspected FeON minimum. *J. Phys. Chem. B*, 109:21118.

[188] Nutt, D. R. and Meuwly, M. (2003). Theoretical investigation of infrared spectra and pocket dynamics of photodissociated carbonmonoxy myoglobin. *Biophys. J.*, 85(6):3612.

[189] Nutt, D. R. and Meuwly, M. (2004a). CO migration in native and mutant myoglobin: atomistic simulations for the understanding of protein function. *Proc. Natl. Acad. Sci.*, 101(16):5998.

[190] Nutt, D. R. and Meuwly, M. (2004b). Ligand dynamics in myoglobin: Calculation of infrared spectra for photodissociated no. *Chem. Phys. Chem.*, 5(11):1710.

[191] Nutt, D. R. and Meuwly, M. (2006). Studying reactive processes with classical dynamics: rebinding dynamics in mbno. *Biophys. J.*, 90(4):1191.

[192] Olkhova, E., Hutter, M. C., Lill, M. A., Helms, V., and Michel, H. (2004). Dynamic water networks in cytochrome c oxidase from paracoccus denitrificans investigated by molecular dynamics simulations. *Biophys. J.*, 86(4):1873.

[193] Oostenbrink, C., Villa, A., Mark, A. E., and Van Gunsteren, W. F. (2004). A biomolecular force field based on the free enthalpy of hydration and solvation: The gromos force-field parameter sets 53a5 and 53a6. *J. Comp. Chem.*, 25(13):1656.

[194] Ostermann, A., Waschipky, R., Parak, F. G., and Nienhaus, G. U. (2000). Ligand binding and conformational motions in myoglobin. *Nature*, 404(6774):205–208.

[195] Ostermeier, C., Harrenga, A., Ermler, U., and Michel, H. (1997). Structure at 2.7 a resolution of the paracoccus denitrificans two-subunit cytochrome c oxidase complexed with an antibody fv fragment. *Proc. Natl. Acad. Sci.*, 94(20):10547.

[196] Ostermeier, C., Iwata, S., and Michel, H. (1996). Cytochrome c oxidase. *Curr. op. struc. biol.*, 6(4):460–466.

[197] Pacher, P., Beckman, J. S., and Liaudet, L. (2007). Nitric oxide and peroxynitrite in health and disease. *Physiol. Rev.*, 87(1):315–424.

[198] Park, E. S., Thomas, M. R., and Boxer, S. G. (2000). Vibrational Stark spectroscopy of NO bound to heme: Effects of protein electrostatic fields on the NO stretch frequency. *J. Am. Chem. Soc.*, 122:12297.

[199] Park, S. and Pan, L. P. (1996). Photoperturbation of the heme a3-cub binuclear center of cytochrome c oxidase CO complex observed by fourier transform infrared spectroscopy. *Biophys. J.*, 71(2):1036.

[200] Patel, S. and Brooks, C. L. (2004). Charmm fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations. *J. Comp. Chem.*, 25(1):1–16.

[201] Perutz, M. F., Rossmann, M. G., Cullis, A. F., Muirhead, H., Will, G., and North, A. C. T. (1960). Structure of haemoglobin: A three-dimensional fourier synthesis at 5.5-[angst]. resolution, obtained by x-ray analysis. *Nature*, 185(4711):416–422.

[202] Perutz, M. F., Wilkinson, A., Paoli, M., and Dodson, G. (1998). The stereochemical mechanism of the cooperative effects in hemoglobin revisited. *Ann. Rev. Biophys. Biomol. Struc.*, 27(1):1–34.

[203] Petersson, G. and Al-Laham, M. A. (1991). A complete basis set model chemistry. ii. open-shell systems and the total energies of the first-row atoms. *J. Chem. Phys.*, 94:6081.

[204] Petersson, G., Bennett, A., Tensfeldt, T., Al-Laham, M. A., Shirley, W. A., and Mantzaris, J. (1988). A complete basis set model chemistry. i. the total energies of closed-shell atoms and hydrides of the first-row atoms. *J. Chem. Phys.*, 89:2193.

[205] Petrich, J. W., Lambry, J. C., Kuczera, K., Karplus, M., Poyart, C., and Martin, J. L. (1991). Ligand binding and protein relaxation in heme proteins: a room temperature analysis of nitric oxide geminate recombination. *Biochem.*, 30(16):3975.

[206] Pierce, L. C., Salomon-Ferrer, R., Augusto F. de Oliveira, C., McCammon, J. A., and Walker, R. C. (2012). Routine access to millisecond time scale events with accelerated molecular dynamics. *Journal of Chemical Theory and Computation*, 8(9):2997–3002.

[207] Pinakoulaki, E., Ohta, T., Soulimane, T., Kitagawa, T., and Varotsis, C. (2004). Simultaneous resonance raman detection of the heme a3-fe-co and cub-co species in CO-bound ba3-cytochrome c oxidase from thermus thermophilus. evidence for a charge transfer cub-co transition. *J. Biol. Chem.*, 279(22):22791.

[208] Plattner, N., Doll, J. D., and Meuwly, M. (2010). Spatial averaging for small molecule diffusion in condensed phase envir onments. *J. Chem. Phys.*, 133:044506.

[209] Plattner, N. and Meuwly, M. (2008). The role of higher CO-multipole moments in understanding the dynamics of photodissociated carbonmonoxide in myoglobin. *Biophys. J.*, 94(7):2505.

[210] Plattner, N. and Meuwly, M. (2012). Quantifying the importance of protein conformation on ligand migration in myoglobin. *Biophys. J.*, 102(2):333.

[211] Porrini, M., Daskalakis, V., Farantos, S. C., and Varotsis, C. (2009). Heme cavity dynamics of photodissociated CO from ba3-cytochrome c oxidase: the role of ring-d propionate. *J. Phys. Chem. B*, 113(35):12129.

[212] Privalle, C. T., Crivello, J. F., and Jefcoate, C. R. (1983). Regulation of intramitochondrial cholesterol transfer to side-chain cleavage cytochrome p-450 in rat adrenal gland. *PNAS*, 80(3):702–706.

[213] Raphael, A. L. and Gray, H. B. (1991). Semisynthesis of axial-ligand (position 80) mutants of cytochrome c. *J. Am. Chem. Soc.*, 113(3):1038–1040.

[214] Reedy, C. J., Elvekrog, M. M., and Gibney, B. R. (2008). Development of a heme protein structure-electrochemical function database. *Nucl. Acids Res.*, 36(suppl 1):D307–D313.

[215] Reedy, C. J. and Gibney, B. R. (2004). Heme protein assemblies. *Chem. Rev.*, 104(2):617–650.

[216] Ren, P. and Ponder, J. W. (2003). Polarizable atomic multipole water model for molecular mechanics simulation. *J. Phys. Chem. B*, 107(24):5933–5947.

[217] Rich, A. M., Armstrong, R. S., Ellis, P. J., and Lay, P. A. (1998). Determination of the fe-ligand bond lengths and fe-n-o bond angles in horse heart ferric and ferrous nitrosylmyoglobin using multiple-scattering xafs analyses. *J. Am. Chem. Soc.*, 120:10827.

[218] Rieger, M. and Westheimer, F. H. (1950). The calculation and determination of the buttressing effect for the racemization of 2,2′,3,3′-tetraiodo-5,5′-dicarboxybiphenyl. *J. Am. Chem. Soc.*, 72(1):19.

[219] Rodgers, K. R. (1999). Heme-based sensors in biological systems. *Curr. op. chem. biol.*, 3(2):158–167.

[220] Rovira, C., Schulze, B., Eichinger, M., Evanseck, J. D., and Parrinello, M. (2001). Influence of the heme pocket conformation on the structure and vibrations of the fe-co bond in Myoglobin: a qm/mm density functional st udy. *Biophys. J.*, 81:435–445.

[221] Ruscio, J. Z., Kumar, D., Shukla, M., Prisant, M. G., Murali, T. M., and Onufriev, A. V. (2008). Atomic level computational identification of ligand migration pathways between solvent and binding site in Myoglobin. *Proc. Natl. Acad. Sci.*, 105:9204–9209.

[222] Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Chem. Phys.*, 13:327.

[223] Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1988). Primer-directed enzymatic amplification of dna with a thermostable dna polymerase. *Science*, 239(4839):487–491.

[224] Schaad, O., Zhou, H.-X., Szabo, A., Eaton, W. A., and Henry, E. R. (1993). Simulation of the kinetics of ligand binding to a protein by molecular dynamics: geminate rebinding of nitric oxide to myoglobin. *Proc. Natl. Acad. Sci.*, 90:9547.

[225] Schlichting, I., Berendzen, J., Phillips, G., and Sweet, R. (1994). Crystal structure of photolysed carbonmonoxy-myoglobin. *Nature*, 371(6500):808–812.

[226] Schmitt, U. and Voth, G. (1998). Multistate empirical valence bond model for proton transport in water. *J. Phys. Chem. B*, 102:5547.

[227] Scholander, P. (1960). Oxygen transport through hemoglobin solutions. *Science*, 131(3400):585–590.

[228] Schölkopf, B., Burges, C., and Vapnik, V. (1995). Extracting support data for a given task. In Fayyad, U. M. and Uthurusamy, R., editors, *Proceedings, First International Conference on Knowledge Discovery & Data Mining*, Menlo Park. AAAI Press.

[229] Schotte, F., Lim, M., Jackson, T. A., Smirnov, A. V., Soman, J., Olson, J. S., Phillips, G. N., Wulff, M., and Anfinrud, P. A. (2003). Watching a protein as it functions with 150-ps time-resolved x-ray crystallography. *Science (New York, N.Y.)*, 300(5627):1944.

[230] Schuresko, D. and Webb, W. (1978). Carboxylation kinetics of hemoglobin and myoglobin: linear transient response to step perturbation by laser photolysis. *Biophys. J.*, 24(1):382–383.

[231] Scott, R. A. and Mauk, A. G. (1996). *Cytochrome c: a multidisciplinary approach*. Univ Science Books.

[232] Shaw, D. E., Dror, R. O., Salmon, J. K., Grossman, J. P., Mackenzie, K. M., Bank, J. A., Young, C., Deneroff, M. M., Batson, B., Bowers, K. J., Chow, E., Eastwood, M. P., Ierardi, D. J., Klepeis, J. L., Kuskin, J. S., Larson, R. H., Lindorff-Larsen, K., Maragakis, P., Moraes, M. A., Piana, S., Shan, Y., and Towles, B. (2009). Millisecond-scale molecular dynamics simulations on anton. In *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*, SC '09, pages 39:1–39:11, New York, NY, USA. ACM.

[233] Shelnutt, J. A., Medforth, C. J., et al. (1998). Nonplanar porphyrins and their significance in proteins. *Chem. Soc. Rev.*, 27(1):31–42.

[234] Sigman, J. A., Kim, H. K., Zhao, X., Carey, J. R., and Lu, Y. (2003). The role of copper and protons in heme-copper oxidases: Kinetic study of an engineered heme-copper center in myoglobin. *Proc. Natl. Acad. Sci.*, 100(7):3629–3634.

[235] Sigman, J. A., Kwok, B. C., and Lu, Y. (2000). From myoglobin to heme-copper oxidase : Design and engineering of a cu b center into sperm whale myoglobin. *J. Am. Chem. Soc.*, 122(34):8192.

[236] Silatani, M., Lima, F. A., Penfold, T. J., Rittman, J., Reinhard, M., Rittmann-Frank, H., Borca, C., Grolimund, D., Milne, C. J., and Chergui, M. (2015). No binding kinetics in myoglobin investigated by picosecond fe k-edge absorption spectroscopy. *Proc. Natl. Acad. Sci.*, 112():12922 – 12927.

[237] Šrajer, V., Ren, Z., Teng, T.-Y., Schmidt, M., Ursby, T., Bourgeois, D., Pradervand, C., Schildkamp, W., Wulff, M., and Moffat, K. (2001). Protein conformational relaxation and ligand migration in myoglobin: a nanosecond to millisecond molecular movie from time-resolved laue x-ray diffraction. *Biochem.*, 40(46):13802–13815.

[238] Šrajer, V., Teng, T.-y., Ursby, T., Pradervand, C., Ren, Z., Adachi, S.-i., Schildkamp, W., Bourgeois, D., Wulff, M., and Moffat, K. (1996). Photolysis of the carbon monoxide complex of myoglobin: nanosecond time-resolved crystallography. *Science*, 274(5293):1726–1729.

[239] Stephens, P., Devlin, F., Chabalowski, C., and Frisch, M. J. (1994). Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem.*, 98(45):11623–11627.

[240] Stokes, G. G. (1863). On the reduction and oxidation of the colouring matter of the blood. *Proc. Royal Soc. London*, 13:pp. 355–364.

[241] Stoutland, P., Lambry, J., Martin, J., and Woodruff, W. (1991). Femtosecond dynamics of reduced cytochrome-oxidase and its CO derivative. *J. Phys. Chem.*, 95(17):6406.

[242] Strachan, A., Kober, E. M., van Duin, A. C. T., Oxgaard, J., and Goddard, W. A. (2005). Thermal decomposition of rdx from reactive molecular dynamics. *J. Chem. Phys.*, 122(5):054502–1 – 054502–10.

[243] Strickland, N. and Harvey, J. N. (2007). Spin-forbidden ligand binding to the ferrous-heme group: Ab initio and dft studies. *J. Phys. Chem. B*, 111(4):841.

[244] Swope, W. C., Andersen, H. C., Berens, P. H., and Wilson, K. R. (1982). A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J. Chem. Phys.*, 76(1):637–649.

[245] Tezcan, F. A., Winkler, J. R., and Gray, H. B. (1998). Effects of ligation and folding on reduction potentials of heme proteins. *J. Am. Chem. Soc.*, 120(51):13383–13388.

[246] Tiefenbrunn, T., Liu, W., Chen, Y., Katritch, V., Stout, C. D., Fee, J. A., and Cherezov, V. (2011). High resolution structure of the *ba3* cytochrome *c* oxidase from *thermus thermophilus* in a lipidic environment. *PLoS ONE*, 6(7):e22348.

[247] Tikhonov, A. N. (1943). [On the stability of inverse problems]. *Doklady Akademii Nauk SSSR*, 39(5):195.

[248] Tikhonov, A. N. (1963). [Solution of incorrectly formulated problems and the regularization method]. *Doklady Akademii Nauk SSSR*, 151(52):501.

[249] Tilton, R. F., Kuntz, I. D., and Petsko, G. A. (1984). Cavities in proteins: structure of a metmyoglobin xenon complex solved to 1.9 .ang. *Biochem.*, 23(13):2849.

[250] Toxvaerd, S. and Olsen, O. H. (1990). Canonical molecular dynamics of molecules with internal degrees of freedom. *Ber. Buns. Phys. Chem.*, 94(3):274.

[251] Traylor, T. and Sharma, V. (1992). WHY NO. *Biochem.*, 31(11):2847–2849.

[252] Treuffet, J., Kubarych, K. J., Lambry, J.-C., Pilet, E., Masson, J.-B., Martin, J.-L., Vos, M. H., Joffre, M., and Alexandrou, A. (2007). Direct observation of ligand transfer and bond formation in cytochrome c oxidase by using mid-infrared chirped-pulse upconversion. *Proc. Natl. Acad. Sci.*, 104(40):15705.

[253] Truhlar, D. G., Garrett, B. C., and Klippenstein, S. J. (1996). Current Status of Transition-State Theory. *J. Phys. Chem.*, 100(31):12771.

[254] van Duin, A. C. T., Dasgupta, S., Lorant, F., and Goddard, W. A. (2001). Reaxff: A reactive force field for hydrocarbons. *J. Phys. Chem. A*, 105(41):9396.

[255] van Duin, A. C. T., Strachan, A., Stewman, S., Zhang, Q., Xu, X., and Goddard, W. A. (2003). Reaxffsio reactive force field for silicon and silicon oxide systems. *J. Phys. Chem. A*, 107(19):3803.

[256] Varadarajan, R., Zewert, T. E., Gray, H. B., and Boxer, S. G. (1989). Effects of buried ionizable amino acids on the reduction potential of recombinant myoglobin. *Science*, 243(4887):69–72.

[257] Čuma, M., Schmitt, U. W., and Voth, G. A. (2001). A multi-state empirical valence bond model for weak acid dissociation in aqueous solution. *J. Phys. Chem. A*, 105(12):2814.

[258] Vega, C., Sanz, E., and Abascal, J. L. F. (2005). The melting temperature of the most common models of water. *J. Chem. Phys.*, 122(11).

[259] Verlet, L. (1967). Computer" experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Phys. rev.*, 159(1):98.

[260] Villà, J. and Warshel, A. (2001). Energetics and dynamics of enzymatic reactions. *J. Phys. Chem. B*, 105(33):7887.

[261] Vogel, K. M., Kozlowski, P. M., Zgierski, M. Z., and Spiro, T. G. (2000). Role of the axial ligand in heme-co backbonding; dft analysis of vibrational data. *Inorg. Chim. Acta*, 297(1-2):11.

[262] Voigt, P. and Knapp, E.-W. (2003). Tuning heme redox potentials in the cytochrome c subunit of photosynthetic reaction centers. *J. Biol. Chem.*

[263] Vojtěchovskỳ, J., Chu, K., Berendzen, J., Sweet, R. M., and Schlichting, I. (1999). Crystal structures of myoglobin-ligand complexes at near-atomic resolution. *Biophysical journal*, 77(4):2153–2174.

[264] Vos, M. H. (2008). Ultrafast dynamics of ligands within heme proteins. *Bioch. Bioph. Acta*, 1777(1):15–31.

[265] Vosko, S., Wilk, L., and Nusair, M. (1980). Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.*, 58(8):1200–1211.

[266] Vuilleumier, R. and Borgis, D. (1998). An extended empirical valence bond model for describing proton transfer in h+(h20)n clusters and liquid water. *Chem. Phys. Lett.*, 284(1):71.

[267] Wang, J., Cieplak, P., and Kollman, P. A. (2000). How well does a restrained electrostatic potential (resp) model perform in calculating conformational energies of organic and biological molecules? *J. Comp. Chem.*, 21(12):1049–1074.

[268] Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004). Development and testing of a general amber force field. *J. Comp. Chem.*, 25(9):1157.

[269] Warshel, A. (2003). Computer simulations of enzyme catalysis: Methods, progress, and insights. *Ann. Rev. Biophys. Biomol. Struc.*, 34:425.

[270] Warshel, A. and W. Parson, W. (2001). Dynamics of biochemical and biophysical reactions: insight from computer simulations. *Quart. Rev. Biophys.*, 34:563.

[271] Warshel, A. and Weiss, R. M. (1980). An empirical valence bond approach for comparing reactions in solutions and in enzymes. *J. Am. Chem. Soc.*, 102:6218.

[272] Warshel, A. and Weiss, R. M. (1981). Empirical valence bond calculations of enzyme catalysis*. *Ann. New York Acad. Sc.*, 367(1):370.

[273] Westheimer, F. H. (1947). A calculation of the energy of activation for the racemization of 2,2'-dibromo-4,4'-dicarboxydiphenyl. *J. Chem. Phys.*, 15(5):252.

[274] Westheimer, F. H. and Mayer, J. E. (1946). The theory of the racemization of optically active derivatives of diphenyl. *J. Chem. Phys.*, 14(12):733.

[275] Wittenberg, B., Wittenberg, J., and Caldwell, P. (1975). Role of myoglobin in the oxygen supply to red skeletal muscle. *J. Biol. Chem.*, 250(23):9038–9043.

[276] Woodcock, L. (1971). Isothermal molecular dynamics calculations for liquid salts. *Chem. Phys. Lett.*, 10:257.

[277] Xu, N., Yi, J., and Richter-Addo, G. B. (2010). Linkage Isomerization in Heme-NOx Compounds: Understanding NO, Nitrit e, and Hyponitrite Interactions with Iron Porphyrins. *Inorg. Chem.*, 49(14):6253–6266.

[278] Ye, X., Demidov, A., and Champion, P. (2002). Measurements of the photodissociation quantum yields of MbNO and MbO(2) and the vibrational relaxation of the six-coordinate heme species. *J. Am. Chem. Soc.*, 124(20):5914–5924.

[279] Yoneya, M., Berendsen, H. J. C., and Hirasawa, K. (1994). A non-iterative matrix method for constraint molecular dynamics simulations. *Mol. Simulation*, 13:395.

[280] Yoo, B.-K., Kruglik, S. G., Lamarre, I., Martin, J.-L., and Negrerie, M. (2012). Absorption Band III Kinetics Probe the Picosecond Heme Iron Motion Triggered by Nitric Oxide Binding to Hemoglobin and Myoglobin. *J. Phys. Chem. B*, 116(13):4106–4114.

[281] Yosa Reyes, J., Nagy, T., and Meuwly, M. (2014). Competitive reaction pathways in vibrationally induced photodissociation of h2so4. *PCCP*, 16:18533.

[282] Yoshikawa, S., Shinzawa-Itoh, K., Nakashima, R., and et al (1998). Redox-coupled crystal structural changes in bovine heart cytochrome c oxidase. *Science*, 280(5370):1723.

[283] Zahran, Z. N., Chooback, L., Copeland, D. M., West, A. H., and Richter-Addo, G. B. (2008). Crystal structures of manganese- and cobalt-substituted myoglobin in complex with {NO} and nitrite reveal unusual ligand conformations. *J. of Inorg. Biochem.*, 102:216.

[284] Zhang, P., Ahn, S. W., and Straub, J. E. (2013). "strange kinetics" in the temperature dependence of methionine ligand rebinding dynamics in cytochrome c. *J. Phys. Chem. B*, 117(24):7190.

[285] Zhang, Y., Feller, S. E., Brooks, B. R., and Pastor, R. W. (1995). Computer simulation of liquid/liquid interfaces. i. theory and application to octane/water. *J. Chem. Phys.*, 103(23):10252.