Computational Investigation of Conformational and Spectroscopic Substates in Neuroglobin

Inauguraldissertation

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

von
Stephan Alexander Lutz
aus Walzenhausen, AR

Basel, 2011

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

auf Antrag von:
Prof. Dr. Markus Meuwly
Prof. Dr. Gerd Ulrich Nienhaus

Basel, den 22. Februar 2011

Prof. Dr. M. Spiess (Dekan)

Acknowledgments

I would like to thank my supervisor Prof. Markus Meuwly for giving me the opportunity to work as a doctoral student in his group and providing me with such a fascinating research project. I appreciated his enthusiasm, the guidance & support, and the inspiring discussions throughout my PhD study.

I would also like to thank Prof. G. Ulrich Nienhaus and Dr. Karin Nienhaus for their interesting and fruitful collaboration on the investigation of conformational substates in photolyzed carbonmonoxy neuroglobin and to Prof. G. Ulrich Nienhaus, especially, for kindly acting as external examiner.

It has been a pleasure to work with the members of the Meuwly group. In particular, I appreciated the constructive discussions with Dr. Nuria Plattner, Dr. Michael Devereux, and Dr. Jonas Danielsson.

Finally, I would like to thank my family and friends for their love and support during the last four years.

Abstract

The family of globins has a strong significance in the history of protein research. Their abundance, spread in nature, and evolutionary diversity are crucial factors for a wide range of approved and suggested physiological functions taking place in almost every organism. The versatility of ligand binding and unbinding and the tendency to act as a catalyst for the oxidation of certain substrates in a selected number of globins is strongly related to specific structural and dynamical properties of the protein.

Neuroglobin, a structurally very related protein to myoglobin, is one of several newly discovered globins which show some unique characteristics in this family of proteins. Most importantly it was found that the heme prosthetic group of the neuroglobin can become hexacoordinated to an internal histidine residue in the absence of an exogenous ligand. This feature, which is meant to control the affinity for external ligand binding, lead to many speculations about the physiological function of neuroglobin. Some suggested functions involve enhanced oxygen supply in neuronal cells, regulation of hypoxia, and radical scavenging.

Here, a collection of specialized molecular dynamics methods is used to investigate the structural and dynamical peculiarities of neuroglobin in more detail. Multipolar force fields are used to energetically, dynamically and spectroscopically determine the migration network of CO in photolyzed carbonmonoxy neuroglobin. A combination of classical molecular dynamics and quantum mechanical techniques elucidate the conformational substates of carbonmonoxy neuroglobin which are experimentally accessible with infrared spectroscopy. Furthermore, the nature of the configurational transitions occurring during the competition for internal and external ligand binding are analyzed by the adiabatic reactive molecular dynamics method.

Contents

Acknowledgments i

Abstract ii

Abbreviations v

Units vi

1 Introduction 1

2 Important Findings in Globin Research 3

2.1 Significance in nature 3

2.2 Protein structural dynamics in Mb 9

2.3 Ligand dissociation and migration dynamics in Mb 13

3 Structure and Function of Neuroglobin 17

3.1 Proposed functional significances 18

3.2 Structural features 19

3.3 Dynamical features 21

4 Theoretical Background 25

4.1 Molecular dynamics simulations 25

4.1.1 Historical notes 25

4.1.2 Basic theory 26

4.2 Combined molecular/quantum methods 29

4.3 Polarizable and multipolar force fields 32

4.3.1 Overview 32

4.3.2 The MTP force field 33

4.4 Adiabatic reactive molecular dynamics 34

4.5 Force field parametrization 38

4.5.1 Heme prosthetic group 39

4.5.2 ARMD specific parameters 42

5 Competitive Rebinding in Neuroglobin 47

5.1 Methods 48
5.1.1 Computational preparations 48
5.1.2 Bis-His dissociation dynamics 49
5.1.3 Umbrella sampling simulations 49
5.2 Heme sliding 50
5.3 Cavity evolution 53
5.3.1 Docking site expansion 53
5.3.2 Water population 57
5.4 Solvent influx and the role of Tyr44 61
5.4.1 Water accessibility to the heme active site 61
5.4.2 The role of Tyr44 65
5.5 Biphasic rebinding barriers 72
5.6 Conclusion 76

6 Published Work and Manuscripts Submitted for Publication 79

6.1 Ligand migration between internal docking sites in photodissociated carbonmonoxy neuroglobin 80
6.2 Structural identification of spectroscopic substates in neuroglobin . . . 91
6.3 Structural characterization of spectroscopical substates in carbonmonoxy neuroglobin 103

7 Conclusion 125

References 128

A Appendix 153

A.1 Parameter tables 153

Abbreviations

ARMD	Adiabatic Reactive Molecular Dynamics
B3LYP	Becke 3-Parameter Lee-Yang-Parr
bis-His	Axially, by two histidine residues, coordinated heme group
CHARMM	Chemistry at HARvard Molecular Mechanics
CO	Carbon monoxide
CS	Conformational substate
cx	Carbonmonoxy coordinated heme state
dx	Pentacoordinated heme state
Hb	Hemoglobin
hNgb	Human Neuroglobin
IR	Infrared
LJ	Lennard-Jones
Mb	Myoglobin
MbCO	Carbonmonoxy Myoglobin
MD	Molecular Dynamics
metNgb	Oxidized (Fe^{3+}) Neuroglobin
MM	Molecular Mechanics
mNgb	Murine Neuroglobin
Ngb	Neuroglobin
NgbCO	Carbonmonoxy Neuroglobin
NPT	Isothermal-isobaric ensemble
NVT	Canonical ensemble
QM	Quantum Mechanics
RMSD	Root Mean Square Deviation
wt	Wild type

Units

Å	Ångström
e	Elementary charge
K	Kelvin
kcal	Kilocalorie
kJ	Kilojoule
fs	Femtosecond
ps	Picosecond
ns	Nanosecond
μs	Microsecond
cm^{-1}	Wavenumber

1. Introduction

"My dear Miss Marianne, you entirely miscomprehend this business. No wonder. How should you know any thing of our legal proceedings, so intricate and difficult to understand, that egad! we are sometimes puzzled to make them out ourselves? But, my dear madam, you are in the hands of your best friends – persons who have the tenderest regard for your interests."

Sir James Bland Burgess, *The Advertisement, 1817*

The understanding of the overarching cooperation between organs, cells and biomolecules taking place inside every living organism is major aim in the field of modern biology and medicine. A countless number of different molecules, mechanisms, conditions, and stimulations, present or taking place inside every biological system, are involved in guiding the organism through its everyday life. The interplay of all these processes are important in many different disciplines in biology which need to be looked at as an entity in order to get a complete survey of the problem. This holistic rather than reductive view on biological systems emerged in recent years into a scientific branch named systems biology (1–5). Some important disciplines associated with systems biology are genomics (study and comparison of DNA sequences in different organisms), transcriptomics (gene expression), proteomics (structure and function of proteins) metabolomics (study of molecular fingerprints deriving from cellular processes), fluxomics (prediction of the cell level of metabolites by studying molecular dynamic changes) and phenomic (organismal variation during the life span). Genomics is by far the most strongly matured discipline. It started to evolve in 1866 by Gregor J. Mendel's experiments on plant hybridization and reached its current climax in 2003 when the complete determination of the human genome sequence was announced (6, 7). Other profound knowledge was obtained in the field of metabolomics which was pushed by the human urge to find cures against metabolic disorders. In between these two stages of the biochemical pathway transcriptomics and proteomics are situated. The first one essentially tries to understand the processes involved in cellular differentiation by comparing individual RNA expression among different types of cells. The second one deals with the translation of the mRNA into the corresponding protein sequence, post-translational modifications and protein folding. At this stage a biocatalyst that forms the principle component of metabolomics has formed. Although tremendous progress in the determination of structure and physiological function of proteins and other biomolecules were made over the last couple of decades, only small pieces of the widespread metabolic puzzle could be solved so far. The main reason for that is the incredible number and the large variation of different biomolecules even present in evolutionary very fundamental organisms. Nevertheless, proteins for which the basic functionality was known many years ago, like Mb, show a much more complex physiological picture nowadays. In the early days of protein structural determination it was believed that the function of the enzyme can easily be determined by structural interpretation. More detailed observations revealed mechanisms like allostery, induced fit and equilibrium fluctuations which all are essential to the proper function of the protein. Today, these dynamical aspects, which are heavily depending on the physiological conditions, are thought of as the major source of complexity.

Ngb has, among several new discovered globins, expanded the range of putative physiological functions in this family of proteins (8–12). On the one hand, the strong sequence diversity of globins found among different organisms is regarded as a key factor for the functional individualism. On the other hand, the tertiary structure of different globins has found to be, in most cases, nearly identical. Therefore, the functional specificity must heavily rely on the dynamical rather than the structural behavior of the protein under the given physiological conditions. Most of todays experimental techniques which uncover protein dynamics use spectroscopic methods. The majority of these measurements can not directly disclose the structural changes as a result of the protein dynamics. This work tries to fill this gap using advanced MD algorithms to interpret spectroscopic observation in a structural as well as in a dynamical point of view. It concentrates on the interpretation of vibrational spectra obtained for different configurations and conformations of NgbCO with classical and quantum mechanical calculations. Furthermore, it tries to give a more detailed picture of the Ngb specific ligand binding kinetics.

2. Important Findings in Globin Research

"Come blow you Eastern Winds, and all you four Points joyne here in one to make a prosperous Gale, that by the vertue of that sweet structure, it may breath some life to my dying Love; wherein I may relate my folly in doing this unto her: Oh joyful sight! she breaths; hold her up! give her more air!"

Thomas Meriton, *The Wandring Lover, 1658*

This chapter gives a short overview on the historical development in globin research and its current state. The first part covers the magnitude of the super-family of globins in nature discovered up to the present, its functional diversity and impact on many known diseases. Subsequently, previous experimental studies, primarily performed for MbCO, related to equilibrium dynamics and ligand binding and photodissociation dynamics are recapitulated. Finally, the theory of protein dynamics is put more into a structural picture focusing on the importance of the internal docking sites to the mechanism, sensitivity and variety of the different physiological functions which are performed by this group of proteins.

2.1. Significance in nature

Globins are probably the best studied group of proteins so far. There are several reasons for this. First, their are highly abundant in rather large quantities in a multitude of organisms (13, 14). According to the UniProt Knowledgebase more than 1100 primary globin structures were determined from eukaryotic and prokaryotic species up to the present (15). Second, due to the large concentration of Hb in vertebrates blood (4.9 to 5.5 mmol/l) (16) the protein was discovered incredibly early in the history of life science by Hünefeld in 1840. Hbs elementary function (reversible binding of O_2) was described a few years later by Hoppe-Seyler who also established its name (17). Furthermore, the large availability and the relatively easy procedure of isolation and crystallization of Mb cleared the way to have its tertiary structure determined as a prototype of all protein crystal structures in 1958 by Kendrew and co-workers (18). Third, globin malfunctions are involved in many diseases like sickle-cell anaemia (19), Rhabdomyolysis (rapid breakdown of skeletal muscle) (20), and delayed vasospasm after subarachnoid hemorrhage (bleeding into cerebrospinal fluid after head trauma) (21) and are therefore also under heavy investigation in medical research.

The general protein structure of the super-family of globins characteristically possesses an iron-protoporphyrin IX as prosthetic group, which is often referred to as the heme group. Most globins are composed of 140 to 160 amino acids with a few exceptions which were mostly discovered in recent years. Although the amino acid sequence homology between distantly related globins is low, they have a very similar tertiary structure. Figure 2.1 shows the crystal structure of sperm whale MbCO which serves as a representative of this typical globin fold, often referred to as myoglobin fold. It consists of eight α-helices (designated as A-H in Figure 2.1) which are oriented in a three-over-three globular arrangement. The heme group is sandwiched between helices E and F and coordinated over its iron atom to the F-helix by an invariant histidine residue in position F8 (eighth amino acid residue in helix F) on the so called proximal site. The discovery of globins in more and more groups of organisms over the years have been accompanied by an increase in the number of known types of globins (14). Despite the often very low amino acid sequence identity between this different sub-families of globins (often less than 25%), there are a few amino acids which are highly or even strictly conserved within the globin super-family. This amino acids are CD1-Phe, E7-His and naturally F8-His (shown in Figure 2.1). E7-His is substituted by glutamic acid in some globins of vertebrates, and by Val, Glu, Leu, Tyr, or Glu in rather more globins of invertebrates. A distinct feature of E7-His, which is located on the so called distal site of the heme group, is the formation of a stable hydrogen-bond with a heme iron-bound CO (see Figure 2.1) or O_2 ligand

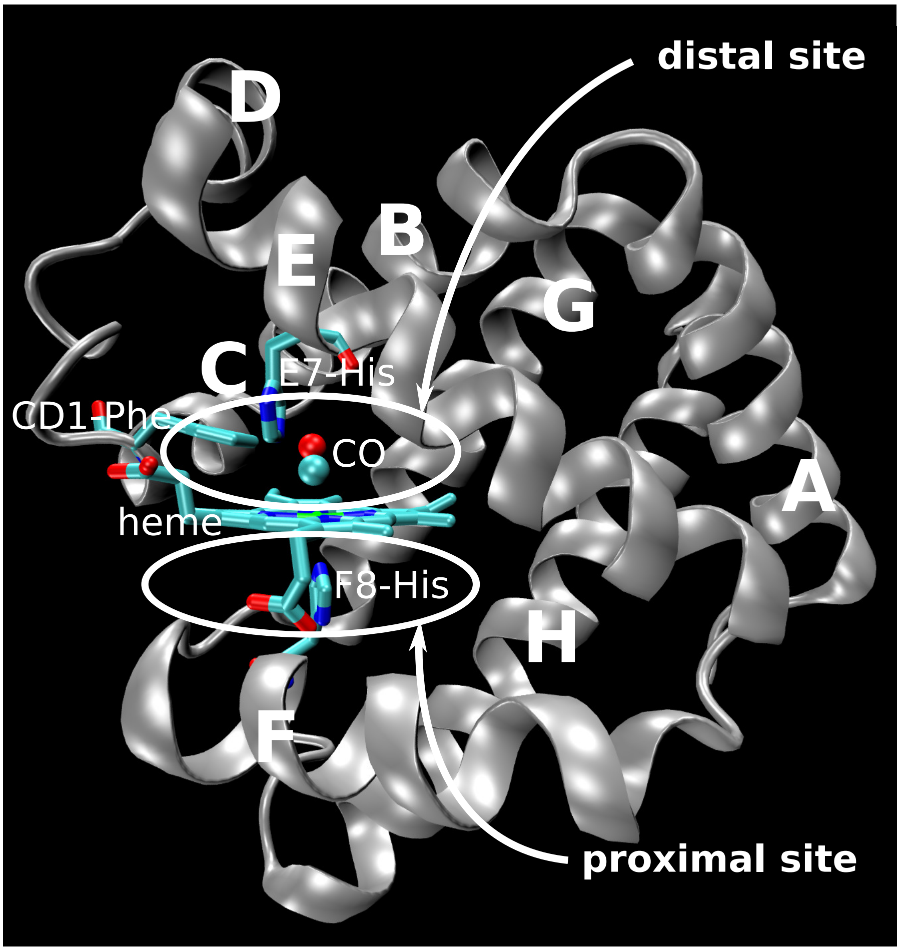
distal site
D
B
E
G
C
E7-His
CD1-Phe
CO
A
heme
F8-His
H
F
proximal site

Figure 2.1.: Crystal structure of sperm whale MbCO. Helices A to H are labeled in bold letters. Amino acids which are conserved within the globin superfamily are marked. The heme prosthetic group divides the huge internal protein cavity into the distal and proximal sites which are indicated as circles.

(22). The formation of a hexacoordinated heme species, in which four coordination sites occupied by the porphyrin nitrogens, one by the Nϵ atom of the proximal F8-His sidechain and the last one by an oxygen atom of O_2 on the distal site, defines the functionally important, reversibly bound O_2 configuration for globins. Having the iron in its ferrous state (Fe^{2+}) is an essential prerequisite for the reversibility of the O_2 binding. The hydrophobic environment formed by the distal site protein cavity is to a large amount responsible for shielding the heme-iron against attacks from oxidatively active species like peroxides. However, under physiological conditions the peroxidative chemistry needs to be additionally kept at a minimum by an extensive antioxidant defense system persisting of heme reductase, glutathione, ascorbate, uric acid, *etc.* to prevent the iron from going from the ferrous to the oxidized ferric (Fe^{3+}) state (23).

The huge variety of globins that has evolved in almost every species in the prokaryotic and eukaryotic kingdoms indicates the importance of O_2 binding for the aerobic as well as anaerobic metabolisms. Among the different families of globins the binding affinity for O_2 and other exogenous ligands is diverse. This is most likely related to versatile functional requirements, different physiological conditions, the iron oxidation state and other differences present in various biological systems. The diversity of O_2 affinity arises primarily from differences in the amino acid sequence of the polypeptide chain. Probably one of the best studied mechanisms for the regulation of O_2 affinity is the allosterically induced conformational transition from the so called T state to the R state in Hb (24, 25). Hb consists of a tetrameric structure of individual globin subunits. Each subunit is able to bind one O_2 ligand. Binding of the first ligand increases the transition rate from the low-affinity T state conformation to the high-affinity R state slightly. Supplementary binding of O_2 to other subunits further stabilizes the R state conformation in every subunit which results in a cooperative O_2 binding affinity for Hb. Other known allosteric cofactors, which are influencing ligand binding affinities even for monomeric globins, are H^+ (pH dependence), CO_2, phosphates, lactate, Ca^{2+} and Mg^{+2} (26–28). The principle of allostery introduces dynamical and environmental aspects to the mechanism of ligand binding as it is not only varying among the different globin species but also depending on the current physiological conditions surrounding it in the host cell. Despite this high complexity, the range of physiological purposes suggested for globins up to the present can roughly be divided into the following sections:

1. O_2 transport
2. Storage of O_2 (oxygen buffer)
3. Facilitation of intracellular O_2 diffusion
4. Biochemical catalysis
5. O_2 scavenging
6. Sensory function (and signal transduction)

The role of O_2 transport in vertebrates is undoubtedly taken by the tetrameric Hb as described above. Its cooperative ligand affinity is favorable for the optimal loading and unloading of O_2 in the lungs and tissues, respectively. Structurally different O_2 transporter Hbs were found in annelids (29, 30). They consist of standard Mb-like 16- to 17-kDa subunits that form multisubunit hexagonal bilayer structures (144 globins and 36 non-Hb linker proteins). Due to their huge size they are stable in an extracellular environment and do not require erythrocytes as it is the case for vertebrates. The most prominent globin belonging to the group of O_2 reservoirs is Mb. It is generally considered to augment O_2 diffusion in highly oxygen demanding and mitochondria-rich muscles, like the heart (31). Other studies complement the function of Mb with a contribution to the spontaneous diffusion of free dissolved O_2 between skeletal muscle cells by an additional flux of O_2 bound Mb (32, 33). The last three groups of physiological functions were mainly discovered in the last twenty years of globin research. Examples for globin mediated biochemical catalysis described so far are the oxidative phosphorylation in cardiac myocytes and the sulfide-reactive Hb from the symbiont-harboring gill of the bivalve mollusc (34). In the latter system hydrogen sulfide destabilizes the ferrous oxidation state of Hb by displacing superoxide in a nucleophilic process to form the ferric protein. Subsequently, the oxidized heme species facilitates binding of hydrogen sulfide. The reactivity of the oxygenated form of globins plays an important role in many other catalytic processes. Hydrogen peroxide, for example, gets reduced by ferrous or ferric Mb to H_2O and leaves the oxygenated ferryl (Fe^{+4}) form or even a cation radical species of the protein (35). Recently, the importance of signal molecules for physiological responses became apparent. These signal molecules often derive from the free-radical based oxidation of lipids which take a key role in the post translational modification of proteins (36). It was found that the ferryl heme iron of Mb and Hb is capable to initiate the oxidation of a wide variety of DNA (37), other proteins (38) and lipids (39). Intense work went into the investigation of oxidation reactions of gaseous NO and nitrite with oxyferrous Hb (40). Furthermore, it is generally accepted that Hb can function as NO transporter for delivery to areas of hypoxia or restricted blood flow (41). The O_2 scavenging function was found in Hb from legumes (leghaemoglobin) and Hb of nodulating nonlegumes (34). These proteins have extremely high O_2 affinities ($\approx$ 100 μM) in order to maintain the concentration of free O_2 at a minimum. A low level of O_2 is an essential precondition for the oxygen-labile nitrogenase to participate in N_2 fixation. FixL, which is a chimeric protein of the root nodule bacteria *R. meliloti* and consists of a heme domain and a kinase domain, binds O_2 reversibly with a very low affinity ($\approx$ 10 nM) (42, 43). Its heme domain senses O_2 and transduces the signal by controlling the phosphorylation of the transcriptional activator FixJ. FixJ, in turn, controls the expression of the symbiotic genes in the bacteria.

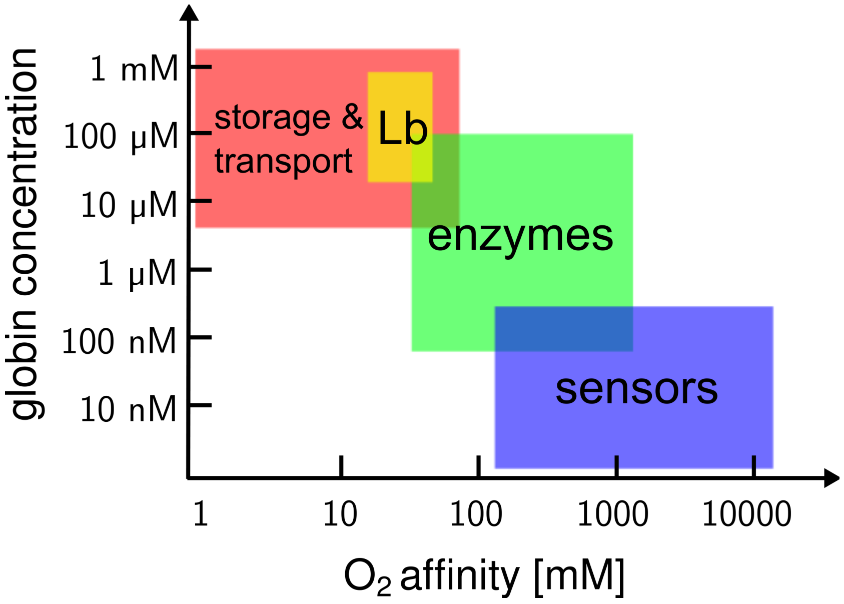

Figure 2.2.: The concentration in cell versus affinity for O_2 of globins performing different functions. Lb refers to leghaemoglobin. Adapted from Reference (14)

The large number of suggested physiological functions for the family of globin proteins is strongly based on the ligand affinity and the susceptibility to reduce the ligand. Generally it was found that the smaller the globin concentration in the cell is, the higher their affinity for ligands becomes. Figure 2.2 roughly clusters the discussed physiological functions of globins into this pattern. Due to the many new discovered functions of globins in recent years the structure-function relationship is getting more and more diverse and complex. Their nature and manifestation under physiological conditions is currently one of the main directions in the overall globin research (14). The interest on the functional mechanisms of globins is driven by medical research due to the many diseases in which globins were found to play a substantial role. The most carefully studied ones are sickle-cell (44) and Alzheimer diseases (45), malaria (46), kidney dysfunction (20), delayed vasospasm (47), fibrotic organ disorder (48) and cancer (49). In addition, it was found that the development of safe Hb-based blood substituents, which was a hopeless endeavor up to the present, heavily relies on the overall knowledge of the redox chemistry of Mb and Hb (50, 51).

2.2. Protein structural dynamics in Mb

The general importance of structural fluctuations to the designated function of biomolecules became evident after the determination of the first protein crystal structure which was sperm whale Mb (18). Its rigid protein conformation does not offer any pathway for a water molecule or a diatomic ligand to enter the distal site cavity from the bulk (52, 53). Subsequent studies characterizing xenon saturated X-ray structures of Mb revealed a network of at least four different docking sites (Xe1 - Xe4) coexisting inside the globular protein fold (54–56). Again, from a static point of view, these xenon accessible pockets do not show any kind of connecting channels between each other which would allow for xenon diffusion into more deeply buried regions of the protein. Deuterium exchange and later NMR experiments emphasized this inferences (57, 58). All these observations suggested quite clearly that Mb needs to undergo conformational transitions in order to allow for intake, migration and binding of O_2, CO, and other ligands. These conformational transitions are most likely of similar nature but less strongly pronounced compared to the T to R state transition in the cooperative O_2 binding in tetrameric Hb (59).

Proteins are structurally and energetically complex. Their primary sequence consists of strong covalent bonds. Hydrogen bonds, which are drastically weaker, and disulfur bonds connect different parts of the polypeptide chain to form a distinct tertiary structure of the protein. However, the weaker these interactions are, the higher is the probability for structural rearrangements to occur over time. Hydrogen bonds from the protein to solvent molecules and different protein side chains might get broken or newly formed in slightly different orientations and the rather flexible side chains often adopt different positions in different molecules. A typical protein sample used in spectroscopic and related experiments consists of an ensemble of $\approx 10^{15}$ protein molecules (60). Individual groups of molecules in this ensemble, clustered according to different overall structure conformations, are commonly called CSs. This feature makes proteins characteristically similar to other complex systems such as glasses, spin glasses (61, 62), or neural-networks (63, 64). The widely accepted opinion has emerged, that the complete set of available CSs of a given protein state perform the same biological function but differ in structural details and possibly in the rate at which the function is performed (65). Hence, for the interpretation of experimental data based on statistical protein ensembles, the classification and theoretical description of CSs is of great importance. Experimental techniques which tremendously helped to identify and characterize the nature and hierarchy of CSs in Mb and related globins are low-temperature flash photolysis coupled with IR, near-IR and Raman spectroscopy (66–69), X-ray diffraction (70, 71), time-dependent X-ray crystallography (72, 73), phonon assisted Mössbauer effect (74, 75) and Mössbauer spectroscopy (76, 77). Neglecting the question of availability for any of these techniques, the choice for a specific method is mainly depending on the time scale of the protein motion of interest. Since the last three decades several studies tried to get a unified picture of the qualitative and quantitative nature of CSs by combining the understandings obtained from the aforementioned experiments (78–83). A view on the structural dynamics in Mb and proteins in general has emerged in which the $3N - 6$ flexible coordinates (where N corresponds to the number of atoms) of the primary sequence are describing a multidimensional energy landscape. As shown in Figure 2.3 this energy landscape is organized hierarchically and is partitioned into at least three tiers of valleys (CS^0, CS^1, CS^2). Cryogenic experiments discovered even lower tiers. These valleys are connected to each other and the protein geometries can interchange by passing one or several transition barriers if needed. The highest tier in Figure 2.3 describes three specific CS (A_0, A_1 and A_3) identified from several studies of MbCO. These CS, more specifically called taxonomic substates, structurally differ in the sidechain orientation of E7-His (89) as shown in Figure 2.4, have different IR spectra (68, 90) and are also supposed to be associated to different biological functions (91). The relative population of every taxonomic substate can be regulated by environmental condi-

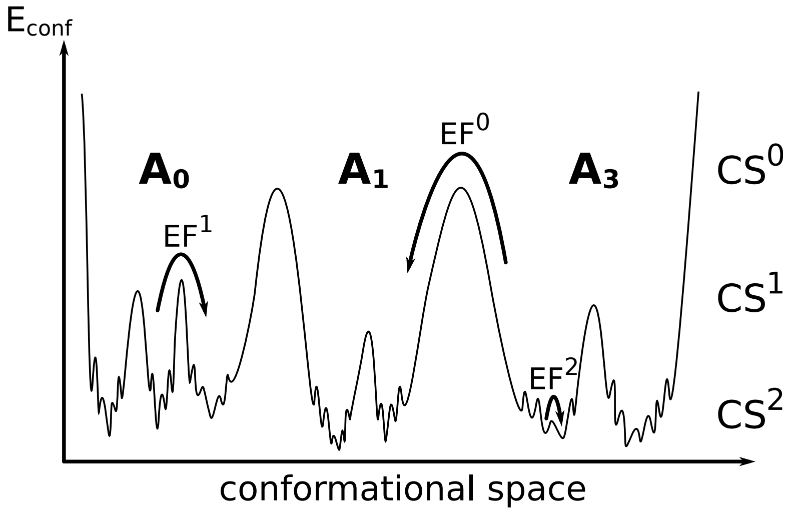

Figure 2.3.: One dimensional illustration describing the hierarchical picture of CS on the energy landscape of MbCO. The spectroscopically identified taxonomic substates A_0, A_1 and A_3 belong to the highest tier CS^0 which can interchange by equilibrium fluctuations (indicated as EF^0). Lower lying substates (CS^1 and CS^2) show minor structural and spectroscopical differences. Their interchange is faster compared to CS^0 due to the smaller barriers related to EF^1 and EF^2.

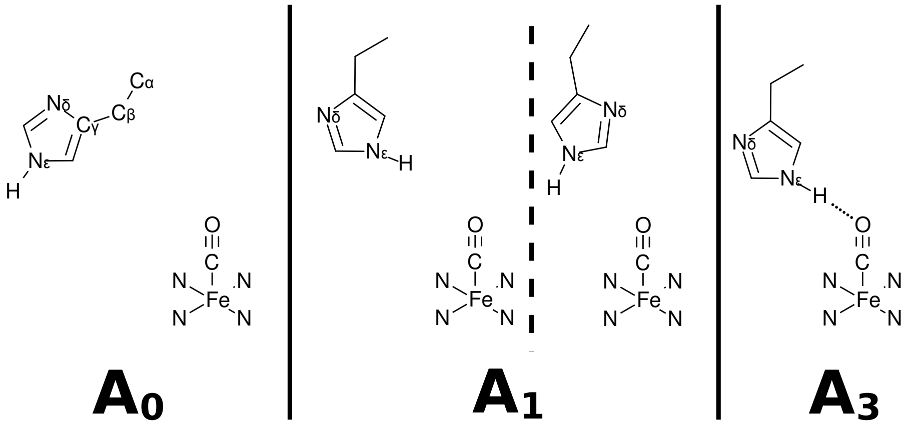

Figure 2.4.: Structural orientations of E7-His suggested for the spectroscopically identified CS A_0, A_1 and A_3 of MbCO. E7-His of the A_0 CS is swung-out from the distal protein cavity without any interaction to CO (84, 85). For the A_3 CS a consensus has emerged that E7-His describes a strong hydrogen bonding interaction between $N_\epsilon H$ and the oxygen atom of CO (85, 86). The E7-His orientation of A_1 is still ambiguous. Some studies suggest a similar arrangement as for A_3 with a less pronounced hydrogen bond (left) (85, 87, 88). Others assigned a ≈ 180° rotated orientation of E7-His along the C_β-C_γ axis for this CS (right) (86).

tions like pH (26, 27), temperature (92) , pressure (93), or activators like iodide, citrate and thiocyanate (28). The number of minima in the lower tiers is much larger and therefore statistically distributed rather than controlled by physical or chemical preferences (94).

More detailed insight into the transition barriers between the different tiers of CSs was gained from rebinding experiments of MbCO (68, 95) and MbO_2 (66). In these experiments the time course of the change of the CO IR and the O_2 Raman absorption frequency is followed after the dissociation of the heme-bound ligand using flash photolysis. The rebinding rate k_{BA} over a single barrier from a dissociated state B to the bound state A can be described by an Arrhenius relation of the form

$$k_{BA} = A_{BA}(T)\exp(-H_{BA}/RT) \tag{2.1}$$

where $A_{BA}(T)$ is the preexponential, T is the temperature, H_{BA} is the barrier height and R is the universal gas constant. Assuming that all protein molecules in the sample are located in the same CS, A_{BA} and H_{BA} are supposed to have unique values which makes the rebinding exponential in time. Thus the survival probability $N(t)$ in state B is exponential in time described by

$$N(t) = \exp(-k_{BA}t). \tag{2.2}$$

In contrast to the suggested rate-law, the experimental rebinding kinetics measured in the temperature range between 60 and 160 K are strictly nonexponential. Two sources for the nonexponential rebinding kinetics are possible. The first one assumes each protein molecule of the ensemble to be locked inside the same CS but rebind over individually large barriers (homogeneous rebinding). In the second case the protein ensemble is thermodynamically distributed over the energy landscape into different CSs all of which have an unique rebinding rate k_{BA} (inhomogeneous rebinding). Multiple-flash experiments using very short laser pulses at a repetition rate in the range of the average rebinding rate unambiguously identified the nonexponential rebinding behavior to derive from an inhomogeneous protein ensemble. At higher temperatures the protein is able to pass the barriers between different CS and can therefore fluctuate among all of them. Above the characteristic or 'glass' temperature T_c these fluctuations become fast compared to the ligand rebinding rate and the ligand rebinds from an average CS. Hence, the observed rebinding rate above T_c derives also from an average over the distributions of all H_{BA} in the protein sample and therefore follows an exponential-law. This behavior is termed “substate symmetry breaking” (66).

2.3. Ligand dissociation and migration dynamics in Mb

Prior to the ligand rebinding process, which is the actually measured magnitude in the photolysis experiments of MbCO and other hexacoordinated Mb species, the ligand first detaches from the heme prosthetic group, migrates between different ligand docking sites of the protein cavity-network and could under certain conditions also exit the protein completely (96). To complete the rebinding, the ligand needs to find its way back into the protein and subsequently into the heme site. Figure 2.5 shows a reaction scheme of the most important stages between ligand dissociation and rebinding. For a consistent picture of all process involved in the rebinding mechanism it is of great interest to know about the dynamical and structural changes taking place while switching between these different stages.

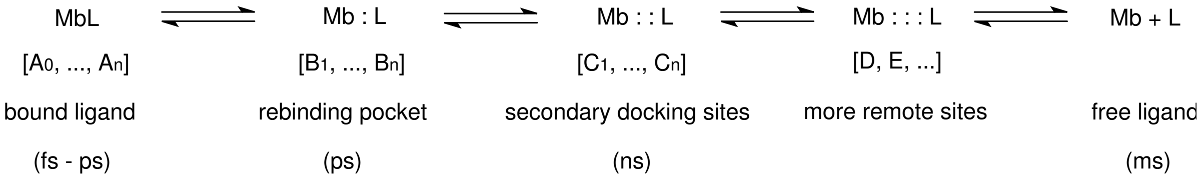

Figure 2.5.: Schematic representation of the different stages of ligand migration between photodissociation and exit from Mb into the solvent. Ligand L can capture different CS (labelled B_i and C_i) while migrating through the protein cavities in the style of the bound MbL states (A_i). The migration network is usually divided into the rebinding site which is close to the heme iron (B) and the more remote sites lying on the way to the protein exit (C, D, E, *etc.*). Typical time scales of ligand transitions for each state are indicated in braces.

The time scales of individual structural changes in each stage can roughly be ordered according to their displacement from the heme. Time-resolved resonance Raman spectroscopy (97, 98) and MD simulations (99, 100) both suggested that the iron atom moves out of its porphyrin in-plane orientation towards the proximal site 1 – 2 ps after photolysis. The Raman spectra additionally revealed an expansion of the heme core size within the same time window. These geometric changes correspond to the pentacoordinated, high-spin heme-state known from the deoxy Mb crystal structure (101, 102). Qualitatively, the amount of heme-doming varies strongly in the individual studies but generally it is found within a range of 0.1 to 0.4 Å. The dissociation of CO (103), NO (104) and other ligands from iron is completed on a time scale of tens to hundreds of picoseconds. X-ray structures of the initial photoproduct of MbCO recorded at 20 K show that the ligand experiences a recoil during this phase of transitional relaxation by capturing a position which is $\approx$ 2 Å away from the iron atom towards the protein interior (105). This photolyzed species belongs to the B_i set of CS in Figure 2.5. Typical crystal structures of this B_i set and IR spectra monitored after polarized excitation (106) both determined the CO to be aligned parallel to the heme plane. In the same time window time-resolved resonance Raman spectra detected an additional downshift of the iron to F8-His stretching band (97, 98). Transient absorption measurements of band III (amide N-H in-plane bending- and C-N stretching-vibrations), which are sensitive to local conformational changes near the heme iron, discovered three different relaxational contributions on time scales of tens, hundreds and thousands of picoseconds (107). The first one can clearly be attributed to the heme-doming motion while the second one fits into the next larger time window of relaxations. A combined interpretation of the structural changes occurring in the iron to F8-His stretching band, after heme-doming has completed, identified the additional bandshift as a structural transitions occurring on the protein surface (108). These changes are transmitted through the complete protein backbone and get detected by the heme through its coordination to the F8-His sidechain.

After heme and backbone of the photolyzed structure have relaxed into an equilibrated dx Mb species, the remaining fluctuations are related to the ligand migration dynamics. The knowledge of location, transfer kinetics and ligand population of the different docking sites remained sparse until the end of the twentieth century. Time resolved IR measurements of photolyzed CO in Mb at room temperature discovered bands appearing after $\approx 1 - 2$ ps and persisting until $\approx$ 200 ns (106, 109). These bands, related to the B_i set of CSs, correspond to different stretch-frequencies of dissociated CO. The observed frequency shifts are a result of the Stark effect (110). It is caused by changes of the electrostatic interaction of the free ligand with the electric field which is induced by the protein environment and occurs when the ligand or polar amino acid residues reorient inside the protein cavity. Many of the earlier studies were focusing on the interpretation of the rebinding kinetics in point-mutated MbCO (27, 100, 111, 112). On the one hand, it was found that an increased amount of ns recombinations take place if the interior volume of the distal pocket is decreased by mutating Val68 to phenylalanine. In this way the ligand molecules are forced to stay inside the primary pocket B. On the other hand, the nanosecond rebinding can be decreased by sterically hindering access to the iron atom by a mutation of Leu29 to phenylalanine. Ligand rebinding in this mutant is mainly observed on the picosecond time scale due to the increased sidechain of residue 29 which forces the ligand inside docking site B closer to the iron atom. Geminate rebinding on time scales > 100 ps is significantly reduced as Phe29 acts as a kinetic barrier for back migration into site B as soon as the ligand reached more remote cavities. The understanding of more remote docking sites (C, D, *etc.*), estimations of migration barriers from and to different docking sites and rebinding barriers to the hexacoordinated state was obtained from temperature-derivative spectroscopy (113–115) and MD simulations (116–120). The conclusions in all of these studies are diverse but the bottom line can be summarized as follows. Rebinding barriers do largely differ between different taxonomic substates. Migrations between different docking sites occur through the thermal motion of the protein and is generally faster than the final step of rebinding or exit to the bulk solvent. Their transfer barriers and the rebinding barrier from the primary pocket B to the hexacoordinated state is depending on the size of the ligand. Additionally, the accessibility and population of specific docking sites is related to the specific composition of residues shaping up the docking sites rather than the species of the ligand.

Starting with the new millennium a remarkable progress was made in this field of research. By trapping intermediate protein conformations at specific timesteps after photolysis and examining their structure using time-resolved X-ray crystallography (72, 73, 121, 122) it became possible to structurally visualize the pathways of ligand migration and relaxation. The crystal structure recorded after 100 ps of photodissociation agrees with earlier low-temperature structures of photolyzed MbCO (102) which identified the CO ligand ≈ 2 Å above the heme in the distal site B. Within ≈ 3 ns CO was found in two different secondary docking sites C. The first one corresponds to the well-known Xe4 site and the other one is located ≈ 6 Å perpendicularly above the heme plane. Both pockets are populated for only tens of picoseconds. Subsequently, CO migrates from Xe4 to the more remote docking sites Xe1 and Xe2 where the ligand remained trapped for microseconds. In addition to the CO dissociation and the heme-doming, the time-dependent X-ray structures resolved other intermediate dynamical features. These involve tilting and translation of the heme, the motion of the distal site residue 29 towards the vacant sixth-coordinate position and displacements in the E-helix and the CD-turn (121).

As a final remark to this chapter it is pointed out that, besides the definite assigned functional role of globins as O_2 transporter and storage, an increasing number of studies started to draw picture of catalytic activity (123–125). In this view, the globular shaped Mb can be regarded as a chemical reactor or a container (91). The iron represents the active center, the heme surrounding cavities are active chambers and the heme itself acts as a cap. Depending on the environmental conditions channels can gradually be opened and closed to allow possible ligands and reactants to enter the container, like through pores. The affinity for specific ligands to enter the protein and migrate into its active center is controlled by the composition of the amino acid sequence along the channel and the constitution of this channel which can be assigned to a certain taxonomic substate. By offering a large affinity for a certain ligand the biological system can maintain a larger concentration of the ligand under the given physiological conditions as it would be possible in the absence of the protein. If the organism demands for a release of the ligand at a certain point, a physical or chemical modification of the protein environment is provoked. The difference of environmental conditions, which acts as an allosteric effector on the protein, redistributes the preferences for the taxonomic substates to one with a reduced ligand affinity and a cavity network with very small migration barriers towards the protein exterior. A probable reason for the evolutionary formation of docking sites in Mb is the functional need to accumulate a second reactant inside these reaction chambers, besides the iron-bound one. In this way the foundation for possible enzymatic reactions in the protein active site is given. Due to the characteristic shape of the cavities the reactants are forced into a specific orientation which speeds up a bimolecular reaction between the two reactants. Formation of the product(s) is expected to induce relaxations of the tertiary protein structure into different CSs having open channels to the protein exterior to simplify the release of the products (94). In case the heme iron became oxidized during this catalytic reaction, it needs to get reduced again to the ferrous state to finalize the catalytic cycle.

3. Structure and Function of Neuroglobin

"City Honours! guard us, ye Pow'rs! pray observe the intoxicating Degrees of City Preferment;– to be Under Scavenger, and shovel up the Street Dirt; Head Church-Warden, to out-snore the whole Parish, and pocket the Poor's Money; but to have the Watchman man bawl out."

Thomas Baker, *Hampstead Heath, 1706*

In contrast to Mb and Hb, which are the most extensively characterized globin proteins, Ngb and cytoglobin (Cygb) were discovered not before the beginning of the new millennium (126). The lower amount of and more localized presence of these proteins in neuronal cells of vertebrates lead to the speculation that they must perform very specific physiological functions. Over the last couple of years, different studies tried to unravel Ngbs functional essence and suggested a manifold of different physiological roles for this protein. Besides the search for a physiological function of Ngb, a lot of effort went into the structural interpretation of Ngb and Cygb. Both proteins possess structural features which sets them apart from their Mb and Hb relatives. The first section of this chapter gives an overview of proposed physiological functions of Ngb. Subsequent sections summarizes the identified structural and dynamical peculiarities of the protein and show how they might be related to the functional role of Ngb.

3.1. Proposed functional significances

In the genetic evolution of the large family of vertebrate globins Ngb diverged around 550 million years ago from the phylogenetic branch containing Mb and Hb (127). Within the set of discovered mammalian Ngb-sequences a point-mutation frequency of $\approx 0.4 \times 10^{-9}$ amino acid changes/year was observed (128). This value is three times smaller compared to Hb and corresponds to 70 conserved amino acid residues among all discovered vertebrate Ngb sequences and indicates a highly specific and structurally tightly associated physiological function of this protein. Although Ngb is probably the best studied globin of all newly discovered species in this family of proteins, its exact physiological function remains uncertain. Suggested functional roles and their rationale can be divided into the following groups:

- O_2 was found to bind reversibly to Ngb with an affinity which is within the range of Mb (126, 129). It combines with the ligand under a high O_2 half saturation pressure and releases it again when the pressure is lowered. Ngb resides in metabolically active neuronal cells and subcellular compartments and was suggested to enhanced O_2 supply in these locations (130).

- Cell cultures exposed to hypoxic conditions were found to moderately induce Ngb expression (131). Additionally, goldfish, an animal that survives extended periods of reduced O_2 availability, maintains an approximately fivefold higher Ngb concentration than the less hypoxia-tolerant zebrafish (132). Given these facts, it was proposed that Ngb is involved in the regulation of hypoxia.

- In analogy to Mb, Ngb was found to protect neurons from reactive oxygen or nitrogen species like NO and H_2O_2 *in vitro* (8, 133). It was suggested that depending on the oxygen partial pressure, Ngb may either decompose or produce NO. This mechanism is important to the control of vaso-constriction or -relaxation and the level of mitochondrial respiration (134). Under a reduced O_2 flow reactive oxygen species are known to be formed in neuronal cells (135). This observation is in line with a neuroprotective effect of Ngb discovered after ischemia and reperfusion (136).

- Several studies considered Ngb to interact with other proteins. An observed binding of Ngb to the α subunit of heterotrimeric G proteins proposed an inhibition of the dissociation of guanosine diphosphate from Gα (137). This fact qualifies Ngb to protect the cell from apoptosis. Recent studies suggested that

> Ngbs neuroprotective effect is the result of the reduction of ferric cytochrome *c* by ferrous Ngb, thereby preventing cytochrome *c*-induced apoptosis (45, 138).

Nevertheless, for most of the proposed physiological functions contradictory studies were published. Some anomalies in these experiments could be addressed to a contamination through Ngb autooxidation (139). Theoretical approaches contain an integrated control of these conditions and can therefore be regarded as a helpful tool to fortify the experimental results.

3.2. Structural features

A comparison of the primary structures of Ngb with Mb and Hb show that Ngb shares less than 25% of conserved residues with Mb and Hb. Among the conserved amino acids B10-Phe, CD1-Phe, CD3-Tyr, E11-Val, F4-Leu, FG1-Val, FG3-Val and G5-Phe are located in the heme binding or ligand interaction sites (140, 141). In spite of the low sequence homology between Mb and Ngb, the tertiary structure of both proteins display the same three over three α-helical sandwich fold. Crystal structures of sperm whale Mb, metNgb, and NgbCO aligned to the C_α backbone atoms are shown in Figure 3.1. The average RMSD for the C_α atoms is 2.0 Å between Mb and metNgb and 1.85 Å between Mb and NgbCO. Nevertheless, the X-ray structure of metNgb has a particular feature setting it apart from Mb. While the heme iron atom in Mb, in the absence of an exogenous ligand, is pentacoordinated to the four porphyrin nitrogen atoms and to the F8-His sidechain on the proximal side of the protein the same position is additionally bound to the distal E7-His sidechain (His64) in metNgb resulting in a hexacoordinated configuration (142).

If exogenous ligands are present in the active site of Ngb the distal histidine coordination is replaced by the ligand. The comparison of NgbCO with metNgb in Figure 3.1 reveals major rearrangements between the two configurations which are mainly located around the heme group (143). Upon disruption of the bis-His coordination metNgb seems to release a positional constraint of the heme and allows it to slide into the proximal site and to tilt towards the distal site. The average RMSD of all heme atoms between metNgb and NgbCO is 4.5 Å whereas the corresponding one between NgbCO and Mb is lowered to 3.2 Å. Hence, the heme group of NgbCO captures a more Mb like position after breakage of the distal bis-His bond. Another

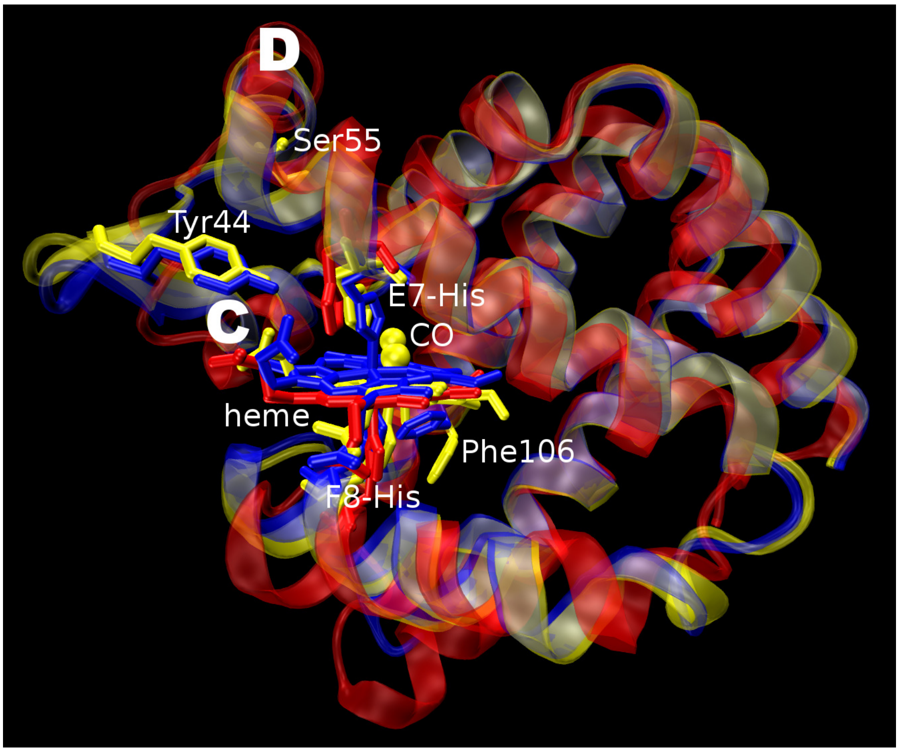

Figure 3.1.: Representations of the C_α-aligned crystal structures of sperm whale Mb (red, PDB code 1VXF), metNgb (blue, PDB code 1Q1F) and NgbCO (yellow, PDB code 1W92). Helices C and D and residues Tyr44, Ser55 and Phe106 mentioned in the text are indicated. The heme group of metNgb is hexacoordinated to E7-His (His64). In the presence of an exogenous ligand (CO) the heme plane reorients towards the proximal site.

distinctive difference observed between the metNgb and NgbCO crystal structures is a reshaping of the large internal cavity network (143). The total volume of this network increases from ≈ 287 Å^3 in metNgb to ≈ 306 Å^3 in NgbCO. While for metNgb distal and proximal sites are connected with each other through the network, the reoriented heme plane of NgbCO and its Phe106 sidechain located at the proximal site leave only a small isolated cavity in this position which corresponds to the Xe1 site in Mb. On the distal site the cavity branch is larger in NgbCO compared to metNgb and features a unique 13.3 Å^3 volume near Ser55 which is located in the D-helix. Tyr44, belonging to the CD-loop, forms a hydrogen bond with one of the heme propionates. The same hydrogen bond is less strongly pronounced in the X-ray structure of NgbCO. The weakening of the Tyr44 to heme interaction is assumed to be transmitted into the CD-loop region inducing a formation of a cavity near Ser55 which was found in the crystal structure of NgbCO (141, 144).

3.3. Dynamical features

Further evidence to the picture of a bis-His coordination in metNgb which was derived from the static crystal structures was found in ligand rebinding studies of photodissociated NgbCO and NgbO$_2$ (129, 145, 146). Contrary to equivalent experiments in MbCO and MbO$_2$ the geminate ligand rebinding process in NgbCO exhibits a minimum two-step process of the form

$$\mathrm{Ngb_H} + \mathrm{L} \rightleftharpoons \mathrm{Ngb_P} + \mathrm{His64} + \mathrm{L} \rightleftharpoons \mathrm{Ngb_L} + \mathrm{His64} \qquad (3.1)$$

where Ngb$_\mathrm{H}$ represents the bis-His hexacoordinated Ngb species, Ngb$_\mathrm{P}$ the pentacoordinated one and L the exogenous ligand which competes for the hexacoordination with the internal His64 sidechain (E7-His). Ngb$_\mathrm{L}$ corresponds to the L coordinated Ngb configuration. A structural representation of the possible transitions is shown in Figure 3.2. A physiological meaning of the bi-phasic rebinding mechanism was suggested for the process of O$_2$ and NO sensing in nervous tissue (133, 147). The functional importance of the stepwise binding procedure in heme proteins gained further support through the discovery of similar processes taking place in other globins (140, 148–151) and even cytochromes (152, 153).

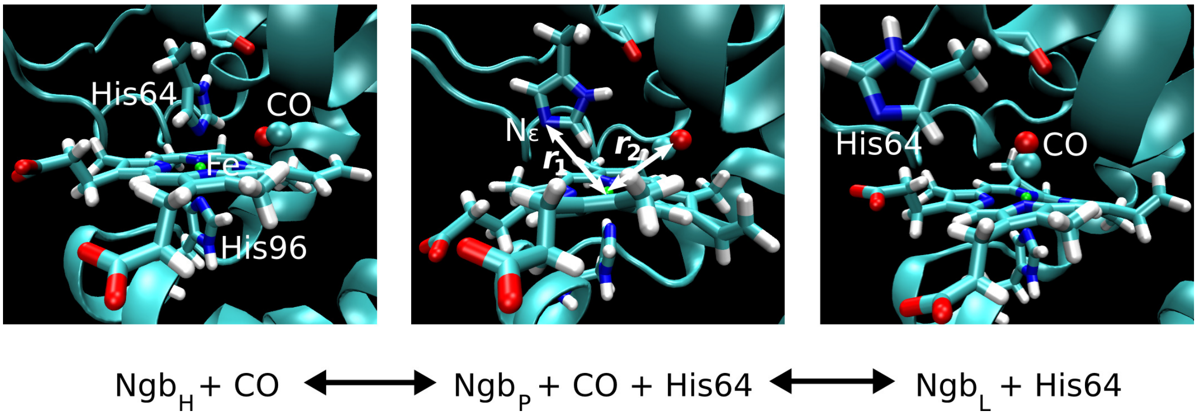

Figure 3.2.: Representation of the experimentally suggested two-step rebinding competition between the internal His64 ligand and an exogenous CO ligand (L). r_1 and r_2 correspond to the reaction coordinates used to calculate the free energy profiles for the individual rebindings $\mathrm{Ngb_H}$ + CO $\rightleftharpoons$ $\mathrm{Ngb_P}$ + His64 + CO and the $\mathrm{Ngb_P}$ + His64 + CO $\rightleftharpoons$ $\mathrm{Ngb_L}$ + His64 discussed in section 5.5.

In wt Ngb residue 55 is consistently represented by a cysteine which, in the X-ray structures, is generally mutated to serine for crystallization reasons. The crystal structure of hNgb (PDB code 1OJ6) has an additional mutation in position 46 where cysteine is replaced by glycine (140). Cysteine residues are rather scarce in globin sequences and they occur at well defined positions, suggesting a specific functional role. Structurally, cysteines can form intra- or intermolecular disulfide bonds or are a place of direct ligand binding. Cys46 and Cys55 in wt hNgb are both situated in the CD-loop region and form an internal disulfide bond under oxidative conditions (154, 155). These residues are highlighted in the MD equilibrated structure of wt $\mathrm{hNgb_P}$ shown in Figure 3.3. Flash photolysis and stopped flow experiments of $\mathrm{NgbO_2}$ and NgbCO for disulfide bound hNgb and disulfide-less wt mNgb show a reduction of the $\mathrm{O_2}$/CO rebinding affinity by a factor larger than two for mNgb (129). The same experiments for disulfide bonded, reduced (without disulfide bond) hNgb and CCC → GSS mutated (no disulfide bond possible) hNgb follow similar trends (154, 156). On the other hand, by reducing the disulfide bond in hNgb it was found that His64 rebinding is slowed down corresponding to an increase of the activation energy for His64 binding (154, 157). From these results it seems that the presence or absence of the disulfide bond plays an important role in the internal versus external ligand rebinding competition of $\mathrm{Ngb_P}$ and therefore also for the physiological function of the protein. An investigation on the structural and dynamical differences between the dissociation of His64 with and without the Cys46-Cys55 bond is part of chapter 5.

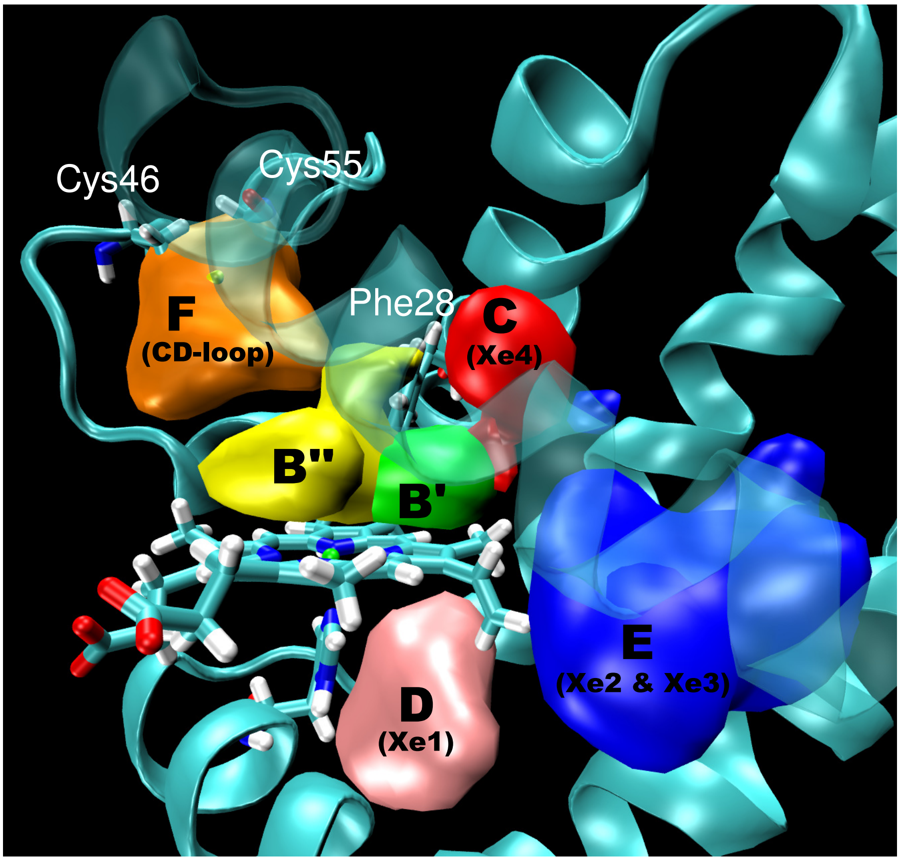

Figure 3.3.: Structure of oxidated wt hNgb featuring a disulfide bond between residues Cys46 and Cys55. Individual docking site cavities which are analyzed in section 5.3 are represented as color-coded volumes.

The Ngb specific ligand migration network was investigated for CO using IR spectroscopy (158, 159) and theoretically for CO, O_2 and NO ligands (160) including the published manuscript enclosed in section 6.1. All of these studies suggest that Ngb describes a ligand migration network which is similar to the one of Mb (161). Regions of individual docking sites which were specifically identified for Ngb in this thesis are shown in Figure 3.3. MD and metadynamics simulations could identify all characteristic Xe pockets known from Mb (54–56). Additionally, the simulations identified populated cavities present in the distal pocket (B′ & B″) and in the region of the CD-loop (F) connecting the former one to the bulk solvent. Experimental as well as theoretical studies found the migration barriers between the sites to be lower compared to Mb (158, 159, 162). IR spectroscopy differentiate docking sites mainly through band shifts induced by the Stark effect. Measuring the ligand rebinding times from these sites gives qualitative information on the distance of each site relative to the rebinding site. More precise information about the locations of ligand occupied cavities under experimental conditions are hard to obtain. In the publication included in section 6.2 a MD approach applying a multipolar force field (discussed in section 4.3) is used to characterize CO docking sites in Ngb. In this work it was successfully shown that experimental IR band shifts can be assigned to specific docking sites and also to specific orientations of CO inside every single pocket.

Taxonomic substates of NgbCO were experimentally analyzed and compared with the previously identified CSs of MbCO by IR spectroscopy (155, 163, 164). They found very similar IR shifts in relation to the A_3 and A_0 bands of MbCO and therefore were called N_3 and N_0. Nevertheless, the bands of NgbCO were found to be significantly broader than in MbCO suggesting a stronger structural heterogeneity. Furthermore, 2D-IR vibrational spectroscopy was able to divide N_0 into two individual sub-CSs showing vibrational relaxations on two different time scales (155). The third manuscript, which is included in section 6.3, tries to structurally and dynamically map out the different CS of NgbCO using a combination of classical MD simulations and QM/MM calculations.

4. Theoretical Background

"They Sir are our Brethren, our intimate Bosome friends. Your Quaker is the Anvil on which we forge our designs, and the Phanatick the person that uses them. They are both the necessary Tools we work with, and indeed the chief Implements of our Trade."

Thomas D'Urfey, *The Royalist, 1682*

The structural and energetical investigation on Ngb performed within this work relies on computational techniques of different levels of theory and accuracy. In the following, the principles of MD simulations are shortly described. Subsequently, specific MD extensions and additional quantum mechanical treatment which help to increase the qualitative and quantitative precision are presented. The last two sections discuss the general description of ARMD used to study the configurational transitions of the heme prosthetic group in Ngb and the detailed force field parametrization of these configurations.

4.1. Molecular dynamics simulations

4.1.1. Historical notes

MD represents a category of computer simulations where a defined amount of particles like atoms or molecules interact with each other over a specific period of time. The striking idea of MD simulations is to give scientists the capability to peer into the motion of individual atoms in a more illustrative way which is hardly possible in a laboratory experiment. First applications of MD simulations started to appear in the late 1950s by Alder and Wainwright who investigated the phase transition of a hard sphere system (165, 166). More interesting applications in a chemical point of view followed by Rahman and Verlet who studied several properties of liquid argon using LJ potentials (167) and a time integration algorithm (168) which both are in use up to the present. Initial MD simulations of water using more realistic potentials started to appear in 1974 by Rahman and Stillinger (169, 170).

Consecutively, the complexity of the MD studied systems evolved in a fast-paced manner. The first all-atom protein simulation was performed in 1977 consisting of the bovine pancreatic trypsin, a globular polypeptide containing 58 amino acid residues (171). Intensive MD studies over years made this protein one of the most thoroughly characterized proteins in terms of structure and folding pathways. An initial MD simulation of Mb was performed by Karplus in 1985 (172). It was found that the long time motions of proteins involve multiple distinct minima as it was experimentally suggested in terms of CSs. Its trajectory was ten times longer (300 ps) than any other protein simulation of similar size at that time. Today, MD simulations take on an important role in biochemical research and material sciences. Recent examples, which marked new milestones in terms of system size, include a 50 ns simulation of the complete satellite tobacco mosaic virus consisting of up to one million atoms (173) and the simulation of nano-twinned copper crystals containing up to 140 million atoms (174).

4.1.2. Basic theory

The basis of every common MD simulation is formed of the Born-Oppenheimer approximation which assumes that the rapid motion of electrons averages out and all atoms can be treated classically (175). The total energy of a molecular system of n atoms is described by the Hamiltonian H:

$$H(x_1, ..., x_n; p_1, ..., p_n) = \sum_{i=1}^{n} T(p_i) + V(x_1, ..., x_n) \tag{4.1}$$

where $T(p_i)$ corresponds to the kinetic energy defined by the momenta p_i of atom i and $V(x_1, ..., x_n)$ to the total potential energy of the system depending on the sum of atom coordinates x_1 to x_n. For molecular systems a popular strategy is to divide

$V(x_1, ..., x_n)$ in terms of bonded (V_b) and nonbonded (V_{nb}) interactions:

$$V(x_1, ..., x_n) = \sum_{i=1}^{n} V(r_b, r_{nb}, \theta, \chi, \phi) = \sum_{i=1}^{n} V_b(r_b, \theta, \chi, \phi) + \sum_{i=1}^{n} V_{nb}(r_{nb}) \tag{4.2}$$

where the arguments are defined by internal coordinates. Bond distances are expressed by r_b, valence angles by θ, torsional angles by χ, out-of-plane angles by ϕ and distances between nonbonded particles by r_{nb}. In the context of MM and MD simulations $V(r_b, r_{nb}, \theta, \chi, \phi)$ combined with an additional set of parameters is generally referred to as a force field. Around a dozen different classical force fields exist such as CHARMM (176), AMBER (177), GROMOS (178), and OPLS (179). The detailed description of $V(r_b, r_{nb}, \theta, \chi, \phi)$ looks basically similar in all these force fields. Intramolecular interactions which are part of V_b describe the strain energy in the molecular structure compared to a relaxed equilibrium structure in terms of harmonic and periodic functions. The potential V_{bond}, reflecting a covalent bond i between two atoms, is commonly given by:

$$V_{bond} = \frac{1}{2} k_{b,i} (r_i - r_{0,i})^2 \tag{4.3}$$

where $k_{b,i}$ represents the force constant, r_i the actual bond distance and $r_{0,i}$ the equilibrium bond-length under relaxed conditions. The harmonic form of the potential characterizes the strain occurring through the separation of the bonded atoms in a simplified manner. Nevertheless, V_{bond} reproduces the bond stretching vibrations sufficiently accurate around the equilibrium. The potential energy restraints of valence and out-of-plane angles are described in a similar way as V_{bond}:

$$V_{bend} = \frac{1}{2} k_{\theta,i} (\theta_i - \theta_{0,i})^2 \tag{4.4}$$

$$V_{improper} = \frac{1}{2} k_{\phi,i} (\phi_i - \phi_{0,i})^2 \tag{4.5}$$

where $k_{\theta,i}$ and $k_{\phi,i}$ are the corresponding force constants, $\theta_{0,i}$ and $\phi_{0,i}$ the equilibrium geometries and θ_i and ϕ_i the internal coordinates for the current atom positions. Torsional modes are expressed through a periodical cosine function depending on the torsional angle χ_i with periodicity n and phase shift δ multiplied by an independent force constant $k_{\chi,i}$

$$V_{torsion} = k_{\chi,i} (1 - \cos(n\chi_i - \delta)). \tag{4.6}$$

The non-bond term V_{nb} of the force field, dealing with the intermolecular interaction between atom pairs separated by three or more bonds can be further divided into Coulomb and LJ terms:

$$V_{nb} = \frac{q_i q_j e^2}{r_{ij}} + 4\epsilon_{ij}\left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 2\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right) \tag{4.7}$$

where q_i and q_j define the partial charges of atoms i and j interacting with each other, e is the elementary charge, r_{ij} the distance between i and j, ϵ_{ij} the depth of the potential well and σ_{ij} the van-der-Waals radius of the atom-pair. As every MD simulation performed in the present work is using the CHARMM force field (176), Equations 4.3 to 4.7 refer to the exact expressions therein.

In the microcanonical ensemble, the MD simulation describes the movement of the atoms in a molecular system during a definite simulation time t_{tot} for a fixed number of atoms n, a constant volume V and a constant energy. The obtained trajectory may be seen as the time evolution of exchanged potential and kinetic energies described in Equation 4.1. A general MD simulation algorithm proceeds by iteratively calculating the forces of $V(x_i)$ acting on an atom i and setting them into relation with Newtons second law of motion:

$$F_i = -\nabla V_i = m_i \dot{v}_i(t) \tag{4.8}$$

where m_i is the mass and $v_i(t)$ is the velocity at time t defined as

$$v_i(t) = \dot{x}_i(t). \tag{4.9}$$

New coordinates at time $t + \Delta t$ are calculated based on the accelerations obtained from the new forces over

$$x_i(t + \Delta t) = x_i(t) + v_i(\Delta t) + \frac{1}{2}\dot{v}\Delta t^2. \tag{4.10}$$

One cycle of iterations is completed by setting $t = t + \Delta t$ and the simulation is terminated when the condition $t = t_{tot}$ has reached. In practice, nearly every MD algorithm, like Verlet (168) or Leap-frog (180), include a predictor and a corrector step in solving the equations of motion. Steps like scaling of temperature and pressure are included additionally to simulate the more meaningful canonical and isothermal-isobaric ensembles from an experimental point of view.

It becomes clear that the functional forms of Equations 4.3 to 4.7 represent a certain reduction of the complexity of the molecular energetics. Nevertheless, they are a good compromise between accuracy and speed. Coarse-graining force fields group multiple atoms of the system into a so-called "pseudo-atom" to reduce the number of degrees of freedom and therefore can reach longer simulation time (181). This approach shows good results in protein folding simulations where fast and strongly localized normal-modes are of minor importance to the problem of interest. Methods and extensions which lead to higher accuracy or overcome certain limitations of the general all-atom based force field exist in a large variety. A selection of such implementations used in the present work are introduced in the following sections.

4.2. Combined molecular/quantum methods

Quantum chemical methods give much more accurate results compared to classical mechanics in almost every aspect. Furthermore, they are appropriate for the description of processes involving bond forming and breaking which is not the case for common force fields used in MD simulations. However, the purely quantum mechanical approach is generally limited to systems of small size and low complexity. The region of space in which chemical reactions, and therefore changes in electronic structure, occur in an enzyme is often relatively small compared to the whole system. This aspect is considered in combined quantum mechanical and molecular mechanical QM/MM models, where a specific part of the system is treated with quantum chemical methods and the surrounding environment is described by a classical MM force field (182–184). The effective Hamiltonian H for the complete system can be written as

$$H = H_{QM} + H_{MM} + H_{QM/MM} \tag{4.11}$$

where H_{QM} represents the interaction energy between all quantum mechanical particles, H_{MM} corresponds to the interaction energy of all classical particles with each other and $H_{QM/MM}$ accounts for the interaction between atoms of the quantum region with the particles of the classical part.

In principle, it is possible to interface any *ab initio* quantum chemistry method, like

Hartree-Fock (HF), density functional theory (DFT) and Post-Hartree-Fock methods (i.e. Møller-Plesset second order), into a MD simulation. Many QM/MM studies focusing on the ligand transfer from the six to the five coordinated heme states in Mb or Hb define large QM regions and make use of high level DFT functionals like B3LYP and substantial basis sets (i.e. 6-31G*) in order to obtain qualitatively and also quantitatively meaningful results (185–187). As these methods are computationally still very demanding and therefore in their application restricted to short simulation times, they are often substituted by semi-empirical quantum chemistry methods. These approaches are based on the HF formalism but make certain approximations and include parameters from empirical data. AM1 (188) and PM3 (189, 190) and more recently RM1 (191) and PM6 (192) are popular semi-empirical methods used in hybrid QM/MM simulations. Another semi-empirical method which became increasingly popular is the self-consistent charge density functional tight-binding (SCC-DFTB) scheme (193). This DFT based approach is widely used as a method of choice in hybrid QM/MM with classical MM force fields in simulations of biological systems like carbonic anhydrase (194) or alcohol dehydrogenase (195).

Car-Parrinello MD (196) is an *ab initio* method which, in contrast to Born-Oppenheimer MD, explicitly introduces electronic degrees of freedom in addition to the nuclear ones. This leads to a system of coupled equations of motion for nuclei and electrons. Therefore, the explicit electronic minimization needs to be run only in the initial step. In any subsequent MD step fictitious dynamics take care of the electrons which keep them on the electronic ground state. In combination with a specialized QM/MM scheme (197) Rovira et al. successfully characterized different CS of MbCO and MbO_2 using Car-Parrinello MD (87, 198).

A method which stands out of the previously discussed hybrid quantal/classical simulations is the empirical valence bond (EVB) method (199, 200). Contrary to any *ab initio* or semi-empirical approach EVB is based on the valence bond theory rather than the molecular orbital theory. Although the general valence bond theory was abandoned in the chemical community since the 1950s, its EVB successor has been further developed and is still used regularly. The basic idea of EVB is to represent the reacting system as a superposition of ionic and covalent resonance forms. A chemical change in the reaction environment is achieved by altering the electrostatic interaction of the ionic resonance forms with its surroundings. EVB has its main field of application in the investigation of enzymatic reactions (200, 201) and proton transfers (202, 203).

Despite the fact that QM/MM methods can simulate bond-formations and breakages in enzymatic systems, their application to condensed phase matter has other benefits which strongly outperform those of pure classical MD approaches. MM force fields are primarily optimized for structural alignment to X-ray and NMR data. Their energetical aspects are usually treated with minor importance. Vibrational frequencies, for example, are only fitted approximately around the equilibrium coordinate of every normal-mode and would deviate from experimental data when leaving this region due to their harmonic representation in the force field. Furthermore, electrostatic interactions in standard classical force fields are only taken care of between point charges. Higher multipole moments describing the structural dependency to the electrostatics of the system in much more detail are not considered at all. An accurate description of vibrational frequencies in the electrostatic field of the protein active site caused by the condensed phase environment is heavily depending on this high-level treatment of the electrostatic interactions. The present study on the structural identification of CSs in NgbCO, included in section 6.3, combines the speed of structural protein sampling, using classical MD simulations, with the energetical accuracy of the DFT method in a step-wise manner. With this approach, the conformational space of the complete protein system gets initially sampled with classical MD for a long enough time to sample a statistically reasonably large distribution of CSs. In the second step, a definite number of snapshots are consistently collected from this MD ensemble. Atoms belonging to a selected QM region are preserved in each snapshot while all remaining atoms representing the classical region are transfered into a point charge representation. The energetics for each of these modified snapshot are then calculated using a standard DFT method incorporating the point charges of the protein and solvent environment to describe the electrostatic field in the background of the QM region (204, 205). In addition to the present NgbCO study the described QM/MM(MD) approach was successfully applied to the structural interpretation of CS in CO bound (88) and CO dissociated MbCO (206).

4.3. Polarizable and multipolar force fields

4.3.1. Overview

The inclusion of polarizability is probably one of the most missed features in standard classical MD force fields. The simulation of increasingly complex biological mechanisms including electron- and proton-transfer reactions and also condensed phase spectroscopy strongly demand for an accurate treatment of polarizability. Due to this, many efforts were put into force field improvement to account for the lack of this feature (207). Current polarizable force fields can roughly be divided into five different groups:

1. X-Pol (208) is a force field which treats the polarizability based on electronic structural theory. It divides the condense-phase system (e.g. the protein) into different pieces or fragments. The molecular wave function of the entire system corresponds to the product of the wave functions of the individual blocks. Hence, the method naturally accounts for the mutual polarization in the system.

2. ENZYMIX (182), which was the first polarizable force field, and PIPF (209) are two independent implementations which both introduce induced dipoles on top of the point-charge interactions.

3. The most simple description of polarizability into a standard force field is introduced by simple fluctuating charges. CHARMM includes this kind of extension through the efforts of Patel and Brooks III (210).

4. The Gaussian electrostatic model is based on density fitting and describes a continuous electrostatic model going from distributed multipoles to densities (211). It allows for an inclusion of short-range quantum effects into the MM energies.

5. Force fields based on distributed multipoles probably gained the strongest attention in recent years. SIBFA (211) is a force field developed for an accurate treatment of transition metals by including ligand field contributions. This allows for an accurate computation of "open-shell" metalloproteins. The AMOEBA force field (212, 213) divides the electrostatic description into permanent and induced multipoles. Its strength over standard force fields was demonstrated in the evaluation of potential energy landscapes of small peptides, the calculation of free energies of hydration, condensed-phase dynamics,

protein-ligand binding and X-ray crystallography refinements (211). ORIENT, developed by Stone and co-workers (214) can describe the electrostatic interactions from simple point charges up to multipoles of rank five. The multipole (MTP) module which is used in the present Ngb studies of dissociated CO (section 6) derives from the Distributed Multipole Analysis (DMA) (215) method and treats multipolar electrostatics up to octopoles. A more detailed description of the method is given in the next section.

4.3.2. The MTP force field

The foundation of the MTP force field lies in the accurateness for the description of electron densities $\rho(\vec{r})$ in molecules obtained from electronic structure calculations. First population analysis methods, capable of separating $\rho(\vec{r})$ into atomic electron populations, only considered the charge population of spatial volume elements or molecular orbitals (216). This information covers only point-charge descriptions corresponding to the first term in a multipole expansion or the Coulomb term in standard classical force fields (see Equation 4.7) (217, 218). A quantitative reproduction of molecular multipole moments additionally requires the inclusion of higher multipole moments into the population analysis (217). For MTP the molecular multipole moments are determined using DMA. It includes a partitioning of the electron density in molecular orbital space and an integration of $\rho(\vec{r})$ over an appropriately defined space for all desired multipole ranks.

The current implementation of the MTP incorporates interactions up to quadrupoles for triatomic molecules of C_{2v}-symmetry and up to octopoles for linear molecules. The Coulomb interaction of a standard force field as described by the first term in Equation 4.7 is replaced in MTP by a description of multipole expansion for a selected number of molecules in the system. As an example, the electrostatic potential in MTP including multipoles up to rank 2 (quadrupoles) on two atoms named a and b can be written as

$$
\begin{aligned}
E_{elec} = E(q_\mathrm{a} q_\mathrm{b}) + \sum_{k=1}^{3} E(q_\mathrm{a} \mu_\mathrm{b}^k) + \sum_{l=1}^{3} E(\mu_\mathrm{a}^l q_\mathrm{b}) \\
+ \sum_{k=1}^{3} \sum_{l=1}^{3} E(\mu_\mathrm{a}^l \mu_\mathrm{b}^k) + \sum_{m=1}^{5} E(q_\mathrm{a} \Theta_\mathrm{b}^m) + \sum_{n=1}^{5} E(\Theta_\mathrm{a}^n q_\mathrm{b}) + \\
\sum_{k=1}^{3} \sum_{m=1}^{5} E(\mu_\mathrm{a}^k \Theta_\mathrm{b}^m) + \sum_{n=1}^{5} \sum_{l=1}^{3} E(\Theta_\mathrm{a}^n \mu_\mathrm{b}^k) + \sum_{m=1}^{5} \sum_{n=1}^{5} E(\Theta_\mathrm{a}^n \Theta_\mathrm{b}^m),
\end{aligned}
\tag{4.12}
$$

where q_a represents the charge of atom a, μ_b^k is the direction k of the magnitude in the dipole moment vector on atom b, and Θ_a^n is the element n of the quadrupole tensor on atom a (219). Other arguments describe related parameters depending on their indices. This representation can easily be extended by adding higher multipole ranks to improve the electrostatic treatment of force fields. The versatility and robustness of MTP has already been demonstrated for different molecular systems (220–223). In these studies MTP successfully unravelled the conformational dependence of vibrational shifts and frequency splittings of CO and water sensed in the electrostatic field of condensed-phase environments.

4.4. Adiabatic reactive molecular dynamics

A major drawback of MD simulations using classical force fields is their limitation to the description of conformational fluctuations or slow diffusional processes. Nonetheless, chemical reactions like the formation and destruction of bonds or the change of oxidation states are key factors in many chemical or biological processes. The mentioned limitation arise from the fact that the number of terms describing $V_b(r_b, \theta, \chi, \phi)$ in Equation 4.2 remains fixed within one MD simulation and can not be interchanged among each other. These terms incorporate the potential formulations for bonds, angles and dihedrals as described in Equations 4.3 to 4.6. The addition or subtraction of some of these components from $V_b(r_b, \theta, \chi, \phi)$ would allow for the study of chemical processes like bond formations and destructions. Furthermore, all potential parameters describing the classical force field as shown in Equations 4.3 to 4.7 (i.e. $k_{b,i}$, $\theta_{0,i}$, or q_i) remain fixed during a running MD simulation. This disallows, for example, for proper simulations of redox reactions and processes involving changes in the equilibrium structural orientation (e.g. cis/trans isomerization). ARMD describes an extension to standard MD simulations which allows to overcome both of these limitations. Its robust, accurate and versatile application was previously demonstrated on the rebinding dynamics of NO in Mb (224), conformational transitions of the heme group in $\mathrm{Ngb_H}$ (225) and the oxidation of NO in truncated Hb (125).

The stepwise procedure of an ARMD simulation is described in more detail for a bond formation reaction between atoms named A and B. A reduced representation of the interaction between A and B for the bound and the unbound states can be given by two individual potentials V_1 and V_2 depending on the distance r_{AB} between the two atoms as depicted in Figure 4.1. ARMD has the capability to define individual force field parameters for V_1 and V_2. Furthermore, V_{bond} described in Equation 4.3 can be completely removed for certain atom pairs like for V_2 in the current example. To account for a possible bond dissociation the harmonic representation of V_{bond} in Equation 4.3 can additionally be replaced by an anharmonic Morse potential V_{Morse} of the form:

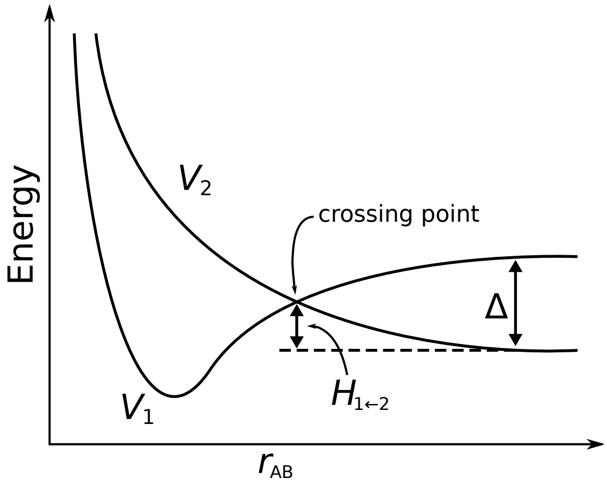

Figure 4.1.: Schematic representation of two potentials V_1 and V_2 representing bonding and non-bonding interactions between two atoms A and B along their displacement coordinate r_{AB}. Δ corresponds to the energetical gap between V_1 and V_2 at their asymptotic limit. $H_{1\leftarrow 2}$ is the activation enthalpy for a transition from V_2 to V_1 defined as the energy difference of the potential minimum on V_2 and the crossing point of V_1 with V_2.

$$V_{Morse} = D_{e,i}(1 - \exp(-\beta_i(r_i - r_{0,i})))^2 \tag{4.13}$$

where $r_{0,i}$ is the equilibrium bond length, $D_{e,i}$ is the dissociation energy and β_i is the anharmonicity parameter. Figure 4.2 shows a schematical representation of the

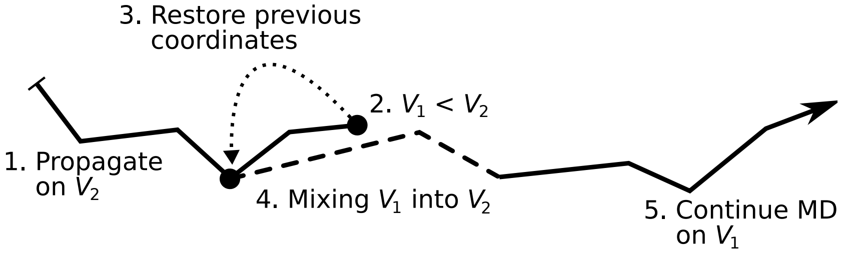

Figure 4.2.: Diagram showing the five steps taking place during an ARMD simulation involving a transition between two potentials V_1 and V_2 along a fictitious trajectory. The dashed line describes the duration m during which V_2 is slowly transfered into V_1.

time course of an ARMD trajectory performing the described bond formation. V_2 represents the potential on which the MD simulation gets initiated in case r_{AB} is large and V_2 is expected to be energetically favored over the bound state V_1 (step 1 in Figure 4.2). In each step of the ongoing simulation V_1 and V_2 are compared in terms of their potential energy difference Δ found in the asymptotic limit of each potential (Figure 4.1). When A and B are approaching each other V_2 starts to increase and V_1 is becoming more stable. At a close enough distance of r_{AB} V_1 and V_2 cross each other. In this case V_1 becomes lower in energy compared to V_2 ($\Delta < 0$) and the ARMD module halts the simulation (step 2 in Figure 4.2) to rewind coordinates and velocities of the system for a user defined amount of timesteps $m/2$ (step 3). Subsequently, the simulation is continued from this point (step 4) for m timesteps where V_2 is replaced by an intermediate potential $V_{1\leftarrow 2}$ of the form:

$$V_{1\leftarrow 2} = V_1 \frac{\tanh(a(t-t_0))+1}{2} + V_2\left(1 - \frac{\tanh(a(t-t_0))+1}{2}\right) \tag{4.14}$$

where t is the current time, t_0 is the time at which the crossing occurred and a is a constant which is related to the chosen mixing time m. The functional form of $V_{1\leftarrow 2}$ describes a smooth transition from V_2 towards V_1 within m timesteps which accounts for an adiabatic process. After the convergence of $V_{1\leftarrow 2}$ into V_1 is completed, the MD simulation is continued for the rest of the simulation time (step 5).

As already pointed out previously, classical force fields are mainly designed to reproduce the correct forces of the molecular system to accurately describe its dynamics. Contrary to QM, MM force fields do not include a reference energy set to which the energy of the system can be compared with. In ARMD the demand for such a reference energy is strongly coming up as the difference of the potential energy Δ between the individual potentials is compared in each MD step. ARMD introduces an additional variable $\Delta_{1\leftarrow 2}$ which modifies Δ between V_1 and V_2 to a meaningful magnitude to circumvent the lack of a reference energy. $\Delta_{1\leftarrow 2}$ represents the potential shift between the two potentials which is, in every ARMD simulation, directly subtracted from V_2. Appropriate values for $\Delta_{1\leftarrow 2}$ can be iteratively fitted to the transition enthalpy $H_{1\leftarrow 2}$ between V_1 and V_2 or free energies of activation $\Delta G^{\ddagger}$ which are related to experimentally accessible transition rates k by the transition-state-theory (226) over

$$\Delta G^{\ddagger} = RT\left(\ln\left(\frac{k_{\mathrm{B}}}{h}\right) - \ln\left(\frac{k}{T}\right)\right) \tag{4.15}$$

where R is the gas constant, T the temperature, k_{B} the Boltzmann constant and h the Planck constant. The latter procedure was used in chapter 5 to describe the transition between the different Ngb configurations $\mathrm{Ngb_L}$, $\mathrm{Ngb_H}$ and $\mathrm{Ngb_P}$.

The CHARMM force field describes dispersion interactions between two nonbonded atoms i and j with a LJ 12-6 potential as specified in the second term of Equation 4.7. Specifically for CHARMM, σ_{ij} is defined as the sum of the individual van-der-Waals radii σ_i and σ_j, and ϵ_{ij} describes the root of the product of the atom related well-depths ϵ_i and ϵ_j in the addressed equation. σ_{ij}, which corresponds to the equilibrium distance of the dispersion interaction, is usually fitted to unnaturally large values. The reason for this overestimation is to prevent clashes between different tertiary structure elements in a protein or at the interface of the protein. However, this form of the potential and its parameters are generally too repulsive for the description of an energetically favorable bond formation. To circumvent this problem, ARMD implements the possibility to replace the potential form of the LJ interaction between bond forming atom pairs, like A and B, by a specific potential of the form:

$$V_{LJ_{\mathrm{A-B}}} = \epsilon_{\mathrm{A-B}}\left[\left(\frac{\sigma_{\mathrm{A-B}}}{r}\right)^{x} - 2\left(\frac{\sigma_{\mathrm{A-B}}}{r}\right)^{y}\right]. \tag{4.16}$$

This potential form allows to replace the LJ 12-6 function coming from the CHARMM force field with a user defined LJ x-y function. x and y have to be defined uniquely for all interactions describing a dissociable bond in a single ARMD simulation. $\epsilon_{\mathrm{A-B}}$ and $\sigma_{\mathrm{A-B}}$ can be assigned individually for each of these atom pairs and correspond to the depth of the potential well and the sum of the van-der-Waals radii between A and B on the dissociated potential. For certain systems, too large values of σ_{ij} can even have an impact on an atom C which is constantly bound (on V_1 and V_2) to atom B and therefore is connected to atom A over two bonds on V_1. To obtain similar benefits for the van-der-Waals interaction between A and C on V_2 as for the interaction between A and B through the introduction of the modified potential (Equation 4.16), the standard CHARMM LJ potential acting between the former atoms on V_2 is similarly replaced by a potential of the form:

$$V_{LJ_{12-6}} = \sqrt{\epsilon_{\mathrm{C}}\epsilon_{\mathrm{A-B}}}\left[\left(\frac{\sigma_{\mathrm{C}} + \sigma_{\mathrm{A-B}}/2}{r}\right)^{12} - 2\left(\frac{\sigma_{\mathrm{C}} + \sigma_{\mathrm{A-B}}/2}{r}\right)^{6}\right] \tag{4.17}$$

where ϵ_{C} and σ_{C} are the general CHARMM force field LJ parameters for atom C and $\epsilon_{\mathrm{A-B}}$ and $\sigma_{\mathrm{A-B}}$ are described by the same parameters used in Equation 4.16. This representation of the van-der-Waals interaction basically corresponds to the standard CHARMM LJ potential where $\epsilon_i = \epsilon_{\mathrm{C}}$, $\epsilon_j = \epsilon_{\mathrm{A-B}}$, $\sigma_i = \sigma_{\mathrm{C}}$ and $\sigma_j = \sigma_{\mathrm{A-B}}/2$. In this way, the van-der-Waals interaction between atoms A and C experiences a similar but weaker degree of scaling as the interaction between A and B which is replaced by Equation 4.16. It assures that a smooth transition of the van-der-Waals interactions between the non-bonding 1-2 and 1-3 interactions is still maintained in case $\sigma_{\mathrm{A-B}}$ is getting significantly reduced over the sum of σ_i and σ_j of atoms A and B.

4.5. Force field parametrization

Previous CHARMM force field parameters for the heme prosthetic group were metal and ligand unspecifically fitted (227, 228) and only crude assumptions were made for hexacoordinated ligands (229, 230). The publications included in chapter 6 on hexacoordinated and photodissociated NgbCO made use of these force field parameter set. However, for the study of explicit transitions and relaxations between different heme configurations in Ngb as it is part of chapter 5, a more profound description of the heme force field in the individual configurations (Ngb$_{\mathrm{P}}$, Ngb$_{\mathrm{L}}$ and Ngb$_{\mathrm{H}}$) is desirable. For this, the heme parameter set was refitted in a consistent manner using recent crystal structure information, data from normal-mode analysis extracted from nuclear resonance vibrational spectroscopy (NRVS) and DFT calculations on heme model systems.

4.5.1. Heme prosthetic group

All QM calculations were performed using the Gaussian 03 suite of programs (231) on a system consisting of a central iron(II)-porphyrin macrocycle with an iron-attached imidazole representing the proximal histidine. On the distal site the cx model system includes additionally CO in coordination distance to the iron and the bis-His system a second imidazole molecule as a representative of the distal histidine (Figure 4.3). The initial coordinates of the porphyrin ring and the proximal imidazole were taken from the X-ray structure with PDB code 1W92. For the initial optimizations all in-plane sidechains were removed and missing hydrogen atoms attached which correspond to the core model system depicted in bold in Figure 4.3. The optimizations were performed using the B3PW91 DFT functional (232) and the Los-Alamos National Lab double-ζ effective core potential (233, 234) (LANL2DZ) on the iron inner-shell electrons. All remaining electrons were described by a 6-31G(d,p) explicit basis. The charge of every core system was set to zero. The minimum-energy spin-states of the hexacoordinated cx and bis-His models is in both cases $S = 0$ whereas the electronic configuration of the pentacoordinated CO unbound system is energetically lowest at $S = 2$ (235). After the optimization of the core structure the heme sidechains were attached as shown in Figure 4.3 and subsequently optimized on the HF/STO-3G level keeping the previously optimized coordinates fixed.

Equilibrium bond distances and angles were extracted from recent high-resolution ferrous heme X-ray structures embedded within globin folded proteins. Equilibrium dihedral angles were only qualitatively analyzed as their main purpose in the force field is to keep the porphyrin system planar (180° for eclipsed, 0° for staggered conformations). For each configuration the equilibrium coordinates were defined as the mean from three such crystal structures: The pentacoordinated, dx-configured heme system is represented by the structures with PDB codes 2DN2 (hemoglobin) (236), 1VXD (myoglobin) (84), and 1GDJ (leghemoglobin) (237). The cx heme coordinates were averaged over PDB structures 1A6G (myoglobin) (85), 1W92 (neuroglobin) (143) and 2DN3 (hemoglobin) (236). Structures with PDB codes 1Q1F (neuroglobin) (141), 2DC3 (cytoglobin) (149) and 2BK9 (hemoglobin) (150) all containing a bis-His config-

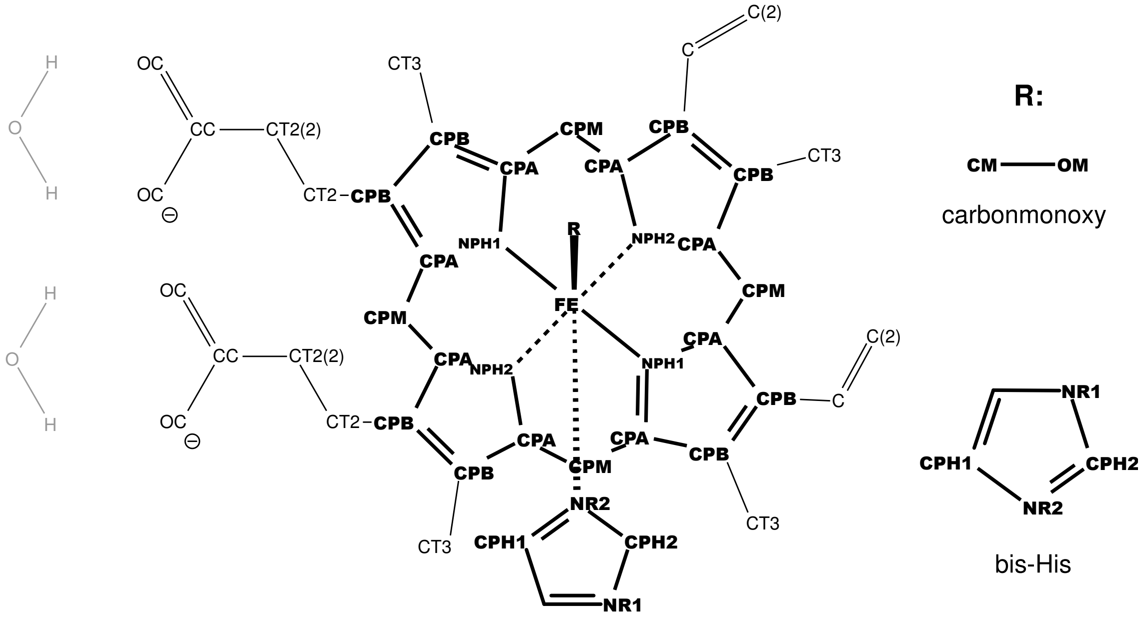
H
O
H
OC
CC
CT2(2)
OC
⊖
CT2-CPB
CT3
CPB
CPA
CPM
CPA
CPB
C
C(2)
CT3
CPB
NPH1
R
NPH2
CPA
CPA
CPM
FE
CPM
CPA
NPH2
NPH1
CPA
C(2)
CPB
C
CT2-CPB
CPA
CPA
CPB
CPM
CPB
NR2
CT3
CT3
CPH1
CPH2
NR1
R:
CM
OM
carbonmonoxy
NR1
CPH1
CPH2
NR2
bis-His

Figure 4.3.: Lewis-structure representation of the model systems used for all *ab initio* calculations in the parameterization procedures for the CHARMM force field and the ARMD specific additions. The structure drawn in bold corresponds to the porphyrin core optimized for each heme configuration individually using B3PW91/6-31G(d,p)/LANL2DZ. Atom names correspond to the CHARMM atom type definitions expect for the charge assignments to atom type C of which the terminal C is specifically named C(2). Hydrogen atoms, which are not shown, are generally of type HA if not specifically defined in the text. Water molecules included to fit the charges on the propionate groups according to the interaction energy are shown in grey.

uration represent the statistical population for the corresponding heme configuration. Table A.1, included in the appendix, summarizes the extracted average geometries which were used as equilibrium values in the simulations for each specific heme configuration. Geometries belonging entirely to the porphyrin sidechains and all geometries including hydrogen atoms were adapted from Kuczera et al. (227). Atom names used in Tables A.1 to A.4 and A.6 coincide with the ones in Figure 4.3. Hydrogen atoms of the porphyrin-ring are all of type HA.

Bond, angle and dihedral angle force constants, which are unspecific of the heme configuration, were converted from a normal-mode analysis deriving from recent nuclear resonance vibrational spectroscopy (NRVS) data on an iron octaethylporphyrin system (238). The complete list of force constants is divided into a section which is common to all three heme configurations (Table A.2) and two parts dealing with the iron modes differing for each configuration and the dihedral torsions involving hydrogen atoms (Tables A.3 and A.4). Modes involved in the penta- and hexa-coordination of the iron-porphyrin which are not available from the NRVS normal-mode analysis (238) were partly adapted from a previous fit to DFT data (239), fitted to DFT frequency calculations on the models and methods described above, or copied from the standard CHARMM force field parameter set for imidazole substituted species. Dihedral torsional modes within the porphyrin ring-system were applied instead of improper torsional modes used in the previous parametrization of Kuczera et al. (227). Therefore, a comparison of the present torsional parameters with the ones from Kuczera is excluded from Tables A.3 and A.4. Another significant difference regarding the torsional treatment between the current and Kuczera's parametrization is the definition of the periodicity n of the porphyrin attached substituents. Kuczera set the multiplicity n of these torsional modes to 2 which retains the substituents in-plane compared to the porphyrin. In contrast, most X-ray structures of porphyrin systems (238) and of the heme prosthetic groups in globins (84, 85, 141, 149, 150, 236, 237) show a staggered conformation of substituents. Accordingly n is set to 6 for these torsions (see Tables A.2 and A.3) which allows the substituents to obtain a staggered 30° orientation. Specific modes influencing the iron in-plane & out-of-plane stretching- and bending-modes were individually fitted for each of the three heme configurations individually by iteratively changing the involved force constants until the iron normal-mode frequencies calculated by CHARMM's VIBRAN interface were in good agreement with NRVS (240, 241) and Raman data (242). Table A.5 shows a comparison of experimental and CHARMM fitted frequencies having the strongest iron normal-mode contributions.

Charges were derived from Mulliken population analysis individually calculated for each QM optimized model. Following the CHARMM parametrization protocol (243) the charges at all nonpolar hydrogens of the heme prosthetic group were set to 0.09 e by transferring the excessive positive Mulliken charge into the adjacent carbon atoms. Charges on equivalent atom types were averaged. As the CHARMM force field is optimized for MD simulations together with the TIP3P water model (244) the charges on the polar propionate sidechains were redistributed to fit the B3PW91/6-31G(d,p) interaction energy with coordinating water molecules. A water molecule was placed in coordination distance of each propionate sidechain (see Figure 4.3) and subsequently optimized on the B3PW91/6-31G(d,p)/LANL2DZ level of theory keeping all remaining coordinates fixed. The interaction energy was computed from the energy difference between the model-systems incorporating the water molecules and the sum of energies of the separated species. The final charge assignment listed in Table A.6 and obtained by fitting the CHARMM force field to this interaction energy resulted in heme-water interaction energies between -38.37 to -38.55 kcal/mol. This corresponds to a difference to the corresponding DFT interaction energy of less than 0.24 kcal/mol.

LJ parameters describing the Van-der-Waals interaction were entirely adopted from the parametrization of Kuczera et al. (227) expect for the atoms involved in ARMD bond dissociation and formation to and from the dx-heme states. The parametrization of the latter is described in the next section.

4.5.2. ARMD specific parameters

For the ARMD studies discussed in section 5.5 the CHARMM force field standard harmonic bond terms of the FE-CM bond in $\mathrm{Ngb_L}$ and the FE-NR2 coordination in $\mathrm{Ngb_H}$ were replaced with Morse potentials as described in Equation 4.13. Parameters for r_0, D_e and β were in both cases obtained by a fit to unrelaxed DFT scans (B3PW91/6-31G(d,p)/LANL2DZ) along the dissociation coordinates (FE-CM and FE-NR2) of the corresponding models shown in Figure 4.3. The resulting potential of the bis-His elongation scan is shown in Figure 4.4A together with the derived Morse potential V_{Morse}^{His}. The fitted parameters r_0^{His}, D_e^{His} and β_{His} are included in Table A.7. Initial ARMD calculations of the CO rebinding in Ngb$_{\mathrm{P}}$ applying a Morse po-

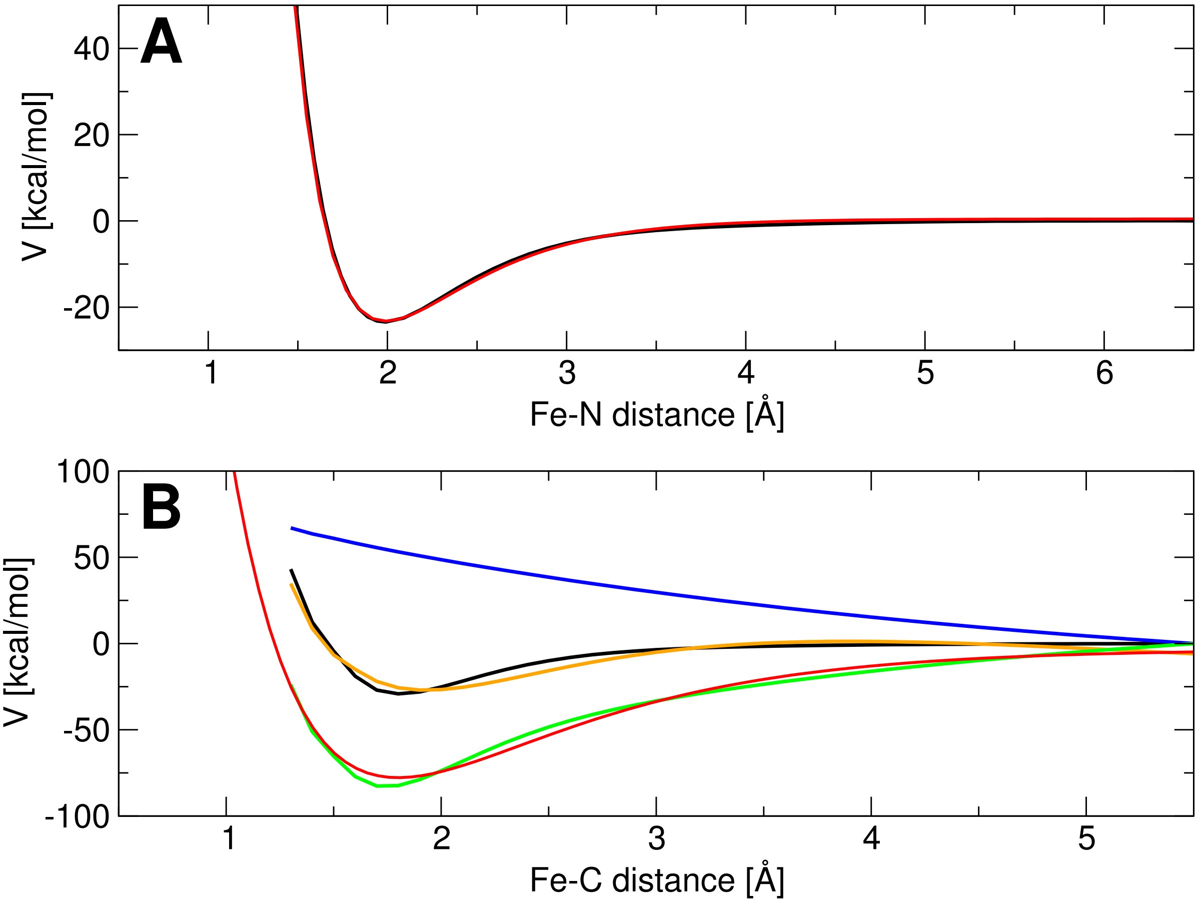

Figure 4.4.: Description of the fitting procedures for V_{Morse}^{His} (A) and V_{Morse}^{CO} (B) describing ARMD bonding interactions between Fe-His64 and Fe-CO. Black: Total potential obtained from the DFT scan. Blue: CHARMM non-bonding potential along the structures of the DFT scan. Green: Difference of the total DFT potential and the CHARMM non-bonding potential. Red: Fitted Morse potential to the black (A) and green (B) curve, respectively. Orange: Potential calculated by the complete CHARMM force field incorporating the fitted Morse function.

tential V_{Morse}^{CO} for the FE-CM which was correspondingly fitted to the related *ab initio* potential (black curve in Figure 4.4B) resulted in chemically unreasonable rebinding events at rather long FE-CM distances ($\gtrapprox$ 3.5 Å). A closer look at the energetic contribution of the CHARMM non-bonding interactions along the scanning coordinate revealed a strong repulsive contribution related to the non-bonding CHARMM potential V_{nb}. V_{nb} includes all non-bonding interactions which are at least three adjacent bonds away from each other (blue curve in Figure 4.4B). The *ab initio* potential represents the total potential V which is, in the CHARMM force field, split into the bonding contribution V_b and V_{nb} (see Equation 4.2). For this reason, the subtraction of V_{nb} from V, the total *ab initio* potential, should result in a partial potential representing only the contribution from V_b (green curve in Figure 4.4B). Refitting V_{Morse}^{CO} to this potential (red curve in Figure 4.4B) resulted in the Morse potential parameters r_0^{CO}, D_e^{CO} and β_{CO} included in Table A.7. Compared to a direct fit to V, D_e^{CO} becomes more than twice as large after subtracting the CHARMM non-bonding contribution. This discrepancy is the main factor preventing the system to perform CO rebindings at absurd distances.

LJ potential parameters between the bond-forming atoms for the transition from $\mathrm{Ngb_P}$ to $\mathrm{Ngb_H}$ and $\mathrm{Ngb_L}$ were fitted in a similar procedure as V_{Morse}^{CO} described above. The interaction of CO and His64 with a dx heme model from Figure 4.3 were calculated using the same DFT method and basis sets. Initial geometries of the model system atoms correspond to the optimized structure in the pentacoordinated state ($S = 2$) which has the iron atom in a domed out-of-plane conformation. Potentials shown in brown in Figures 4.5A and B represent LJ 12-6 functions for the FE—NR2 (4.5A) and FE—CM (4.5B) van-der-Waals interaction with ϵ_{ij} and σ_{ij} taken from the standard CHARMM force field parameter set (227, 243). The DFT computed potentials V from the model scans are shown in black. Both, the equilibrium distances and the well-depths in this potentials look drastically smaller from visual inspection. The potential well from the CHARMM force field starts to rise at $\approx$ 0.25 and $\approx$ 0.5 Å longer FE—NR2 and FE—CM distances compared to the *ab initio* scanned potential V. The blue curves in Figures 4.5A and B correspond to non-bonding CHARMM potentials V_{nb} calculated on the model structures including all non-bonding interactions of the systems expect the LJ interaction between FE and the distal NR2 and between FE and CM, respectively. These potentials were subtracted from V (shown in green in Figures 4.5A and B) to obtain the quantum-mechanical contributions only related to the van-der-Waals interaction between FE—NR2 and FE—CM. The red curves in Figures 4.5A and B are fitted LJ 7-5 functions according to Equation 4.16. Their parameters which were used on the $\mathrm{Ngb_P}$ are listed in Table A.7.

Rate constants for the dissociation and rebinding of CO and distal His64 were de-

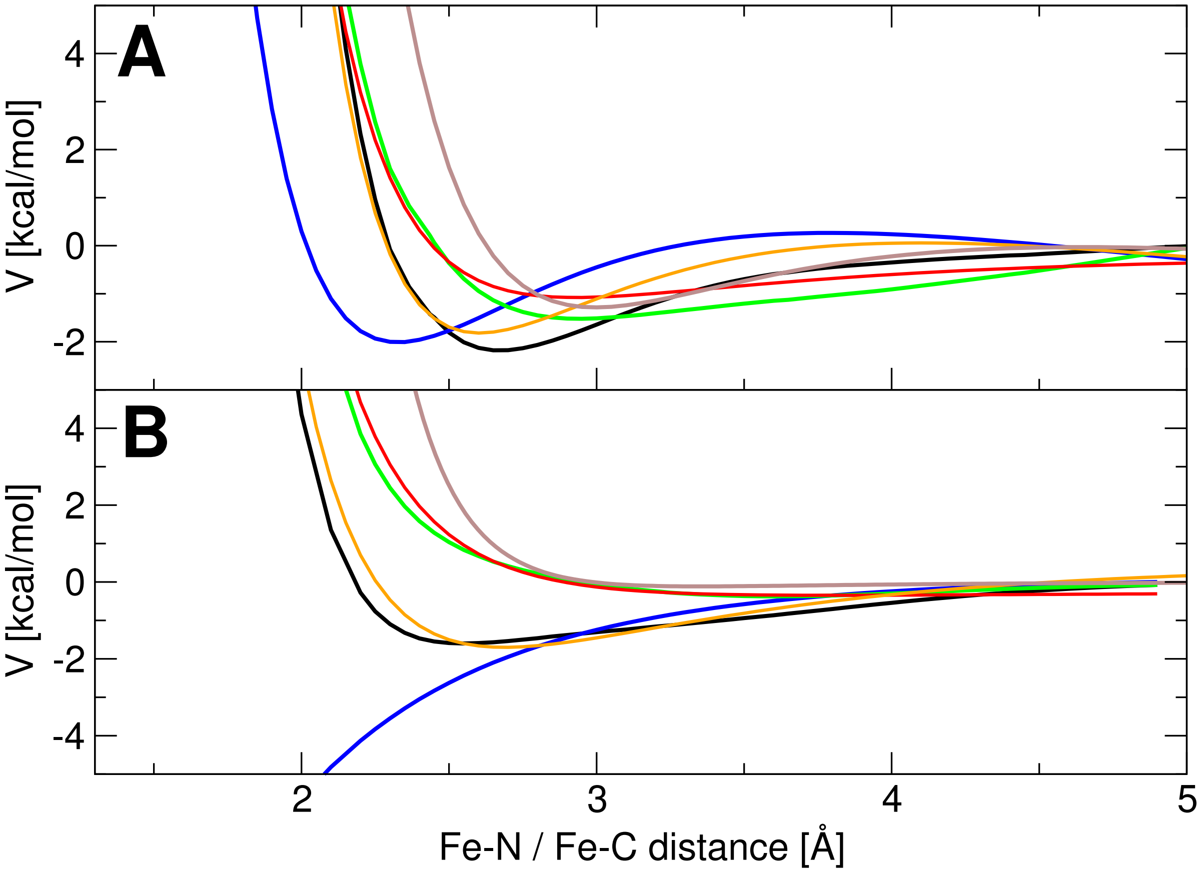

Figure 4.5.: Description of the fitting procedures for the modified non-bonding Van-der-Waals $V_{LJ_{7-5}}$ potentials described by Equation 4.16 between FE—NR2 of His64 (A) and FE—CM of CO (B). Brown: Standard CHARMM LJ potentials. Black: Total potentials obtained from the DFT scan in the model systems. Blue: CHARMM non-bonding potential for the same structures excluding the LJ interactions between FE—NR2 and FE—CM, respectively. Green: Difference of the black and blue potentials. Red: Fitted LJ 7-5 potentials to the green curves. Orange: Potentials calculated by the complete CHARMM force field incorporating the modified LJ functions.

termined in several experimental studies (129, 154, 157). Therefore, meaningful ARMD potential energy differences between the dx and cx heme configured potentials ($\Delta_{\mathrm{Ngb_P}\leftarrow\mathrm{Ngb_L}}$) and the dx and bis-His states ($\Delta_{\mathrm{Ngb_P}\leftarrow\mathrm{Ngb_H}}$) were estimated by fitting the free energies of transition $\Delta G^{\ddagger}$ iteratively to the experimental rate constants $k_{\mathrm{CO_{off}}}$, $k_{\mathrm{CO_{on}}}$, $k_{\mathrm{His_{off}}}$ and $k_{\mathrm{His_{on}}}$ using Equation 4.15. The parameters ($\Delta_{\mathrm{Ngb_P}\leftarrow\mathrm{Ngb_L}}$) and ($\Delta_{\mathrm{Ngb_P}\leftarrow\mathrm{Ngb_H}}$), obtained in this way, are listed in section 5.5 together with a discussion on the derived free energy profiles describing the bi-phasic ligand rebinding competition in photolyzed NgbCO.

5. Competitive Rebinding in Neuroglobin

"When you were bravest, sir, and your sword sharpest, I durst affront you, you know I durst; when the court sun gilded you, and every cry was, The young hopeful Pedro, Alonso's sprightly son! then I durst meet you, when you were master of this mighty fame, and all your glories in the full meridian. Had we then come to competition, which I often sought–"

John Philip Kemble, *The Pilgrim, 1787*

The published work on Ngb which is included in chapter 6 studies the dynamical fluctuations of NgbCO or photodissociated NgbCO individually which corresponds either to the Ngb_L or the Ngb_P configuration. In this chapter, additional simulations from unpublished studies on the relaxation dynamics occurring during the configurational transitions between Ngb_P and Ngb_L or Ngb_H are presented. The initial section of this chapter describes the computational methods for these simulations. Subsequently, the results are discussed in more detail.

One set of simulations looks into the details of relaxational transitions invoked by the His64 dissociation from Ngb_H and its gradual conversion to the Ngb_P configuration. Specific features which were looked at in these simulations include the heme-sliding process suggested from X-ray structure comparisons (141, 143), the rearrangements of the cavity network, the influx of water molecules from the bulk into the internal protein cavities and the rearrangement of the Tyr44 residue which is under equilibrium conditions coordinated to the heme group. Furthermore, relaxational differences found between mNgb and hNgb are discussed and put into relation with experimental observations. The last section of this chapter looks at the energetical profiles for the rebinding competition of Ngb_H with Ngb_L in the Ngb_P state, compares them to experiment and shows how Ngb_P can be stabilized over Ngb_H in the absence of an exogenous ligand.

5.1. Methods

5.1.1. Computational preparations

All MD simulations were carried out with the CHARMM suite of programs (245) and the CHARMM22 force field (243). hNgb was generated from monomer D of the Ngb_H crystal-structure with PDB code 1OJ6 (140). Cys46 and Cys55 which are replaced in 1OJ6 by serines were reintroduced including the connecting disulfide bond. mNgb is built on the X-ray structure of mNgbCO (PDB code 1W92 (143)). Missing H-atoms were added to both systems. The current study includes the bis-His configuration which makes the presence of a neutral His64 side chain protonated on the N_δ atom a prerequisite and was therefore explicitly assigned in both Ngb species. Positions of all remaining hydrogens were chosen by CHARMM's built-in HBUILD algorithm (245). Both proteins were solvated in a preequilibrated periodic TIP3P waterbox with dimensions $41 \times 61 \times 56$ Å^3. After solvation hNgb consists of 2444 protein atoms and 4276 water molecules and mNgbCO of 2465 protein atoms and 4269 waters. All hydrogens in the systems were minimized for 500 steps with the steepest descent method and another 1000 steps with the Adopted Basis Newton-Raphson algorithm. The same procedure was repeated for all water molecules in each system and in hNgb_H additionally for residues 44 to 58 which build up the CD loop including the Cys46 to Cys55 disulfide bond. After heating both proteins from 20 to 300 K for 40 ps and equilibrating for an additional 100 ps at 300 K, the C_α atom RMSD in hNgb_H converged to 1.3 Å compared to the initial crystal structure. In mNgb_L the same RMSD converged to 1.5 Å compared to the X-ray structure it derives from.

Pentacoordinated structures of hNgb and mNgb (hNgb_P & mNgb_P) were generated by either removing the covalent bond between N_ϵ of His64 and Fe of the heme prosthetic group in the equilibrated structure of hNgb_H or by deleting the CO molecule in equilibrated mNgb_L. Both pentacoordinated proteins were further equilibrated at 300 K for 150 ps each. After this the RMSD for the C_α atoms was 1.1 Å for hNgb_P (compared to the bis-His equilibrated structure) and 1.5 Å compared to the crystal structure. The mNgb_P equilibrated structure shows a backbone C_α RMSD of 1.1 Å compared to the previously equilibrated NgbCO structure and of 1.2 Å compared to the initial X-ray geometries.

mNgb$_{\mathrm{H}}$ used for the His64 dissociation simulations was generated by introducing a harmonic potential between N$_\epsilon$ of His64 and the iron atom of the heme group. After an initial 500 step steepest descent and a subsequent 1000 steps Adopted Basis Newton-Raphson minimization of the entire protein, followed by 40 ps of heating from 20 to 300 K, the system was equilibrated at 300 K for 250 ps. The C$_\alpha$ RMSD of this structure is 1.1 Å compared to the equilibrated mNgb$_{\mathrm{P}}$ and 1.4 Å relatively to the crystal structure. Similarly to this procedure, equilibrated systems of V109L, V109F and V109W mNgb$_{\mathrm{H}}$ were generated starting from mNgb$_{\mathrm{P}}$ by replacing the Val109 side chain by the one from the corresponding mutant prior to minimization, heating and final equilibration.

5.1.2. Bis-His dissociation dynamics

Initial structures for ten independent 10 ns MD simulations of a bis-His dissociation in hNgb$_{\mathrm{H}}$ and mNgb$_{\mathrm{H}}$ were collected after every 100 ps of a 1 ns *NPT* simulation for each system at 300 K and 1 bar. The dissociation was initiated by running short (100 fs) trajectories having the bond and angle energy terms between the N$_\epsilon$ atom of His64 and the heme iron replaced by a repulsive potential of the form $V_{\mathrm{Diss}} = \alpha r_{\mathrm{e}}^{12}\, r^{-12}$ where r_{e} is the equilibrium bondlength of the two atoms and α was set to an empirical value of 10 kcal/mol. Subsequently, the standard nonbonding potential energy terms were introduced for the remaining simulation time. For a comparison, 5 ns simulations of hNgb with a broken Cys46-Cys55 bond were calculated from the equilibrated hNgb structures including the disulfide bond.

5.1.3. Umbrella sampling simulations

All umbrella sampling simulations were carried out starting from fully equilibrated systems as described above. In simulations involving unbound CO, the electrostatic interactions were treated with a well tested three-point fluctuating charge model to more accurately capture the protein-ligand interactions (103, 117). Furthermore, the CO bond was treated with a spectroscopically accurate rotational Rydberg-Klein-Rees potential (246, 247). Umbrella sampling windows along every distance dependent reaction coordinate (see below) are separated by 0.1 Å from each other and each window was sampled for 50 ps from NVT MD simulations. To every sampling window a harmonic restraining potential of the form $V_{\mathrm{umb}} = k_{\mathrm{umb}}(r - r_{\mathrm{umb}})^2$ was applied where k_{umb} is the umbrella force constant, r the sampling coordinate and r_{umb} the equilibrium coordinate of the potential. k_{umb} was varied in the range of 5 to 350 kcal/mol Å^2. Transitions between Ngb$_{\mathrm{H}}$ to Ngb$_{\mathrm{P}}$ and between Ngb$_{\mathrm{L}}$ to Ngb$_{\mathrm{P}}$ were treated by ARMD. The reaction coordinate for the former transfer was defined by the His64 N$_\epsilon$ to heme iron distance r_1 (see Figure 3.2) and for the latter transition by the heme iron to CO center-of-mass distance r_2. The data from individual simulations were combined with a weighted histogram analysis (248).

Umbrella scans describing the separation of Tyr44 from its heme propionate hydrogen bond coordination (see Figure 3.1) were performed along the distance $r_{\mathrm{O-O}}$ defined as the Tyr44 oxygen OH to the heme propionate oxygen O1A, starting from a hydrogen bonded conformation. Individual free energy profiles obtained from consecutively elongated 50 ps MD simulations were subsequently combined with weighted histogram analysis.

5.2. Heme sliding

This section analyzes the heme-sliding process in terms of the average coordinate fluctuations from each set of ten individual 10 ns MD simulations and 5 ns for hNgb having a broken disulfide bond when the bis-His coordination of the equilibrated structure gets disrupted. The structural rearrangements of the heme prosthetic group are in the following compared along its surface normal $\vec{n}$ (Figure 5.1A) and the angle γ between the distal E– and proximal F–helices (Figure 5.1B). Time series discussed in the following were extracted from structures aligned to the C$_\alpha$ atoms of the protein backbone before the bis-His dissociation was initiated (Ngb$_{\mathrm{H}}$). Figure 5.1C shows the averaged time series of Δn_t^z which is the z-component n_t^z of $\vec{n}_t$ relative to its initial Ngb$_{\mathrm{H}}$ position n_0^z. Relaxation times for this coordinate differ strongly between mNgb and hNgb. While mNgb shows an initial relaxation following His64 dissociation already after 800 ps by returning more into the proximal site (-0.29 Å) the same process takes twice as long in hNgb with the disulfide bond. In the same time frame, hNgb with a broken Cys46-Cys55 bond, closely follows the relaxation behavior of mNgb.

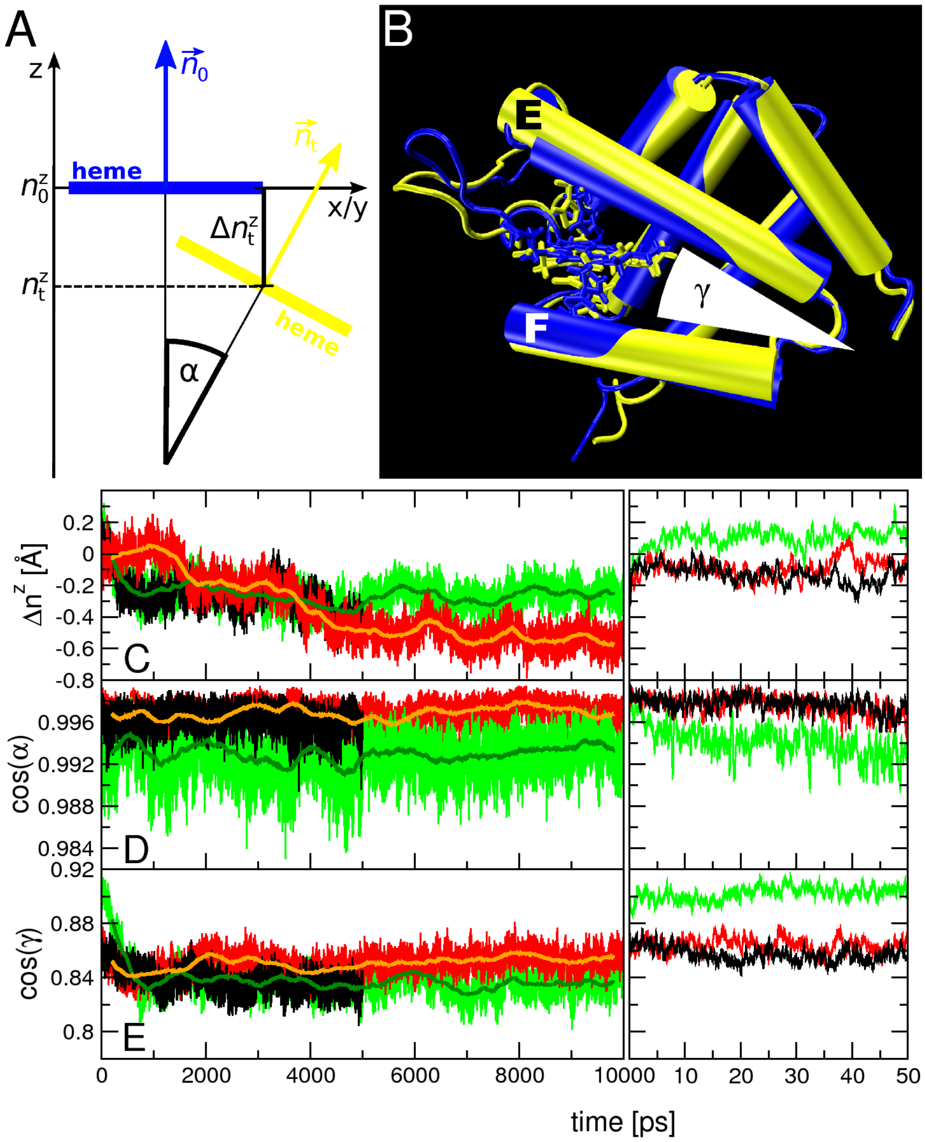

Figure 5.1.: A: Illustration of the heme rearrangements related to the heme plane normal $\vec{n}$ before (blue, $\vec{n}_0$) and at time t after the His64 dissociation (yellow, $\vec{n}_t$). B: Superimposed structures of mNgb$_{\mathrm{H}}$ (blue) and mNgb$_{\mathrm{P}}$ (yellow, taken after 10 ns of simulation). The angle γ represents the scalar product of the two vectors described by helices E and F. C: Evolution of the z-component displacement Δn_t^z at the current position n_t^z relative to its initial Ngb$_{\mathrm{H}}$-magnitude n_0^z. A negative value describes a heme movement towards the proximal site and a positive towards the distal site. D: Evolution of the scalar product of $\vec{n}_0$ and $\vec{n}_t$. E: Evolution of the scalar product between the vectors defining the orientation of helices E and F. Each timeseries represents an average over the ten individual ones of mNgb shown in green, hNgb with Cys46-Cys55 bond shown in red and hNgb with broken disulfide bond in black. Left: Complete 10 ns representation, right: Detailed view on the first 50 ps.

This behavior suggests that the difference in the relaxation time of the initial heme movement towards the proximal site is predominantly caused by the presence or absence of the disulfide bond. Subsequently, for all three systems, the heme plane slides more slowly into the proximal site until it reaches $\Delta n_t^z = -0.3$ Å at 3.7 ns after which it exhibits a different behavior again. The relaxation of Δn_t^z in mNgb is finished in the beginning of the second half of the simulation and fluctuates at ≈ -0.33 Å whereas in the same time window hNgb, having a disulfide bond, relaxes further down into the proximal site and stabilizes at $\Delta n_t^z \approx -0.55$ Å. This suggests that Cys46-Cys55 bound hNgb is able to offer a larger volume to the distal site cavity. On the ps timescale (right hand graph of Figure 5.1C) the heme plane in both proteins advances the proximal site ($\Delta n_t^z \approx -0.2$ Å) within 1 ps. After 5 ps the vertical displacement of the heme group in both hNgb systems is relaxing again to its initial position. In mNgb this back drift is more strongly pronounced reaching a position of $\Delta n_t^z \approx +0.1$ Å.

The angular displacement of $\vec{n}_t$ at time t relative to the initial orientation $\vec{n}_0$ is plotted in Figure 5.1D in terms of its scalar product. mNgb and hNgb both experience a roughly periodic motion along this coordinate which is almost in phase between each other showing a minimum at $\approx$ 1100 ps and $\approx$ 3600 ps. The frequency of this oscillating motion lies around 0.4 GHz. After 5 ns these periodic fluctuations are nearly absorbed by the environment which likely correlate with the convergence of the vertical heme movement in Figure 5.1C. The main difference between mNgb and hNgb (with and without Cys46-Cys55 bond) is the average angle α to which the two systems converge to. mNgb reorients by roughly 2° more strongly compared to its initial orientation than both species of hNgb do. The initial relaxation of α takes place within $\approx$ 10 ps in all three analyzed proteins (right graph of Figure 5.1D). The fact that the angle α relaxes similarly in both hNgb systems and on every timescale (ps and ns) suggests that the angular reorientation of the heme sliding process is controlled by characteristics unrelated to the Cys46-Cys55 bond. Figure 5.1E shows the evolution of $\gamma(t)$ between helices E and F. In the initial $\mathrm{Ngb_H}$ configuration, $\gamma(0)$ between mNgb and both hNgb differs by $\approx 5°$. $\gamma(0)$ in $\mathrm{hNgb_H}$ is constraint to $\approx 31.2°$ independently from the presence or absence of the disulfide bond, whereas in mNgb the angle is smaller ($\gamma(0) = 26.3°$). During the transition from the $\mathrm{Ngb_H}$ to $\mathrm{Ngb_P}$ state γ relaxes to similar amplitudes of $\approx 34°$ in disulfide broken hNgb as well as mNgb within $\approx$ 1 ns. On the other hand, Cys46-Cys55 bound hNgb converges to $\gamma \approx 31°$ after around 2 ns of His64 dissociation.

The two crystal structures from which the heme-sliding motion was originally suggested (Figure 3.1) (143) show γ angles which are in the same range as found in the present relaxation of mNgb. In the metNgb X-ray structure of mNgb (PDB code 1Q1F) which corresponds to the $\mathrm{Ngb_H}$ configuration, γ captures an angle of 28° which differs by less than 2° from the MD equilibrated $\mathrm{mNgb_H}$ value. The NgbCO crystal structure (PDB code 1W92) has a angle γ of 35° which is only 1° larger compared to the 10 ns relaxed γ angle of $\mathrm{mNgb_P}$. The heme related coordinates Δn^z and α differ more strongly between the two X-ray structures than observed from the relaxations in the 10 ns MD simulations. In the mNgbCO crystal structure Δn^z is -0.91 Å compared to the metNgb one and $\cos(\alpha)$ lies at 0.96. These differences in the reorientation between the crystal structure from metNgb to NgbCO and the MD simulations describing the transition from the $\mathrm{Ngb_H}$ to the $\mathrm{Ngb_P}$ state most likely arise from an additional contribution of the bound CO ligand in the heme slidden X-ray structure. The inclusion of CO can therefore be expected to introduce steric interactions on top of the $\mathrm{Ngb_P}$ configuration leading to a continuation of the heme sliding motion. Further MD simulations would be needed to verify this proposition and to estimate the timescale of this process.

5.3. Cavity evolution

5.3.1. Docking site expansion

From every 10 ns $\mathrm{Ngb_H}$ dissociation MD simulation structures of mNgb and disulfide-bonded hNgb were extracted every 100 ps. Protein cavities in each of these structures were identified using the SURFNET suite of programs (249). The pocket network was divided into six different cavities which are indicated in Figure 3.3. It follows a similar docking site network description for Ngb as already discovered in the publication included in section 6.1. The evolution of each individual volume ($V_{\mathrm{B'}}$, $V_{\mathrm{B''}}$, V_{C}, V_{D}, V_{E} and V_{F}) and the total cavity volume V_{tot} over time is shown in the upper and center graphs of Figure 5.2. The lower graph shows the relative changes in volume over time compared to the initial volume $V(0)$ for the cavities which are closest to the bulk site and therefore are expected to be most strongly involved in the absorption of external ligands like CO, O_2, or H_2O. In the initial bis-His state of mNgb V_{tot} increases

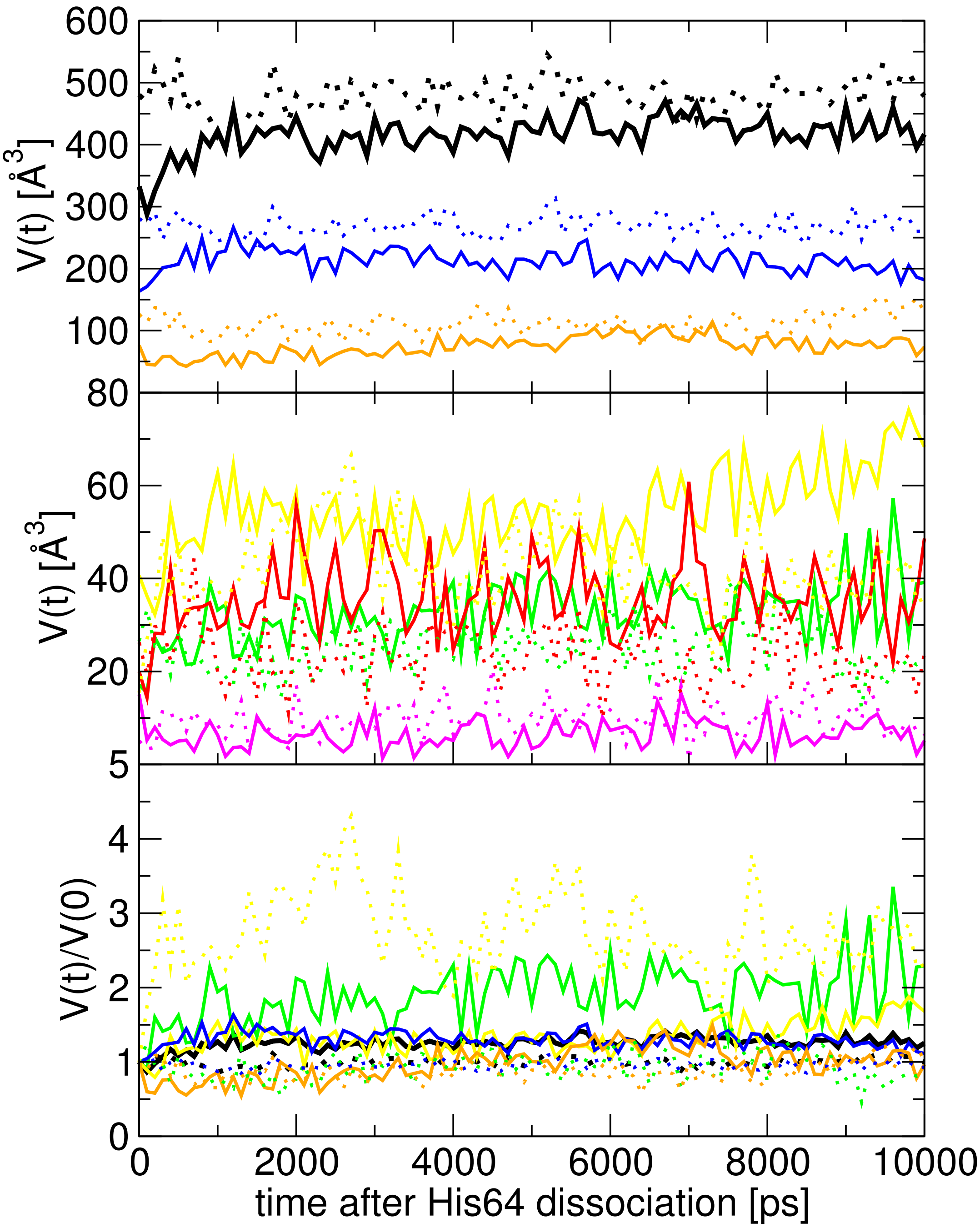

Figure 5.2.: Averaged evolution of individual cavity volumes after His64 dissociation for mNgb (solid lines) and hNgb (dotted lines). Top and center graph: Absolute volumes, Bottom graph: Volume relative to the initial Ngb_H configuration. Green: $V_{B'}$, yellow: $V_{B''}$, red: V_C, blue: V_D, purple: V_E, orange: V_F, black: V_{tot}.

from 332 Å^3 within $\approx$ 1.2 ns to 430 Å^3 where it roughly resides for the remaining simulation time. The equivalent volume $V_{\text{tot}}(0)$ of hNgb is 480 Å^3. Contrary to mNgb $V_{\text{tot}}(t)$ does not relax significantly and allocates an average volume of $\approx$ 480 Å^3 over the complete 10 ns of dissociated MD. V_{tot} calculated inhere is $\approx$ 100 Å^3 larger compared to the reported volumes of the crystal structures (141, 143). This discrepancy derives from the fact that the docking sites which are located in the CD-loop region (pocket F) are an integral part of the current analysis, whereas the interpretation of the X-ray structure volume contributions in this region were not considered previously. Inspection of V_{F} for mNgb and hNgb shows that it accounts roughly for the 100 Å^3 overestimation of V_{tot} compared to the crystal structure. Docking site F has a large influence on the water intake into the protein, as shown in the next section, and therefore is regarded to be crucial for the solvent and ligand diffusion from and into the protein. Individual volumes which contribute most strongly to the total difference in the intial Ngb_{H} states of mNgb and hNgb are V_{E} and V_{F}. Additionally, the volume of both cavities is constantly larger for hNgb over the complete 10 ns simulation. The rapid volume increase in mNgb originates mainly from $V_{\text{B}'}$, V_{C}, V_{E}, and V_{F}. In hNgb the only cavity which shows a significant change of volume over time is the primary docking site B″ which increases by a factor of 3 to 4 compared to $V_{\text{B}''}(0)$. This pocket, which is adjacent to the heme iron atom, is regarded as the ligand rebinding site and can therefore be expected to have the strongest influence on the ligand coordination kinetics.

Differences in cavity sizes and evolution between mNgb and Cys46-Cys55 bound hNgb can also be identified from structure. Figure 5.3 shows averaged cavity densities at 0, 2 and 10 ns after the initiated His64—Fe disruption. In hNgb the most strongly noticeable change is a dilution of site F (orange isosurface in Figure 5.3). The pocket is more open to the protein surface compared to mNgb and shows a stronger population near the protein surface. After the disruption of His64 site F expands (2 ns) and its high density isosurface (shown in white) moves closer towards the protein surface (10 ns). The same space in mNgb is only sparsely populated at 0 ns and gains only little density during the 10 ns relaxation. Similarly, the primary docking site B″ (yellow) shows only smaller changes in mNgb over time which also agrees with the evolution of $V_{\text{B}''}$ in the lower graph of Figure 5.2. In hNgb the same docking site shows a remarkable increase of $V_{\text{B}''}$ of the factor of three. It expands towards site F and its density rises in the region near the heme plane propionates. The volume

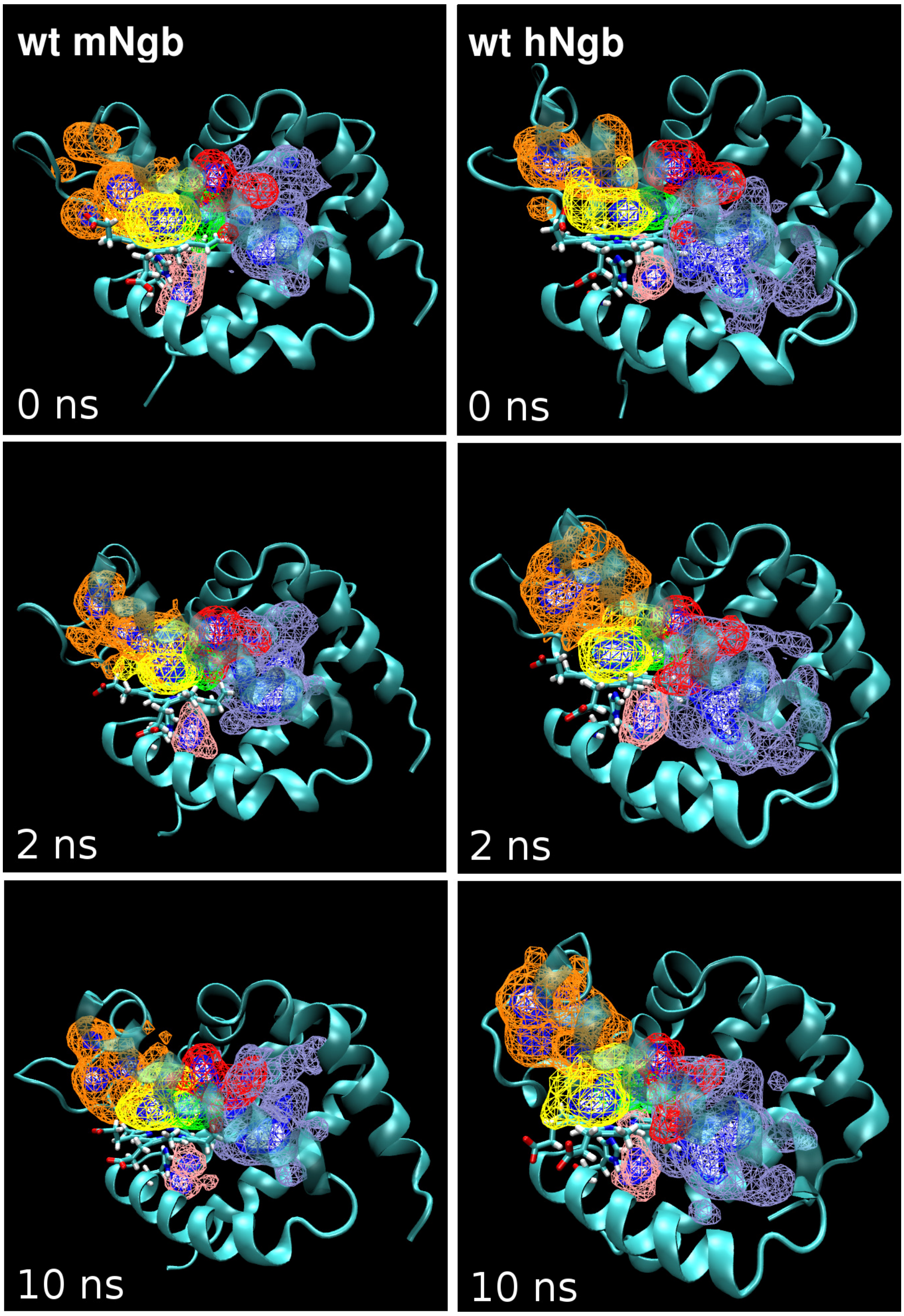

Figure 5.3.: Collection of snapshots showing the individual cavity population densities in mNgb and Cys46-Cys55 bound hNgb for 0, 2 and 10 ns after the His64 dissociation. The isosurfaces describe individual volumes as relative population densities identified from the total of ten independent trajectories of each protein. The outermost isosurface of each cavity is color-coded as labeled in Figure 5.2 and encloses the total volume of each docking site. White: 50 %, blue: 75 %, color-coded: 100 % population density.

V_{E} which is larger in hNgb compared to mNgb is more expanded in each timestep as shown in Figure 5.3.

Figure 5.4 shows the population density distribution of the total volume V_{tot} for mNgb and hNgb. The $\mathrm{Ngb_H}$ configuration at 0 ns indicates a stronger overall cavity density in the CD-loop region of hNgb compared to mNgb as already postulated for the individual population density of site F between the two proteins. Furthermore, it becomes apparent from this unified representation of the internal cavity density that the regions of docking sites B′, B″, E and F are more strongly interconnected in hNgb than in mNgb. In mNgb the population densities look more concentrated after 2 and 10 ns of His64 dissociated dynamics. Regions showing a high total density (white isosurfaces) are only found in site E and B′, and in the snapshots at 10 ns also in B′. In return, the volume occupying site C (red cavity in Figures 3.3 and 5.3) has a higher total density in mNgb compared to hNgb. This is indicated by the appearance of the blue isosurfaces in this region of mNgb which are missing for hNgb.

5.3.2. Water population

The accessibility of docking sites within a protein is not only determined by their size and shape but also by their connectivity and hydrophobicity. In the following the fluctuation of the water occupancy over the complete 10 ns His64 dissociation MD simulations of mNgb and disulfide bound hNgb were analyzed to estimate how rapidly and to what extent the different cavities in Ngb are populated by external ligands. Figure 5.5 shows the sum of water molecules over all ten individual simulations found in each docking site of mNgb and Cys46-Cys55 bound hNgb.

None of the 20 different 10 ns trajectories initially have water molecules at sites B′ or B″, which are occupied by the iron-bound His64 side chain. After $\approx$ 500 ps of His64 dissociated dynamics the two sites start to get populated. Pocket B′, which is buried more deeply inside the protein structure, shows a maximum population of two waters in mNgb over the sum of ten trajectories. In hNgb B′ is even less frequently occupied (1 water/10 trajectories). B″ is generally more attractive to water than B′. The total population rises to 8 water molecules within 2.5 ns for hNgb and 7 water molecules within 4 ns for mNgb. Subsequently, the B″ population of hNgb

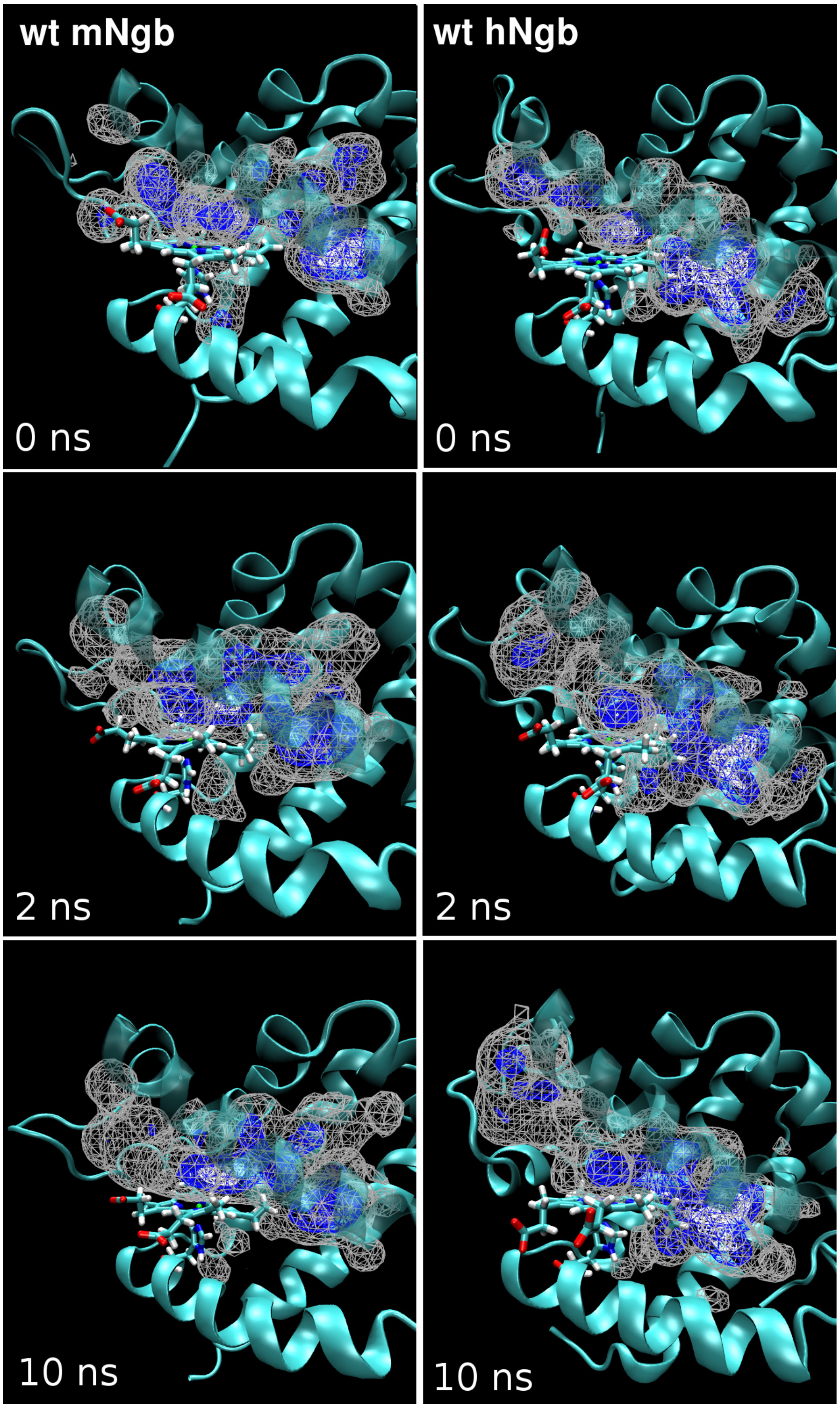

Figure 5.4.: Collection of snapshots showing the total cavity population densities of mNgb and hNgb at 0, 2 and 10 ns after the His64 dissociation. The isosurfaces describe the combined docking site volume V_{tot} as the relative population density of the total of ten independent trajectories for each protein. White: 25 %, blue: 66 % and grey: 100 % population density.

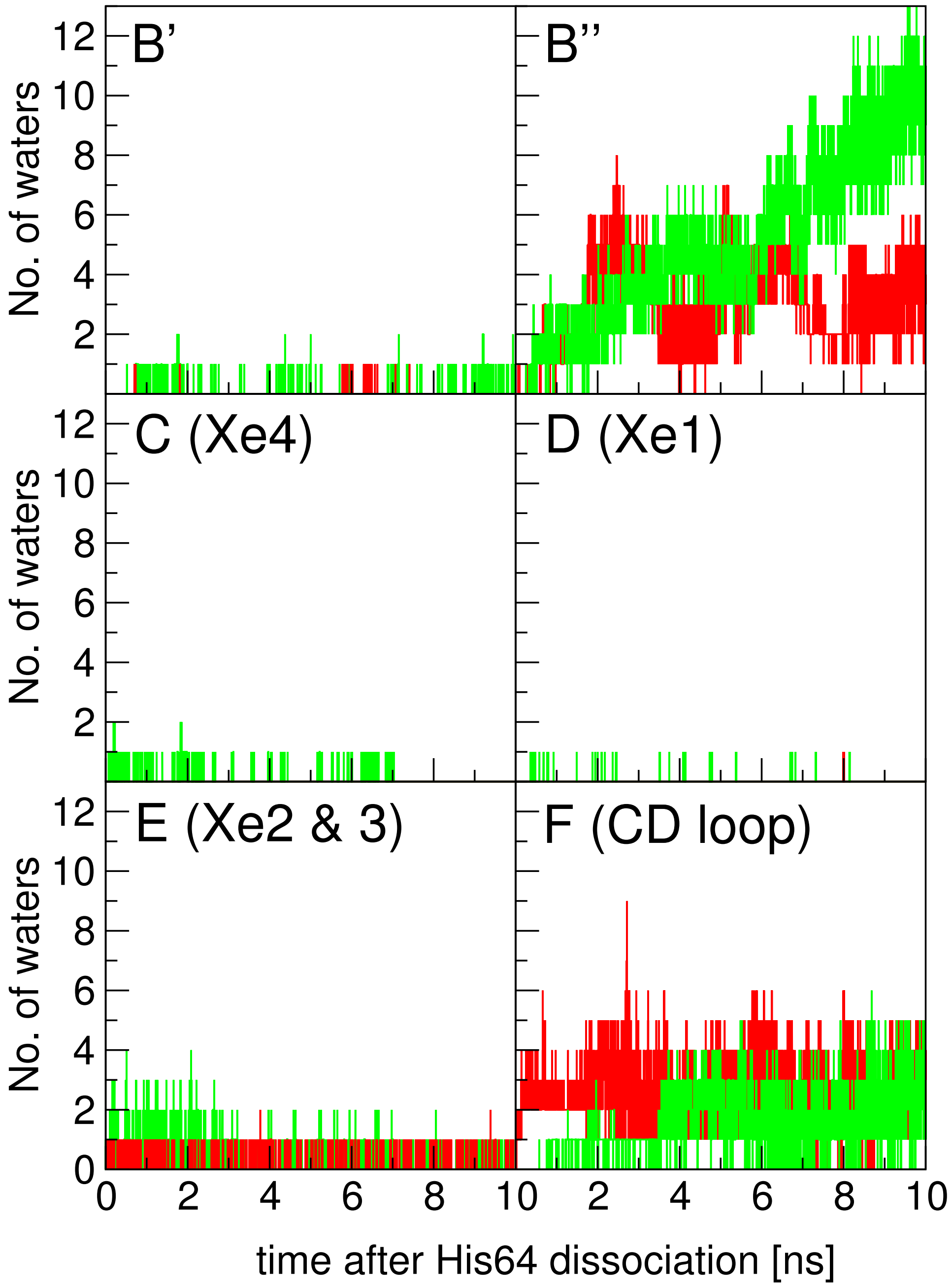

Figure 5.5.: Total amount of water molecules within each of the identified Ngb docking sites (Figure 3.3) along the ten individual 10 ns MD simulation of mNgb and hNgb. Green: mNgb timeseries, red: disulfide bound hNgb timeseries.

remains stable and fluctuates between 0 and 7 water molecules for the remaining simulation time with a periodicity of $\approx$ 2.5 ns. This strong fluctuation is an indication of a very fast water transition to and from this docking site. On the other hand, water molecules in the B″ pocket are getting slowly but continuously accumulated after the His64 dissociation as the occupation rises almost linearly up to 10 ns where a total amount of 13 water molecules are located within B″.

Sites C and D show only minor water population. In mNgb a maximum of two water molecules reach site C and remain there for $\approx$ 500 ps before leaving the pocket again. In hNgb site C does not get populated by water at all. Site D in mNgb and hNgb gets occupied by water molecules only for very short periods of time ($\leq$ 3 ps). It is indicative that site D has no or only a weak affinity for polar particles. In fact, both pockets (C and D) have hydrophobic methionine residues in immediate vicinity (Met69 in docking site C, Met144 in docking site D).

Site E (Xe2 & Xe3) is moderately occupied in both proteins already from the beginning of the His64 dissociation due to the large size of this cavity which allowed to accumulate water already within the heating and equilibration phase of the $\mathrm{Ngb_H}$ configurations. After $\approx$ 500 ps up to four water molecules occupy this pocket in mNgb occasionally. The rise originates from a number of water transitions between site E and sites C and D. After 2 ns the H_2O population drops down again to the range of 0 to 2 molecules by irrevocable water transfers into site C and subsequently into sites B′ and B″ which can be seen in the corresponding graphs of Figure 5.5. The water population of site E in hNgb remains stable over the complete simulation time and fluctuates in the same region. Differences in the H_2O population of sites C, D and E and transitions between these pockets for mNgb and hNgb is likely related to the more pronounced relaxations of the interhelical angle γ between helices E and F in mNgb (Figure 5.1E) compared to hNgb. The stronger tertiary structure rearrangements might be used to absorb water molecules more effectively in certain cavities.

Site F (CD-loop) in hNgb incorporates two water molecules in total prior to the E7-His dissociation. The population increases to six waters after 500 ps of dissociated MD. After 3800 ps a maximum of nine waters occupy the cavity which is nearly one water per protein. The same cavity in mNgb remains empty until 800 ps of His64 dissociation. After that, the water occupation increases only slowly until it reaches a maximum of six molecules after 8500 ps of simulation. These observation coincides with the bigger cavity volume of site F in hNgb compared to mNgb as identified above.

5.4. Solvent influx and the role of Tyr44

5.4.1. Water accessibility to the heme active site

To identify from which directions the water molecules preferably enter the rebinding environment B′ and B″ the two 10 ns bis-His dissociation trajectories with the strongest amount of water occupation in these sites for mNgb and disulfide bound hNgb were analyzed with the program Hop (250). The software identifies regions in the protein-bulk interface and inside the protein which show a high density of water molecules and separates these sites from the less dense bulk site region in a subsequent step.

Figure 5.6A shows two C_α aligned structures of mNgb (green) and Cys46-Cys55 bound hNgb (red) together with a network of specific high water density sites which connect the distal site cavity of the protein with the bulk in a most direct way. The water density sites between the two proteins are located at different positions. While in mNgb the water network starting in the vicinity of the heme leaves the active site preferentially via a channel directing towards the E-loop, the water density map in hNgb has most of its concentration located between the C- and F-helices. The same region in mNgb is occupied by the sidechain of Tyr44 and therefore blocks water influx from this side. Due to the presence of the disulfide bond in the CD-loop of hNgb Tyr44 aligns closer to the E-helix.

A more detailed picture on the connectivity between the different high water density sites is shown in Figures 5.7A and B for mNgb and in Figures 5.7C and D for hNgb. For each protein the sites are individually ordered by Roman numerals starting with the water site I located in the primary pocket B″ of the heme and progressing towards the bulk site. As these water sites do generally not overlap between the two proteins they are compared independently. Figures 5.7B and D show the transitions between all identified water density sites together with the exchange rates between those from

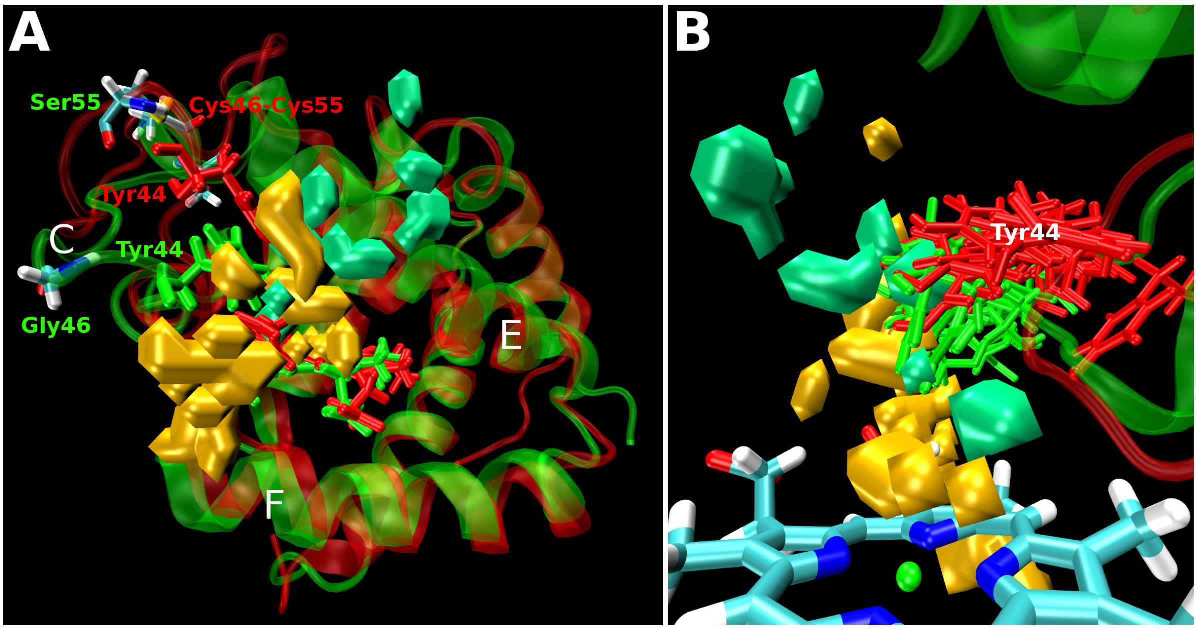

Figure 5.6.: A: Water density hopping networks reaching into the heme cavity constructed from one 10 ns bis-His dissociation MD simulation each for mNgb (green) and Cys46-Cys55 bound hNgb (red). For a comparison of the two proteins, the densities are plotted on top of the backbone C_α-aligned protein structures. B: The same water density networks shown from inside the heme cavity. All 10 final structures of the Tyr44 sidechain are shown for both proteins individually in their corresponding color-code. The structures were aligned according to the superimposition of all heme prosthetic groups.

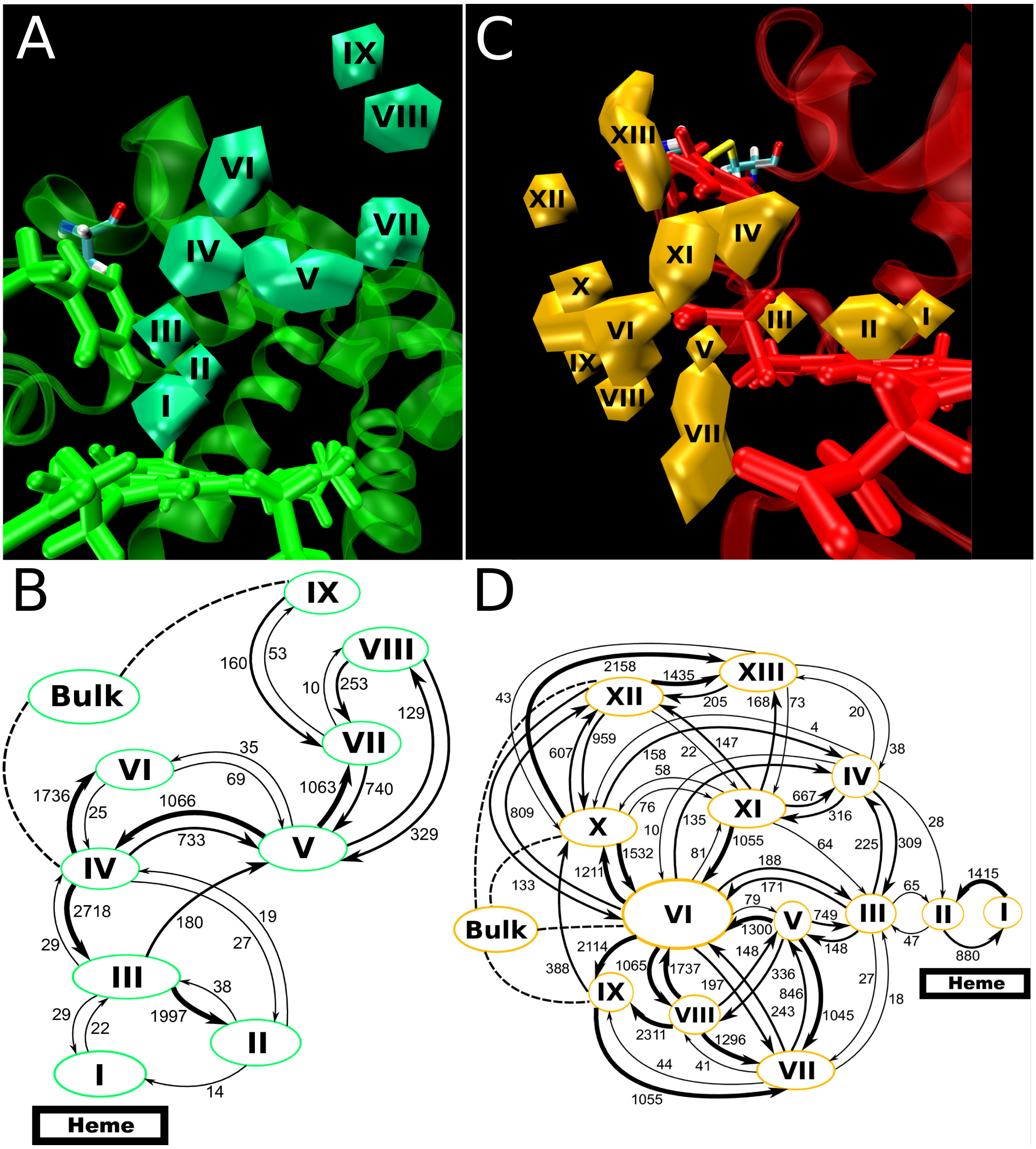
A
IX
VIII
VI
VII
IV
V
III
II
I
C
XIII
XII
IV
XI
X
VI
III
II
I
V
IX
VIII
VII
B
IX
Bulk
VIII
VII
VI
V
IV
III
II
I
Heme
160
53
10
253
129
35
69
1063
740
1736
25
1066
733
329
2718
180
19
27
29
38
1997
29
22
14
D
XIII
XII
XI
X
IV
VI
V
III
II
I
IX
VIII
VII
Bulk
Heme
2158
1435
205
168
73
20
43
607
959
147
4
38
158
22
58
667
316
809
76
135
10
64
225
309
28
1532
81
1055
188
133
1211
171
1415
79
65
749
1300
148
47
880
2114
148
388
1065
1737
197
336
27
846
18
243
1045
2311
1296
44
41
1055

Figure 5.7.: A: Detailed representation of the water density network discovered for mNgb. Densities are ordered according to the vicinity to the heme cavity using Roman numerals. B: Scheme of the water densities from A showing the interconnection between the different sites together with the calculated transition rates in s^{-1}. Transitions from the bulk site showing a large rate ($> 100\ s^{-1}$) are indicated in dashed lines. C: Same as A but for Cys46-Cys55 bound hNgb. D: Same as B but for Cys46-Cys55 bound hNgb. Transitions from the bulk into the specified sites showing a large rate ($> 200\ s^{-1}$) are indicated in dashed lines.

exponential fits to the site survival function

$$S_{ji} = < \Theta(t_{ij} - t) > \qquad (5.1)$$

where i and j correspond to two connected water sites. These hopping networks look rather complex with up to seventeen possible transitions to and from a specific site. For simplicity the following discussion focuses only on those transitions showing large exchange rates of $> 1000\ \mathrm{s}^{-1}$. As suggested above, the water influx towards the heme prosthetic group in mNgb (Figure 5.7B) origins mainly from water sites which are located near the E-helix on top of the heme plane. The fastest ingestion of bulk waters were found at sites IV (136 s^{-1}) and IX (115 s^{-1}). From both sites the migration occurs predominantly towards the heme plane with the largest rates found from site IV to III (2718 s^{-1}) and subsequently from III to II (1997 s^{-1}). The last hop into site I is drastically slower (14 s^{-1}) but irreversible in return. As the location of water density I corresponds to the one of the binding site B″ (Figure 3.3) the slow but continuous aggregation of H_2O molecules in this pocket (Figure 5.5) can be linked to the only weakly reversible (from III) water transfer into I. The same analysis for hNgb (Figure 5.7D) looks slightly more complex. Bulk water preferentially enters the water density network at sites VI (226 s^{-1}), IX (227 s^{-1}), X (308 s^{-1}) and XII (213 s^{-1}). The whole network can roughly be divided into one region located above, one below and one in plane with the heme rebinding site. These regions all intersect in density site VI which is located in hydrogen bonding distance to one of the heme propionates. Most of the densities of the lower region tend to have a large water hopping frequency into site VI and also among each other in contrast to the upper region where most of the transitions are rather slow. The only significant contributions to the water exchange rate from the upper region into site VI are from XI (1532 s^{-1}) and IX (1055 s^{-1}). In addition, almost every reversible transition found in the upper region has its equilibrium on the site which is closer to the bulk and pointing away from site VI. The last steps of water influx in hNgb is dominated by hops from site VI to V or III, site VII to VI and site IV to III. Finally, the water transfers from III to II or I which are both located within the primary docking site of the protein.

5.4.2. The role of Tyr44

The analysis of the water hopping network into the rebinding site B″ of Ngb discussed in the previous section revealed distinct differences in the hopping channels for mNgb and disulfide bound hNgb. Water density VI takes on an important role in the water transfer process of hNgb as it serves as the central connection between the solvent shell waters on the protein surface and the water densities V and III directing towards site B″ (see Figure 5.7D). Site VI is located next to one of the heme propionate sidechains which attracts water molecules due to its negative charge. In mNgb a high density of water molecules was similarly detected near the same heme propionate group. In contrast to hNgb it shows no contribution to the water hopping network pointing into the heme cavity and therefore its presence was neglected in Figures 5.6 and 5.7A and the related discussion. In the following, the lack of contribution of this water density site in mNgb will be explained by comparing the structural relaxation of the Tyr44 sidechain in mNgb and hNgb.

In the crystal structure of both, mNgb (PDB code 1W92) and hNgb (PDB code 1OJ6) the hydroxyl group of Tyr44 is in hydrogen bonding distance to one of the oxygen atoms of the heme propionate discussed here (see Figure 3.1). Figure 5.8A shows the averaged timeseries from the 10 ns $\mathrm{Ngb_H}$ dissociation simulations of the Tyr44-OH oxygen to heme propionate oxygen distance $r_{\mathrm{O-O}}$ which is initially smaller (to insure that the distance always refers to that propionate oxygen which initially describes the hydrogen bond to the hydroxyl group of Tyr44). At 0 ns $r_{\mathrm{O-O}}$ captures a hydrogen bonding distance in both proteins. On average $r_{\mathrm{O-O}}(0)$ equals 2.8 Å in hNgb and 3.4 Å in mNgb. The slightly larger average distance of $r_{\mathrm{O-O}}(0)$ in mNgb derives from the contributions of three trajectories which started from equilibrium structures with a slightly broken Tyr44 to heme propionate hydrogen bond. Following the complete 10 ns timeseries a similar relaxational behavior of $r_{\mathrm{O-O}}$ is observed in both proteins. $r_{\mathrm{O-O}}$ increases until it reaches convergence at around 8 Å after $\approx$ 5 ns of His64 dissociated dynamics. As the stability of a hydrogen bond does typically not exceed $\approx$ 4 kcal/mol, the disruption of the Tyr44 to heme interaction as a result of the other structural relaxations occurring through the $\mathrm{Ngb_H}$ to $\mathrm{Ngb_P}$ transition (Figure 5.1) is of no big surprise.

The nearly identical progression of $r_{\mathrm{O-O}}$ found in mNgb and hNgb can not explain a possible relation of Tyr44 to the different water density networks ranging into

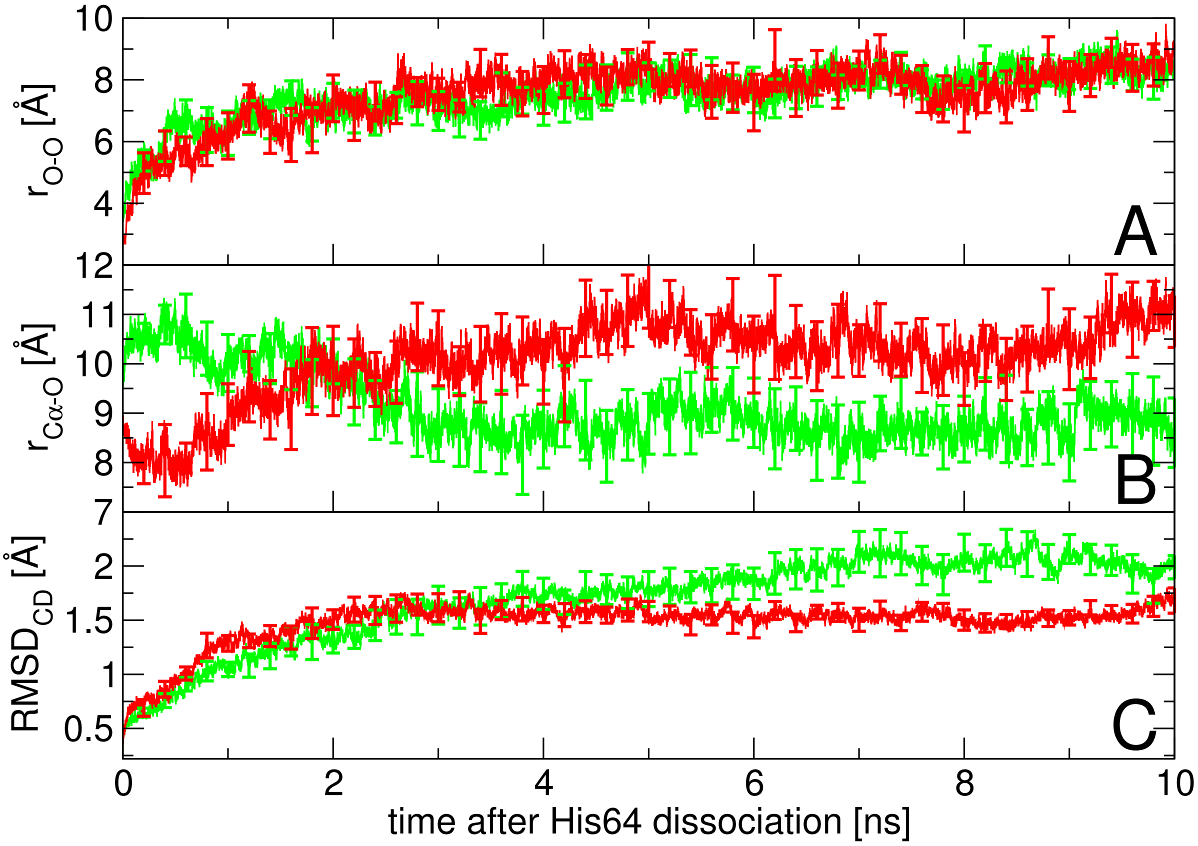

Figure 5.8.: Averaged timeseries extracted from the ten individual 10 ns bis-His dissociation MD simulations of mNgb (green) and Cys46-Cys55 bound hNgb (red). Error bars represent the standard error of each timeseries. A: Evolution of the Tyr44-OH oxygen atom to heme propionate oxygen atom which is initially in hydrogen bonding distance to Tyr44 r_{O-O}. B: Evolution of the Tyr44 C_α atom to center-of-mass of the two heme propionate atoms $r_{C_\alpha-O}$. C: Evolution of the CD-loop Cα RMSD (residues 44 to 59) aligned on the backbone C_α atoms of the complete amino acid sequence of every initial structure (Ngb_H).

the heme active site of both proteins. The protein structures of mNgb (green) and disulfide-bound hNgb (red) which are shown in Figure 5.6 correspond to snapshots taken at the end of each 10 ns MD simulation from which the individual water networks were constructed. The two structures are aligned along the C_α atoms of their protein backbone. Their Tyr44 sidechains have a large average RMSD of 3.7 Å between each other and therefore capture different orientations. In hNgb the C_α atom of Tyr44 is positioned closely to the DE-helical corner and its sidechain is pointing almost perpendicularly towards the heme plane. In mNgb the C_α of Tyr44 is more closely aligned to the heme plane and the sidechain orients more horizontally relative to the heme group. The orientation of Tyr44 in mNgb hinders the entrance of water from a potential site located near the heme propionate (site VI in Figure 5.7C and D) whereas for hNgb the formation of a water channel from this position into the heme pocket is feasible. To show that the picture of different Tyr44 orientations in mNgb and hNgb roughly correspond to the statistical average of the complete set of ten trajectories in each protein, the sum of all structures of both proteins at 10 ns after bis-His dissociation are presented in Figure 5.6B. All structures were RMSD-aligned according to the heme group atoms. The average RMSDs of the Tyr44 sidechain including only the mNgb structures, is 2.79 Å and 2.92 Å when only considering Tyr44 of hNgb. The average RMSD calculated for the combined structures is significantly larger (3.3 Å) which is an indication for clustering of this sidechain between the two proteins. This structural separation is observable in Figure 5.6B where almost all Tyr44 sidechains of mNgb (green) occupy a region closer to the heme group compared to the majority of the hNgb (red) Tyr44 residues. The latter ones leave a gap in the region between the heme propionate group and the CD-loop through which water can approach the heme cavity more easily compared to mNgb as it is indicated by the orange water densities. The progression for the Tyr44 separation between mNgb and hNgb after the bis-His dissociation can be followed in the ten-trajectory-averaged timeseries of the Tyr44-C_α to heme propionate oxygens center-of-mass distance $r_{C_\alpha - O}$ shown in Figure 5.8B. In the $\mathrm{Ngb_H}$ state, present at 0 ns, $r_{C_\alpha - O}$ is very similar in mNgb (9.5 Å) and hNgb (8.9 Å). After the His64 dissociation $r_{C_\alpha - O}$ diverges between the two proteins within $\approx$ 0.5 ns. The distance increases in mNgb up to 11.4 Å whereas in hNgb it decreases to 7.5 Å. After 2 ns both trajectories are converging again and cross each other at $r_{C_\alpha - O} \approx 9.8$ Å. Subsequently, $r_{C_\alpha - O}$ relaxes to approximately 8 Å in mNgb and to a notably larger distance of 11 Å in hNgb.

The 3 Å longer $r_{\mathrm{C}_\alpha-\mathrm{O}}$ distance in disulfide bound hNgb compared to mNgb can be attributed to a stronger repulsion of Tyr44 related to the close-by Cys46-Cys55 bond which implies an additional structural constraint in the CD-loop due to the reduced number of degrees of freedom in the disulfide bonded system. This behavior is shown in Figure 5.8C which describes the timeseries of the average RMSD of the C_α atoms of the CD-loop (residues 44 to 59, $\mathrm{RMSD}_{\mathrm{CD}}$) in each protein. Up to 2 ns the $\mathrm{RMSD}_{\mathrm{CD}}$ increases nearly linearly in mNgb and increases more slowly for the rest of the simulation. After approximately 8 ns a certain amount of convergence seems to have reached at $\mathrm{RMSD}_{\mathrm{CD}} \approx 2.1$ Å but fluctuations in the range of ± 0.25 Å are still observable. The initial increase of $\mathrm{RMSD}_{\mathrm{CD}}$ in hNgb is faster compared to mNgb due to the released strain implied through the Cys46-Cys55 bond. Consequently, convergence of $\mathrm{RMSD}_{\mathrm{CD}}$ in hNgb is already reached after approximately 3 ns. Subsequent fluctuations of this coordinate become significantly smaller compared to mNgb due to the reduced number of degrees of freedom of its CD-loop. This is also indicated by the error bars in Figure 5.8C which become, at this stage, explicitly weaker for hNgb whereas for mNgb the standard error of this averaged timeseries remains nearly unchanged over the complete simulation time.

The relevance of the Cys46-Cys55 bond to the dynamics of Tyr44 and the CD-loop it belongs to can be estimated by evaluating the energy landscape of the protein described along the Tyr44 separation from the heme. Figure 5.9 shows the free energy profiles of $r_{\mathrm{O-O}}$ for mNgb (green), hNgb including the Cys46-Cys55 bond (red) and hNgb with a broken disulfide bond (black) for the $\mathrm{Ngb_H}$ (solid lines) and $\mathrm{Ngb_P}$ (dashed lines) configurations. In all cases the profiles were constructed from consecutive umbrella sampling simulations starting in the hydrogen bonded distance of $r_{\mathrm{O-O}} = 2.7$ Å and the profiles were shifted to 0 kcal/mol for the local minimum showing up in this position. The $\mathrm{Ngb_P}$ profiles of all three proteins have relatively weak barriers to break the Tyr44-O-H $\cdots$ $^{-}$OOC-heme hydrogen bond which are in the range of $\Delta G^{\ddagger} = 0.4$ to 1.7 kcal/mol. This conforms to the immediately observed increase of $r_{\mathrm{O-O}}$ in Figure 5.8A during the transition from the $\mathrm{Ngb_H}$ to the $\mathrm{Ngb_P}$ state. After passing the hydrogen bond barrier, located between 3.2 and 3.5 Å on the free energy profiles, further elongation of $r_{\mathrm{O-O}}$ in the $\mathrm{Ngb_P}$ systems results in relatively flat potentials up to the limit of the calculated free energy scan of $r_{\mathrm{O-O}} = 9.0$ Å. At this coordinate the free energy profile of the Cys46-Cys55 broken $\mathrm{hNgb_P}$ is lowest in energy lying only 1.3 kcal/mol above the heme coordinated conformation. At the

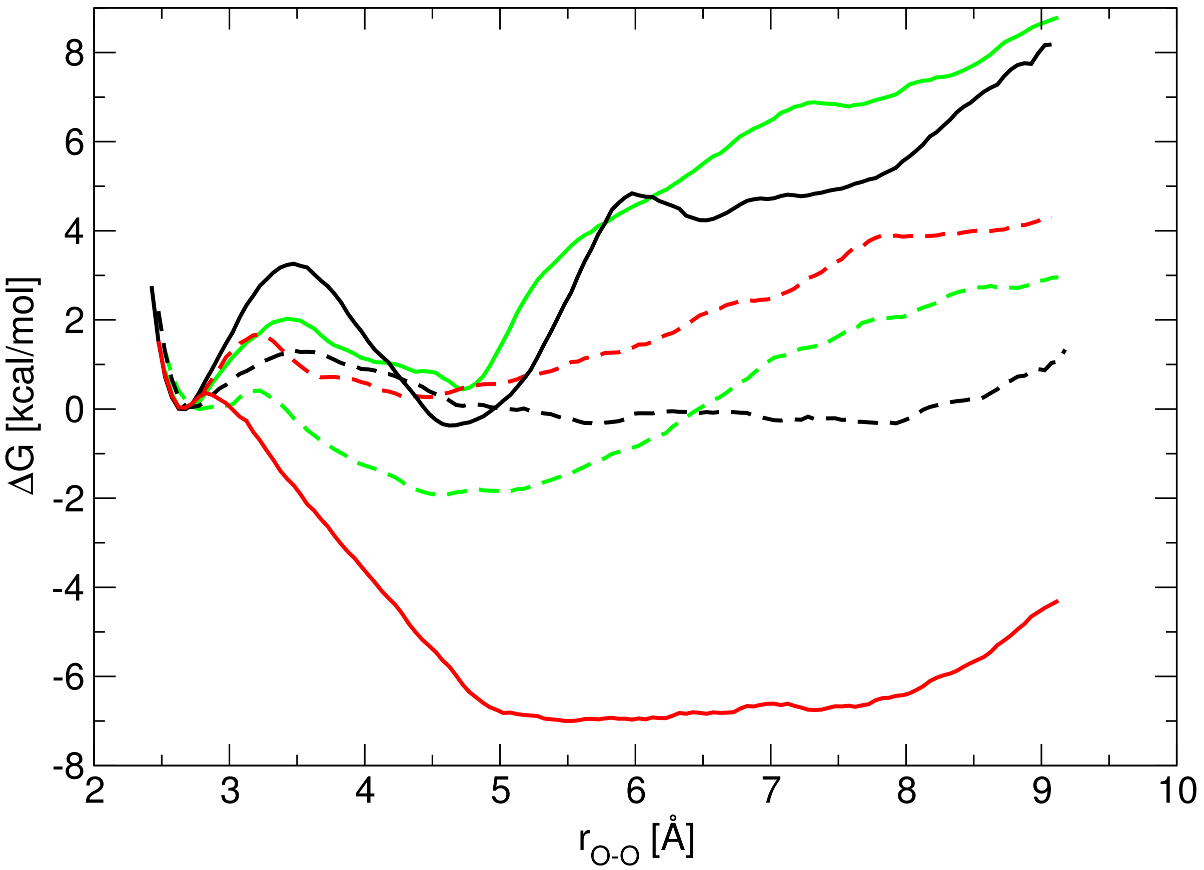

Figure 5.9.: Free energy profiles of mNgb (green), Cys46-Cys55 bound hNgb (red) and Cys46-Cys55 broken hNgb (black) along an elongation of the Tyr44-OH oxygen to heme-propionate oxygen distance r_{O-O}. Solid lines: Ngb_H configuration. Dashed lines: Ngb_P configuration.

same $r_{\mathrm{O-O}}$ mNgb$_{\mathrm{P}}$ is next higher in energy ($\Delta G = 3.0$ kcal/mol) and hNgb$_{\mathrm{P}}$ including the disulfide bond is the least stable one relative to the minimum at $r_{\mathrm{O-O}} = 2.7$ Å ($\Delta G = 4.3$ kcal/mol). The minima at more distant $r_{\mathrm{O-O}}$ are located at $r_{\mathrm{O-O}} = 4.6$ Å & $\Delta G = -2.0$ kcal/mol for mNgb$_{\mathrm{P}}$, $r_{\mathrm{O-O}} = 5.8$ Å & $\Delta G = -0.4$ kcal/mol for disulfide-less hNgb$_{\mathrm{P}}$ and at $r_{\mathrm{O-O}} = 4.4$ Å & $\Delta G = 0.2$ kcal/mol for Cys46-Cys55 bound hNgb$_{\mathrm{P}}$. In contrast to these rather small differences on the potential landscapes of the pentacoordinated Ngb species the free energy profiles of the bis-His coordinated systems are drastically different between Cys46-Cys55 bound hNgb$_{\mathrm{H}}$ and the proteins missing the disulfide bond. Both of the latter systems show two distinct minima at $r_{\mathrm{O-O}} = 2.7$ Å and $r_{\mathrm{O-O}} = 4.7$ Å which show more or less equivalent stabilities. The barrier connecting these two minima in disulfide-less hNgb is $\Delta G^{\ddagger} = 3.3$ kcal/mol whereas in mNgb it is 1.3 kcal/mol smaller. Beyond $r_{\mathrm{O-O}} = 4.7$ Å both potentials steeply rise up to $r_{\mathrm{O-O}} = 9.0$ Å where ΔG reaches approximately 8.5 kcal/mol with no significant indication for another stabilized conformation. Cys46-Cys55 bound hNgb$_{\mathrm{H}}$ has the smallest free energy barrier of $\Delta G^{\ddagger} = 0.3$ kcal/mol to break the Tyr44-O-H $\cdots$ $^{-}$OOC-heme hydrogen bond. After passing the transition state its potential experiences a strong energy stabilization down to $\Delta G = -7.0$ kcal/mol at $r_{\mathrm{O-O}} = 5.0$ Å where it roughly persists up to a distance of $r_{\mathrm{O-O}} \approx 7.5$ Å and after which it starts to rise again.

Figures 5.10A and B compare the structures of Ngb$_{\mathrm{H}}$ and Ngb$_{\mathrm{P}}$ taken from the umbrella sampling MD simulation at $r_{\mathrm{O-O}} = 5.0$ Å for mNgb (green), hNgb with (red) and without the Cys46-Cys55 bond (grey). Like in the final structure comparison of the bis-His dissociation MD simulations (see Figure 5.6B) hNgb including the disulfide bond arranges quite differently compared to mNgb. The RMSD deviation of the Tyr44 sidechains between the two proteins in Figures 5.10A and B are 4.17 Å and 3.59 Å. On the other hand, Tyr44 in disulfide-less hNgb aligns more closely to mNgb with a RMSD of only 1.79 Å (Ngb$_{\mathrm{H}}$) and 1.25 Å (Ngb$_{\mathrm{P}}$). This comparison confirms the strong influence of a Cys46-Cys55 disulfide bond on the orientational relaxation of Tyr44 after breaking the coordination to the heme propionate. As already shown in Figure 5.6B the orientation of Tyr44 present in Cys46-Cys55 bound hNgb makes the heme cavity more accessible for water entrance and therefore more likely for other ligands too.

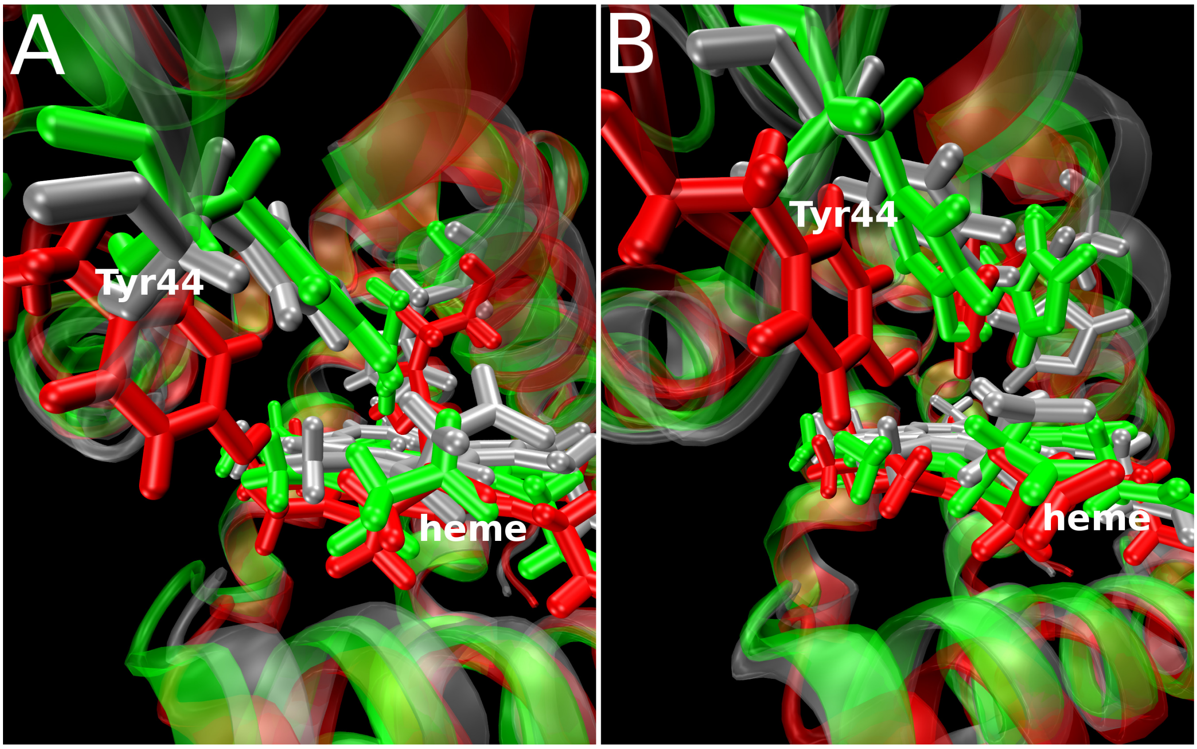

Figure 5.10.: A: C_α aligned structures from the umbrella sampling simulations along r_{O-O} taken at $r_{O-O} = 5.0$ Å for the $\mathrm{Ngb_H}$ configurations. Green corresponds to the mNgb system, red to the Cys46-Cys55 bound hNgb system and grey to hNgb with a broken disulfide bond. B: Same as A but for the $\mathrm{Ngb_P}$ configuration.

5.5. Biphasic rebinding barriers

As described in section 3.3 $\mathrm{Ngb_P}$ has two possibilities to undergo a hexacoordination when a exogenous ligand L is present inside the docking site network of the protein. Either the heme-iron binds to L or the hexacoordination occurs intramolecular with the His64 sidechain to form $\mathrm{Ngb_H}$. The latter configuration is native in wt Ngb and therefore most stable in the absence of L. In the following, the potential energy distribution and the free energy profiles are evaluated for the transitions between $\mathrm{Ngb_P}$ and $\mathrm{Ngb_H}$ or $\mathrm{Ngb_L}$, where L corresponds to CO, using ARMD in combination with umbrella sampling simulations (see section 5.1). Binding and unbinding barriers to $\mathrm{Ngb_H}$ are computed for mNgb, disulfide bonded hNgb and three mutants (V109L, V109F and V109W) of wt mNgb.

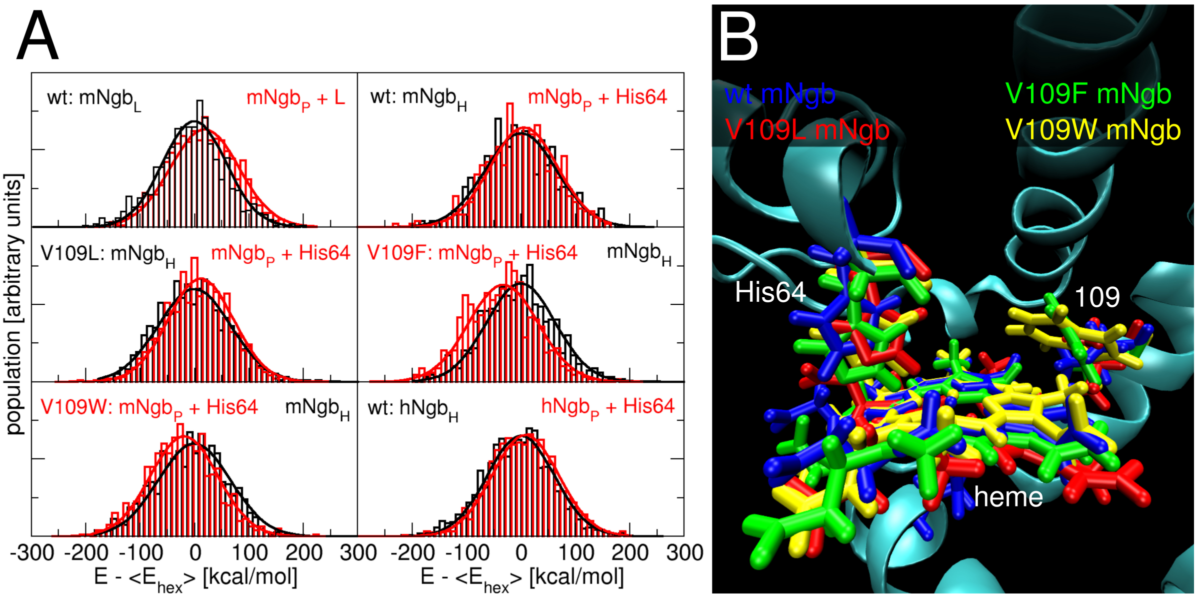

Figure 5.11.: A: Population analysis of the potential energy distribution in different heme configurations ($\mathrm{Ngb_L}$, $\mathrm{Ngb_P}$, $\mathrm{Ngb_H}$) and different protein species (wt mNgb, V109L mNgb, V109F mNgb, V109W mNgb, Cys46-Cys55 bound wt hNgb). Each distribution was calculated from an equilibrated 100 ps ARMD simulation where $\Delta_{\mathrm{Ngb_P}\leftarrow\mathrm{Ngb_H}} = -223$ kcal/mol and $\Delta_{\mathrm{Ngb_P}\leftarrow\mathrm{Ngb_L}} = 390$ kcal/mol. B: C_α aligned, $\mathrm{mNgb_P}$ snapshots of wt and all V109 mNgb mutants analyzed. Each set of distributions is aligned relatively to the mean potential energy $< E_{\mathrm{hex}} >$ in the corresponding hexacoordinated state ($\mathrm{Ngb_L}$ or $\mathrm{Ngb_H}$).

Individual rebinding barriers of His64, O_2 and CO on/off rates in Ngb were experimentally available by several groups (129, 145, 154, 157). The His64 and CO binding and dissociation rates of mNgb used for fitting $\Delta_{\mathrm{Ngb_P \leftarrow Ngb_H}}$ and $\Delta_{\mathrm{Ngb_P \leftarrow Ngb_L}}$ inhere are $k_{on}^{\mathrm{His}} = 2000\ \mathrm{s}^{-1}$, $k_{off}^{\mathrm{His}} = 1.2\ \mathrm{s}^{-1}$, $k_{on}^{\mathrm{CO}} = 72 \times 10^6\ \mathrm{s}^{-1}$ and $k_{off}^{\mathrm{CO}} = 0.013 \times 10^6\ \mathrm{s}^{-1}$ evaluated by Dewilde et al. (129). The rates can be put into relation to free energy of activation $\Delta G^{\ddagger}$ over Equation 4.15 and further to free energy of stabilization $\Delta\Delta G$ corresponding to equilibrium coefficients K. With this data in mind initial ARMD potential shifts $\Delta_{\mathrm{Ngb_P \leftarrow Ngb_L}}$ and $\Delta_{\mathrm{Ngb_P \leftarrow Ngb_H}}$ were estimated. Energy distributions of equilibrated MD simulations in each hexa- and pentacoordinated heme states are shown in Figure 5.11A for every studied transition represented as histograms. $\Delta_{\mathrm{Ngb_P \leftarrow Ngb_H}}$ is set to -223 kcal/mol and $\Delta_{\mathrm{Ngb_P \leftarrow Ngb_L}} = 300$ kcal/mol for these energy comparisons. In all graphs, expect for V109F and V109W mNgb, the hexacoordinated heme species is more stable on average, although the difference is generally very small compared to the total energy distribution of the trajectories. The energetical preference of V109F and V109W mNgb for the $\mathrm{Ngb_P}$ state becomes evident by a visual comparison of corresponding MD equilibrated protein structures shown in Figure 5.11B. Larger residues in position 109 have a stronger steric impact on the heme plane as with increasing size of the side chain the residue can cover additional space on top of the heme. These observations qualitatively agree with rebinding experiments where formation of $\mathrm{Ngb_H}$ could be weakened or even inhibited by replacing residue 109 with amino acid having larger side chains (from private communication with G. U. Nienhaus). The mean energy differences between hexa- and pentacoordinated Ngb states for each of the compared transitions in Figure 5.11A are from left to right and top-down: $+48.9$, $+5.0$, $+11.7$, -32.1, -19.0 and $+8.4$ kcal/mol.

The free energy profiles calculated with ARMD for both rebinding transitions is shown in Figures 5.12A and B. Experimental rate constants observed in the transitions of mNgb taken from Ref. (129) and converted into free energies of activation are represented as vertical bars in each graph of Figures 5.12A and B. Values of $\Delta_{\mathrm{Ngb_P \leftarrow Ngb_H}}$ and $\Delta_{\mathrm{Ngb_P \leftarrow Ngb_L}}$ obtained from the potential energy distribution in each equilibrium state (Figure 5.11A) were more thoroughly fitted to the experimental barriers of wt mNgb resulting in $\Delta_{\mathrm{Ngb_P \leftarrow Ngb_H}} = -145$ kcal/mol and $\Delta_{\mathrm{Ngb_P \leftarrow Ngb_L}} = 380$ kcal/mol. Quantitatively, the obtained theoretical barriers for the wt $\mathrm{mNgb_H} \rightleftharpoons$ wt $\mathrm{mNgb_P}$ transition, which are $\Delta G^{\ddagger}_{\mathrm{Ngb_H \rightarrow Ngb_P}} = 19.3$ and $\Delta G^{\ddagger}_{\mathrm{Ngb_H \leftarrow Ngb_P}} = 10.4$ kcal/mol, differ only by ≈ 2 kcal/mol compared to the experimental ones (Figure 5.12A). For the wt $\mathrm{mNgb_L} \rightleftharpoons$ wt $\mathrm{mNgb_P}$ transfers (Figure 5.12B) the barriers of $G^{\ddagger}_{\mathrm{Ngb_P \rightarrow Ngb_L}} =$

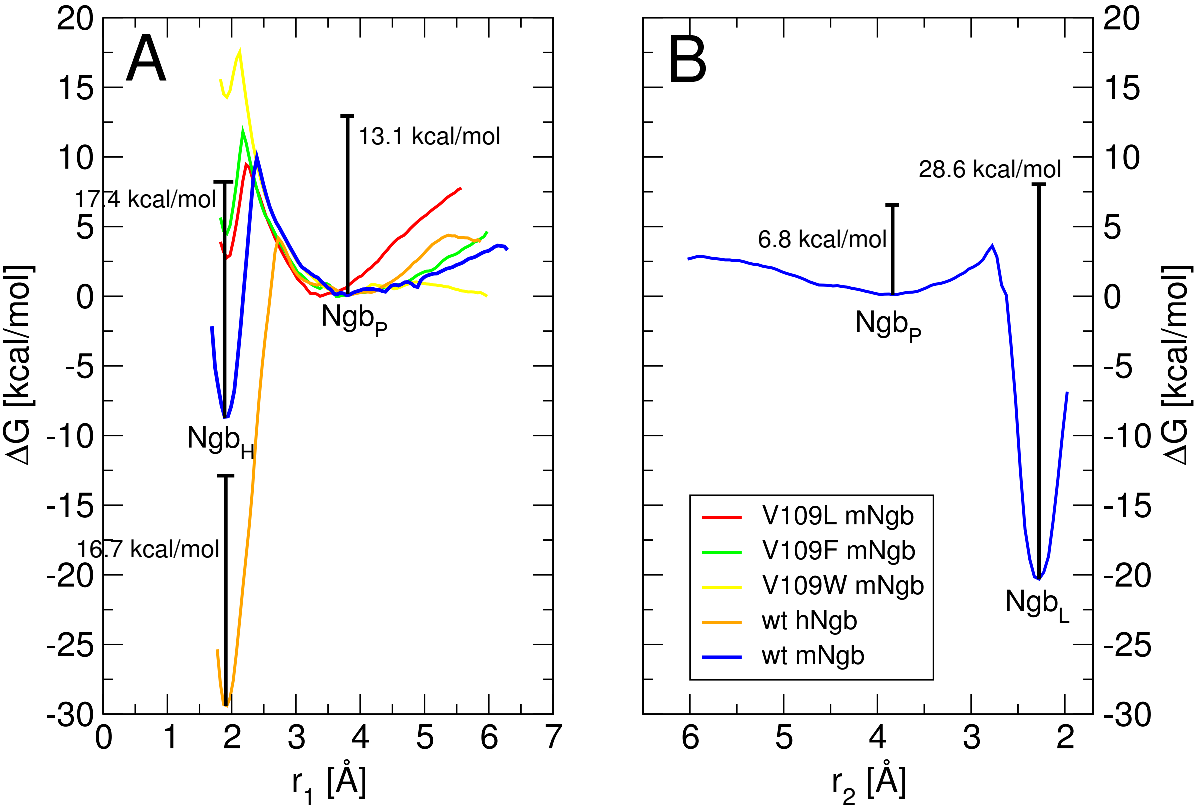

Figure 5.12.: A: Free energy profiles for the $\mathrm{Ngb_H}$ to the $\mathrm{Ngb_P}$ state along r_1 (described in Figure 3.2) calculated with ARMD. The different colors represent potentials of different mutations in position 109 and for disulfide bound wt hNgb. B: Free energy profile for the transition from the $\mathrm{Ngb_P}$ to the $\mathrm{Ngb_L}$ state calculated for wt mNgb along r_2. The bars represent corresponding barriers evaluated from experimental rate constants (129).

3.7 kcal/mol and $G^{\ddagger}_{\mathrm{Ngb_P \leftarrow Ngb_L}} = 24.1$ kcal/mol are both slightly underestimated by 3.0 and 4.5 kcal/mol compared to the experiment.

The free energy stabilization of the mutated $\mathrm{mNgb_H}$ and $\mathrm{mNgb_P}$ species follow similar trends as already observed from the equilibrium energy distributions in Figure 5.11A. While in wt mNgb the free energy minimum of the $\mathrm{Ngb_H}$ configuration is 9.0 kcal/mol more stable than the one of $\mathrm{Ngb_P}$, the V109L mutant shows a decrease in stability of $\Delta\Delta G = 2.6$ kcal/mol. In V109F and V109W mNgb the $\mathrm{Ngb_P}$ state is even 4.2 kcal/mol and 14.1 kcal/mol more stable than the $\mathrm{Ngb_H}$ one. The His64 rebinding barrier $G^{\ddagger}_{\mathrm{Ngb_H \leftarrow Ngb_P}}$ follow similar trends (wt mNgb: 10.6 kcal/mol, V109L mNgb: 9.6 kcal/mol, V109F mNgb: 12.3 kcal/mol, V109W mNgb: 17.6 kcal/mol).

In contrast to the stabilization of the $\mathrm{Ngb_P}$ state observed above for the valine 109 mutants of mNgb, the free energy profile of the disulfide bound $\mathrm{hNgb_H}$ to $\mathrm{hNgb_P}$ transfer in Figure 5.12A reveals a stability increase of the $\mathrm{Ngb_H}$ state compared to wt mNgb of a factor of ≈ 2. The stabilization of the $\mathrm{hNgb_H}$ state over the $\mathrm{hNgb_P}$ corresponds to $\Delta\Delta G = 29.4$ kcal/mol. The equilibrium potential energy distributions of $\mathrm{Ngb_H}$ and $\mathrm{Ngb_P}$ for mNgb and hNgb in Figure 5.11A do not show such a large difference. It is therefore strongly suggested that the reduced number of degrees of freedom in hNgb, due to the incorporation of the disulfide bond, is strongly influencing the entropic contribution of the free energy of activation in this transition. The experimental His64 dissociation rate constant of wt $\mathrm{hNgb_H}$ ($k^{\mathrm{His}}_{off} = 4.5\,\mathrm{s}^{-1}$) was determined to be larger compared to $\mathrm{mNgb_H}$ which corresponds to a slightly smaller dissociation barrier of $\Delta G^{\ddagger} = 16.7$ kcal/mol as indicated in Figure 5.12A. This is quantitatively and qualitatively different from the calculated free energy profile in Figure 5.11A which has a larger His64 dissociation barrier of $G^{\ddagger}_{\mathrm{Ngb_H \rightarrow Ngb_P}} = 33.5$ kcal/mol. This large difference might derive from inaccuracies in the description of the force field or other errors in the corresponding simulations but it needs additionally to be pointed out that the experimental rate constants are generally determined from CO or O_2 photolysis and stopped-flow experiments at different concentrations of these ligands. Their derived binding and unbinding rates of His64 are therefore only indirectly estimated from approximated rate laws applied to the rebinding rates of CO and O_2 (157). This indirect determination previously lead to stronger discrepancies in the assignment of k^{His}_{off} and k^{His}_{on} between different experimental studies (129, 145, 157). For example, Trent et al. (145) reported the equilibrium coefficient $K_H = k^{\mathrm{His}}_{off}/k^{\mathrm{His}}_{on}$ of states $\mathrm{Ngb_H}$ and $\mathrm{Ngb_P}$ to be in the order of 1 whereas Dewilde et al. found K_H to be smaller by a factor of one thousand (129). Furthermore, some studies reported k_{on}^{His} to be $\approx$ 2 times larger in hNgb (157) compared to mNgb while other studies assigned the exactly same rate constant to both proteins and both cysteine oxidation states in hNgb (154). However, in section 3.3 it was pointed out that experimentally disulfide bound hNgb experiences a faster rebinding to the $\mathrm{Ngb_H}$ state compared to disulfide unbound Ngb (154, 157). This conclusion was justified by the direct detection of a more pronounced slow phase CO rebinding taking place over a $\mathrm{Ngb_H}$ intermediate configuration (see Equation 3.1) in disulfide bound hNgb compared to disulfide-less mNgb. It means that the population of the $\mathrm{Ngb_H}$ state has increased faster in hNgb than in mNgb after the CO photolysis. In the calculated free energy profile of Figure 5.12A the His64 binding barrier of $G^{\ddagger}_{\mathrm{Ngb_H} \leftarrow \mathrm{Ngb_P}} = 4.1$ kcal/mol for $\mathrm{hNgb_H}$ is more than half as large compared to $\mathrm{mNgb_H}$. This corresponds to a faster binding rate of $k_{on}^{\mathrm{His}} = 6.44 \times 10^9\ \mathrm{s}^{-1}$ and qualitatively agrees with the experimental observation of a faster population of the $\mathrm{Ngb_H}$ state in Cys46-Cys55 bound hNgb.

5.6. Conclusion

The present analysis of non-equilibrium transitions occurring during the ligand rebinding competition of Ngb draws a multifaceted dynamical picture for its mechanism. Small differences in the structural dynamics between mNgb and Cys46-Cys55 bound hNgb can already be encountered in the equilibrium $\mathrm{Ngb_H}$ state where the disulfide bond in hNgb imposes a stronger structural constraint to the CD-loop residues and forces the sidechain of Tyr44 into a different alignment compared to mNgb. The structural arrangements initiated by the dissociation of His64 from the heme iron can be classified into the ones located at the heme group and into the more remote ones which occupy larger and more diffuse areas of the protein. Differences between the disulfide-bound and disulfide-less systems are even present in the former ones (e.g. heme sliding in Figure 5.1C) although they are not directly connected to residues 46 and 55. This implies a strong coupling of the disulfide normal mode to the heme normal modes. The coupling does not come into effect before hundreds of ps as one can see in the right-hand side graph of the same figure. Similarly, the evolution of angle γ between helices E and F (Figure 5.1E) of Cys46-Cys55 bound hNgb and the other proteins can not be distinguished before 1 ns of His64 dissociation.

The different structural evolution between the analyzed protein species has also an impact on the evolution of the internal cavity network and the water diffusion into the protein and among its different sites. This is most strongly visible for cavities B′, B″, C and F which form the distal site cavity. As was shown in Figure 5.2, site F is constantly larger by ≈ 20 Å^3 over the complete 10 ns His64 dissociation simulation in disulfide bound hNgb compared to mNgb and its size remains untouched from any relaxational equilibration in the former protein. This also affects the water influx into site F which gets initiated immediately after the His64 dissociation in hNgb (Figure 5.5) whereas mNgb needs roughly 1 ns to accumulate water molecules in this site. All these deviations can, to a large extent, be assigned to the disulfide bond forming residues Cys46 and Cys55 which are part of this docking site. The volume of cavity B″ increases in both proteins with nearly the same speed and up to the first 5 ns of dissociated MD with roughly the same absolute volume. On the other hand, the strong volume increase relative to the $\mathrm{Ngb_H}$ state observed in Cys46-Cys55 bound hNgb likely induces a strong pulling force on external water molecules to enter this docking site faster in this protein compared to mNgb. This effect can be seen in Figures 5.5 and 5.7D. The structural context for these differences are firstly situated in the initially slower but, after a time lag of 2 ns His64 dissociated MD, faster and more pronounced sliding of the heme plane into the proximal cavity in disulfide bound hNgb relatively to the disulfide-less proteins. Secondly, the preferential reorientation of the Tyr44 sidechain into a position located closer to the CD-loop in hNgb (Figure 5.6B) increases the volume of the B″ cavity and simultaneously opens the channel for bulk water influx into the site. The structural origins of the relative volume increase of docking site B′ after the His64 dissociation observed for mNgb is less clearly assessable from the present analysis. A possible determinant is the angular evolution α between vectors $\vec{n}_0$ and $\vec{n}_t$ (Figure 5.1D) which describes a strong bending of the heme plane in mNgb and by this stronger fluctuations in the adjacent B′ cavity.

Finally, it should be pointed out that the present chapter has demonstrated how important even small structural differences in the same protein family can become when looking at dynamical processes. Even if these differences have only a small impact between two equilibrated proteins, they can initiate a chain of events when instantaneously put into an non-equilibrium state. Factors like structural constraints, protein to solvent interactions, shaping, size, connectivity and ligand occupancy of internal protein cavities and ligand diffusion are in a strong interplay with each other and have also a high impact on measurable quantities like ligand binding. It is of a huge interest to investigate these factors which trigger the protein specific ligand affinities to gain more detailed insight into the mechanism of ligand binding and unbinding in globin proteins.

6. Published Work and Manuscripts Submitted for Publication

"Something I must have done or he would not have changed so; would not so entirely have suspended our intimacy. Those long evening rambles when he ceased from his labours, and opened to me such new worlds of thought and enjoyment; poured out all that knowledge of human life and achievement–which made these few hours span the bright life of ages; why has it all ceased? If my pride has forbade my asking, still is the change generous? When he knows–when he must know, how dear he has become to me–these are creditable tears that I should shed them without cause."

William Bayle Bernard, *The Tide of Time, 1859*

6.1. Ligand migration between internal docking sites in photodissociated carbonmonoxy neuroglobin

The following paper appeared as

S. Lutz, K. Nienhaus, G. U. Nienhaus and M. Meuwly, *J. Phys. Chem. B*, **113(46)** 15334 – 15343 (2009).

15334 *J. Phys. Chem. B* **2009,** *113,* 15334–15343

Ligand Migration between Internal Docking Sites in Photodissociated Carbonmonoxy Neuroglobin

Stephan Lutz,[†] Karin Nienhaus,[‡] G. Ulrich Nienhaus,[*,‡,§,||] and Markus Meuwly[*,†]

Department of Chemistry, University of Basel, Klingelbergstrasse 80, 4056 Basel, Switzerland; Institute of Biophysics, University of Ulm, Albert-Einstein-Allee 11, 89081 Ulm, Germany; Institute of Applied Physics and Center for Functional Nanostructures, Karlsruhe Institute of Technology, Karlsruhe, Germany; and Department of Physics, University of Illinois at Urbana−Champaign, 1110 West Green Street, Urbana, Illinois 61801

Received: June 17, 2009; Revised Manuscript Received: September 14, 2009

Neuroglobin (Ngb) belongs to the large family of globular heme proteins capable of binding small gaseous ligands such as O_2, CO, or NO within their active site. In this work, we have analyzed CO migration pathways in photolyzed NgbCO using molecular dynamics (MD) simulations in combination with Fourier transform infrared temperature derivative spectroscopy (FTIR-TDS). A total of 55 ns of MD simulation was analyzed to explore the ≈300 Å^3 internal Ngb cavity. Overall, the simulations differentiated between eight possible docking sites, three of which were also identified experimentally. Low-temperature FTIR-TDS experiments on wild-type (wt) and F28W mutant NgbCO revealed that a small fraction of ligands migrates from site B to site C from which they rebound after slow cool illumination. For the F28L mutant, however, population of site C was not observed. In agreement with these findings, the simulations at 20 K showed ligand transfer between sites B and C for wt Ngb, but not for the F28L mutant. The ligand migration network could be mapped out and two key gate residues, Phe28 and Pro52, were identified. Ligand population analysis from the MD simulations revealed a direct relation between the size of the B10 side chain (Phe28 in wild-type Ngb) and the barrier against migration. Barriers for the transition of photodissociated CO from the distal pocket to the Xe4 site in Ngb are lower by up to 4 kcal/mol compared to myoglobin, suggesting that ligand migration between different docking sites is more facile in Ngb than in myoglobin.

I. Introduction

Neuroglobin (Ngb), a small heme protein from the globin family, is expressed in neuronal cells of vertebrates.[1] The physiological function of Ngb is still under debate.[2] Low O_2 affinity under physiological conditions compared to myoglobin (Mb) and hemoglobin (Hb) and a higher autoxidation tendency disqualifies Ngb as a O_2 transporter or reservoir.[3] Greenberg et al. suggested a role in neuroprotection.[4] They revealed that both neuronal hypoxia and cerebral ischemia induce the expression of Ngb. Other suggestions of alternative functions include the signaling of hypoxia[5] and radical scavenging.[6,7]

Primary structure comparisons of Ngb with the structurally related Mb and Hb show less than 25% sequence identity.[8] Only ∼0.4 × 10^{-9} amino acid changes/year are observed within different mammalian sequences of Ngb, whereas a 3-fold higher rate is observed for Hb. Additionally, residue conservation of more than 46% between Ngbs from different vertebrates suggests tight constraints on amino acid replacements during evolution to fulfill functional requirements. Despite the low sequence identity, both Ngb and sperm whale Mb display the same characteristic three-over-three globin fold[9] (Figure 1), with a rmsd of the C_α atoms of 2.0 Å. In the ferric form, the Ngb heme iron is hexacoordinated by the distal and proximal histidines, His64 and His96. In contrast, the sixth coordination site of Mb is occupied by a water molecule (hydroxide ion at high pH). Upon reduction, the iron remains hexacoordinated in Ngb, whereas the Mb iron becomes pentacoordinated and thus accessible for dioxygen. For an exogenous ligand such as CO to bind to Ngb, the His64-iron bond has to be ruptured.[10−12] Subsequently, the heme plane slides deeper into a spacious internal cavity, which connects the distal and proximal heme sides.[13,14] The presence of this large open 300 Å^3 volume is energetically unfavorable and therefore suggests that it is predesigned to allow for a favorable interaction between heme and ligand, for instance by providing transient ligand storage sites. Recent Fourier transform infrared (FTIR) experiments have

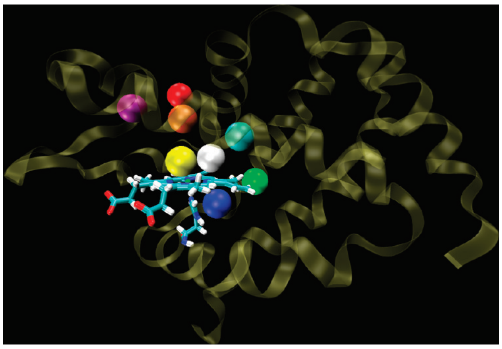

Figure 1. Structure of Ngb showing the docking sites identified by MD simulations as colored spheres. Color code: white (site B′), yellow (site B″), cyan (site C, Xe4), blue (site D, Xe1), green (site E, Xe2), orange (site F), red (site G), and purple (site H).

---

[*] To whom correspondence should be addressed.
[†] University of Basel.
[‡] University of Ulm.
[§] Karlsruhe Institute of Technology.
[||] University of Illinois at Urbana−Champaign.

10.1021/jp905673p CCC: $40.75 © 2009 American Chemical Society
Published on Web 10/23/2009

distinguished two intermediate sites in Ngb. However, the spectroscopic data could not provide structural information on these sites.[15]

In Mb, four small hydrophobic cavities exist.[16] These so-called xenon cavities (Xe1, Xe2, Xe3, Xe4) were shown to be part of the ligand migration pathway through the protein matrix.[17−21] These sites enable the ligand to remain unbound within the protein to wait for large-scale fluctuations that open up escape pathways. Blockage of these sites changes the ligand binding kinetics and thus the physiological function.[22] In the truncated hemoglobins, ligand access to and escape from the active site is realized by a hydrophobic tunnel. This direct pathway results in a 15-fold faster oxidation kinetics of NO as compared to horse heart Mb.[23−25] As the scavenging of NO and other radicals is one of the proposed functions of Ngb, the determination of its ligand migration network and the migration dynamics is expected to be of fundamental importance.

In the present work, we combine atomistic simulations and FTIR spectroscopy experiments to elucidate possible CO migration pathways in Ngb. Molecular dynamics (MD) simulations reveal a network of transient docking sites. The relative stability of each identified docking site is quantitatively characterized by umbrella sampling. To complement the theoretical work, temperature-derivative spectroscopy (TDS) experiments were performed at cryogenic temperatures. This technique has proven to be an excellent tool for elucidating the number of docking sites. Finally, the findings from theory and experiment are compared and discussed.

II. Theoretical and Experimental Methods

A. Molecular Dynamics Simulations. All molecular dynamics simulations were carried out with the CHARMM program[26] and the CHARMM27 force field.[27] Initial carboxy Ngb (NgbCO) coordinates were taken from the X-ray structure by Nienhaus and co-workers[13] (PDB-code 1W92). Hydrogen atoms were added to form both possible neutral tautomers of His64, $\mathrm{His}_\delta 64$, and $\mathrm{His}_\varepsilon 64$. Simulations were carried out for both protonation states also because in myoglobin the protonation state of His64 has been somewhat controversial. Recent work, however, found $\mathrm{His}_\varepsilon 64$ to be the more likely one in Mb.[28−30] For Ngb, recent experimental and spectroscopic studies also favor $\mathrm{His}_\varepsilon 64$.[31] Additionally, single mutants F28L and F28W were generated by replacing Phe28 of wild-type (wt) $\mathrm{His}_\delta 64$ NgbCO with the corresponding amino acid. The simulation is focused on the protein active site consisting of the CO-binding heme group and its immediate environment; therefore, the stochastic boundary method was used to increase computational efficiency.[32] The heme pocket was solvated by three sequential layers of water molecules, centered on the center of mass of the four porphyrin nitrogens of the heme prosthetic group, and a solvent boundary potential with a radius of 25 Å was applied to constrain the water molecules. A "reaction region" of radius 16 Å around the heme was defined, inside which the system was propagated with Newtonian dynamics. The dynamics of the buffer region between 16 and 20 Å from the center was described using Langevin dynamics. After solvation, the wild-type systems consisted of 2465 atoms of NgbCO and 1450 water molecules, which leads to 6815 atoms in total.

The interatomic stretching potential of the photodissociated CO ligand is given by an anharmonic rotational Rydberg−Klein−Rees (RRKR) potential.[33,34] To describe the electrostatic interactions, the fluctuating three-point-charge model, which involves placing charges on the C and O atoms and an additional point charge at the center of mass of CO[35,36] was applied to the CO in the unbound state. This model is inspired by an earlier fixed three-point-charge model.[37]

Initially, NgbCO was heated from 100 to 300 K for 50 ps and equilibrated for 100 ps. The final rmsd of the C_α atoms with respect to the crystal structure was 1.7 Å. The photodissociated system (Ngb−CO) was prepared following an additional equilibration simulation (200 ps) by storing 20 snapshots, each separated by 10 ps, from which the photodissociation event was initiated. Photodissociation was invoked by the "sudden" approximation:[38] The Fe−C bond is deleted and the potential parameters describing the bound state are replaced by those of the dissociated state. A repulsive term of the form r^{-12}, where r is the distance between heme iron and the CO center of mass, was added and all nonbonded interactions (electrostatic and van der Waals) between the CO and the heme plane were switched off. Force field parameters were taken from previous work on MbCO.[38,39] After 0.1 ps of dynamics, the Fe−C bond was fully dissociated. The repulsive term was removed and the nonbonded interactions were reintroduced. Subsequently, individual 1 ns trajectories (20 wt $\mathrm{His}_\delta 64$ Ngb−CO, 15 wt $\mathrm{His}_\varepsilon 64$ NgbCO, 10 F28L Ngb−CO, and 10 F28W Ngb−CO, respectively) of photodissociated Ngb−CO were run at 300 K.

Umbrella sampling MD simulations were carried out for wt $\mathrm{His}_\delta 64$ and $\mathrm{His}_\varepsilon 64$ Ngb−CO and the corresponding F28L mutant starting from different snapshots along a transfer observed in the 1 ns unbiased MD simulations. Each free energy profile was calculated from 10 to 15 overlapping windows with a simulation time of 100 ps each. Merging of the windows was performed by the weighted histogram analysis method (WHAM).[40]

To analyze the influence of temperature on the migration network, 10 snapshots of the 300 K simulations of wt $\mathrm{His}_\delta 64$ and F28L $\mathrm{His}_\delta 64$ NgbCO were cooled to 20 K within 50 ps and equilibrated for another 30 ps. Subsequently, the ligand was dissociated as described above. As the *B*-factor analysis showed only small fluctuations of the protein backbone ($\lesssim 7.0$ Å^2 at 20 K) the protein atoms were frozen in these simulations. Due to the reduced number of degrees of freedom, the MD simulations of the dissociated species could be extended to 100 ns.

B. Experimental Procedures. ***Sample Preparation.*** Cysteine-depleted murine Ngb (C55S/C120S) is referred to as wt Ngb in the following. It was used as a template for site-directed mutagenesis, using the Quikchange mutagenesis kit (Stratagene Europe, Amsterdam). Custom-designed primers were ordered from MWG (MWG-Biotech GmbH, Ebersberg, Germany). The mutant proteins were expressed in *E. coli* and purified according to the published protocol.[41] For the FTIR experiments, lyophilized protein was dissolved at a concentration of ∼15 mM in cryosolvent (75% glycerol/25% potassium phosphate buffer (v/v), pH 8), stirred under a CO atmosphere, and reduced with a 2-fold molar excess of sodium dithionite solution.

Temperature-Derivative Spectroscopy (TDS). TDS is an experimental protocol designed to investigate thermally activated rate processes with distributed enthalpy barriers.[42−44] Such distributions have been inferred from the nonexponential rebinding of ligands in heme and other proteins at cryogenic temperature; they result from conformational heterogeneity.[45,46] The sample is photodissociated prior to the TDS experiment, and the measurement is started at the lowest temperature chosen. FTIR transmittance spectra, $I(v,T)$, are taken every 1 K while increasing the sample temperature T at a rate of 0.3 K/min. FTIR

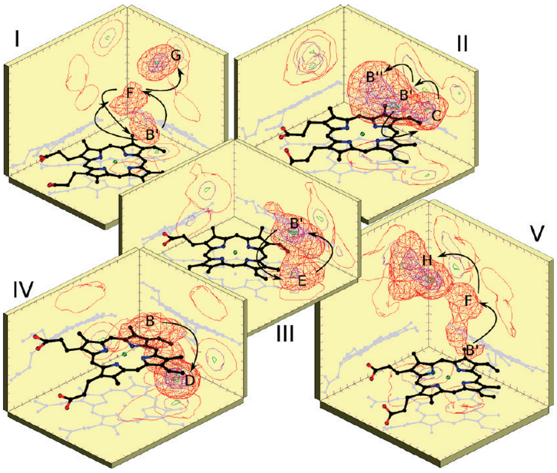

Figure 2. Typical probability distributions of CO migration between specific docking sites in Ngb−CO. The heme is shown as a wire frame with the central iron atom as a green sphere. Arrows indicate migration pathways. Projections onto the *xy*, *xz*, and *yz* planes are also given. Panel I: migration from site B′ over F to G. Panel II: starting from site B′ to B″ back to B′ and finally to site C. Panel III: transfer from site B′ to E. Panel IV: from site B″ quickly into docking site D of F28L Ngb−CO. Panel V: from site B′ to H.

absorbance difference spectra at frequency ν, $A(\nu,T)$, are calculated from transmittance spectra at successive temperatures:

$$A(\nu, T) = \log\left(\frac{I(\nu, T + 1/2K)}{I(\nu, T - 1/2K)}\right) \qquad (1)$$

For ligand rebinding, the change in spectral area is taken proportional to the fraction of ligands that rebind during acquisition of two successive spectra. However, absorption changes may also arise from ligand dynamics within a docking site,[47,48] ligand migration to different docking sites,[22,49] and conformational changes of the protein.[50] The temperature ramp protocol ensures that rebinding occurs sequentially with respect to the temperature at which the different processes become activated on the time scale of the experiment. TDS is a so-called “rate-window method” which shifts all processes sequentially into a rate (or time) window that can be controlled by the temperature ramp rate which, in our experiments, is typically $\sim$0.01 s^{-1} (100 s).[43,51] TDS data are usually displayed as contour plots of the absorbance change on a surface spanned by the wavenumber and temperature axes. Black/red contours indicate an absorption increase/decrease, and logarithmic contour spacing is chosen to emphasize small features.

III. Results

A. Molecular Dynamics Simulations. ***Docking Site Network.*** Molecular dynamics simulations at 300 K show that, after photodissociation, the CO ligand samples eight well-defined regions within the protein on the nanosecond time scale (colored spheres in Figure 1). Density plots of wt His$_\delta$64 Ngb−CO (panels I, II, III, and V in Figure 2) and F28L Ngb−CO (panel IV in Figure 2) provide representative CO populations, each from an independent 1 ns simulation. Corresponding time series for the migration of CO are provided in Figure 3. The graphs show the x, y, and z components of the time series along $\vec{r}$, where $\vec{r}$ is the vector pointing from the Fe atom to the CO center of mass. For some of the detected sites, a corresponding site exists in Mb.[19,52] Thus, we adopt the customary nomenclature B, C (Xe4), and D (Xe1) for these sites also for Ngb. Additional docking sites unique to Ngb follow in alphabetical order. In Table 1, the relative CO docking site populations from the complete set of unbiased MD trajectories are summarized.

After photodissociation, the CO ligand in wt His$_\delta$64 Ngb predominantly migrates to a primary docking site B. Occupation times typically range from 50 to 500 ps although, in a few cases, the ligand spends an entire 1 ns (full length of trajectories) in this region. In wt and F28W Ngb−CO, the primary docking site B actually consists of two sites, B′ and B″ (white and yellow spheres in Figure 1). The former is displaced by ≈1.8 to 2.8 Å from the heme center, whereas the latter is located on top of the heme iron. The relative populations of B′ and B″ are related to the dynamics of residue Phe28, which is found to play the role of a gate. If the ligand is in B′, a translation or rotation of Phe28 relative to the heme is required to allow transfer from B′ to B″ (Figure 4A). The time series depicted in black (top graph in Figure 4C) indicates a small (≈1 Å) and very short lifting of Phe28 from the heme plane 370 ps after photodissociation. Concomitantly, the Phe28 side chain adopts a more parallel orientation relative to the heme plane (≈50°) 30 ps prior to ligand migration and relaxes again to ≈70° after the transfer is accomplished (bottom graph in Figure 4C). If residue 28 moves toward the protein interior, the ligand can diffuse into site C (Xe4), as seen from the two selected trajectories involving migration between sites B and C of wt His$_\delta$64 Ngb in Figure 4C. As shown in Table 1, the CO molecule spends more time in site C with increasing bulkiness of residue 28.

Even the two wt systems (His$_\delta$64 and His$_\varepsilon$64) show differences in their relative docking site populations (Table 1). The CO population in site C is 9% smaller for His$_\varepsilon$64 than for His$_\delta$64 because the population density of the former system redistributes to secondary docking sites F, G, and H on the distal side of the protein. The redistribution originates from an isomerization of His$_\varepsilon$64 during equilibration, which is apparent from the average time series after CO dissociation shown in Figure 5. The imidazole side chain of His$_\varepsilon$64 rotates by 180° compared with respect to the 1W92 crystal structure. This conformation allows the N$_\delta$ atom to form a hydrogen bond to the solvent; the N$_\varepsilon$ proton points toward the heme iron, which leads to a stable His64 conformation. The presence of the protonated N$_\varepsilon$ close to the heme iron appears to promote ligand transfer to sites F, G, and H. In contrast, the conformationally less stable His$_\delta$64 imidazole side chain relaxes after CO dissociation into a position where its ring is aligned in a rather parallel fashion to the heme plane. The rearrangement favors secondary docking site C over site F.

Sites H and G are both directly populated by CO transfer from site F (Figure 6). Here, residue Pro52 is found to play the role of a gate, similar to residue Phe28. Depending on its actual conformation, the Pro52 side chain occupies either site G or site H. Hence, for a ligand transfer from site F to site G (H) to occur, the side chain of Pro52 has to move into site H (G). Figure 4B shows three aligned snapshots with CO in docking sites F, G, and H. Site H coincides with the small, 13.3 Å^3 cavity near residue Ser55 that was found in the crystal structure of CO-ligated Ngb.[53]

Figure 6 summarizes the docking site network derived from the combined 55 ns of unbiased MD simulation. The transfers observed in MD simulations are represented by arrows. Solid arrows describe frequent transfers (>15), long-dashed arrows

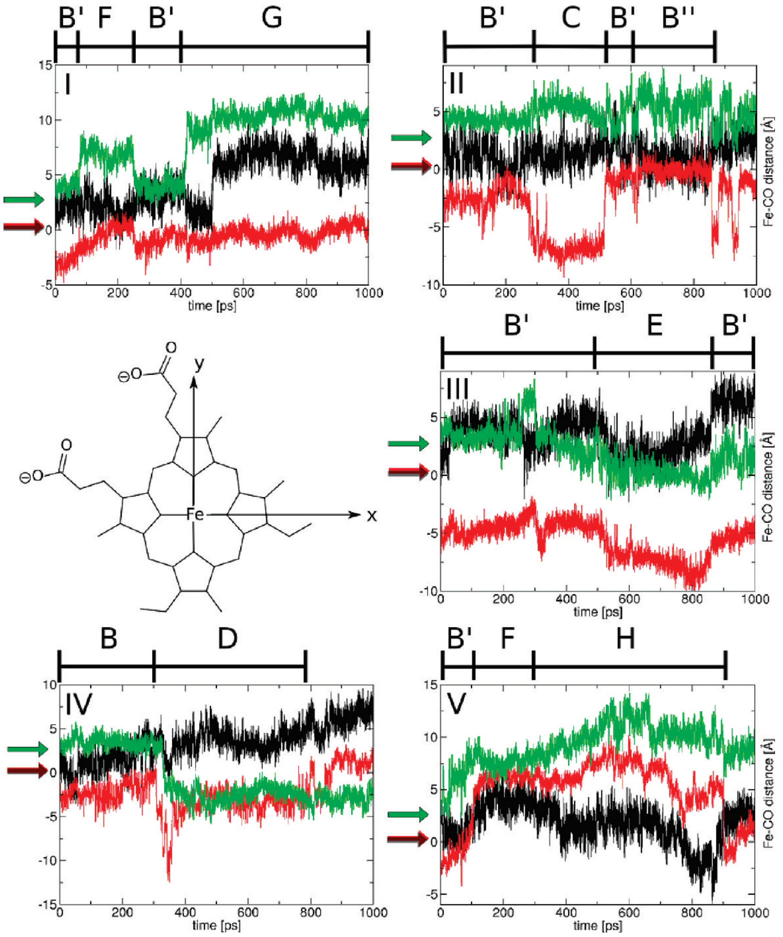

Figure 3. Time series of the CO center of mass to heme iron distance for the trajectories representing the probability distributions in Figure 2. Labels on top of each panel correspond to the ligand sampling a particular docking site. Arrows indicate the starting positions of the bound ligand at the beginning of the simulation. The heme plane defines the origin of the coordinate system (black, *x*-coordinate; red, *y*-coordinate; green, *z*-coordinate, where positive values correspond to the distal side and negative ones to the proximal side of the heme group).

indicate occasional transfers, and short-dashed arrows rarely observed ligand migrations.

Free Energy Profiles (FEP). Each identified docking site is described by the CO center-of-mass distance to the heme iron. For transitions between neighboring sites (see Figure 6) inspection of the trajectories showed that the Fe−CO distance is a useful local progression coordinate. In previous work it was also found that this distance is able to describe migration from the bound state via the distal pocket to pocket C (Xe4).[54,55] However, the coordinate does not distinguish between all docking sites. The relative stabilities of the different CO docking sites is determined from free energy simulations using the umbrella sampling method.[56] Docking site B″ is taken as the zero of energy, and the stability of all other sites is reported relative to this site. Here we are primarily interested in the relative stabilities of the different docking sites which should assist in differentiating between more or less probable ligand diffusion pathways.

Migration profiles to docking sites C (Xe4), D (Xe1), and E (Xe2) are shown in Figure 7, A, C, and D. These free energy profiles (FEPs) correspond to averaged potentials from three independent simulations, each starting from a different snapshot by assigning different initial velocity distributions. This approach allows error bars to be estimated. In the following, $\Delta G_{X/Y}$ denotes the relative stabilization energy of sites X and Y involved, and the barrier separating them is $G_{X\rightarrow Y}$.

In the wt His$_{\delta}$64 protein, the B to C migration with $G_{B\rightarrow C}$ = 0.8 kcal/mol (Figure 7A) has a transition state (TS) at a Fe−CO displacement of ≈4 Å. State B′ (identified by the Fe···CO separation) corresponds to a plateau in the FEP. Therefore, it is unstable and constitutes no further barrier for reaching docking site C. The FEP of wt His$_{\varepsilon}$64 Ngb−CO shows a stable B′ state with the same initial transfer barrier as in the wt His$_{\delta}$64 configuration ($G_{B''\rightarrow B'}$ = 0.8 kcal/mol). Further migration to docking site C involves an additional barrier of $G_{B'\rightarrow C}$ = 0.7 kcal/mol. The value of the reaction coordinate corresponding to site B″ in wt His$_{\varepsilon}$64 Ngb−CO is increased by ~0.7 Å compared to the His$_{\delta}$64 configuration due to N$_{\varepsilon}$H interacting with the iron atom. For the F28L mutant, B′ corresponds to a TS and the $G_{B\rightarrow C}$ barrier is 1.3 kcal/mol. The barrier difference between wt His$_{\delta}$64 and F28L Ngb−CO is in qualitative agreement with the docking site population distributions obtained from the unbiased ligand dynamics (Table 1). The reverse barrier, $G_{B\leftarrow C}$, is 0.4 kcal/mol smaller for F28L compared to that of the wt His$_{\delta}$64 protein which suggests a faster CO escape from this site in the mutant.

TABLE 1: Relative Occupation (%) of the Transient Docking Sites after Photodissociation of His$_\delta$64, His$_\epsilon$64, F28L, and F28W Ngb−CO[a]

docking site	surrounding residues	wt His$_\delta$64 NgbCO	wt His$_\epsilon$64 NgbCO	F28L NgbCO	F28W NgbCO
B′	Phe28, Leu31 Phe32	37	27	71	58
B″	Phe28, His64 Val68	16	28		3
C (Xe4)	Gly24, Leu27 Phe28, Ile65 Val68, Met69	20	11	13	38
D (Xe1)	Leu92, His96 Phe106, Met144	0	0	6	0
E (Xe2)	Ile72, Tyr137 Val140	5	0	5	1
F	Phe28, Phe42 Pro52, Phe61 Ile65	15	25	5	0
G	Thr25, Pro52 Glu53	4	6	0	0
H	Leu39, Pro52 Ser55	3	3	0	0

[a] For F28L, Ngb−CO docking sites B′ and B″ are unified because of the missing separation introduced in wt Ngb by the Phe28 side chain.

Figure 7B shows the FEPs for CO transfer in wt His$_\delta$64 (black) and wt His$_\epsilon$64 (green) Ngb−CO from site B″ via site F to G (solid) and H (dashed), respectively. These FEPs were only calculated for a single observed transfer as it is a rare event in the unbiased 1 ns trajectories. Migration from site F to G in the His$_\delta$64 configuration has an insignificant barrier, $G_{\mathrm{F}\rightarrow\mathrm{G}} \approx 0.2$ kcal/mol; the barrier from F to H is significantly larger ($G_{\mathrm{F}\rightarrow\mathrm{H}} = 1.0$ kcal/mol). Compared to site B, sites G and H are stabilized by $\Delta G_{\mathrm{B/H}} = 5.8$ kcal/mol. For His$_\epsilon$64, docking site F is more pronounced than for His$_\delta$64. Both pathways (B″ to G via F) and (B″ to H via F) involve considerable barriers between sites F and H ($G_{\mathrm{F}\rightarrow\mathrm{H}} = 2.3$ kcal/mol) or G ($G_{\mathrm{F}\rightarrow\mathrm{G}} = 4.3$ kcal/mol), respectively, for His$_\epsilon$64, whereas they are largely absent for His$_\delta$64. For His$_\epsilon$64 the pathway from B″ to H via F is unlikely because B″ and F are separated by a considerable barrier of $G_{\mathrm{B''}\rightarrow\mathrm{F}} = 4.1$ kcal/mol. Transfers to sites F, G, and H in the unbiased simulations are rarely observed because they require opening of the corresponding transfer channels (formed by residues Phe28, Phe32, Pro52, and Ile65), which only occurs infrequently.

Direct migration from site B to E for wt His$_\delta$64 (Figure 7C) has a single TS with $G_{\mathrm{B}\rightarrow\mathrm{E}} = 1.3$ kcal/mol. Conversely, the FEP of the F28L mutant has an intermediate minimum located at the position of the native $G_{\mathrm{B}\rightarrow\mathrm{E}}$ TS arising from reduced steric hindrance at position 28. The overall barrier in the mutant is significantly larger, however, with $G_{\mathrm{B}\rightarrow\mathrm{E}} = 3.1$ kcal/mol. Moreover, the error bars in F28L Ngb−CO are roughly 2 times larger than those of wt Ngb. Both observations suggest that, in the F28L mutant, fluctuations are larger and rearrangement is required for ligand migration from site B to E to occur.

For the wt His$_\delta$64 protein, none of the unbiased simulations lead to CO migration to site D (Xe1). Umbrella sampling of this transfer showed a rather large barrier of $G_{\mathrm{E}\rightarrow\mathrm{D}} = 2.7$ kcal/mol ending on a plateau (Figure 7D) and, thus, Xe1 does not appear to be a stable substate in the wild-type protein. In F28L Ngb−CO, the same transfer shows a stabilization by $\Delta G_{\mathrm{E/D}} = 1.6$ kcal/mol after crossing over a barrier of $G_{\mathrm{E}\rightarrow\mathrm{D}} = 1.9$ kcal/mol, to reach site D. Simulations for His$_\epsilon$64 Ngb−CO showed no transfer to sites E and D within the 1 ns of simulation time; thus, no FEPs for these transfers were calculated.

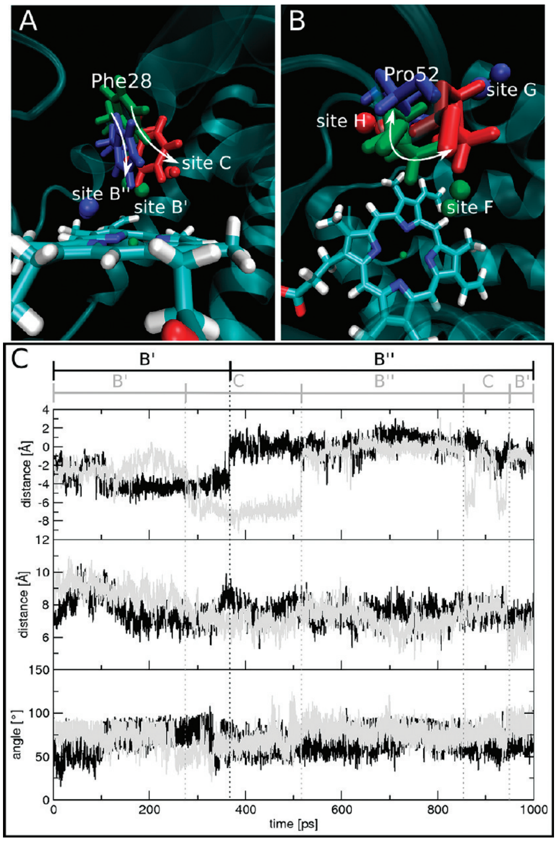

Figure 4. Illustration of the gating function of residues Phe28 and Pro52. (A) Aligned snapshots for deoxy wt Ngb−CO trajectories to illustrate the role of Phe28 (green, CO in site B′; blue, CO in site B″; red, CO on its way to site C). (B) Snapshot to highlight the role of Pro52 taken after transfer from site B″ to F (green), after transfers from site F to G (blue) and to H (red). (C, top graph) Two trajectories (black and gray) showing docking site transfers by the CO (center of mass displacement) along the y-axis. (C, center graph) Distance between His64 ring center and heme-iron for both trajectories. (C, lower graph) His64 ring tilting angle relative to the heme plane for both trajectories.

Table 2 summarizes the free energies of all docking sites relative to site B″. Both primary docking sites (B′, B″) are higher in energy than most of the secondary and tertiary docking sites. It is important to recall that details of the pathway energetics depend somewhat on the fact that they were estimated along observed migration paths. These energies as well as the barriers separating them represent only a subset of the entire reaction network of the protein. Therefore, stability and TS energies should not be directly compared with the docking site occurrences described in Table 1 which were derived from unbiased simulations. Finally, umbrella sampling has the tendency of inducing additional relaxation processes which may bias stabilities and transition state energies.[54,55]

20 K Simulations. The FTIR-TDS experiments described below were started at $T < 10$ K. To allow for a direct comparison of experimental and computational results, additional simulations were performed at temperatures close to the experimental conditions. The fact that protein fluctuations are no longer activated[57] was accounted for by running each simulation in a different fixed conformational substate of Ngb (cooled 300 K

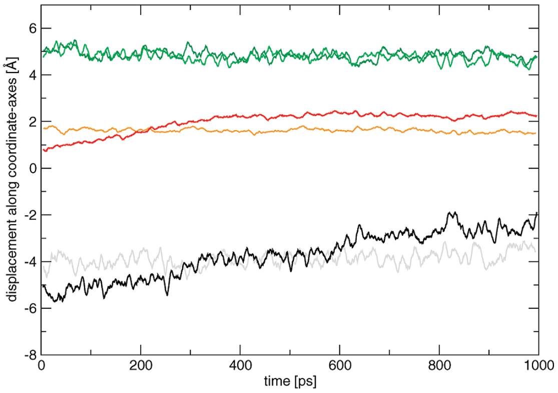

Figure 5. Average time series of the projected distance between the N_ε atom of His64 and the heme iron along the *x*-, *y*-, and *z*-axis for $His_\delta64$ Ngb−CO (black, red, light green) and $His_\varepsilon64$ Ngb−CO (gray, orange, dark green) after CO dissociation. $His_\varepsilon64$ has already isomerized (see text) and the time series show that this conformation is stable compared to $His_\delta64$ for which drifts in the *x*- and *y*-coordinate are found. The orientation of the coordinate system corresponds to the one from Figure 3.

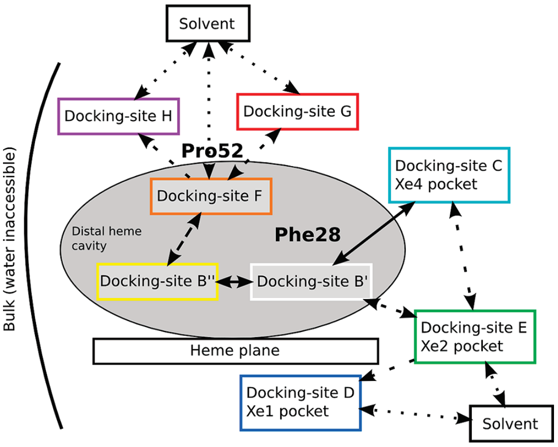

Figure 6. Schematic representation of all docking sites and their connectivities observed in the simulations. Sites C, D, and E coincide with xenon pockets Xe4, Xe1, and Xe2 in Mb. Transfers observed in MD simulations are indicated by arrows. Solid arrows describe frequent transfers (>15), long-dashed arrows occasional transfers, and short-dashed arrows very rarely observed ligand migrations. Transfers to the bulk solvent were not observed; possible connections to the exterior are illustrated by dotted lines.

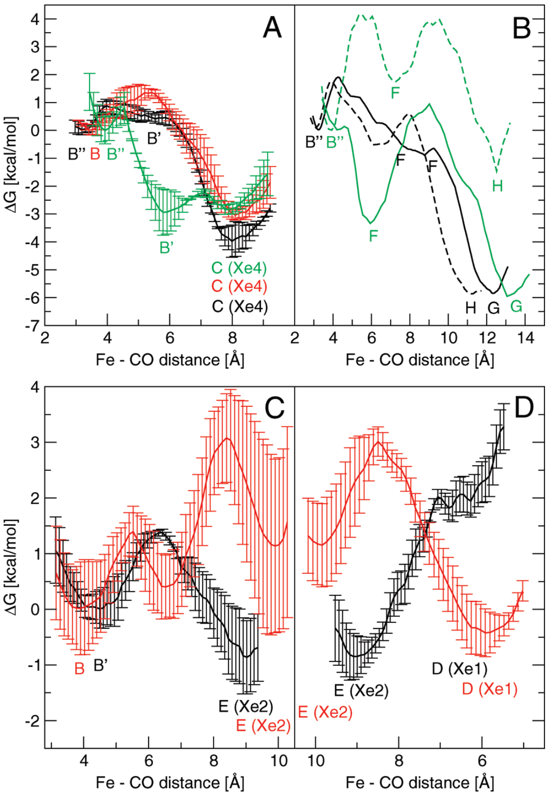

Figure 7. Free energy profiles for ligand migration between different sites from umbrella sampling simulations. Color codes: wt $His_\delta64$ (black), wt $His_\varepsilon64$ (green), and F28L Ngb−CO (red). In the different panels, specific migration paths are illustrated: CO migration from site B to C (Xe4) (panel A); transfer from site B″ over F to G (solid) and H (dashed) (panel B); CO migration from site B to E (panel C); CO migration from site E to D (panel D). Note that panel B shows two pathways (see also Figure 6): B″ to G via F (solid lines) and B″ to H via F (dashed lines).

TABLE 2: Average Stability of the Docking Sites in Wt $His_\delta64$, Wt $His_\epsilon64$, and F28L $His_\delta64$ Ngb−CO According to Their Relative Free Energies Compared to Site B″, Calculated from Umbrella Sampling MD Simulations at 300 K

	ΔG (kcal/mol)		
docking site	wt $His_\delta64$ NgbCO	wt $His_\varepsilon64$ NgbCO	F28L $His_\delta64$ NgbCO
G	–5.8	–6.0	–
H	–5.8	–1.5	–
C (Xe4)	–3.3	–2.9	–2.5
F	–0.7	–3.3	–
E (Xe2)	–0.9	–	1.1
B″	0.0	0.0	0.0
B′	0.5	–2.9	–
D (Xe1)	1.8	–	–0.4

snapshots). This procedure mimics the experimental conditions in a simplified fashion.

The relative populations of the primary docking site B evaluated from these calculations (59% in wt and 71% in F28L Ngb−CO) are similar to those obtained from the simulations at 300 K (see Table 1). Migration from site B′ to B″ was observed only in one specific wt Ngb−CO trajectory and was only of short duration (≈200 ps). In the corresponding F28L Ngb−CO samplings, the carbon atom of the ligand in site B strongly interacts with the Fe (site B″-like arrangement). The rarely observed CO transfers (three times within the total 1 μs of simulation) to docking site C in F28L Ngb−CO happened already within a few hundred picoseconds after dissociation. At this time, the ligand still had excess kinetic energy from the dissociation event. In wt Ngb−CO, two of in total five observed migrations to secondary docking sites took place 24 and 53 ns after the dissociation event. Once in site C, back transfer to the primary docking sites was not observed within the 100 ns of simulations resulting in a still moderately populated secondary docking site C (41% and 29% for wt and F28L Ngb−CO, respectively). This compares with B-site populations in wt (59%) and F28L Ngb−CO (71%) from simulations at 300 K (see Table 1). The previous finding of a larger barrier against

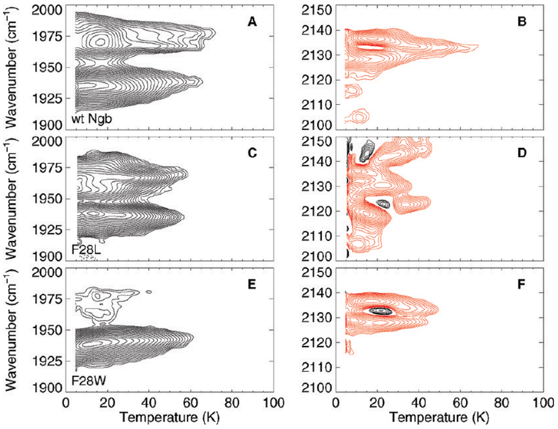

Figure 8. TDS contour maps of (A,B) wt NgbCO, (C,D) F28L NgbCO, and (E,F) F28W NgbCO obtained after 1 s illumination at 4 K. Left column: absorption changes in the bands of bound CO. Right column: absorption changes in the photoproduct bands. Solid black lines indicate an absorption increase; solid red lines indicate a loss in absorption. Contours are spaced logarithmically.

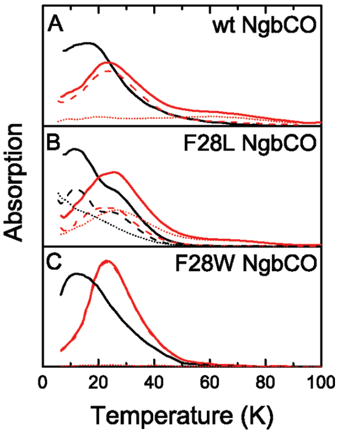

Figure 9. Integrated absorption changes calculated from the TDS contour maps in the left columns of Figure 8 (black lines) and Figure 9 (red lines). Solid lines: overall absorption change. Dashed lines: changes in A_1. Dotted lines: changes in A_2 and A_0.

migration to docking site C in F28L Ngb−CO thus corroborates the present observations at 20 K. Transitions from the primary docking site to more remote docking sites E (Xe2) and G were only observed for wt Ngb−CO.

B. FTIR-TDS at Cryogenic Temperatures. ***Photodissociation of Ngb.*** All NgbCO samples display several IR bands of heme-bound CO. Each band represents a particular conformational substate. In wt NgbCO, they have been denoted as A_1 (~1930 cm^{-1}), which is the dominant substate, A_2 (~1980 cm^{-1}), and A_0 (~1965 cm^{-1}). The iron−ligand bond can be ruptured by light. If ligand dissociation is initiated by a 1 s laser pulse (532 nm, 300 mW) at 4 K, the ligands are expected to migrate preferentially to the primary docking site B. This brief illumination has proven sufficient to photolyze all CO ligands in Mb. However, the photoproduct yield in wt Ngb−CO amounts to only ~50% for each of the A substates. In mutant F28L NgbCO, ~75% of the A_1 substate is photolyzed, but only ~40% of A_0 and A_2, suggesting that site B in the A_1 substate has become better suited to trap a ligand. In F28W NgbCO, ~65% of A_1 is photodissociated. The population in A_2 and A_0 is essentially negligible. To screen for all possible docking sites, the samples were cooled under steady illumination from 140 to 4 K (ramp rate 0.3 K/min).[42,58] However, even this so-called slow cooling protocol did not result in complete photolysis, indicating very low barriers against rebinding.

Upon photolysis, a multitude of photoproduct bands appear. In general, both their number and the frequencies depend on the illumination protocol. The bands may represent CO in different orientations at one transient docking site, CO at the same site in different protein conformations and CO in different sites. It is difficult to predict putative transient docking sites in Ngb because of the space available to the photolyzed ligands (~300 Å^3). To sort the photolyzed ligands according to their enthalpy barriers against rebinding, TDS experiments were started immediately after illumination.

Ligand Rebinding From the Primary Docking Site. Figure 8 shows the absorption changes in the bands of heme-bound and photolyzed CO, respectively, monitored after 1 s illumination. wt Ngb−CO shows strong rebinding features in A_1 already at 4 K (Figure 8A). The enthalpy barrier distribution for rebinding was fitted with a gamma function peaking at 1.1 kJ/mol[15]. In the corresponding photoproduct map, we see a concomitant absorption decrease of essentially two bands. Rebinding from the photoproduct state represented by the band at 2129 cm^{-1} peaks at 7 K, whereas most ligands associated with the band at 2136 cm^{-1} recombine at 20 K. We suggest that these two main photoproduct bands arise from Stark splitting of the CO band in the local electric field at the primary docking site B of A_1. A few CO ligands do not return to the heme iron but instead migrate to site C of A_1. This process is only adumbrated by the contour-free area at 20 K and 2132 cm^{-1} of Figure 8B, but is more apparent from the black contours in Figure 8F. Photoproduct bands associated with the minor substates A_0 and A_2 could not be resolved because of too low amplitudes. Please note that only about 30% of the molecules adopt these subconformations.

The A state contour plot of F28L NgbCO again shows that rebinding in all subconformations starts already at the lowest temperature (Figure 8C). The photoproduct map of F28L Ngb−CO obtained after 1-s illumination (Figure 8D) reveals an increased manifold of photoproduct states as compared to that of the wt protein (Figure 8B). At least five bands at 2106, 2119, 2129, 2136, and 2142 cm^{-1} can be found in the mutant spectrum already at 4 K. Additional photoproduct states are occupied in the course of the TDS experiment. We cannot decide whether the concomitant red and black contours at particular temperatures (Figure 8D) represent ligand reorientation at a particular site or migration between different sites. An assignment of the different photoproduct bands to the A substates is also extremely difficult (i) because of their extensive spectral overlap and (ii) because of the ligand dynamics.

The TDS maps of F28W NgbCO are similar to those of wt Ngb (Figure 8E,F). Although the contours associated with A_0 and A_2 are small (Figure 8E), they still allow us to estimate the peak positions of A_0 (1965 cm^{-1}) and A_2 (1978 cm^{-1}). The photoproduct map indicates rebinding from site B and migration to site C in molecules in the A_1 conformation. The latter process is more pronounced than in wt Ngb; it is clearly visible from the black contours. Photoproduct bands associated with A_0 and A_2 are again too small to be detected.

To compare the temperature dependencies of the rebinding process in the different samples, we have integrated the TDS data along the wavenumber axis (Figure 9). At first sight, a

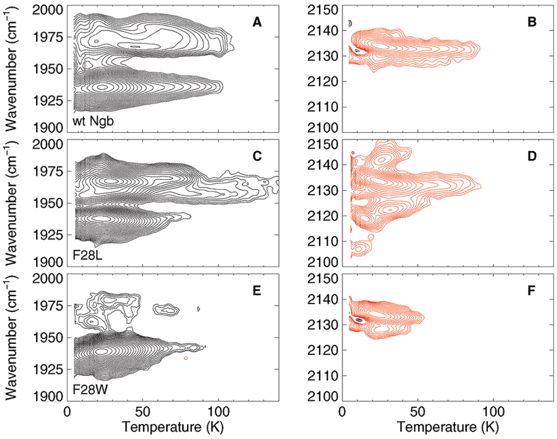

Figure 10. TDS contour maps of (A,B) wt NgbCO, (C,D) F28L NgbCO, and (E,F) F28W NgbCO obtained after slow cool illumination from 140 to 4 K. Left column: absorption changes in the bands of bound CO. Right column: absorption changes in the photoproduct bands. Solid black lines indicate an absorption increase; solid red lines indicate a loss in absorption. Contours are spaced logarithmically.

very similar rebinding behavior is observed for all three samples. Rebinding is already significant at the lowest temperatures and has a maximum between 10 and 20 K. A closer look reveals, however, a distinct two-step rebinding behavior of F28L Ngb−CO. This effect is related only to A_1, as can be seen from a comparison of the overall absorption change (solid black line in Figure 9B) and the integrated absorption changes of A_1 (dashed black line) and A_0 (dotted black line), suggesting that, only in A_1, two separate docking sites are populated.

Ligand Rebinding From Secondary Docking Sites. After slow cooling of wt Ngb−CO under illumination from 140 to 4 K, most ligands rebind from the secondary docking site C which is associated with the photoproduct band at 2134 cm^{-1} (Figure 10A,B). Recombination is maximal at ~25 K (Figure 9A), which suggests a low barrier for the return to the binding site. An additional rebinding process that extends from ≈ 50 to 100 K might represent rebinding from still another docking site, which is represented by a band at 2133 cm^{-1} (Figure 10B).

After slow-cool illumination of F28L Ngb−CO, the TDS features of the wt Ngb−CO map are essentially retained (Figure 10C,D), implying that the secondary sites present in wt Ngb are unchanged in the mutant sample. However, additional contours peaking at 2119, 2123, and 2142 cm^{-1} indicate rebinding from sites unique to this mutant. Overall, the temperature of maximal rebinding is shifted to higher temperature, implying that the CO ligands rebind from a secondary site (Figure 9). As already observed after 1 s illumination, the TDS maps of F28W Ngb−CO resemble those of wt Ngb−CO (Figure 10E,F).

IV. Discussion and Conclusion

Ngb is a small globular heme protein with a spacious internal cavity. Its heme iron is hexacoordinated by the proximal and distal histidines in the absence of an exogenous ligand; binding of the exogenous ligand is made possible by rupturing the covalent bond between the heme iron and the distal histidine, which is accompanied by a sliding motion of the heme and the concomitant reshaping of the large open volume.[13,14] In the ligand-bound state, the active site of NgbCO is structurally heterogeneous, as can be inferred from multiple stretching bands of heme-bound CO in the FTIR spectra (see Figure 8). Up to four CO bands are observed, depending on pH, which may arise from different orientations of the His64 side chain.[41] Their detailed structural interpretation is, however, difficult and will not be attempted here. For MbCO, it is well established that the three stretching bands are associated with different orientations and protonation states of the His64 side chain.[59,60] A recent study confirmed the different structural origins of A_0 (open conformation of His64) and A_1/A_3 (closed conformation), in agreement with experimental work.[61] A_1 and A_3 are structurally similar; their precise structural differences, however, still remain elusive.[59−62]

Ligand association and dissociation, migration, cavity rearrangement, and the competition between exogenous ligand and His64 for the binding site are key issues involved in the functional peculiarities of Ngb. Hence, profound investigations of the dynamics and structural rearrangements involved in these processes are necessary for a deeper insight into the function of Ngb. Obtaining direct experimental evidence on kinetic processes such as migration of a ligand between internal docking sites in a protein is challenging and requires techniques such as kinetic trapping X-ray crystallography[19,63] or picosecond time-resolved X-ray crystallography.[20,64] Here, an alternative approach has been taken. Combining the complementary strengths and merits of atomistic simulations and FTIR-TDS spectroscopy, the docking site network for CO migration in wt and mutant Ngb samples has been mapped out. The simulations at 300 K find eight docking sites, compared with three detected from experiment. On the other hand, simulations at the temperatures used in the experiments indicate that ligand migration to sites other than B and C appear not possible, which is in agreement with the spectroscopic data. In the low-temperature simulations, it is found that docking site F is occasionally populated from docking site B″; however, no direct evidence of this site exists from experiment. It is possible that the spectroscopic features of this site overlap with those of another site.

In an initial experimental study on ligand migration in wt Ngb−CO,[15] two transient docking sites were unambiguously resolved; evidence of a third site was present, but weak. On the basis of the rebinding behavior of the CO ligand, they were tentatively denoted as primary site B and secondary site C. The location of these sites could, however, not be determined on the basis of the FTIR-TDS data alone. Due to the low rebinding barriers, it was suggested that these sites are in close vicinity to the binding site. The present MD simulations at 300 K have provided the necessary additional insight into the nature of these sites. Site B is indeed the analogue of the primary docking site B seen in wt Mb, represented by the white sphere in Figure 1. This was found from detailed comparison between experimentally and computationally determined IR spectra obtained from extensive MD simulations of CO in different pockets of wild-type and mutant Ngb.[31] The present simulations indicate that, for both wt and F28W Ngb, docking site B consists of two separate sites B′ and B″. Compared to wt Mb, the MD simulations of photolyzed Ngb reveal a slightly larger displacement of docking site B′ from the heme iron (≈1.8 to 2.8 Å versus ≈1.7 Å). Concomitantly, the distance between the His64 side chain and the CO ligand is increased, leading to a smaller Stark splitting of the corresponding photoproduct bands (8 cm^{-1} in Ngb as compared to 14 cm^{-1} in Mb).[65] The spatial separation caused by the aromatic side chain of residue B10 is missing in mutant F28L. Therefore, only a single B site could be resolved in the MD simulations (Figure 7A).

The additional site B″ is not unique to Ngb, but was earlier described in MD simulations on the L29F Mb mutant by Anfinrud and co-workers.[66] The large Stark splitting observed with FTIR spectroscopy for CO trapped in the primary site in L29F Mb (24 cm^{-1}) and the observation that rebinding starts already at 4 K imply that, at cryogenic temperatures, the CO ligands are trapped at site B″, right on top of the heme iron.[67] At ambient temperatures, CO migration to site B″ of Mb is coupled to fluctuations of Phe29, in agreement with the present findings for Ngb (Figure 4C). In the L29W MbCO mutant, two active site conformations, A_I and A_{II}, exist that differ in the orientation of the Trp29 side chain.[19,68] Below 180 K, large-scale protein motions are arrested and transitions between these conformations are not possible.[69,70] Whereas in conformation A_I only site B′ is available to the photolyzed ligand at 4 K, in A_{II} only site B″ can be populated. Structurally, the orientation of the indole side chain in A_{II} corresponds to that of the Phe29 side chain in L29F Mb.

The FTIR-TDS data collected after 1 s illumination reveal that most ligands rebound from the primary site B. A temperature increase to ≈20 K during the TDS experiment enabled a few ligands to migrate from site B to site C (Figure 8). After slow cool illumination of wt NgbCO, essentially all ligands rebound from site C (Figure 10). The same recombination behavior was observed for mutant F28W. In contrast, in mutant F28L, no significant fraction of ligands could be trapped at site C upon 1 s illumination. CO rather returned from site B. In agreement with these experimental data, the MD simulations at 20 K also showed ligand transfer only between two docking sites in wt and F28L Ngb−CO, namely sites B and C. And, as observed in the experiment, migration to site C was less frequent in the simulations on F28L. This is also reflected in the calculated barriers $G_{B\to C}^{T=300}$ for the B → C migration of 0.8 and 1.3 kcal/mol for the wt and the F28L mutant, respectively. Based on this remarkable agreement, we identify the secondary site C seen in the experiment as the Xe4 analogue, represented by the cyan sphere (see Figure 1).

An analogous behavior is found in wt Mb and mutants with an aromatic residue at position 29, where we also observe increased trapping in the Xe cavities.[71−73] Apparently, it is a common phenomenon in globins that bulky residues at position B10 favor ligand migration away from the primary docking site. In L29F MbCO, migration from site B to the Xe4 pocket occurs on the subnanosecond time scale,[20,36,66] which is approximately 1 order of magnitude faster than in wt MbCO.[36,54] For the even larger tryptophan residue in L29W MbCO, this trend was found to continue.[19,21] Our theoretical ligand population analysis (see Table 1) and the FEPs in Ngb also support a direct relation between the size of the B10 side chain and the barrier against migration. The umbrella sampling calculations on photodissociated Ngb−CO suggest overall reduced barriers (by up to 4 kcal/mol) compared to wt MbCO[19,20,54,73] governing migration to docking site C and probably other remote sites.

Experimental evidence of additional docking sites is rather weak. The integrated absorption changes obtained after slow cooling shown in Figure 9 indicate rebinding up to about 100 K in both wt Ngb and mutant F28L, whereas the recombination process in F28W is complete at 60 K. The respective contour maps indicate show that recombination at higher temperatures is mainly related to the open conformation (Figure 10). It is of interest to note that the F28W mutant essentially lacks this open conformation. To locate the related site, we refer to the results from umbrella sampling simulations on F28L (Figure 7B). The calculations predict a large barrier of 7.1 kcal/mol for the transfer from site H via site F to the primary site B from which CO would finally rebind. In an open conformation, the distal site is opened up, which would most likely simplify ligand transfer to sites F, G, and H. Therefore, we suggest that the FTIR-TDS data of mutant F28L represent ligand rebinding from sites B, C, and from above the distal heme cavity (G, H).

In summary, the present study has mapped out parts of the ligand migration network in Ngb by combining computational and experimental techniques. Whereas the experimental data can distinguish at most three ligand docking sites, all of which are rather close to the binding site, the MD simulations at 300 K have revealed a total of eight possible docking sites. The transfer barriers toward the more remote Xe1 (D), Xe2 (E), and Xe4 (C) sites are considerably lower in Ngb−CO compared to Mb−CO, most likely due to the larger internal protein cavity and the more pronounced active site fluctuations in the photodissociated system. These fluctuations are potentially relevant for bis-His-hexacoordination, which, in turn, may be important for the functional differences between Ngb and Mb.

Acknowledgment. This work was supported by the Schweizerischer Nationalfonds (S.L. and M.M.) through grant 200021-117810 (to M.M.) and by the Deutsche Forschungsgemeinschaft (DFG) and the State of Baden-Württemberg through the Center for Functional Nanostructures (CFN), and DFG grant Ni 291/3 (to G.U.N.). We thank Uwe Theilen for expression and purification of the recombinant proteins used in this study.

References and Notes

(1) Burmester, T.; Weich, B.; Reinhardt, S.; Hankeln, T. *Nature* **2000**, *407*, 520–523.

(2) Nienhaus, K.; Nienhaus, G. U. *IUBMB LIFE* **2007**, *59*, 490–497.

(3) Fago, A.; Hundahl, C.; Dewilde, S.; Gilany, K.; Moens, L.; Weber, R. E. *J. Biol. Chem.* **2004**, *279*, 44417–44426.

(4) Greenberg, D. A.; Jin, K.; Khan, A. A. *Curr. Opin. Pharmacol.* **2008**, *8*, 20–24.

(5) Esplugues, J. V. *Br. J. Pharmacol.* **2002**, *135*, 1079–1095.

(6) Herold, S.; Fago, A.; Weber, R. E.; Dewilde, S.; Moens, L. *J. Biol. Chem.* **2004**, *279*, 22841–22847.

(7) Weiland, T. R.; Kundu, S.; Trent, J. T.; Hoy, J. A.; Hargrove, M. S. *J. Am. Chem. Soc.* **2004**, *126*, 11930–11935.

(8) Roesner, A.; Fuchs, C.; Hankeln, T.; Burmester, T. *Mol. Biol. Evol.* **2005**, *22*, 12–20.

(9) Pesce, A.; Dewilde, S.; Nardini, M.; Moens, L.; Ascenzi, P.; Hankeln, T.; Burmester, T.; Bolognesi, M. *Structure* **2003**, *11*, 1087–1095.

(10) Dewilde, S.; Kiger, L.; Burmester, T.; Hankeln, T.; Baudin-Creuza, V.; Aerts, T.; Marden, M. C.; Caubergs, R.; Moens, L. *J. Biol. Chem.* **2001**, *276*, 38949–38955.

(11) Fago, A.; Mathews, A. J.; Dewilde, S.; Moens, L.; Brittain, T. *J. Inorg. Biochem.* **2006**, *100*, 1339–1343.

(12) Brunori, M.; Giuffre, A.; Nienhaus, K.; Nienhaus, G. U.; Scandurra, F. M.; Vallone, B. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 8483–8488.

(13) Vallone, B.; Nienhaus, K.; Matthes, A.; Brunori, M.; Nienhaus, G. U. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17351–17356.

(14) Vallone, B.; Nienhaus, K.; Brunori, M.; Nienhaus, G. U. *Proteins* **2004**, *56*, 85–92.

(15) Nienhaus, K.; Nienhaus, G. U. *J. Biol. Phys.* **2005**, *31*, 417–432.

(16) Tilton, R. F.; Kuntz, I. D.; Petsko, G. A. *Biochemistry* **1984**, *23*, 2849–2857.

(17) Schlichting, I.; Berendzen, J.; Phillips, G. N.; Sweet, R. M. *Nature* **1994**, *371*, 808–812.

(18) Hartmann, H.; Zinser, S.; Kominos, P.; Schneider, R. T.; Nienhaus, G. U.; Parak, F. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 7013–7016.

(19) Ostermann, A.; Waschipky, R.; Parak, F. G.; Nienhaus, G. U. *Nature* **2000**, *404*, 205–208.

(20) Schotte, F.; Lim, M. H.; Jackson, T. A.; Smirnov, A. V.; Soman, J.; Olson, J. S.; Phillips, G. N.; Wulff, M.; Anfinrud, P. A. *Science* **2003**, *300*, 1944–1947.

(21) Schmidt, M.; Nienhaus, K.; Pahl, R.; Krasselt, A.; Anderson, S.; Parak, F.; Nienhaus, G. U.; Srajer, V. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 11704–11709.

(22) Nienhaus, K.; Deng, P.; Kriegl, J. M.; Nienhaus, G. U. *Biochemistry* **2003**, *42*, 9647–9658.

(23) Ouellett, H.; Ouellett, Y.; Richard, C.; Labarre, M.; Wittenberg, B.; Wittenberg, J.; Guertin, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5902–5907.

(24) Mishra, S.; Meuwly, M. *Biophys. J.* **2009**, *96*, 2105–2118.

(25) Couture, M.; Yeh, S. R.; Wittenberg, B. A.; Wittenberg, J. B.; Ouellet, Y.; Rousseau, D. L.; Guertin, M. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11223–11228.

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

(27) Mackerel Jr., A.; Brooks III, C.; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. In *CHARMM: The energy function and its parameterization with an overview of the program.*; Schleyer, P. v. R., et al., Eds.; John Wiley & Sons: Chichester, UK, 1998; Vol. 1, pp 271−277.

(28) Johnson, J.; Lamb, D.; Frauenfelder, H.; Müller, J. D.; McMahon, B.; Nienhaus, G.; Young, R. D. *Biophys. J.* **1996**, *71*, 1563–1573.

(29) Rovira, C.; Schulze, B.; Eichinger, M.; Evanseck, J. D.; Parrinello, M. *Biophys. J.* **2001**, *81*, 435–445.

(30) Merchant, K. A.; Noid, W. G.; Thompson, D. E.; Akiyama, R.; Loring, R. F.; Fayer, M. D. *J. Phys. Chem. B* **2003**, *107*, 4–7.

(31) Nienhaus, K.; Lutz, S.; Meuwly, M.; Nienhaus, G. U. *ChemPhysChem* **2009**, in press.

(32) Brooks, C. L.; Karplus, M. *J. Chem. Phys.* **1983**, *79*, 6312–6325.

(33) Huffaker, J. N. *J. Chem. Phys.* **1976**, *64*, 3175–3181.

(34) Huffaker, J. N. *J. Chem. Phys.* **1976**, *64*, 4564–4570.

(35) Nutt, D. R.; Meuwly, M. *Biophys. J.* **2003**, *85*, 3612–3623.

(36) Nutt, D. R.; Meuwly, M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 5998–6002.

(37) Straub, J. E.; Karplus, M. *Chem. Phys.* **1991**, *158*, 221–248.

(38) Meuwly, M.; Becker, O. M.; Stote, R.; Karplus, M. *Biophys. Chem.* **2002**, *98*, 183–207.

(39) Kuczera, K.; Kuriyan, J.; Karplus, M. *J. Mol. Biol.* **1990**, *213*, 351–373.

(40) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Chem. Comput.* **1995**, *16*, 1339–1350.

(41) Kriegl, J. M.; Bhattacharyya, A. J.; Nienhaus, K.; Deng, P.; Minkow, O.; Nienhaus, G. U. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 7992–7997.

(42) Nienhaus, G. U.; Mourant, J. R.; Chu, K.; Frauenfelder, H. *Biochemistry* **1994**, *33*, 13413–13430.

(43) Berendzen, J.; Braunstein, D. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 1–5.

(44) Mourant, J. R.; Braunstein, D. P.; Chu, K.; Frauenfelder, H.; Nienhaus, G. U.; Ormos, P.; Young, R. D. *Biophys. J.* **1993**, *65*, 1496–1507.

(45) Austin, R. H.; Beeson, K. W.; Eisenstein, L.; Frauenfelder, H.; Gunsalus, I. C. *Biochemistry* **1975**, *14*, 5355–5373.

(46) Ehrenstein, D.; Nienhaus, G. U. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 9681–9685.

(47) Kriegl, J. M.; Nienhaus, K.; Deng, P.; Fuchs, J.; Nienhaus, G. U. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 7069–7074.

(48) Lehle, H.; Kriegl, J. M.; Nienhaus, K.; Deng, P.; Fengler, S.; Nienhaus, G. U. *Biophys. J.* **2005**, *88*, 1978–1990.

(49) Lamb, D. C.; Nienhaus, K.; Arcovito, A.; Draghi, F.; Miele, A. E.; Brunori, M.; Nienhaus, G. U. *J. Biol. Chem.* **2002**, *277*, 11636–11644.

(50) Nienhaus, K.; Maes, E. M.; Weichsel, A.; Montfort, W. R.; Nienhaus, G. U. *J. Biol. Chem.* **2004**, *279*, 39401–39407.

(51) Nienhaus, G.; Nienhaus, G. U. Ligand dynamics in heme proteins observed by Fourier transform infrared spectroscopy at cryogenic temperatures. In *Globins and other nitric oxide-reactive proteins, part B*; Elsevier Academic Press Inc: New York, 2008; Vol. 437, pp 347−378.

(52) Frauenfelder, H.; McMahon, B. H.; Austin, R. H.; Chu, K.; Groves, J. T. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 2370–2374.

(53) Brunori, M.; Vallone, B. *Cell. Mol. Life Sci.* **2007**, *64*, 1259–1268.

(54) Banushkina, P.; Meuwly, M. *J. Phys. Chem. B* **2005**, *109*, 16911–16917.

(55) Banushkina, P.; Meuwly, M. *J. Chem. Phys.* **2007**, *127*, 135101.

(56) Kottalam, J.; Case, D. A. *J. Am. Chem. Soc.* **1988**, *110*, 7690–7697.

(57) Frauenfelder, H.; Parak, F.; Young, R. D. *Annu. Rev. Biophys. Biochem.* **1988**, *17*, 451–479.

(58) Chu, K.; Ernst, R. M.; Frauenfelder, H.; Mourant, J. R.; Nienhaus, G. U.; Philipp, R. *Phys. Rev. Lett.* **1995**, *74*, 2607–2610.

(59) Yang, F.; Phillips, G. N. *J. Mol. Biol.* **1996**, *256*, 762–774.

(60) Müller, J. D.; McMahon, B. H.; Chien, E. Y. T.; Sligar, S. G.; Nienhaus, G. U. *Biophys. J.* **1999**, *77*, 1036–1051.

(61) Devereux, M.; Meuwly, M. *Biophys. J.* **2009**, *96*, 4363–4375.

(62) Ishikawa, H.; Kwak, K.; Chung, J. K.; Kim, S.; Fayer, M. D. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 8619–8624.

(63) Schlichting, I.; Chu, K. *Curr. Opin. Struct. Biol.* **2000**, *10*, 744–752.

(64) Key, J.; Srajer, V.; Pahl, R.; Moffat, K. *Biochemistry* **2007**, *46*, 4706–4715.

(65) Nienhaus, K.; Olson, J. S.; Franzen, S.; Nienhaus, G. U. *J. Am. Chem. Soc.* **2005**, *127*, 40–41.

(66) Hummer, G.; Schotte, F.; Anfinrud, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15330–15334.

(67) Nienhaus, K.; Nienhaus, G. U. *J. Biol. Phys.* **2007**, *33*, 357–370.

(68) Nienhaus, K.; Ostermann, A.; Nienhaus, G. U.; Parak, F. G.; Schmidt, M. *Biochemistry* **2005**, *44*, 5095–5105.

(69) Frauenfelder, H.; Nienhaus, G. U.; Johnson, J. B. *Ber. Bunsen. Phys. Chem.* **1991**, *95*, 272–278.

(70) Nienhaus, G. U.; Heinzl, J.; Huenges, E.; Parak, F. *Nature* **1989**, *338*, 665–666.

(71) Olson, J. S.; Soman, J.; Phillips, G. N. *IUBMB LIFE* **2007**, *59*, 552–562.

(72) Lamb, D. C.; Arcovito, A.; Nienhaus, K.; Minkow, O.; Draghi, F.; Brunori, M.; Nienhaus, G. U. *Biophys. Chem.* **2004**, *109*, 41–58.

(73) Nienhaus, G. U.; Nienhaus, K. *J. Biol. Phys.* **2002**, *28*, 163–172.

JP905673P

6.2. Structural identification of spectroscopic substates in neuroglobin

The following paper appeared as

K. Nienhaus, S. Lutz, M. Meuwly and G. U. Nienhaus, *ChemPhysChem*, **11(1)** 119 – 129 (2010).

DOI: 10.1002/cphc.200900637

Structural Identification of Spectroscopic Substates in Neuroglobin

Karin Nienhaus,[a] Stephan Lutz,[b] Markus Meuwly,*[b, c] and G. Ulrich Nienhaus*[a, d]

The structural origins of infrared absorptions of photodissociated CO in murine neuroglobin (Ngb) are determined by combining Fourier transform infrared (FTIR) spectroscopy and molecular dynamics (MD) simulations. Such an approach allows to identify and characterize both the different conformations of the Ngb active site and the transient ligand docking sites. To capture the influence of the protein environment on the spectroscopy and dynamics, experiments and simulations are carried out for the wild type protein and its F28L and F28W mutants. It is found that a voluminous side chain at position 28 divides site B into two subsites, B′ and B″. At low temperatures, CO in wt Ngb only migrates to site B′ from where it can rebind, and B″ is not populated. The spectra of CO in site B′ for wt Ngb from simulations and experiments are very similar in spectral shift and shape. They both show doublets, red-shifted with respect to gas-phase CO and split by $\approx 8\ \mathrm{cm}^{-1}$. The FTIR spectra of the F28L mutant show additional bands which are also found in the simulations and can be attributed to CO located in substate B″. The different bands are mainly related to different orientations of the His64 side chain with respect to the CO ligand. Large red-shifts arise from strong interactions between the Histidine–NH and the CO oxygen. After dissociation from the heme iron, the CO ligand visits multiple docking sites. The locations of the primary docking site B and a secondary site C, which corresponds to the Mb Xe4 cavity, could be identified unambiguously. Finally, by comparing experiment and simulations it is also possible to identify protonation of its ε position (His$_\varepsilon$64 NgbCO) as the preferred heme-bound conformation in the wild type protein with a signal at $1935\ \mathrm{cm}^{-1}$.

1. Introduction

Globins are small respiratory heme proteins, and hemoglobin (Hb) and myoglobin (Mb) are arguably the most prominent members. Neuroglobin (Ngb) has only recently joined this family. The small heme protein is predominantly expressed in neuronal cells of vertebrates.[1] Its physiological role is not yet known.[2] However, it most likely involves the binding of a small ligand such as dioxygen, nitric oxide or carbon monoxide at the heme iron. Greenberg et al. suggested that Ngb plays a role in neuroprotection.[3] They revealed that both neuronal hypoxia and cerebral ischemia induce the expression of Ngb. Other suggestions of possible functions include the signaling of hypoxia[4] and radical scavenging.[5,6] A function as transporter and reservoir for O_2 is very unlikely due to the low O_2 affinity at physiological conditions and the higher autooxidation tendency compared to Mb and Hb.[7]

The polypetide chains of both Mb and Ngb adopt the characteristic three-over-three globin fold,[8] with a RMSD of the Cα atoms of 2.0 Å. In contrast to Mb, ferric and ferrous Ngb are hexacoordinated by the distal and proximal histidines His64 and His96. Hexacoordination has been proposed as a mechanism to regulate ligand affinity of heme proteins.[9,10,11,12,13,14] For an exogenous ligand such as CO to bind, the His64-iron bond has to be broken.[9,15,16] The pH-dependence observed in infrared experiments[10] suggested that the simplest description of ligand binding involves a two-step mechanism shown in Equation (1),

$$\mathrm{Ngb_H} \underset{k_{\mathrm{H}}}{\overset{k_{-\mathrm{H}}}{\rightleftharpoons}} \mathrm{Ngb_P} + \mathrm{CO} \underset{k'_{\mathrm{CO}}}{\overset{k_{\mathrm{CO}}}{\rightleftharpoons}} \mathrm{NgbCO}, \tag{1}$$

where $\mathrm{Ngb_H}$ represents the bis-His-hexacoordinated and $\mathrm{Ngb_P}$ the pentacoordinated system. A graphical illustration of Equation (1) is provided in Figure 1 which also illustrates the differences between His64 protonated at its δ and ε position (His$_\delta$64 and His$_\varepsilon$64).

Initially, it was assumed that the His64 side chain swings away from the heme iron to provide space for the incoming ligand. This flexibility of the imidazole side chain is known from other heme proteins such as Mb[17,18] and dehaloperoxidase,[19,20] where it induces conformational heterogeneity at the active site. This heterogeneity, in return, has pronounced influence on the physiological function. Comparison of the

---

[a] Dr. K. Nienhaus, Prof. G. U. Nienhaus
Institute of Applied Physics and Center for Functional Nanostructures
Karlsruhe Institute of Technology, 76128 Karlsruhe (Germany)
E-mail: uli.nienhaus@kit.edu

[b] S. Lutz, Prof. M. Meuwly
Department of Chemistry, University of Basel
Klingelbergstrasse 80, 4056 Basel (Switzerland)
Fax: (+41)-61-2673855
E-mail: m.meuwly@unibas.ch

[c] Prof. M. Meuwly
Department of Chemistry, Brown University
Providence RI 02912 (USA)

[d] Prof. G. U. Nienhaus
Department of Physics, University of Illinois at Urbana-Champaign
1110 West Green Street, Urbana, IL 61801 (USA)

© 2010 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

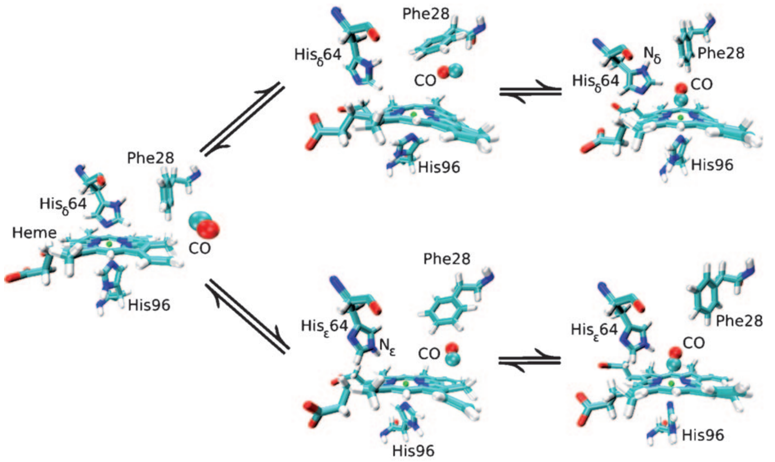

Figure 1. Graphical representation of Equation (1): the top trace illustrates the equilibrium between the bis-histidine-coordinated heme (far left) with His$_\delta$64, the penta-coordinated heme, and the CO-bound heme (far right; N$_\delta$ labelled). The lower trace is for His64 protonated at its ε position. Note that for the lower trace the protonation state has to change from δ to ε which is not shown explicitly.

three-dimensional structures of ferric and CO-ligated Ngb (NgbCO) revealed, however, that upon binding of an exogenous ligand, the His64 side chain essentially maintains its position and instead, the heme group slides deeper into the protein matrix to enable ligand binding.[21,22] Still, NgbCO has significant active-site heterogeneity as was shown by FTIR spectroscopy. The IR stretching frequency of heme-bound CO is extremely sensitive to even small changes of the local electric field. Low/high frequency bands of heme-bound CO are observed whenever a residue with a positive/negative partial charge is close to the oxygen atom of the CO molecule. In MbCO, for example, three stretching bands can be resolved that are associated with different orientations of the His64 side chain.[17,18] A structural interpretation based on mixed quantum mechanics/molecular mechanics calculations together with sampling from molecular dynamics simulations has recently confirmed the different structural origins of A_0 (open conformation of His64) and A_1/A_3 (closed conformation), in agreement with experimental work.[23] However, the differences between A_1 and A_3 still remain elusive except for the fact that the two states are structurally related.[17, 18, 23, 24] FTIR spectra of wt NgbCO display up to four stretching bands, dependent on pH. It was proposed that these bands also reflect different orientations of the His64 side chain.[14] A structural assignment of these states is, however, still lacking.

The bond between heme iron and CO can be broken by light. If photolysis is performed at cryogenic temperatures, the photodissociated CO ligands cannot escape from the protein matrix. Instead, the ligands concentrate in transient docking sites and give rise to narrow photoproduct bands between 2100 and 2160 cm^{-1}. Recent FTIR experiments on photodissociated wt NgbCO (Ngb–CO) have distinguished two transient CO docking sites: the primary docking site located most likely close to the binding site and a second, more remote site.[25] The spectral fingerprint alone is, however, not sufficient to determine the actual locations of these sites within the protein matrix. In Mb, both experimental and theoretical approaches have provided detailed information about location and orientation of the free ligand within these sites.[26,27,28] It should be noted that relating structure and spectroscopy is less problematic from computations for the photodissociated ligands than for Fe-bound CO.

The aim of the present study is to obtain more specific information on the active-site heterogeneity, to locate the intermediate docking sites, and to assign the spectral bands to specific sites and structural features. This knowledge will aid elucidating the molecular details of the binding process and, finally, the physiological function. For Mb, the structure-based interpretation of the IR spectra required intensive experimental and computational efforts.[29,26,27] By comparing a large number of mutant Mb proteins it was shown that different active site conformations have different ligand binding properties and that the docking sites are associated with migration pathways of the exogenous ligand into and out of the protein.[30,31,32,33,34]

Aromatic amino acids at position B10 of Mb had the most pronounced effects on the ligand binding properties. A comparison of wt Mb and Mb mutant L29F, which can be considered the Mb analog of wt Ngb, has shown that the bulky residue at position 29 (B10) shifts the stretching band of heme-bound CO and enhances the CO population in the more remote docking sites.[35] This 'trapping' effect was even more pronounced in Mb mutant L29W.[36] Based on those results, we have chosen wt Ngb, with a phenylalanine at position 28 (B10), and mutants F28L and F28W for this initial study on transient docking sites in Ngb. The F28L replacement was introduced to mimic the Mb distal side. We have included the F28W mutant to elucidate whether also in Ngb the size of residue 28 is a major determinant of the occupancies of the different transient docking sites after photodissociation.

In the present work, FTIR spectroscopy and atomistic simulations are combined to elucidate the structural origins of the split spectroscopic bands. Both, IR spectra of the Fe-bound and the photodissociated ligand in native, F28L, and F28W proteins have been recorded. They are then compared to results from molecular dynamics (MD) simulations at ambient and low temperature, corresponding to the conditions in the experiments. Finally, the findings are compared and discussed.

Methods

Sample Preparation: Site-directed mutagenesis of murine Ngb was performed using the Quikchange mutagenesis kit (Stratagene Europe, Amsterdam). Cysteine-depleted Ngb (C55S/C120S), which is referred to as wild-type (wt) Ngb in the following, was used as a template. Custom designed primers were ordered from MWG (MWG-Biotech GmbH, Ebersberg, Germany). The Ngb proteins were expressed in *E. coli* and purified according to Kriegl et al.[14] For the IR experiments, lyophilized protein was dissolved at a concentration of ~15 mM in cryosolvent (75 % glycerol/25 % potassium phosphate buffer (v/v), pH 8), stirred under a CO atmosphere and reduced with a two-fold molar excess of an anaerobically prepared sodium dithionite solution.

Fourier Transform Infrared (FTIR) Spectroscopy: FTIR transmission spectra were collected between 1800 cm^{-1} and 2400 cm^{-1} with a resolution of 2 cm^{-1}, using a FTIR spectrometer (IFS 66v/S, Bruker, Karlsruhe, Germany) equipped with an InSb detector. Sample loading and cryogenic equipment have been described previously.[37,38]

Photolysis difference spectra were calculated from transmission spectra taken before and after photolysis. We plot the bands of the heme-bound CO with negative absorption to indicate that the signal is missing after photolysis. Each band represents one so-called "taxonomic" or "A" substate[39] characterized by a particular heme pocket structure. The different stretching bands arise from electrostatic interactions of the CO dipole with distinctly different active-site structures. Upon photodissociation at low temperatures, the CO cannot escape, but only migrate within the protein because large-scale protein motions are arrested.[40] If well-defined transient docking sites exist, narrow photoproduct bands will emerge. Their frequencies again depend on the interactions between the CO dipole and the electric field at the intermediate site.[26,41,42] As these photoproduct bands emerge upon photodissociation, they are plotted with positive amplitudes. Note that the spectra have been scaled to the same areas. Typically, the integrated absorption is ~20x smaller in the CO photoproduct bands than in the bands of the heme-bound form.

Computational Procedures and Molecular Dynamics Simulations

All molecular dynamics (MD) simulations were carried out with the CHARMM program[43] and the CHARMM27 force field.[44] The initial NgbCO structure was generated by adding hydrogen atoms to the X-ray structure resolved by Nienhaus and co-workers[22] (PDB code 1W92). Both possible neutral tautomers of His64 ($His_\delta64$ and $His_\varepsilon64$) were investigated. The simulation is focused on the region surrounding the heme group; therefore, the stochastic boundary method was used to increase computational efficiency.[45] The heme pocket was solvated by three sequential layers of water molecules, forming a 25 Å sphere around the center of mass of the four porphyrin nitrogens of the heme prosthetic group, and a solvent boundary potential with a radius of 25 Å was applied to constrain the water molecules. A "reaction region" of radius 16 Å centered on the heme was defined, inside which the system was propagated with Newtonian dynamics. The dynamics of the buffer region between 16 and 20 Å from the center was described using Langevin dynamics. After solvation, the system consisted of 2465 atoms of NgbCO and 1450 water molecules, which leads to 6815 atoms in total. Single mutants F28L and F28W were generated by replacing Phe28 of wt NgbCO with the corresponding side chain. The interatomic stretching potential of the photodissociated CO ligand is given by an anharmonic rotational Rydberg-Klein-Rees (RRKR) potential.[46,47] The electrostatic interaction of CO with its environment is based on distributed multipole analysis (DMA)[48] which has been developed and successfully applied to simulations of the spectroscopic properties of CO[28,49,50] and to structurally assign spectroscopic features in photodissociated Mb–CO. In particular, an expansion of the electrostatic potential up to an atomic quadrupole moment on the C atom and an octupole moment on the O atom resulted in IR spectra in very good agreement with experimental data of CO in Mb.

Initially, the NgbCO was heated from 100 to 300 K for 50 ps and equilibrated for 100 ps. The final RMSD of the C_α atoms with respect to the crystal structure was 1.7 Å. The photodissociated system (Ngb–CO) was prepared following an additional equilibration simulation (200 ps) by storing 20 snapshots, each separated by 10 ps, from which the photodissociation event was initiated. The photodissociation was modeled by the "sudden" approximation:[51] The Fe–C bond was deleted and the potential parameters describing the bound state were replaced by those of the dissociated state. A repulsive term of the form r^{-12}, where *r* is the distance between heme iron and the CO center of mass, was added and all nonbonded interactions (electrostatic and van der Waals) between the CO and the heme plane were switched off. Force field parameters were taken from previous work on MbCO.[51,52] After 0.1 ps of dynamics, the Fe–C bond was fully dissociated. The repulsive term was removed and the nonbonded interactions were reintroduced. Subsequently, several 1 ns trajectories of photodissociated Ngb–CO were run at 300 K. Infrared spectra were extracted from 15 trajectories of wt Ngb–CO, the F28W and the F28L mutant with His64 in its δ ($His_\delta64$) and ε ($His_\varepsilon64$) protonation state. In the following, these realizations are denoted $WT_{\delta/\varepsilon}$, $F28W_{\delta/\varepsilon}$, and $F28L_{\delta/\varepsilon}$, respectively.

Calculation of Theoretical CO Infrared Spectra: The IR spectrum of the CO ligand was calculated from the Fourier transform of the dipole–dipole autocorrelation function $C_\mu(t)$ (μ is the dipole moment). The absorption lineshape can then be derived from Equation (2):[53]

$$\begin{aligned} I(\omega) &= \omega(1 - e^{-\hbar\omega/kT}) \int_0^\infty C_\mu(\tau) e^{-i\omega\tau} d\tau \\ C_\mu(\tau) &= < \mu(t_0)\mu(t_0+\tau) >_{t_0} \end{aligned} \quad (2)$$

Here, the $\langle\rangle_{t_0}$ symbol denotes averaging over all starting points t_0, and $C_\mu(\tau)$ is normalized. Furthermore, $\hbar$ is the reduced Planck constant, *k* is the Boltzmann constant and *T* is the temperature in Kelvin. In practice, $\mu(t)$ was accumulated over 2^n time origins, where *n* is an integer so that 2^n corresponds to between 1/3 and 1/2 of the trajectory with time origins separated by 1 fs leading to $C(\tau)$. The autocorrelation function can then be transformed using a fast Fourier transform and a Blackman filter to minimize noise.[54] A CO gas-phase spectrum obtained following this procedure results in a sharp C–O stretching peak at 2183 cm^{-1}. The difference of 39 cm^{-1} compared to the experimentally determined band at 2143 cm^{-1} is related to the classical treatment of the Morse potential and the use of the Verlet integrator in MD with a 1 fs timestep.[53] The integration of the equations of motions using the velocity–verlet algorithm with a time step of 1 fs shifts the frequency by 16 cm^{-1}. The remaining 23 cm^{-1} can be attributed to the classical treatment of the anharmonic oscillator, which can be shown to have an energy-dependent fundamental frequency.[55]

2. Results

2.1 FTIR Spectra Recorded at Cryogenic Temperatures

2.1.1 Active-Site Conformations in Ngb

Wild-type NgbCO displays three stretching bands of heme-bound CO at pH 7.2 (Figure 2), A_1 peaking at 1935 cm^{-1}, A_0 (1968 cm^{-1}) and A_2 (1978 cm^{-1}). (Note that the peak positions differ slightly from those reported in ref. [25] because spectra of the cysteine-substituted variants are presented in this study). The dominant band, which we had earlier denoted by A_1,[14] is likely related to the A_3 band that is present in those MbCO mutants that carry an aromatic residue at position B10.[56,57] In this conformation, the protonated $His_{\epsilon}64$ forms a strong hydrogen bond to the bound ligand, resulting in a substantial red-shift of the spectral band. In the A_0 conformation, the His64 side chain is expected to be protonated and rotated away from the bound ligand,[10] which renders the binding site less polar.[25] The high frequency of A_2 suggests that a negative partial charge is located near the CO oxygen. F28L NgbCO (pH 7.0) also displays three active-site conformations (Figure 2). The dominant A_1 band has shifted from 1935 to 1937 cm^{-1}; the narrower line indicates a more homogeneous conformation. A second band resides at 1957 cm^{-1}. As in the wild-type protein, there are also (at least) two high-frequency bands. We suggest that A_0 at 1967 cm^{-1} represents a species with minimal interaction between the CO and the distal heme pocket environment also in the mutant. An extended tail towards higher frequencies suggests that another band is present at ≈1974 cm^{-1}. The fraction of the A_2 and A_0 bands amounts to 58% as compared to 32% in wt Ngb. To obtain further information on these conformations, we have lowered the sample pH to 5 and again measured the absorption spectrum. It shows three bands, centered at 1937, 1957 and 1973 cm^{-1}. The contribution at 1967 cm^{-1} is no longer present. A comparison with the spectrum at 290 K suggests, however, that the two species at 1957 and 1973 cm^{-1} are related to the band at 1967 cm^{-1} and, therefore, are most likely associated with a protonated His64. F28W NgbCO displays a dominant band at 1939 cm^{-1} (Figure 2). There are also high-frequency bands in the range 1960–1980 cm^{-1}, but they are not resolvable at 4 K because of their very low amplitude (these bands have altogether only ~5% of the total area). All three samples have in common that the fraction of the A_0 species is reduced upon cooling (Figure 2). The effect increases in the order Leu28 < Phe28 < Trp28, implying that a large aromatic entity at position B10 stabilizes the closed conformation of the protein.

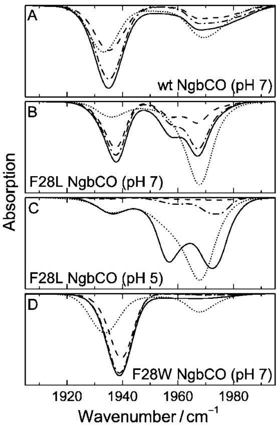

Figure 2. Experimental FTIR absorption spectra of A) wt NgbCO, pH 7, B) F28L NgbCO, pH 7, C) F28L NgbCO, pH 5 and D) F28W NgbCO, pH 7 in the spectral region of heme-bound CO (1900–2000 cm^{-1}). Negative amplitudes indicate that the absorption due to heme-bound CO is missing after photolysis. -----: fraction photolyzed after 1 s illumination at 4 K. –•–•: fraction photolyzed after slow cooling from 140 to 4 K under continuous illumination. ——: overall bound CO (4 K). •••••: overall bound CO (290 K). Solid and dotted lines serve as reference and therefore are also plotted with negative amplitude.

2.1.2 Photoproduct Spectra of Ngb

It is difficult to predict putative transient docking sites in Ngb because the protein has a large internal cavity of ~300 Å^3, which could provide several docking sites. In addition, a small, apolar cavity corresponding to the Xe4 cavity in Mb exists. In the hexa-coordinate state, the huge open volume connects the distal and proximal sides of the heme and has a direct opening to the bulk solvent.[21] This tunnel is maintained upon binding of an exogenous ligand, however, its accessibility from the bulk is restricted because of decreased flexibility of the protein.[22]

We have performed two different illumination protocols to probe possible CO docking sites in Ngb. Initially, the NgbCO samples were illuminated for 1 s at 4 K, which transfers the system into the Ngb–CO state where CO is expected to populate preferentially the primary site B close the binding site. In addition, the samples were also cooled under steady illumination from 140 to 4 K (ramp rate 0.3 Kmin^{-1}), which is a convenient method to screen for other, more remote transient docking sites.[58,59]

The photolysis yields achieved by the different illumination protocols are apparent from Figure 2. The solid lines represent the total bound CO at 4 K; dashed and dashed-dotted lines show the fractions which are photolyzed by 1 s and slow-cool illumination. It is obvious that, in contrast to Mb, 1 s illumination is not sufficient to keep all CO ligands in Ngb photodissociated for the time it takes to record a spectrum (100 s). Even after extended illumination a small fraction of CO remains bound to the heme iron. The photoproduct yield is particularly low in F28L at pH 5 (Figure 2C).

The photoproduct spectra taken after 1 s illumination at 4 K and after slow cooling from 140 to 4 K under continuous illumination are shown in Figures 3A and B. After 1 s-illumination, the photoproduct spectrum of wt Ngb–CO shows a doublet of bands at 2128 and 2136 cm^{-1} (Figure 3A). Exact peak positions were determined by fits with multiple Gaussians and are compiled in Table 1. The same bands are present in the spectra of mutants F28W and F28L. In addition, mutant F28L Ngb–CO ex-

© 2010 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 1. Band positions of the IR stretching absorption of heme-bound and photodissociated CO in mutant Ngb samples, as determined at 4 K and pH 7, with an estimated experimental error of ±0.5 cm^{-1}. The numbers in parentheses indicate the percentage of each spectral band contributing to the total integrated absorbance.

		Heme-bound CO				Photolyzed CO					
WT	4 K illum.	1935 (62)	1953 (5)	1966 (23)	1976 (9)	2105 (5)	2114 (5)	2123 (7)	2129 (37)	2131 (7)	2136 (39)
	Slow cool								2128 (19)	2134 (75)	2139 (6)
F28L	4 K illum.	1937 (42)	1957 (20)	1967 (38)		2107 (6)	2120 (39)	2130 (24)	2136 (11)	2142 (17)	2149 (3)
	Slow cool					2108 (2)	2122 (23)	2131 (36)	2134 (26)	2141 (11)	2148 (2)
F28W	4 K illum.	1939 (95)		1965 (~2)	1978 (~2)			2117 (4)	2128 (46)	2136 (50)	
	Slow cool								2128 (16)	2134 (82)	2137 (2)

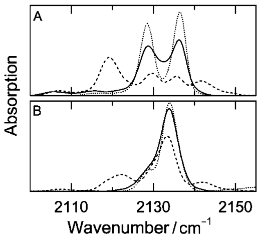

Figure 3. Experimental FTIR photolysis difference spectra at 4 K in the spectral region of the photolyzed CO (2100–2160 cm^{-1}). Positive amplitudes indicate that these bands are created by photolysis. A) FTIR photoproduct bands after 1 s illumination at 4 K. B) photoproduct bands after slow-cool illumination from 140 to 4 K. All spectra have been scaled to unit area. (——): wt Ngb, (- - - -): F28L, (•••••): F28W.

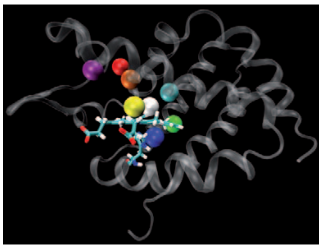

Figure 4. Structure of Ngb showing the docking sites identified by MD simulations as colored spheres. Color code: white (site B′), yellow (site B″), orange (site F), red (site G), purple (site H), cyan (site C, Xe4), green (site E, Xe2), blue (site D, Xe1).

hibits a large band at 2119 and a smaller one at 2142 cm^{-1}. After cooling the samples from 140 to 4 K under continuous illumination, the photoproduct spectra have changed significantly, indicating that different docking sites have been populated (Figure 3 B). For wt Ngb–CO, the bands at 2128 and 2136 cm^{-1} have disappeared. Instead, a dominant band at 2134 cm^{-1} has been created. Essentially the same changes are observed for F28L and F28W Ngb–CO. However, in F28L Ngb–CO, residual intensity of the photoproduct bands measured after 1 s illumination is still observed after slow-cool illumination. These data clearly indicate that wt Ngb–CO and F28W Ngb–CO share very similar transient docking sites. In F28L Ngb–CO, additional sites are populated as indicated by the photoproduct bands at 2119 and 2142 cm^{-1}.

2.2 Molecular Dynamics Simulations at 300 K

To obtain structural information about the photodissociated CO in Ngb—that is, to relate structure and spectroscopy—MD simulations were carried out at different temperatures for wt Ngb and the F28W and F28L mutants. The 1 ns trajectories indicate that at 300 K, the free CO ligand can sample a number of transient sites that are shown as colored spheres in Figure 4. Spectra corresponding to specific docking sites can be extracted from trajectories in which a particular site is sampled for at least a consecutive 400 ps (resolution of 0.03 cm^{-1}). It was previously found that such an approach is a meaningful way to correlate spectroscopy with structural information in myoglobin.[28] Only the four sites closest to the heme iron (B′ shown in white, B″ in yellow, C in cyan, and F in orange) are sufficiently sampled to allow spectroscopic data to be extracted. The averaged IR spectra of CO at these sites are shown in Figure 5, labeled according to the different mutant samples (see caption of Figure 5). Individual peak positions (indicated by the arrows) were determined by fitting Gaussian functions to the calculated band profiles *I*(ω). They are summarized in Table 2.

Infrared spectra for CO in docking site B′ are shown in Figure 5 A. For WT$_\varepsilon$ (black trace) the spectrum shows two bands, red-shifted by 11 and 2 cm^{-1} relative to gas-phase CO. The intensity of the two peaks is unequal and may depend on the limited conformational sampling. The spectral signature for the F28W$_\varepsilon$ mutant is similar but with a more pronounced low-frequency peak. The two peaks are again shifted to the red from gas-phase CO and the splitting is 12 cm^{-1}, 3 cm^{-1} larger than for WT$_\varepsilon$, which is in qualitative agreement with experiment (see Figure 3). Conversely, the spectrum for F28W$_\delta$ (gray) consists of two peaks, one blue- and the other one red-shifted from gas-phase CO, split by only 5 cm^{-1} and does not correspond to the experimental spectrum. In Figure 6, spectra of dissociated CO obtained from individual runs together with the corresponding averaged structures are shown. The line color of a spectrum is repeated in the frame of the related structure plot. The nature of residue 28 (F, W, or L) is indicated in the structures. Figure 6 A shows two individual spectra of CO at site B′, extracted from WT$_\varepsilon$. The spectra in black and red

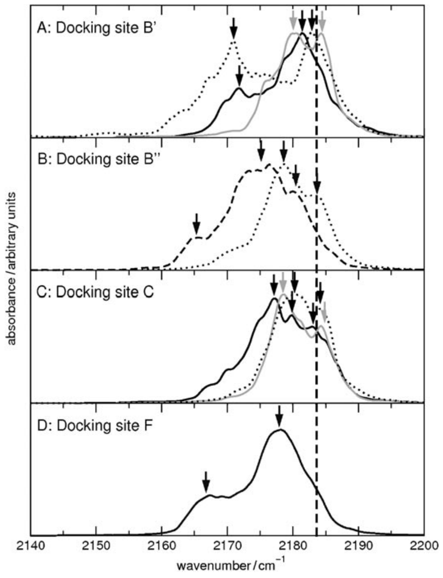

Figure 5. Docking site specific IR spectra of CO calculated from molecular dynamics simulations at 300 K at sites B′, B″, C and F. Each spectrum is averaged over three to six single spectra extracted from 400 ps of independent MD simulations each. ——: WT$_\varepsilon$, -----: F28L$_\varepsilon$, •••••: F28W$_\varepsilon$, grey: F28W$_\delta$. The vertical dashed line is the computed stretching band for free CO. Line styles correspond to those from Figure 3.

Table 2. Docking site specific band positions (in cm^{-1}) of the theoretical spectra for all Ngb–CO systems studied (wt, F28L and F28W mutants each with both His64 tautomers). Bands from the selected spectra shown in Figure 5 are highlighted in boldface.

System/Site	B′	B″	C	F
WT$_\delta$	2170 2175 2179	2173 2181		2169 2173 2177 2181
WT$_\varepsilon$	**2172** **2181**	2164 2168 2171 2178	**2177** **2180** **2183**	**2167** **2178**
F28L$_\delta$		2173 2177 2180		
F28L$_\varepsilon$		**2165** **2175** **2181**		
F28W$_\delta$	**2180** **2185**		**2179** **2184**	
F28W$_\varepsilon$	**2171** **2183**	**2179** **2184**	**2180** **2184**	

(Figure 6 A) have a dominant peak at 13 cm^{-1} and at 5 cm^{-1} relative to the theoretical gas-phase CO band (2183 cm^{-1}). In both structures, the CO is located above the heme-pyrrol C (white sphere in Figure 4). The main structural difference is the relative orientation of the CO ligand and residue Phe28. In the black-framed averaged structure the CO ligand has a distance of 5.7 Å to the N$_\varepsilon$ of His64, the CO-carbon atom points towards the heme-iron and the Phe28 side chain ring is perpendicular to the heme plane. In the red-framed structure, the distance between CO and the His64 side chain is increased by 0.5 Å. More importantly, the ligand is rotated by ≈180°, with the oxygen atom pointing towards the heme-iron. The Phe28 side chain is located on top of the CO and parallel to the heme plane.

Figure 5 B shows spectra for F28W$_\varepsilon$ (•••••) and F28L$_\varepsilon$ (-----). Comparison with panel A for the F28W mutant establishes that the experimental spectrum corresponds to site B′. For site B″ the F28W$_\varepsilon$ spectrum is essentially unshifted from gas-phase CO and no definite splitting is found. For the F28L$_\varepsilon$ mutant a rather broad, strongly red-shifted spectrum with a band shifted by almost 30 cm^{-1} is found. Because for this mutant site B′ is not extensively sampled, no corresponding spectra could be calculated. However, the strongly red-shifted signal for F28L$_\varepsilon$ hints towards assigning the experimental 2119 cm^{-1} to a B″ site. Other spectra of structures with His$_\varepsilon$ configuration calculated for site B″ are rather broad, barely red-shifted and unstructured. Site B″ is located almost on top of the heme iron (yellow sphere in Figure 4). The black and blue spectra in Figure 6 B represent essentially identical conformational substates of WT$_\varepsilon$ but without/with water interacting with the carbon atom of CO. A single water molecule in close contact to the carbon atom of CO in the gas phase was found to lead to blue-shifts of ≈12 cm^{-1} of the calculated IR bands (not shown), which is consistent with the

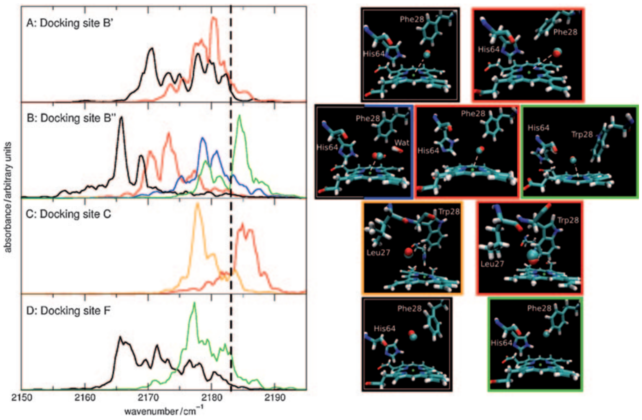

Figure 6. Left: Selected site-specific photoproduct spectra at 300 K from 400 ps of MD simulation each. The collection of bands in each present docking site covers the complete bandwidth from the corresponding average spectra in Figure 5. Right: Average structures from the trajectories sampling the spectra on the left. The color-coded borders relate structure and spectrum. The set of spectra displayed consists of native and mutant Ngb systems.

large difference encountered between the black and the blue spectra in Figure 6 B. In the red-framed structure (Figure 6 B), His$_\delta$64 has adopted a slightly bent average orientation, with N$_\varepsilon$ pointing towards the solvent. Therefore, the electrostatic interaction with the CO is reduced and the corresponding spectrum is less red-shifted (≈ -10 cm^{-1}) compared to the black one. The green spectrum represents a typical subconformation of F28W$_\varepsilon$ with a completely swung out His$_\varepsilon$64 side chain, which considerably reduces the interaction between His64 and CO. As a result, only small spectral shifts of -5 to 0 cm^{-1} relative the gas-phase band are observed. Overall, small details in the orientation of the ligand relative to the protein are found to have a pronounced influence on the IR spectra. This observation is in line with previous investigations on CO in Mb.[28,60,61]

Averaged docking site C spectra are shown in Figure 5 C for WT$_\varepsilon$ (solid black), F28W$_\delta$ (solid gray), and F28W$_\varepsilon$ (dotted black). They are consistently unstructured, and their maximum is shifted by -6, -4 and -3 cm^{-1} to the red from free CO. Comparison with the experimental spectrum shows that the maximum intensity peak is expected to lie between the maxima of the two peaks for WT and F28W. This is also found from the simulations. For WT$_\varepsilon$, the maxima of the two peaks in B′ are red-shifted by -2 and -11 cm^{-1}, whereas the maximum intensity peak in docking site C is shifted by -3 cm^{-1}. The same holds for the F28W$_\varepsilon$ spectra. For F28L no spectra of site C were calculated at 300 K because the residence time of the ligand in site C was not sufficiently long to extract meaningful data. Individual contributions to the average spectrum of CO in docking site C calculated for mutant F28W are shown in Figure 6 C. The location within the protein is indicated by the cyan-colored sphere in Figure 4. It corresponds to the Xe4 site in Mb. The two average structures shown represent the two species which contribute to the overall average spectrum of this docking site (Figure 5 C). They display very similar arrangements of the amino acid side chains but differ in the orientation of the CO. The main band, red-shifted by ≈ -4 cm^{-1}, represents a CO molecule in the Fe–CO conformation. The high-frequency band (red) corresponds to a CO molecule in the opposite orientation (Figure 6).

For docking site F, the line shape for WT$_\varepsilon$ (black) is shown in Figure 5 D. The spectrum is dominated by two peaks, red-shifted by -16 cm^{-1} and -5 cm^{-1}. Individual spectra of CO in site F of WT$_\varepsilon$ and averaged conformations of the amino acids are shown in Figure 6 D. In the structure corresponding to the high frequency spectrum (green, Figure 6 D) the CO-carbon atom loosely points towards the His64 residue. In the other structure (black frame and black spectrum in Figure 6 D) the CO is oriented in a parallel fashion to the edge of the imidazole ring, pointing to the active site. A minor structural difference arises from the Phe28 side chain which is arranged slightly more perpendicular towards the heme plane in the first structure.

In summary, the 300 K simulations provide strong evidence that the experimental spectra for WT and the F28W mutant in Figure 3 A correspond to site B′. This is clearly borne out for F28W, and less clearly (due to the low intensity of the stronger red-shifted peak for the WT) for the wild type protein. In addition, the results also establish that His64 is protonated at its ε-position because all spectra for His64$_\delta$ bear little resemblance with the experimental findings and, more generally, the spectra are essentially unshifted relative to gas-phase CO. For the F28L mutant only trajectories in site B″ could be analyzed in a meaningful way. They lead to the conclusion that, most likely, the absorption at 2119 cm^{-1} found in the experiment corresponds to site B″. All other structure in the spectrum is washed out and sampling at lower temperature (see next section) may provide additional insight. Finally, the shape of the site C spectra is simpler for WT$_\varepsilon$ and F28W$_\varepsilon$ and the maximum intensity of the band falls between the two maxima of the B′ spectra, which agrees also with experiment (see Figure 3 B).

2.3 Molecular Dynamics Simulations at 20 K

FTIR photolysis difference spectra are usually measured at $T < 10$ K so as to prevent geminate rebinding within the time it takes to collect a spectrum of the photolyzed state (100 s). In contrast, MD simulations are typically performed at 300 K. The different premises of experimental and computational approaches potentially make direct comparison difficult. Therefore, additional simulations were performed at low temperature.

At 20 K, only the migration to docking sites B′ and B″ was observed in the trajectories. Figure 7 shows photoproduct spectra of WT$_\varepsilon$ (upper graph) and F28L$_\varepsilon$ (lower graph) from 1 ns of MD at 20 K. The corresponding structures are displayed above and below, again with color-coded frames. All spectral bands are narrow. The calculated photoproduct bands of WT$_\varepsilon$ Ngb cover a range of ≈ -10 cm^{-1} predominantly to the red of free CO, which compares favorably with the simulations at 300 K. Moreover, the snapshots for $T = 20$ K have CO located within site B′, thus leading to similar spectral shifts as in the 300 K simulations (see Figure 5 A). Contrary to the average 300 K structures the carbon atom of CO points towards His64 in all the wt 20 K simulations indicating that considerably smaller conformational differences can still lead to appreciable spectral shifts.

The low temperature photoproduct spectra of F28L$_\varepsilon$ cover a significantly wider frequency range (from 2167 (black) to 2185 cm^{-1} (blue)). They are in good qualitative agreement with the site B″ spectra from simulations at 300 K, shown in Figure 5 B. Because CO in these simulations is even closer to His64 (N$_\varepsilon$H–CO center-of-mass ~3.1 Å) and the heme iron (Fe–CO center-of-mass ~3.5 Å) than in the corresponding 20 K simulations for WT$_\varepsilon$ (~4.3 Å and 4.0 Å) the spectroscopy is particularly susceptible to small structural changes.

Comparison of the low-temperature spectra with the 300 K spectra shows differences in the width and, therefore, in the resolution of the individually calculated spectra. However, the frequency range and the relative shifts of all 20 K spectra in each system correspond reasonably well with their 300 K counterparts in Figures 5 A and B, suggesting that spectra from room temperature simulations (better sampling) contain meaningful information to interpret the experimental low-temperature data.

© 2010 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

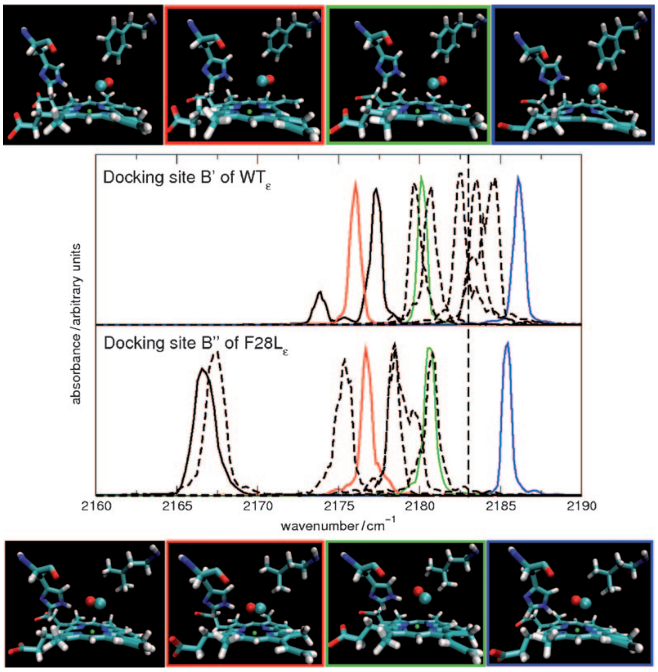

Figure 7. IR spectra of CO in docking site B' of $WT_ε$ Ngb–CO (top) and in docking site B'' of $F28L_ε$ Ngb–CO (bottom), obtained from simulations at 20 K. Color-coded snapshots of structures associated with particular bands (solid lines) are plotted on top and below. Additional 20 K spectra for $WT_ε$ Ngb–CO and $F28L_ε$ Ngb–CO from independent 1 ns simulations are shown (-----) to provide an impression of the subtle structural differences that lead to the width of the observed experimental spectra (Figure 3). The vertical dashed line represents the location of the gas-phase CO stretching mode.

3. Discussion and Conclusions

We have combined experimental (FTIR spectroscopy) and theoretical (MD simulations) approaches to characterize the different conformational substates present in heme-bound NgbCO and photodissociated Ngb–CO. The existence, identification and structural assignment of ligand-bound and photodissociated states is essential to analyze and characterize ligand migration pathways and rebinding trajectories which are most likely directly related to the function of Ngb. Both techniques have their advantages and their method-related drawbacks. FTIR photoproduct spectra can provide information on the number of conformational substates and also the number of visited transient docking sites, but lack direct structural information. Here, site-specific spectra obtained from MD simulations can assist in interpreting the experimental data. Simulations, on the other hand, are limited in terms of ensemble size and the computationally accessible timescale.

Simulations were carried out for wt Ngb–CO using both tautomers, $His_δ64$ and $His_ε64$, of the neutral His64. The different tautomerization, protonation and isomerization states of this residue are known to constitute structural heterogeneity at the active site of heme proteins.[26] During the equilibration of $WT_ε$ at 300 K, the imidazole side chain rotated by 180° around the $C_β$–$C_γ$ axis. As a result, the free electron pair on $N_δ$ can form a hydrogen bond to a solvent molecule. Moreover, the protonated $N_ε$ is positioned slightly closer to the bound ligand, which allows for a stronger interaction. The same conformation of His64 is seen in L29F MbCO.[62] Both, L29F MbCO and $WT_ε$ NgbCO in the rotated form, have a dominant A substate band at about 1930 cm^{-1}. Therefore, we suggest that the A_1 band corresponds to the $His_ε64$ NgbCO isomer. The discrepancy compared to the crystal structure is explained by the fact that protein crystallography is not able to distinguish between N and C atoms. Hence, the atom assignment in NgbCO was adopted from the one of the ferric form (1Q1F), where the mapping was unambiguous due to the covalent bond between the $N_ε$ and the heme iron.[21] A_0 represents the open conformation, with His64 protonated and rotated out of the distal pocket.

Two possible conformations may be considered for the high-frequency, A_2 band:[25] i) The π-electron system of Phe28 interacts with the bound ligand; we note that the F28L mutation would abolish this interaction. ii) A neutral $His_δ64$ tautomer points its non-bonding electron pair on $N_ε$ at the bound ligand. This conformation is physiologically of particular relevance because only the $His_δ64$ tautomer can form the hexa-coordinate ferrous (bis-histidine) species after dissociation of the exogenous ligand. A combination of both effects is also conceivable: tautomerism of His64 may induce reorientation of Phe28, which in turn could result in the electrostatic interaction between its π-electron system and the bound ligand. The absorption spectra of the F28L mutant indicate a marked decrease of the A_2 species, which supports possibility (i). However, in mutant F28W, the population in A_2 is negligible, which contradicts the hypothesis that this band arises from an interaction of the CO and the π-electron system of residue 28. Possibility (ii)—interaction between a neutral $His_δ64$ and CO—is, however, also not fully supported by the data. A comparison of the FTIR spectra of F28L (pH 5) measured at 4 and 290 K shows that both A_2 and the band at 1957 cm^{-1} are A_0-type bands and, therefore, most likely represent conformations with a protonated His64 side chain. Preliminary analysis of the MD trajectories reveal that, indeed, Phe28 can interact with CO via π-interaction with a corresponding spectroscopic A_2 state whereas Trp28 is pushed away (i.e. no π-interaction) from CO through interactions with neighboring residues which has a similar effect as the F28L mutation. This agrees qualitatively with model (i) in that π-interaction with Phe28 is responsible for the spectroscopic state A_2. However, additional effects, such as the relative position and orientation of $\text{His}_\varepsilon 64$ changes in the active site should be taken into account for a more complete analysis.

When the Fe-CO bond is ruptured at cryogenic temperatures, the CO cannot escape from the protein matrix but rather becomes trapped at specific internal sites. Each site is characterized by a photoproduct band. Often, a doublet of bands is seen due to different orientations of the CO in the site.[41,42] In an experimental study of ligand migration in wt Ngb–CO,[25] two transient docking sites could unambiguously be resolved; evidence of a third site was present, but weaker. Based on the spectra and rebinding behavior of the CO, they were tentatively denoted as primary site B and secondary site C.

The experimental photoproduct spectrum of wt Ngb–CO associated with site B displays two prominent bands at 2128 and 2136 cm^{-1} which are also found for mutants F28W (prominent) and F28L (less prominent, see below), suggesting that the electric field at the docking site is little influenced by the present modifications of residue 28 (Figure 3). The MD simulations at 300 K suggest that a voluminous side chain at position 28 divides site B into two subsites, B′ and B″ whereas low-temperature trajectories of wt Ngb–CO show migration to site B′ only. Apparently, site B″ is not populated at cryogenic conditions, which implies that the experiment probes ligand rebinding from site B′. Photoproduct spectra of WT_ε Ngb for site B′ from the MD simulations at 300 K (Figure 5 A) show close similarity to the experimental data, both with respect to spectral shift and shape and thus allow to relate spectroscopy with structure. They display two regions with increased band densities ($\approx -11\ \text{cm}^{-1}$ and $\approx -2\ \text{cm}^{-1}$ red-shifted from gas-phase CO), which are likely to be related to the bands at $-15\ \text{cm}^{-1}$ and $-7\ \text{cm}^{-1}$ seen in the experiment. Both doublets show almost identical band splitting between 8 and 9 cm^{-1}. For F28W Ngb–CO the same but more equally pronounced band splitting compared to the wild type protein (Figure 3 A) is observed experimentally. This is also seen in the simulated spectra of WT_ε and F28W_ε but not in the corresponding F28W_δ mutant (Figure 5 A). Additionally, the splitting for F28W_ε compared to WT_ε is increased by 3 cm^{-1}, in agreement with experiment. Although the B′ spectrum of F28W_δ shows two bands, the splitting is clearly too small and not sufficiently red-shifted to be related to the experimental spectrum (see Figure 3 A). This observation is a further reassurance for the major His_ε fraction in the experimental protein sample.

Structures corresponding to specific WT_ε B′ spectra are compared in Figure 6 A. The CO molecule is slightly displaced from the heme-center towards the protein interior and populates two different orientations. The more red-shifted band arises from a Fe–CO conformation. This scenario resembles the situation in wt Mb, where it was possible to unambiguously relate each spectral band with a particular CO orientation.[27,28] Contrary to Ngb, however, the more red-shifted band was assigned to Fe–OC in Mb. However, as also discussed below, the population of two distinct substates (Fe–CO and Fe–OC) is less pronounced in Ngb than in Mb, probably because of the larger and more solvent-accessible active site in Ngb.

The FTIR spectrum of F28L Ngb obtained after 1 s illumination shows two additional photoproduct bands, a major band at $-24\ \text{cm}^{-1}$ (2119 cm^{-1}) and an additional high-energy band at $-1\ \text{cm}^{-1}$ (2142 cm^{-1}), yielding a rather broad overall spectrum. This finding is reflected in the spectrum obtained from simulations at 20 K (Figure 7). It also spans a significantly wider frequency range than that of wt Ngb–CO at cryogenic temperatures. The structures corresponding to the calculated B site spectra of F28L Ngb–CO at 20 K (Figure 7) indicate that CO is located more towards the heme center, at a more B″-like position. As the aromatic side chain at position 28 is missing in F28L Ngb, only a single B-site exists. The different bands are mainly related to different orientations of the His64 side chain with respect to the CO ligand. Large red-shifts arise from strong interactions between His64-NH and the CO-oxygen whereas small red-shifts or even blue-shifts occur if the His64 side chain swings out of the heme pocket.

The additional site B″ is not unique to Ngb, but was earlier described in MD simulations on L29F MbCO.[63] The very large Stark splitting observed with FTIR spectroscopy for CO trapped in the primary site in L29F Mb–CO (24 cm^{-1}) and the observation that rebinding starts already at 4 K imply that, at cryogenic temperatures, the CO ligands are trapped at site B″, right on top of the heme iron.[35] In the L29W MbCO mutant, two active site conformations, A_{I} and A_{II}, exist that differ in the orientation of the Trp29 side chain.[32,64] Below 180 K, the protein is frozen in one of the major conformations; transitions between these conformations are not possible.[39,40] Whereas in conformation A_{I} only site B′ is available to the photolyzed ligand at 4 K, in A_{II} only site B″ can be populated. Structurally, the orientation of the indole side chain in A_{II} corresponds to that of the Phe29 side chain in L29F Mb.

The FTIR photoproduct band associated with site C (see Figure 3 B) corresponds best to the site C spectrum determined by MD simulations (Figure 5 C), although here, the major peak is less red-shifted than in the experiment. Site C is equivalent to the Xe4 site in Mb. The highest relative occupancy of site C was found for mutant F28W Ngb–CO. Theoretical spectra for $\text{His}_\delta 64$ and $\text{His}_\varepsilon 64$ protonation are not significantly different, probably due to the rather large distance of CO to His64. In F28L Ngb–CO, owing to the considerably smaller residue at position 28, a significant fraction of CO ligands remained at the primary B-site. If we compare ligand migration in wt Mb and mutants with an aromatic residue at position 29, we also observe increased trapping in the Xe cavities. For docking site F, calculated IR bands appear within the same spectral range as the bands of site B″. From experiment no direct evidence is available for population of this site and the spectroscopic features may be partially obscured by the bands of sites B and C. Moreover, based on the location of site F, one would expect it to be populated most likely via migration from site B″ which is essentially not available in wt Ngb–CO (see above). Only in mutant F28L Ngb–CO, the CO is found at a B″-like site at 20 K.

In summary, the combination of experimental and theoretical methods used in this work allows the structural interpretation of the CO photoproduct spectra in Ngb and provides a rationale for the heme-bound spectra. It is found that His64 is present in its His64$_\varepsilon$ protonation state and that photodissociated CO populates docking site B′ for WT and F28W, whereas docking site B″ is also available for the F28L mutant. The C-site spectra could also be assigned for WT and F28W. Despite the very different physiological functions, the detailed analysis revealed marked similarities between Ngb and Mb. As in MbCO, the His$_\varepsilon$64 protonation state of His64 is favored in NgbCO over its His$_\delta$64 tautomer. After photodissociation, the CO ligands initially become trapped at site B. In Ngb, a large residue at position 28 divides this site into two subsites, B′ and B″, which is reminiscent of Mb. In addition, the voluminous side chain enhances the ligand population in site C. This is also seen in Mb, in particular in L29F.[33,61,63]

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) and the State of Baden-Württemberg through the Center for Functional Nanostructures (CFN), and DFG grant Ni 291/3 (to GUN), and by the Schweizerischer Nationalfonds grant 200021-117810 (to MM). We thank Uwe Theilen for expression and purification of the recombinant proteins used in this study.

Keywords: distributed multipole · globins · IR spectroscopy · ligand migration · molecular dynamics

[1] T. Burmester, B. Weich, S. Reinhardt, T. Hankeln, *Nature* **2000**, *407*, 520–523.
[2] K. Nienhaus, G. U. Nienhaus, *IUBMB LIFE* **2007**, *59*, 490–497.
[3] D. A. Greenberg, K. Jin, A. A. Khan, *Curr. Opin. Pharmacol.* **2008**, *8*, 20–24.
[4] J. V. Esplugues, *Br. J. Pharmacol.* **2002**, *135*, 1079–1095.
[5] S. Herold, A. Fago, R. E. Weber, S. Dewilde, L. Moens, *J. Biol. Chem.* **2004**, *279*, 22841–22847.
[6] T. R. Weiland, S. Kundu, J. T. Trent, J. A. Hoy, M. S. Hargrove, *J. Am. Chem. Soc.* **2004**, *126*, 11930–11935.
[7] A. Fago, C. Hundahl, S. Dewilde, K. Gilany, L. Moens, R. E. Weber, *J. Biol. Chem.* **2004**, *279*, 44417–44426.
[8] A. Pesce, S. Dewilde, M. Nardini, L. Moens, P. Ascenzi, T. Hankeln, T. Burmester, M. Bolognesi, *Structure* **2003**, *11*, 1087–1095.
[9] S. Dewilde, L. Kiger, T. Burmester, T. Hankeln, V. Baudin-Creuza, T. Aerts, M. C. Marden, R. Caubergs, L. Moens, *J. Biol. Chem.* **2001**, *276*, 38949–38955.
[10] K. Nienhaus, J. M. Kriegl, G. U. Nienhaus, *J. Biol. Chem.* **2004**, *279*, 22944–22952.
[11] M. S. Hargrove, E. A. Brucker, B. Stec, G. Sarath, R. Arredondo-Peter, R. V. Klucas, J. S. Olson, G. N. Phillips, *Structure* **2000**, *8*, 1005–1014.
[12] J. B. Wittenberg, M. Bolognesi, B. A. Wittenberg, M. Guertin, *J. Biol. Chem.* **2002**, *277*, 871–874.
[13] J. T. Trent, A. N. Hvitved, M. S. Hargrove, *Biochemistry* **2001**, *40*, 6155–6163.
[14] J. M. Kriegl, A. J. Bhattacharyya, K. Nienhaus, P. Deng, O. Minkow, G. U. Nienhaus, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 7992–7997.
[15] A. Fago, A. J. Mathews, S. Dewilde, L. Moens, T. Brittain, *J. Inorg. Biochem.* **2006**, *100*, 1339–1343.
[16] M. Brunori, A. Giuffre, K. Nienhaus, G. U. Nienhaus, F. M. Scandurra, B. Vallone, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 8483–8488.
[17] F. Yang, G. N. Phillips, *J. Mol. Biol.* **1996**, *256*, 762–774.
[18] J. D. Muller, B. H. McMahon, E. Y. T. Chien, S. G. Sligar, G. U. Nienhaus, *Biophys. J.* **1999**, *77*, 1036–1051.
[19] K. Nienhaus, E. Nickel, M. F. Davis, S. Franzen, G. U. Nienhaus, *Biochemistry* **2008**, *47*, 12985–12994.
[20] K. Nienhaus, P. C. Deng, J. Belyea, S. Franzen, G. U. Nienhaus, *J. Phys. Chem. B* **2006**, *110*, 13264–13276.
[21] B. Vallone, K. Nienhaus, M. Brunori, G. U. Nienhaus, *Proteins Struct. Funct. Bioinf.* **2004**, *56*, 85–92.
[22] B. Vallone, K. Nienhaus, A. Matthes, M. Brunori, G. U. Nienhaus, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 17351–17356.
[23] M. Devereux, M. Meuwly, *Biophys. J.* **2009**, *96*, 4363–4375.
[24] H. Ishikawa, K. Kwak, J. K. Chung, S. Kim, M. D. Fayer, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 8619–8624.
[25] K. Nienhaus, G. U. Nienhaus, *J. Biol. Phys.* **2005**, *31*, 417–432.
[26] K. Nienhaus, J. S. Olson, S. Franzen, G. U. Nienhaus, *J. Am. Chem. Soc.* **2005**, *127*, 40–41.
[27] M. Meuwly, *ChemPhysChem* **2006**, *7*, 2061–2063.
[28] N. Plattner, M. Meuwly, *Biophys. J.* **2008**, *94*, 2505–2515.
[29] M. Lim, T. Jackson, P. Anfinrud, *J. Chem. Phys.* **1995**, *102*, 4355–4366.
[30] I. Schlichting, J. Berendzen, G. N. Phillips, R. M. Sweet, *Nature* **1994**, *371*, 808–812.
[31] H. Hartmann, S. Zinser, P. Kominos, R. T. Schneider, G. U. Nienhaus, F. Parak, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 7013–7016.
[32] A. Ostermann, R. Waschipky, F. G. Parak, G. U. Nienhaus, *Nature* **2000**, *404*, 205–208.
[33] F. Schotte, M. H. Lim, T. A. Jackson, A. V. Smirnov, J. Soman, J. S. Olson, G. N. Phillips, M. Wulff, P. A. Anfinrud, *Science* **2003**, *300*, 1944–1947.
[34] M. Schmidt, K. Nienhaus, R. Pahl, A. Krasselt, S. Anderson, F. Parak, G. U. Nienhaus, V. Srajer, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 11704–11709.
[35] K. Nienhaus, G. U. Nienhaus, *J. Biol. Phys.* **2007**, *33*, 357–370.
[36] K. Nienhaus, P. Deng, J. M. Kriegl, G. U. Nienhaus, *Biochemistry* **2003**, *42*, 9647–9658.
[37] G. U. Nienhaus, K. Nienhaus, *J. Biol. Phys.* **2002**, *28*, 163–172.
[38] K. Nienhaus, D. C. Lamb, P. Deng, G. U. Nienhaus, *Biophys. J.* **2002**, *82*, 1059–1067.
[39] H. Frauenfelder, G. U. Nienhaus, J. B. Johnson, *Ber. Bunsen Phys. Chem.* **1991**, *95*, 272–278.
[40] G. U. Nienhaus, J. Heinzl, E. Huenges, F. Parak, *Nature* **1989**, *338*, 665–666.
[41] J. M. Kriegl, K. Nienhaus, P. Deng, J. Fuchs, G. U. Nienhaus, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 7069–7074.
[42] H. Lehle, J. M. Kriegl, K. Nienhaus, P. Deng, S. Fengler, G. U. Nienhaus, *Biophys. J.* **2005**, *88*, 1978–1990.
[43] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus, *J. Comp. Chem.* **1983**, *4*, 187–217.
[44] A. Mackerel Jr., C. Brooks III, L. Nilsson, B. Roux, Y. Won, M. Karplus In *CHARMM: The Energy Function and its Parameterization with an Overview of the Program*; v. R. Schleyer et P. al., Ed.; John Wiley & Sons: Chichester, **1998**; Vol. 1, pp. 271–277.
[45] C. L. Brooks, M. Karplus, *J. Chem. Phys.* **1983**, *79*, 6312–6325.
[46] J. N. Huffaker, *J. Chem. Phys.* **1976**, *64*, 3175–3181.
[47] J. N. Huffaker, *J. Chem. Phys.* **1976**, *64*, 4564–4570.
[48] A. J. Stone, *J. Chem. Theory Comput.* **2005**, *1*, 1128–1132.
[49] N. Plattner, M. Meuwly, *ChemPhysChem* **2008**, *9*, 1271–1277.
[50] N. Plattner, T. Bandi, J. D. Doll, D. L. Freeman, M. Meuwly, *Mol. Phys.* **2008**, *106*, 1675–1684.
[51] M. Meuwly, O. M. Becker, R. Stote, M. Karplus, *Biophys. Chem.* **2002**, *98*, 183–207.
[52] K. Kuczera, J. Kuriyan, M. Karplus, *J. Mol. Biol.* **1990**, *213*, 351–373.
[53] D. A. McQuarrie in *Statistical Mechanics*; Harper and Row, New York, **1976**.
[54] M. P. Allen, D. J. Tildesley in *Computer Simulations of Liquids*; Clarendon Press, Oxford, **1989**.
[55] J. Danielsson, M. Meuwly, *ChemPhysChem* **2007**, *8*, 1077–1084.
[56] D. C. Lamb, K. Nienhaus, A. Arcovito, F. Draghi, A. E. Miele, M. Brunori, G. U. Nienhaus, *J. Biol. Chem.* **2002**, *277*, 11636–11644.
[57] T. S. Li, M. L. Quillin, G. N. Phillips, J. S. Olson, *Biochemistry* **1994**, *33*, 1433–1446.
[58] G. U. Nienhaus, J. R. Mourant, K. Chu, H. Frauenfelder, *Biochemistry* **1994**, *33*, 13413–13430.
[59] K. Chu, R. M. Ernst, H. Frauenfelder, J. R. Mourant, G. U. Nienhaus, R. Philipp, *Phys. Rev. Lett.* **1995**, *74*, 2607–2610.
[60] D. R. Nutt, M. Meuwly, *Biophys. J.* **2003**, *85*, 3612–3623.

© 2010 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

[61] D. R. Nutt, M. Meuwly, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 5998–6002.
[62] R. Aranda, E. J. Levin, F. Schotte, P. A. Anfinrud, G. N. Phillips, *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **2006**, *62*, 776–783.
[63] G. Hummer, F. Schotte, P. A. Anfinrud, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 15330–15334.
[64] K. Nienhaus, A. Ostermann, G. U. Nienhaus, F. G. Parak, M. Schmidt, *Biochemistry* **2005**, *44*, 5095–5105.

Received: August 11, 2009
Published online on December 4, 2009

© 2010 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

6.3. Structural characterization of spectroscopical substates in carbonmonoxy neuroglobin

The following paper is submitted for publication in Faraday Discussion 150: Frontiers in Spectroscopy

CREATED USING THE RSC LaTeX PCCP ARTICLE TEMPLATE – SEE www.rsc.org/electronicfiles FOR DETAILS
PAPER www. rsc.org/faraday_d | Faraday Discussions

Structural characterization of spectroscopic substates in carbonmonoxy neuroglobin

Stephan Lutz,[a] and Markus Meuwly,*[b]

Relating structure and spectroscopy is fundamental in characterizing the conformational dynamics and elucidating function at an atomistic level in condensed-phase environments. In particular, the combination of infrared spectroscopy and atomistic simulations has provided fundamental insight into structural assignments of spectroscopic bands. Infrared spectroscopy and Molecular Dynamics (MD) simulations on carbonmonoxy myoglobin (MbCO) were able to partially identify three major CO infrared bands (A_0, A_1 and A_3) which are related to different conformational substates in the active site environment of the protein. Recently, two similar CO bands were identified from 2D-IR experiments in human carbonmonoxy neuroglobin (NgbCO), named N_0 and N_3. Time-dependent frequency changes found in the N_0 band and a large variation of relaxation times for these bands made the characterization of these substates considerably more difficult compared to MbCO. In this work we discuss the structure-spectroscopy relationship for three different His64 protonation states in human and murine NgbCO using MD simulations and density functional theory (DFT) calculations. The present work assigns the N_3 band to the His$_\varepsilon$64 tautomer having its sidechain in hydrogen bonding contact to CO. Frequencies of the corresponding His64H$^+$ and His$_\delta$64 tautomers show characteristic contributions to the N_0 band. Spectroscopic line broadening, most strongly observed for human neuroglobin, is likely to be attributed to increased solvent accessibility.

1 Introduction

The bond between a ligand and the heme-iron atom in proteins such as myoglobin (Mb), hemoglobin (Hb), neuroglobin (Ngb) or truncated hemoglobin (trHb) is photolabile and can be broken by electronically exciting the Soret band.[1] Biologically, this bond is of central importance, because it can be used to better understand reactivity, binding and rebinding dynamics in a well defined biological context. In less than 100 fs after excitation, the Fe-CO bond is broken and as a result the heme unit relaxes to a structure that prevents immediate rebinding of CO.[1] The excess energy is primarily transferred into low frequency modes, including CO translation and rotation and vibrational modes of the heme. Eventually, the energy is further released to both, the protein and the solvent and leads to a temperature increase in both.[2]

---

[a] *Department of Chemistry, University of Basel, Klingelbergstrasse 80, CH-4056 Basel, Switzerland. Fax: +41 (0)61 267 38 55; Tel: +41 (0)61 267 38 15; E-mail: stephan.lutz@unibas.ch*
[b] *Department of Chemistry, University of Basel, Klingelbergstrasse 80, CH-4056 Basel, Switzerland. Fax: +41 (0)61 267 38 55; Tel: +41 (0)61 267 38 21; E-mail: m.meuwly@unibas.ch*

Infrared (IR) spectroscopy has served as a useful tool to characterize protein and ligand dynamics.[3–8]. By sensing the iron-bound CO vibrations in carbonmonoxy myoglobin (MbCO), several CO-stretching bands were observed. Contrary to electronic spectroscopy, vibrational spectra of reactants and products of CO dissociation contain signatures related to geometrical structures. These changes can be directly associated with particular chemical bonding patterns and are therefore more readily interpreted in terms of chemical structure. Thus, an important goal of IR studies of proteins is concerned with the structural interpretation of such spectra. This is, however, a difficult undertaking in most situations because structure and spectroscopy can only be recorded at the same time under special circumstances. One example is picosecond time-resolved X-ray diffraction of the L29F mutant which, in conjunction with femtosecond time-resolved mid-IR spectroscopy, provided information on structures and spectroscopy at the same time.[9] However, this method is highly specialized and not applicable to an arbitrary problem. More traditional methods such as linear and - more recently - 2D–infrared spectroscopy also provide information that can be used to infer structural changes. However, analysis of the spectroscopic signatures in terms of local structure has to be done either by exploring analogies with previously investigated systems,[10,11] or structure and spectroscopy can be related by employing atomistic simulations with accurate force fields. This has been successfully pursued over the past few years and has allowed to structurally assign vibrational spectroscopic signatures of unbound CO in Mb and Ngb.[12–16] These studies along with the corresponding experimental efforts have opened the possibility to use small diatomic ligands as sensitive structural spectroscopic probes of heterogeneous environments such as the primary (heme pocket) and several secondary (Xenon pockets) docking sites in Mb.

Likewise, the Fe-bound ligand can be used to probe the arrangement of the protein side chains in the heme pocket. Over the years, in MbCO the three main bands (A_0, A_1 and A_3) were characterized in more detail.[7,17] Multidimensional vibrational echo experiments[18] and Molecular Dynamics (MD) simulations[19] also coupled with DFT calculations[20,21] played an important role in the structural assignment of these IR bands. Today, A_0 is unambiguously considered to represent a conformational substate (CS) with the distal His64 sidechain rotated out from the proximal heme-CO active site towards the solvent. This conformation dominates at low pH having His64 in its doubly protonated His64H$^+$ state. Structural assignment of bands A_1 and A_3 is still elusive[6,17,21,22] but the majority of all recent studies suggest His64 to be singly protonated at N_ε for both bands but with different orientations and distances to CO.[19] However, even improved force fields have not yet allowed to unambiguously assign the three spectroscopically known substates (A_0, A_1, A_3) in MbCO.[21] On the other hand, MD simulations together with electronic structure calculations have been quite successful in assigning these substates to particular bonding situations although a full characterization is also still lacking.

Another relevant aspect in using simulation and spectroscopy to explore and characterize proteins at an atomistic level is the notion that proteins typically exist in a finite number of thermally accessible substates. It is quite likely that these CSs are related to particular functions of the protein. However, relating structure, dynamics, spectroscopy and function requires a multi-faceted approach and even for the protein probably best characterized so far - myoglobin - a widely accepted functional interpretation of the various CSs is still lacking. The CSs correspond to local minima on the free energy landscape and are hierarchically organized.[23–25] This network can nowadays be mapped out and analyzed by using MD simulations.[15,26,27] The thermodynamically most stable minima can be differentiated by vibrational spectroscopy as seen for the A_0, A_1, and A_3 band in MbCO.[3,28] The complete ensemble of CS is interconnected on the $(3N-6)-$dimensional free energy landscape where every CS is located in one of the minima in the first tier. Above the critical glass-temperature T_c transitions between two different CSs are possible. Thermal fluctuations in the lower tiers of the energy landscape give rise to processes such as opening of channels and gates[29,30] and small ligand diffusion through the protein.[31]

Another protein for which considerable experimental data relevant to the present discussion has become available is Ngb which was discovered only recently in neuronal cells of men and mice.[32–34] It is classified as a member of the vertebrate globin family of proteins. Although the amino acid sequence of Ngb differs by more than 75 % compared to Mb, both tertiary X-ray structures have the characteristic three-over-three globin fold.[35] Despite this structural agreement, Ngb was found to behave differently compared to Mb in the presence of exogenous ligands such as O_2, CO or NO and consequently a different physiological role was suggested for Ngb. For oxygenated Ngb it was found that a reaction with NO to a peroxynitrite intermediate occurs with a fast first order rate of 200 s^{-1} which subsequently decays to nitrate[36]. From this observation a possible physiological function as radical scavenger was suggested for Ngb[37]. Ngb and Mb also differ in their ligand-free crystal structure: While the iron atom of the heme prosthetic group in Mb remains pentacoordinated to the four porphyrin nitrogens and to the N_ε atom of the proximal His93(F8) residue, the distal His64(E7) in Ngb forms a sixth coordination to the heme-iron[38]. Several rebinding studies support the existence of this hexacoordinated configuration by suggesting biphasic rebinding rates for exogenous ligands where in case of the slower rebinding rate the His64 to iron coordination needs to be broken prior to ligand rebinding[39]. Moreover, human Ngb (hNgb) which, in contrast to mouse Ngb (mNgb), displays a disulfide bond in the CD-loop was found to lower its O_2 affinity when removing the S=S bond.[40] All these observations lead to the assumption that the functional differences between Mb and among different Ngb families are strongly related to protein dynamics rather than to structural differences.

To better understand the dynamics of the possible CSs in bound NgbCO, MD simulations were used together with electronic structure calculations. The following section presents the computational strategies employed. Subsequently the results are discussed and compared with recent 2D-IR experiments.[10,11]

2 Methods

All MD simulations were carried out with the CHARMM program[41] and the CHARMM27 force field[42]. For the present study the bis-His-coordinated structure for human neuroglobin (hNgb, PDB code 1OJ6[35]) and the CO-bound structure for mouse neuroglobin (mNgb, PDB code 1W92[43]) were used. In the absence of a suitable X-ray structure, hNgb (PDB 1OJ6) was modified by replacing Ser46 and Ser55 by cysteines, forming the disulfide bond between the two residues, breaking the His64-Fe bond, introducing the bound CO-ligand, and extensively equilibrating the structure (see below). This structure is referred to as N (native) in the following. For the mutant M the X-ray structure of mNgb was used as it is the only one with a bound CO-ligand, the tertiary structures between PDB codes 1OJ6 and 1W92 do not differ considerably (backbone RMSD = 2.1 Å), and because it shares a high degree of sequence identity (94 %) with hNgb. Proteins N and M allow us to studying systems closely related to the ones used in the spectroscopic work.[10,11]

Structure N was prepared in the following way: Missing H-atoms were added to monomer D of PDB code 1OJ6. The protein was solvated in a periodic waterbox with dimensions $41 \times 61 \times 56$ Å^3. All hydrogen atoms in the systems were minimized for 500 steps with the steepest descent method and another 1000 steps with the Adopted Basis Newton-Raphson algorithm.[44] The same procedure was repeated for all water molecules and for residues 44 to 58 which include the CD loop and residues Cys46 and Cys55 linked through a disulfide bond. After heating the system from 20 to 300 K for 40 ps and equilibrating for an additional 250 ps at 300 K the C_α atom RMSD converged to 1.3 Å compared to the initial structure. Subsequently a heme-bound CO was added to this structure and the equilibration was continued for another 250 ps after which the backbone RMSD stabilized at 1.1 Å and 1.5 Å compared to the previous equilibration and the initial structure, respectively.

After adding H-atoms to the crystal structure of M (PDB 1W92) both proteins (N and M) were treated identically. All protonation states for His64, including protonation at N_ε (His$_\varepsilon$64), N_δ (His$_\delta$64), and the doubly protonated histidine (His64H$^+$), were generated. As the simulation is focused on the region surrounding the heme group, stochastic boundaries[45] were used for the MD production phase to increase computational efficiency. The heme pocket was solvated by three sequential layers of water molecules around the center of mass of the four porphyrin nitrogens of the heme prosthetic group, and a solvent boundary potential with a radius of 25 Å was applied to constrain the water molecules. A "reaction region" of radius 16 Å centered on the heme was defined, inside which the system was propagated with Newtonian dynamics. The dynamics of the buffer region between 16 and 20 Å from the center was described using Langevin dynamics. After solvation, the N system consisted of 2446 protein atoms (2447 for His64H$^+$) and 1713 water molecules whereas M contains 2465 protein atoms (2466 for His64H$^+$) and 1450 water molecules. Each system was heated from 100 to 300 K for 50 ps and equilibrated for another 100 ps. The final RMSD of the C_α atoms with respect to the crystal structures (PDB codes 1OJ6 and 1W92) was 1.9 Å for N and 1.7 Å for M, respectively. Subsequently, forty snapshots from 2 ns of equilibrium MD simulations were analyzed further for each system (i.e., M and N and all His64 protonation states) giving a total of 240 analyzed structures.

To analyze more distant fluctuations around the CD-loop and cavity volumes, His64H$^+$ M and MbCO with periodic boundary conditions were prepared following the same protocol as described above for the preparation of N. For MbCO the initial coordinates were from the crystal structure with PDB code 1A6G. After heating and equilibration, the RMSD of the C_α atoms in MbCO converged at ≈ 1.2 Å compared to the X-ray structure. In M the RMSD stabilized at ≈ 1.5 Å.

All electronic structure calculations were carried out with the Gaussian 03 suite of programs[46] and DFT with the hybrid B3LYP functional.[47,48] The heme group, bound CO, distal His64, and proximal His93 residue were treated quantum mechanically. His64 was represented by its imidazole side chain, capped with a methyl group. A double-ζ VDZ basis set was used for the CO ligand, Fe, and N atoms, whereas the remainder of the QM system was treated with the smaller 3-21G basis set. Remaining protein and solvent atoms were represented as point charges from the CHARMM27 force field placed at the respective atomic coordinates. The same approach was successfully used in previous studies on bound and photodissociated MbCO.[21,49] Subsequently, the fundamental vibration frequency were obtained from the one-dimensional Schrödinger equation: The CO bond energy was sampled at eleven 0.025 Å intervals of r, the CO bond distance centered at its equilibrium distance r_e and the energies $V(r)$ were fitted to a Morse function

$$V(r) = D_e(1 - \exp(-\beta(\mathrm{r} - \mathrm{r_e}))) \tag{1}$$

where β and D_e were allowed to vary. Bound vibrational states supported by $V(r)$ were then calculated with LEVEL[50]. Vibrational frequencies calculated in this way allow quantitative estimation of vibrational frequency shifts.

3 Results

First of all this section separately presents the theoretical spectra obtained from the DFT calculations for M and N. Then the identified band positions are structurally analyzed.

3.1 Theoretical NgbCO spectra

mouse NgbCO: The calculated band distribution for M together with a superposition of four independent Gauss-functions is shown in Fig. 1A and B. Due to the larger number of water molecules within the active site of Ngb compared to Mb, the QM scans were performed in two different ways: once including all charges of the water molecules (“wet”) with their fixed geometries in the corresponding snapshot (Fig. 1A) and once with all waters removed (“dry”) within a radius of 5 Å around the CO molecule (Fig. 1B). In the following the calculated His64H$^+$ band-center, estimated from the frequency distribution of all 40 snapshots (at 2094.8 cm^{-1} in the “wet” and at 2095.1 cm^{-1} in the “dry” state) is taken as reference. All other frequencies are reported relative to these values and are summarized in Table 1.

The His$_\delta$64 band distributions in Fig. 1A and B each have two peaks. The two band centers can be fit with a sum of two Gaussian functions centered at -9 and $+7$ cm^{-1} (Fig. 1A) and at -13 and $+16$ cm^{-1} (Fig. 1B). The frequency range of His$_\delta$64 covers roughly the same region as His64H$^+$ although it is significantly broader. The distribution of computed IR bands from the His$_\varepsilon$64 snapshots is predominantly redshifted (Fig. 1A) with a maximum at ≈ -17 cm^{-1} for the “wet” protein and at ≈ -29 cm^{-1} for the “dry” one. Although there may be a splitting for “wet” His$_\varepsilon$64 tautomer (-19 and -30 cm^{-1}), it is not found for

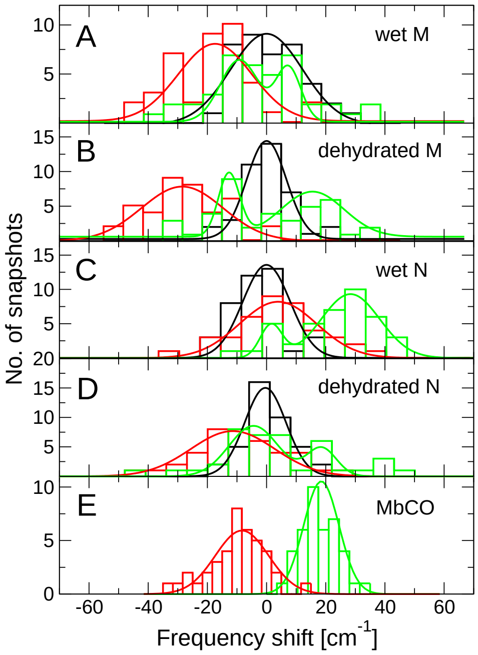

Fig. 1 Theoretical IR CO stretching spectra calculated as described in the method section. A: Relative frequency shifts to the calculated His64H$^+$ (black) band maximum in mNgbCO (M) for the His$_\varepsilon$64 (red) and His$_\delta$64 (green) tautomers. B: Same as A but having all water atoms within the active site removed from the snapshots prior to the frequency calculation. C: Relative frequency shifts to the calculated His64H$^+$ (black) band maximum in hNgbCO (N) for the His$_\varepsilon$64 (red) and His$_\delta$64 (green) tautomers. D: Same as C but having all water atoms within the active site removed from the snapshots prior to the frequency calculation. E: Relative frequency shifts to the calculated His64H$^+$ band position in wt MbCO for the His$_\varepsilon$64 (red) and His$_\delta$64 (green) MbCO tautomer taken from Ref.[21].

CO band shift relative to His64H$^+$ peak (FWHM)	**Protein**					
	His$_\varepsilon$64 M	His$_\delta$64 M	His$_\varepsilon$64 N	His$_\delta$64 N	His$_\varepsilon$64 MbCO	His$_\delta$64 MbCO
fully solvated [cm^{-1}]	-17 (9)	-9 (5), +7 (3)	+4 (13)	+2 (4), +29 (10)	-12 (9)	+16 (6)
dry active site [cm^{-1}]	-29 (10)	-13 (3), +16 (8)	-11 (14)	-4 (8), +18 (5)	-	-

Table 1 Band shifts of His$_\varepsilon$64 and His$_\delta$64 protonation states relative to His64H$^+$ in M, N and MbCO calculated from MD snapshots and their DFT CO stretching potential as described in the methods section. Values for MbCO are taken from Ref.[21]. Numbers in parentheses are full width at half maximum extracted from the Gaussian fits of each frequency distribution in Fig. 1.

the corresponding "dry" form (Fig. 1C). This suggests that the splitting arises from interaction with water and does not represent an intrinsic property of one or several protein conformations.

human NgbCO: Fig. 1C and D show calculated spectra for the "wet" and the "dry" N system. Again, the His64H$^+$ band which shows one distinct maximum peak (2081.5 cm^{-1} "wet", 2092.9 cm^{-1} "dry") is the reference and the spectra of the two tautomeric forms are reported relative to it. Absolute band peaks of the His64H$^+$ frequency distribution in "dry" N differ only by -1.9 cm^{-1} compared to the corresponding one in M although the distribution is broader with a full width at half maximum (FWHM) of 7 cm^{-1} versus 5 cm^{-1} in M. The same distribution in full solvation is redshifted by -13.3 cm^{-1} compared to M which is likely due to the increased presence of water molecules within the active-site in N (90 deleted water molecules for the His64H$^+$ spectrum in Fig. 1D vs. 1C compared to only 56 waters in Fig. 1B vs. 1A).

The His$_\delta$64 band distributions show distinct separations into two peaks. The spectrum (Fig. 1C) for the "wet" protein is split by 27 cm^{-1} with both peaks shifted to the blue relative to His64H$^+$. The His$_\delta$64 band separation for the "dry" system in Fig. 1D is 22 cm^{-1}. Its low energy peak is redshifted by -4 cm^{-1} and the high energy band is blueshifted by $+18$ cm^{-1}. The difference of two blueshifts in spectra of the "wet" and the "dry" systems is related to the difference of the absolute band positions of His64H$^+$ as mentioned in the previous paragraph.

The computed His$_\varepsilon$64 N IR spectra in Fig. 1 are single peaked having the band center located at $+4.2$ cm^{-1} ("wet") and -11 cm^{-1} ("dry"). These bandshifts are weaker compared to the corresponding ones in M (Table 1). On the other hand, the widths of the frequency distributions in His$_\varepsilon$64 N are 4 cm^{-1} broader than in the M system with FWHMs of 13 cm^{-1} ("wet") and 14 cm^{-1} ("dry"), respectively.

3.2 Structural comparison of theoretical peak positions

Double protonation of His64, which introduces a positive charge on this sidechain, is known to render this residue more hydrophilic.[28] The positive charge on His64H$^+$ is expected to expose its sidechain preferably to the solvent.[7,21] The present MD simulations show this for both Ngb variants, M and N, Ngb. Starting from a "closed" (strongly interacting with the bound CO) His64H$^+$ NgbCO conformation the protein equilibrates after heating to 300 K within a few hundred ps into an "open" (solvent exposed) one. The Gaussian fitted spectra in Fig. 1A and B show one distinct peak. Representative heme-aligned snapshots from the low-energy (white and blue), band center (red) and high-energy (green) region of the calculated His64H$^+$ M band distribution are shown in Fig. 2. In all cases, His64H$^+$ is found in its solvent-exposed, open conformation and both, His64H$^+$ and Phe28, the two residues closest to CO fluctuate considerably. For example, for the white and blue structures in Fig. 2 which both correspond to spectral features redshifted by ≈ -20 cm^{-1} the RMSD of the sidechain atoms of His64H$^+$ and Phe28 is 5.7 Å and 6.2 Å respectively. Thus, even with strong structural rearrangements the open conformation in NgbCO does only moderately affect the CO infrared spectrum.

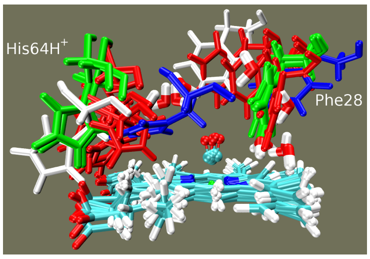

Fig. 2 Collection of low-frequency (white and blue), ≈ 0 cm^{-1} shift frequency (red) and high-frequency (green) CO stretching structures from His64H$^+$ M snapshots. The total of eleven water molecules located within a radius of 5 Å around CO taken from every single of the totally eight snapshots are shown additionally.

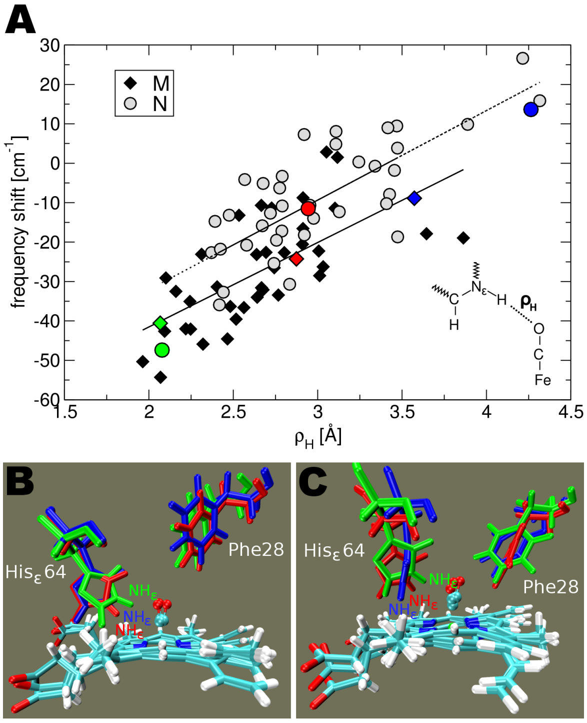

Fig. 3 A: CO-stretch frequency shift distribution of all "dry" His$_\varepsilon$64 M (diamonds) and N (circles) snapshots relatively to the His64-N$_\varepsilon$H to CO oxygen distance ρ_H. The solid (M) and dashed (N) lines represents corresponding linear regressions to these distributions. B: Representative His$_\varepsilon$64 M structures selected from different regions of Figure A (color-coded) C: The same as in B but for the His$_\varepsilon$64 N system.

Fig. 3A displays the 40 frequency shifts for "dry" His$_\varepsilon$64 M (diamonds) and N (circles) as a function of the $H_{N_\varepsilon,\text{His64}}$ to O_{CO} distance ρ_H. The straight lines are regression fits $y = a + b\rho$ to the data points. The fit yields $b_M = 21.4$ cm^{-1}/Å with a standard deviation of $\sigma_{b,M} = 3.6$ cm^{-1}/Å where subscript refers to the mutant structure M. For N the parameters are $b_{b,N} = 22.7$ cm^{-1}/Å and $\sigma_{b,N} = 2.8$ cm^{-1}/Å. Therefore, the CO stretching frequency shifts equally strong with varying His64-NH$_\varepsilon$ to CO oxygen distances in both proteins. The intercepts a differ by 6.9 cm^{-1} which suggests that the CO stretching frequencies in M are typically more blueshifted compared to the configurations in N for the same distance ρ_H. The blueshift not only depends on ρ_H but also on the strength of the N$_\varepsilon$H–O hydrogen bond. The average angle θ_{NHO} over all forty snapshots of M is 138.1° whereas for N $\theta_{NHO} = 99.2°$ which is far away from typical H-bonding geometries. Fig. 3B and C show each three typical structures for the His$_\varepsilon$64 protonation state from different regions in Fig. 3A (indicated by the corresponding color-code). These representations illustrate the orientational trends described above. For both proteins the green structures are related to CO bands with strong blueshifts (-40 to -50 cm^{-1}). Both structures also have a short ($\rho_H \approx 2.1$ Å) and predominantly linear H-bond ($\theta_{NHO} = 162°$ for M and $\theta_{NHO} = 147°$ for N), respectively. The structures shown in red in Fig. 3B and C are representative for the center of each frequency distribution in Fig. 3A and their H-bond appears weaker ($\rho_H \approx 3.0$ Å and $\theta_{NHO} = 115°$ in M and $\theta_{NHO} = 87°$ in N) whereas the blue structures (Fig. 3B and C) correspond to weak or nonexistent H-bonds with ($\rho_H = 3.6$ Å, $\theta_{NHO} = 107°$ (M) and $\rho_H = 4.3$ Å, $\theta_{NHO} = 63°$ (N), respectively). The CO stretching frequency calculated from the blue N structure is one of the rare redshifted ones of this protein configuration. This is related to the fact that the His$_\varepsilon$64-imidazole sidechain is rotated by ≈ 90° away from an optimal interaction of N$_\varepsilon$H with the CO oxygen.

Fig. 4A shows two graphs of the CO stretch frequency shifts in "dry" His$_\delta$64 M (top) and N (bottom) as a function of the $N_{\varepsilon,\text{His64}}$ to O_{CO} distance ρ_N. In the figure, the structures are also grouped into those having either the N$_\varepsilon$ (triangles) or the N$_\delta$ (squares) of His64 closer to the O_{CO}. This separation corresponds to clustering the data into two rotameric forms of the His$_\delta$64 sidechain for which representative structures are shown in red and green in Fig. 4B and C. The linear regression coefficients are $b_M = -11.9$ cm^{-1}/Å and $b_N = -24.6$ cm^{-1}/Å with standard errors $\sigma_{b,M} = \sigma_{b,N} = 2.4$ cm^{-1}/Å. Therefore the frequency shift is more sensitive to ρ_N for native protein than for mutant. The blue structures in Fig. 4B and C represent structures which are either reminiscent of His$_\varepsilon$64 protonation (for M, see Fig. 3A) or strongly deviate from the linear regression curve (for N).

In the M structure the imidazole ring of the blue His$_\delta$64 M sidechain is oriented in an intermediate conformation between the red and green rotamers and therefore does not properly fit into any of the two groups. The blue N structure in Fig. 4C is reminiscent of the open conformation and comparable to the His64H$^+$ configurations. Therefore it deviates more strongly from the linear regression curve. In both proteins (His$_\delta$64 M and N) the sidechain of Phe28 is also found to be correlated with the frequency distribution and the His$_\delta$64 orientational distribution. In the high frequency structures (rotamers as described by the red structures in Fig. 4B and C) the phenyl ring of residue 28 is closer to the bound CO while in the low frequency structures (green in Fig. 4B and

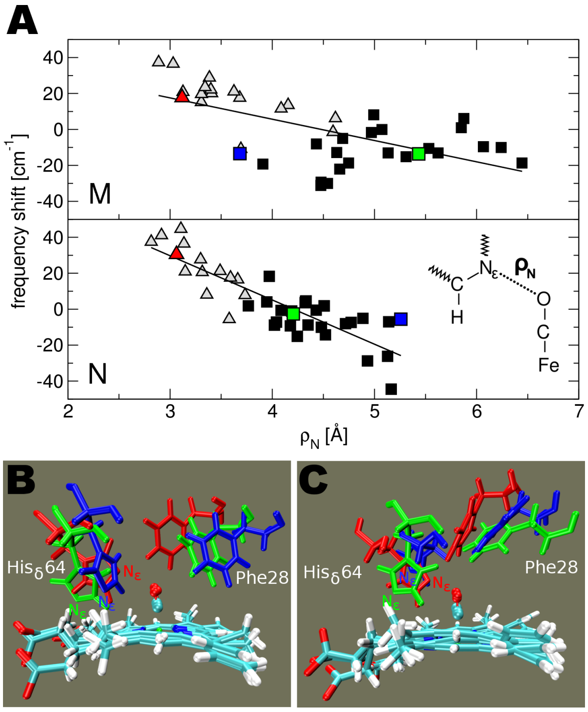

Fig. 4 A: CO-stretch frequency shift distribution of all "dry" His$_\delta$64 M (top graph) and N (bottom graph) snapshots relatively to the His64-N$_\varepsilon$ to CO oxygen distance ρ_N. Frequencies marked by triangles correspond to structures having N$_\varepsilon$ atom of His64 oriented closer to the CO oxygen than the N$_\delta$ atom and those marked by squares are showing the inverse His64 orientation. The solid line corresponds to the individual linear regression curve for each distribution. B: Representative structures from different regions of the frequency distribution of His$_\delta$64 M in Figure A (color-coded). C: Same as B but for His$_\delta$64 N.

C) it is more oriented towards the B' docking site recently discovered for this protein.[15] Regarding a possible CO dissociation event the green structures are therefore more likely to offer an initial migration of CO into the docking sites B" and F, identified in the same work, whereas residues His64 and Phe28 in the red CS rather close the passage to this site and guide the dissociated ligand towards sites C (Xe4), D (Xe1) and E (Xe2).

4 Discussion

This section discusses the results from the present study with the findings of the experimental 2D-IR echo experiments.[10,11] Furthermore, the results are also discussed in relation to MbCO. Finally, the dynamics observed in the different IR spectra for M, N, and MbCO are related to structural characteristics. The spectroscopic studies were carried out with two sequences: native hNgb which differs from the crystallized sequence by the presence of a Cys46-Cys55 disulfide bond, and a mutated sequence in which the two cysteines are replaced by serines.[10,11] In the following, these two sequences are associated with N and M, respectively.

4.1 Comparison with experimental NgbCO spectra

FTIR experiments of NgbCO found two major CO stretching bands N_0 and N_3 (see Table 2) which were discussed by comparing with the A_0 and A_3 bands in MbCO.[10,11,17] To provide structural insight, assignments based on analogies to potentially related bands in MbCO mutants were explored. The frequency of the N_3 band (1933.4 cm^{-1}) found in the spectrum of N is near the A_3 substate (1931.8 cm^{-1}) in L29F MbCO and N_0 (at 1968.1 cm^{-1}) has almost the same frequency as the A_0 state (1968.0 cm^{-1}) in H64V MbCO. At lower pH, the N_0 intensity is larger than that of N_3[51] suggesting that the N_0 band may be structurally related to His64H$^+$ NgbCO. Similarly, substate N_3 was proposed to originate from N_εH interacting via H-bonding with the bound CO in His$_\varepsilon$64 NgbCO.[10,51]

The structures of His$_\varepsilon$64 M and N in Fig. 3A and B are supportive of such assignments. Furthermore, the frequency shift between the calculated spectra of the His64H$^+$ and His$_\varepsilon$64 protonation states (Table 1) are generally in qualitative agreement with the experimental N_0 - N_3 band splitting of -35 cm^{-1} which corroborate the assignment of the two experimental bands to an open His64H$^+$ and a closed His$_\varepsilon$64 tautomeric substate. Quantitatively the computed M and N band splittings rather underestimate the experimental band separation. The "dry" form of M (Fig. 1B) best reproduces a His64H$^+$ to His$_\varepsilon$64 band splitting of -29 cm^{-1}. The same splitting in "dry" N is only -11 cm^{-1} (Fig. 1D) and becomes even slightly positive in the "wet" form of N (Fig. 1C). As previously pointed out this can be caused by the more pronounced solvation of N compared to M.

Using frequency-frequency correlation analysis[52] Fayer and co-workers detected two components, N_0' and N_0", of the N_0 peak and attributed them to two different but related conformations.[10] N_0' and N_0" were found to have two different dynamical components, one in the range of tens of ps and a second one occurring on time scales of several hundreds of ps. According to their interpretation the rapid time scale is associated with smaller conformational changes of side chains most likely localized around CO. The long time scale involves

Protein/band	ν_{CO}, cm^{-1}	FWHM, cm^{-1}
wt MbCO A_0	0	-
wt MbCO A_1	-16	-
wt MbCO A_3	-32	-
H64V MbCO	+3	8.8
L29F MbCO	-33	9.6
N N_0	+4	9.8
M N_0’	-11	6.4
M N_0”	0	14.6
N N_3	-32	12.1
M N_3	-32	11.1

Table 2 Collection of experimental CO stretching modes relative to the A_0 band wt MbCO. The full width at half maximum for the different mutants of MbCO and NgbCO are shown in the second column. The absolute frequencies taken from the related references are: wt MbCO $A_0 = 1965$ cm^{-1},[17] $A_1 = 1949$ cm^{-1},[17] $A_3 = 1933$ cm^{-1};[17] H64V MbCO 1968 cm^{-1};[10] L29F MbCO 1932 cm^{-1};[10] N $N_0 = 1969$ cm^{-1},[11] $N_3 = 1933$ cm^{-1};[11] M N_0’ $= 1954$ cm^{-1},[11] N_0” $= 1965$ cm^{-1},[11] $N_3 = 1933$ cm^{-1}.[11]

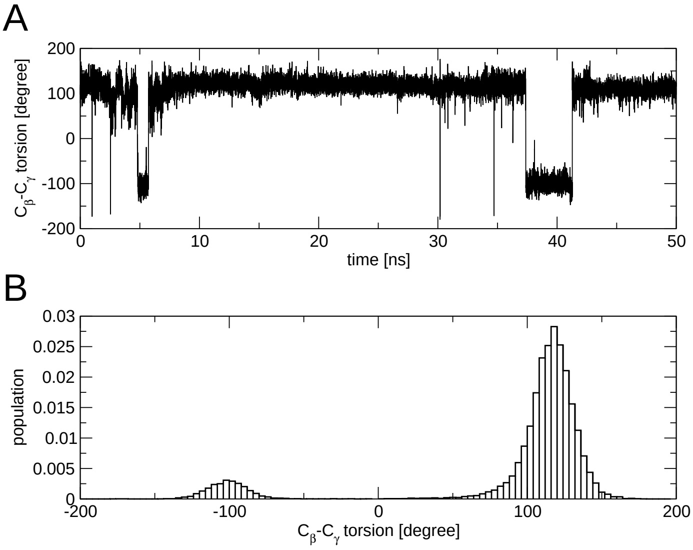

Fig. 5 A: Timeseries of the C_β-C_γ torsional angle of residue $His_\delta64$ for 50 ns of MD simulation. Structures at $\approx -100°$ reflect the rotameric conformation shown as green snapshot in Fig. 4B and angles of $\approx 100°$ represent the red structure in the same Figure. B: Normalized population distribution of the $His_\delta64$ torsional angle from A.

conformational transitions with larger barriers and is of a more global character. Averaging the two bands (negative and positive shift, respectively, see Fig. 1) corresponding to the His$_\delta$64 M and N IR spectra it is found that they overlap with the His64H$^+$ band distribution positioned at 0 cm^{-1}. This is particularly pronounced for the "dry" states (Fig. 1). It is of interest to explore whether one of the two experimentally detected substates (N$_0$', N$_0$") is related to dynamical transitions taking place in His$_\delta$64 NgbCO in addition to the one which is sampled by protein dynamics in the open His64H$^+$ NgbCO state. An additional 50 ns MD simulation of His$_\delta$64 N sampled six transitions between structures with N$_\delta$ (green structures in Fig. 4B) or N$_\varepsilon$ closer to the CO ligand (red structures in Fig. 4B), respectively. The two states correspond to a C$_\alpha$-C$_\beta$-C$_\gamma$-C$_{\delta 2}$ dihedral angle of $\approx +100°$ and $\approx -100°$, respectively (see Fig. 5A and B). The analysis in Fig. 5B suggests that the green rotamer is preferred over the red one by a factor of $\approx$ 10 which corresponds to an equilibrium constant $K = 0.1$ for the transition between the two conformations. From

$$\Delta G^{\ominus} = -RT ln(K), \qquad (2)$$

a differential stabilization energy between the two rotamers of $\Delta G^{\ominus} = 1.3$ kcal/mol is obtained, where R is the gas constant and T is the simulation temperature (300 K). The average residence time of the less stable conformation (red snapshots in Fig. 4B) is around 600 ps which is in the range of the experimentally reported slow decay time window (slower than 100 ps) for the N$_0$ band. Hence, the agreement between experimental N$_0$ band positions and the computed His$_\delta$64 M and N frequencies suggest that additional contributions to N$_0$ arise from at least one His$_\delta$64 rotameric state. As is discussed further below, the existence and population of His$_\delta$64 is also required in Ngb for functional reasons, contrary to the situation in Mb where the majority of recent work established that protonation at His$_\delta$64 is unlikely.

4.2 Comparison with MbCO

The close structural relationship between NgbCO and MbCO and the extensive experimental and theoretical data available for MbCO makes this protein an ideal candidate for comparison with NgbCO. The L29F mutation in MbCO mimics the active site environment of closed NgbCO with a unique IR absorption band at 1932 cm^{-1}, whereas the H64V MbCO mutant lacks the His64 side chain and represents an open active site conformation of MbCO with a CO IR band at 1968 cm^{-1}.[53] The individual IR peak positions of L29F and H64V MbCO have similar wavelengths compared to the corresponding A$_3$ and A$_0$ states in wt MbCO. Similarly, the N$_0$ and N$_3$ bands of N also appear in the same region blueshifted only by 1 cm^{-1} compared to L29F and H64V MbCO. Moreover, FTIR experiments on NgbCO which have the distal His64 residue substituted by nonpolar amino acids like valine or alanine show also an increase in the population of the N$_0$ substate.[51]

A comparison of the computed NgbCO band positions here with those in MbCO[21] show several differences. First, the CO frequency distribution for the His64H$^+$ configurations in wt MbCO peaks at 2103 cm^{-1} whereas for His64H$^+$ in M and N it is at $\approx$ 2094 cm^{-1}. This corresponds to a redshift of –9 cm^{-1}.

Other differences between the computed band positions of wt MbCO and M/N are mainly found for the His$_\delta$64 configurations. Fig. 1E shows the calculated bandshifts for His$_\varepsilon$64 (red) and His$_\delta$64 (green) relative to His64H$^+$ wt MbCO from Ref.[21]. The His$_\varepsilon$64 frequency distribution in wt MbCO is redshifted as it is found for both "dry" forms of M and N (Fig. 1B and D) and shows one single peak. Contrary to the spectra of His$_\delta$64 M and N the MbCO frequency range is totally blueshifted, narrow and shows only one distinct peak. More detailed analysis of the positions of the experimentally determined N$_0$' and N$_0$" bands reveal a -11 cm^{-1} redshift for N$_0$' relative to wt MbCO A$_0$. This agrees favourably with the computed bandshift of His64H$^+$ M/N versus His64H$^+$ MbCO. Thus, it is quite possible that the suggested open conformation in NgbCO is primarily related to the N$_0$' substate and not to N$_0$" although it peaks at the same wavelength as A$_0$ in wt MbCO.

In MbCO, the calculated spectra for His$_\delta$64 are blueshifted relative to His64H$^+$ (Fig. 1E) but experimentally no IR bands were observed with energies higher than the A$_0$ band. Thus, the presence of a His$_\delta$64 configuration was excluded in recent simulations.[21] This is also consistent with His$_\varepsilon$64 in Mb found by experiment and simulations.[5,54] Two possible rotamers of His$_\varepsilon$64 MbCO have previously been assigned to the (weak) A$_1$ and (dominant) A$_3$ bands using QM/MM[17] and pure MM simulations.[19] Both studies independently suggested similar conformations for A$_1$ and A$_3$. The A$_3$ conformation has the imidazole sidechain of His64 oriented with the N$_\varepsilon$-H pointing towards the iron-bound CO and the unprotonated N$_\delta$ to the solvent exposed side of the protein. This His64 orientation corresponds to the majority of sampled His$_\varepsilon$64 M and N conformations in the present work (green structures in Fig. 3A and B) which we assigned to the N$_3$ band. Corresponding to the weak A$_1$ band in MbCO an even weaker N$_{1,2}$ band has been reported from FTIR studies.[51] However, no such state was found in the present simulations.

Contrary to MbCO, the His$_\delta$64 configuration is expected to be present to some extent in NgbCO because it is required for formation of the bis-His hexacoordinated state in the absence of an exogenous ligand.[38] The position of the hydrogen atom bound to N$_\delta$ (N$_\delta$H) in His$_\delta$64 NgbCO is not suitable to establish a stable H-bond to a heme-bound CO similar to the N$_\varepsilon$H in His$_\varepsilon$64 NgbCO. Therefore the transition barrier between the rotamers is expected to be smaller in His$_\delta$64 NgbCO compared to the His$_\varepsilon$64 NgbCO. Surprisingly, none of the previous experimental NgbCO FTIR studies suggested contributions from either one or both of the His$_\delta$64 NgbCO rotamers. Nevertheless, the present simulations of the His$_\delta$64 tautomers in M and N found both rotameric forms of His$_\delta$64 and assigned the conformation with the N$_\varepsilon$ facing towards the solvent side of the protein to the low-energy peak in Fig. 1A to D and one with N$_\varepsilon$ towards the protein interior to the high-energy peak.

A more detailed comparison of the two rotamers shows that the low energy peaks of the His$_\delta$64 M and N spectra (Table 1 and green squares in Fig. 4A) better agree with the N$_0$' band position and the N$_0$" FTIR band could correspond to the blueshifted His$_\delta$64 M and N peak (red triangles in Fig. 4A). The observed bandsplitting between N$_0$' and N$_0$" is only 11 cm^{-1} whereas the computed splitting is ≈ 2 times larger. It should be noted that one of the two substates (N$_0$') is not shifted relative to the spectra corresponding to His64H$^+$ and could therefore be masked by the dominant His64H$^+$ protonation state. Similarly, in MbCO it was suggested that small contributions of the His$_\varepsilon$64 rotamer (responsible for the A$_1$ band) contribute to the A$_0$ band.[17]

4.3 Structural rearrangements

The experimentally observed FWHM of the IR bands increase in going from MbCO to M and N (Table 2). The broadening implies an increased structural heterogeneity in the CSs of NgbCO. The computed band distributions show increasing FWHMs in line with the measured widths Table (1). An explicit example of increased heterogeneity is the rotational disorder of the His$_\varepsilon$64 sidechain in Fig. 3B and C. In M the imidazole ring of His64 is predominantly aligned in plane to the bound CO to establish an H-bond between N$_\varepsilon$H and the oxygen atom of CO. In N the blue His64 sidechain (high frequency contribution in Fig. 1D) shows increased flexibility by rotating almost 90° away from the ideal H-bonding distance.

Shiro and co-workers[51] found that exogenous ligand binding for the open conformation in Ngb and cytoglobin is more rapid compared to Mb which requires a more accessible active site. Fig. 6 compares the averaged distal site cavity network of equilibrated MbCO, M and N along a 500 ps MD simulation with periodic boundaries at 300 K. In MbCO (with His64H$^+$) a large primary docking site on top of the heme plane is found with a conformationally averaged (over 500 ps) volume of $V = 164$ Å^3. For the Ngbs the volumina are $V = 85$ Å^3 (in the mutant) and $V = 65$ Å^3 (in the native protein). Partly, these differences arise from the additional space occupied by residue Phe28 in both Ngbs compared to the smaller Leu29 residue in MbCO at the same position.

As was pointed out above, individual water molecules were able to enter the active site of M and N. This can be rationalized by considering the protein cavities in Fig. 6 which are located near the CD-loop. In MbCO this region encloses an average volume of 16 Å^3. In Mb, the sidechain of Arg45 is able to form a salt bridge to one of the heme propionate groups and thus can block water access from the solvent-exposed side of the heme group. In M and N, Arg45 is replaced by Asn45 and the neighboring residue (Tyr44) is able to form a hydrogen bond with the heme propionates. However, this H-bond is weak and easily broken at ambient temperatures. In M, Tyr44 occupies a position extending into the distal site which allows water molecules to build a hydrogen bonded network from the bulk into the active site (indicated by dotted lines in Fig. 6). The average volume near the CD-loop in M is 18 Å^3 which offers somewhat more space for one water molecule whereas in N the cavity is the largest with an average volume of 24 Å^3. The hydrogen bond network in N is more extended and reaches even the CD-loop cavity. The reduced degrees of freedom in N compared to M (Cys46-Cys55 bond) positions Tyr44 preferentially towards the bulk. This CS allows formation of a water channel from the CD-loop to the primary docking site. The average number of water molecules in the distal heme cavity along the 500 ps MD simulation is 1.5, 1.7, and 2.2 molecules for MbCO, M, and N, respectively. In MbCO and M one water molecule enters and one leaves the active site during the simulation, in N two enter and one leaves. The presence of the water channel in N might also be an explanation for the experimentally observed lowering of O$_2$ affinity when

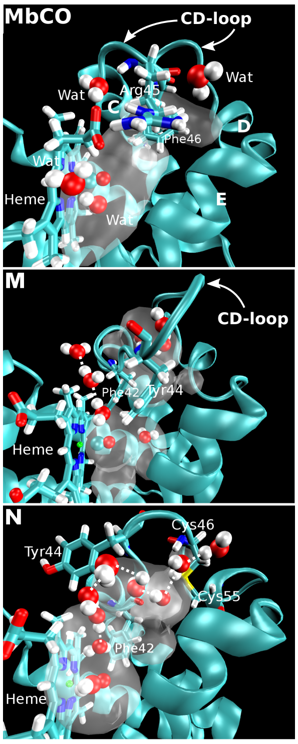
MbCO
CD-loop
Wat
Wat
Arg45
C
Phe46
D
Wat
Heme
E
Wat
M
CD-loop
Phe42
Tyr44
Heme
N
Cys46
Tyr44
Cys55
Phe42
Heme

Fig. 6 Combined distal site cavity network extracted from ten individual snapshots along an equilibrated 500 ps MD simulations each for His64H$^+$ MbCO, M and N at 300 K. The transparent white volume maps correspond to 10% population density of the cavity. Pockets were identified by the program SURFNET.[55] Water molecules, extracted from one characteristic snapshot in each system, which are occupying the pockets or are located at the protein interface next to a pocket are shown in CPK representation. Hydrogen bonds between the waters indicate the increasing formation of hydrogen bond network from the bulk into the heme active site going from MbCO over M to N.

the disulfide bond is removed.[56] Without the channel not only water molecules can less easily access the active site but also an exogenous ligand.

Decay times of the N_0 and N_3 bands extracted from 2D-IR spectra of NgbCO were found to have at least one additional, much slower time component compared to the related A_0 and A_3 IR-bands found in L29F and H64V MbCO.[10,11] The A_3 relaxation of L29F MbCO, for example, has a slowest relaxation component of $\tau_2 = 66$ ps which is in the time range for switching between A_1 - A_3.[57] The slowest decays of the N_3 and N_0 bands in NgbCO were found to be slower than the experimental time window of ≈ 100 ps. They were related to global rather than local changes around the active site compared to MbCO.[11] Because the slow rearrangements were found to speed up in M compared to N, their origin was suspected to be located close to residues 46 and 55 and the CD-loop. A comparison of the backbone fluctuations between two independent 1 ns MD simulation of M and N in a waterbox at 300 K confirm the dynamical origin suggested from experiment. The average RMSD of the backbone atoms in the CD-loop (residues 44 to 58) in N is 0.86 Å compared to 1.36 Å in M. This is indicative of increased local motion and may lead to the different dynamics between N and M. Conversely, the protein backbone fluctuations in N and M are considerably more similar with an average RMSD difference of only 0.16 Å. Again, the increased number of degrees-of-freedom in M compared to N, due to the lack of the disulfide bond, are probably responsible for the altered dynamics between M and N.

5 Conclusion

More complex structural dynamics and a larger active site cavity makes the structural interpretation of spectroscopic states in NgbCO considerably more challenging compared to Mb. Here, we used MD simulations in conjunction with QM/MM calculations to shed more light into the structural and dynamical interpretation of recent FTIR and 2D-IR vibrational echo experiments on NgbCO. Previously, Ngb spectra were analyzed indirectly by comparing with and capitalizing on analogies with wt and mutant Myoglobins. Assignments from analogies may be valid but a more direct approach is preferable, as it can not only provide a more fundamental understanding, but also scrutinize the validity of “assignment from analogy”.

The calculations support the previous assignment of the N_3 peak to a closed $\mathrm{His}_\varepsilon 64$ conformation with a strong $\mathrm{His}_\varepsilon 64$ N_ε-H to CO interaction. The N_0 band (with two substates N_0’ and N_0”) was assigned to an open active site structure with the His64 residue oriented towards the solvent and no interaction with the bound CO-ligand. The present MD simulations of His64H$^+$ NgbCO represent this CS closest. More detailed comparison of the computed spectra with MbCO suggests the conformational sampling of His64H$^+$ M and N to be more closely related to N_0’. In contrast to MbCO, contributions of the $\mathrm{His}_\delta 64$ tautomer in the IR spectra of NgbCO are expected to be present to allow formation of the bis-His coordination. The calculations for $\mathrm{His}_\delta 64$ M and N revealed a band splitting into two specific rotamers for $\mathrm{His}_\delta 64$ which were not found experimentally so far. The low-energy peak in the present theoretical spectrum is close to the IR-band maximum computed for His64H$^+$ NgbCO and we therefore suggest this conformation to also contribute to the N_0' band. However, the band is masked, probably due to spectral overlap. Such spectral masking of the bands corresponding to the two different His64 protonation states may also be responsible for the different relaxation times of the N_0 state. The high-energy band representing the other His64 rotamer of His_δ64 NgbCO was found in any case to be blueshifted relative to His64H$^+$ and therefore matching the corresponding band position of N_0".

Finally, the simulations found a marked slowdown of the conformational dynamics between M and N which may be related to the disulfide bond in N. This is consistent with experiments[10,11] where the rate for protein fluctuations was found to slow down in going from MbCO over M to N. Broadening of related IR-bands going from MbCO over M to N is observed in the experimental as well the computed spectra. A larger distal site cavity for M and especially N makes the active site more accessible for bulk water molecules. Such water molecules located closely to the bound CO or to protein atoms interacting with CO (i.e. His_ε64) can broaden the IR spectrum.

In summary, atomistic simulations have provided novel insight into the relationship between conformational dynamics and spectroscopy in NgbCO. With the approach chosen here it is possible to verify previous assignments which had been done based on analogies. Together with the insight gained for unligated CO in other globins, MD simulations with accurate force fields and ab initio methods offer an attractive computational route to relate structure and spectroscopy at an atomistic level.

References

1 J. W. Petrich, C. Poyart and J. L. Martin, *Biochemistry*, 1988, **27**, 4049 – 4060.

2 T. Q. Lian, B. Locke, Y. Kholodenko and R. M. Hochstrasser, *J. Phys. Chem.*, 1994, **98**, 11648 – 11656.

3 A. Ansari, J. Berendzen, D. Braunstein, B. R. Cowen, H. Frauenfelder, M. K. Hong, I. E. T. Iben, J. B. Johnson, P. Ormos, T. B. Sauke, R. Scholl, A. Schulte, P. J. Steinbach, J. Vittitow and R. D. Young, *Biophys. Chem.*, 1987, **26**, 337 – 355.

4 P. A. Anfinrud, C. Han and R. M. Hochstrasser, *Proc. Nat. Acad. Sci. USA*, 1989, **86**, 8387 – 8391.

5 J. B. Johnson, D. C. Lamb, H. Frauenfelder, J. D. Müller, B. McMahon, G. U. Nienhaus and R. D. Young, *Biophys. J.*, 1996, **71**, 1563 – 1573.

6 G. N. Phillips, M. L. Teodoro, T. Li, B. Smith and J. S. Olson, *J. Phys. Chem. B*, 1999, **103**, 8817 – 8829.

7 D. Morikis, P. M. Champion, B. A. Springer and S. G. Sligar, *Biochemistry*, 1989, **28**, 4791 – 4800.

8 T. S. Li, M. L. Quillin, G. N. Phillips and J. S. Olson, *Biochemistry*, 1994, **33**, 1433 – 1446.

9 F. Schotte, M. H. Lim, T. A. Jackson, A. V. Smirnov, J. Soman, J. S. Olson, G. N. Phillips, M. Wulff and P. A. Anfinrud, *Science*, 2003, **300**, 1944 – 1947.

10 H. Ishikawa, I. J. Finkelstein, S. Kim, K. Kwak, J. K. Chung, K. Wakasugi, A. M. Massari and M. D. Fayer, *Proc. Nat. Acad. Sci. USA*, 2007, **104**, 16116 – 16121.

11 H. Ishikawa, S. Kim, K. Kwak, K. Wakasugi and M. D. Fayer, *Proc. Nat. Acad. Sci. USA*, 2007, **104**, 19309 – 19314.

12 D. R. Nutt and M. Meuwly, *Biophys. J.*, 2003, **85**, 3612 – 3623.

13 D. R. Nutt and M. Meuwly, *Proc. Nat. Acad. Sci. USA*, 2004, **101**, 5998 – 6002.

14 N. Plattner and M. Meuwly, *Biophys. J.*, 2008, **94**, 2505 – 2515.

15 S. Lutz, K. Nienhaus, G. U. Nienhaus and M. Meuwly, *J. Phys. Chem. B*, 2009, **113**, 15334 –15343.

16 K. Nienhaus, S. Lutz, M. Meuwly and G. U. Nienhaus, *ChemPhysChem*, 2010, **11**, 119 – 129.

17 C. Rovira, B. Schulze, M. Eichinger, J. D. Evanseck and M. Parrinello, *Biophys. J.*, 2001, **81**, 435 – 445.

18 N. T. Hunt, *Chem. Soc. Rev.*, 2009, **38**, 1837 – 1848.

19 K. A. Merchant, W. G. Noid, D. E. Thompson, R. Akiyama, R. F. Loring and M. D. Fayer, *J. Phys. Chem. B*, 2003, **107**, 4 – 7.

20 S. Franzen, *J. Am. Chem. Soc.*, 2002, **124**, 13271 – 13281.

21 M. Devereux and M. Meuwly, *Biophys. J.*, 2009, **96**, 4363–4375.

22 D. Ivanov, J. T. Sage, M. Keim, J. R. Powell, S. A. Asher and P. M. Champion, *J. Am. Chem. Soc.*, 1994, **116**, 4139 – 4140.

23 A. Ansari, J. Berendzen, S. F. Bowne, H. Frauenfelder, I. E. T. Iben, T. B. Sauke, E. Shyamsunder and R. D. Young, *Proc. Nat. Acad. Sci. USA*, 1985, **82**, 5000 – 5004.

24 H. Frauenfelder, F. Parak and R. D. Young, *Annu. Rev. Biophys. Bio.*, 1988, **17**, 451 – 479.

25 H. Frauenfelder, S. G. Sligar and P. G. Wolynes, 1991, **254**, 1598 – 1603.

26 S. Mishra and M. Meuwly, *Biophys. J.*, 2009, **96**, 2105–2118.

27 S. Mishra and M. Meuwly, *Biophys. J.*, 2010, **in print**,.

28 F. Yang and G. N. George N. Phillips, *J. Mol. Biol.*, 1996, **256**, 762 – 774.

29 P. W. Fenimore, H. Frauenfelder, B. H. McMahon and F. G. Parak, *Proc. Nat. Acad. Sci. USA*, 2002, **99**, 16047 – 16051.

30 H. Frauenfelder, P. W. Fenimore, G. Chen and B. H. McMahon, *Proc. Nat. Acad. Sci. USA*, 2006, **103**, 15469 – 15472.

31 D. A. Case and M. Karplus, *J. Mol. Biol.*, 1979, **132**, 343 – 368.

32 T. Burmester, B. Weich, S. Reinhardt and T. Hankeln, *Nature*, 2000, **407**, 520 – 523.

33 A. Pesce, M. Bolognesi, A. Bocedi, P. Ascenzi, S. Dewilde, L. Moens, T. Hankeln and T. Burmester, *EMBO Reports*, 2002, **3**, 1146 – 1151.

34 M. Brunori and B. Vallone, *Cell. Mol. Life Sci.*, 2007, **64**, 1259 – 1268.

35 A. Pesce, S. Dewilde, M. Nardini, L. Moens, P. Ascenzi, T. Hankeln, T. Burmester and M. Bolognesi, *Structure*, 2003, **11**, 1087 – 1095.

36 D. A. Greenberg, K. Jin and A. A. Khan, *Curr. Opin. Pharmacol.*, 2008, **8**, 20 – 24.

37 S. Herold, A. Fago, R. E. Weber, S. Dewilde and L. Moens, *J. Biol. Chem.*, 2004, **279**, 22841 – 22847.

38 A. Pesce, S. Dewilde, M. Nardini, L. Moens, P. Ascenzi, T. Hankeln, T. Burmester and M. Bolognesi, *Micron*, 2004, **35**, 63 – 65.

39 K. Nienhaus, J. M. Kriegl and G. U. Nienhaus, *J. Biol. Chem.*, 2004, **279**, 22944 – 22952.

40 D. Hamdane, L. Kiger, S. Dewilde, B. N. Green, A. Pesce, J. Uzan, T. Burmester, T. Hankeln, M. Bolognesi, L. Moens and M. C. Marden, *Micron*, 2004, **35**, 59 – 62.

41 B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J. Chem. Comp.*, 1983, **4**, 187 – 217.

42 A. Mackerel Jr., C. Brooks III, L. Nilsson, B. Roux, Y. Won and M. Karplus, in *CHARMM: The energy function and its parameterization with an overview of the program*, ed. P. v. R. Schleyer et al., John Wiley & Sons: Chichester, 1998, vol. 1, pp. 271–277.

43 B. Vallone, K. Nienhaus, A. Matthes, M. Brunori and G. U. Nienhaus, *Proc. Nat. Acad. Sci. USA*, 2004, **101**, 17351 – 17356.

44 J. W. Chu, B. L. Trout and B. R. Brooks, *J. Chem. Phys.*, 2003, **119**, 12708 – 12717.

45 C. L. Brooks and M. Karplus, *J. Chem. Phys.*, 1983, **79**, 6312 – 6325.

46 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. J. Montgomery, T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A., M. Nanayakkara, P. M. W. Challacombe, B. Gill, Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, *Gaussian 03, Revision C.01*, Gaussian, Inc., Wallingford CT, U.S.A., 2004.

47 C. T. Lee, W. T. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785 – 789.

48 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648 – 5652.

49 M. Meuwly, *ChemPhysChem*, 2006, **7**, 2061 – 2063.

50 R. J. Le Roy, *Chem. Phys. Res. Reports*, 1996, **CP-555R**,.

51 H. Sawai, M. Makino, Y. Mizutani, T. Ohta, H. Sugimoto, T. Uno, N. Kawada, K. Yoshizato, T. Kitagawa and Y. Shiro, *Biochemistry*, 2005, **44**, 13257 – 13265.

52 S. Mukamel, *Annu. Rev. Phys. Chem.*, 2000, **51**, 691 – 729.

53 I. J. Finkelstein, A. Goj, B. L. McClain, A. M. Massari, K. A. Merchant, R. F. Loring and M. D. Fayer, *J. Phys. Chem. B*, 2005, **109**, 16959 – 16966.

54 M. Meuwly, *ChemPhysChem*, 2006, **7**, 2061 – 2063.

55 R. A. Laskowski, *J. Mol. Graphics*, 1995, **13**, 323 – 330.

56 D. Hamdane, L. Kiger, S. Dewilde, B. N. Green, A. Pesce, J. Uzan, T. Burmester, T. Hankeln, M. Bolognesi, L. Moens and M. C. Marden, *J. Biol. Chem.*, 2003, **278**, 51713 – 51721.

57 H. Ishikawa, K. Kwak, S. K. Chung and M. D. Fayer, *Proc. Nat. Acad. Sci. USA*, 2008, **105**, 8619 – 8624.

7. Conclusion

"That apprehension has also disturbed my mind, and to put an end to this disagreeable suspense, I have resolved to acquaint him with our attachment; surely, there is nothing in my family or fortune to disentitle me to the honor of aspiring to your hand."

Joseph Stirling Coyne, *The Queer Subject, 1837*

The research of globin proteins has a long standing history. During its evolution, ranging over many decades, a more and more detailed picture of the physiological function of these proteins and the mechanisms which are driving them has emerged. At the same time, a continuative discovery of new globin species raised many new questions and extended the view on their functional role and mechanism. As an example, new fields of research dealing with energy landscapes and catalytic functions have been established. Many questions asked in current globin research are related to the differences found in structure, dynamics and function of different sub families of globins. The connection of these three features is complex but seems to be essential for the diversity of suggested functional roles in globins.

The discovery of Mb and Hb and their subsequent investigation had a strong impact on the development of experimental procedures used in protein science. Especially X-ray crystallography, IR and Raman spectroscopy, elastic neutron scattering and kinetic measurements experienced large improvements through the research on proteins. Although these methods were able to unravel many general features of proteins and could draw a more or less complete picture on the processes of ligand binding and unbinding, they show certain limitations and drawbacks which are difficult to bypass. One elementary problem arising in some situations is the impossibility to preserve the *in vivo* physiological conditions during the experimental measurement. This is especially true for the determination of the X-ray structure which needs to have the protein sample in a dry state in order to be able to crystallize. Under these conditions the protein is in nearly every case functionally inactive. Furthermore, *in vitro* experiments can sometimes result in contradictory or strongly different results if not processed in the exactly same manner (possibility for autooxidation *etc.*). IR and Raman spectroscopy are more versatile in their application but have the drawback that they can not give a direct structural picture of the functional processes. Finally, the determination of rate constants generally rely on the description of a kinetic law to which the data can be fitted to. A complete picture of all structural and dynamical attributes of a certain protein can only be achieved through the combination of multiple experiments. MD simulations have proven as a good approach to overcome most of these experimental drawbacks. They can determine nearly every property of a protein which might be of physiological interest. The only limitation which remains is the accessible timescale. A possible way to circumvent this problem is a balanced adjustment between accuracy and speed of the calculation depending on the problem of interest. The computational methods used in this work were individually chosen to obtain sufficiently accurate answers to specific questions in the dynamics of Ngb within meaningful simulation time. With the MTP polarizable force field the individual locations and orientations which photolyzed CO captures within Ngb could be elucidated with reliable accuracy. Similarly, a network of CO visited docking sites was resolved together with free energies for migration between these cavities. CS dependent IR spectra of CO bound Ngb which are problematic to determine with MTP, could be resolved using a combination of standard classical MD and QM/MM potential scans. Finally, configurational transitions occurring during the bi-phasic rebinding of Ngb were energetically and kinetically analyzed with ARMD.

In this work some noticeable differences between Ngb and Mb, to which most of the present data was compared to, could be encountered. First of all, it was shown that the presence of Phe28 in wt Ngb lowers the transition barrier for CO migration into docking sites Xe2, Xe3 and Xe4 compared to the smaller leucine residue located at the same position in Mb. Phe28 therefore takes up the position of a “gatekeeper” who modulates the relative ligand population between the distal and proximal docking sites. The distal site of Ngb was found to include additional cavities in the CD-loop region which were not reported for Mb up to the present. Although this site looks like a promising stopover for ligands in the process between entering the protein from the bulk and transferring to the primary docking site, it could not be observed from unconstrained MD simulations. Nevertheless, a high population of water molecules in the CD-loop region simplifies the assembly of a hydrogen bond network between solvent molecules which is ranging from the bulk exposed site of the heme group into its active site. Another factor which eases the buildup of the water network is the presence of Tyr44 which, in Ngb, forms a relatively easily breakable coordination to a heme propionate group. This coordination usually takes up the role of shielding the heme active site from solvent and other molecules able to oxidate the heme iron. In Mb this coordination is described by a salt bridge to Arg45 which can less easily be broken. The fact that Ngb can form a hexacoordination with His64 (His64_{H}) imposes the need to have the sidechain of this residue to be, to a larger extent, protonated on the Nδ atom. In contrast to MbCO, where all discussed CS were assigned to the Nϵ tautomer of His64, the present calculations suggest that CSs related to the Nδ protonated tautomeric form of His64 are likely to exist in NgbCO.

Clearly not all structurally important features for the function of Ngb could be answered in this work as this subject is highly complex and extends many aspects of protein dynamics. Nonetheless, it could give new insights for possible mechanisms which control ligand binding and unbinding in globin proteins. Experimental and theoretical studies picking up and continuing on these ideas might strengthen the importance of the observed dynamical features in Ngb and put them into a more general relation with the functional diversity of the large family of globin proteins.

Bibliography

1. L. Alberghina and H. Westerhoff. *Systems Biology: Definitions and Perspectives*. Springer, 2005.

2. K. Kaneko. *Life: An Introduction to Complex Systems Biology*. Springer, 2006.

3. U. Alon. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman & Hall, 2006.

4. B. Palsson. *Systems Biology: Properties of Reconstructed Networks*. Cambridge University Press, 2006.

5. U. Sauer, M. Heinemann, and N. Zamboni. Genetics - Getting closer to the whole picture. *Science*, 316(5824):550 – 551, 2007.

6. R. H. Waterston, E. S. Lander, and J. E. Sulston. On the sequencing of the human genome. *Proc. Nat. Acad. Sci. USA*, 99(6):3712 – 3716, 2002.

7. R. H. Waterston, E. S. Lander, and J. E. Sulston. More on the sequencing of the human genome. *Proc. Nat. Acad. Sci. USA*, 100(6):3022 – 3024, 2003.

8. S. Herold, A. Fago, R. E. Weber, S. Dewilde, and L. Moens. Reactivity studies of the Fe(III) and Fe(II)NO forms of human neuroglobin reveal a potential role against oxidative stress. *J. Biol. Chem.*, 279(22):22841 – 22847, 2004.

9. A. Giuffre, T. Moschetti, B. Vallone, and M. Brunori. Is there an answer? Is neuroglobin a signal transducer? *IUBMB LIFE*, 60(6):410 – 413, 2008.

10. J. Kelsen, C. A. Hundahl, and A. Hay-Schmidt. Neuroglobin: Endogenous neuroprotectant or maintenance of homeostasis? *Stroke*, 39(11):E177 – E178, 2008.

11. T. Burmester and T. Hankeln. What is the function of neuroglobin? *J. Exp. Biol.*, 212(10):1423 – 1428, 2009.

12. K. Henty, Y. Yosaatmadja, S. Bonding, J. Skommer, N. Birch, and T. Brittain. The role of neuroglobin in the prevention of cell death. *FEBS J.*, 276:199 – 200, 2009.

13. K. Shikama and A. Matsuoka. Structure-function relationships in unusual nonvertebrate globins. *Crit. Rev. Biochem. Mol.*, 39(4):217 – 259, 2004.

14. O. V. Kosmachevskaya and A. F. Topunov. Hemoglobins: Diversity of structures and functions. *Appl. Biochem. Microbiol.*, 45(6):563 – 587, 2009.

15. R. Apweiler *et al.* The universal protein resource (uniprot) in 2010. *Nucleic Acids Res.*, 38:D142 – D148, 2010.

16. M. C. P. Van Beekvelt, W. N. J. M. Colier, R. A. Wevers, and B. G. M. Van Engelen. Performance of near-infrared spectroscopy in measuring local O-2 consumption and blood flow in skeletal muscle. *J. Appl. Physiol.*, 90(2):511 – 519, 2001.

17. F. Hoppe-Seyler. Über die Oxydation in lebendem Blute. *Med-chem Untersuch Lab.*, 1(2865):133 – 140, 1866.

18. J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff, and D. C. Phillips. 3-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature*, 181(4610):662 – 666, 1958.

19. L. Pauling, H. A. Itano, S. J. Singer, and I. C. Wells. Sickle cell anemia, a molecular disease. *Science*, 110(2865):543 – 548, 1949.

20. K. P. Moore, S. G. Holt, R. P. Patel, D. A. Svistunenko, W. Zackert, D. Goodier, B. J. Reeder, M. Clozel, R. Anand, C. E. Cooper, J. D. Morrow, M. T. Wilson, V. Darley-Usmar, and L. J. Roberts. A causative role for redox cycling of myoglobin and its inhibition by alkalinization in the pathogenesis and treatment of rhabdomyolysis-induced renal failure. *J. Biol. Chem.*, 273(48):31731 – 31737, 1998.

21. J. van Gijn, R. S. Kerr, and G. J. E. Rinkel. Subarachnoid haemorrhage. *Lancet*, 369(9558):306 – 318, 2007.

22. S. E. V. Phillips and B. P. Schoenborn. Neutron-diffraction reveals oxygen-histidine hydrogen-bond in oxymyoglobin. *Nature*, 292(5818):81 – 82, 1981.

23. B. J. Reeder. The redox activity of hemoglobins: From physiologic functions to pathologic mechanisms. *Antioxid. Redox Sign.*, 13(7):1087 – 1123, 2010.

24. M. F. Perutz. Stereochemistry of cooperative effects in haemoglobin. *Nature*, 228(5273):726 – 734, 1970.

25. J. E. Hall. *Guyton and Hall Textbook of Medical Physiology*. Saunders, Philadelphia, 12 edition, 2010.

26. W. Doster, D. Beece, S. F. Bowne, E. E. Dilorio, L. Eisenstein, H. Frauenfelder, L. Reinisch, E. Shyamsunder, K. H. Winterhalter, and K. T. Yue. Control and pH-dependence of ligand-binding to heme-proteins. *Biochemistry*, 21(20):4831 – 4839, 1982.

27. J. D. Müller, B. H. McMahon, E. Y. T. Chien, S. G. Sligar, and G. U. Nienhaus. Connection between the taxonomic substates and protonation of histidines 64 and 97 in carbonmonoxy myoglobin. *Biophys. J.*, 77(2):1036 – 1051, 1999.

28. J. D. Müller, B. McMahon, and G. U. Nienhaus. The influence of pH and salts on the A-states of myoglobin. *Biophys. J.*, 72(2):TH448 – TH448, 1997.

29. R. L. Garlick. Structure of annelid high molecular-weight hemoglobins (erythrocruorins). *Am. Zool.*, 20(1):69 – 77, 1980.

30. W. E. Royer, K. Strand, M. van Heel, and W. A. Hendrickson. Structural hierarchy in erythrocruorin, the giant respiratory assemblage of annelids. *Proc. Nat. Acad. Sci. USA*, 97(13):7107 – 7111, 2000.

31. M. T. Wilson and B. J. Reeder. Oxygen-binding haem proteins. *Exp. Physiol.*, 93(1):128 – 132, 2008.

32. J. B. Wittenberg. Myoglobin-facilitated oxygen diffusion: role of myoglobin in oxygen entry into muscle. *Physiol. Rev.*, 50(4):559 – 636, 1970.

33. K. D. Jürgens, T. Peters, and G. Gros. Diffusivity of myoglobin in intact skeletal-muscle cells. *Proc. Nat. Acad. Sci. USA*, 91(9):3829 – 3833, 1994.

34. J. B. Wittenberg and B. A. Wittenberg. Mechanisms of cytoplasmic hemoglobin and myoglobin function. *Annu. Rev. Biophys. Bio.*, 19:217 – 241, 1990.

35. P. George and D. H. Irvine. The reaction of metmyoglobin with strong oxidizing agents. *Biochem. J.*, 58(2):188 – 195, 1954.

36. G. Poli, G. Leonarduzzi, F. Biasi, and E. Chiarpotto. Oxidative stress and cell signalling. *Curr. Med. Chem.*, 11(9):1163 – 1182, 2004.

37. M. Glei, G. O. Latunde-Dada, A. Klinder, T. W. Becker, U. Hermann, K. Voigt, and B. L. Pool-Zobel. Iron-overload induces oxidative DNA damage in the human colon carcinoma cell line HT29 clone 19A. *Mutat. Res-Gen. Tox. En.*, 519(1-2):151 – 161, 2002.

38. B. J. Reeder, D. A. Svistunenko, M. A. Sharpe, and M. T. Wilson. Characteristics and mechanism of formation of peroxide-induced heme to protein cross-linking in myoglobin. *Biochemistry*, 41(1):367 – 375, 2002.

39. B. J. Reeder and M. T. Wilson. Mechanism of reaction of myoglobin with the lipid hydroperoxide hydroperoxyoctadecadienoic acid. *Biochem. J.*, 330:1317 – 1323, 1998.

40. M. T. Gladwin, J. R. Lancaster, B. A. Freeman, and A. N. Schechter. Nitric oxide's reactions with hemoglobin: A view through the SNO-storm. *Nat. Med.*, 9(5):496 – 500, 2003.

41. A. J. Gow. Nitric oxide, hemoglobin, and hypoxic vasodilation. *Am. J. Resp. Cell Mol.*, 32(6):479 – 482, 2005.

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

43. W. C. Winkler, G. Gonzalez, J. B. Wittenberg, R. Hille, N. Dakappagari, A. Jacob, L. A. Gonzalez, and M. A. Gillesgonzalez. Nonsteric factors dominate binding of nitric oxide, azide, imidazole, cyanide, and fluoride to the rhizobial heme-based oxygen sensor FixL. *Chem. Biol.*, 3(10):841 – 850, 1996.

44. E. Nagababu, M. E. Fabry, R. L. Nagel, and J. M. Rifkind. Heme degradation and oxidative stress in murine models for hemoglobinopathies: Thalassemia, sickle cell disease and hemoglobin C disease. *Blood Cell Mol. Dis.*, 41(1):60 – 66, 2008.

45. S. Raychaudhuri, J. Skommer, K. Henty, N. Birch, and T. Brittain. Neuroglobin protects nerve cells from apoptosis by inhibiting the intrinsic pathway of cell death. *Apoptosis*, 15(4):401 – 411, 2010.

46. D. E. Goldberg, A. F. G. Slater, A. Cerami, and G. B. Henderson. Hemoglobin degradation in the malaria parasite Plasmodium-falciparum: An ordered process in a unique organelle. *Proc. Nat. Acad. Sci. USA*, 87(8):2931 – 2935, 1990.

47. A. S. Arthur, A. H. Fergus, G. Lanzino, J. Mathys, N. F. Kassell, and K. S. Lee. Systemic administration of the iron chelator deferiprone attenuates subarachnoid hemorrhage-induced cerebral vasospasm in the rabbit. *Neurosurgery*, 41(6):1385 – 1391, 1997.

48. Y. H. Lv, Q. Z. Wang, Y. Diao, and R. A. Xu. Cytoglobin: A novel potential gene medicine for fibrosis and cancer therapy. *Curr. Gene Ther.*, 8(4):287 – 294, 2008.

49. R. J. Shaw, M. M. Omar, S. Rokadiya, F. A. Kogera, D. Lowe, G. L. Hall, J. A. Woolgar, J. Homer, T. Liloglou, J. K. Field, and J. M. Risk. Cytoglobin is upregulated by tumour hypoxia and silenced by promoter hypermethylation in head and neck cancer. *Brit. J. Cancer*, 101(1):139 – 144, 2009.

50. J. Y. Chen, M. Scerbo, and G. Kramer. A review of blood substitutes: Examining the history, clinical trial results, and ethics of hemoglobin-based oxygen carriers. *Clinics*, 64(8):803 – 813, 2009.

51. P. W. Buehler and F. D'Agnillo. Toxicological consequences of extracellular hemoglobin: biochemical and physiological perspectives. *Antioxid. Redox Signal.*, 12(2):275 – 291, 2010.

52. M. F. Perutz and F. S. Mathews. An X-ray study of azide methaemoglobin. *J. Mol. Biol.*, 21(1):199 – 202, 1966.

53. D. A. Case and M. Karplus. Dynamics of ligand-binding to heme-proteins. *J. Mol. Biol.*, 132(3):343 – 368, 1979.

54. Schoenborn B. P., H. C. Watson, and J. C. Kendrew. Binding of xenon to sperm whale myoglobin. *Nature*, 207(4992):28 – 30, 1965.

55. B. P. Schoenborn and C. L. Nobbs. Binding of xenon to sperm whale deoxymyoglobin. *Mol. Pharmacol.*, 2(5):495 – 498, 1966.

56. R. F. Tilton, I. D. Kuntz, and G. A. Petsko. Cavities in proteins: structure of a metmyoglobin xenon complex solved to 1.9 Å. *Biochemistry*, 23(13):2849 – 2857, 1984.

57. G. Weber. Deuterium exchange of poly-DL-alanine in aqueous solution. *Arch. Biochem. Biophys.*, 69:106 – 118, 1957.

58. R. F. Tilton and I. D. Kuntz. Nuclear magnetic-resonance studies of Xe-129 with myoglobin and hemoglobin. *Biochemistry*, 21(26):6850 – 6857, 1982.

59. G. Weber. Ligand binding and internal equilibria in proteins. *Biochemistry*, 11(5):864 – 878, 1972.

60. F. G. Parak. Proteins in action: The physics of structural fluctuations and conformational changes. *Curr. Opin. Struct. Biol.*, 13(5):552 – 557, 2003.

61. J. L. Green, J. Fan, and C. A. Angell. The protein-class analogy - Some insights from homopeptide comparisons. *J. Phys. Chem.*, 98(51):13780 – 13790, 1994.

62. D. Vitkup, D. Ringe, G. A. Petsko, and M. Karplus. Solvent mobility and the protein 'glass' transition. *Nat. Struct. Biol.*, 7(1):34 – 38, 2000.

63. L. H. Holley and M. Karplus. Protein secondary structure prediction with a neural network. *Proc. Nat. Acad. Sci. USA*, 86(1):152 – 156, 1989.

64. J. D. Hirst and M. J. E. Sternberg. Prediction of structural and functional features of protein and nucleic-acid sequences by artificial neural networks. *Biochemistry*, 31(32):7211 – 7218, 1992.

65. H. Frauenfelder, F. Parak, and R. D. Young. Conformational substates in proteins. *Annu. Rev. Biophys. Bio.*, 17:451 – 479, 1988.

66. R. H. Austin, K. W. Beeson, L. Eisenstein, H. Frauenfelder, and I. C. Gunsalus. Dynamics of ligand-binding to myoglobin. *Biochemistry*, 14(24):5355 – 5373, 1975.

67. A. Ansari, J. Berendzen, S. F. Bowne, H. Frauenfelder, I. E. T. Iben, T. B. Sauke, E. Shyamsunder, and R. D. Young. Protein states and protein quakes. *Proc. Nat. Acad. Sci. USA*, 82(15):5000 – 5004, 1985.

68. A. Ansari, J. Berendzen, D. Braunstein, B. R. Cowen, H. Frauenfelder, M. K. Hong, I. E. T. Iben, J. B. Johnson, P. Ormos, T. B. Sauke, R. Scholl, A. Schulte, P. J. Steinbach, J. Vittitow, and R. D. Young. Rebinding and relaxation in the myoglobin pocket. *Biophys. Chem.*, 26(2-3):337 – 355, 1987.

69. D. L. Rousseau and P. V. Argade. Metastable photoproducts from carbonmonoxide myoglobin. *Proc. Nat. Acad. Sci. USA*, 83(5):1310 – 1314, 1986.

70. H. Frauenfelder, G. A. Petsko, and D. Tsernoglou. Temperature-dependent X-ray-diffraction as a probe of protein structural dynamics. *Nature*, 280(5723):558 – 563, 1979.

71. H. Hartmann, F. Parak, W. Steigemann, G. A. Petsko, D. R. Ponzi, and H. Frauenfelder. Conformational substates in a protein: Structure and dynamics of metmyoglobin at 80 K. *Proc. Nat. Acad. Sci. USA*, 79(16):4967 – 4971, 1982.

72. F. Schotte, M. H. Lim, T. A. Jackson, A. V. Smirnov, J. Soman, J. S. Olson, G. N. Phillips, M. Wulff, and P. A. Anfinrud. Watching a protein as it functions with 150-ps time-resolved X-ray crystallography. *Science*, 300(5627):1944 – 1947, 2003.

73. D. Bourgeois, B. Vallone, A. Arcovito, G. Sciara, F. Schotte, P. A. Anfinrud, and M. Brunori. Extended subnanosecond structural dynamics of myoglobin revealed by laue crystallography. *Proc. Nat. Acad. Sci. USA*, 103(13):4924 – 4929, 2006.

74. C. Keppler, K. Achterhold, A. Ostermann, U. van Bürck, W. Potzel, A. I. Chumakov, A. Q. R. Baron, R. Ruffer, and F. Parak. Determination of the phonon spectrum of iron in myoglobin using inelastic X-ray scattering of synchrotron radiation. *Eur. Biophys. J. Biophy.*, 25(3):221 – 224, 1997.

75. K. Achterhold and F. G. Parak. Protein dynamics: Determination of anisotropic vibrations at the haem iron of myoglobin. *J. Phys-Condens. Mat.*, 15(18):S1683 – S1692, 2003.

76. F. Parak and L. Reinisch. Mössbauer effect in the study of structure dynamics. *Methods Enzymol.*, 131:568 – 607, 1986.

77. G. U. Nienhaus, H. Hartmann, F. Parak, J. Heinzl, and E. Huenges. Angular dependent Rayleigh-scattering of Mössbauer radiation on proteins. *Hyperfine Interact.*, 47-8(1-4):299 – 310, 1989.

78. A. Ehrenberg, R. Rigler, and A. Graslund. *Structure, dynamics, and function of biomolecules*, volume 1 of *Springer series in biophysics*. Springer Verlag, New York, 1987.

79. H. Frauenfelder, S. G. Sligar, and P. G. Wolynes. The energy landscapes and motions of proteins. *Science*, 254(5038):1598 – 1603, 1991.

80. D. T. Leeson, D. A. Wiersma, K. Fritsch, and J. Friedrich. The energy landscape of myoglobin: An optical study. *J. Phys. Chem. B*, 101(33):6331 – 6340, 1997.

81. B. H. McMahon, J. D. Müller, C. A. Wraight, and G. U. Nienhaus. Electron transfer and protein dynamics in the photosynthetic reaction center. *Biophys. J.*, 74(5):2567 – 2587, 1998.

82. H. Frauenfelder, P. G. Wolynes, and R. H. Austin. Biological physics. *Rev. Mod. Phys.*, 71(2):S419 – S430, 1999.

83. F. G. Parak, K. Achterhold, M. Schmidt, V. Prusakov, and S. Croci. Protein dynamics on different timescales. *J. Non-Cryst. Solids*, 352(42-49):4371 – 4378, 2006.

84. F. Yang and G. N. George N. Phillips. Crystal structures of CO-, deoxy- and met-myoglobins at various pH values. *J. Mol. Biol.*, 256(4):762 – 774, 1996.

85. J. Vojtechovsky, K. Chu, J. Berendzen, R. M. Sweet, and I. Schlichting. Crystal structures of myoglobin-ligand complexes at near-atomic resolution. *Biophys. J.*, 77(4):2153 – 2174, 1999.

86. K. A. Merchant, W. G. Noid, D. E. Thompson, R. Akiyama, R. F. Loring, and M. D. Fayer. Structural assignments and dynamics of the a substates of MbCO: Spectrally resolved vibrational echo experiments and molecular dynamics simulations. *J. Phys. Chem. B*, 107(1):4 – 7, 2003.

87. C. Rovira, B. Schulze, M. Eichinger, J. D. Evanseck, and M. Parrinello. Influence of the heme pocket conformation on the structure and vibrations of the Fe-CO bond in myoglobin: A QM/MM density functional study. *Biophys. J.*, 81(1):435 – 445, 2001.

88. M. Devereux and M. Meuwly. Structural assignment of spectra by characterization of conformational substates in bound MbCO. *Biophys. J.*, 96:4363–4375, 2009.

89. K. A. Merchant, W. G. Noid, R. Akiyama, I. J. Finkelstein, A. Goun, B. L. Mcclain, R. F. Loring, and M. D. Fayer. Myoglobin-CO substate structures and dynamics: Multidimensional vibrational echoes and molecular dynamics simulations. *J. Am. Chem. Soc.*, 125(45):13804 – 13818, 2003.

90. J. O. Alben, D. Beece, S. F. Bowne, W. Doster, L. Eisenstein, H. Frauenfelder, D. Good, J. D. McDonald, M. C. Marden, P. P. Moh, L. Reinisch, A. H. Reynolds, E. Shyamsunder, and K. T. Yue. Infrared-spectroscopy of photodissociated carboxymyoglobin at low-temperatures. *Proc. Nat. Acad. Sci. USA*, 79(12):3744 – 3748, 1982.

91. H. Frauenfelder, B. H. McMahon, R. H. Austin, K. Chu, and J. T. Groves. The role of structure, energy landscape, dynamics, and allostery in the enzymatic function of myoglobin. *Proc. Nat. Acad. Sci. USA*, 98(5):2370 – 2374, 2001.

92. R. D. Young, H. Frauenfelder, J. B. Johnson, D. C. Lamb, G. U. Nienhaus, R. Philipp, and R. Scholl. Time-dependence and temperature-dependence of large-scale conformational transitions in myoglobin. *Chem. Phys.*, 158(2-3):315 – 327, 1991.

93. M. K. Hong, D. Braunstein, B. R. Cowen, H. Frauenfelder, I. E. T. Iben, J. R. Mourant, P. Ormos, R. Scholl, A. Schulte, P. J. Steinbach, A. H. Xie, and R. D. Young. Conformational substates and motions in myoglobin. External influence on structure and dynamics. *Biophys. J.*, 58(2):429 – 436, 1990.

94. H. Frauenfelder, P. W. Fenimore, and R. D. Young. Protein dynamics and function: Insights from the energy landscape and solvent slaving. *IUBMB LIFE*, 59(8-9):506 – 512, 2007.

95. M. R. Chance, B. F. Campbell, R. Hoover, and J. M. Friedman. Myoglobin recombination at low-temperature. Two phases revealed by Fourier transform infrared spectroscopy. *J. Biol. Chem.*, 262(15):6959 – 6961, 1987.

96. J. S. Olson and G. N. Phillips. Kinetic pathways and barriers for ligand binding to myoglobin. *J. Biol. Chem.*, 271(30):17593 – 17596, 1996.

97. Y. Mizutani and T. Kitagawa. Ultrafast dynamics of myoglobin probed by time-resolved resonance Raman spectroscopy. *Chem. Rec.*, 1(3):258 – 275, 2001.

98. Y. Mizutani and T. Kitagawa. Ultrafast structural relaxation of myoglobin following photodissociation of carbon monoxide probed by time-resolved resonance Raman spectroscopy. *J. Phys. Chem. B*, 105(44):10992 – 10999, 2001.

99. J. W. Petrich, J. C. Lambry, K. Kuczera, M. Karplus, C. Poyart, and J. L. Martin. Ligand binding and protein relaxation in heme proteins: A room temperature analysis of NO geminate recombination. *Biochemistry*, 30(16):3975 – 3987, 1991.

100. Q. H. Gibson, R. Regan, R. Elber, J. S. Olson, and T. E. Carver. 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, 1992.

101. T. Y. Teng, V. Srajer, and K. Moffat. Photolysis-induced structural-changes in single-crystals of carbonmonoxy myoglobin at 40 K. *Nat. Struct. Biol.*, 1(10):701 – 705, 1994.

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

103. D. R. Nutt and M Meuwly. Theoretical investigation of infrared spectra and pocket dynamics of photodissociated carbonmonoxy myoglobin. *Biophys. J.*, 85(6):3612 – 3623, 2003.

104. D. R. Nutt and M. Meuwly. Ligand dynamics in myoglobin: Calculation of infrared spectra for photodissociated NO. *ChemPhysChem*, 5(11):1710 – 1718, 2004.

105. I. Schlichting, J. Berendzen, G. N. Phillips, and R. M. Sweet. Crystal-structure of photolyzed carbonmonoxy-myoglobin. *Nature*, 371(6500):808 – 812, 1994.

106. M. H. Lim, T. A. Jackson, and P. A. Anfinrud. Midinfrared 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 – 4366, 1995.

107. T. A. Jackson, M. Lim, and P. A. Anfinrud. Complex nonexponential relaxation in myoglobin after photodissociation of MbCO: Measurement and analysis from 2 ps to 56 μs. *Chem. Phys.*, 180(2-3):131 – 140, 1994.

108. T. Kitagawa, N. Haruta, and Y. Mizutani. Time-resolved resonance Raman study on ultrafast structural relaxation and vibrational cooling of photodissociated carbonmonoxy myoglobin. *Biopolymers*, 67(4-5):207 – 213, 2002.

109. M. Lim, T. A. Jackson, and P. A. Anfinrud. Binding of CO to myoglobin from a heme pocket docking site to form nearly linear FE-C-O. *Science*, 269(5226):962 – 966, 1995.

110. S. Franzen. An electrostatic model for the frequency shifts in the carbonmonoxy stretching band of myoglobin: Correlation of hydrogen bonding and the Stark tuning rate. *J. Am. Chem. Soc.*, 124(44):13271 – 13281, 2002.

111. Y. Abadan, E. Y. T. Chien, K. Chu, C. D. Eng, G. U. Nienhaus, and S. G. Sligar. Ligand binding to heme proteins. V. Light-induced relaxation in proximal mutants L89I and H97F of carbonmonoxymyoglobin. *Biophys. J.*, 68(6):2497 – 2504, 1995.

112. M. L. Carlson, R. M. Regan, and Q. H. Gibson. Distal cavity fluctuations in myoglobin: Protein motion and ligand diffusion. *Biochemistry*, 35(4):1125 – 1136, 1996.

113. G. U. Nienhaus, K. Chu, and K. Jesse. Structural heterogeneity and ligand binding in carbonmonoxy myoglobin crystals at cryogenic temperatures. *Biochemistry*, 37(19):6819 – 6823, 1998.

114. D. C. Lamb, K. Nienhaus, A. Arcovito, F. Draghi, A. E. Miele, M. Brunori, and G. U. Nienhaus. Structural dynamics of myoglobin - Ligand migration among protein cavities studied by Fourier transform infrared/temperature derivative spectroscopy. *J. Biol. Chem.*, 277(14):11636 – 11644, 2002.

115. K. Nienhaus, P. C. Deng, J. M. Kriegl, and G. U. Nienhaus. Structural dynamics of myoglobin: Effect of internal cavities on ligand migration and binding. *Biochemistry*, 42(32):9647 – 9658, 2003.

116. R. Elber and M. Karplus. Multiple conformational states of proteins: a molecular dynamics analysis of myoglobin. *Science*, 235(4786):318 – 321, 1987.

117. D. R. Nutt and M Meuwly. Migration in native and mutant myoglobin: Atomistic simulations for the understanding of protein function. *Proc. Nat. Acad. Sci. USA*, 101(16):5998 – 6002, 2004.

118. P. Banushkina and M. Meuwly. Free-energy barriers in MbCO rebinding. *J. Phys. Chem. B*, 109(35):16911 – 16917, 2005.

119. J. Cohen, A. Arkhipov, R. Braun, and K. Schulten. Imaging the migration pathways for O_2, CO, NO, and Xe inside myoglobin. *Biophys. J.*, 91(5):1844 – 1857, 2006.

120. J. Cohen and K. Schulten. O_2 migration pathways are not conserved across proteins of a similar fold. *Biophys. J.*, 93(10):3591 – 3600, 2007.

121. D. Bourgeois, B. Vallone, F. Schotte, A. Arcovito, A. E. Miele, G. Sciara, M. Wulff, P. Anfinrud, and M. Brunori. Complex landscape of protein structural dynamics unveiled by nanosecond laue crystallography. *Proc. Nat. Acad. Sci. USA*, 100(15):8704 – 8709, 2003.

122. M. Schmidt, K. Nienhaus, R. Pahl, A. Krasselt, S. Anderson, F. Parak, G. U. Nienhaus, and V. Srajer. Ligand migration pathway and protein dynamics in myoglobin: A time-resolved crystallographic study on L29W MbCO. *Proc. Nat. Acad. Sci. USA*, 102(33):11704 – 11709, 2005.

123. E. De Marinis, L. Casella, C. Ciaccio, M. Coletta, P. Visca, and P. Ascenzi. Catalytic peroxidation of nitrogen monoxide and peroxynitrite by globins. *IUBMB LIFE*, 61(1):62 – 73, 2009.

124. U. B. Hendgen-Cotta, M. Kelm, and T. Rassaf. A highlight of myoglobin diversity: The nitrite reductase activity during myocardial ischemia-reperfusion. *Nitric Oxide-Biol Ch.*, 22(2):75 – 82, 2010.

125. S. Mishra and M. Meuwly. Atomistic simulation of NO dioxygenation in group I truncated hemoglobin. *J. Am. Chem. Soc.*, 132(9):2968 – 2982, 2010.

126. T. Burmester, B. Weich, S. Reinhardt, and T. Hankeln. A vertebrate globin expressed in the brain. *Nature*, 407(6803):520 – 523, 2000.

127. T. Hankeln, B. Ebner, C. Fuchs, F. Gerlach, M. Haberkamp, T. L. Laufs, A. Roesner, M. Schmidt, B. Weich, S. Wystub, S. Saaler-Reinhardt, S. Reuss, M. Bolognesi, D. de Sanctis, M. C. Marden, L. Kiger, L. Moens, S. Dewilde, E. Nevo, A. Avivi, R. E. Weber, A. Fago, and T. Burmester. Neuroglobin and cytoglobin in search of their role in the vertebrate globin family. *J. Inorg. Biochem.*, 99(1):110 – 119, 2005.

128. A. Roesner, C. Fuchs, T. Hankeln, and T. Burmester. A globin gene of ancient evolutionary origin in lower vertebrates: Evidence for two distinct globin families in animals. *Mol. Biol. Evol.*, 22(1):12 – 20, 2005.

129. S. Dewilde, L. Kiger, T. Burmester, T. Hankeln, V. Baudin-Creuza, T. Aerts, M. C. Marden, R. Caubergs, and L. Moens. Biochemical characterization and ligand binding properties of neuroglobin, a novel member of the globin family. *J. Biol. Chem.*, 276(42):38949 – 38955, 2001.

130. M. Schmidt, A. Giessl, T. Laufs, T. Hankeln, U. Wolfrum, and T. Burmester. How does the eye breathe? Evidence for neuroglobin-mediated oxygen supply in the mammalian retina. *J. Biol. Chem.*, 278(3):1932 – 1935, 2003.

131. R. Schmidt-Kastner, M. Haberkamp, C. Schmitz, T. Hankeln, and T. Burmester. Neuroglobin mRNA expression after transient global brain ischemia and prolonged hypoxia in cell culture. *Brain Res.*, 1103:173 – 180, 2006.

132. A. Roesner, S. A. Mitz, T. Hankeln, and T. Burmester. Globins and hypoxia adaptation in the goldfish, Carassius auratus. *FEBS J.*, 275(14):3633 – 3643, 2008.

133. M. Brunori, A. Giuffre, K. Nienhaus, G. U. Nienhaus, F. M. Scandurra, and B. Vallone. Neuroglobin, nitric oxide, and oxygen: Functional pathways and conformational changes. *Proc. Nat. Acad. Sci. USA*, 102(24):8483–8488, 2005.

134. M. G. Petersen, S. Dewilde, and A. Fago. Reactions of ferrous neuroglobin and cytoglobin with nitrite under anaerobic conditions. *J. Inorg. Biochem.*, 102(9):1777 – 1782, 2008.

135. U. Flögel, A. Gödecke, L. O. Klotz, and J. Schrader. Role of myoglobin in the antioxidant defense of the heart. *FASEB J.*, 18(7):1156 – +, 2004.

136. A. A. Khan, Y. M. Wang, Y. J. Sun, X. O. Mao, L. Xie, E. Miles, J. Graboski, S. Chen, L. M. Ellerby, K. L. Jin, and D. A. Greenberg. Neuroglobin-overexpressing transgenic mice are resistant to cerebral and myocardial ischemia. *Proc. Nat. Acad. Sci. USA*, 103(47):17944 – 17948, 2006.

137. S. Watanabe and K. Wakasugi. Neuroprotective function of human neuroglobin is correlated with its guanine nucleotide dissociation inhibitor activity. *Biochem. Bioph. Res. Co.*, 369(2):695 – 700, 2008.

138. A. Fago, A. J. Mathews, L. Moens, S. Dewilde, and T. Brittain. The reaction of neuroglobin with potential redox protein partners cytochrome b(5) and cytochrome c. *FEBS Lett.*, 580(20):4884 – 4888, 2006.

139. A. Fago, C. Hundahl, S. Dewilde, K. Gilany, L. Moens, and R. E. Weber. Allosteric regulation and temperature dependence of oxygen binding in human neuroglobin and cytoglobin - Molecular mechanisms and physiological significance. *J. Biol. Chem.*, 279(43):44417 – 44426, 2004.

140. A. Pesce, S. Dewilde, M. Nardini, L. Moens, P. Ascenzi, T. Hankeln, T. Burmester, and M. Bolognesi. Human brain neuroglobin structure reveals a distinct mode of controlling oxygen affinity. *Structure*, 11(9):1087 – 1095, 2003.

141. B Vallone, K Nienhaus, M Brunori, and GU Nienhaus. The structure of murine neuroglobin: Novel pathways for ligand migration and binding. *Proteins*, 56(1):85 – 92, 2004.

142. A. Pesce, S. Dewilde, M. Nardini, L. Moens, P. Ascenzi, T. Hankeln, T. Burmester, and M. Bolognesi. The human brain hexacoordinated neuroglobin three-dimensional structure. *Micron*, 35(1-2):63 – 65, 2004.

143. B. Vallone, K. Nienhaus, A. Matthes, M. Brunori, and G. U. Nienhaus. The structure of carbonmonoxy neuroglobin reveals a heme-sliding mechanism for control of ligand affinity. *Proc. Nat. Acad. Sci. USA*, 101(50):17351 – 17356, 2004.

144. M. Brunori and B. Vallone. Neuroglobin, seven years after. *Cell. Mol. Life Sci.*, 64(10):1259 – 1268, 2007.

145. J. T. Trent, A. N. Hvitved, and M. S. Hargrove. A model for ligand binding to hexacoordinate hemoglobins. *Biochemistry*, 40(20):6155 – 6163, 2001.

146. K. Nienhaus, J. M. Kriegl, and G. U. Nienhaus. Structural dynamics in the active site of murine neuroglobin and its effects on ligand binding. *J. Biol. Chem.*, 279(22):22944 – 22952, 2004.

147. J. M. Kriegl, A. J. Bhattacharyya, K. Nienhaus, P. Deng, O. Minkow, and G. U. Nienhaus. Ligand binding and protein dynamics in neuroglobin. *Proc. Nat. Acad. Sci. USA*, 99(12):7992 – 7997, 2002.

148. D. de Sanctis, S. Dewilde, A. Pesce, L. Moens, P. Ascenzi, T. Hankeln, T. Burmester, and M. Bolognesi. Crystal structure of cytoglobin: The fourth globin type discovered in man displays heme hexa-coordination. *J. Mol. Biol.*, 336(4):917 – 927, 2004.

149. M. Makino, H. Sugimoto, H. Sawai, N. Kawada, K. Yoshizato, and Y. Shiro. High-resolution structure of human cytoglobin: identification of extra N- and C-termini and a new dimerization mode. *Acta Crystallogr. Sect. D: Biol. Crystallogr.*, 62:671 – 677, 2006.

150. D. de Sanctis, S. Dewilde, C. Vonrhein, A. Pesce, L. Moens, P. Ascenzi, T. Hankeln, T. Burmester, M. Ponassi, M. Nardini, and M. Bolognesi. Bishistidyl heme hexacoordination, a key structural property in drosophila melanogaster hemoglobin. *J. Biol. Chem.*, 280(29):27222 – 27229, 2005.

151. J. Yoon, M. A. Herzik, M. B. Winter, R. Tran, C. Olea, and M. A. Marletta. Structure and properties of a bis-histidyl ligated globin from caenorhabditis elegans. *Biochemistry*, 49(27):5662 – 5670, 2010.

152. F. A. J. Rotsaert, B. M. Hallberg, S. de Vries, P. Moenne-Loccoz, C. Divne, V. Renganathan, and M. H. Gold. Biophysical and structural analysis of a novel heme B iron ligation in the flavocytochrome cellobiose dehydrogenase. *J. Biol. Chem.*, 278(35):33224 – 33231, 2003.

153. N. Nakanishi, F. Takeuchi, H. Okamoto, A. Tamura, H. Hori, and M. Tsubaki. Characterization of heme-coordinating histidyl residues of cytochrome b(5) based on the reactivity with diethylpyrocarbonate: A mechanism for the opening of axial imidazole rings. *J. Biochem.*, 140(4):561 – 571, 2006.

154. D. Hamdane, L. Kiger, S. Dewilde, B. N. Green, A. Pesce, J. Uzan, T. Burmester, T. Hankeln, M. Bolognesi, L. Moens, and M. C. Marden. The redox state of the cell regulates the ligand binding affinity of human neuroglobin and cytoglobin. *J. Biol. Chem.*, 278(51):51713 – 51721, 2003.

155. H. Ishikawa, S. Kim, K. Kwak, K. Wakasugi, and M. D. Fayer. Disulfide bond influence on protein structural dynamics probed with 2d-ir vibrational echo spectroscopy. *Proc. Nat. Acad. Sci. USA*, 104(49):19309 – 19314, 2007.

156. D. Hamdane, L. Kiger, S. Dewilde, B. N. Green, A. Pesce, J. Uzan, T. Burmester, T. Hankeln, M. Bolognesi, L. Moens, and M. C. Marden. Coupling of the heme and an internal disulfide bond in human neuroglobin. *Micron*, 35(1-2):59 – 62, 2004.

157. L. Kiger, J. Uzan, S. Dewilde, T. Burmester, T. Hankeln, L. Moens, D. Hamdane, V. Baudin-Creuza, and M. C. Marden. Neuroglobin ligand binding kinetics. *IUBMB LIFE*, 56(11-12):709 – 719, 2004.

158. K. Nienhaus and G. U. Nienhaus. A spectroscopic study of structural heterogeneity and carbon monoxide binding in neuroglobin. *J. Biol. Phys.*, 31(3-4):417 – 432, 2005.

159. S. Abbruzzetti, S. Faggiano, S. Bruno, F. Spyrakis, A. Mozzarelli, S. Dewilde, L. Moens, and C. Viappiani. Ligand migration through the internal hydrophobic cavities in human neuroglobin. *Proc. Nat. Acad. Sci. USA*, 106(45):18984 – 18989, 2009.

160. A. Bocahut, S. Bernad, P. Sebban, and S. Sacquin-Mora. Relating the diffusion of small ligands in human neuroglobin to its structural and mechanical properties. *J. Phys. Chem. B*, 113(50):16257 – 16267, 2009.

161. R. Elber and Q. H. Gibson. Toward quantitative simulations of carbon monoxide escape pathways in myoglobin. *J. Phys. Chem. B*, 112(19):6147 – 6154, 2008.

162. S. Lutz, K. Nienhaus, G. U. Nienhaus, and M. Meuwly. Ligand migration between internal docking sites in photodissociated carbonmonoxy neuroglobin. *J. Phys. Chem. B*, 113(46):15334 –15343, 2009.

163. H. Sawai, M. Makino, Y. Mizutani, T. Ohta, H. Sugimoto, T. Uno, N. Kawada, K. Yoshizato, T. Kitagawa, and Y. Shiro. Structural characterization of the proximal and distal histidine environment of cytoglobin and neuroglobin. *Biochemistry*, 44(40):13257 – 13265, 2005.

164. H. Ishikawa, I. J. Finkelstein, S. Kim, K. Kwak, J. K. Chung, K. Wakasugi, A. M. Massari, and M. D. Fayer. Neuroglobin dynamics observed with ultrafast 2D-IR vibrational echo spectroscopy. *Proc. Nat. Acad. Sci. USA*, 104(41):16116 – 16121, 2007.

165. B. J. Alder and T. E. Wainwright. Phase transition for a hard sphere system. *J. Chem. Phys.*, 27(5):1208 – 1209, 1957.

166. B. J. Alder and T. E. Wainwright. Studies in molecular dynamics. i. general method. *J. Chem. Phys.*, 31(2):459 – 466, 1959.

167. A. Rahman. Correlations in motion of atoms in liquid argon. *Phys. Rev. A-Gen. Phys.*, 136(2A):A405 – A411, 1964.

168. L. Verlet. Computer “experiments” on classical fluids. i. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.*, 159(1):98 – 103, 1967.

169. F. H. Stillinger and A. Rahman. Improved simulation of liquid water by molecular-dynamics. *J. Chem. Phys.*, 60(4):1545 – 1557, 1974.

170. F. H. Stillinger and A. Rahman. Molecular-dynamics study of liquid water under high compression. *J. Chem. Phys.*, 61(12):4973 – 4980, 1974.

171. J. A. McCammon, B. R. Gelin, and M. Karplus. Dynamics of folded proteins. *Nature*, 267(5612):585 – 590, 1977.

172. R. M. Levy, R. P. Sheridan, J. W. Keepers, G. S. Dubey, S. Swaminathan, and M. Karplus. Molecular-dynamics of myoglobin at 298 degrees K. Results from a 300-ps computer simulation. *Biophys. J.*, 48(3):509 – 518, 1985.

173. P. L. Freddolino, A. S. Arkhipov, S. B. Larson, A. McPherson, and K. Schulten. Molecular dynamics simulations of the complete satellite tobacco mosaic virus. *Structure*, 14(3):437 – 449, 2006.

174. X. Y. Li, Y. J. Wei, L. Lu, K. Lu, and H. J. Gao. Dislocation nucleation governed softening and maximum strength in nano-twinned metals. *Nature*, 464(7290):877 – 880, 2010.

175. M. Born and R. Oppenheimer. Quantum theory of molecules. *Ann. Phys.-Berlin*, 84(20):0457 – 0484, 1927.

176. B. R. Brooks *et al.* CHARMM: The biomolecular simulation program. *J. Chem. Comp.*, 30(10):1545 – 1614, 2009.

177. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. A 2nd generation

force-field for the simulation of proteins, nucleic-acids, and organic-molecules. *J. Am. Chem. Soc.*, 117(19):5179 – 5197, 1995.

178. C. Oostenbrink, A. Villa, A. E. Mark, and W. F. Van Gunsteren. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J. Chem. Comp.*, 25(13):1656 – 1676, 2004.

179. W. L. Jorgensen and J. Tirado-Rives. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.*, 110(6):1657 – 1666, 1988.

180. R. D. Ruth. A canonical integration technique. *IEEE T. Nucl. Sci.*, 30(4):2669 – 2671, 1983.

181. A. Kolinski and J. Skolnick. Discretized model of proteins. I. Monte-Carlo study of cooperativity in homopolypeptides. *J. Chem. Phys.*, 97(12):9412 – 9426, 1992.

182. Warshel A. and Levitt M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.*, 103(2):227 – 249, 1976.

183. P. A. Bash, M. J. Field, and M. Karplus. Free-energy pertubation method for chemical reactions in the condensed phase: A dynamical approach based on a combined quantum and molecular mechanics potential. *J. Am. Chem. Soc.*, 109(26):8092 – 8094, 1987.

184. J. L. Gao. Absolute free-energy of solvation from Monte-Carlo simulations using combined quantum and molecular mechanical potentials. *J. Phys. Chem.*, 96(2):537 – 540, 1992.

185. V. Guallar, A. A. Jarzecki, R. A. Friesner, and T. G. Spiro. Modeling of ligation-induced helix/loop displacements in myoglobin: Toward an understanding of hemoglobin allostery. *J. Am. Chem. Soc.*, 128(16):5427 – 5435, 2006.

186. R. E. Alcantara, C. Xu, T. G. Spiro, and V. Guallar. A quantum-chemical picture of hemoglobin affinity. *Proc. Nat. Acad. Sci. USA*, 104(47):18451 – 18455, 2007.

187. V. Guallar, C. Y. Lu, K. Borrelli, T. Y. Egawa, and S. R. Yeh. Ligand migration in the truncated hemoglobin-II from mycobacterium tuberculosis: The role of G8 tryptophan. *J. Biol. Chem.*, 284(5):3106 – 3116, 2009.

188. M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart. Development and use of quantum mechanical molecular models. 76. AM1: A new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.*, 107(13):3902 – 3909, 1985.

189. J. J. P. Stewart. Optimization of parameters for semiempirical methods I. Method. *J. Chem. Comp.*, 10(2):209 – 220, 1989.

190. J. J. P. Stewart. Optimization of parameters for semiempirical methods II. Applications. *J. Chem. Comp.*, 10(2):221 – 264, 1989.

191. G. B. Rocha, R. O. Freire, A. M. Simas, and J. J. P. Stewart. RM1: A reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I. *J. Chem. Comp.*, 27(10):1101 – 1111, 2006.

192. J. J. P. Stewart. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *J. Mol. Model.*, 13(12):1173 – 1213, 2007.

193. T. Frauenheim, D. Porezag, M. Elstner, G. Jungnickel, J. Elsner, M. Haugk, and A. Sieck. An ab initio two-center tight-binding approach to simulations of complex materials properties. *Mater. Res. Soc. Symp. P.*, 491:91 – 104, 1998.

194. D. Riccardi, P. Schaefer, Y. Yang, H. B. Yu, N. Ghosh, X. Prat-Resina, P. König, G. H. Li, D. G. Xu, H. Guo, M. Elstner, and Q. Cui. Development of effective quantum mechanical/molecular mechanical (QM/MM) methods for complex biological processes. *J. Phys. Chem. B*, 110(13):6458 – 6469, 2006.

195. Q. Cui, M. Elstner, and M. Karplus. A theoretical analysis of the proton and hydride transfer in liver alcohol dehydrogenase (LADH). *J. Phys. Chem. B*, 106(10):2721 – 2740, 2002.

196. R. Car and M. Parrinello. Unified approach for molecular-dynamics and density-functional theory. *Phys. Rev. Lett.*, 55(22):2471 – 2474, 1985.

197. M. Eichinger, P. Tavan, J. Hutter, and M. Parrinello. A hybrid method for solutes in complex solvents: Density functional theory combined with empirical force fields. *J. Chem. Phys.*, 110(21):10452 – 10467, 1999.

198. C. Rovira. Role of the His64 residue on the properties of the Fe-CO and Fe-O-2 bonds in myoglobin. A CHARMM/DFT study. *J. Mol. Struc-Theochem.*, 632:309 – 321, 2003.

199. A. Warshel. Dynamics of enzymatic-reactions. *Proc. Nat. Acad. Sci. USA*, 81(2):444 – 448, 1984.

200. A. Warshel and S. Russell. Theoretical correlation of structure and energetics in the catalytic reaction of trypsin. *J. Am. Chem. Soc.*, 108(21):6569 – 6579, 1986.

201. S. C. L. Kamerlin and A. Warshel. The EVB as a quantitative tool for formulating simulations and analyzing biological and chemical reactions. *Faraday Discuss.*, 145:71 – 106, 2010.

202. C. M. Maupin, R. McKenna, D. N. Silverman, and G. A. Voth. Elucidation of the proton transport mechanism in human carbonic anhydrase II. *J. Am. Chem. Soc.*, 131(22):7598 – 7608, 2009.

203. S. Braun-Sand, M. Strajbl, and A. Warshel. Studies of proton translocations in biological systems: Simulating proton transport in carbonic anhydrase by EVB-based models. *Biophys. J.*, 87(4):2221 – 2239, 2004.

204. G. G. Hall and C. M. Smith. Fitting electron-densities of molecules. *Int. J. Quantum Chem.*, 25(5):881 – 890, 1984.

205. C. M. Smith and G. G. Hall. The approximation of electron-densities. *Theor. Chim. Acta*, 69(1):63 – 69, 1986.

206. M. Meuwly. On the influence of the local environment on the CO stretching frequencies in native myoglobin: Assignment of the B-states in MbCO. *ChemPhysChem*, 7(10):2061 – 2063, 2006.

207. A. Warshel, M. Kato, and A. V. Pisliakov. Polarizable force fields: History, test cases, and prospects. *J. Chem. Theory Comput.*, 3(6):2034 – 2045, 2007.

208. J. L. Gao. Toward a molecular orbital derived empirical potential for liquid simulations. *J. Phys. Chem. B*, 101(4):657 – 663, 1997.

209. J. L. Gao, D. Habibollazadeh, and L. Shao. A polarizable intermolecular potential function for simulation of liquid alcohols. *J. Phys. Chem.*, 99(44):16460 – 16467, 1995.

210. S. Patel and C. L. Brooks. CHARMM fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations. *J. Chem. Comp.*, 25(1):1 – 15, 2004.

211. N. Gresh, G. A. Cisneros, T. A. Darden, and J. P. Piquemal. Anisotropic, polarizable molecular mechanics studies of inter- and intramoecular interactions and ligand-macromolecule complexes. A bottom-up strategy. *J. Chem. Theory Comput.*, 3(6):1960 – 1986, 2007.

212. P. Y. Ren and J. W. Ponder. Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations. *J. Chem. Comp.*, 23(16):1497 – 1506, 2002.

213. J. W. Ponder, C. J. Wu, P. Y. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. Distasio, M. Head-Gordon, G. N. I. Clark, M. E. Johnson, and T. Head-Gordon. Current status of the AMOEBA polarizable force field. *J. Phys. Chem. B*, 114(8):2549 – 2564, 2010.

214. A. J. Stone and S. L. Price. Some new ideas in the theory of intermolecular forces - Anisotropic atom atom potentials. *J. Phys. Chem.*, 92(12):3325 – 3335, 1988.

215. A. J. Stone and M. Alderton. Distributed multipole analysis: Methods and applications. *Mol. Phys.*, 56(5):1047 – 1064, 1985.

216. R. S. Mulliken. Electronic population analysis on LCAO-MO molecular wave functions. *J. Chem. Phys.*, 23(10):1833 – 1840, 1955.

217. A. J. Stone. Distributed multipole analysis, or how to describe a molecular charge distribution. *Chem. Phys. Lett.*, 83(2):233 – 239, 1981.

218. W. A. Sokalski and R. A. Poirier. Cumulative atomic multipole representation of the molecular charge distribution and its basis set dependence. *Chem. Phys. Lett.*, 98(1):86 – 92, 1983.

219. A. J. Stone. *The Theory of Intermolecular Forces*. Clarendon Press, Oxford, 1996.

220. N. Plattner and M. Meuwly. The role of higher CO-multipole moments in understanding the dynamics of photodissociated carbonmonoxide in myoglobin. *Biophys. J.*, 94(7):2505 – 2515, 2008.

221. N. Plattner and M. Meuwly. Atomistic simulations of CO vibrations in ices relevant to astrochemistry. *ChemPhysChem*, 9:1271 – 1277, 2008.

222. N. Plattner, T. Bandi, J. D. Doll, D. L. Freeman, and M. Meuwly. MD simulations using distributed multipole electrostatics: Structural and spectroscopic properties of CO- and methane-containing clathrates. *Mol. Phys.*, 106(12-13):1675 – 1684, 2008.

223. K. Nienhaus, S. Lutz, M. Meuwly, and G. U. Nienhaus. Structural identification of spectroscopic substates in neuroglobin. *ChemPhysChem*, 11(1):119 – 129, 2010.

224. D. R. Nutt and M. Meuwly. Studying reactive processes with classical dynamics: Rebinding dynamics in MbNO. *Biophys. J.*, 90(4):1191 – 1201, 2006.

225. J. Danielsson and M. Meuwly. Atomistic simulation of adiabatic reactive processes based on multi-state potential energy surfaces. *J. Chem. Theory Comput.*, 4(7):1083 – 1093, 2008.

226. D. G. Truhlar, B. C. Garrett, and S. J. Klippenstein. Current status of transition-state theory. *J. Phys. Chem.*, 100(31):12771 – 12800, 1996.

227. K. Kuczera, J. Kuriyan, and M. Karplus. Temperature dependence of the structure and dynamics of myoglobin. A simulation approach. *J. Mol. Biol.*, 213(2):351 – 373, 1990.

228. X. Y. Li and M. Z. Zgierski. Iron motion in a 5-coordinated heme model. *Chem. Phys. Lett.*, 188(1-2):16 – 20, 1992.

229. J. E. Straub and M. Karplus. Molecular dynamics study of the photodissociation of carbon monoxide from myoglobin: Ligand dynamics in the first 10 ps. *Chem. Phys.*, 158(2-3):221 – 248, 1991.

230. M. A. Lopez and P. A. Kollman. Application of molecular dynamics and free energy pertubation methods to metalloporphyrin-ligand systems II: CO and dioxygen binding to myoglobin. *Protein Sci.*, 2(11):1975 – 1986, 1993.

231. M. J. Frisch *et al.* *Gaussian 03, Revision C.01*. Gaussian, Inc., Wallingford CT, U.S.A., 2004.

232. J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais. Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Phys. Rev. B*, 46(11):6671 – 6687, Sep 1992.

233. P. J. Hay and W. R. Wadt. *Ab initio* effective core potentials for molecular calculations. Potentials for the transition metal atoms Sc to Hg. *J. Chem. Phys.*, 82(1):270 – 283, 1985.

234. P. J. Hay and W. R. Wadt. *Ab initio* effective core potentials for molecular calculations. Potentials for K to Au including the outermost core orbitals. *J. Chem. Phys.*, 82(1):299 – 310, 1985.

235. B. H. McMahon, B. P. Stojkovic, P. J. Hay, R. L. Martin, and A. E. Garcia. Microscopic model of carbon monoxide binding to myoglobin. *J. Chem. Phys.*, 113(16):6831 – 6850, 2000.

236. S. Y. Park, T. Yokoyama, N. Shibayama, Y. Shiro, and J. R. H. Tame. 1.25 Å resolution crystal structures of human haemoglobin in the oxy, deoxy and carbonmonoxy forms. *J. Mol. Biol.*, 360(3):690 – 701, 2006.

237. E. H. Harutyunyan, T. N. Safonova, I. P. Kuranova, A. N. Popov, A. V. Teplyakov, G. V. Obmolova, A. A. Rusakov, B. K. Vainshtein, G. G. Dodson, J. C. Wilson, and M. F. Perutz. The structure of deoxy-leghemoglobin and oxy-leghemoglobin from lupin. *J. Mol. Biol.*, 251(1):104 – 115, 1995.

238. V. Starovoitova, G. R. A. Wyllie, W. R. Scheidt, W. Sturhahn, E. E. Alp, and S. M. Durbin. Intermolecular dynamics in crystalline iron octaethylporphyrin (feoep). *J. Phys. Chem. B*, 112(40):12656 – 12661, 2008.

239. F. Autenrieth, E. Tajkhorshid, J. Baudry, and Z. Luthey-Schulten. Classical force field parameters for the heme prosthetic group of cytochrome c. *J. Chem. Comp.*, 25(13):1613 – 1622, 2004.

240. B. K. Rai, S. M. Durbin, E. W. Prohofsky, J. T. Sage, M. K. Ellison, W. R. Scheidt, W. Sturhahn, and E. E. Alp. Iron normal mode dynamics in a porphyrin-imidazole model for deoxyheme proteins. *Phys. Rev. E*, 66(5):051904, 2002.

241. B. M. Leu, N. J. Silvernail, M. Z. Zgierski, G. R. A. Wyllie, M. K. Ellison, W. R. Scheidt, J. Y. Zhao, W. Sturhahn, E. E. Alp, and J. T. Sage. Quantitative vibrational dynamics of iron in carbonyl porphyrins. *Biophys. J.*, 92(11):3764 – 3783, 2007.

242. M. L. Mitchell, X. Y. Li, J. R. Kincaid, and T. G. Spiro. Axial ligand and out-of-plane vibrations for bis(imidazolyl)heme: Raman and infrared iron-54, nitrogen-15, and deuterium isotope shifts and normal coordinate calculations. *J. Phys. Chem.*, 91(18):4690 – 4696, 1987.

243. A.D. Mackerel Jr., C.L. Brooks III, L. Nilsson, B. Roux, Y. Won, and M. Karplus. *CHARMM: The energy function and its parameterization with an overview of the program*, volume 1 of *The encyclopedia of computational chemistry*, pages 271 – 277. John Wiley & Sons: Chichester, 1998.

244. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79(2):926 – 935, 1983.

245. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Chem. Comp.*, 4(2):187 – 217, 1983.

246. J. N. Huffaker. Diatomic-molecules as perturbed Morse oscillators. I. Energy-levels. *J. Chem. Phys.*, 64(8):3175 – 3181, 1976.

247. J. N. Huffaker. Diatomic-molecules as perturbed Morse oscillators. II. Extension to higher-order parameters. *J. Chem. Phys.*, 64(11):4564 – 4570, 1976.

248. S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman. Multidimensional free-energy calculations using the weighted histogram analysis method. *J. Chem. Comp.*, 16(11):1339 – 1350, 1995.

249. R. A. Laskowski. SURFNET: A program for visualizing molecular-surfaces, cavities, and intermolecular interactions. *J. Mol. Graph.*, 13(5):323 – 330, 1995.

250. O. Beckstein, N. Michaud-Agrawal, and Woolf T. B. Quantitative analysis of water dynamics in and near proteins. *Biophys. J.*, 96(3, Supplement 1):601a – 601a, 2009.

A. Appendix

A.1. Parameter tables

Coordinate	dx [Å, °]	cx [Å, °]	bis-His [Å, °]	from Ref. (227) [Å, °]
CPB-C	1.48	1.51	1.51	1.3800
CPB-CPA	1.42	1.44	1.44	1.4432
CPB-CPB	1.36	1.36	1.37	1.3464
CT2-CPB	1.58	1.53	1.53	1.4900
CT3-CPB	1.53	1.53	1.53	1.4900
FE-CM	—	1.77	—	1.9000
NPH1-CPA	1.39	1.38	1.38	1.3757
NPH2-CPA	1.39	1.38	1.38	1.3757
NPH1-FE	2.04	2.01	2.02	1.9580
NPH2-FE	2.04	2.01	2.02	1.9580
NR2-FE	2.22	2.09	2.04	2.2000
OM-CM	—	1.15	—	1.1280
CPM-CPA	1.39	1.38	1.37	1.3716
CPA-CPB-C	122.5	124.7	124.7	126.74
CPA-CPM-CPA	126.0	125.3	127.0	125.12
CPA-NPH1-CPA	105.3	106.0	106.2	103.90
CPA-NPH2-CPA	105.3	106.0	106.2	103.90
CPB-C-C	126.1	126.5	126.5	121.50
CPB-CPB-C	128.4	128.4	128.4	126.75
CPB-CPB-CPA	107.2	107.2	107.0	106.51
CPM-CPA-CPB	125.6	125.8	125.5	124.07
CT2-CPB-CPA	122.5	124.7	124.7	126.74
CT2-CPB-CPB	128.4	128.4	128.4	126.75
CT2-CT2-CPB	110.0	110.9	110.9	113.00
CT3-CPB-CPA	122.5	124.7	124.7	126.74
CT3-CPB-CPB	128.4	128.4	128.4	126.75
FE-NPH1-CPA	127.0	127.2	126.8	128.05
FE-NPH2-CPA	127.0	127.2	126.8	128.05
FE-NR2-CPH1	127.4	127.8	127.8	133.00
FE-NR2-CPH2	126.1	125.7	125.7	123.00
NR2-FE-NPH1	98.2	90.0	90.0	90.00
NR2-FE-NPH2	98.2	90.0	90.0	90.00
NPH1-CPA-CPB	109.7	109.8	110.0	111.54
NPH2-CPA-CPB	109.7	109.8	110.0	111.54
NPH1-CPA-CPM	124.4	124.5	124.5	124.39
NPH2-CPA-CPM	124.4	124.5	124.5	124.39
NPH1-FE-NPH2	89.0	90.0	89.0	90.00
NPH1-FE-NPH1	180.0	180.0	180.0	180.0
NPH2-FE-NPH2	180.0	180.0	180.0	180.0

Table A.1.: Averaged internal coordinates obtained from three high-resolution globin X-ray structures each for a dx, a cx and a bis-His configured heme system.

Stretching-mode	k_b from Ref. (238)	k_b from Ref. (227)	Torsional-mode	k_χ from Ref. (238)	n	δ
CPB-C	388.6	441.3	NPH1-FE-NPH1-CPA	1.44	2	180.0
CPB-CPA	388.6	299.8	NPH2-FE-NPH2-CPA	1.44	2	180.0
CPB-CPB	367.0	340.7	NPH1-CPA-CPB-CPB	1.08	2	180.0
CT2-CPB	388.6	441.3	NPH2-CPA-CPB-CPB	1.08	2	180.0
CT3-CPB	388.6	441.3	NPH1-CPA-CPB-C	1.08	2	180.0
NPH1-CPA	475.0	377.2	NPH2-CPA-CPB-C	1.08	2	180.0
NPH2-CPA	475.0	377.2	NPH1-CPA-CPB-CT2	1.08	2	180.0
NPH1-FE	52.5	270.2	NPH2-CPA-CPB-CT2	1.08	2	180.0
NPH2-FE	52.5	270.2	NPH1-CPA-CPB-CT3	1.08	2	180.0
CPM-CPA	273.5	360.0	NPH2-CPA-CPB-CT3	1.08	2	180.0
Bending-mode	**k_θ from Ref. (238)**	**k_θ from Ref. (227)**	NPH1-CPA-CPM-CPA	2.16	2	180.0
CPA-CPB-C	12.2	70.0	NPH2-CPA-CPM-CPA	2.16	2	180.0
CPA-CPM-CPA	4.3	94.2	CPA-NPH1-CPA-CPB	1.14	2	180.0
CPA-NPH1-CPA	179.9	139.3	CPA-NPH2-CPA-CPB	1.14	2	180.0
CPA-NPH2-CPA	179.9	139.3	CPA-NPH1-CPA-CPM	0.07	2	180.0
CPB-C-C	5.8	70.0	CPA-NPH2-CPA-CPM	0.07	2	180.0
CPB-CPB-C	3.6	70.0	CPA-CPB-CPB-CPA	0.0	2	180.0
CPB-CPB-CPA	30.9	30.08	CPA-CPB-CPB-C	0.22	2	180.0
CPM-CPA-CPB	12.2	61.6	CPA-CPB-CPB-CT2	0.22	2	180.0
CT2-CPB-CPA	12.2	65.0	CPA-CPB-CPB-CT3	0.22	2	180.0
CT2-CPB-CPB	3.6	65.0	CPA-CPM-CPA-CPB	1.08	2	180.0
CT2-CT2-CPB	5.8	70.0	CPB-CPB-CPA-CPM	0.22	2	180.0
CT3-CPB-CPA	12.2	65.0	CPM-CPA-CPB-C	0.22	2	180.0
CT3-CPB-CPB	3.6	65.0	CPM-CPA-CPB-CT2	0.22	2	180.0
FE-NPH1-CPA	15.8	96.15	CPM-CPA-CPB-CT3	0.22	2	180.0
FE-NPH2-CPA	15.8	96.15	CT3-CPB-CPB-C	5.18	2	180.0
FE-NR2-CPH1	9.5[a]	30.0	CT3-CPB-CPB-CT2	5.18	2	180.0
FE-NR2-CPH2	9.5[a]	30.0	NPH1-FE-NR2-CPH1[b]	0.047	4	0.0
NR2-FE-NPH1	27.0[a]	50.0	NPH2-FE-NR2-CPH1[b]	0.047	4	0.0
NR2-FE-NPH2	27.0[a]	50.0	NPH1-FE-NR2-CPH2[b]	0.047	4	0.0
NPH1-CPA-CPB	2.9	122.0	NPH2-FE-NR2-CPH2[b]	0.047	4	0.0
NPH2-CPA-CPB	2.9	122.0	CPA-NPH1-FE-NR2[c]	0.5	2	0.0
NPH1-CPA-CPM	8.6	88.0	CPA-NPH2-FE-NR2[c]	0.5	2	0.0
NPH2-CPA-CPM	8.6	88.0	CPA-CPB-C-C	0.14	6	0.0
NPH1-FE-NPH2	8.6	50.0	CPA-CPB-CT2-CT2	0.14	6	0.0
NPH1-FE-NPH1	0.0	0.0	CPB-CPB-C-C	1.51	6	0.0
NPH2-FE-NPH2	0.0	0.0	CPB-CPB-CT2-CT2	1.51	6	0.0

Table A.2.: Comparison of heme stretching-, bending- and proper torsional-mode force-constants common to all three heme configurations in the present parameterization and the parameterization of Kuczera et al. (227) The force-constants are converted from the normal-mode analysis by Starovoitova et al.(238) if not mentioned otherwise. Units are kcal mol^{-1} Å^{-2} for k_b, kcal mol^{-1} rad^{-2} for k_θ, kcal mol^{-1} for k_χ and ° for δ. [a] Obtained by fitting to NRVS frequencies analysis(240, 241); [b] adapted from Autenrieth et al. (239); [c] adapted from related CHARMM parameters for the THF-imidazole system.

Stretching-mode	**configuration**	k_b			**Torsional-mode**	**configuration**	k_χ	n	δ
FE-NR2	dx	72.0			CPA-NPH1-FE-NPH2	dx	0.72	2	180.0
	cx	59.0				cx	0.72	2	180.0
	bis-His	102.0				bis-His	48.7	2	180.0
FE-CM	cx	205.0			CPA-NPH2-FE-NPH1	dx	0.72	2	180.0
Bending-mode	**configuration**	k_θ				cx	0.72	2	180.0
NPH1-FE-CM	cx	27.0				bis-His	48.7	2	180.0
NPH2-FE-CM	cx	50.0			FE-NPH1-CPA-CPB	dx	0.0	2	180.0
OM-CM-FE	cx	42.5				cx	0.86	2	180.0
NR2-FE-CM	cx	50.0				bis-His	0.86	2	180.0
Torsional-mode	**configuration**	k_χ	n	δ	FE-NPH2-CPA-CPB	dx	0.0	2	180.0
NPH1-CPA-CPM-HA	all	14.0	2	180.0		cx	0.86	2	180.0
NPH2-CPA-CPM-HA	all	14.0	2	180.0		bis-His	0.86	2	180.0
CPB-CPA-CPM-HA	all	3.0	2	180.0	FE-NPH1-CPA-CPM	dx	0.0	2	180.0
CPA-CPB-CT2-HA	all	0.19	6	0.0		cx	0.43	2	180.0
CPA-CPB-CT3-HA	all	0.19	6	0.0		bis-His	0.43	2	180.0
CPA-CPB-C-HA	all	0.14	6	0.0	FE-NPH2-CPA-CPM	dx	0.0	2	180.0
CPB-CPB-CT2-HA	all	0.19	6	0.0		cx	0.43	2	180.0
CPB-CPB-CT3-HA	all	0.19	6	0.0		bis-His	0.43	2	180.0
CPB-CPB-C-HA	all	1.51	6	0.0	CM-FE-NPH1-CPA	cx	0.5	2	0.0
CPB-CT2-CT2-HA	all	0.195	3	0.0					

Table A.3.: Heme configuration specific stretching-, bending- and proper torsional-mode parameters and remaining proper torsional parameters connected to porphyrin hydrogen atoms. Units are kcal mol^{-1} Å^{-2} for k_b, kcal mol^{-1} rad^{-2} for k_θ, kcal mol^{-1} for k_χ and ° for δ.

Improper	**configuration**	k_ϕ	**Improper from Ref. (227)**	k_ϕ
NPH1-FE-CPA-CPA	all	42.5	HA-CPA-CPA-CPM	29.4
NPH2-FE-CPA-CPA	all	42.5	HA-CPB-C-C	20.0
CPA-NPH1-CPB-CPM	all	46.8	CPB-X-X-C	90.0
CPA-NPH2-CPB-CPM	all	46.8	CPB-X-X-CT2	90.0
CPB-CPA-CPB-C	all	37.4	CPB-X-X-CT3	90.0
CPB-CPA-CPB-CT2	all	37.4		
CPB-CPA-CPB-CT3	all	37.4		
NR2-CPH1-CPH2-FE[a]	all	29.4		
FE-NPH1-NPH2-NPH1	dx	0.0		
	cx	5.2		
	bis-His	5.2		
FE-NPH2-NPH1-NPH2	dx	0.0		
	cx	5.2		
	bis-His	5.2		

Table A.4.: List of all improper torsion force-constants k_ϕ used in the reparametrized force field. Units are in kcal mol^{-1} rad^{-2}. Impropers describing the porphyrin substituents out-of-plane wagging motion are adapted from the parametrization of Kuczera et al. (227) All improper force-constants restricting the heme-doming motion of the dx configuration are set to 0.0 kcal mol^{-1} rad^{-2}.

Configuration	mode	experimental freq. [cm^{-1}]	calculated freq. [cm^{-1}]
dx	$\delta_{\mathrm{N_{porph}-Fe-N_{Im}}}$	79, 110	83, 119
	$\nu_{\mathrm{Fe-N_{Im}}}$	215, 248	236, 244
	$\nu_{\mathrm{Fe-N_{porph}}}$	288, 293	290, 296
cx	$\delta_{\mathrm{N_{porph}-Fe-N_{Im}}}$	64, 127	60, 129
	$\nu_{\mathrm{Fe-N_{Im}}}$	222	223
	$\nu_{\mathrm{Fe-N_{porph}}}$	320, 338	319, 323
	$\delta_{\mathrm{Fe-C-O}}$	586	586
	$\nu_{\mathrm{Fe-C}}$	507	505
	$\nu_{\mathrm{C-O}}$	1969	1969
bis-His	$\nu_{\mathrm{N_{Im}-Fe-N_{Im}}}$ (sym)	200	199
	$\nu_{\mathrm{N_{Im}-Fe-N_{Im}}}$ (asym)	385	387

Table A.5.: Frequencies showing the strongest iron stretching- (ν) and bending-mode (δ) contributions in each heme configuration. Experimental frequencies of the dx configuration derive from NRVS data of 2-methylimidazole tetraphenylporphinato iron(II) ([Fe(TPP)(2-MeImH)]) (240), of the cx state from NRVS data of [Fe(TPP)(1-MeImH)CO](241) and the bis-His configuration from resonance Raman- and infrared data of a bis(imidazole)(octaethylporphyrin)iron(II) complex(242). The calculated frequencies were obtained by iteratively fitting normal-mode force constants involving the iron with the heme model systems shown in Figure 4.3.

Atom type	**charge [e]**		
	dx	**cx**	**bis-His**
FE	0.84	0.19	1.38
NPH1, NPH2	−0.68	−0.57	−0.77
CPA	0.31	0.28	0.30
CPB	0.03	0.05	0.05
CPM	−0.22	−0.21	−0.21
C	−0.15	−0.06	−0.06
C(2)	−0.25	−0.29	−0.29
CT2	−0.28	−0.28	−0.28
CT2(2)	−0.28	−0.28	−0.28
CT3	−0.33	−0.33	−0.33
CC	0.45	0.50	0.50
OC	−0.62	−0.69	−0.69
CT2-HA	0.10	0.11	0.11
CT2(2)-HA	0.10	0.12	0.12
NR2	−0.46	−0.39	−0.52
CPH1	0.05	0.04	0.03
CPH1-HR3	0.16	0.10	0.09
CPH2	0.28	0.29	0.23
CPH2-HR1	0.20	0.14	0.25
NR1	−0.49	−0.45	−0.60
NR1-H	0.30	0.29	0.35
CM	-	0.49	-
OM	-	−0.32	-

Table A.6.: CHARMM force field charges fitted from Mulliken population analysis for each heme configuration. The atom type definition follows the nomenclature shown in Figure 4.3. If not listed, hydrogen atoms have assigned a charge of 0.09 e. Remaining hydrogens are listed with the atoms they are attached to.

configuration	$V^{\mathrm{His}}_{Morse}/V^{\mathrm{CO}}_{Morse}$			
	D_e **[kcal/mol]**	r_0 **[Å]**	β **[Å]**$^{-1}$	
bis-His	23.68	2.079	1.985	
cx	74.50	1.22	1.798	
	$V_{LJ_{7-5}}$			
	$\epsilon_{\mathrm{FE-CM}}$	$\sigma_{\mathrm{FE-CM}}$	$\epsilon_{\mathrm{FE-NR2}}$	$\sigma_{\mathrm{FE-NR2}}$
dx	0.059 kcal/mol	4.536 Å	0.628 kcal/mol	3.500 Å

Table A.7.: ARMD specific force field parameters for the dissociable bonds in the biphasic rebinding process of Ngb$_{\mathrm{P}}$: V^{CO}_{Morse} and V^{CO}_{Morse} correspond to the Morse potentials describing the dissociable FE-NR2 and FE-CM bonds in the bis-His and cx configurations. $V_{LJ_{7-5}}$ represents the non-bonding Van-der-Waals potential defined as a LJ 7-5 function which is present between the same atom pairs on the pentacoordinated dx state.

Curriculum vitae

Personal Data

Address:	Stephan Lutz Birkenstrasse 2 CH-4323 Wallbach
Email:	Stephan.Lutz@unibas.ch
Nationality:	Swiss, Canadian

Education

1/2007 – 2/2011	**PhD Studies, Department of Chemistry, University of Basel**[*] Thesis: Computational Investigation of Conformational and Configurational Substates in Neuroglobin Supervisor: Prof. M. Meuwly, University of Basel External Examiner: Prof. G. U. Nienhaus, Karlsruher Institut für Technologie
09/2001 – 8/2006	**Studies in Chemistry, University of Basel** Thesis: Studying Reactive Processes with Classical Force Fields: Application to Chemically and Biologically Relevant Problems Supervision: Prof. M. Meuwly, University of Basel
08/1997 – 12/2000	**Gymnasium (Typus C) in Muttenz / BL**
1988 – 1996	**Primary Formation in Wallbach and Möhlin / AG**

---

[*] During the time of my PhD thesis I attended the lectures of Prof. H. Huber, Prof. M. Meuwly, Prof. S. Borner, and Prof. E. C. Constable.

Publications

- **Reactive Force Fields for Proton Transfer Dynamics**
  S. Lammers, S. Lutz, M. Meuwly
  J. Comp. Chem. **2008**, *29*: 1048 – 1063

- **Ligand Migration between Internal Docking Sites in Photodissociated Carbonmonoxy Neuroglobin**
  S. Lutz, K. Nienhaus, G. U. Nienhaus, M. Meuwly
  J. Phys. Chem. B **2009**, *113*:15334 – 15343

- **Structural Identification of Spectroscopic Substates in Neuroglobin**
  K. Nienhaus, S. Lutz, M. Meuwly, G. U. Nienhaus
  ChemPhysChem **2010**, *11*:119 – 129

Submitted for publication:

- **Structural Characterization of Spectroscopic Substates in Carbonmonoxy Neuroglobin**
  S. Lutz, M. Meuwly

- **Water-Assisted Proton Transfer in Ferredoxin I**
  S. Lutz, I. Tubert-Brohman, Y. Yang, M. Meuwly