Structural characterization of spectroscopic substates in carbonmonoxy neuroglobin

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

Received 16th November 2010, Accepted 28th January 2011
DOI: 10.1039/c0fd00003e

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 of the protein. Recently, two similar CO bands were identified from 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 side chain in hydrogen bonding contact to CO. Frequencies of the corresponding $His64H^+$ and $His_{\delta}64$ tautomers show characteristic contributions to the N_0 band.

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]

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

[a]*Department of Chemistry, University of Basel, Klingelbergstrasse 80, 4056 Basel, Switzerland*
[b]*Department of Chemistry, University of Basel, Klingelbergstrasse 80, 4056 Basel, Switzerland. E-mail: m.meuwly@unibas.ch*

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 of Mb 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 protein. 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 made it possible 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 side chain 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 made it possible 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]

Neuroglobin is another protein for which considerable experimental data relevant to the present discussion has become available. It is classified as a member of the vertebrate globin family of proteins and was discovered only recently in neuronal

cells of men and mice.[32–34] 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 penta-coordinated 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 the 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), features 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 relationship between spectroscopy and the underlying 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 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 study 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 $\approx$1.2 Å compared to the X-ray structure. In M the RMSD stabilized at $\approx$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. This approach was successful in previous studies on bound and photodissociated MbCO.[21,49] Subsequently, the fundamental vibration frequency was obtained from the one-dimensional Schrödinger equation: The CO bond energy was sampled at eleven 0.025 Å intervals of r, around r_e, the CO equilibrium distance and the energies $V(r)$ were fitted to a Morse function

$$V(r) = D_e(1 - \exp(-\beta(r - r_e))) \quad (1)$$

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 are referred to as DFT/QM and allow quantitative computation of CO-vibrational frequency shifts induced by the conformational and electrostatic environment of the protein.

3. Results

This section separately presents the theoretical spectra obtained from the DFT/QM calculations for M and N. Then the identified band positions are analyzed in structural terms.

3.1 Computed 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

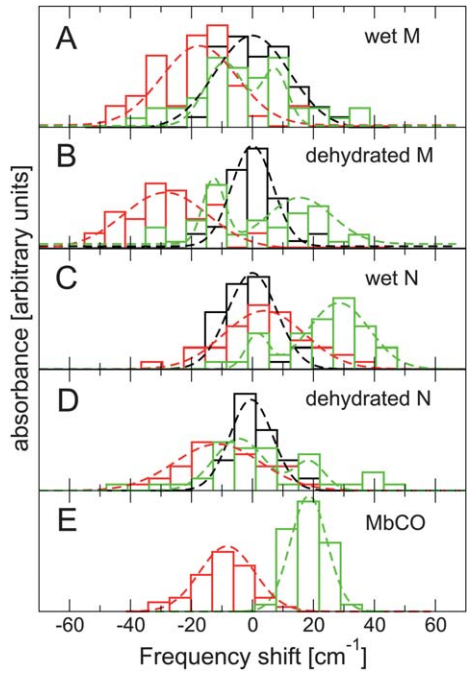

Fig. 1 Theoretical IR CO stretching spectra represented as histograms and calculated as described in the Methods section. The dashed lines are Gaussian fits to guide the eye. A: Relative frequency shifts to the calculated His64H$^+$ (black) band maximum in mNgbCO (M) for the His$_\varepsilon$64 (red) and His$_\delta$64 (green) tautomers. B: Same as A but all water molecules within the active site are removed from the snapshots prior to the frequency calculation. C: Relative frequency shifts to the calculated His64H$^+$ (black) band maximum in hNgbCO (N) for the His$_\varepsilon$64 (red) and His$_\delta$64 (green) tautomers. D: Same as C but all water molecules within the active site removed from the snapshots prior to the frequency calculation. E: Relative frequency shifts to the calculated His64H$^+$ band position in wt MbCO for the His$_\varepsilon$64 (red) and His$_\delta$64 (green) MbCO tautomer taken from Ref. 21.

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

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

CO band shift relative to	Protein					
the $His64H^+$ peak	$His_{\varepsilon}64$ 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, +7	+4	+2, +29	−12	+16
dry active site [cm^{-1}]	−29	−13, +16	−11	−4, +18	—	—

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. 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.* Fig. 1C compared to only 56 waters in Fig. 1B *vs.* Fig. 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 smaller compared to the corresponding ones in M (Table 1).

3.2 Structural assignments of the peak positions

Double protonation of His64, which introduces a positive charge on this side chain, is known to render this residue more hydrophilic.[28] The positive charge on $His64H^+$ is expected to expose its side chain preferably to the solvent.[7,21] The present MD simulations show this for both Ngb variants, M and N. 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 side chain 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. 3A displays the 40 frequency shifts for "dry" $His_{\varepsilon}64$ M (diamonds) and N (circles) as a function of the H 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} $Å^{-1}$ with a standard deviation of $\sigma_{b,M} = 3.6$ cm^{-1} $Å^{-1}$ where the subscript text refers to the mutant M. For N the parameters are $b_N = 22.7$ cm^{-1} $Å^{-1}$ and $\sigma_{b,N} = 2.8$ cm^{-1} $Å^{-1}$. Therefore, the CO stretching frequency shifts equally strongly with varying $His64\text{-}NH_{\varepsilon}$ to CO

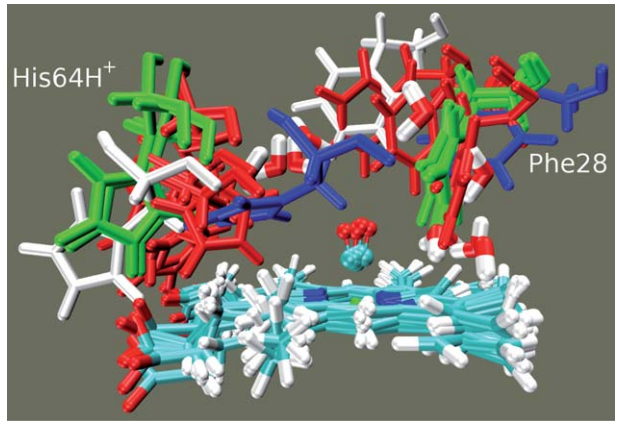

Fig. 2 Representative His64H$^+$ M structures corresponding to low-frequency (white and blue), ≈ 0 cm^{-1} shift frequency (red) and high-frequency (green) CO stretching vibrations. Eleven water molecules within a radius of 5 Å around CO are shown additionally.

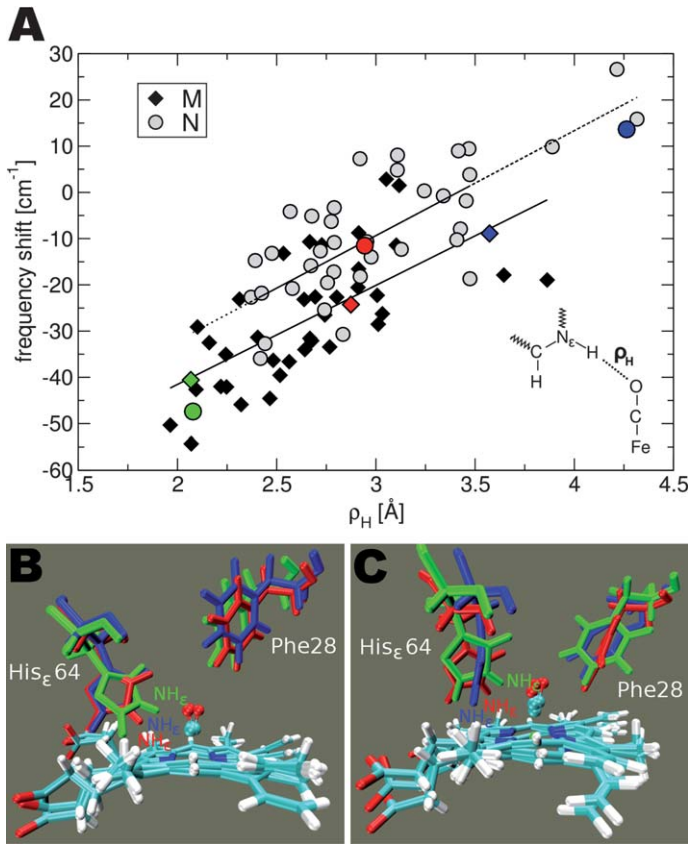

Fig. 3 A: CO-stretch frequency shift distribution of all "dry" His$_\varepsilon$64 M (diamonds) and N (circles) snapshots as a function of the His64-N$_\varepsilon$H to CO oxygen distance ρ_H. The solid (M) and dashed (N) lines represent linear regressions. B: Representative His$_\varepsilon$64 M structures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

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 (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 side chain 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 $N_{\varepsilon,His64}$ to O_{CO} distance ρ_N. In the figure, the structures are also grouped into those having either the N_ε (triangles) or the N_δ (squares) of His64 closer to the O_{CO}. This separation corresponds to clustering the data into two rotameric forms of the $His_\delta 64$ side chain 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} $Å^{-1}$ and $b_N = -24.6$ cm^{-1} $Å^{-1}$ with standard errors $\sigma_{b,M} = \sigma_{b,N} = 2.4$ cm^{-1} $Å^{-1}$. Therefore the frequency shift is more sensitive to ρ_N for the native protein than for the 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 $His_\delta 64$ side chain 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 side chain 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 Phe28 is closer to the bound CO while in the low frequency structures (green in Fig. 4B and 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 spectroscopic experiments.[10,11] Furthermore, the results are also discussed in relation to MbCO. The spectroscopic studies were carried out with two sequences: native hNgb (called N in the following) which differs from the crystallized sequence by the presence of a Cys46–Cys55 disulfide bond, and a mutated sequence (M) in which the two cysteines are replaced by serines.[10,11]

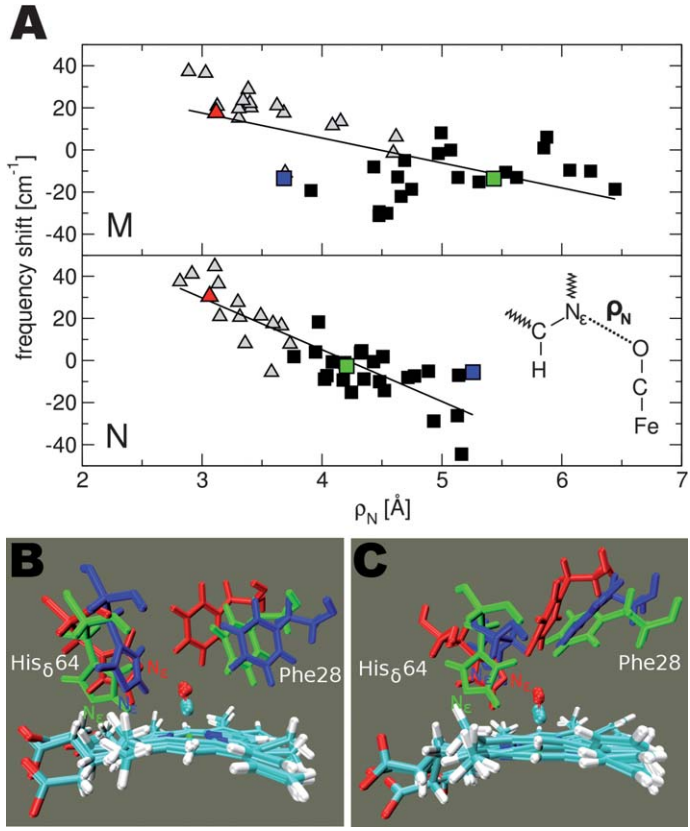

Fig. 4 A: CO-stretch frequency shift distribution of all "dry" His$_\delta$64 M (top graph) and N (bottom graph) snapshots as a function of the His64-N$_\varepsilon$ to CO oxygen distance ρ_N. Frequencies marked by triangles correspond to structures with His64N$_\varepsilon$ closer to the oxygen of CO than His64N$_\delta$. Squares correspond to the reverse situation. The solid lines are linear regressions for each distribution. B: Representative structures from different regions of the frequency distribution of His$_\delta$64 M in Fig. A (color-coded). C: Same as B but for His$_\delta$64 N.

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}) 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$_\varepsilon$H 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$64 protonation states in Table 1 qualitatively agree with the experimental N_0–N_3 band splitting of ≈ -28 (M) and -36 cm^{-1} (N), respectively. This corroborates the assignment of the two experimental bands to an open His64H$^+$ and a closed His$_\varepsilon$64 tautomeric substate. Quantitatively the computed M and N band splittings somewhat underestimate the experimental band separation.

Table 2 Collection of experimental CO stretching modes relative to the A_0 band wt MbCO. 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])

Protein/band	ν_{CO}, cm^{-1}
wt MbCO A_0	0
wt MbCO A_1	−16
wt MbCO A_3	−32
H64V MbCO	+3
L29F MbCO	−33
N N_0	+4
M N_0'	−11
M N_0''	0
N N_3	−32
M N_3	−32

The "dry" form of M (Fig. 1B) yields a $His64H^+$ to $His_\varepsilon64$ band splitting of −29 cm^{-1} whereas in "dry" N it is reduced to −11 cm^{-1} (Fig. 1D) and becomes even slightly positive for "wet" N (Fig. 1C). This can be caused by the more pronounced solvation of N compared to M. An analysis of the protein cavities near the CD-loop (see Fig. 5) shows the average volume in M is 18 $Å^3$ whereas in N it is 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 water molecules between the CD-loop and 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, whereas in N two enter and one leaves. The more pronounced presence of water in N may also be an explanation for the experimentally observed lowering of O_2 affinity when the disulfide bond is removed.[52] Without the available space, not only water molecules can less easily access the active site but also an exogenous ligand.

The computed spectra for the native and the mutant protein in their $His64H^+$ and $His_\delta64$ states show that they partly overlap. This is particularly pronounced for the "dry" states (Fig. 1). As the $His_\delta64$ spectra show a bimodal distribution throughout, it is of interest to explore whether conformational dynamics between the two states can be observed and on which time scale this occurs. An additional 60 ns MD simulation of $His_\delta64$ N sampled around fifty transitions between structures with N_δ (green structures in Fig. 4B) or N_ε closer to the CO ligand (red structures in Fig. 4B), respectively. The two states correspond to a C_α-C_β-C_γ-$C_{\delta2}$ dihedral angle of $\approx +100°$ and $\approx -100°$, respectively. Fig. 6 shows that the rotamer with N_δ closer to the CO ligand is preferred over that with N_ε closer to CO by a factor of ≈ 10 which corresponds to an equilibrium constant $K = 0.1$ for transitions between the two conformations. From

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

a differential stabilization energy $\Delta G^\ominus = 1.3$ kcal mol^{-1} is obtained, where R is the gas constant and $T = 300$ K is the simulation temperature. The average residence time of the less stable conformation (red snapshots in Fig. 4B) is around 600 ps. Such a time scale for interconversion between two different but structurally related states N_0' and N_0'' of the N_0 peak, is in qualitative agreement with observing a static

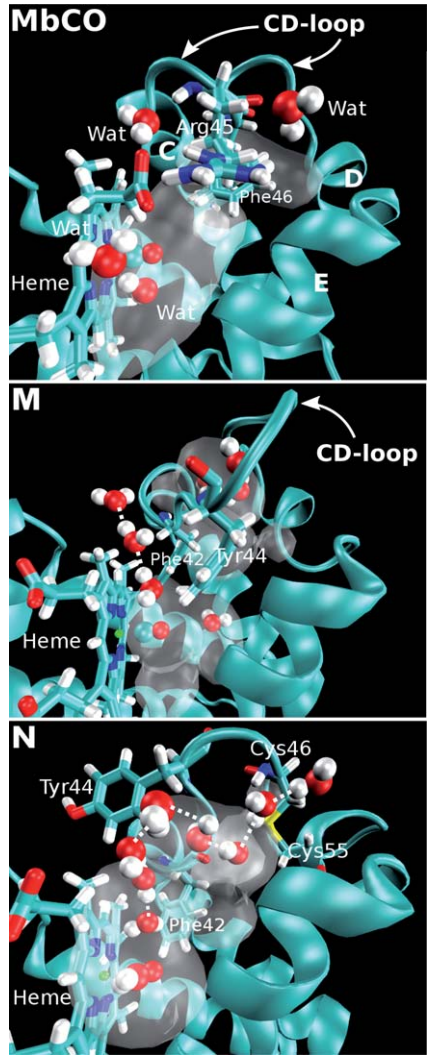

Fig. 5 Combined distal site cavity network extracted from ten individual snapshots along an equilibrated 500 ps MD simulations for His64H$^+$ MbCO, M and N at 300 K. The transparent white volume maps correspond to 10% population density of the cavity. Pockets were identified by the program SURFNET.[57] 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 a hydrogen bond network from the bulk into the heme active site in going from MbCO over M to N.

component in the frequency–frequency correlation functions (FFCFs) indicative of sufficiently slow (several 100 ps) structural dynamics of the protein.[10] Given that spectroscopically the His64H$^+$ and His$_\delta$64 states partly overlap – with the His$_\delta$64 contribution occluded by the more prevalent His64H$^+$ state – and that the two components of the His$_\delta$64 state interconvert on a time scale of several hundred picoseconds in the simulations, supports the existence of both, His64H$^+$ and His$_\delta$64 protonation. Further evidence for His$_\delta$64 protonation comes from earlier IR experiments.[51] They observed that the N_0 substate in NgbCO at neutral pH (only a smaller amount of His64 is doubly protonated (His64H$^+$) and hence in the open

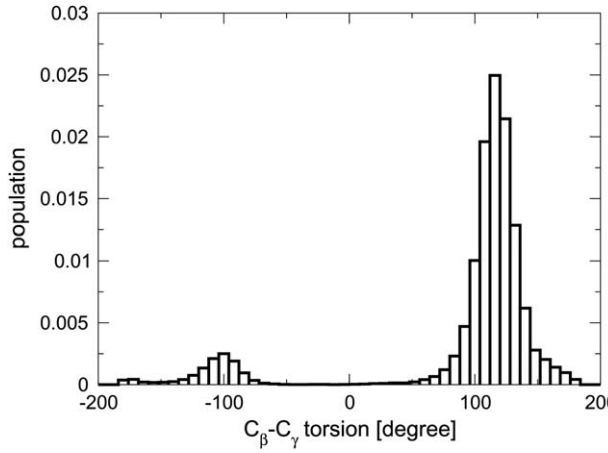

Fig. 6 Normalized population distribution of the C_β–C_γ torsional angle of residue $His_\delta 64$ for 60 ns of MD simulation. Structures at $\approx -100°$ reflect the rotameric conformation shown as a red snapshot in Fig. 4B and angles of $\approx 100°$ represent the green structures in the same Figure.

conformation) shows a significantly larger contribution (40%) compared to the corresponding A_0 substate in MbCO. Therefore, the additional N_0 contribution at pH 7 most likely derives from the singly protonated His64 tautomer (*e.g.* $His_\delta 64$). Finally, as will be discussed further below, the existence and population of $His_\delta 64$ is also required in Ngb for functional reasons.

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 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 ≈ 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 (1B and D) and shows one single peak. Contrary to the spectra of $His_\delta 64$ M and N, the corresponding frequency in MbCO is blueshifted 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 $His_\delta64$ are blueshifted relative to $His64H^+$ (1*E*) but experimentally no IR bands were observed with energies higher than the A_0 band. Thus, the presence of a $His_\delta64$ configuration was excluded in recent simulations.[21] This is also consistent with $His_\varepsilon64$ in Mb found by experiment and simulations.[5,54] Two possible rotamers of $His_\varepsilon64$ 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 side chain of His64 oriented with the N_ε–H pointing towards the iron-bound CO and the unprotonated N_δ to the solvent exposed side of the protein. This His64 orientation corresponds to the majority of sampled $His_\varepsilon64$ 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 $His_\delta64$ 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_δ (N_δH) in $His_\delta64$ NgbCO is not suitable to establish a stable H-bond to a heme-bound CO similar to the N_εH in $His_\varepsilon64$ NgbCO. Therefore the transition barrier between the rotamers is expected to be smaller in $His_\delta64$ NgbCO compared to the $His_\varepsilon64$ NgbCO. Surprisingly, none of the previous experimental NgbCO FTIR studies suggested contributions from either one or both of the $His_\delta64$ NgbCO rotamers. Nevertheless, the present simulations of the $His_\delta64$ tautomers in M and N found both rotameric forms of $His_\delta64$ and assigned the conformation with the N_ε facing towards the solvent side of the protein to the low-energy peak in Fig. 1A to D and one with N_ε 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 $His_\delta64$ 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 $His_\delta64$ M and N peak (red triangles in Fig. 4A). The observed band-splitting between N_0' and N_0'' is only 11 cm^{-1} whereas the computed splitting is $\approx$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 (see discussion above). Similarly, in MbCO it was suggested that small contributions of the $His_\varepsilon64$ rotamer (responsible for the A_1 band) contribute to the A_0 band.[17]

5. Conclusion

Increased structural dynamics and a larger active site cavity makes the structural interpretation of spectroscopic states in NgbCO considerably more challenging compared to MbCO. Here, we used MD simulations in conjunction with QM/MM calculations to shed more light into the structural interpretation of spectroscopic experiments on NgbCO. Previously, Ngb spectra were analyzed indirectly by comparing with and capitalizing on analogies with wt and mutant Myoglobins.[10,11] 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 $His_\varepsilon64$ conformation with a strong $His_\varepsilon64$ N_ε–H to CO interaction. The N_0 band (with two substates N_0' and N_0'') is 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, the His$_\delta$64 tautomer is expected for NgbCO to allow formation of the bis-His coordinated state.[38] The calculations for His$_\delta$64 M and N revealed a band splitting into two specific rotamers which were not found experimentally so far. The low-energy peak in the computed spectrum is close to the IR-band maximum computed for His64H$^+$ 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 second His64 rotamer of His$_\delta$64 NgbCO was always blueshifted relative to His64H$^+$ and is assigned to the N_0'' substate.

Finally, the simulations find considerably reduced conformational dynamics in N compared to M due to the disulfide bond in N. Experimentally, the rate of structural fluctuations in wild type (with disulfide bond), reduced wild type (without disulfide bond) and a serine mutant has been found to be very different on the time scale of several ten picoseconds by following the spectroscopy of the N_0 and the N_3 states.[11] Two independent 1 ns MD simulations of fully solvated M and N at 300 K show an average fluctuation of RMSD = 0.86 Å of the backbone atoms in the CD-loop (residues 44 to 58) for N compared to 1.36 Å in M whereas the overall protein backbone fluctuations in N and M differ considerably less (RMSD difference 0.16 Å). This, however, does not provide any information on the time scale of the fluctuations. A larger distal site cavity for M and especially N makes the active site more accessible for bulk water molecules which may contribute to a slowing down of the relaxational dynamics through vibrational coupling. However, to fully characterize such effects a complete analysis of the lineshapes[55] is required which is not possible with the present approach due to the large QM region employed and outside the scope of the present work. Alternatively, improved force fields – similar to those developed for unligated CO[12,14] – capable of capturing details of the electrostatic interactions in the active site and the coupling between different vibrational degrees of freedom may be envisaged.[56]

In summary, atomistic simulations have provided structural assignments of the spectroscopic features N_0 (open His64H$^+$) and N_3 (closed His$_\varepsilon$64) and support the presence of a His$_\delta$64 tautomer which is involved in the substates N_0' and N_0''. The dynamics between N_0' and N_0'' (N_δ or N_ε closer to the CO ligand, respectively) occurs on a time scale of several hundred picoseconds which corresponds to experimentally observed conformational dynamics of the N_0 peak from the static component of the FFCF. These results give independent insight into the relationship between conformational dynamics and spectroscopy in NgbCO and provides the basis for further understanding. With the approach chosen here it is possible to corroborate previous but indirectly inferred assignments and to extend the characterization of the active site dynamics in NgbCO.

Acknowledgements

This work is supported by the Swiss National Science Foundation (project 200021-117810 and the NCCR MUST). The authors thank Dr. H. Ishikawa and Prof. M. D. Fayer for helpful correspondence and discussions.

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. Natl. Acad. Sci. U. S. A.*, 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. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 16116–16121.
11 H. Ishikawa, S. Kim, K. Kwak, K. Wakasugi and M. D. Fayer, *Proc. Natl. Acad. Sci. U. S. A.*, 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. Natl. Acad. Sci. U. S. A.*, 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. Natl. Acad. Sci. U. S. A.*, 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, pp. 1598–1603.
26 S. Mishra and M. Meuwly, *Biophys. J.*, 2009, **96**, 2105–2118.
27 S. Mishra and M. Meuwly, *Biophys. J.*, 2010, **99**, 3969–3978.
28 F. Yang and G. N. George, *J. Mol. Biol.*, 1996, **256**, 762–774.
29 P. W. Fenimore, H. Frauenfelder, B. H. McMahon and F. G. Parak, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 16047–16051.
30 H. Frauenfelder, P. W. Fenimore, G. Chen and B. H. McMahon, *Proc. Natl. Acad. Sci. U. S. A.*, 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 Rep.*, 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. Comput. Chem.*, 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 A. Vallone, K. Nienhaus, A. Matthes, M. Brunori and G. U. Nienhaus, *Proc. Natl. Acad. Sci. U. S. A.*, 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, *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 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.
53 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 S. Mukamel, *Annu. Rev. Phys. Chem.*, 2000, **51**, 691–729.
56 M. Devereux and M. Meuwly, *J. Chem. Inf. Model.*, 2010, **50**, 349–357.
57 R. A. Laskowski, *J. Mol. Graphics*, 1995, **13**, 323–330.

