Stationary and time-dependent carbon monoxide stretching mode features in carboxy Myoglobin: a theoretical-computational reappraisal.

Andrea Amadei[1] *and Massimiliano Aschi[2] †

December 10, 2021

[1] Dipartimento di Scienze e Tecnologie Chimiche, Università di Roma "Tor Vergata", via della Ricerca Scientifica 1, 00133 Roma, Italia.
E-mail: andrea.amadei@uniroma2.it

[2]Dipartimento di Scienze Fisiche e Chimiche Università de l'Aquila, via Vetoio (Coppito 1), 67010 l'Aquila, Italia.

*author to whom the correspondence should be addressed
†author to whom the correspondence should be addressed

E-mail: massimiliano.aschi@univaq.it

Abstract

The stationary and time-dependent infrared spectrum (IR) of the CO stretching mode (ν_{CO}) in Carboxymyoglobin (MbCO), a long-standing problem of biophysical chemistry, has been modelled through a theoretical-computational method specifically designed for simulating quantum observables in complex atomic-molecular systems and based on a combined application of long timescale Molecular Dynamics simulations and Quantum-Chemical calculations. This study is basically focused on two aspects: (i) the origin of the stationary IR sub-states (termed as A_0, A_1 and A_3) and (ii) the modelling and the interpretation of the ν_{CO} energy relaxation. The results, strengthened by a more than satisfactory agreement with the experimental data, concisely indicate that: (i) the conformational His64-FeCO relevant sub-states, i.e. characterized by the formation-disruption of the H-bond between the above moieties, are the main responsible of the presence of two distinct and well separated (A_0 and A_1/A_3) spectroscopic regions; (ii) the characteristic bimodal shape of the A_1/A_3 spectral region, according to our model, is the result of the fluctuation of the electric field pattern as provided by the protein-solvent framework perturbing the bound His64-CO-Heme complex; (iii) the electric field pattern, in conjunction with the relatively high density of MbCO vibrational states, is also the main determinant of the ν_{CO}

energy relaxation, characterizing its kinetic efficiency.

1 Introduction

The binding of carbon monoxide (CO), and more in general of small molecules, to the sixth coordination site of iron(II) in the heme prosthetic group of Myoglobin(1; 2; 3; 4; 5) (both horse-heart and sperm-whale) has received a great deal of attention from basic research. A huge number of fundamental studies, focused on a large variety of structural, spectroscopic, thermodynamic and kinetic aspects has been indeed devoted to Carboxymyoglobin (MbCO) and Oxy-Myoglobin (MbO2), as well as their mutants, both experimentally (6; 7; 8; 9; 10; 11; 12; 13; 14; 15; 16; 17; 18; 19; 20; 21; 22; 23; 24) and computationally. (25; 26; 27; 28; 29; 30; 31; 32; 33; 34; 35; 36; 37; 38; 39) In this context, it was particularly interesting the stationary and time-dependent Infra Red (IR) spectral behavior of the heme-bound CO stretching frequencies, i.e. both the one associated to the $Fe - C$ mode (hereafter ν_{Fe-C}) and the one associated to $C - O$ mode (hereafter ν_{CO}), whose spectral signal interpretation turned out to be crucial for understanding the mechanisms underlying the binding differences between CO and O_2 to iron(II). (40; 41; 42; 43; 44; 45; 46; 47) Specifically, the IR signal of ν_{CO} in MbCO is characterized by three different sub-states, denoted as A_0

($\simeq$ 1965cm^{-1}), A_1 ($\simeq$ 1945cm^{-1}) and A_3 ($\simeq$ 1934cm^{-1})(48) also showing other relevant features such as different CO re-binding rates(6), different dephasing rates(49) and different sensitivity to pH (54). A fourth sub-state, usually indicated as Ax, appears (55) as scarcely relevant in physiological conditions. In wild-type (wt) MbCO in aqueous solution the dominant peak is represented by A_1 with A_0 showing a spectral intensity more than 10 times lower. The third peak, A_3, appears as a shoulder in the red tail of A_1 and it is not always well resolved (51; 52; 53). The situation is almost perfectly reversed in the case of the T67R/S92D double mutant(60) in which the A_3 peak becomes the dominant one with A_1 representing a blue-shifted shoulder. A number of experimental and computational studies have been addressed in the last two decades for interpreting these results. (41; 56; 57; 7; 58; 59; 60; 61; 62; 63; 64; 50) It is nowadays widely accepted that A_0 and A_1 sub-states are associated to the His64 open and closed conformations driven by an hydrogen bond (H-bond) between the ϵ protonated His64 imidazole ring and FeCO moiety.(65; 64; 8; 44; 72; 48) Such an interaction, although weaker than that observed in MbO2 (67; 68; 44; 69; 34; 64; 66) is evidently strong enough not only to justify the well resolved A_1 and A_0 spectral maxima but also to explain their rather large intensity ratio in wt MbCO. On the other hand more debated is the origin of the A_3 sub-state initially ascribed to a modulation of ν_{CO} frequency as a function of the Fe-

C-O angle. In the last years this hypothesis has been discarded (70; 64), and other alternative, not mutually excluding, explanations have been proposed ranging from (i) changes in the specific local interactions between FeCO and the His64 essentially due to different conformational sub-states or ring rotation (71; 73; 8; 74; 75; 60) (ii) presence of different sub-states in regions next to the distal pocket characterized by differently sized cavities(76) or even (iii) local field-effect induced by other charged aminoacids such as Arg45 in the heme-pocket. (77) The effect of the mechanical and structural features of both the protein framework and the solvent has been also hypothesized by several experimental and computational studies to be important in the modulation of the vibrational lifetime of ν_{CO}. Fayer and coworkers in their seminal paper of 1994(14) and more recently in 2010 (60) and 2011(78), as well as Kwak and Cho(47) in 2018 have showed that the vibrational relaxation rates of excited ν_{CO} is strongly modulated by different protein conformational sub-states, different mutations as well as the presence of solvent molecules in the vicinity of the prosthetic group. As a part of our ongoing interest in the theoretical-computational study of vibrational stationary spectra(79) and vibrational energy relaxation (80) (VER) of chromophores embedded in complex atomic-molecular environments we decided to reappraise the modelling of the above mentioned spectral features of ν_{CO} in MbCO. Although most of the subtle effects underlying the experimental observations have been elu-

cidated by the already cited theoretical-computational studies, we feel that the peculiar character of our computational approach, somewhat complementary to the most popular mixed Quantum Mechanics/Molecular Mechanics (QM/MM) methods (81) and comparably accurate at lower computational cost,(82) might provide an additional contribution to a deeper understanding of such an important issue. In particular, our primary purpose is to deepen the analysis of the plausible role of the conformational fluctuations, even on relatively large space-time scale, of the entire atomic-molecular environment on the stationary and time-dependent ν_{CO} spectral features. Moreover in the present study we also wish to show how it is nowadays possible to address a relatively complicated issue - such as VER - in a very complex atomistic environment without departing from the quantum description required by the intrinsic nature of the vibrational mode (modes) we are focussing on. This study is organized as follows. After a concise methodological outline in which we report both the computational strategy and the related technical details we describe the results concerning the stationary IR spectrum, also necessary to assess the quality of the method, and subsequently the modelling of ν_{CO} relaxation kinetics.

2 Computational strategy and details

2.1 Stationary IR spectrum

We evaluated the IR stationary spectrum by using a theoretical-computational procedure, hereafter termed as IR-Perturbed Matrix Method (IR-PMM) whose details and basic approximations can be found in the literature (84; 85; 86; 87; 88). In brief, (i) a classical Molecular Dynamics (MD) simulation is first produced with MbCO, water and counterions all described by the same (in this case empirical) force-field. (ii) A sub-portion of the whole simulated system is selected as Quantum-Centre (QC), i.e. the part of the system whose quantum (perturbed) observables we want to model. (iii) The electric field and potential exerted by the protein framework, solvent and counterions on the QC center of mass is then evaluated at each MD frame (QC-based expansion (85)). (iv) On the basis of the selected QC a single structure, or a number of structures in the case of non-rigid QC (89), is considered for calculating the unperturbed quantum observables with standard electronic structure methods. Note that this latter step is intimately connected to step (i) as below shown. (v) The perturbed QC frequency is calculated at each MD frame by using the instantaneous electric field and potential of the point (iii) and the unperturbed observables of point (iv); in practice, as also further explained below, at each MD frame the perturbed electronic energy along the

unperturbed vibrational mode/modes of interest is calculated thus allowing to estimate the corresponding instantaneous (perturbed) energy minimum and frequency. Note that the possible mixing of the unperturbed vibrational modes induced by the perturbation can be in general disregarded unless when considering extremely high perturbation, by far exceeding the typical atomic molecular electric field(90) (vi) The perturbed spectrum is finally modelled by using the collection of the QC perturbed frequencies, whose number is obviously equal to the number of MD frames, as more thoroughly explained below.

Concerning the steps (i) and (iv) two independent simulations of wt MbCO, termed as MD-free and MD-constrained simulations have been carried out for mimicking the effects on the ν_{CO} perturbed frequency of the open and the closed His64 conformation, respectively. As reported in Figure 1, the two simulation setups (free and constrained for both the wild type and double mutant T67R/S92D) have implied the selection of two different QCs - termed as QC-small and QC-large - and, consequently, two different structures (highlighted with sticks in the same Figure and with methyl groups used as capping-groups) were used for the calculation of the unperturbed properties. In the case of the MD-constrained simulation we decided to apply an ideal constraint for the hydrogen-bond length and a restraint for the corresponding bond angles to mimic the His-closed conformation involving a

stable hydrogen bond. This means that along such a simulation the distance between His64/Nϵ and FeCO/oxygen was constrained and the His64/Nϵ-H-O and H-O-C angle fluctuations were relevantly limited by the added restraining term. Moreover, to disregard the unrealistically large His64-FeCO complex internal fluctuations as provided by the classical MD simulation and resulting in an excessively broad spectrum of the His-closed state (see Figure 3), we always considered a rigid QC-large thus defined by a single reference structure for calculating the unperturbed properties in the step (iv). Such a choice is well motivated by the relatively high frequencies of most of the QC-large modes involving His64-CO relative rototranslations (larger than 1000 cm^{-1}), thus implying that at room temperature (about 300 K) no classical vibrations along such modes can be present (i.e. fully quantum mechanical modes) and then a basically rigid QC-large should be used when modeling the stable His64-FeCO complex. Additional information/analysis of the above simulations are reported in the Supplementary Information (S.I. Section S2). In the MD-free simulation no constraint and/or restraint for His64-FeCO interaction were used and the corresponding QC-small (not including His64) was also treated as a rigid QC defined by a single reference structure. Note that although a rather low internal mobility is observed in the QC in all the simulations and the dihedral angle for the rotation around the hydrogen bond is basically unable to affect the IR spectrum (see S.I. Section S5),

the small internal deformations due to the hydrogen-bond angles and the other internal degrees of freedom are largely responsible for the significant and unrealistic broadening of the IR spectrum of the His-closed conformation when using the QC-small in PMM calculations (see Fig S10 Section 8 in the S.I.). The same simulation setups of the wt MbCO were utilized for modelling the stationary spectrum of the double mutant (hereafter simply termed as DM) T67R/S92D. We refer to these simulations as MD-free-DM and MD-constrained-DM. In this respect we wish to underline that in the case of the DM system we limited our attention to the A_1/A_3 region, hence, only considering the His-closed conformation.

In the case of QC-small we performed an unperturbed full-optimization which, as expected, resulted in a perfectly linear Fe-CO angle.(91) The effect of the Fe-CO angle on the stationary IR spectral features has been extensively debated in the past (see for example reference (64)) and, hence, re-addressed in the present study as shown more in detail in the S.I. (Section S3). However its role revealed practically negligible (see S.I. Section S4) and for this reason hereafter disregarded by our analysis/discussion.

Concerning the QC-large unperturbed geometry we carried out a constrained optimization starting from one configuration (including heme, CO, His64-side-chain and His93-side-chain as reported in Figure 1) extracted from the closed His64 sub-ensemble of the MD-free simulation, corresponding to

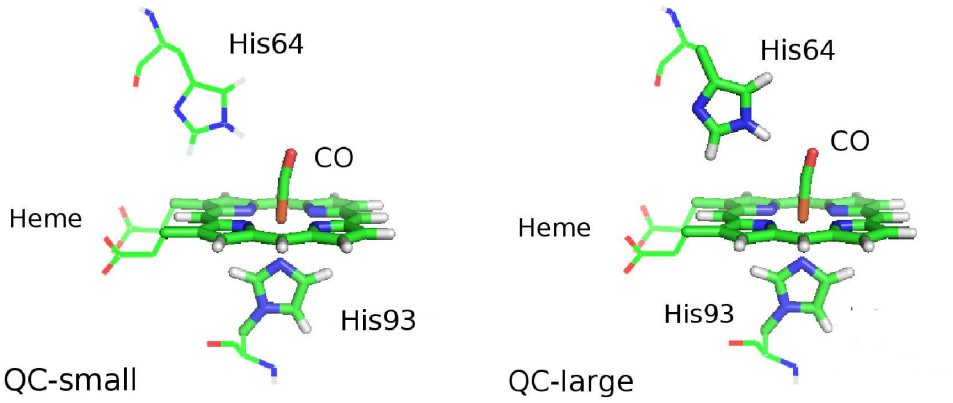

Figure 1: Schematic view of the QCs used for the unperturbed calculations (sticks) for the MD-free (left panel) and MD-constrained (right panel) simulations. The parts reported with thinner lines (wireframe) are not included in the unperturbed calculations, in which capping hydrogen atoms have been used, and included with the rest of the protein and the solvent in the perturbing environment.

the frequency of the maximum spectral signal (ν_{max}) as obtained from IR-PMM calculations (see Results section). The QCs (both small and large) unperturbed harmonic frequencies were calculated adopting the above optimized, either fully or constrained, structures. The corresponding mass-weighted Hessian matrices were then used for calculating the perturbed frequencies as explained in detail (and here only concisely outlined) in the referenced papers.(79; 86) Briefly: a number of distorted structures (21 for QC-small and 13 for QC-large) (83) were first generated by projecting the coordinates of the optimized structure along the the eigenvector corresponding to the ν_{CO}. For each of the above distorted structures we then calculated the vertical electronic-energies and the electric dipole matrix to be used for constructing, at each MD frame, the perturbed electronic Hamiltonian ($\widetilde{H}$) matrices (hence 21 $\widetilde{H}$ for QC-small and 13 $\widetilde{H}$ for QC-large)(84)

$$\widetilde{H} \simeq \widetilde{H}^0 + \widetilde{I} q_T \mathcal{V} + \widetilde{Z}_1 + \Delta V \widetilde{I} \tag{1}$$

$$[\widetilde{Z}_1]_{j,j'} = -\mathbf{E} \cdot \langle \Phi_j^0 | \widehat{\boldsymbol{\mu}} | \Phi_{j'}^0 \rangle \tag{2}$$

In the above expressions Φ_j^0 are the unperturbed electronic eigenstates (parametrically functions of the nuclear coordinates) furnishing the basis set of the Hamiltonian matrix, q_T and $\widehat{\boldsymbol{\mu}}$ are the QC total charge dipole operator, $\mathcal{V}$ is the electric potential and $\mathbf{E}$ is the electric potential field, both exerted by the environment on the QC centre of mass at each frame of the

simulation. Finally, ΔV includes all the other terms treated as a simple short range potential and $\widetilde{I}$ is the identity matrix. The ground and the first three vertical excited electronic states have been used at this purpose (note that the results turned out to be virtually the same also using two excited states). All the details of the above calculations are reported in the S.I. Section S1. At each MD frame, by diagonalizing all the corresponding $\widetilde{H}$ matrices, we then obtained the perturbed ground state energies for all the points along the CO-stretching mode hence providing the associated instantaneous perturbed potential energy function (i.e. we obtained an instantaneous perturbed Morse-curve) utilized, through a polinomial fitting, for calculating the instantaneous perturbed harmonic and anharmonic frequencies. From the complete ensemble of the above perturbed instantaneous frequencies, as obtained by the MD frames, in conjunction with the (unperturbed) ground-state electric dipole first derivative along the stretching mode we can evaluate at each frequency bin the corresponding transition probability and thus the complete lineshape of the infrared spectrum. Note that the intensity, the broadening, the frequency position and the lineshape of the spectrum are then all explicitly evaluated without the inclusion of any empirical or semiempirical parameter.(86) It follows that in all the Figures reported hereafter the spectral linesahapes, and not the frequency distributions, are systematically reported.

2.2 Vibrational Energy Relaxation kinetics

The VER kinetics of ν_{CO} was modelled using a recently developed theoretical-computational method(80) concisely described in the S.I. (Section S10). The method, in the same spirit of IR-PMM, exploits the perturbing electric field fluctuations acting on the same QC utilized for the equilibrium IR spectra. In particular, we limited our vibrational relaxation modeling to the His-closed conformation thus considering only the QC-large system which mimics the heme-CO His64 hydrogen bonded condition. For the application of this method we make use of (i) a large set of independent MD trajectories as obtained by using the same force field (the electronic ground state force field) of the simulations employed in IR spectra calculations, i.e no QC electronic rearrangement upon vibrational (ground to first excited state) transition; (ii) the first and mixed second derivatives (see S.I. Section S10) of the unperturbed electric dipole with respect to the mode coordinates, both obtained using the set of QC structures and type of (vacuum) QM calculations above reported for IR spectra. We clearly consider as the initial state of the kinetics the first excited state of the ν_{CO}, with the set of independent MD trajectories providing the reactive ensemble employed to reconstruct the time-dependent density operator of the vibrational states. The first and second dipole derivatives (together with the perturbing field time derivative) determine the transition to the ground state or to the other excited states,

respectively, with each vibrational state transition induced by the field fluctuation components with frequency resonant with the excitation frequency, similarly to the process involved in light induced excitation-emission. Note that such a relaxation process as occurring with the perturbing field provided by usual MD trajectories, corresponds to vibrational state transitions as occurring at fixed classical kinetic energy (i.e. neglecting any quantum-classical energy exchange). Such a condition, where the classical degrees of freedom determine the quantum state transitions via the perturbing field fluctuations but are uncoupled from the quantum state changes (i.e. the quantum state dynamics is coupled to the classical trajectory but not viceversa), is physically equivalent to a real ensemble with a work flux instantaneously removing the classical kinetic energy variations due to the quantum state transitions. Therefore, the relaxation kinetics obtained in the employed reactive ensemble (invariant free energy ensemble (80)) need to be corrected for the missing effects of the quantum-classical energy coupling, and when properly corrected the computed relaxation rates can be compared to the experimental data.

2.3 Computational details

All the MD simulations were carried out in the NVT ensemble with a timestep of 2.0 fs with the program Gromacs (92; 93) version 5.1.2. MbCO initial structure with the His64 in the closed state, taken from the literature (9)

(pdb code: 1bzr), was inserted at the center of a cubic box (367.07 nm^3) filled with 11289 water molecules (94) and one chloride ion for ensuring the electroneutrality. The density of the system (i.e. the box size) was adjusted to reproduce the average pressure obtained from a simulation of an identical box only containing the same number of water molecules at the experimental density, at 298 K and 1.0 bar. The protein and heme were described using the Gromos force-field(96) (G53a6 version) with the residues in their protonation state at neutral pH. For the heme-CO moiety we utilized the charges, adapted to the Gromos force field as reported in the literature(97) showing, specifically in the FeCO moiety, 0.6 atomic units for iron, 0.17 atomic units for carbon and -0.17 atomic units for oxygen. In the S.I. (Section S9) we re-evaluated the accuracy of such a force field in describing the FeCO-His64 interaction including H-bond potential energy. We wish to remark that in the present study we have considered the ϵ protonated His64 as the only tautomer significantly populated(44; 48) at least in the pH and temperature conditions here addressed. The temperature was kept constant using the Parrinello thermostat (98), the bond lengths were constrained using Lincs algorithm(99) and long range electrostatic interactions were computed by the Particle Mesh Ewald method(100) with 34 wave vectors in each dimension and a 4th order cubic interpolation and a cut-off of 1.1 nm was used. In the case of the MD-constrained-DM simulation the wt starting structure

was modified by performing the T67R/S92D mutations with the program Molden.(95) Both the wt simulations were extended up to 210 ns whereas for the DM simulations we have performed 50 ns trajectory. In all the cases the first 10 ns were systematically discarded. A test concerning the appropriateness of the simulations lengths was carried out by checking the convergence of the observable of interest (i.e. the stationary spectrum) as reported in the S.I. Section S6. All the ground state quantum-chemical (QM) calculations for unperturbed QC-small and QC-large, i.e. geometry optimization and Hessian matrix, were performed in the framework of Density Functional Theory (101) using the hybrid Becke3LYP functional (102) in conjunction with the 6-31G(d) atomic basis set for all the atoms and Hay and Wadt basis set and Effective Core Potential (103) for iron. Vertical excited states energies, dipoles and transition dipoles (necessary for $\widetilde{H}$ matrix) were calculated on the unperturbed minimum energy structures, on all the structures generated along the Hessian eigenvector (see S.I. section S1) and, for comparison, on the deformed-QC structures (see S.I. section S5), using Time Dependent DFT (TD-DFT) (104) adopting the same functional and basis set. In this respect it is important to note that, although well aware of possible drawbacks deriving from, e.g., the single-reference character of the TD-DFT method, (105) the relatively high number of unperturbed states (see above) requested for IR-PMM approach led us to select a relatively cheap yet sufficiently ac-

curate level of theory. We wish also to underline that possible limitations deriving from the use of ECP basis set for iron were excluded by performing an all-electron calculation on the unperturbed QC-small minimum energy which virtually produced the same vertical excitation energies within a maximum (relative) error of 5 percent and the same transition dipoles within a maximum (relative) error of 2 percent. Such a level of theory was utilized also for evaluating the first and second derivatives of electric dipoles, needed for the VER modelling, using finite differences along a grid on the space of the selected Hessian eigenvectors (see S.I. section S10). All the calculations were performed with the package Gaussian09. (106)

3 Results and Discussion.

The first goal of this study is the modelling and reconstruction of the wt MbCO ν_{CO} IR stationary spectrum in the two conformational His64 conditions, i.e. open and closed, widely accepted to correspond to the A_0 and A_1/A_3 MbCO IR sub-states. At this purpose we have initially localized such conditions by monitoring, along the MD-free simulation, the His64-heme distance as reported in Figure 2. From the Figure, and in line with literature data, (29) the His-closed state turns out to be rather stable within the first 50.0 ns of simulation. During this initial simulation time we can also observe

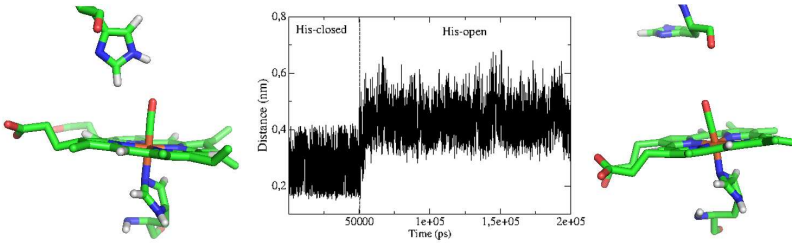

Figure 2: His64-heme minimum distance along the MD-free simulation.

temporary increases in the distance up to about 0.4 nm which, however, correspond to a loss of the H-bond due to the rotation of the His64 imidazole ring which remains in the distal pocket. In this respect by using the Gromacs criterion(93) we have found that within the first 50.0 ns of MD-free trajectory, the H-bond is considered as formed for about the 85 percent of such a simulation time. Beyond the first 50.0 ns a conformational transition can be observed (see Figure 2) corresponding to the His64 sidechain exit from the distal pocket adopting a conformation hereafter termed as His-open. In correspondance of these two conditions, i.e. using the two MD-free simulation sub-portions shown in Figure 2, we obtained the ν_{CO} spectra, i.e. the A_1/A_3 and A_0 sub-states, by means of the IR-PMM approach (see Figure 3(a)). Moreover, the ν_{CO} spectrum of wt MbCO for the His-closed state, hence the A_1/A_3 sub-state, has been further modelled using the MD-constrained/QC-large and MD-constrained-DM/QC-large simulations for the wt and for the DM, respectively (see Figures 3(b), 3(c) and 3(d)).

		A_0	A_1/A_3	
	$\nu_{gas-phase}$	ν_{max}	ν_{max}	fwhm
MD-free	2118	2098 (1997)	2077 (1977)	40
MD-constrained	2082	-	2077 (1977)	21
wt-exp(48)		1965	1945	14
MD-constrained-DM	2082	-	2067 (1968)	23
DM-exp(60)		1965	1933	21

Table 1: Becke3LYP/6-31G(d)/ECP(Fe) calculated unperturbed harmonic frequency ($\nu_{gas-phase}$) for QC-small (used in the MD-free simulation) and QC-large (used in MD-constrained and MD-constrained-DM simulations), MbCO calculated and experimental maximum absorption perturbed frequency ν_{max} for A_0 and A_1/A_3 spectral regions and full-width at half maximum (fwhm) of the A_1/A_3 region for wt and double-mutant (DM); note that fwhm for the A_0 region is not reported due to the lack of the experimental value (too low intensity of the signal) and the wt and DM ν_{max} essentially correspond to the maxima of the A_1 and A_3 bands, respectively. Note also that for the experimental A_1/A_3 fwhm we considered the overlapping A_1 and A_3 bands as a single unresolved peak reporting the corresponding maximum (for wt and DM essentially the A_1 and A_3 band maximum, respectively) as provided by the referenced (experimental) paper. The computed maximum frequency values corrected using the recommended Frequency Scale Factor(107) are also indicated in parenthesis. All the data are reported in wavenumbers.

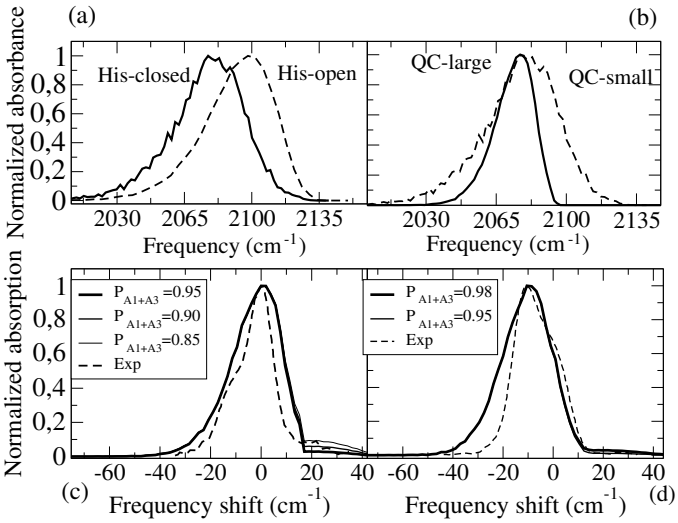

Figure 3: Panel a: Normalized wild type MbCO ν_{CO} spectra using the His-closed (solid line, QC-small, first 50.0 ns) and His-open (dashed line, QC-small, from 50 to 210 ns) portions of the MD-free trajectory (see Figure 2). Panel b: Normalized wild type MbCO ν_{CO} His-closed spectra from first 50.0 ns of the MD-free trajectory (see Figure 2) (dashed line, QC-small) and from the MD-constrained simulation (solid line, QC-large). Panel c: Superposition of the experimental(48) (dotted line) and calculated (solid line) wild type MbCO ν_{CO} spectra, using the corresponding ν_{max} (the A_1 frequency) as reference frequency, i.e. the zero of the abscissa. The computed spectrum was obtained using different relative weights (i.e probabilities P) for the His-closed (QC-large, MD-constrained trajectory) spectrum and His-open (QC-small, MD-free simulation from 50.0 ns to 210 ns) spectrum. Panel d: Comparison of the experimental(60) (dotted line) and calculated (solid line) T67R/S92D DM ν_{CO} spectra, using the wt ν_{max} (the A_1 frequency) as the reference frequency, i.e. the zero of the abscissa. The computed spectrum was obtained using different relative weights for the His-closed (QC-large, MD-constrained-DM trajectory) spectrum and His-open (QC-small, MD-free simulation from 50.0 ns to 210 ns) spectrum.

Before entering into the details of the Results we wish to highlight that when a single structure representing the QC is utilized, as in the present case, the spectral features of the perturbed chromophore mostly reflect the electric field produced by the atomistic environment hence influencing the maximum position (i.e. the ν_{max} value) as well as the spectrum width and lineshape. All the Results concerning the stationary spectrum are summarized in Table 1 and compared to the corresponding experimental data. Note that in this study, unlike the previous applications of the present approach to peptide backbone vibrational excitation (i.e. Amide I excitation)(86; 87), the anharmonic correction proved to be relevant not only for the ν_{max} position but also for properly reproducing the experimental lineshape of the IR signal in the A_1/A_3 region(47) (see S.I. Section S4 for additional details). From Table 1 it is evident that although inherently affected by the overestimated value of the unperturbed ν_{CO} frequency as obtained at the DFT level(108; 109), our model nicely reproduces the experimental $\Delta\nu_{max}$ between A_0 and A_1/A_3 in the wt simulation (20-21 cm^{-1} calculated vs. 20 cm^{-1} experimental) as well as the $\Delta\nu_{max}$ between A_1 and A_3 as determined (see below) by the wt and DM maxima shift (9-10 cm^{-1} calculated vs. 12 cm^{-1} experimental) essentially corresponding to the $A_1 - A_3$ band shift. Note that very similar results ($\Delta\nu_max \approx 14$ cm^1) are obtained when evaluating the closed conformation spectra for wt and DM by using the MD-free and MD-free-DM

simulations. These results, beyond confirming the widely accepted idea of the correspondence between the A_0 and A_1/A_3 IR sub-states with the His-open and His-closed conformations, also indicate that the A_1 and A_3 signals (in our model both involving the His-CO hydrogen bond), should necessarily be considered as the outcome of different perturbation patterns as provided by the QC perturbing environment (protein, solvent and counterion). From Table 1 it is also clear that, although our model nicely reproduces the experimental spectral lineshape of the A_1/A_3 region, the computed spectra are affected by a systematic overestimation of the signal width, resulting in unresolved A_1 and A_3 bands. Comparison of the experimental data with the computational results obtained for the wt and DM His-closed conformation using the MD-free and MD-free-DM simulations (QC-small) and the MD-constrained and MD-constrained-DM simulations (QC-large), indicates that such a width overestimation is mainly due to the excessive mobility of the His64-CO relative rototranslations as determined by the unrealistic classical motions along the high frequency modes within the His-CO-Heme complex, as provided by the classical MD simulations. See also S.I. (Sections S7 and S8) for additional details. In this respect from Figures 3(b), 3(c) and 3(d) it is evident that when for IR-PMM we use MD-contrained with explicit constraints between His64 and heme-CO in conjunction with the (rigid) QC-large, hence disregarding the His-CO-Heme internal structural fluctuations

due to the classical MD, the A_1/A_3 computed signal both for wt and DM, although still too large to discriminate between the A_1 and A_3 bands, is only slightly wider than the experimental one. Therefore, we assumed the MD-constrained simulation and the use of the (rigid) QC-large as providing the proper A_1/A_3 spectrum by means of the IR-PMM procedure. Interestingly, from Figures 3(c) and 3(d) we also estimated the relative stability between the His-open and His-closed conformations providing the A_0 and A_1/A_3 band probabilities (P_0 and $1-P_0$, respectively). From Figure 3(c) it is evident that for the wt a good correspondence between the computed spectrum (obtained by the weighted sum of the A_0 and A_1/A_3 computed spectra) and the experimental one is obtained when assuming the His-open conformation (i.e. the A_0 band) probability P_0 at $\approx$ 10-15 per cent, corresponding to a free energy change $\Delta A = -RT \ln \frac{P_0}{1-P_0} \approx$ 4-5 kJ/mol very close to the value of H-bond potential energy estimated by a Ryde and coworkers(67) but lower than the QM/MM value reported by Rovira and coworkers (ca. 12.7 kJ/mol). (44) A larger free energy change of about 9-10 kJ/mole (corresponding to $P_0 \approx 2-3$ per cent) is found for the DM as reported in Figure 3(d).

In order to characterize the perturbing electric field patterns involved in the A_1 and A_3 spectral signals, we identified the field direction mostly affecting the ν_{CO} frequency (defined by the unit vector $\mathbf{n}$) by means of a linear fitting of the frequency versus the field components as expressed within the QC

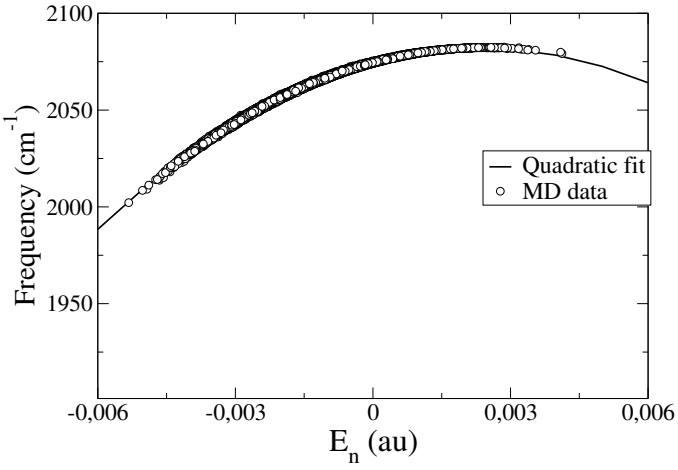

Figure 4: Calculated ν_{CO} frequencies (circles) as obtained by the wild type MD-constrained simulation versus the corresponding electric field component along the unit vector **n** (i.e. the perturbing electric field direction determining the stretching frequency, see text). In the figure it is also shown the quadratic fit (solid line) of the ν_{CO} frequencies as a function of the electric field component.

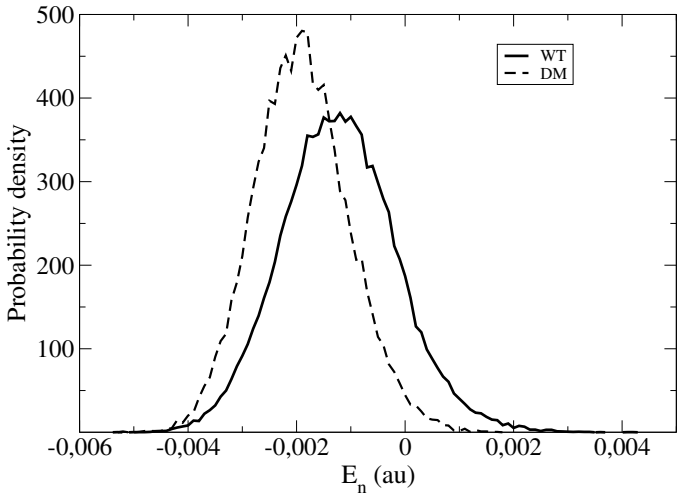

Figure 5: Probability distributions of E_n (the perturbing electric field component determining the stretching frequency, see text) as obtained by the wt (solid line) and T67R/S92D DM (dashed line) MD-constrained simulations.

coordinate system, basically corresponding to the mean frequency gradient within such a field component space. Such analysis provided a unit vector **n** virtually coinciding with the CO bond direction. In Figure 4 we show the calculated stretching frequencies versus the corresponding field along **n** (i.e. the field component E_n) as obtained by the wt MD-constrained simulation. From the figure it is evident that E_n virtually fully determines the ν_{CO} frequency, with the latter well expressed via a quadratic function of E_n. Identical results were obtained by using the DM MD-constrained simulation. From Figure 4 it is clear that different ν_{CO} spectral bands can only arise from different E_n fluctuation distributions characterized by significantly different mean values, as indeed shown in Figure 5 where the E_n distributions obtained by the wt and DM MD-constrained simulations are presented. From this last figure it is clear that the A_1 and A_3 bands, mostly characterizing the wt and DM spectrum respectively, are each corresponding to a specific E_n region with the whole His-closed spectrum determined by the weighted overposition of the corresponding two perturbation field distributions. It is worth noting that such results suggest that a map-based approach(110) could be a good tool for a computationally very efficient evaluation of the IR spectra for the same QC in different environments. Interestingly, when dissecting the relevant perturbation field component (i.e. E_n) into its various contributions in terms of protein residues, solvent and counter-ions, we obtain (see

Figure 6) that for both wt and DM most of the contribution concerning the mean intensity of E_n should be ascribed to the charged residues close to the heme (including the propionate groups) as well as to the solvent, while the counter-ions (corresponding to the last index number in Figure 6) are virtually irrelevant as always too far from the QC. As expected we also observe a not negligible contribution of the mutated (charged) residue in the case of DM. Note that since in our model the only relevant difference between the MD-free and MD-constrained simulations is the inclusion of the His64 residue into the QC, and hence its exclusion from the perturbing environment, we limited this last analysis to the MD-constrained and MD-constrained-DM systems.

A similar scenario can be observed in Figure 7 where we reported the root mean square fluctuation of the electric field component E_n due to the protein residues, the solvent and the counter-ions. We observe that almost all the chemical groups providing the larger mean E_n (Figure 6) also correspond to the larger field component fluctuations, hence being mainly responsible of the spectral signal shape and width. In this respect it is also interesting to note that, as reported more in details in the S.I. (Section S11), chloride ion perturbation provides a slight but not negligible effect on the spectral width. In fact, in the absence of the counterion perturbing field, the stationary IR spectrum is found the be slightly larger than the one reported in the Figure

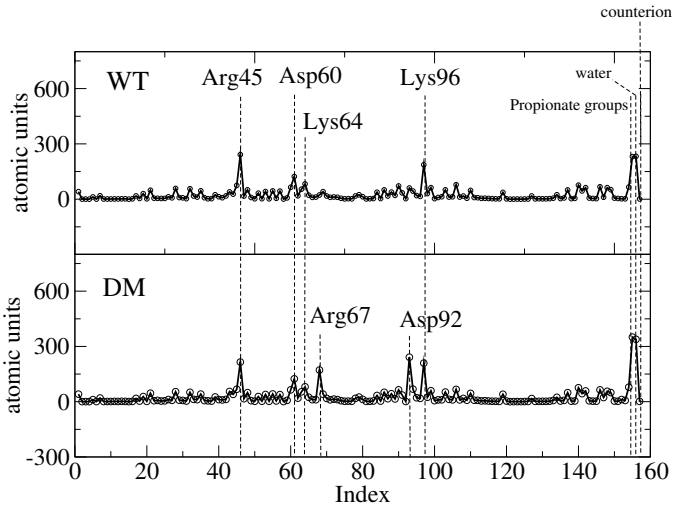

Figure 6: Average intensity (x 10^5) per protein residue, solvent and counterions (the last index number) of the perturbing electric field component E_n, as obtained for the wt (upper panel) and T67R/S92D double mutant (lower panel) MD-constrained simulations.

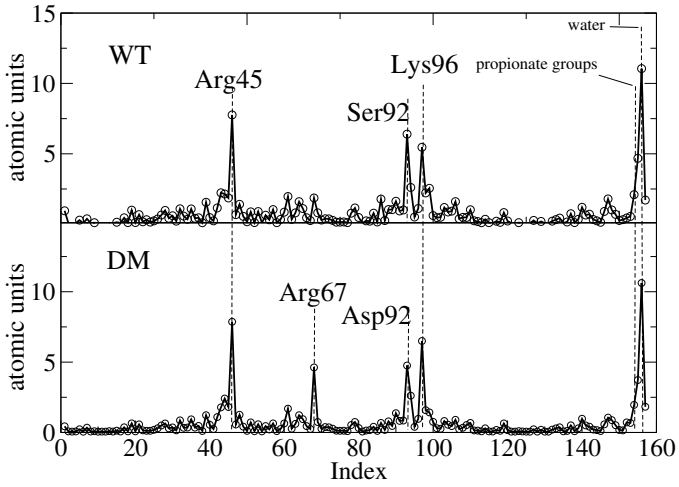

Figure 7: Root Mean Square Fluctuation per protein residue, solvent and counter-ions (the last index number) of the perturbing electric field component E_n, as obtained for the wt (upper panel) and T67R/S92D double mutant (lower panel) MD-constrained simulations.

3. Finally we also wish to underline that residue 92 (one of the mutated residues) relevantly contributes to the E_n fluctuation both in the WT and in the DM, i.e. irrespective of its chemical nature.

In order to assess the reliability of our results and possibly unveil the ν_{CO} VER mechanism and relaxation channels, we calculated the wt A_1/A_3 sub-state relaxation kinetics using the His-closed MD-constrained simulation (QC-large) which, according to our previous results, should be the best model for reconstructing the A_1/A_3 behavior. As briefly described in the Methods section, and more in details in the S.I. (Section S10), the approach we use (80) provides the relaxation kinetics for each possible channel and in particular for the direct first excited to ground state transition as well as for the excited to excited state transitions (mode-mode transitions): we used as final states all the vibrational eigenstates with energy lower or equal to the ν_{CO} first excited state, excluding only those modes with a frequency close or below the thermal energy one as they were considered as classical vibrational modes. When summing the transition rate constants as provided by all the possible relaxation channels we obtained as overall relaxation rate constant $k_{CO} = 0.048 \pm 0.010$ ps^{-1} corresponding to a mean lifetime $\tau_{CO} = 1/k_{CO} = 21 \pm 4$ ps, remarkably well reproducing the experimental estimates (47; 14) ($\tau_{CO} \approx 17 - 20$ ps). Interestingly, the direct first excited to ground state relaxation channel mostly contributes to the overall VER with a mean lifetime of ≈ 59

ps, while each of the other (mode-mode) channels has always a mean lifetime in the nanosecond range, thus corresponding to an almost negligible rate contribution. However, the large number of such excited to excited state transitions (equal to 138) results in an overall rate contribution which is even slightly larger than the direct excited to ground state relaxation rate (i.e. the overall mean lifetime of these summed rate constants is $\approx$ 33 ps). This is clearly indicating that when considering QC's with a high density of excited vibrational states, even when corresponding to very slow relaxation channels, the excited to excited state transitions can provide a very efficient relaxation mechanism. Finally, the large rate constant of the excited to ground state direct relaxation shows that in soft condensed systems characterized by large electrostatic interaction fluctuations the QC environment provides several perturbing field fluctuation components with frequencies close to the first vibrational excitation frequency (resonant perturbing field fluctuations (80)), clearly indicating that for such systems the direct transition to ground state can be the major relaxation channel.

Concluding Remarks.

Although at the center of an intense and longstanding theoretical and experimental activity, the nature of the sub-states of the stationary IR spectrum

of the C-O stretching in MbCO (ν_{CO}) has not yet been completely clarified. For this reason, in the present study we have readdressed such an issue using a theoretical-computational methodology (IR-PMM), intrinsically different from the most popular QM/MM techniques and specifically designed for modelling quantum observables in complex atomic-molecular systems. The wild-type (wt) MbCO as well as the T67R/S92D double mutant (DM) have been utilized at this purpose. Our study allowed for the first time to rigorously reconstruct the ν_{CO} stationary and time-dependent spectral features without departing from their quantum nature; moreover our results confirmed, on one hand, some of the previously reported hypothesis on the origin of the A_0 spectral band and, on the other hand, suggested new insights into the actual features of the A_1/A_3 band shape and also into the VER mechanism. More specifically our results can be summarized through the following points:

- Our calculations nicely reproduce the experimental A_0 to A_1/A_3 signal maxima shift (in the wt ≈ 20 cm^{-1}) as well as the experimental A_1/A_3 maximum shift between the wt and the DM spectra (≈ 10 cm^{-1}).

- Also the shape of the A_1/A_3 spectrum, showing a stretched red-tail, is satisfactorily reproduced although the experimental linewidth (14cm^{-1} for wt and 21 cm^{-1} for DM) is overestimated (40cm^{-1}) when including the effects of all the classical motions as provided by the MD-free and

MD-free-DM simulations (QC-small) and only slightly overestimated (21cm^{-1} for wt and 23 cm^{-1} for DM) when disregarding the unrealistic classical vibrations within the His64-FeCO-Heme complex (i.e. using the rigid QC-large and the constrained MD simulations).

- In line with most of literature data (see for example references (64; 8; 44; 72; 48)), our model confirms that the origin of the spectroscopic states A_0 and A_1/A_3 can be entirely ascribed to the presence of a H-bond between His64 and FeCO moieties giving rise to the His-open and His-closed conformational states, respectively.

- Using the experimental A_0 and A_1/A_3 relative instensities, we have also estimated a relative stability (free energy difference) of 4-5 kJ/mol between the His-open and His-closed conformations in the wt-MbCO. Our estimation of the H-bond free energy is in reasonable agreement with the value of the H-bond potential energy obtained by Ryde and coworkers (67) from calculations performed on the His64-FeCO-Heme complex in the presence of the protein rigid framework, but lower than the value reported by a different QM/MM study by Rovira and coworkers.(44)

- The A_1 and A_3 signals, both involved in the His-closed spectrum, are determined by different perturbation patterns as provided by the QC electrostatic environment (protein and solvent), rather than being due

to different structures of the His64-CO-Heme complex as previously suggested (48; 60).

- The systematic overestimation of the A_1/A_3 spectral widths corresponds, in our model, to an overestimation of the perturbing electric field fluctuations acting on the QC. In particular, when using in PMM calculations the QC-small we found that the relative motions of His64 (in the His-closed conformation H-bound FeCO) as provided by the classical MD simulations provide an unrealistically broad electric field pattern, mainly due to the unrealistic classical vibrations along modes which should be treated as fully quantum mechanical degrees of freedom (i.e. classically corresponding to constrained coordinates).

- The His64-CO-Heme complex perturbing environment, as also emerged from other studies(49) was also found to be the main determinant of the VER of ν_{CO}, which we modelled by using a recently developed theoretical-computational approach based on the explicit evaluation of the effects of the perturbation fluctuations on the vibrational states (80). Our calculations produce a mean lifetime of the ν_{CO} first vibrational excited state equal to $21\pm$ 4 ps in very good agreement with the experimental values of $\approx 17-20$ ps Interestingly, according to our model, the kinetic efficiency of the ν_{CO} VER is due both to the direct

relaxation to the ground state (the main relaxation channel) as well as to the mode-to-mode relaxations that, due to the high number of vibrational modes, provide an efficient alternative relaxation route.

Acknowledgments

We gratefully acknowledge CINECA (Italy) for an Iscra-C project.

4 Supporting Information

Details and results of unperturbed QM calculations. Additional information concerning MD simulations PMM-based evaluation free-energy associated to the Fe-C-O angle distortion. Comparison between IR-PMM stationary spectra with harmonic and anharmonic approximation. Comparison between MD-PMM stationary spectra with different QC-models. Analysis of the convergence for MD-constrained spectrum. IR spectrum for the His-closed state from unconstrained simulations. IR spectrum for the His-closed state from MD-free and MD-constrained using QC-small. Evaluation of empirical force-field ability in describing FeCO-His64 H-bond potential energy. Details of Theoretical Model for Vibrational Energy Relaxation. Effect of chloride ion on the stationary IR-spectrum.

References

[1] Antonini, E. and Brunori, M. *Hemoglobin and Myoglobin in Their Reactions with Ligands*. North-Holland, Amsterdam, 1971.

[2] Wilson, M.T.; Reeder, B.J. in *Encyclopedia of Respiratory Medicine.* G.J.Laurent and S.D.Shapiro Editors Academic-Pressi. 2006; pp 73-76

[3] Ordway, G.A.; Garry, D.J. Myoglobin: an essential hemoprotein in striated muscle. *J. Exp. Biol.* **2004** *207*, 3441-3446.

[4] Frauenfelder, H.; McMahon, B.H.; Fenimore, P.W. Myoglobin: the hydrogen atom of biology and a paradigm of complexity *Proc.Natl.Acad.Sci. USA* **2003** *100*, 8615-8617

[5] Sage, J.T.; Champion, P.M. Small substrate recognition in heme proteins. in: textitComprehensive Supramolecular Chemistry. 5. Pergamon, Oxford, UK, 1996; pp 171-213.

[6] Ansari, A.; Berendzen, J.; Braunstein, D.; Cowen, B.R.; Frauenfelder, H.; Kyung Hong, M.; Iben, E.T.; Johnson, J.B.; Ormos, P.; Sauke, T.B. et al. Rebinding and relaxation in the myoglobin pocket. Biophys.Chem. **1987** *26* 337-355.

[7] Yang, F.; Phillips, Jr., G.N. Crystal structures of CO-, deoxy- and Met-myoglobins at various pH values. *Biophys. J.* **1996**, *256* 762-774.

[8] Vojtechovsky, J.; Chu, K.; Berendzen, J.; Sweet, R.M.; Schlichting, I. Crystal Structures of Myoglobin-Ligand Complexes at Near-Atomic Resolution *Biophys. J.* **1999** *77* 2153-2174.

[9] Kachalova, G.S.; Popov, A.N.; Bartunik, H.D. A steric mechanism for inhibition of CO binding to heme proteins *Science* **1999** *284* 473-476

[10] Smith, R.D.; Blouin, G.C.; Johnson, K.A.; Phillips Jr, G.N.; Olson, J.S. Straight-chain alkyl isocyanides open the distal histidine gate in crystal structures of myoglobin *Biochemistry.* **2010** *49* 4977-4986.

[11] Barends, T.R.M.; Foucar, L.; Ardevol, A.; Nass, K.; Aquila, A.; Botha, S.; Doak, R.; Falahati, K.; Hartmann, E. ; Hilpert, M. et al. Direct observation of ultrafast collective motions in CO myoglobin upon ligand dissociation *Science* **2015** *350* 445-450

[12] Li, T.; Quillin, M.L.; Phillips Jr, G.N.; Olson, J.S. Structural determinants of the stretching frequency of CO bound to myoglobin *Biochemistry* **1994** *33* 1433-1446.

[13] Achterhold, K.; Keppler, C.; Ostermann, A.; Van Burck, U.; Sturhahn, W.; Alp, E.E.; Parak, F.G. Vibrational dynamics of myoglobin determined by the phonon-assisted Mssbauer effect *Phys Rev E* **2002** *65* 051916-051928.

[14] Hill, J.R.; Tokmakoff, A.; Peterson, K.A.; Santer, B.; Zimdars, I.; Dlott, D.D.; Fayer, M.D. Vibrational Dynamics of Carbon Monoxide at the Active Site of Myoglobin: Picosecond Infrared Free-Electron Laser Pump-Probe Experiments *J. Phys. Chem.* **1994** *98* 11213-11219

[15] Egeberg, K.D.; Springer, B.A.; Sligar, S.G.; Carver, T.E.; Rohlfs, R.J.; Olson, J.S. The role of Val68(E11) in ligand binding to sperm whale myoglobin. Site-directed mutagenesis of a synthetic gene *J Biol Chem.* **1990** *265* 11788-11795

[16] Carver, T.E.; Rohlfs, R.J.; Olson, J.S.; Gibson, Q.H.; Blackmore, R.S.; Springer, B.A.; Sligar, S.G. Analysis of the kinetic barriers for ligand binding to sperm whale myoglobin using site-directed mutagenesis and laser photolysis techniques *J Biol Chem.* **1990** *265* 20007-20020.

[17] Lim, M.; Jackson, T.A.; Anfinrud, P.A. Mid-infrared vibra- tional spectrum of CO after photodissociation from heme: evidence for a ligand docking site in the heme pocket of hemoglobin and myoglobin. *J. Chem. Phys.* **1995** *102* 4355-4366.

[18] Miksovska, J.; Day, J.H.; Larsen, R.W. Volume and enthalpy profiles of CO rebinding to horse heart myoglobin *J Biol Inorg Chem* **2003** *8* 621-625

[19] Srajer, V.; Teng, T.Y.; Ursby, T.; Pradervand, C.; Ren, Z.; Adachi, S.; Schildkamp, W.; Bourgeois, D.; Wulff, M. and Moffat, K. Photolysis of the carbon monoxide complex of myoglobin: nanosecond time-resolved crystallography. *Science* **1996** *274* 1726-1729

[20] Scott, E.E.; Gibson, Q.H. Ligand migration in sperm whale myoglobin. *Biochemistry* **1997** *36* 11909-11917

[21] Schlichting, I.; Berendzen, J.; Phillips, G.N.; Sweet, R.M. Crystal structure of photolysed carbonmonoxy-myoglobin. *Nature* **1994** *371* 808-812.

[22] Nienhaus, K.; Ostermann, A.; Nienhaus, G.U.; Parak, F.G.; Schmidt, M. Ligand Migration and Protein Fluctuations in Myoglobin Mutant L29W, Biochemistry **2005** *44* 5095-5105

[23] Sakakura, M.; Morishima, I.; Terazima, M. The Structural Dynamics and Ligand Releasing Process after the Photodissociation of Sperm Whale Carboxymyoglobin. *J. Phys. Chem. B* **2001** *105* 10424-10434

[24] Osapay, K.; Theriault, Y.; Wright, P.E.; Case, D.A. Solution structure of carbonmonoxy myoglobin determined from nuclear magnetic resonance distance and chemical shift constraints. *J. Mol. Biol.* **199** *244* 183-197.

[25] Steinbach, P.J.; Brooks, B.R. Protein hydration elucidated by molecular dynamics simulation *Proc. Nat. Ac. Sci.* **1993** *90* 9135-9139

[26] Cottoni, G.; Cordone, L.; Ciccotti, G. Molecular Dynamics Simulation of Carboxy-Myoglobin Embedded in a Trehalose-Water Matrix *Biophys. J.* **80** *2001* 931-938

[27] Vitkup, D.; Petsko, G.A.; Karplus, M. A comparison between molecular dynamics and X-ray results for dissociated CO in myoglobin. *Nat. Struct. Biol.* **1990** *4* 202-208

[28] Chillemi, G.; Anselmi, M.; Sanna, N.; Padrin, C.; Balducci, L.; Cammarata, M.; Pace, E.; Chergui, M.; Benfatto, M. Dynamic multiple-scattering treatment of X-ray absorption: Parameterization of a new molecular dynamics force field for myoglobin Struct Dyn. **2018** *5* 054101-054112.

[29] Bossa, C.; Anselmi, M.; Roccatano, D.; Amadei, A.; Vallone, B.; Brunori, M.; Di Nola, A. Extended molecular dynamics simulation of the carbon monoxide migration in sperm whale myoglobin, *Biophys. J.* **2004** *86* 3855-3862.

[30] Straub, J.E.; Karplus, M. Molecular dynamics study of the photodissociation of carbon monoxide from myoglobin: ligand dynamics in the first 10 ps. *Chem. Phys.* **1991** *158* 221-224.

[31] Merchant, K.A.; Noid, W.G.; Thompson, D.E.; Akiyama, R.; Loring,

R.F.; Fayer, M.D. Structural assignments and dynamics of the a substates of MbCo: Spectrally resolved vibrational echo experiments and molecular dynamics simulations *J. Phys. Chem. B* **2003** *107* 4-7.

[32] Sagnella, D.E.; Straub, J.E. A study of vibrational relaxation of B-state carbon monoxide in the heme pocket of photolyzed carboxymyoglobin *Biophys. J* 1999 *77* 70-84.

[33] Finkelstein, I.J.; Goj, A.; McClain, B.L.; Massari, A.M.; Merchant, K.A.; Loring, R.F.; Fayer, M.D. Ultrafast Dynamics of Myoglobin without the Distal Histidine: Stimulated Vibrational Echo Experiments and Molecular Dynamics Simulations *J. Phys. Chem. B* **2005** *109* 16959-16966

[34] De Angelis, F.; Jarzecki, A.A.; Car, R.; Spiro, T. Quantum Chemical Evaluation of Protein Control over Heme Ligation: CO/O2 Discrimination in Myoglobin *J. Phys. Chem. B* **2005** *109* 3065-3070

[35] Chen, H.; Ikeda-Saito, M.; Shaik, S. Nature of the Fe-O2 Bonding in Oxy-Myoglobin: Effect of the Protein *J. Am. Chem. Soc.* **2008** *130* 14778-14790

[36] Nutt, D.R.; Meuwly, M. CO migration in native and mutant myoglobin: atomistic simulations for the understanding of protein func- tion. *Proc. Natl. Acad. Sci. USA.* **2004** *101* 5998-6002.

[37] Meuwly, M. On the influence of the local environment on the CO stretching frequencies in native myoglobin: assignment of the B-states in MbCO. *ChemPhysChem* **2006** *10* 2061-2063.

[38] Plattner, N.; Meuwly, M. The role of higher CO-multipole moments in understanding the dynamics of photodissociated carbon- monoxide in myoglobin. *Biophys. J.* **2008** *94* 2505-2515.

[39] Wang, X.; Lu, C.; Yang, M. The Impact of Electron Correlation on Describing QM/MM Interactions in the Attendant Molecular Dynamics Simulations of CO in Myoglobin *Sci. Rep.* **2020** *10* 8539-8551

[40] Springer, B.A.; Sligar, S.G.; Olson, J.S.; Phillips, Jr, G.N. Mechanisms of Ligand Recognition in Myoglobin *Chem. Rev.* **1994** *94* 699714

[41] Phillips, G.N.; Teodoro, M.T.: Li, T.; Smith, B.; Olson, J.S. Bound CO is a molecular probe of electrostatic potential in the distal pocket of myoglobin. *J. Phys. Chem. B.* **1999** *103* 8817-8829.

[42] Tsai, A.-L-; Berka, V.; Martin, E.; Olson, J.S. A Sliding Scale Rule for Selectivity among NO, CO, and O2 by Heme Protein Sensors. *Biochemistry* **2012** *51* 172-186

[43] Tran, R.; Weinert, E.E.; Boon, E.M.; Mathies, R.A.; Marletta, M.A. .

Determinants of the Heme-CO Vibrational Modes in the H-NOX Family. *Biochemistry* **2011** *50* 6519-6530

[44] Rovira, C.; Schulze, B.; Eichinger, M.; Evanseck, J.D.; Parrinello, M. Influence of the heme pocket conformation on the structure and vibrations of the Fe-CO bond in myoglobin: a QM/MM density func- tional study. *Biophys. J.* **2001** *81* 435-445.

[45] Collman, J. P.; Fu, L. Synthetic models for hemoglobin and myoglobin *Acc. Chem. Res.* **1999** *32* 455-463.

[46] Rovira, C. Role of the His64 residue on the properties of the Fe-CO and Fe-O2 bonds in myoglobin. A CHARMM/DFT study *J. Mol. Struct. (THEOCHEM)* **2003** 632 309-321.

[47] Kossowska, D.; Kwak, K.; Cho, M: Do Osmolytes Impact the Structure and Dynamics of Myoglobin? *Molecules* **2018** *23* 3189-3193.

[48] Merchant, K. A.; Noid, W.G.; Akiyama, R.; Finkelstein, I.J.; Goun, A.; McClain, B.L.; Loring, R.F.; Fayer, M.D. Myoglobin-CO substate structures and dynamics: multidimensional vibrational echoes and molecular dynamics simulations *J. Am. Chem. Soc.* **2003**, *125* 13804-13818

[49] Merchant,K.A.; Thompson,D.E.; Xu, Q-H; Williams, R.B.; Loring,R.F.;

Fayer, M.D. Myoglobin-CO Conformational Substate Dynamics: 2D Vibrational Echoes and MD Simulations *Biophys. J.* **2002**, *82* 3277-3288.

[50] Cui, Q.; Karplus, M. Molecular properties from combined QM/MM methods. I. Analytical second derivative and vibrational calculations. *J. Chem. Phys.* **2000** *112* 1133-1149

[51] Caughey, W.S.; Shimada, H.; Choc, M.G.; Tucker, M.P. Dynamic protein structures: infrared evidence for four discrete rapidly interconverting conformers at the carbon monoxide binding site of bovine heart myoglobin. *Proc. Natl. Acad. Sci. U.S.A.* **1981** *78* 2903-2907.

[52] Anderton, C. L.; Hester, R. E.; Moore, J. N. A chemometric analysis of the resonance Raman spectra of mutant carbonmonoxy-myoglobins reveals the effects of polarity *Biochim. Biophys. Acta, Prot. Struct. Mol. Enzymol.* **1997** *1338*, 107-120.

[53] Iben I.E.T., Braunstein D., Doster W., Frauenfelder H., Hong M.K., Johnson J.B., Luck S., Ormos P., Schulte A., Steinbach P.J. et al. Glassy behavior of a protein. Phys. Rev. Lett. 1989; 62: 1916.

[54] Fuchsman, W.H.; Appleby, C.A. CO and O2 complexes of soybean leghemoglobins: pH effects upon infrared and visible spectra: comparison with CO and O2 complexes of myoglobin and hemoglobin. *Biochemistry* **1979** *18* 1309-1321

[55] Muller, J.D.; McMahon, B.J.; Chien, E.Y.T.; Sligar, S.G.; Nienhaus, G.U. Connection between the taxonomic substates and protonation of histidines 64 and 97 in carbonmonoxy myoglobin. *Biophys. J.* **1999** *77* 1036-1051.

[56] Johnson, J.B.; Lamb, D.C.; Frauenfelder, H.; Mueller, J.D.; McMahon, B. Ligand binding to heme proteins. VI. Interconversion of taxonomic substates in carbonmonoxymyoglobin. *Biophys. J.* **1996** *71* 1563-1573.

[57] Morikis, D.; Champion,P.M.; Springer, B.A.; Sligar, S.G. Resonance Raman investigations of site-directed mutants of myoglobin: effects of distal histidine replacement. *Biochemistry* **1989** *28* 4791-4800.

[58] Braunstein, D.P.; Chu, K.; Egeberg, K.D.; Frauenfelder, H.; Mourant, J.R. Ligand binding to heme proteins. III. FTIR studies of His-E7 and Val-E11 mutants of carbonmonoxymyoglobin. *Biophys. J.* **1993** *65* 2447-2454.

[59] Devereux, M.; Meuwly, M. Structural Assignment of Spectra by Characterization of Conformational Substates in Bound MbCO *Biophys. J* **2009** *96* 4363-4375

[60] Bagchi, S.; Nebgen, B.T.; Loring, R.F.; Fayer, M.D. Dynamics of a Myoglobin Mutant Enzyme: 2D IR Vibrational Echo Experiments and Simulations *J. Am. Chem. Soc.* **2010** *132* 18367-18376

[61] Oldfield, E.; Guo, K.; Augspurger, J. D.; Dykstra, C. E. A molecular model for the major conformational substates in heme proteins *J. Am. Chem. Soc.* **1991** *113* 7537-7541.

[62] Kushkuley, B.; Stavrov, S. S. Theoretical Study of the Electrostatic and Steric Effects on the Spectroscopic Characteristics of the Metal-Ligand Unit of Heme Proteins. 2. C-O Vibrational Frequencies, 17-O Isotropic Chemical Shifts, and Nuclear Quadrupole Coupling Constants *Biophys. J.* 1997 *72* 899-912.

[63] Franzen, S. 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.* **2002** *124* 13271-13281.

[64] Rovira, C.; Schulze, B.; Eichinger, M.; Evanseck, J.D.; Parrinello, M. 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. **2001**, *81* 435-445

[65] Unno, M.; Christian, J.F.; Olson, J.S.; Sage, J.T.; Champion, P.M. Evidence for hydrogen bonding effects in the iron ligand vibrations of carbonmonoxy myoglobin *J. Am. Chem. Soc.* **1998** *120* 2670-2671

[66] Cole, D.J.; ORegan, D.D.; Payne, M.C. Ligand Discrimination in Myo-

globin from Linear-Scaling DFT+U *J. Phys. Chem. Lett.* **2012** *11* 14481452

[67] Sigfridsson,E.; Ryde, U. Theoretical study of the discrimination between O2 and CO by myoglobin *J Inorg Biochem* **2002** *91* 101-115

[68] Momenteau, M.; Reed, C. A. Synthetic heme-dioxygen complexes Chem. Rev. **1994** *94* 659-698.

[69] Spiro, T. G.; Zgierski, M. Z.; Kozlowski, P. M. Stereoelectronic factors in CO, NO and O2 binding to heme from vibrational spectroscopy and DFT analysis *Coord. Chem. Rev.* **2001** *219* 923-936.

[70] Ivanov, D.; Sage, J.T.; Keim, M.; Powell, J.R.; Asher, S.A. Determination of CO orientation in myoglobin by single-crystal infrared linear dichroism. *J. Am. Chem. Soc.* **1994** *116* 4139-4140

[71] Choi, J-H.; Kwak, J.W.; Cho, M. Computational Infrared and Two-Dimensional Infrared Photon Echo Spectroscopy of Both Wild-Type and Double Mutant Myoglobin-CO Proteins *J. Phys. Chem. B* **2013** *117* 15462-15478

[72] Loccisano, A.; Acevedo, O.; DeChancie, J.; Schulze, B.; Evanseck, J. Enhanced sampling by multiple molecular dynamics trajectories: carbonmonoxy myoglobin 10 microseconds $A0$ to $A1-3$ transition from

ten 400 picosecond simulations. *J. Mol. Graph. Model.* **2004** *22* 369-376.

[73] Ray G.B., Li X.-Y., Ibers J.A.,Sessler J.L., Spiro T.G. How far proteins bend the FeCO unit? Distal polar and steric effects in heme proteins and models. *J. Am. Chem. Soc.* **1994** *116* 162-176

[74] Oldfield E., Augspurger J.D., Dykstra C.E. A molecular model for the major conformational substates in heme proteins. *J. Am. Chem. Soc.* **1991** *113* 7537-7541

[75] Sigfridson E., Ryde U. On the significance of hydrogen bonds for the discrimination between CO and O2 by myoglobin. *J. Biol. Inorg. Chem* **1999** *4* 99-110

[76] Teeter, M. M. Myoglobin cavities provide interior ligand pathway *Protein Sci.* **2004** *13* 313-321.

[77] Schulze, B.; Evanseck, J.D Cooperative role of Arg45 and His64 in the spectroscopic A3 state of carbonmonoxy myoglobin: molecular dynamics simulations, multivariate analysis, and quantum mechanical computations. *J. Am. Chem. Soc.* **1999** *121* 6444-6454.

[78] Thielges, M.C.; Axup, J.Y.; Wong, D.; Lee, H.S.; Chung, J.K.; Schultz, P.G.; Fayer, M.D. Two-dimensional IR spectroscopy of protein dynamics

using two vibrational labels: A site-specific genetically encoded unnatural amino acid and an active site ligand. *J. Phys. Chem. B* **2011** *115* 11294-11304.

[79] Daidone, I.; Aschi, M.; Zanetti-Polzi, L.; Di Nola, A.: Amadei, A. On the origin of IR spectral changes upon folding. *Chem.Phys.Lett.* **2010** *488* 213-218.

[80] Amadei,A.; Aschi, M. Modelling vibrational relaxation in complex molecular systems *Phys Chem Chem Phys* **2019** 21, 20003-20017.

[81] Morzan, U.N.; Alonso de Armino, D,J.; Foglia, N.O.; Ramrez, F.; Gonzlez Lebrero, M.C.; Scherlis, D.A:; Estrin D.A. Spectroscopy in Complex Environments from QM-MM Simulations *Chem. Rev.* **2018** *118* 4071-4113

[82] Del Galdo, S.; Mancini, G.; Daidone, I.; Zanetti Polzi, L.; Amadei, A.; Barone, V. Tyrosine absorption spectroscopy: Backbone protonation effects on the side chain electronic properties *J. Comput. Chem.* **2018** *39* 1747-1756

[83] The different number of states is justified by the fact the in the case of QC-open we performed a sensitivity analysis upon base-size, which clearly indicated that the convergence of the result was already reached with 13 states.

[84] Aschi, M.; Spezia, R.; Di Nola, A.; Amadei, A. A first-principles method to model perturbed electronic wavefunctions: the effect of an external homogeneous electric field *Chem Phys Lett* **2001**, *344* 374-380

[85] Zanetti-Polzi, L.; Del Galdo, S.; Daidone, I.; D'Abramo, M.; Barone, V.; Aschi, M.; Amadei, A. Extending the perturbed matrix method beyond the dipolar approximation: comparison of different levels of theory. *Phys Chem Chem Phys* **2018** *20*, 24369-24378.

[86] Amadei, A.; Daidone, I.; Zanetti-Polzi, L.; Aschi, M. Modeling quantum vibrational excitations in condensed-phase molecular systems. *Theor. Chem. Acc.* **2011** *129*, 31-43.

[87] Daidone, I., Thukral, L., Smith, J.C., Amadei, A. Monitoring the Folding Kinetics of a -Hairpin by Time-Resolved IR Spectroscopy in Silico. *J.Phys. Chem. B* 2015 *119* 4849-4856.

[88] Pinto, V.; Tasinato, N.: Barone, V.; Amadei, A.; Zanetti-Polzi, L.; Daidone, I. Modeling amino-acid side chain infrared spectra: The case of carboxylic residues. *Phys Chem Chem Phys* **2020** *22* 3008-3016.

[89] Aschi, M.; Barone, V.; Carlotti, B.; Daidone, I.; Elisei, F.; Amadei, A. Photoexcitation and relaxation kinetics of molecular systems in solution: towards a complete in silico model. *Phys Chem Chem Phys* **2016** *18* 28919-28931.

[90] Del Galdo, S.; Aschi, M.; Amadei, A. IR spectroscopy of condensed phase systems: can the environment induce vibrational mode coupling? *Chem Phys Lett* **2020** *763* 138168-138173.

[91] C. Rovira The structure and dynamics of the FeCO bond in myoglobin *J. Phys.: Condens. Matter* **2003** *15* S1809S1822

[92] Berendsen, H.J.C.; van der Spoel, D.; van Druned, R. GROMACS: A message-passing parallel molecular dynamics implementation, *Computer Physics Communications* **1995** *91* 43-56.

[93] D. van der Spoel, E. Lindahl, B. Hess, A. R. van Buuren, E. Apol, P. J. Meulenhoff, D. P. Tieleman, A. L. T. M. Sijbers, K. A. Feenstra, R. van Drunen and H. J. C. Berendsen, Gromacs User Manual version 4.5.6, www.gromacs.org (2010)

[94] Berendsen, H.J.C.; Postma, J.P.M.; van Gunsteren, W.F.; Hermans, J. In *Intermolecular Forces* edited by B. Pullman (Reidel, Dordrecht) **1981**, p 331.

[95] Schaftenaar G.; Noordik J. H. Molden: a pre- and post-processing program for molecular and electronic structures *J Computer-Aided Molec Des* **2000**, *14* 123-134.

[96] van Gunsteren, W.F.; Billeter, S.R.; Eising, A.A.; Hünenberger, P.H.;

Krüger, P.; Mark, A.E.; Scott, W.R.P.; Tironi, I.G. Biomolecular Simulation: The GROMOS96 Manual and User Guide Hochschulverlag AG an der ETH Zürich, **1996**

[97] Anselmi, M; Brunori, M; Vallone, B; Di Nola, A. Molecular Dynamics Simulation of Deoxy and Carboxy Murine Neuroglobin in Water. *Biophys. J. Volume* **2007**, *93* 434441

[98] Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2017** *126*, 014101-014109.

[99] Hess, B.; Bekker, H.; Berendsen, H.J.C:; Frajie, J.C.E.M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997** *81*, 1463-1472.

[100] Darden,T.A. and D. M. York and L. G. Pedersen, Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems, *J. Chem. Phys.* **1993** *98*, 10089-10092.

[101] Parr, R.G.; Yang, W. *Density functional theory of atoms and molecules* Oxford University Press, New York, **1999**

[102] Lee, C.; Yang, W.; Parr, R.G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys Rev B* **1988** *37*, 785-789.

[103] Hay, P.J.; Wadt, W.R. Ab initio effective core potentials for molecular calculations - potentials for K to Au including the outermost core orbitals. *J. Chem. Phys.* 1985, *82* 299-310.

[104] Runge, E.; Gross, E. K. U. Density-Functional Theory for Time-Dependent Systems *Phys Rev Lett* **1984**, *52* 997-1000.

[105] Blomberg, M.R.A.; Borowski, T,; Himo, F.; Liao, R-Z.; Siegbahn, P.E.M. Quantum Chemical Studies of Mechanisms for Metalloenzymes *Chem Rev* **2014**, *114* 3601-3658.

[106] Gaussian 09, Revision E.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, et al. Gaussian, Inc., Wallingford CT, 2013.

[107] Kanchanakungwankul, S.; Bao, J.L.; Zheng, J.; Alecu, I.M.; Lynch, B.J. Zhao,Y.; Truhlar, D.G. *Database of Frequency Scale Factors for Electronic Model Chemistries-Version 4* (https://comp.chem.umn.edu/freqscale/index.html)

[108] In this study we have also repeated the frequency calculation of the unperturbed QCs, without any significant improvement of the outcome, using different functionals, in particular: HSEH1PBE and PBE0.

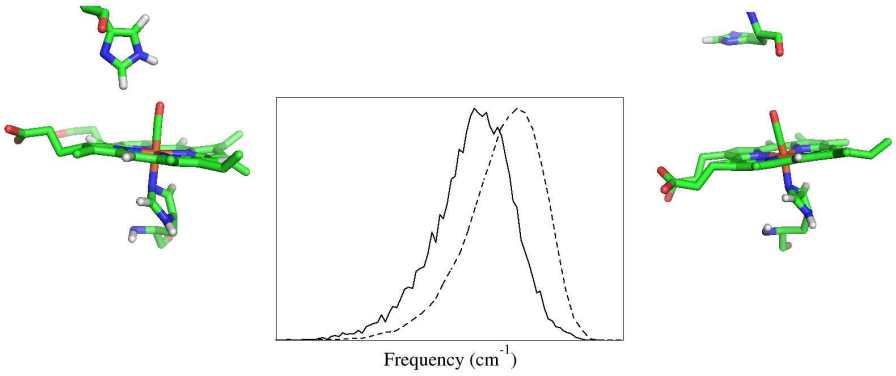

Figure 8: For Table of Content Only.

[109] Spiro, T.G.; Soldatova, A.V.; Balakrishnan, G. CO, NO and O2 as vibrational probes of heme protein interactions *Coord. Chem. Rev* **2013** *257* 511-527

[110] la Cour Jansen, T.; Knoester, J. A transferable electrostatic map for solvation effects on amide I vibrations and its application to linear and two-dimensional spectroscopy *J. Chem. Phys* **2006** *124* 044502-044513.