



OPEN

# The Impact of Electron Correlation on Describing QM/MM Interactions in the Attendant Molecular Dynamics Simulations of CO in Myoglobin

Xianwei Wang<sup>1</sup>✉, Chenhui Lu<sup>2</sup> & Maoyou Yang<sup>3,4</sup>✉

The impact of the dispersion and electron correlation effects on describing quantum mechanics/molecular mechanics (QM/MM) interactions in QM/MM molecular dynamics (MD) simulations was explored by performing a series of up to 2 ns QM/MM MD simulations on the B states of the myoglobin-carbon monoxide (MbCO) system. The results indicate that both dispersion and electron correlations play significant roles in the simulation of the ratios of two B states ( $B_1/B_2$ ), which suggests that the inclusion of the electron correlation effects is essential for accurately modeling the interactions between QM and MM subsystems. We found that the QM/MM interaction energies between the CO and the surroundings statistically present a linear correlation with the electric fields along the CO bond. This indicates that QM/MM interactions can be described by a simple physical model of a dipole with constant moment under the action of the electric fields. The treatment provides us with an accurate and effective approach to account for the electron correlation effects in QM/MM MD simulations.

Mixed multiscale QM/MM<sup>1–5</sup> models are routinely utilized in studies exploring the structure, reactivity, and electronic properties of proteins and other biological molecules<sup>6–11</sup>. In such models, the part of a large system that is of primary interest (e.g., the active site of an enzyme) is treated using an electronic structure method for capturing more accurate interaction energies, with the surroundings treated using an MM approach. Generally, the current QM/MM approaches use a single hybrid Hamiltonian to describe the system:

$$H = H_{QM} + H_{MM} + H_{QM/MM} \quad (1)$$

where  $H_{QM}$  is the Hamiltonian describing the QM subsystem,  $H_{MM}$  is the MM Hamiltonian, and  $H_{QM/MM}$  is the coupling term in the Hamiltonian describing the QM and MM interactions. Under this representation, the total energy of a given system can be obtained from the computation of the lowest eigenvalue of the Hamiltonian in Eq. 1.

$$E = E_{QM} + E_{MM} + E_{QM/MM} \quad (2)$$

The QM/MM formulation is flexible enough to accommodate almost all types of QM methods (e.g., semi-empirical methods, density functional theory (DFT)-based QM methods and *ab initio* (wave function (WF)-based) methods) in combination with an MM method (Amber, CHARMM, or a number of polarizable force fields)<sup>12–15</sup>.

Although the QM/MM treatment of the interactions is not exactly the same as that of the QM method, the change in the interactions (e.g., during a reaction or the dynamics of a biomolecular hydrogen bond) are modeled

<sup>1</sup>College of Science, Zhejiang University of Technology, Hangzhou, Zhejiang, 310023, China. <sup>2</sup>College of Mechanical Engineering, Shanghai University of Engineering Science, Shanghai, 201620, China. <sup>3</sup>School of Electronic and Information Engineering (Department of Physics), Qilu University of Technology (Shandong Academy of Sciences), Jinan, Shandong, 250353, China. <sup>4</sup>State Key Laboratory of Precision Spectroscopy, East China Normal University, Shanghai, 200062, China. ✉e-mail: [xwwang@zjut.edu.cn](mailto:xwwang@zjut.edu.cn); [ymy@qlu.edu.cn](mailto:ymy@qlu.edu.cn)

well by the QM/MM method using the electrostatic embedding approach<sup>16</sup>. This is as expected, because most of the protein functions are primarily achieved by the electrostatic<sup>17–21</sup>. For example, as previous works demonstrate, the electric field that a protease exerts at the active site is one of the most important origins of its catalytic power<sup>22–26</sup>.

With the advances in computer hardware and computational methods<sup>27–31</sup>, a number of QM/MM molecular dynamics (MD) studies employing various implementations of the expensive (prohibitive time/computational costs) Møller–Plesset perturbation theory (MP2) have been conducted<sup>32–43</sup>. However, it remains challenging to apply potentially highly accurate correlated *ab initio* QM methods (e.g., coupled clustering) in QM/MM MD simulations with a large number of QM atoms. Therefore, semi-empirical, DFT, or even Hartree–Fock (HF) methods have been employed in most QM/MM studies. However, DFT and HF methods suffer from a well-known limitation; that is, the Coulomb correlation is missing in the mean-field HF approach, which results in its inability to account for the dispersion effects, whereas many DFT-based QM methods cannot account for these effects in precise terms because the exact correlation functional is unknown<sup>44,45</sup>. Consequently, DFT methods do not always provide correct results in QM/MM simulations<sup>46–49</sup>, whereas HF method can only work well in the QM/MM simulation of specific systems (e.g., methyl-transfer reactions)<sup>50</sup>. Great efforts have been made to account for the dispersion effects in QM methods over the past decade, with a number of empirical dispersion methods routinely used for the QM calculations<sup>45,51–53</sup>.

The inclusion of the electron correlation and dispersion effects is mandatory for QM studies related to chemical systems and various intramolecular phenomena<sup>27,44</sup>. Investigation of the importance of these effects on the accurate description of the QM/MM interactions is a primary issue for realistic MD simulations of large or condensed systems with a QM/MM protocol. For this, first, an appropriate model system is required. Such a system should have a characteristic quantity (e.g., reaction barriers) that is easily handled by the QM/MM simulation with a small-sized QM region to facilitate the convergence of the simulation, and the characteristic quantity should be sensitive to the electrostatic interactions. It is preferable not to have dangling bonds between the QM and MM subsystems to avoid introducing unnecessary effects from the QM/MM boundary<sup>54</sup>.

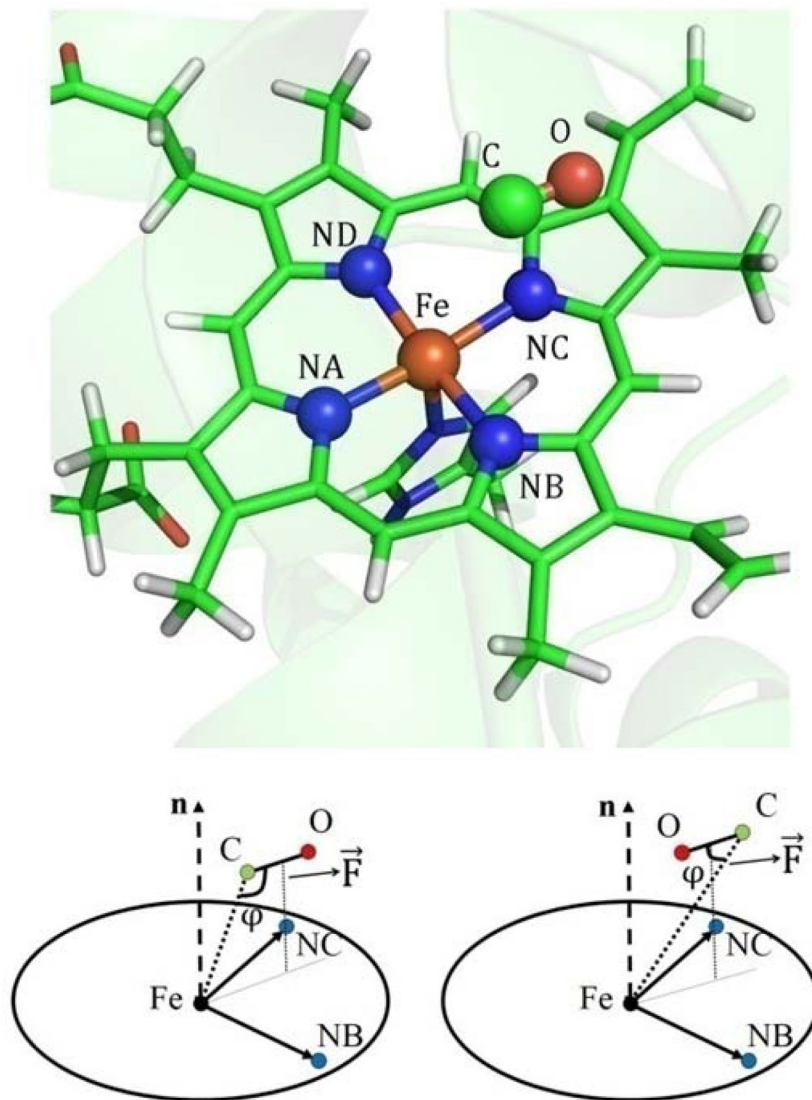
The MbCO system is one of the most studied proteins for investigating ligand binding, migration or other biologically relevant processes<sup>55–74</sup>. The spectroscopic characterization of CO has been used to explore the relationship between the dynamics, structure, and function of proteins in numerous experimental studies<sup>63,71,73–75</sup>. The molecular migration pathway for the CO to leave the myoglobin involves a set of transition steps between certain specific cavities, including the so-called distal heme pocket (the “docking” site) and four sites (Xe1–4) identified by the host xenon<sup>75–77</sup>. The CO binds itself to the myoglobin by forming a covalent bond with the center iron (Fe) of the heme before it then moves to the docking site adjacent to the heme at the edge of the distal pocket following photodissociation<sup>62,68,70,78,79</sup>. A previous study<sup>75</sup> from Alben *et al.* demonstrated that the features of the time-resolved polarized absorbance spectra of the CO in the bound and docking sites of myoglobin are different and were accordingly labeled as states A and B, respectively. Meanwhile, further splitting of the infrared spectrum with two well-resolved bands of photodissociated CO in the docking site (B state) was observed by Lim *et al.*<sup>71</sup>. Previous theoretical studies<sup>78–81</sup> have demonstrated that the splitting is a result of two different orientations of CO (B<sub>1</sub> and B<sub>2</sub>, which correspond to high frequency and low frequency, respectively) within the same site of the distal heme pocket with the carbon or oxygen atom closer to the center Fe of the heme. The relative locations of the CO and the heme for the B<sub>1</sub> state and a graphical representation of the B<sub>1</sub> and B<sub>2</sub> states are shown in Fig. 1. In our previous study<sup>80</sup>, we demonstrated that the obtuse angle of Fe···C≡O (labeled as  $\varphi$  in Fig. 1) with the most likely population of around 115° corresponded with the B<sub>1</sub> state, whereas the acute angle of  $\varphi$  with the most likely population of around 50° corresponded with the B<sub>2</sub> state. Meanwhile, Lim *et al.* demonstrated that the ratio of the two states (B<sub>1</sub>/B<sub>2</sub>) was not 1:1 but ~1.7:1<sup>72</sup>, with the ratio sensitive to the electrostatic interactions (including the polarization effect and the multipole electrostatic interactions) between the CO and its surroundings, which has been confirmed in numerous theoretical studies<sup>78–80</sup>. Therefore, the MbCO system is a particularly striking model for exploring the role of electron correlation effects in QM/MM MD simulations.

The HF method is a mean-field electronic structure method. The calculated molecular dipole moment using such an approach would be overestimated because of the failure to consider the electron correlation effects<sup>82</sup>. DFT-based methods such as the B3LYP are often expected to provide a more accurate description of the electrostatic interactions because they account for the electron correlation in their theoretical basis<sup>83</sup>. However, B3LYP functionals cannot provide a sufficient description of the dispersion<sup>44,45</sup>. Although the WF-based singles and doubles coupled-cluster method with perturbative triples (CCSD(T)) is regarded as the current gold standard in quantum chemistry, it remains impractical to utilize such a method in QM/MM MD simulations due to the huge computational costs. The MP2 offers a good combination of accuracy and low computational costs and has the capacity to provide a good estimation of the correlation energies. In practice, the MP2 has been used as benchmark calculations in dispersion-dominated complexes<sup>84,85</sup>.

In this study, a series of up to 2 ns QM/MM MD simulations using HF, B3LYP, and MP2 methods were performed in view of investigating the impact of electron correlations on the dynamics of the CO in the docking site of myoglobin. The physical fundamental of the MbCO system possesses two B states with different distribution ratios, and the correlations between the QM/MM electrostatic interactions and the electric fields are thus explored. We also discuss how a better description of the QM/MM interactions can be achieved.

## Methods

**MD simulations.** The A state of the MbCO structure (PDB id: 1MBC<sup>86</sup>), in which the CO binds itself to the center Fe of the heme, was used as the initial simulation model. The LEaP module<sup>87</sup> of the Amber program was utilized to add the missing hydrogen atoms. The amine groups of all lysine (Lys) and arginine (Arg) residues were protonated, whereas the carboxylic were deprotonated for all aspartic acid (Asp) and glutamic acid (Glu) residues.



**Figure 1.** The location of the CO in the docking site of the myoglobin for the B state (upper). The structure was extracted from the QM/MM MD simulation. Graphical representations of the B<sub>1</sub> state (lower left panel) and B<sub>2</sub> state (lower right panel). Here,  $\varphi$  is the angle that is defined as  $\text{Fe}\cdots\text{C}\equiv\text{O}$ . The heme is depicted as a circular plate, and  $\mathbf{n}$  is the normal vector of the plate plane, and  $\vec{F}$  represents the electric field exerted by the environment.

All histidine (His) residues were left neutral and protonated at the N $\delta$ 1 position based on the local electrostatic environment, with the exception of His93, which was protonated at the N $\epsilon$ 2 position. Meanwhile, the N-terminal was protonated, whereas the C-terminal was deprotonated. The force field parameter given by Giammona<sup>88</sup> was used to simulate the heme and the CO. The protein was placed in a periodic truncated octahedron box of TIP3P water molecules. The closest distance between the surface of the boxer and the protein atom was set to 12 Å. Three minimization steps based on the Amber ff99SB force field and short MD simulations were employed to relax the system and bring the CO molecule from the bound site (A state) to the docking site (B state). The bonding interaction between the CO and the center Fe of the heme was removed so that the CO would be dissociated under the action of the repulsive term of the van der Waals interactions. First, the solvent molecules were optimized with all protein atoms constrained to their initial structure. Second, the backbone protein atoms and all the heme and CO atoms were constrained, and the other protein atoms and the solvent molecules were relaxed. Third, all the protein and solvent atoms were energy minimized, and the heme and the CO were still constrained to their initial structure. Then, the system was heated to 300 K at 100 ps with all the heme and CO atoms constrained. Following this, a 100-ps MD simulation without constraint was carried out to simulate the photodissociation with a periodic boundary condition at 300 K and 1 atm. The time integration step was set to 2.0 fs with the application of the SHAKE<sup>89</sup> algorithm to maintain all hydrogen atoms in reasonable positions. The particle-mesh Ewald method<sup>90</sup> was applied to treat long-range electrostatic interactions. A 10 Å cut-off was implemented for the van der Waals interactions. Langevin dynamics<sup>91</sup> was used to regulate the temperature with a collision frequency of 1.0 ps<sup>-1</sup>.

A series of up to 2 ns QM/MM MD simulations was independently performed following the classical equilibrium simulation. Only the CO was partitioned into the QM subsystem, whereas the remainder of the system, including the heme, protein, and solvent was partitioned into the MM subsystem. Corresponding force field parameters that were similar to the classical equilibrium simulation were used for the MM subsystem. The QM subsystem was treated by the B3LYP hybrid density functional with various basis sets of 6–31 G, 6–31 G\*\*, 6–31++G\*\*, aug-cc-pVDZ and aug-cc-pVTZ in the QM/MM MD simulations to test the effect of the size of the basis sets. The other two MD simulations were performed at the HF/aug-cc-pVTZ/amber99SB and MP2/aug-cc-pVTZ/amber99SB QM/MM levels. In all QM/MM MD simulations, a 25 Å cutoff was utilized to treat QM/MM electrostatic interactions. The electronic coupling between the QM and MM subsystems was treated by including MM charges in the QM Hamiltonian. The integration step was set to 1.0 fs and trajectories were generated with the structures saved every 200 fs. The angles of the Fe···C≡O were calculated using the saved structures ( $1 \times 10^4$  configurations in total) and were subsequently used to analyze the ratios of the two B states. The Amber12 program<sup>87</sup> was used to perform the MD simulations. The sander module with an interface to the Gaussian09 program<sup>92</sup> was utilized to perform the QM/MM MD simulations.

**Calculation of the electric field and QM/MM interactions.** The electric field along the CO bond and the electrostatic interaction energies between the QM and MM subsystems were calculated and applied for exploring the foundation of that how the inclusion of the electron correlation and dispersion effects influences QM/MM MD simulations. Although the QM/MM MD simulations at various QM/MM levels generated a series of trajectories, the calculations of the electric field and the interaction energies were performed on the basis of only one of them to test the correlations between the relatively independent calculations and the simulated results. The obtained ( $B_1/B_2$ ) ratio is ~1.9:1 based on the MD simulation at the B3LYP/6–31 G\*\* QM/MM level. This ratio is somewhat modest for studying both the  $B_1$  and  $B_2$  states. Therefore, the QM/MM interactions were calculated using the HF, B3LYP and MP2 QM methods combined with the aug-cc-pVTZ basis set based on this trajectory and the calculated results were plotted as the function of the calculated electric fields, and then compared with the QM/MM simulated results (the distribution of the angle of the Fe···C≡O) at the HF/aug-cc-pVTZ, B3LYP/aug-cc-pVTZ and MP2/aug-cc-pVTZ QM/MM levels.

**Calculation of the electric field.** A total of 10,000 configurations generated from the 2 ns QM/MM MD simulations were used to calculate the electric fields along the CO bond using the charge model of the corresponding MM force field. Here, the electrostatic potential at the center of the C and O atoms of the CO molecule was first calculated as follows,

$$\varphi(\vec{r}) = \frac{1}{4\pi\epsilon_{\text{eff}}} \sum_i \sum_{j \in i} \frac{q_{ij}}{|\vec{r} - \vec{r}_{ij}|} \quad (3)$$

where  $i$  denotes the residue number that runs over all the residues of the protein, heme and solvent molecules (6,855 waters);  $j$  is the atom number in residue  $i$ ;  $q_{ij}$  and  $\vec{r}_{ij}$  are the atomic charge and position vector of atom  $j$  respectively;  $\vec{r}$  is the position vector C or O atoms of the CO molecule;  $\epsilon_{\text{eff}}$  denotes the effective dielectric constant and was set to 1.0 since the explicit water model was utilized<sup>93</sup>.

The mean electric field along the CO bond was obtained using the following expression,

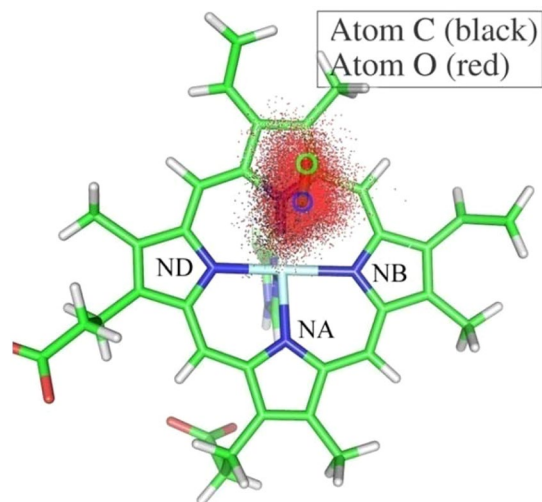
$$\vec{F}(\vec{r}_{\text{CO}}) = \frac{\varphi(\vec{r}_{\text{C}}) - \varphi(\vec{r}_{\text{O}})}{|\vec{r}_{\text{CO}}|} \quad (4)$$

where  $\varphi(\vec{r}_{\text{C}})$  and  $\varphi(\vec{r}_{\text{O}})$  are the electrostatic potentials at the center of the C and O atoms of the CO, which were calculated using Eq. 3, and  $|\vec{r}_{\text{CO}}|$  is the bond length of the CO.

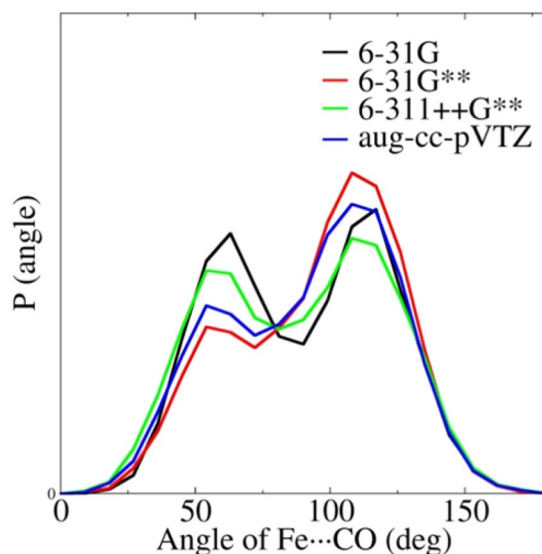
**Calculation of the electrostatic interactions.** Two schemes were utilized to calculate the electrostatic interaction energies between the QM (CO molecule) and MM subsystems (the surroundings). The first was based on the point charge model, whereas the ESP charges derived from the QM/MM calculations were used for the CO to ensure the polarization effect was taken into account. The charges of the MM subsystem were taken from the corresponding force field parameter. One limitation of this scheme is that it does not account for the electrostatic multipole interactions. The second scheme was based on the QM/MM calculations. Here, a QM energy was first obtained from the QM/MM calculations of the QM subsystem in the protein environment. The obtained energy (labeled as  $E_{\text{total}}$ ) contained the self-energy of the QM subsystem (labeled as  $E_{\text{self}}$ ) and the electrostatic interactions between the QM and MM subsystems (labeled as  $E_{\text{EI}}$ ). The  $E_{\text{self}}$  of the QM subsystem could be obtained from the QM calculations in gas phase, and the  $E_{\text{EI}}$  could then be obtained by deducting the  $E_{\text{self}}$  from the  $E_{\text{total}}$ . The calculated interaction energies were plotted as a function of the electric fields to provide a statistics analysis of the foundation of the dynamic behavior of CO in the docking site of myoglobin and to allow us to explore the correlation relationship between the QM/MM interaction energies and the electric fields.

## Results and Discussion

In previous works<sup>78,94</sup>, it was observed that photodissociated CO only remains in the docking site for nanoseconds before it then rapidly moves to other regions of the protein. To confirm that the CO molecule was located in the docking site of the myoglobin in all 2 ns MD simulations, contour plots presenting the cumulative distribution projected onto the heme plane of the C and O atoms during the 2 ns QM/MM MD simulation were created as shown in Fig. 2, and an X-ray experimental structure<sup>95</sup> of the  $B_1$  state is also shown in Fig. 2. It can be seen from



**Figure 2.** Contour plots of the probability distribution of the C and O atoms of the CO molecule on the heme plane during one of the trajectories generated by the 2 ns QM/MM MD simulation. The relative locations of the  $B_1$  state of the CO obtained from the X-ray experiment<sup>95</sup> are marked by the blue circle (C atom) and the green circle (O atom).

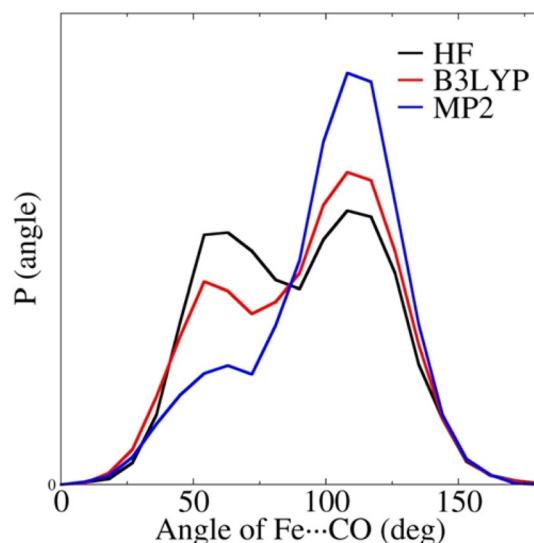


**Figure 3.** The distribution of the  $\text{Fe}\cdots\text{C}\equiv\text{O}$  angle for the QM/MM MD simulations at the B3LYP with various basis sets of 6-31 G, 6-31 G\*\*, 6-311++G\*\*, and aug-cc-pVTZ QM/MM levels.

the contour plot that the distributions of the C and O atoms were both concentrated. The predicted locations of the  $B_1$  state in the 2 ns QM/MM MD simulation were at the edge of the distal heme pocket adjacent to the porphyrin ring *c*, which is in good agreement with the X-ray experiment. Meanwhile, the CO molecule lay approximately parallel to the heme plane in the 2 ns QM/MM MD simulation, which is also in good agreement with the X-ray experiment. More detailed analysis indicated that the distribution of the center of mass of the CO for both the  $B_1$  and the  $B_2$  states was concentrated in the same location, with the CO molecule displaying an antiparallel orientation. The C and O atoms of the CO alternately formed a weak hydrogen bond with the H atom of the  $-\text{NH}$  group of the His64 during the MD simulations, which provided an anchoring force for trapping the CO. In fact, the CO molecule remained in the docking site following 4 ns QM/MM MD simulations. Therefore, 2 ns MD simulations will not make the CO molecule move out of the docking site.

The B3LYP is one of the most widely used density functionals<sup>44</sup> and is expected to provide an accurate calculation of the electrostatic potentials for protein. The QM/MM MD simulation was thus first performed using the B3LYP method. To test the effects of the size of the basis sets, a series of up to 2 ns QM/MM MD simulations with a QM subsystem treated via B3LYP/6-31 G, B3LYP/6-31 G\*\*, B3LYP/6-311++G\*\*, and B3LYP/aug-cc-pVTZ were carried out, with the results shown in Fig. 3. It can be seen from the figure that the size of the basis sets in the QM/MM MD simulations had an obvious effect on the simulated results. The predicted ratio of the  $B_1$



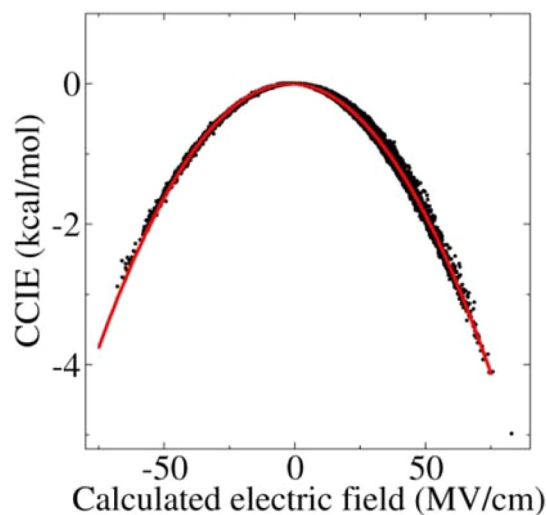


**Figure 4.** The distribution of the Fe...C≡O angle for the QM/MM MD simulations at the HF/aug-cc-pVTZ/Amberff99SB, B3LYP/aug-cc-pVTZ/Amberff99SB, and MP2/aug-cc-pVTZ/Amberff99SB QM/MM levels.

(corresponding to the  $\sim 115^\circ$  angle of the Fe...C≡O) and  $B_2$  (corresponding to the  $\sim 55^\circ$  angle of the Fe...C≡O) states using the B3LYP/6-31 G QM method was  $\sim 1.1:1$ , and the proportion of the  $B_1$  state enhanced the ratio, which became  $\sim 1.8:1$  when the polarization function was introduced (6-31 G\*\* basis set). The results demonstrate that the introduction of the polarization function in the basis sets had a significant impact on the simulations using the QM/MM method. However, the ratio fell back to  $\sim 1.1:1$  when the diffuse function was added (the result of 6-311++G\*\* in Fig. 3), which indicates that the diffuse function and the polarization function have different roles in impacting the QM/MM simulations. The ratios of the two B states obtained via the B3LYP/aug-cc-pVDZ and B3LYP/aug-cc-pVTZ methods were almost the same (see Figure S1), which indicates that the further improvement in the basis sets after the basis set of aug-cc-pVDZ was achieved had very little influence on the simulations. The predicted ratio of the two B states was  $\sim 1.5:1$  using the B3LYP/aug-cc-pVTZ method, which was closest to that using the B3LYP/6-31 G\*\*, meaning the basis set of 6-31 G\*\* may provide a good balance in terms of accuracy and speed due to the error cancellation.

Although the density functional B3LYP is regarded as providing a more accurate description of molecular electrostatic potentials than the HF method due to the consideration of certain electron correlation effects, the B3LYP does not account for the dispersion. Thus, to explore the role of dispersion and electron correlation in QM/MM MD simulations, two 2 ns QM/MM MD simulations of MbCO with a QM subsystem treated via HF/aug-cc-pVTZ and MP2/aug-cc-pVTZ were performed. The obtained distribution of the Fe...C≡O angles is plotted in Fig. 4. For comparison, the results based on the B3LYP/aug-cc-pVTZ/Amberff99SB QM/MM MD simulation were also plotted. It can be seen from Fig. 4 that there was a significant difference in the ratios of the two B states predicted by the HF/aug-cc-pVTZ/Amberff99SB, B3LYP/aug-cc-pVTZ/Amberff99SB, and MP2/aug-cc-pVTZ/Amberff99SB QM/MM MD simulations. In terms of the results obtained from the HF/aug-cc-pVTZ/Amberff99SB QM/MM MD simulation, the  $B_1/B_2$  ratio increased from  $\sim 1.1:1$  to  $\sim 1.5:1$  for the simulation at the B3LYP/aug-cc-pVTZ/Amberff99SB QM/MM level, which indicates that the inclusion of the electron correlation is essential in QM/MM MD simulations. The  $B_1/B_2$  ratio calculated from the MP2/aug-cc-pVTZ/Amberff99SB QM/MM MD simulation was  $\sim 3.5:1$ , a more than threefold increase on that obtained via the HF method and a more than twofold increase on that obtained via the B3LYP method. The obtained ratio at the MP2/aug-cc-pVTZ/Amberff99SB QM/MM level demonstrated a more significant change in that at the B3LYP/aug-cc-pVTZ/Amberff99SB QM/MM level compared with the result of the HF/aug-cc-pVTZ/Amberff99SB QM/MM simulation. Therefore, it can be stated that it is essential to adopt a QM method that accounts for the dispersion effect in QM/MM MD simulations to obtain accurate results. However, in most QM/MM MD simulations, dozens or even hundreds of atoms have to be partitioned into a QM subsystem. Although correlated WF methods (such as MP2 or CCSD(T)) used in combination with a moderate-sized basis set, such as a aug-cc-pVTZ, could provide accurate results for the simulations, it remains impractical to obtain convergent results using these methods due to the huge computational cost. It is thus essential to make efforts to develop accurate and effective methods that account for the electron correlation effects in QM/MM MD simulations.

It is worth noting that the results based on the B3LYP/aug-cc-pVTZ/Amberff99SB QM/MM model had the best agreement with the experimental measurement of 1.7:1, but this does not prove that the B3LYP performs better than the MP2 for QM/MM MD simulations. In fact, it may indicate that a larger QM region is required for a more accurate simulation of the ratio of the two B states and that the QM region should include the heme, given the strong QM interactions between the center Fe of the heme and the CO molecule. To accurately model the QM interactions, a new MM method may need to be developed. However, it is important to note that the aim of this work was to explore how the QM/MM interactions can be accurately modeled. Here the MbCO system served as



**Figure 5.** Distribution of the CCIEs between the QM (CO molecule) and the MM (the surroundings) as a function of the electric fields along the CO bond, which exhibits a quadratic function relationship. The influence on the CCIEs of the changes in the bond length of the CO was eliminated.

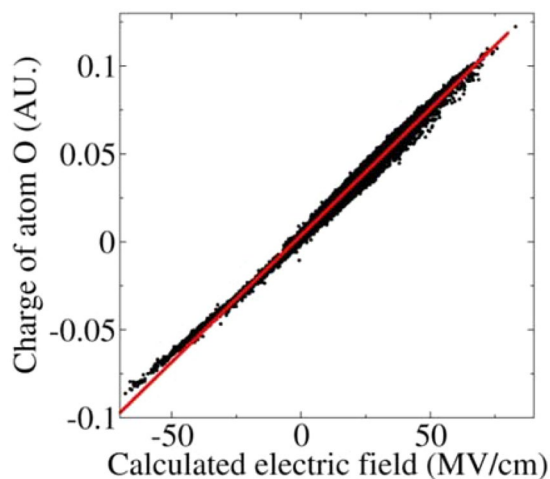
a model system rather than a real system, meaning we simply focused on the differences in the simulated results of various QM/MM models.

Thus far, we have examined the role of electron correlation in the QM/MM simulations of the dynamics of the CO molecule in the docking site of myoglobin and have indicated its significance for determining the ratio of the two B states. Next, the theoretical fundamental of possessing two MbCO B states is explored. In previous works<sup>22,74,78</sup>, it was indicated that the two B states are stabilized by the electrostatic interactions between the CO and its surroundings and that the splitting of the infrared spectrum is derived from the Stark effect where the photodissociated CO in the same docking site experiences the electric field with antiparallel orientations. The conclusions suggest that there may be a specific relationship between the electrostatic interactions and the electric fields. An investigation of this relationship could help us to better understand the dynamics of the CO in the docking site of myoglobin and could reveal new physical insights into molecular interactions.

The classical electrostatic interactions (calculated on the basis of the point charge model) as a function of the electric fields are shown in Fig. 5. It can be seen from the figure that the classical electrostatic interactions present an exact quadratic function relationship with the electric fields and that the quadratic function relationship has a y-axis (vertical axis) symmetry. Although the electrostatic energies presented in Fig. 5 were mainly derived from the polarization effect, in reality, the calculated electrostatic interactions were also influenced by the change in bond length of the CO, with the exception of the polarization effect of the electric field. Meanwhile, the influence on the calculated interaction energies from the change in the bond length of the CO was small and relatively independent from that of the polarization effect. Therefore, the quadratic function relationship remained true and exhibited a y-axis (vertical axis) symmetry, even given the effect of the change in the bond length of the CO (see Figure S2), whereas the zero point of the quadratic function was almost unchanged because the ESP charge of the CO in the gas phase calculated at the B3LYP/aug-cc-pVTZ QM level was very small. The distribution of the classical interaction energies was symmetrical in the positive and negative electric fields, meaning it should not result in a different ratio for the two B states. As such, it is not a primary factor in the B state of the MbCO system possessing a different population, which is consistent with previous research<sup>78,79</sup>.

Corresponding with the quadratic function relationship between the interaction energies and the electric fields, the ESP-based atomic charges of the O atom of the CO molecule presented an exact linear relationship with the electric fields, as shown in Fig. 6. Similarly, the charge was also affected by the bond length (see panel A in Figure S3). Meanwhile, the fluctuation of the charges also had a linear relationship with bond length (see panel B in Figure S3). The positive direction of the electric field along the CO bond calculated using Eq. 4 pointed from the C toward the O. It can be seen from Fig. 6 that the O atom takes a positive and a negative charge in the positive and negative electric fields, respectively, which resulted in the CO molecule exhibiting  $C^-O^+$  and  $C^+O^-$  states. The linear relationship between the atomic charge and the electric fields may be utilized to introduce the polarization effect into the point charge model of classical force fields.

In actual QM/MM MD simulations, the interactions between the QM and MM subsystems undoubtedly contain multipole interactions that go beyond the classical point charge interactions, because these interactions are obtained using electronic structure methods. To statistically explore the physical fundamental of the two B states, we plotted the calculated QM/MM interactions at the HF/aug-cc-pVTZ/Amberff99SB, B3LYP/aug-cc-pVTZ/Amberff99SB, and MP2/aug-cc-pVTZ/Amberff99SB QM/MM levels as a function of the electric fields, with the results shown in Fig. 7. It can be seen from the figure that the obtained QM/MM electrostatic interactions of all the QM/MM models exhibited an approximate linear relationship relative to the positive or negative electric fields. Although different from the classical point charge interactions, the distribution of the QM/MM interactions in the positive and negative electric fields did not exhibit a symmetry that corresponded exactly with the



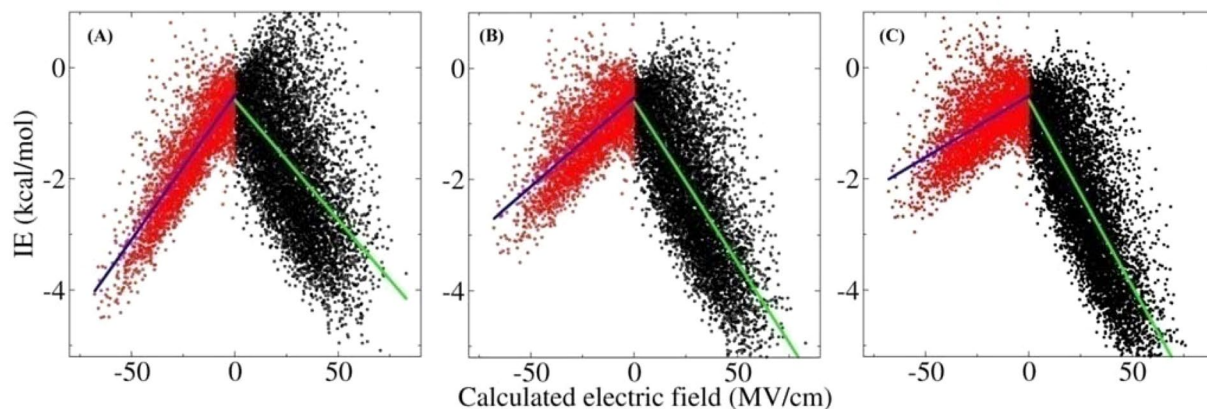
**Figure 6.** The obtained ESP charges in the O atom of the CO molecule derived from the QM/MM calculations in the MD simulation as a function of the electric fields along the CO bond, which exhibits a linear function relationship. The influence of the changes in the bond length of the CO was eliminated.

different ratio of the two B states. To further understand the correlation between the distribution of the QM/MM interactions relative to the electric fields and the ratio of the two B states, the QM/MM interaction data were simply partitioned into two segments on the basis that the corresponding electric field was positive or negative and fitted with a linear function. The ratios of the two slopes of the fitted curves ( $k_1$  and  $k_2$  correspond to a positive and a negative electric field, respectively) for the results obtained at the HF/aug-cc-pVTZ/Amberff99SB, B3LYP/aug-cc-pVTZ/Amberff99SB, and MP2/aug-cc-pVTZ/Amberff99SB QM/MM levels are presented in Table 1. For comparison, the predicted  $B_1/B_2$  ratio using the HF/aug-cc-pVTZ/Amberff99SB, B3LYP/aug-cc-pVTZ/Amberff99SB, and MP2/aug-cc-pVTZ/Amberff99SB QM/MM MD simulations are also presented in Table 1. The  $k_1/k_2$  of the results for the HF/aug-cc-pVTZ/Amberff99SB, B3LYP/aug-cc-pVTZ/Amberff99SB, and MP2/aug-cc-pVTZ/Amberff99SB QM/MM calculations were 0.83, 1.80, and 3.02, respectively, which demonstrates a good correlation with the predicted  $B_1/B_2$  ratios of 1.09, 1.54, and 3.45. Therefore, we can conclude that the physical fundamental that there are two B states with different proportions is a discrepancy in the QM interactions between the CO and the surroundings in antiparallel orientations.

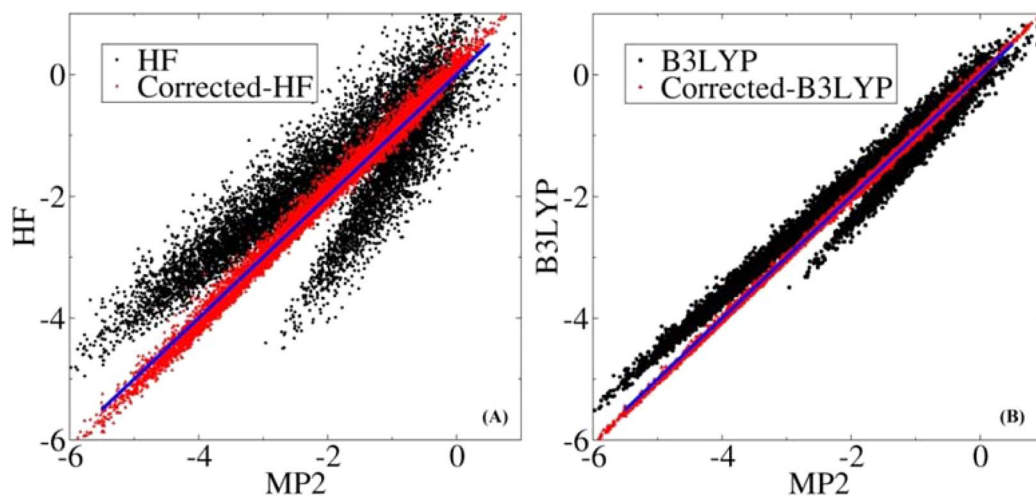
It is worth noting that the QM/MM interactions relative to the electric fields statistically presented a linear distribution as shown in Fig. 7, which indicates that the electrostatic interactions between the QM group and its surroundings can be described using a simple physical model of a dipole with constant moment under the action of an electric field. In the treatment, the interaction energy can be calculated using the formula  $E = \vec{\mu} \cdot \vec{F}$ , where  $\vec{F}$  is the electric field,  $\vec{\mu}$  is the dipole moment of the QM group (CO here) and  $\vec{\mu} = q\vec{l}$ , and  $q$  and  $\vec{l}$  are the electric-field-based equivalent atomic charge (EAC) and the bond length of the CO molecule, respectively. If the bond length adopted an approximate constant value of 1.14 Å, the obtained EAC will be  $C^{-0.1638}O^{0.1638}$  (under the action of the positive electric field) and  $C^{0.1979}O^{-0.1979}$  (under the action of the negative electric field) for the HF/aug-cc-pVTZ/Amberff99SB QM/MM calculations,  $C^{-0.2196}O^{0.2196}$  (under the action of the positive electric field) and  $C^{0.1222}O^{-0.1222}$  (under the action of the negative electric field) for the B3LYP/aug-cc-pVTZ/Amberff99SB QM/MM calculations, and  $C^{-0.2541}O^{0.2541}$  (under the action of the positive electric field) and  $C^{0.0843}O^{-0.0843}$  (under the action of the negative electric field) for the MP2/aug-cc-pVTZ/Amberff99SB QM/MM calculations. Here, it is worth noting that the EAC was statistically obtained unlike with the ESP-based charge model. The obtained  $B_1/B_2$  ratio using the MP2/aug-cc-pVTZ/Amberff99SB QM/MM simulation was 3.45:1, which indicates that the  $C^-O^+$  state played a dominant role in the interaction of the CO with the electric field. This result can explain the experimentally observed  $C^-O^+$  polarity<sup>96,97</sup>.

The obtained EACs for the CO were different across the various QM/MM calculations (HF, B3LYP, and MP2), and these were utilized to empirically introduce the electron correlation effects in the QM/MM MD simulation by correcting the QM/MM interactions using calculated interactions with the EACs. We tested the performance of this approach on the calculated QM/MM interaction energies at the HF/aug-cc-pVTZ/Amberff99SB and B3LYP/aug-cc-pVTZ/Amberff99SB QM/MM levels, with the interaction energies corrected by deducting the calculated interaction energies from the corresponding (HF or B3LYP QM/MM model) EACs and adding the interaction energies to the MP2 EACs. The correlation between the corrected HF and B3LYP interaction energies and the MP2 QM/MM calculations is shown in Fig. 8. For comparison, the correlation between the original HF and B3LYP QM/MM interaction energies and the MP2 QM/MM calculations is also plotted. It can be seen from Fig. 8 that the corrected interaction energies for both the HF and the B3LYP QM/MM calculations were in exact agreement with the MP2 QM/MM calculations, with a correlation coefficient ( $R^2$ ) of 0.9921 for the corrected HF and a  $R^2$  of 0.9995 for the corrected B3LYP. The mean unsigned error between them was only 0.101 kcal/mol for the corrected HF and 0.026 kcal/mol for the corrected B3LYP. Therefore, the approach is an accurate and effective method for introducing the electron correlation effects in QM/MM MD simulations. The physical foundation of the EACs is that the interactions of a chemical bond with the electric field possess an approximate linear





**Figure 7.** Distribution of the calculated interaction energies between the QM (CO molecule) and the MM (the surroundings) subsystems at the (A) HF/aug-cc-pVTZ/Amberff99SB, (B) B3LYP/aug-cc-pVTZ/Amberff99SB, and (C) MP2/aug-cc-pVTZ/Amberff99SB QM/MM levels as a function of the electric fields along the CO bond. The interaction energies exhibit a linear relationship with the electric fields. The green and blue lines represent the regression curves of the interaction energies relative to the positive and negative electric fields.



**Figure 8.** Correlation of the calculated QM/MM interaction energies (in kcal/mol) between the (A) HF/aug-cc-pVTZ/Amberff99SB, (B) B3LYP/aug-cc-pVTZ/Amberff99SB, and MP2/aug-cc-pVTZ/Amberff99SB QM/MM calculations. The corrected-HF and corrected-B3LYP mean that the obtained QM/MM interaction energies at the HF/aug-cc-pVTZ/Amberff99SB and B3LYP/aug-cc-pVTZ/Amberff99SB QM/MM levels were corrected with the obtained EFEAC charges. The blue line represents the strict correlation curve.

Ratio	HF	B3LYP	MP2
$B_1/B_2$	1.09	1.54	3.45
$^a k_1/k_2$	0.83	1.80	3.02

**Table 1.** The obtained  $B_1/B_2$  ratio from the QM/MM MD simulations at the HF/aug-cc-pVTZ/Amberff99SB, B3LYP/aug-cc-pVTZ/Amberff99SB, and MP2/aug-cc-pVTZ/Amberff99SB QM/MM levels. Here,  $k_1/k_2$  is the ratio of the slope of regression curves for the QM/MM interaction energies relative to the electric fields, and  $k_1$  and  $k_2$  are the slopes that correspond to the green and blue lines in Fig. 7, respectively. <sup>a</sup>Unsigned values are adopted.

relationship. Although this was only tested in terms of the sample MbCO system, many chemical bond groups (e.g., -OH, -NH, -SH, -CH, -C≡N, and >C=O) on various small molecules displayed a linear response to an electric field, which has been experimentally confirmed<sup>23,25,98,99</sup>. Therefore, the approach has a potential application for tackling various complicated QM/MM systems.

## Conclusions

In this work, QM/MM MD simulations of the ratio of the two B states of the MbCO system were performed to investigate how we can accurately model the QM/MM interactions with a QM/MM model. The calculated  $B_1/B_2$  ratio was significantly sensitive to the size of the basis sets and was close to convergent after a basis set of aug-cc-pVTZ was reached. The introduction of both the polarization function and the diffuse function in the basis sets had a significant impact on the simulations. By comparing the obtained results at the HF/aug-cc-pVTZ/Amberff99SB, B3LYP/aug-cc-pVTZ/Amberff99SB, and MP2/aug-cc-pVTZ/Amberff99SB QM/MM levels, we demonstrated that the inclusion of the electron correlation effects is essential for accurately modeling the interactions between the QM and MM subsystems in QM/MM MD simulations.

The calculated electrostatic polarization energies based on the classical point charge model between the CO and the surroundings exhibited an exact quadratic function relationship with the electric fields along the CO bond. Meanwhile, the QM/MM interactions between the CO and the surroundings exhibited an asymmetric distribution in the positive and negative electric fields and presented a linear relationship with the electric fields. The QM/MM interactions could be described by a simple physical model of a dipole with constant moment under the action of the electric fields according to its statistical distribution relative to the electric fields. The treatment provides us with an accurate and effective approach to account for the electron correlation effects in QM/MM MD simulations. Given the limitation of our theoretical approach, the current study provided strong but incomplete evidence that there (statistically) exists an inevitable correlation between QM/MM interaction energies and the electric fields along the chemical bonds, which, nonetheless, could be utilized to account for the electron correlation effects in QM/MM studies.

Received: 22 October 2019; Accepted: 5 May 2020;

Published online: 22 May 2020

## References

1. Warshel, A. & Levitt, M. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **103**, 227–249 (1976).
2. Field, M. J., Bash, P. A. & Karplus, M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comput. Chem.* **11**, 700–733 (1990).
3. Bakowies, D. & Thiel, W. Hybrid models for combined quantum mechanical and molecular mechanical approaches. *J. Phys. Chem.* **100**, 10580–10594 (1996).
4. Monard, G. & Merz, K. M. Combined quantum mechanical/molecular mechanical methodologies applied to biomolecular systems. *Accounts Chem. Res.* **32**, 904–911 (1999).
5. Gao, J. & Truhlar, D. G. Quantum mechanical methods for enzyme kinetics. *Annu. Rev. Phys. Chem.* **53**, 467–505 (2002).
6. Senn, H. M. & Thiel, W. QM/MM methods for biomolecular systems. *Angew. Chem., Int. Ed.* **48**, 1198–1229 (2009).
7. van der Kamp, M. W. & Mulholland, A. J. Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology. *Biochemistry* **52**, 2708–2728 (2013).
8. Brunk, E. & Rothlisberger, U. Mixed quantum mechanical/molecular mechanical molecular dynamics simulations of biological systems in ground and electronically excited states. *Chem. Rev.* **115**, 6217–6263 (2015).
9. Ryde, U. QM/MM calculations on proteins. *Methods Enzymol.* **577**, 119–158 (2016).
10. Boulanger, E. & Harvey, J. N. QM/MM methods for free energies and photochemistry. *Curr. Opin. Struct. Biol.* **49**, 72–76 (2018).
11. Morzan, U. N. *et al.* Spectroscopy in complex environments from QM–MM simulations. *Chem. Rev.* **118**, 4071–4113 (2018).
12. Ganguly, A., Boulanger, E. & Thiel, W. Importance of MM Polarization in QM/MM Studies of Enzymatic Reactions: Assessment of the QM/MM Drude Oscillator Model. *J. Chem. Theory Comput.* **13**, 2954–2961 (2017).
13. Seabra, A. E. *et al.* Implementation of the SCC-DFTB method for hybrid QM/MM simulations within the Amber molecular dynamics package. *J. Phys. Chem. A* **111**, 5655–5664 (2007).
14. Woodcock, H. L. 3rd *et al.* Interfacing Q-Chem and CHARMM to perform QM/MM reaction path calculations. *J. Comput. Chem.* **28**, 1485–1502 (2007).
15. Loco, D. *et al.* A QM/MM approach using the AMOEBA polarizable embedding: from ground state energies to electronic excitations. *J. Chem. Theory Comput.* **12**, 3654–3661 (2016).
16. Szczyk, B., Mulholland, A. J., Ranaghan, K. E. & Sokalski, W. A. Differential transition-state stabilization in enzyme catalysis: quantum chemical analysis of interactions in the chorismate mutase reaction and prediction of the optimal catalytic field. *J. Am. Chem. Soc.* **126**, 16148–16159 (2004).
17. Warshel, A. & Russell, S. T. Calculations of electrostatic interactions in biological systems and in solutions. *Q. Rev. Biophys.* **17**, 283–422 (1984).
18. Davis, M. E. & McCammon, J. A. Electrostatics in biomolecular structure and dynamics. *Chem. Rev.* **90**, 509–521 (1990).
19. Honig, B. & Nicholls, A. Classical electrostatics in biology and chemistry. *Science* **268**, 1144–1149 (1995).
20. Perutz, M. Electrostatic effects in proteins. *Science* **201**, 1187–1191 (1978).
21. Matthew, J. B. Electrostatic effects in proteins. *Annu. Rev. Biophys. Biophys. Chem.* **14**, 387–417 (1985).
22. Wang, X. & He, X. An Ab Initio QM/MM Study of the Electrostatic Contribution to Catalysis in the Active Site of Ketosteroid Isomerase. *Molecules* **23**, 2410 (2018).
23. Fried, S. D., Bagchi, S. & Boxer, S. G. Extreme electric fields power catalysis in the active site of ketosteroid isomerase. *Science* **346**, 1510–1514 (2014).
24. Warshel, A. & Bora, R. P. Perspective: Defining and quantifying the role of dynamics in enzyme catalysis. *J. Chem. Phys.* **144**, 180901 (2016).
25. Fried, S. D. & Boxer, S. G. Electric fields and enzyme catalysis. *Annu. Rev. Biochem.* **86**, 387–415 (2017).
26. Wu, Y. & Boxer, S. G. A critical test of the electrostatic contribution to catalysis with noncanonical amino acids in ketosteroid isomerase. *J. Am. Chem. Soc.* **138**, 11890–11895 (2016).
27. Wang, X., Zhang, J. Z. & He, X. Quantum mechanical calculation of electric fields and vibrational Stark shifts at active site of human aldose reductase. *J. Chem. Phys.* **143**, 184111 (2015).
28. Collins, M. A. & Bettens, R. P. Energy-based molecular fragmentation methods. *Chem. Rev.* **115**, 5607–5642 (2015).
29. Liu, J., Zhu, T., Wang, X., He, X. & Zhang, J. Z. Quantum fragment based ab initio molecular dynamics for proteins. *J. Chem. Theory Comput.* **11**, 5897–5905 (2015).
30. Raghavachari, K. & Saha, A. Accurate composite and fragment-based quantum chemical models for large molecules. *Chem. Rev.* **115**, 5643–5677 (2015).
31. Wang, X. *et al.* A quantum mechanical computational method for modeling electrostatic and solvation effects of protein. *Sci. Rep.* **8**, 5475 (2018).
32. Riahi, S. & Rowley, C. N. The CHARMM–TURBOMOLE interface for efficient and accurate QM/MM molecular dynamics, free energies, and excited state properties. *J. Comput. Chem.* **35**, 2076–2086 (2014).

33. Schwenk, C. F. & Rode, B. M. Influence of electron correlation effects on the solvation of Cu<sup>2+</sup>. *J. Am. Chem. Soc.* **126**, 12786–12787 (2004).
34. Somtua, T. & Tongraar, A. Correlation effects on the structure and dynamics of the H<sub>3</sub>O<sup>+</sup> hydrate: B3LYP/MM and MP2/MM MD simulations. *Phys. Chem. Chem. Phys.* **13**, 16190–16196 (2011).
35. Uddin, N., Choi, T. H. & Choi, C. H. Direct Absolute pK a Predictions and Proton Transfer Mechanisms of Small Molecules in Aqueous Solution by QM/MM-MD. *J. Phys. Chem. B* **117**, 6269–6275 (2013).
36. Xenides, D., Randolph, B. & Rode, B. Hydrogen bonding in liquid water: An ab initio QM/MM MD simulation study. *J. Mol. Liq.* **123**, 61–67 (2006).
37. Prasetyo, N. & Hofer, T. S. Carbon dioxide in liquid ammonia: An ab initio Quantum Mechanical/Molecular Mechanical Molecular Dynamics Thermodynamic Integration (QM/MM MD TI) simulation study on structure, dynamics and thermodynamics of solvation. *J. Mol. Liq.* **304** (2020).
38. Prasetyo, N. & Hofer, T. S. Structure, Dynamics, and Hydration Free Energy of Carbon Dioxide in Aqueous Solution: A Quantum Mechanical/Molecular Mechanics Molecular Dynamics Thermodynamic Integration (QM/MM MD TI) Simulation Study. *J. Chem. Theory Comput.* **14**, 6472–6483 (2018).
39. Mondal, A. & Balasubramanian, S. Understanding SO<sub>2</sub> Capture by Ionic Liquids. *J. Phys. Chem. B* **120**, 4457–4466 (2016).
40. Del Ben, M., Hutter, J. & VandeVondele, J. Forces and stress in second order Moller-Plesset perturbation theory for condensed phase systems within the resolution-of-identity Gaussian and plane waves approach. *J. Chem. Phys.* **143** (2015).
41. Kaledin, M. & Adedeji, D. T. Driven Molecular Dynamics Studies of the Shared Proton Motion in the H<sub>5</sub>O<sub>2</sub><sup>+</sup>-center dot Ar Cluster: The Effect of Argon Tagging and Deuteration on Vibrational Spectra. *J. Phys. Chem. A* **119**, 1875–1884 (2015).
42. Duignan, T. T., Parsons, D. F. & Ninham, B. W. Ion Interactions with the Air-Water Interface Using a Continuum Solvent Model. *J. Phys. Chem. B* **118**, 8700–8710 (2014).
43. Yang, Z., Lin, H., Gui, T., Zhou, R. F. & Chen, X. S. Infrared spectroscopy of N-methylacetamide in water from high-level QM/MM calculations. *Chin. Chem. Lett.* **25**, 107–110 (2014).
44. Cohen, A. J., Mori-Sánchez, P. & Yang, W. Challenges for density functional theory. *Chem. Rev.* **112**, 289–320 (2011).
45. Grimme, S., Hansen, A., Brandenburg, J. G. & Bannwarth, C. Dispersion-corrected mean-field electronic structure methods. *Chem. Rev.* **116**, 5105–5154 (2016).
46. Ranaghan, K. E. *et al.* Projector-based embedding eliminates density functional dependence for QM/MM calculations of reactions in enzymes and solution. *J. Chem. Inf. Model.* **59**, 2063–2078 (2019).
47. Lawan, N., Ranaghan, K. E., Manby, F. R. & Mulholland, A. J. Comparison of DFT and ab initio QM/MM methods for modelling reaction in chorismate synthase. *Chem. Phys. Lett.* **608**, 380–385 (2014).
48. van der Kamp, M. W., Perruccio, F. & Mulholland, A. J. High-level QM/MM modelling predicts an arginine as the acid in the condensation reaction catalysed by citrate synthase. *Chem. Commun.*, 1874–1876 (2008).
49. van der Kamp, M. W., Zurek, J., Manby, F. R., Harvey, J. N. & Mulholland, A. J. Testing high-level QM/MM methods for modeling enzyme reactions: acetyl-CoA deprotonation in citrate synthase. *J. Phys. Chem. B* **114**, 11303–11314 (2010).
50. Wang, S., Hu, P. & Zhang, Y. Ab initio quantum mechanical/molecular mechanical molecular dynamics simulation of enzyme catalysis: The case of histone lysine methyltransferase SET7/9. *J. Phys. Chem. B* **111**, 3758–3764 (2007).
51. Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* **27**, 1787–1799 (2006).
52. Grimme, S., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **132**, 154104 (2010).
53. Grimme, S., Ehrlich, S. & Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **32**, 1456–1465 (2011).
54. Senthilkumar, K. *et al.* Analysis of polarization in QM/MM modelling of biologically relevant hydrogen bonds. *J. R. Soc. Interface* **5**, 207–216 (2008).
55. Barends, T. R. *et al.* Direct observation of ultrafast collective motions in CO myoglobin upon ligand dissociation. *Science* **350**, 445–450 (2015).
56. Levantino, M. *et al.* Ultrafast myoglobin structural dynamics observed with an X-ray free-electron laser. *Nat. Commun.* **6**, 6772 (2015).
57. Olson, J. S. & Phillips, G. N. Jr. Myoglobin discriminates between O<sub>2</sub>, NO, and CO by electrostatic interactions with the bound ligand. *J. Biol. Inorg. Chem.* **2**, 544–552 (1997).
58. Spiro, T. G. & Kozlowski, P. M. Is the CO adduct of myoglobin bent, and does it matter? *Accounts Chem. Res.* **34**, 137–144 (2001).
59. Phillips, G. N. *et al.* (ACS Publications, 1999).
60. Springer, B. A., Sligar, S. G., Olson, J. S. & Phillips, G. N. J. Mechanisms of ligand recognition in myoglobin. *Chem. Rev.* **94**, 699–714 (1994).
61. Merchant, K. A. *et al.* Myoglobin-CO conformational substate dynamics: 2D vibrational echoes and MD simulations. *Biophys. J.* **82**, 3277–3288 (2002).
62. Maragliano, L., Cottone, G., Ciccotti, G. & Vanden-Eijnden, E. Mapping the network of pathways of CO diffusion in myoglobin. *J. Am. Chem. Soc.* **132**, 1010–1017 (2009).
63. Lim, M., Jackson, T. A. & Anfinsen, P. A. Binding of CO to myoglobin from a heme pocket docking site to form nearly linear Fe-CO. *Science* **269**, 962–966 (1995).
64. Banushkina, P. & Meuwly, M. Free-energy barriers in MbCO rebinding. *J. Phys. Chem. B* **109**, 16911–16917 (2005).
65. Bossa, C. *et al.* Molecular dynamics simulation of sperm whale myoglobin: effects of mutations and trapped CO on the structure and dynamics of cavities. *Biophys. J.* **89**, 465–474 (2005).
66. Kiyota, Y. *et al.* Theoretical study of CO escaping pathway in myoglobin with the 3D-RISM theory. *J. Am. Chem. Soc.* **131**, 3852–3853 (2009).
67. Schmidt, M. *et al.* Ligand migration pathway and protein dynamics in myoglobin: a time-resolved crystallographic study on L29W MbCO. *Proc. Natl. Acad. Sci. USA* **102**, 11704–11709 (2005).
68. Lapelosa, M. & Abrams, C. F. A computational study of water and CO migration sites and channels inside myoglobin. *J. Chem. Theory Comput.* **9**, 1265–1271 (2013).
69. Wang, P.-H., Bruschi, M., De Gioia, L. & Blumberger, J. Uncovering a dynamically formed substrate access tunnel in carbon monoxide dehydrogenase/acetyl-CoA synthase. *J. Am. Chem. Soc.* **135**, 9493–9502 (2013).
70. Yu, T.-Q., Lapelosa, M., Vanden-Eijnden, E. & Abrams, C. F. Full kinetics of CO entry, internal diffusion, and exit in myoglobin from transition-path theory simulations. *J. Am. Chem. Soc.* **137**, 3041–3050 (2015).
71. Lim, M., Jackson, T. A. & Anfinsen, P. A. Mid-infrared vibrational spectrum of CO after photodissociation from heme: Evidence for a ligand docking site in the heme pocket of hemoglobin and myoglobin. *J. Chem. Phys.* **102**, 4355–4366 (1995).
72. Lim, M., Jackson, T. A. & Anfinsen, P. A. Ultrafast rotation and trapping of carbon monoxide dissociated from myoglobin. *Nat. Struct. Biol.* **4**, 209 (1997).
73. Lim, M., Jackson, T. A. & Anfinsen, P. A. Nonexponential protein relaxation: dynamics of conformational change in myoglobin. *Proc. Natl. Acad. Sci. USA* **90**, 5801–5804 (1993).
74. Park, E. S., Andrews, S. S., Hu, R. B. & Boxer, S. G. Vibrational stark spectroscopy in proteins: A probe and calibration for electrostatic fields. *J. Phys. Chem. B* **103**, 9813–9817 (1999).

75. Alben, J. *et al.* Infrared spectroscopy of photodissociated carboxymyoglobin at low temperatures. *Proc. Natl. Acad. Sci. USA* **79**, 3744–3748 (1982).
76. Tetreau, C., Blouquit, Y., Novikov, E., Quiniou, E. & Lavalette, D. Competition with xenon elicits ligand migration and escape pathways in myoglobin. *Biophys. J.* **86**, 435–447 (2004).
77. Ostermann, A., Waschipyk, R., Parak, F. G. & Nienhaus, G. U. Ligand binding and conformational motions in myoglobin. *Nature* **404**, 205 (2000).
78. Plattner, N. & Meuwly, M. The role of higher CO-multipole moments in understanding the dynamics of photodissociated carbonmonoxide in myoglobin. *Biophys. J.* **94**, 2505–2515 (2008).
79. Nutt, D. R. & Meuwly, M. Theoretical investigation of infrared spectra and pocket dynamics of photodissociated carbonmonoxy myoglobin. *Biophys. J.* **85**, 3612–3623 (2003).
80. Wang, X.-W., Zhang, J. Z. & He, X. Ab initio Quantum Mechanics/Molecular Mechanics Molecular Dynamics Simulation of CO in the Heme Distal Pocket of Myoglobin. *Chin. J. Chem. Phys.* **30**, 705 (2018).
81. 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* **7**, 2061–2063 (2006).
82. Cornell, W. D. *et al.* A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**, 5179–5197 (1995).
83. Duan, Y. *et al.* A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **24**, 1999–2012 (2003).
84. Hobza, P., Sponer, J. & Polasek, M. H-bonded and stacked DNA base pairs: Cytosine dimer. An ab initio second-order Moeller-Plesset study. *J. Am. Chem. Soc.* **117**, 792–798 (1995).
85. Hobza, P. & Sponer, J. Structure, energetics, and dynamics of the nucleic acid base pairs: nonempirical ab initio calculations. *Chem. Rev.* **99**, 3247–3276 (1999).
86. Kuriyan, J., Wilz, S., Karplus, M. & Petsko, G. A. X-ray structure and refinement of carbon-monoxo (Fe II)-myoglobin at 1.5 Å resolution. *J. Mol. Biol.* **192**, 133–154 (1986).
87. Case, D. A. *et al.* The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688 (2005).
88. D.A., G. Heme Parameters, University of California, Davis: Davis, CA (1984).
89. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341 (1977).
90. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
91. Pastor, R. W., Brooks, B. R. & Szabo, A. An analysis of the accuracy of Langevin and molecular dynamics algorithms. *Mol. Phys.* **65**, 1409–1419 (1988).
92. Frisch M. J. *et al.* GAUSSIAN 09, Revision B.01, Wallingford, CT: Gaussian, Inc. (2010).
93. Schutz, C. N. & Warshel, A. What are the dielectric “constants” of proteins and how to validate electrostatic models? *Proteins* **44**, 400–417 (2001).
94. Schotte, F. *et al.* Watching a protein as it functions with 150-ps time-resolved x-ray crystallography. *Science* **300**, 1944–1947 (2003).
95. Chu, K. *et al.* Structure of a ligand-binding intermediate in wild-type carbonmonoxy myoglobin. *Nature* **403**, 921 (2000).
96. Muentner, J. Electric dipole moment of carbon monoxide. *J. Mol. Spectrosc.* **55**, 490–491 (1975).
97. Meerts, W., De Leeuw, F. & Dymanus, A. Electric and magnetic properties of carbon monoxide by molecular-beam electric-resonance spectroscopy. *Chem. Phys.* **22**, 319–324 (1977).
98. Saggi, M., Levinson, N. M. & Boxer, S. G. Direct measurements of electric fields in weak OH... $\pi$  hydrogen bonds. *J. Am. Chem. Soc.* **133**, 17414–17419 (2011).
99. Saggi, M., Levinson, N. M. & Boxer, S. G. Experimental quantification of electrostatics in X–H... $\pi$  hydrogen bonds. *J. Am. Chem. Soc.* **134**, 18986–18997 (2012).

## Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant Nos. 21703206), Zhejiang Provincial Natural Science Foundation (Grant No. LY17B030008). We also thank the Supercomputer Center of East China Normal University for providing us computational time.

## Author contributions

In this work, Xianwei Wang prepared all the figures and wrote the main manuscript text. Xianwei Wang and Maoyou Yang provided ideas. Chenhui Lu helped with data analysis and all authors reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-65475-2>.

**Correspondence** and requests for materials should be addressed to X.W. or M.Y.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020