CHALLENGES IN COMPUTATIONAL ENZYMOLOGY

EDITED BY: Vicent Moliner and Fahmi Himo

frontiers Research Topics

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

CHALLENGES IN COMPUTATIONAL ENZYMOLOGY

Topic Editors:
Vicent Moliner, University of Jaume I, Spain
Fahmi Himo, Stockholm University, Sweden

Table of Contents

Editorial: Challenges in Computational Enzymology

Vicente Moliner[1*†] and Fahmi Himo[2*†]

[1] Departamento de Química Física y Analítica, Universidad Jaume I, Castellón de la Plana, Spain, [2] Arrhenius Laboratory, Department of Organic Chemistry, Stockholm University, Stockholm, Sweden

Editorial on the Research Topic

Challenges in Computational Enzymology

Living organisms utilize enzymes to accelerate and control chemical reactions, and the attraction of these biological catalysts lies not only in their extraordinary efficiency, but also in their selectivity and their ability to function at mild conditions. Because of their involvement in fundamental biological processes, enzymes are common targets for the development of drug molecules in the pharmaceutical industry. In addition, enzymes are today increasingly employed as synthetic tools for industrial production of high value chemicals. It is therefore of great importance to understand in depth how these fascinating molecular machines perform their reactions, both from a fundamental scientific point of view and also for technical applications. To this end, computational approaches have in recent years become an indispensable tool in this endeavor. A number of powerful methodologies have been developed that have allowed for breakthroughs in the mechanistic understanding of enzymes. The exponential growth of computer power has, of course, been of key importance here. The field has also benefited greatly from the fruitful collaborations between theoreticians and experimentalists, where results derived from the simulations are used to interpret the experiments, and also to predict the behavior of natural and altered enzymes.

This Research Topic is concerned with the general theme of computational enzymology. It consists of 13 contributions covering various aspects of the field, spanning from benchmarking of methods to high-end applications and reviews of recent work.

In a review paper by Wei et al., the authors summarized their work on the modeling of metalloenzymes with both the quantum chemical cluster approach and the hybrid quantum mechanics/molecular mechanics (QM/MM) method. This review discussed recent progress in the computational understanding of mechanisms and various selectivities, such as chemo-, regio-, and stereoselectivity, in this important class of enzymes. In a more focused mini-review, Caddell Haatveit et al. described a computational protocol for modeling cytochrome P450 and exploring its reactivity and selectivity. The approach leads to predictions of enzyme variants that are more efficient and selective. Rovaletti et al. discussed in another mini-review the modeling of an illustrative example of redox-active enzymes, namely Mo/Cu-dependent CO dehydrogenase. The challenges concerned with both the DFT and the QM/MM approaches are highlighted and promising future directions are mentioned. In the paper of Siegbahn and Blomberg, problems associated with the study of mechanisms of redox processes in enzymes by DFT methods were discussed. An alternative systematic approach was proposed, where the amount of exact exchange in the B3LYP functional can be used as a parameter, rather than using a large number of different functionals. The limitations of DFT methods to study catalytic reaction mechanisms involving Fe(III)/Fe(II) oxidation states were discussed in the paper of Listyarini et al. The benchmark study, based on QM and QM/MM computational methods, provides guidelines to estimate the inaccuracies coming from the density functionals and to choose the most appropriate ones.

On the applications side, a number of very interesting papers are included in the Research Topic. Marques et al. presented a molecular dynamics study, employing metadynamics and adaptive sampling techniques, to estimate the rates for unbinding of the product molecule from the enzyme haloalkane dehalogenase DhaA and its mutant DhaA31. The simulations provided insights into the energetic bottlenecks in the rate-limiting unbinding process, which is of great help for the design of improved biocatalysts. Alonso-Gil et al. unraveled the hydrolytic reaction mechanism of *Neisseria polysaccharea* amylosucrase (NpAS), a member of GH13 family, based on QM(DFT)/MM metadynamics simulations. The results provide an atomistic picture of the active site reorganization along the catalytic double-displacement reaction, consistent with the general conformational itinerary observed for α-glucosidases. Kumar et al. investigated the binding modes of two inhibitors of the two metalloproteases (human JAMM deubiquitinylases Rpn11 and CSN5) using MD simulations and binding energy analysis and found that it was necessary to include larger heterodimeric protein-protein complexes in order to avoid unrealistic structural changes.

Alonso-Cotchico et al. combined force field-based techniques with a variety of trajectory convergence analyses to study the effect of cofactor binding on the conformational plasticity of the Lactococcal multidrug resistance regulator (LmrR) protein, a crucial aspect in order to design efficient artificial metalloenzymes. Martí et al. used QM/MM calculations to explore the mechanisms of the inactivation process of kanamycin A catalyzed by 4′-O-nucleotidyltransferase. Free energy perturbation techniques were employed and primary and secondary ^{18}O kinetic isotope effects were computed and compared to available experiments to elucidate the detailed reaction mechanism. Romero-Téllez et al. used QM(DFT)/MM calculations to compare the glycosylation, hydrolysis and transglycosylation steps catalyzed by wild type *Thermus*

thermophilus β-glycosidase. They showed how the molecular understanding of similarities and differences between hydrolysis and transglycosylation steps may be of help in the design of new biocatalysts for glycan synthesis. Prejanò et al. described based on QM and QM/MM calculations the reaction mechanism of the inhibition mechanism of hydrolyzed piperlongumine (hPL), an anticancer compound whose activity is related to the inhibition of human glutathione transferase of pi class (GSTP1). Finally, Timmins et al. studied the reaction mechanism of the HctB, a non-heme iron halogenase, using a combination of MM, MD, QM/MM, and DFT techniques. The effect of the substrate position in the active site on the halogenation vs. hydroxylation selectivity is discussed and comparison is made to other non-heme iron enzymes.

We believe that the new insights presented in these contributions will lead to advances in both fundamental understanding and practical applications in, for example, the design of new drug compounds and the developments of new biocatalysts for the chemical industry. We wish finally to thank all the contributors for their excellent work and all the reviewers for their comments that helped to improve the contents.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

Check for updates

Computational Studies on the Inhibitor Selectivity of Human JAMM Deubiquitinylases Rpn11 and CSN5

Vikash Kumar [1,2], Michael Naumann [1] and Matthias Stein [2]*

[1] Institute of Experimental and Internal Medicine, Medical Faculty, Otto von Guericke University, Magdeburg, Germany,
[2] Molecular Simulations and Design Group, Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany

Deubiquitinylases (DUBs) are highly specialized enzymes which are responsible for removal of covalently attached ubiquitin(s) from the targeted proteins. DUBs play an important role in maintaining the protein homeodynamics. Recently, DUBs have emerged as novel therapeutic targets in cancer, inflammation, diabetes, and neurodegeneration. Among the different families of DUBs, the metalloprotease group or JAB1/MOV34/MPR1 (JAMMs) proteases are unique in terms of catalytic mechanism. JAMMs exhibit a Zn^{2+}-dependent deubiquitinylase activity. Within the JAMM family, deubiquitinylases Rpn11 and CSN5 are constituents of large bimolecular complexes, namely the 26S proteasome and COP9 signalosome (CSN), respectively. Rpn11 and CSN5 are potential drug targets in cancer and selective inhibitors of both proteins have been reported in the literature. However, the selectivity of JAMM inhibitors (capzimin for RPN11 and CSN5i-3 for CSN5) has not been structurally resolved yet. In the present work, we have explored the binding modes of capzimin and CSN5i-3 and rationalize their selectivity for Rpn11 and CSN5 targets. We found that capzimin interacts with the active site Zn^{+2} of Rpn11 in a bidentate manner and also interacts with the residues in the distal ubiquitin binding site. MD simulations studies and binding energy analysis revealed that the selective binding of the inhibitors can be only explained by the consideration of larger heterodimeric complexes of Rpn11 (Rpn8-Rpn11) and CSN5 (CSN5-CSN6). Simulation of these protein-protein complexes is necessary to avoid unrealistic large conformational changes. The selective binding of inhibitors is mainly governed by residues in the distal ubiquitin binding site. This study demonstrates that selective inhibitor binding design for Rpn11 and CSN5 JAMM proteases requires consideration of heterodimeric protein-protein target structures.

Keywords: computational drug design, deubiquitinase (DUB), selectivities, protein-protein interaction, molecular dynamics (MD), ligand binding

INTRODUCTION

Ubiquitinylation is one of the major post-translational modifications of proteins. Ubiquitin is a 76 amino acids protein which is covalently attached to the lysine residue of the substrate by consecutive action of three enzymes i.e., activating (E1), conjugating (E2), and ligating (E3) enzymes (Nandi et al., 2006). Ubiquitinylated proteins participate in various cellular processes. In

order to remove the ubiquitinylation mark, separate families of proteins exist and they have been named as deubiquitinylases. In contrast to ubiquitinylation, deubiquitinylation requires action of a single enzyme only (Reyes-Turcu et al., 2009). DUBs have been classified into 5 families i.e., ubiquitin specific proteases (USPs), ovarian tumor proteases (OTUs), Machado-Joseph disease proteases (MJDs), and JAB1/MOV34/MPR1 (JAMM) proteases (Komander et al., 2009; Mevissen and Komander, 2017). Except for JAMM proteases which have Zn^{2+} in the catalytic site, all other DUBs are cysteine proteases. Members of JAMM family have a common Zn^{2+} binding motif which contains three conserved residues (one aspartate and two histidines) (Berndt et al., 2002).

The human genome encodes for 14 JAMM proteins and only seven of them have the full set of conserved residues required for Zn^{+2} binding (Shrestha et al., 2014). Six JAMM proteins have deubiquitinylase activities which are AMSH, AMSH-LP, BRCC36, CSN5, MYSM1, and Rpn11 (also known as POH1 or PSMD14) (Shrestha et al., 2014). CSN5 is a component of the CSN and it also possesses nedylation activity (Lee et al., 2011; Echalier et al., 2013). Monomeric CSN5 shows no DUB activity and requires the presence of its non-active binding partner (CSN6) to gain full activity (Echalier et al., 2013). Rpn11 is a component of the 19S regulatory subunit of the large proteasome complex and only displays deubiquitinylase activity (Verma et al., 2002). Similar to the CSN5, a monomeric Rpn11 is not catalytically active (Yao and Cohen, 2002; Pathare et al., 2014).

Recently, Rpn11 has emerged as a potential drug target in human cancers (Li et al., 2017). Rpn11 is responsible for deubiquitinylation of proteasomal substrates (Verma et al., 2002) Inhibition of Rpn11 has been reported to overcome bortezomib resistance and induce apoptosis in multiple myeloma cells (Song et al., 2017). Capzimin is a potent and selective inhibitor of Rpn11 (**Figure 1A**) and it has been suggested that capzimin chelates the Zn^{2+} ion in the active site of Rpn11 (Li et al., 2017; Perez et al., 2017). Capzimin shows an 80-fold selectivity for Rpn11 over CSN5 (Li et al., 2017). Similarly, CSN5i-3, a potent inhibitor of CSN5 (**Figure 1B**), shows 10,000-fold selectivity for CSN5 over Rpn11. However, the structural basis of these inhibitor selectivities is not known. From the drug design perspective, it is important to rationalize the binding mode of capzimin and structural elements responsible for imparting DUB selectivity. In the present study, we used a workflow of molecular docking, refinement and ligand binding stability studies by molecular dynamics simulations and binding energy calculations to investigate the structural basis of selective inhibition of Rpn11 and CSN5.

MATERIALS AND METHODS
Monomer and Heterodimeric Target Structures

Cryo-EM structure of human 26S proteasome (PDB_code: 5GJR) was obtained from PDB (Huang et al., 2016). With the exception of the chains U (Rpn8 or PSMD7) and V (Rpn11 or PSMD14), other were deleted. The Rpn-Rpn11 heterodimer in

FIGURE 1 | 2D structure of the Rpn11 selective inhibitor, capzimin **(A)** and the CSN5 selective inhibitor, CSN5i-3 **(B)**. Both are targeting JAMMs but their selectivity of action cannot be rationalized.
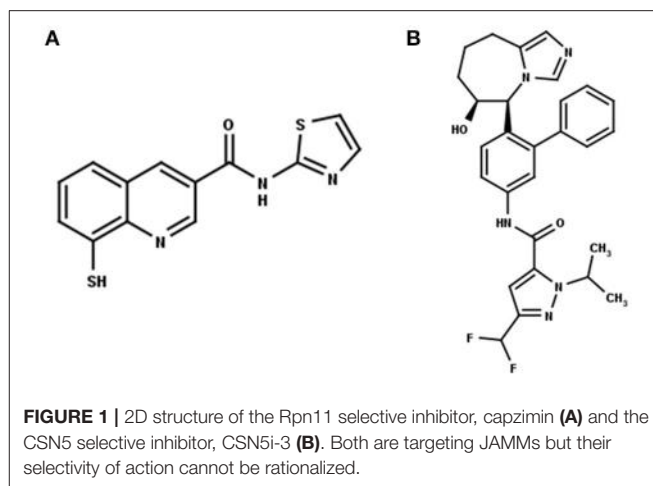
the unprocessed form is shown in **Figure S1A**. We noticed that RPN11 does not have Zn^{2+} ion in the catalytic site. Rpn8 and Rpn11 were processed separately. We deleted the C-terminal region of Rpn8 (182–295) and Rpn11 (210–316) since they are not relevant for enzymatic catalysis. With the help of UCSF Chimera (Pettersen et al., 2004) interface to MODELLER (Sali and Blundell, 1993; Webb and Sali, 2017), missing regions in Rpn8 (143–151) and Rpn11 (164–189) were modeled as loops (**Figure 2A** and **Figure S1**). Ins-1 loop (76–88) of the Rpn11 was initially in the closed conformation hence we generated 10 alternative conformations of Ins-1 loop (**Figure S1B**). We selected a loop conformation which was pointing away from the catalytic site (**Figure 2A**). The final Rpn8-Rpn11 heterodimer model is shown is **Figure 3A** and **Figure S1C**. A Zn^{2+} ion was transferred to the catalytic site of Rpn11 by aligning the Rpn11 structure with the crystal structure of CSN5. After transferring Zn^{2+} ion to the Rpn11, we used both the Rpn11 monomer (**Figure 2A**) and the Rpn8-Rpn11 heterodimer (**Figure 3A**) structures for subsequent analysis.

For the CSN5 monomer (**Figure 2B**), the recently solved crystal structure of CSN5 with CSN5i-3 (PDB_code: 5JOG) (Schlierf et al., 2016) was considered. We modeled the missing Ins-1 loop region (**Figure S2A**) by analogy to Rpn11 (see above) and out of 5 loop conformations (**Figure S2B**), a loop conformation which did not show any steric clash and contact with the bound inhibitor was considered (**Figure S2C**). The structure of the CSN5-CSN6 heterodimer (**Figure S3A**) was extracted from the crystal structure of human COP9 signalasome (PDB_code: 4D10) (Lingaraju et al., 2014) C-terminal regions of CSN5 (258–333) and CSN6 (215–316) were also deleted since they are not involved in catalysis and heterodimer formation can take place without them. Here, Ins-1 loop (98–113) of CSN5 was found to be in a closed conformation. We generated 10 alternative conformations of Ins-1 loop (**Figure 2C** and **Figure S3B**). We eventually selected a CSN5 model in which the Ins-1 loop is in an open conformation and does not obstruct ligand access to the catalytic site. The processed CSN5-CSN6 heterodimer is shown in **Figure 3B** and **Figure S3C**.
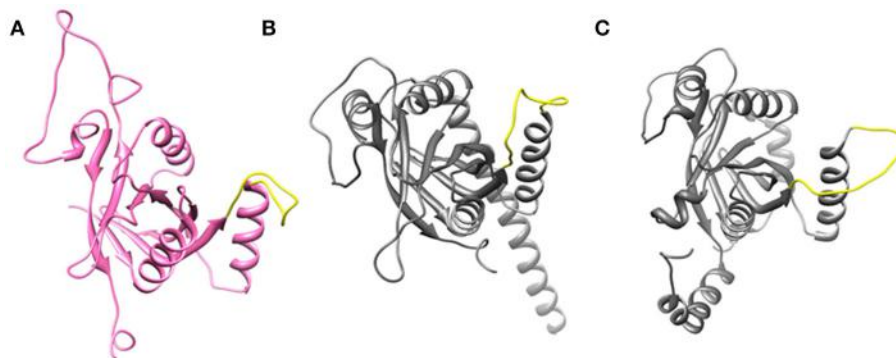
FIGURE 2 | Open conformation of Ins-1 loop (yellow) in **(A)** Rpn11 monomer **(B)** CSN5 (monomeric state) and **(C)** CSN5 (as in heterodimeric state).

Inhibitor Positioning

The 3D structure of capzimin was generated with the help of MarvinSketch program (https://chemaxon.com/products/marvin). The sulfur thiolate has a negative charge. Ligand docking of capzimin was carried out with AUTODOCK4.2 (Morris et al., 2009) Kollman charges were assigned to the atoms of Rpn11 and CSN5. On Zn, +2 charge was manually assigned. Partial charges on capzimin were calculated using the SwissParam server (Zoete et al., 2011). For both Rpn11 and CSN5, a grid was centered on the catalytic Zn^{2+}. To enclose the binding site in RPN11, the size of the 3D grid was set to 46, 50, and 52 grid points in x, y and z directions, respectively, with a default spacing of 0.375 Å. In case of CSN5, binding site was enclosed in a grid consisting of 48, 50, and 64 grid points in x, y and z directions, respectively. In each case, the Lamarckian Genetic Algorithm (LGA) was used to generate 100 docked conformations of capzimin. Binding mode of CSN5i-3 in the CSN5 monomer is known (Schlierf et al., 2016) hence we used the co-crystalized conformation of CSN5i-3 to generate CSN5i-3 bound CSN5-CSN6 heterodimeric, Rpn11 monomeric and Rpn8-Rpn11 heterodimeric complexes by manual docking (structural superimposition).

Molecular Dynamics Simulations

All MD simulations were carried out using GROMACS-5.1.2 (Van Der Spoel et al., 2005). For monomeric Rpn11, CSN5, and heterodimeric complexes Rpn11-Rpn8 and CSN5-CSN6 the all-atom CHARMM27 force field (which has CHARMM22 and CMAP for proteins) (Mackerell et al., 1998, 2004) provided in the GROMACS package was used. CHARMM27 force field provides non-bonded parameters for Zn^{2+}. Optimized forcefield parameters for Zn^{+2} were taken from Stote and Karplus (1995) and were shown to give reliable coordination geometries. The topology files for capzimin and CSN5i-3 (see **Supplementary File** for more information) were generated with the help of the SwissParam server (Zoete et al., 2011). All complexes were enclosed in triclinic boxes (see **Table S1**). The TIP3P water (Jorgensen, 1981; Mark and Nilsson, 2001) model was used to solvate all complexes. Ions (Na^+ and Cl^-) were added at 0.15 M concentration to neutralize the systems. After neutralization,
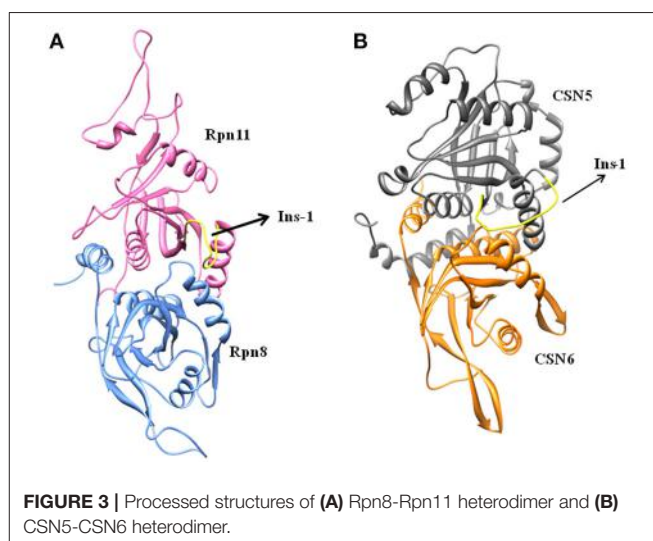
FIGURE 3 | Processed structures of **(A)** Rpn8-Rpn11 heterodimer and **(B)** CSN5-CSN6 heterodimer.

the systems were subjected to 5000 steps of steepest decent minimization. Minimized systems were further equilibrated under both NVT and NPT conditions for 1 and 2 ns, respectively. During equilibration, position restrains were applied to both protein (including Zn^{+2}) and ligand atoms. Temperature (310 K) and pressure (1 atm) were controlled by the velocity rescaling thermostat (Bussi et al., 2007) and Parrinello-Rahman barostat (Parrinello and Rahman, 1981), respectively. The equilibrated systems were finally subjected to the 100 ns production phase under NPT condition without any position restraints. Three independent simulations were carried out for each of the complexes.

Binding Energy Calculation

Binding energies of inhibitors were calculated with the help of the linear interaction energy (LIE) methodology. LIE methodology has been reported to predict reliable binding energies (Hansson et al., 1998; Aqvist and Marelius, 2001). In the LIE methodology, the free energy of transfer of ligand from water to the protein environment is giving the binding energy. In simple terms, the LIE equation can be given as:

$$\Delta G_{bind} \ (\text{ligand}) = G^{bound}_{sol} \ (\text{ligand}) - G^{free}_{sol} \ (\text{ligand}) \quad (1)$$

Where ΔG_{bind} (ligand) is binding energy of ligand, G^{bounds}_{sol} (ligand) is the energy of ligand in the solvated protein-ligand complex and G^{free}_{sol} (ligand) is the energy of the free ligand in water. LIE calculation is generally carried out in combination with MD or Monte Carlo (MC) simulation. LIE has two components, i.e., electrostatic (el) and van der Waals (vdW) interactions[30]. Hence, Equation (1) can be rewritten as

$$\Delta G_{bind} = \alpha(\langle U^{vdW}_{lig-surr}\rangle_{protein} - \langle U^{vdW}_{lig-surr}\rangle_{water})$$
$$+ \beta(\langle U^{el}_{lig-surr}\rangle_{protein} - \langle U^{el}_{lig-surr}\rangle_{water})$$
$$= \alpha\Delta U^{vdW} + \beta\Delta U^{el} \quad (2)$$

Where brackets $<>$ indicate thermodynamic averages of the interaction energies of the ligand with its surroundings (Aqvist and Marelius, 2001). α is an empirically derived non-polar scaling factor and β is a polar scaling factor (Aqvist and Marelius, 2001). Almlöf et al. have suggested $\beta_0 = 0.43$ for neutral compounds and correction factors ($\Delta\beta$) for different ligands (Almlöf et al., 2007). For calculation of ligand binding energies we have used $\alpha = 0.18$. β values for the thiolate form of capzimin (anion) and CSN5i-3 (alcohol) were obtained after applying functional group-specific correction factors to the β_0 (0.43+0.02 = 0.45 for capzimin and 0.43−0.06 = 0.37 for CSN5i-3) (Almlöf et al., 2007; Gutiérrez-De-Terán and Aqvist, 2012). Values of β depend on the chemical nature of the ligand (Hansson et al., 1998; Rinaldi et al., 2004). In order to calculate the energy terms (vdW and el) of capzimin and CSN5i-3 in the water, we have carried out separate 50 ns MD simulation for each.

RESULTS AND DISCUSSION

Open Conformations of Rpn11 and CSN5

Ins-1 region of CSN5 (97–131) and RPN11 (74–106) has been reported to be flexible and its flexibility is important for the binding of distal ubiquitin (Echalier et al., 2013; Worden et al., 2014). In the cryo-EM structure of Rpn11, the Ins-1 loop (76–88) obstructs the distal ubiquitin binding pocket. Previous studies suggest that the Ins-1 loop is flexible and very important for the regulation of enzymatic activity of zinc metalloproteases. It seems that during ligand binding this Ins-1 loop can adopt different conformations. Keeping the above fact in mind, we generated alternative loop conformations so that the distal ubiquitin binding pocket becomes accessible for the binding of inhibitors. In the open conformation, Ins-1 loop is pointing away from the catalytic site (**Figure 2**). We have used two different crystal structures to represent CSN5 in monomeric and heterodimeric states. In the monomeric state, CSN5 is already complexed with a potent inhibitor CSN5i-3 and a part of Ins-1 loop (100–106) is missing but we have modeled the Ins-1 loop in open conformation. In the crystal structure of CSN5-CSN6 heterodimer, initial Ins-1 loop (98–109) conformation blocks the distal ubiquitin binding site which

we also observed in cryo-EM structure of Rpn11. Hence, Ins-1 loop was remodeled in the open conformation (**Figures 2B,C**). Rpn8-RPN11 and CSN5-CSN6 heterodimeric states are shown in **Figure 3**.

Initial Docked Conformations of Capzimin

The recently published crystal structure of CSN5 with CSN5i-3 provides a picture of inhibitor occupation of the distal ubiquitin binding site and possible interference with the catalytic activity of CSN5. Due to the conservation of residues in the distal ubiquitin binding site between CSN5 and Rpn11, we assumed that capzimin might also occupy the distal ubiquitin binding site of Rpn11. In the CSN5i-3 bound crystal structure of CSN5, the nitrogen atom of the azole ring makes a coordinate bond with Zn^{2+}. In the previous study (Perez et al., 2017), the parent compound of capzimin that is 8-thiquinoline (8TQ) has been shown to interact with the Zn^{2+} in a bidentate manner and it has been proposed that capzimin inhibits Rpn11 also via chelation of the catalytic Zn^{+2}. Interestingly, the top binding pose generated by AUTODOCK4.2 showed that the 8TQ moiety of capzimin also makes a monodentate interaction with Zn^{2+} of CSN5 and Rpn11 (**Figure 4**). Another possible coordinating atom of the 8TQ moiety was remote from the Zn^{2+} ion. The binding modes of capzimin in both Rpn11 and CSN5 appear similar but we observed few differences in the distal ubiquitin binding site. In RPN11, the amide portion of capzimin makes H bond with Thr129, and the thiazole moiety makes hydrophobic and van der Waals interactions with side chains of Met54 and Asp88, respectively (**Figure 4A**). In CSN5, H-bond with Thr154 was absent and thiazole moiety of capzimin makes H-bond with side chain of Asn158 and shows hydrophobic interaction with the sidechains of Met78 and Trp136 (**Figure 4B**).

Binding of Capzimin and CSN5i-3 to Monomeric Rpn11 and Rpn8-Rpn11 Heterodimer

Being an integral part of the proteasome machinery, Rpn11 works in coordination with other subunits. Monomeric Rpn11 lacks deubiquitinylase activity and is active only in the presence of Rpn8. Here, we have investigated the binding of Rpn11 inhibitors in the absence and presence of Rpn8 (**Figures 5–9**). $C\alpha$ root mean square deviation ($C\alpha$-RMSD) analysis revealed that compared to the monomer capzimin-Rpn11 complex, Rpn11 in the heterodimeric capzimin-Rpn8-Rpn11 complex shows lower RMSD (**Figure 5A**) and is thus stabilized. $C\alpha$ root mean square fluctuation ($C\alpha$-RMSF plot) (**Figure 5B**) showed that in the absence of Rpn8, residues belonging to the Ins-1 loop and $\alpha2$ helix undergo larger fluctuations. Structural analysis also revealed that in the capzimin-Rpn11 complex, the Ins-1 loop and $\alpha2$ helix exhibited large movements. Particularly, the $\alpha2$ helix came closer to the $\alpha3$ helix (**Figure 9A**). The presence of Rpn8 restricts this movement of $\alpha2$ helix as well as that of the Ins-1 loop. Both crystal structures of the yeast Rpn8-Rpn11 heterodimer (Pathare et al., 2014) and the cryo-EM structure of human 26S proteasome (Huang et al., 2016) show that the
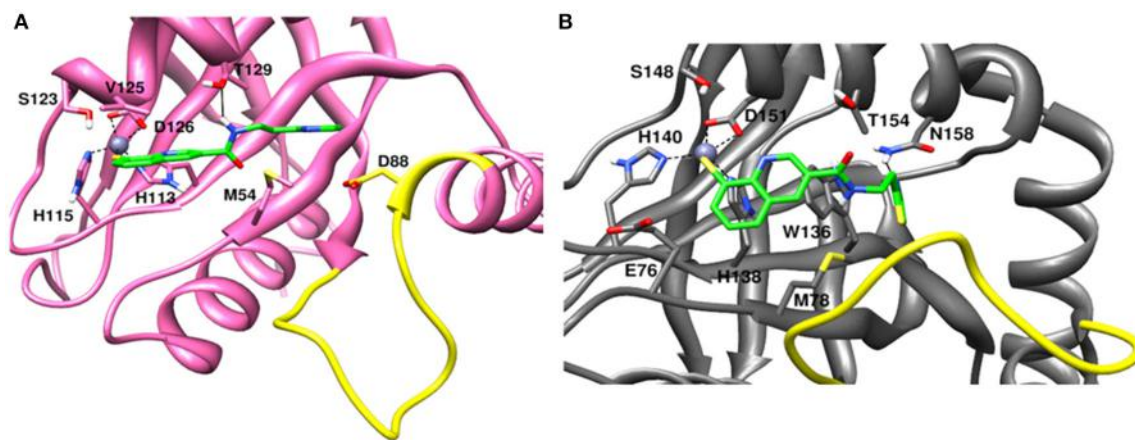
FIGURE 4 | Docked conformation of capzimin in the **(A)** Rpn11 binding pocket and **(B)** CSN5 binding pocket.
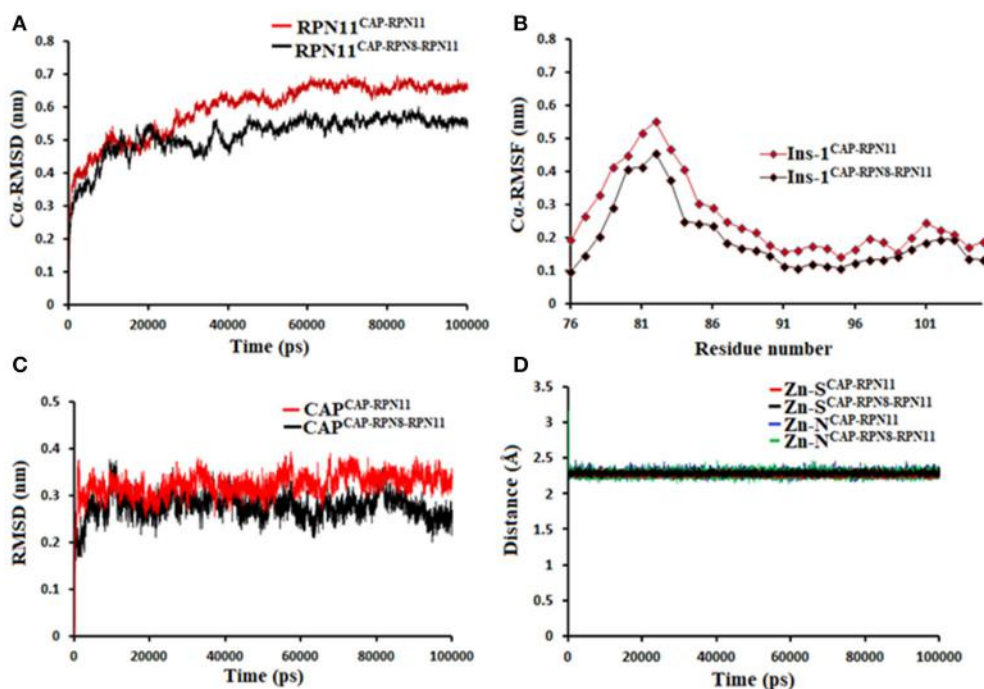
FIGURE 5 | MD simulation data of Rpn11 bound to capzimin in the absence (red) and presence of Rpn8 (black). **(A)** Cα-RMSD of Rpn11 **(B)** Cα-RMSF of Ins-1 region of Rpn11 **(C)** RMSD of capzimin and **(D)** distances between coordinating atoms of capzimin and Zn^{+2}. Zn^{+2}-S distances in capzimin bound Rpn11 and Rpn8-Rpn11 heterodimer are shown in blue and green color, respectively.

α2 helices of both proteins makes extensive contacts with each other.

Molecular dynamics simulations refined the binding position of capzimin. The average RMSDs of the refined structures to the starting were 0.32 and 0.27 Å in Rpn11 monomer and Rpn8-Rpn11 heterodimer, respectively. In the heterodimeric state, capzimin showed overall less deviation in (**Figure 5C**). It seems that during the start of simulation, capzimin tries to optimize interactions with the residues of the binding pocket and this leads to the deviation from

initial conformation. However, in both complexes, $S-Zn^{2+}$ and $N-Zn^{2+}$ distances (**Table 1**) were stable (**Figure 5D**) and comparable to the experimentally determined distances reported in $[(Tp^{Me,Ph})Zn(8TQ)]$ complex (Perez et al., 2017). The H-bond between capzimin and the side chain of Thr129 was maintained during most of the part of trajectory (**Figures S4A,B**). In **Figures 6A,B**, we see that the 8TQ fragment of the capzimin interacts with the catalytic Zn^{+2} and the additional amide moiety interacts with Thr129. Leu56, Pro89, and Phe133 provides hydrophobic interactions to the azole moiety of the
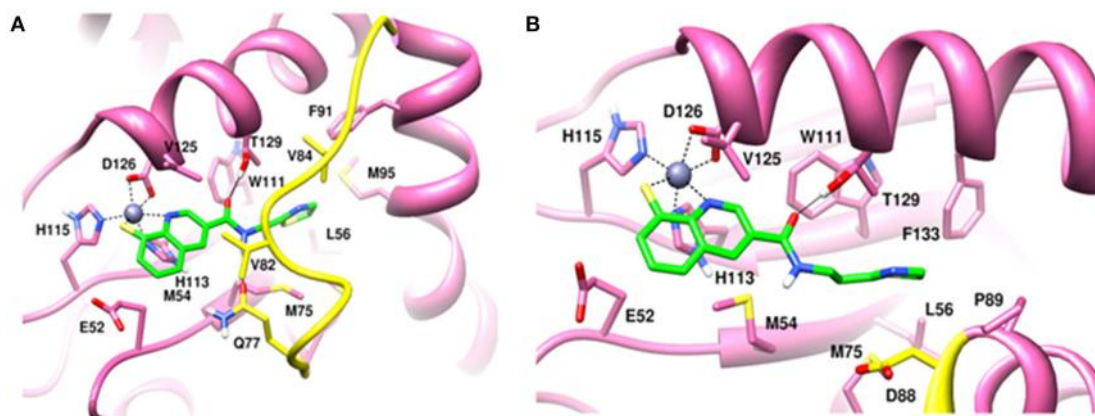
FIGURE 6 | MD snapshots showing interaction of **(A)** capzimin with monomeric Rpn11 and **(B)** capzimin with Rpn11 in the presence of Rpn8. Yellow region corresponds to Ins-1 loop. H-bonds are represented by solid black line.
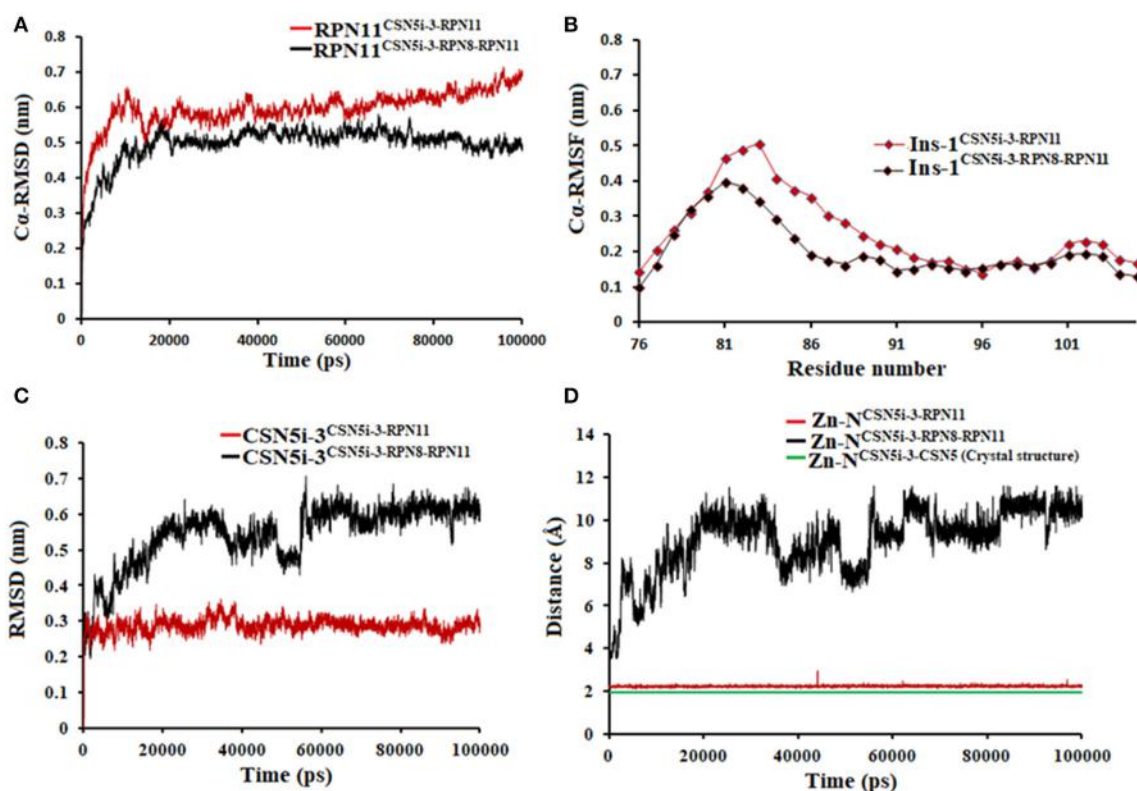
FIGURE 7 | MD simulation data of Rpn11 bound to CSN5i-3 in the absence (red) and presence of Rpn8 (black). **(A)** Cα-RMSD of Rpn11 **(B)** Cα-RMSF of Ins-1 region of Rpn11 **(C)** RMSD of CSN5i-3 and **(D)** distances between coordinating atoms of CSN5i-3 and Zn^{+2}. Zn^{+2}-N distance in the crystal structure of CSN5i-3 bound CSN5 is shown in green color.

capzimin. Capzimin showed almost similar binding affinity to monomeric Rpn11 as well as Rpn8-Rpn11 heterodimer (**Table 2**).

We also investigated the binding of CSN5i-3, which is a very weak inhibitor of Rpn11. The starting structures of CSN5i-3 bound to the monomeric Rpn11 and Rpn8-Rpn11

heterodimer were generated by manual docking. The CSN5i-3-Rpn11 complex showed a very high Cα-RMSD (**Figure 7A**) suggesting that Rpn11 undergoes large conformational changes (**Figure 9B**). However, in the presence of Rpn8, the Cα-RMSD of Rpn11 was comparatively low (**Figure 7A**). The Ins-1 loop was more flexible in the absence of Rpn8 (**Figure 7B**). In
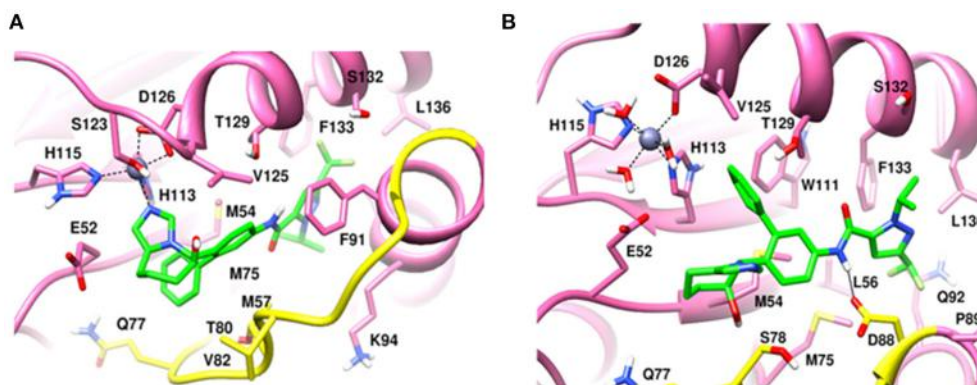
FIGURE 8 | MD snapshots showing interaction of **(A)** CSN5i-3 with monomeric Rpn11 and **(B)** CSN5i-3 with Rpn11 in the presence of Rpn8. Yellow region corresponds to Ins-1 loop. H-bonds are represented by solid black line.
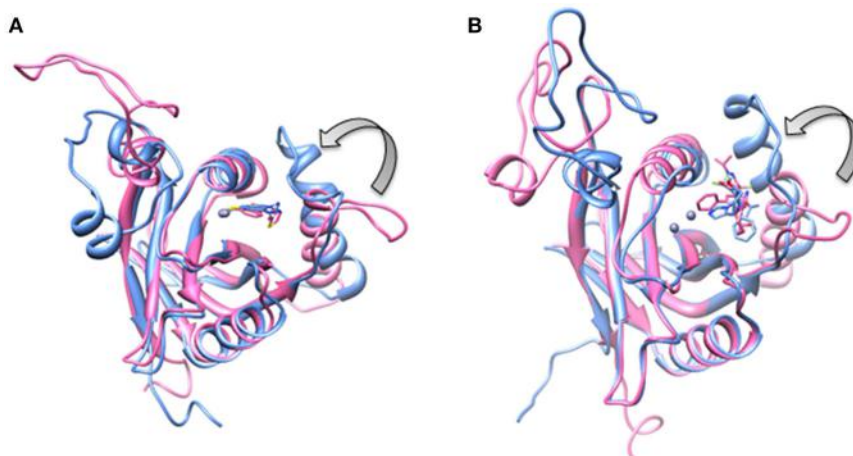
FIGURE 9 | Aligned structures of inhibitors bound Rpn11 in the absence (blue) and presence of Rpn8 (hot pink) in which Ins-1 loop has moved (curved arrow) toward the distal ubiquitn binding site **(A)** capzimin bound Rpn11 and **(B)** CSN5i-3 bound Rpn11.

CSN5i-3-Rpn11 complex, we observed that CSN5i-3 was stable (**Figure 7C**). The N-Zn^{2+} distance was very close to the distance reported in the crystal structure of CSN5 crystallized with CSN5i-3 (**Figure 7D** and **Table 1**). This simulated binding of CSN5i-3 to Rpn11 would not be in agreement with experiment and explain its low inhibitor activity. Thus, we investigated whether this was an unrealistic over binding, a protein-protein complex environment would be able to reproduce the selectivity.

In the heterodimeric Rpn8-Rpn11 complex, CSN5i-3 binding in the pocket was not stable (**Figure 7C**), the distance to zinc increases continously and eventually the ligand lost coordination with Zn^{+2} (**Figure 7D** and **Table 1**) which shows that the influence of Rpn8 on the capzimin bound Rpn11 is prominent. Interaction of CSN5i-3 in monomeric Rpn11 is shown in **Figure 8A**. In both complexes, H-bond with Thr129 was absent (**Figure 8** and **Figures S4C,D**). We see that binding of CSN5i-3 to Rpn11 is significantly influenced by the presence of Rpn8 (**Figures 7C,D**, **8B**) and the consideration of this heterodimeric state off Rpn11-Rpn8 is necessary to explain capzimin binding

and CSN5i-3 non-binding. LIE calculations (**Table 2**) revealed that CSN5i-3 binds moderately to the monomeric Rpn11 and very weakly to the Rpn8-Rpn11 heterodimer. Outside the proteasome, Rpn11 also plays an important role in different cellular activities.[33] Therefore, binding of capzimin and CSN5i-3 to the monomeric Rpn11 are physiologically possible. In recent study, CSN5i-3 has been shown to bind with the recombinant monomeric CSN5 and a co-crystallized structure has been obtained (Schlierf et al., 2016).

Binding of Capzimin and CSN5i-3 to Monomeric CSN5 and CSN5-CSN6 Heterodimer

MD simulation of capzimin with monomeric CSN5 and CSN5-CSN6 heterodimer revealed that its binding to CSN5 is also influenced by the CSN6 (**Figure 10**). In the presence of CSN6, the Ins-1 loop showed a larger flexibility (**Figure 10B**). Capzimin shows a higher ligand RMSD in the monomeric

TABLE 1 | Distances of coordinating atoms of capzimin and CSN5i-3 from catalytic Zn^{+2} of Rpn11 and CSN5.

	Distance (Å)	
System	**N-Zn^{+2}**	**S-Zn^{+2}**
Capzimin-Rpn11	2.28 ± 0.04	2.27 ± 0.02
Capzimin-Rpn8-Rpn11	2.27 ± 0.04	2.28 ± 0.02
Capzimin-CSN5	2.30 ± 0.07	2.30 ± 0.02
Capzimin-CSN5-CSN6	4.04 ± 0.14	2.28 ± 0.02
CSN5i-3-Rpn11	2.23 ± 0.04	
CSN5i-3-Rpn8-Rpn11	9.10 ± 1.51	
CSN5i-3-CSN5	2.22 ± 0.02	
CSN5i-3-CSN5-CSN6	2.18 ± 0.02	

Data represented here are averages of three independent MD simulation runs.

TABLE 2 | Binding energies of ligands calculated using LIE method.

	Binding energy (kcal/mole)	
System	**Capzimin**	**CSN5i-3**
Rpn11	−37.815 ± 2.898	−5.401 ± 2.023
Rpn8-Rpn11	**−38.265 ± 3.006**	**0.103 ± 2.023**
CSN5	−36.724 ± 3.007	−6.043 ± 2.010
CSN5-CSN6	**−26.429 ± 3.054**	**−8.710 ± 1.975**

The final 20 ns of each trajectory were used to calculate LIE. Data represented here are averages of three independent MD simulation runs. (α = 018, β = 0.37 for CSN5i-3 and 0.45 for capzimin). Bold values refer to protein-protein complexes.

CSN5 (**Figure 10C**). In the monomeric CSN5, capzimin shows bidentate interactions with the catalytic zinc (**Figures 10D, 11A**) and stable H-bond between amide NH and Thr154 (**Figure 11A** and **Figure S5A**). Intermittent H-bonds with Glu101 and Tyr143 were also observed. However, in the presence of CSN6, capzimin only showed a mono-dentate coordination with catalytic zinc (**Figures 10D, 11B**) and low occupancy H-bond with Met78, Arg106, and Asn158 (**Figure S5B**). Thiazole and amide moieties of capzimin showed H-bonding with the side chains of Asn158 and Met78, respectively. The side chain of Arg106 of Ins-1 loop showed H-bond with the 8TQ fragment of capzimin. Apart from H-bond, the thiazole moiety also showed hydrophobic interactions with side chains of Met78, Leu80 and Phe165. LIE calculations show that capzimin binds more strongly to the monomeric CSN5 (**Table 2**).

If we consider the crystal structure of CSN5 with CSN5i-3, a part of Ins-1 loop (100–106) is missing and residues of α4 helix do not interact with the CSN5i-3 ligand. The orientation of the α4 helix suggests that residues in the Ins-1 loop will also not make any interaction with the CSN5i-3. The MD simulation results (**Figure 12**) show that the Cα-RMSDs of CSN5 in both complexes do not vary much and they converge near the end of simulations (**Figure 12A**). In the presence of CSN6, Residues (101–110) in the Ins-1 loop (98–113) of CSN5 showed more flexibility but α4 helix (111–131) was comparatively less flexible (**Figure 12B**). Previous MD simulations study on monomeric CNS5 suggests

that portions of the Ins-1 region show high flexibility (Echalier et al., 2013). In the CSN5-CSN6 heterodimer we observed that CSN5i-3 is stable (**Figure 12C**) and its conformation is very close to the conformation reported in the crystal structure (**Figure S6**). The N-Zn^{2+} distances are shown in **Figure 12D**. They are stable over time for the monomer and heterodimer states and only slightly longer than in the crystal structure. In both complexes, residues in Ins-1 loop make contact with the CSN5i-3 (**Figure 13**). We observed that interaction of the α2 helix of CSN6 affects the movement of the Ins-1 loop and α4 helix of CSN5.

In the monomeric CSN5, CSN5i-3 makes a low occupancy H-bond with Thr154 and relatively stable H-bond with Asn158 (**Figure 13A** and **Figure S5C**). However, in the CSN5-CSN6 heterodimer, CSN5i-3 forms stable H-bonds with both Thr154 and Asn158 (**Figure S5D** and **Figure 13B**). In the crystal structure, the H-bond between the azole ring of CSN5i-3 and Asn158 is not present but our MD refinement showed a stable H-bond. The difluoromethyl group projects toward Leu157 and makes hydrophobic interactions with it (**Figure 13**). In both capzimin and CSN5i-3 bound monomeric CSN5, we observed that Ins-1 region slightly moves toward inhibitors and a part of Ins-1 loop changes into α helix (**Figure 14**). This may be the reason behind lower flexibility of Ins-1 loop in the monomeric CSN5 compared to the CSN6 bound CSN5.

Structural Elements Responsible for Ligand-Selective Inhibition of Rpn11 and CSN5

Based on the MD simulation results of Rpn11 and CSN5 with capzimin and CSN5i-3, we can explain the selective inhibition of these two proteins. At the sequence level, the MPN domain of CSN5 and Rpn11 are moderately conserved. However, the Zn^{2+} binding residues are fully conserved. The distal ubiquitin binding region is large therefore we mainly focused on the residues which interact with capzimin and CSN5i-3. We found that Leu98, Val100, Thr105, Arg106, Gln110, Ala112, Ala113, Tyr114, Glu115, Tyr116, Met117, A119, Ile150, Leu157 and Asn158, and Phe161 (in CSN5) are substituted by Met75, Gln77, Val82, Ser83, Glu85, Val87, Asp88, Pro89, Val90, Phe91, Gln92, Lys94, Val125, Ser132, Phe133, and Leu136, respectively (in Rpn). We observed that most of the residues in the Ins-1 loop region are not conserved. The Ins-1 region plays an important role in positioning of C-terminus of the distal ubiquitin for cleavage of iso-peptide bond (Worden et al., 2014). Hence Ins-1 region appears very promising for the design of selective JAMM inhibitors.

If we consider the calculated LIE of capzimin in the monomeric Rpn11 and CSN5 proteins, we see that capzimin has a slightly higher affinity for Rpn11 (1 kcal/mole more than CSN5). Selective binding is more pronounced and can be rationalized when we consider their respective binding partners, Rpn8 and CSN6. Similarly, selective binding of CSN5i-3 is more pronounced when we consider the heterodimeric states of Rpn11 and CSN5.

FIGURE 10 | MD simulation data of capzimin bound to CSN5 in the absence (red) and presence of CSN6 (black). **(A)** Cα-RMSD of CSN5 **(B)** Cα-RMSF of Ins-1 region of CSN5 **(C)** RMSD of capzimin and **(D)** distances between coordinating atoms of capzimin and Zn^{+2}. Zn^{+2}-S distances in capzimin bound CSN5 and CSN5-CSN6 heterodimer are shown in blue and green color, respectively.

FIGURE 11 | MD snapshots showing interactions of **(A)** capzimin with the CSN5 monomer and **(B)** capzimin with CSN5 in the presence of CSN6. Yellow region corresponds to Ins-1 loop. H-bonds are represented by solid black line.

Thus, we need to consider the heterodimeric states of Rpn11 and CSN5 to explain the selectivity of capzimin and CSN5i-3. In the presence of binding partners, the Ins-1 regions of RPN11 and CSN5 show different flexibility which in turn affects the binding of capzimin and CSN5i-3.

Capzimin is 80-fold more selective toward Rpn11 and its thiazole moiety appears important for this selectivity. In the Rpn8-Rpn11 heterodimer, capzimin showed a bidentate coordination with Zn^{2+} and a stable H-bond with side chain of Thr129 (**Figure S4B**). However, in the CSN5-CSN6 heterodimer, capzimin displayed a monodentate coordination with Zn^{2+} and only low occupancy H-bonds with side chains of Met78, Arg106, and Asn158 (**Figure S5B**). The low affinity of capzimin for CSN5 can be attributed to the lack of an extra N-Zn^{2+} coordination and
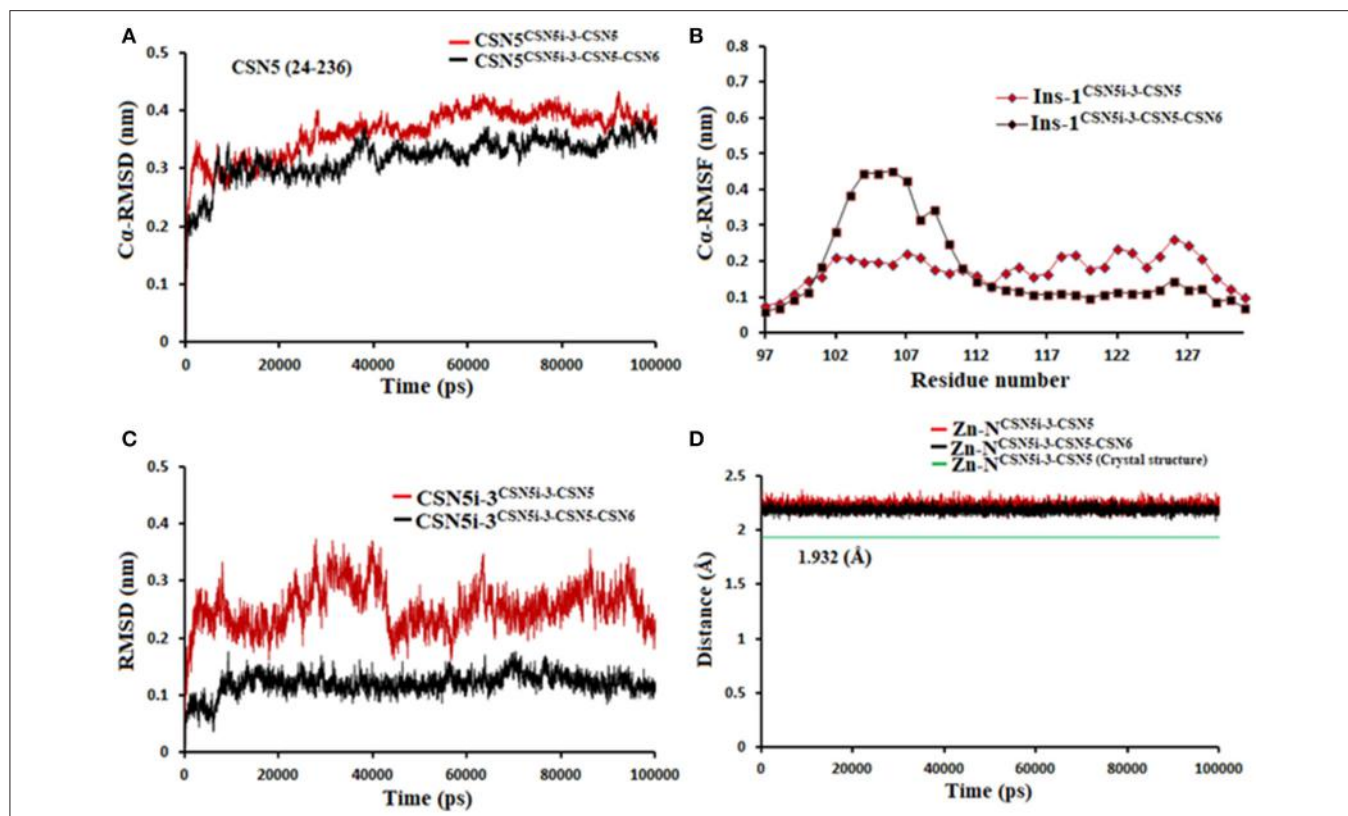
FIGURE 12 | MD simulation data of CSN5i-3 bound to CSN5 in the absence (red) and presence of CSN6 (black). **(A)** Cα-RMSD of CSN5 **(B)** Cα-RMSF of Ins-1 region of CSN5 **(C)** RMSD of CSN5i-3 and **(D)** distances between coordinating atoms of CSN5i-3 and Zn^{+2}.
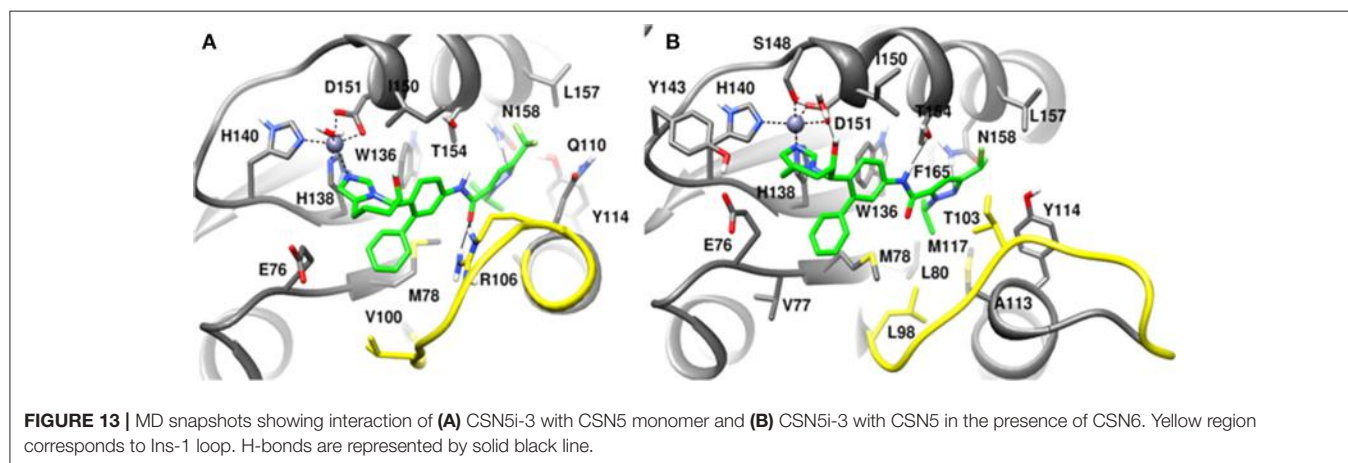
FIGURE 13 | MD snapshots showing interaction of **(A)** CSN5i-3 with CSN5 monomer and **(B)** CSN5i-3 with CSN5 in the presence of CSN6. Yellow region corresponds to Ins-1 loop. H-bonds are represented by solid black line.

H-bond with Thr154. We observed that in the case of capzimin, the binding energies were overestimated. The major reason for this overestimation can be attributed to the presence of a net negative charge on the capzimin ligand. A previous study has also reported very negative binding energy values for negatively charged ligands (Genheden and Ryde, 2015). It should be noted that LIE calculations are very sensitive to the β parameter. Values of β depend on the chemical nature of ligands and can have different values for different ligands. In present study, we have

used LIE data only to compare relative binding affinities between monomeric and heterodimeric protein structures.

As for selective binding of CSN5i-3 to CSN5, we observed that in the presence of CSN6, CSN5i-3 binds strongly to the protein-protein complex. However in the case of the Rpn8-Rpn11 heterodimer, CSN5i-3 showed only very weak interaction. With the progress of the MD simulations, CSN5i-3 loses coordination with Zn^{2+} and H-bond with Thr129. In Rpn11, the side chain of Phe133 cannot form H-bond with

FIGURE 14 | Aligned structures of inhibitors bound CSN5 in the absence (gray) and presence of CSN6 (orange) in which Ins-1 loop has moved (curved arrow) toward the distal ubiquitn binding site **(A)** capzimin bound CSN5 and **(B)** CSN5i-3 bound CSN5.
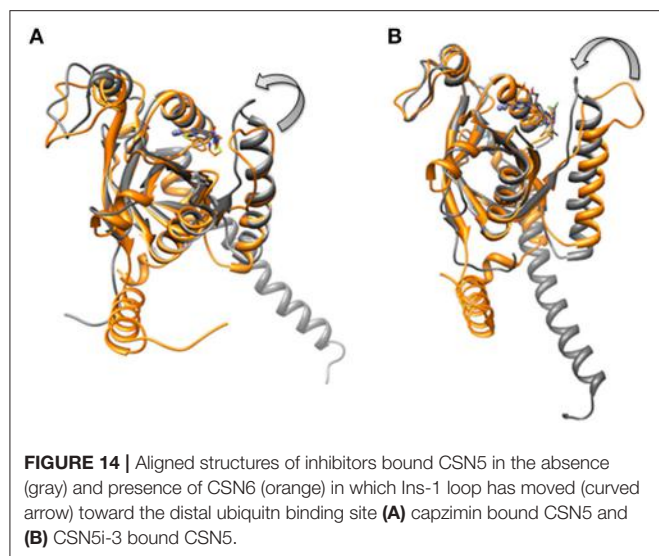
the azole ring of CSN5i-3. Instead, the azole ring showed hydrophobic interaction with side chain of Leu56, Met75 and Phe133 and van der Waals interactions with the sidechain of Asp88.

It seems that establishing interactions of the azole ring with nearby sidechains resulted in loss of coordination with the active site Zn^{2+} ion. In the case of the Rpn8-Rpn11 heterodimer, LIE calculations showed only a very low binding energy of CSN5i-3 (**Table 2**) which is in agreement with the reported ligand binding selectivity. Overall our findings suggest that the heterodimeric protein states of both CSN5 and Rpn11 targets have to be considered to explain the selectivity of capzimin and CSN5i-3 ligands.

CONCLUSIONS

In the present study the computational analysis of binding modes and selectivities of reported metallo-deubiquitinylase

inhibitors, capzimin and CSN5i-3, was performed in detail. Capzimin is a selective inhibitor of Rpn11 while CSN5i-3 is selective for CSN5. We found that capzimin binds to RPN11 via chelation of the active site Zn^{2+} and its interaction extends to the distal ubiquitin binding site. Our MD studies suggest that compared to the monomeric states, the heterodimeric protein-protein complexes of Rpn11 and CSN5 are conformationally stable. Considering the Rpn8-Rpn11 and CSN5-CSN6 heterodimers, we found that residues in the distal ubiquitin site are responsible for selectivity and must be taken into consideration for the design of selective inhibitors of CSN5 and Rpn11 in future studies. Additionally, we have shown that flexibility of Ins-1 region in Rpn11 and CSN5 is significantly affected by the presence of their respective protein binding partners.

AUTHOR CONTRIBUTIONS

MS and MN designed the study. VK performed simulations in this study. MS, MN, and VK analyzed and discussed the results and wrote the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2018.00480/full#supplementary-material

REFERENCES

Almlöf, M., Carlsson, J., and Åqvist, J. (2007). Improving the accuracy of the linear interaction energy method for solvation free energies. *J. Chem. Theory Comput.* 3, 2162–2175. doi: 10.1021/ct700106b

Aqvist, J., and Marelius, J. (2001). The linear interaction energy method for predicting ligand binding free energies. *Comb. Chem. High Throughput Screen* 4, 613–626. doi: 10.2174/1386207013330661

Berndt, C., Bech-Otschir, D., Dubiel, W., and Seeger, M. (2002). Ubiquitin system: JAMMing in the name of the lid. *Curr. Biol.* 12, R815–R817. doi: 10.1016/S0960-9822(02)01317-9

Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126:014101. doi: 10.1063/1.2408420

Echalier, A., Pan, Y., Birol, M., Tavernier, N., Pintard, L., Hoh, F., et al. (2013). Insights into the regulation of the human COP9 signalosome catalytic subunit, CSN5/Jab1. *Proc. Natl. Acad. Sci. U.S.A.* 110, 1273–1278. doi: 10.1073/pnas.1209345110

Genheden, S., and Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* 10, 449–461. doi: 10.1517/17460441.2015.1032936

Gutiérrez-De-Terán, H., and Aqvist, J. (2012). Linear interaction energy: method and applications in drug design. *Methods Mol. Biol.* 819, 305–323. doi: 10.1007/978-1-61779-465-0_20

Hansson, T., Marelius, J., and Aqvist, J. (1998). Ligand binding affinity prediction by linear interaction energy methods. *J. Comput. Aided Mol. Des.* 12, 27–35. doi: 10.1023/A:1007930623000

Huang, X., Luan, B., Wu, J., and Shi, Y. (2016). An atomic structure of the human 26S proteasome. *Nat. Struct. Mol. Biol.* 23, 778–785. doi: 10.1038/nsmb.3273

Jorgensen, W. (1981). Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. *J. Am. Chem. Soc.* 103, 335–340.

Komander, D., Clague, M. J., and Urbé, S. (2009). Breaking the chains: structure and function of the deubiquitinases. *Nat. Rev. Mol. Cell Biol.* 10, 550–563. doi: 10.1038/nrm2731

Lee, M. H., Zhao, R., Phan, L., and Yeung, S. C. (2011). Roles of COP9 signalosome in cancer. *Cell Cycle* 10, 3057–3066. doi: 10.4161/cc.10.18.17320

Li, J., Yakushi, T., Parlati, F., Mackinnon, A. L., Perez, C., Ma, Y., et al. (2017). Capzimin is a potent and specific inhibitor of proteasome isopeptidase Rpn11. *Nat. Chem. Biol.* 13, 486–493. doi: 10.1038/nchembio.2326

Lingaraju, G. M., Bunker, R. D., Cavadini, S., Hess, D., Hassiepen, U., Renatus, M., et al. (2014). Crystal structure of the human COP9 signalosome. *Nature* 512, 161–165. doi: 10.1038/nature13566

Mackerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., et al. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* 102, 3586–3616. doi: 10.1021/jp973084f

Mackerell, A. D., Feig, M., and Brooks, C. L. (2004). Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* 25, 1400–1415. doi: 10.1002/jcc.20065

Mark, P., and Nilsson, L. (2001). Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *J. Phys. Chem. A* 105, 9954–9960. doi: 10.1021/jp003020w

Mevissen, T. E. T., and Komander, D. (2017). Mechanisms of deubiquitinase specificity and regulation. *Annu. Rev. Biochem.* 86, 159–192. doi: 10.1146/annurev-biochem-061516-044916

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., et al. (2009). AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* 30, 2785–2791. doi: 10.1002/jcc.21256

Nandi, D., Tahiliani, P., Kumar, A., and Chandu, D. (2006). The ubiquitin-proteasome system. *J. Biosci.* 31, 137–155. doi: 10.1007/BF02705243

Parrinello, M., and Rahman, A. (1981). Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* 52, 7182–7190. doi: 10.1063/1.328693

Pathare, G. R., Nagy, I., Sledz, P., Anderson, D. J., Zhou, H. J., Pardon, E., et al. and Baumeister, W. (2014). Crystal structure of the proteasomal deubiquitylation module Rpn8-Rpn11. *Proc. Natl. Acad. Sci. U.S.A.* 111, 2984–2989. doi: 10.1073/pnas.1400546111

Perez, C., Li, J., Parlati, F., Rouffet, M., Ma, Y., Mackinnon, A. L., et al. (2017). Discovery of an inhibitor of the proteasome subunit Rpn11. *J. Med. Chem.* 60, 1343–1361. doi: 10.1021/acs.jmedchem.6b01379

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., et al. (2004). UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612. doi: 10.1002/jcc.20084

Reyes-Turcu, F. E., Ventii, K. H., and Wilkinson, K. D. (2009). Regulation and cellular roles of ubiquitin-specific deubiquitinating enzymes. *Annu. Rev. Biochem.* 78, 363–397. doi: 10.1146/annurev.biochem.78.082307.091526

Rinaldi, T., Pick, E., Gambadoro, A., Zilli, S., Maytal-Kivity, V., Frontali, L., et al. (2004). Participation of the proteasomal lid subunit Rpn11 in mitochondrial morphology and function is mapped to a distinct C-terminal domain. *Biochem. J.* 381, 275–285. doi: 10.1042/BJ20040008

Sali, A., and Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234, 779–815. doi: 10.1006/jmbi.1993.1626

Schlierf, A., Altmann, E., Quancard, J., Jefferson, A. B., Assenberg, R., Renatus, M., et al. (2016). Targeted inhibition of the COP9 signalosome for treatment of cancer. *Nat. Commun.* 7:13166. doi: 10.1038/ncomms13166

Shrestha, R. K., Ronau, J. A., Davies, C. W., Guenette, R. G., Strieter, E. R., Paul, L. N., et al. (2014). Insights into the mechanism of deubiquitination by JAMM deubiquitinases from cocrystal structures of the enzyme with the substrate and product. *Biochemistry* 53, 3199–3217. doi: 10.1021/bi5003162

Song, Y., Li, S., Ray, A., Das, D. S., Qi, J., Samur, M. K., et al. (2017). Blockade of deubiquitylating enzyme Rpn11 triggers apoptosis in multiple myeloma cells and overcomes bortezomib resistance. *Oncogene* 36, 5631–5638. doi: 10.1038/onc.2017.172

Stote, R. H., and Karplus, M. (1995). Zinc binding in proteins and solution: a simple but accurate nonbonded representation. *Proteins* 23, 12–31. doi: 10.1002/prot.340230104

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. (2005). GROMACS: fast, flexible, and free. *J. Comput. Chem.* 26, 1701–1718. doi: 10.1002/jcc.20291

Verma, R., Aravind, L., Oania, R., Mcdonald, W. H., Yates, J. R., Koonin, E. V., et al. (2002). Role of Rpn11 metalloprotease in deubiquitination and degradation by the 26S proteasome. *Science* 298, 611–615. doi: 10.1126/science.1075898

Webb, B., and Sali, A. (2017). Protein structure modeling with MODELLER. *Methods Mol. Biol.* 1654, 39–54. doi: 10.1007/978-1-4939-7231-9_4

Worden, E. J., Padovani, C., and Martin, A. (2014). Structure of the Rpn11-Rpn8 dimer reveals mechanisms of substrate deubiquitination during proteasomal degradation. *Nat. Struct. Mol. Biol.* 21, 220–227. doi: 10.1038/nsmb.2771

Yao, T., and Cohen, R. E. (2002). A cryptic protease couples deubiquitination and degradation by the proteasome. *Nature* 419, 403–407. doi: 10.1038/nature01071

Zoete, V., Cuendet, M. A., Grosdidier, A., and Michielin, O. (2011). SwissParam: a fast force field generation tool for small organic molecules. *J. Comput. Chem.* 32, 2359–2368. doi: 10.1002/jcc.21816

Check for updates

Does Substrate Positioning Affect the Selectivity and Reactivity in the Hectochlorin Biosynthesis Halogenase?

Amy Timmins[1], Nicholas J. Fowler[2], Jim Warwicker[2], Grit D. Straganz[3,4] and Sam P. de Visser[1]*

[1] The Manchester Institute of Biotechnology and School of Chemical Engineering and Analytical Science, University of Manchester, Manchester, United Kingdom, [2] The Manchester Institute of Biotechnology and School of Chemistry, University of Manchester, Manchester, United Kingdom, [3] Institute of Biochemistry, Graz University of Technology, Graz, Austria, [4] Institute of Molecular Biosciences, Graz University, Graz, Austria

In this work we present the first computational study on the hectochlorin biosynthesis enzyme HctB, which is a unique three-domain halogenase that activates non-amino acid moieties tethered to an acyl-carrier, and as such may have biotechnological relevance beyond other halogenases. We use a combination of small cluster models and full enzyme structures calculated with quantum mechanics/molecular mechanics methods. Our work reveals that the reaction is initiated with a rate-determining hydrogen atom abstraction from substrate by an iron (IV)-oxo species, which creates an iron (III)-hydroxo intermediate. In a subsequent step the reaction can bifurcate to either halogenation or hydroxylation of substrate, but substrate binding and positioning drives the reaction to optimal substrate halogenation. Furthermore, several key residues in the protein have been identified for their involvement in charge-dipole interactions and induced electric field effects. In particular, two charged second coordination sphere amino acid residues (Glu$_{223}$ and Arg$_{245}$) appear to influence the charge density on the Cl ligand and push the mechanism toward halogenation. Our studies, therefore, conclude that nonheme iron halogenases have a chemical structure that induces an electric field on the active site that affects the halide and iron charge distributions and enable efficient halogenation. As such, HctB is intricately designed for a substrate halogenation and operates distinctly different from other nonheme iron halogenases.

Keywords: nonheme iron, enzyme catalysis, reaction mechanism, QM/MM, density functional theory, halogenation, hydroxylation

INTRODUCTION

Enzymatic C–Cl bond formation is a rare process in Nature, yet over the past few decades a range of haloperoxidases and halogenases have been discovered (Gribble, 2003; Vaillancourt et al., 2006; van Pée et al., 2006; Butler and Sandy, 2009; Wagner et al., 2009; Weichold et al., 2016; Agarwal et al., 2017; Schnepel and Sewald, 2017; Timmins and de Visser, 2018). Their catalytic mechanism, however, is still subject to controversies and understanding the fundamental details of these processes may have an impact on biotechnological advances as well as drug development. Three

different classes of halogenation enzymes appear in Nature, namely the heme haloperoxidases, the vanadium-dependent nonheme haloperoxidases and the α-ketoglutarate dependent nonheme iron halogenases (Gribble, 2003; Vaillancourt et al., 2006; van Pée et al., 2006; Butler and Sandy, 2009; Wagner et al., 2009; Weichold et al., 2016; Agarwal et al., 2017; Schnepel and Sewald, 2017; Timmins and de Visser, 2018).

Heme haloperoxidases, (Sundaramoorthy et al., 1995; Wagenknecht and Woggon, 1997; Green et al., 2004; Kim et al., 2006) such as chloroperoxidase bind H_2O_2 on an iron(III)-heme center, which is then converted into an iron(IV)-oxo heme cation radical species called Compound I, (Meunier et al., 2004; Denisov et al., 2005; Shaik et al., 2005; Rittle and Green, 2010) with the help of a proton shuttle machinery. Compound I subsequently reacts with chloride to form OCl^- products. The product drifts out of the active site and reacts with substrates through halogenation. A second class of haloperoxidases has a nonheme vanadium co-factor that also utilizes hydrogen peroxide and halide in a catalytic cycle to form OCl^- products (Messerschmidt and Wever, 1996; Chen and van Pée, 2008). The vanadium haloperoxidases have been characterized in marine algae and are believed to have functions related to natural product synthesis associated with defense mechanisms (Martinez et al., 2001).

The final class of halogenases are the α-ketoglutarate (αKG) dependent nonheme iron halogenases, (Blasiak et al., 2006; Buongiorno and Straganz, 2013; Huang and Groves, 2017) which show structural and functional similarities to the corresponding nonheme iron—α-ketoglutarate-dependent dioxygenases (Solomon et al., 2000; Bugg, 2001; Ryle and Hausinger, 2002; Costas et al., 2004; Abu-Omar et al., 2005; Bruijnincx et al., 2008). These nonheme iron halogenases contain an iron(II) center in the resting state that is bound to the side chains of two histidine residues. During the catalytic cycle, halide and αKG bind to the iron(II) ion prior to molecular oxygen binding. It is believed dioxygen attacks αKG to form succinate and a high-valent iron(IV)-oxo species similarly to the αKG-dependent hydroxylases (Schofield and Zhang, 1999; Bollinger et al., 2005). The iron(IV)-oxo species of several nonheme iron halogenases have been spectroscopically trapped and characterized and shown to react with substrates with a rate-determining hydrogen atom abstraction barrier (Galonić Fujimori et al., 2007; Neidig et al., 2007; Matthews et al., 2009; Wong et al., 2013; Srnec et al., 2016; Srnec and Solomon, 2017). The subsequent pathway leading to the halogenated product; however, is controversial as a thermodynamically much more favorable hydroxylation pathway is prevented. How the enzyme manages to perform this unfavorable thermodynamic reaction pathway is under much debate. One proposal suggested links to substrate binding and orientation (Matthews et al., 2006). Thus, in the antibiotic biosynthesis protein SyrB2 the substrate L-Thr is linked to an acyl-carrier protein (SyrB1) via a phosphopantetheinyl (PPT) bridge. Replacement of L-Thr by L-norvaline changed the chemoselectivity of substrate halogenation to hydroxylation due to substrate positioning in the active site (Matthews et al., 2006). On the other hand, computational modeling of the Borowski group proposed a rotation step in the iron(IV)-oxo(halide) group, whereby after

hydrogen atom abstraction, the positions of the hydroxo and halide ions switched position leading to easier halide rebound (Borowski et al., 2010). Clearly, the mechanism of aliphatic halogenation remains controversial. Because of this, synthetic (biomimetic) model complexes of these nonheme iron halogenases and haloperoxidases have been developed and studied with experimental (Podgoršek et al., 2009; Comba and Wunderlich, 2010; Liu and Groves, 2010; Chatterjee and Paine, 2016; Puri et al., 2016; Wang et al., 2016) and computational approaches (Noack and Siegbahn, 2007; de Visser and Latifi, 2009; Kulik et al., 2009; Pandian et al., 2009; Quesne and de Visser, 2012; Senn, 2014; Huang et al., 2016; Timmins et al., 2018). Thus, if the halogenation pathway is dependent on substrate positioning then biomimetic models that lack the protein and do not bind substrates in fixed orientations may not be able to react via substrate halogenation efficiently.

A recently discovered halogenase (HctB) from *Lyngbya majuscula* is involved in the biosynthesis of hectochlorin, whereby a fatty acyl substrate is dihalogenated on a nonheme iron center (Ramaswamy et al., 2007; Pratter et al., 2014a). In contrast to the nonheme iron halogenase SyrB2, HctB activates a non-amino acid group as a substrate. This would give HctB biotechnological applicability that is beyond that of halogenases like SyrB2. As such, HctB appears to have certain flexibility in the activation of an alkyl chain; therefore, we decided to investigate its mechanism using computational methods and compare the mechanism and active site features with those described for SyrB2 previously. The mechanism of O_2 activation for this mononuclear nonheme iron halogenase was characterized with stopped-flow and spectroscopic (circular dichroism and magnetic circular dichroism) studies (Pratter et al., 2014b).

Functional characterization of the enzyme together with analysis of its primary structure reveals that it is a unique three-domain halogenase containing an acyl carrier protein (ACP), which binds the substrate covalently via a phosphopantetheinyl bridge (**Figure 1**). The other terminus of the ACP group is connected to the C-terminus of an acyl-Coenzyme-A subunit, while the N-terminus is linked to the halogenase domain. As such the protein has an intricate set-up and arrangement that enables dihalogenation of the tethered substrate. Interestingly, no evidence of substrate hydroxylation is available, but the activation of hexanoic acid by HctB apart from dihalogenation also gave products from oxygenation leading to vinyl-chloride and ketone products (Pratter et al., 2014a). HctB shows a certain degree of sequence similarity with SyrB2 (see **Supporting Information** for an overlay and sequence alignment), (Pratter et al., 2014a) and binds iron in a nonheme ligand configuration and utilizes α-ketoglutarate (αKG). The metal in HctB is linked to the protein via two histidine groups (His_{111} and His_{227}) in a pentacoordinate environment. However, there are major differences in the overall structure of SyrB2 vs. HctB, which are currently understood and warrant a detailed computational study.

After αKG and halide binding, the metal binds molecular oxygen, and it has been hypothesized to react with αKG to form an iron(IV)-oxo species and succinate upon release of CO_2. Unfortunately, the iron(IV)-oxo species in HctB has never been trapped and characterized and details of the halogenation

FIGURE 1 | (A) Structure of the resting state of HctB as taken from Pratter et al. (2014b). The different domains of the secondary structure are identified and a close-up of the active site structure is given. **(B)** General reaction mechanism of the first halogenation step in HctB investigated here.
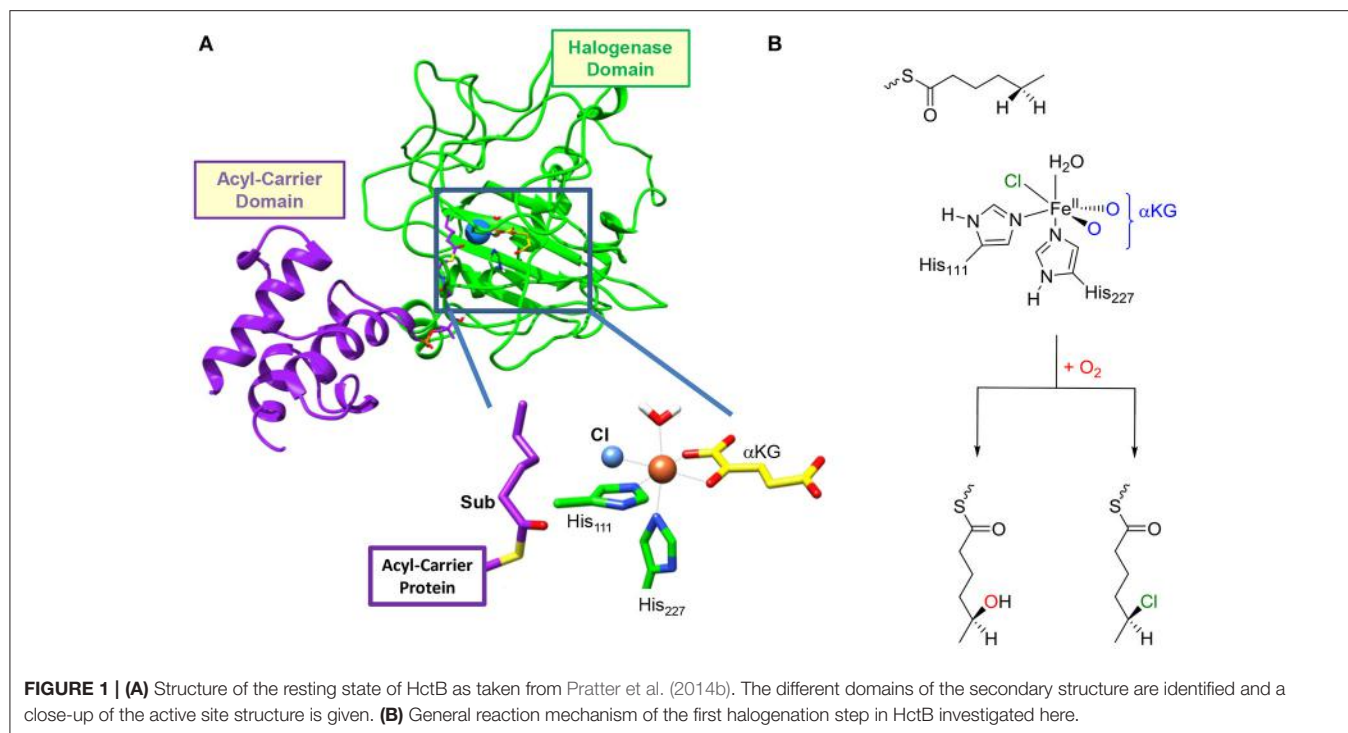
mechanism are unknown. In the next stage of the catalytic cycle it is expected that a hydrogen atom abstraction occurs to form an iron(III)-hydroxo species, which is elusive as well. Technically, however, the halide should rebound to the substrate radical to form halogenated products, but thermodynamically it has been shown that hydroxo rebound is energetically favorable (Timmins and de Visser, 2015). How the enzyme avoids the low-energy hydroxylation pathway in favor of the higher-energy halogenation remains a matter of discussion. In order to understand the effect of substrate binding on the chemoselectivity of substrate halogenation vs. hydroxylation in HctB we employed a detailed molecular mechanics (MM) and quantum mechanics/molecular mechanics (QM/MM) study. We located two substrate entrance channels and have calculated the substrate halogenation and hydroxylation pathways with a substrate in these positions. The work shows that dramatic differences in halogenation vs. hydroxylation product ratios should be expected based upon substrate positioning. Moreover, our calculations predict that the protein induces an electric field effect that withdraws electron density from the halide toward the metal during the reaction in order to make the halogenation process favorable.

RESULTS

Model Selection and Reactant Benchmarking

So far, no computational modeling studies have been reported on HctB and little is known how different (if at all) it is to the well-studied SyrB2. Obviously, both nonheme iron halogenases

utilize a different substrate although both are tethered to a carrier protein. Conversely, SyrB2 catalyzes a monohalogenation of the substrate, whereas HctB performs a dihalogenation instead. A comparison of the amino acid sequences of HctB vs. SyrB2 shows major deviations which must result in differences in secondary structure. Therefore, their mechanisms may be quite different and we decided to do a computational study on the mechanism of halogenation vs. hydroxylation of HctB and how it compares with previous experimental and computational studies of analogous nonheme iron halogenases. In particular, we focused our computational study on the first halogenation step of the fatty acyl tethered substrate of a HctB model and took the structure reported by Pratter et al. (2014a). Our set-up of QM/MM models was reviewed thoroughly recently; therefore, we will summarize the main issues only briefly (Hernández-Ortega et al., 2015; Quesne et al., 2016a; Hofer and de Visser, 2018). The model was altered from an iron(II)-water α-ketoglutarate bound structure into an iron(IV)-oxo with ligated succinate. Prior to the full set-up of the complete QM/MM chemical system, however, we investigated possible substrate binding positions of the tethered hexanoyl-PPT moiety. In particular, we searched for alternative substrate entrance channels into the active site.

Figure 2 shows highlights of the two substrate entrance channels we identified, which are narrow channels that should fit the linear terminal chain of the substrate-carrier protein. The acyl-carrier group of the substrate protein then latches onto the surface of the protein and inserts the tethered hexanoyl-PPT group into the active site. Thus, substrate entrance channel II represents the model from Pratter et al. (2014a) with substrate position **1**. In model **1** with the substrate entering through channel II, the substrate is located parallel to the iron(IV)-oxo

FIGURE 2 | Substrate entrance channels I (in purple) and II (in cyan) into the active site of the halogenase domain of HctB. The yellow circle highlights the active site of the halogenase in the protein. The right-hand-side of the figure gives an overlay of the four substrate binding positions, where we only show the hexanoyl group and highlight the transferring hydrogen atom in white. Substrate binding position 1 is located in entrance channel II and substrate binding position 2, 3, and 4 are bound in entrance channel I.

group in a position very similar to the methylated DNA strand in the AlkB repair enzyme (Quesne et al., 2014). The substrate approaches the iron(IV)-oxo group in between the Val_{113} and Glu_{223} side chains in the corner where the halide group is also located. The entrance channel is located well below the iron(IV)-oxo group.

During analysis of the structure, we identified another entrance channel and attempted to latch the acyl carrier protein onto this and manually inserted the tethered hexanoyl-PPT chain. In the analogous halogenase SyrB2 a similar substrate entrance channel is seen in the same position, (Matthews et al., 2006) but due to the lack of a substrate bound crystal structure of HctB we will consider all possible substrate orientations and entrance channels. Entrance channel I is located above the iron(IV)-oxo group and inserts the tethered hexanoyl group into a large open space (possibly filled with water molecules) and hence we created three starting orientations for the substrate, namely models **2**, **3,** and **4**. In model **2** the terminus of the tethered hexanoyl group points down and hangs in between the Val_{113} and Arg_{245} residues, whereas in model **3** it is found in between Lys_{126} and Pro_{127}. The final substrate position (**4**) through channel I is located on the side of the iron(IV)-oxo group and also brings the terminus of the tethered hexanoyl group into the active site nearby the Val_{113} and Glu_{223} residues. The substrate binding orientation through channel I model **4** brings it in close proximity to both the oxo and halide groups. These three substrate binding positions are distinctly different and guided by hydrogen bonding interactions of the thioester moiety of the hexanoyl-PPT moiety with amino acid residues aligning the channel walls. As such we do not expect easy interconversion between the three substrate binding models.

Subsequently, we investigated the reaction mechanism from each of these different substrate starting orientations and set up QM/MM models with the substrate bound in positions

1, **2**, **3** and **4**. The set-up of the QM/MM models follows previously reported and benchmarked methods (Quesne et al., 2014, 2016a; Hernández-Ortega et al., 2015; Hofer and de Visser, 2018). Firstly, we added hydrogen atoms using known pK_a values and visually inspected polar residues for correct protonation states. Thereafter, we applied an iterative solvation procedure and followed this with an equilibration and heating run to a temperature of 298 K. During these set-up steps the protein backbone was fixed, but in the final molecular dynamics simulation all atoms were allowed to move. We initially ran MD simulations for a period of up to 200 ns; however, the run stabilized in all cases after about 5 ns. Clearly, HctB is a very rigid protein with substrate and active site in tight binding orientation with little flexibility as seen from the MD runs. Therefore, for all subsequent systems described here only a 10 ns MD simulation was taken, see **Supporting Information Figure S1**. From the MD simulations we selected a low energy snapshot after 5 ns as starting structures for the actual QM/MM calculations.

Finally, we bisected the full chemical structure of protein, substrate and water layer into a QM and MM region and included key residues in the QM region that form covalent or hydrogen bonding interactions with substrate and oxidant. Initial, exploratory calculations were done with a small QM region containing only the iron(IV)-oxo(chloro) group and the first-coordination sphere of ligands to the metal, namely the imidazole groups of His_{111} and His_{227}, the acetate terminus of succinate (Succ) and the thiohexanoic acid arm of the substrate as our minimal QM region **A**, see **Figure 3**. This model contains 48 QM atoms and is overall charge neutral. To test the effect of the second coordination sphere we also calculated the full mechanism with a larger QM region that in addition to QM region **A** included the amino acid side chains of residues within 6 Å of the iron(IV)-oxo-chloro structure, i.e. QM region **AB**. For model **1**, the large QM region **AB** contained the amino acid side chains of Ile_{108}, Val_{113}, Asp_{200}, Asp_{202}, Glu_{223}, Val_{225}, Met_{226},

FIGURE 3 | Description of QM region A and AB for Model **1**. Ball-and-stick model of QM region AB is shown on the right-hand-side.
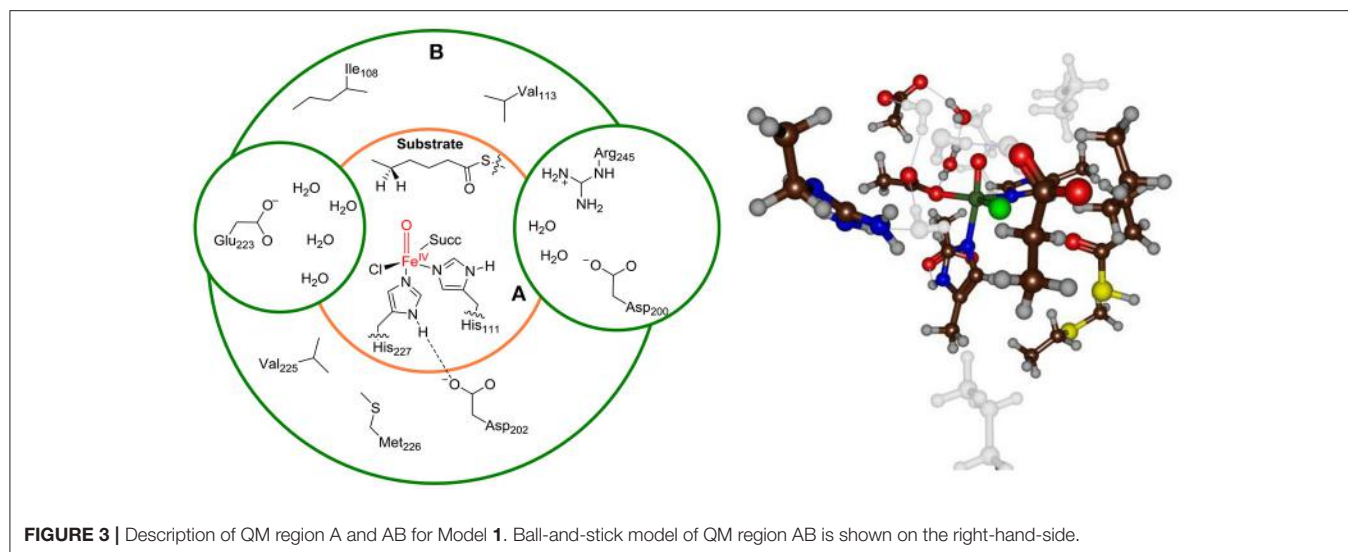
Arg$_{245}$, and six water molecules and has a total of 160 atoms in the QM region.

Thereafter, QM/MM geometry optimizations of the iron(IV)-oxo(chloro) species (**Re**) were performed in the singlet, triplet and quintet spin states. Note that in the label of the structure we give the spin multiplicity in superscript before the label and the substrate binding position and the QM region (**A** or **AB**) in subscript after the label. The full set of results that were obtained is given in the **Supporting Information**, while we focus in the main text on the low-energy pathways only.

We started the work with extensive validation and benchmarking of the methods. Unfortunately, there are no experimental rate constants and spectroscopic data for the iron(IV)-oxo species. Previously, we calibrated thioanisole sulfoxidation free energies of activation of a biomimetic nonheme iron(IV)-oxo complex against experimental data and tested 50 different computational methods and techniques (Cantú Reinhard et al., 2016a). The best agreement with experiment was found for the PBE0/BS2 and B3LYP/BS2 methods with a solvent model included. In particular, free energies of activation were reproduced within 4 kcal mol^{-1} from experiment (Vardhaman et al., 2011, 2013; de Visser et al., 2014; Sainna et al., 2015; Barman et al., 2016a; Cantú Reinhard et al., 2016b). Furthermore, the methods reproduced experimental product distributions of bifurcation processes well (Ji et al., 2015; Kaczmarek et al., 2018). Finally, for the nonheme iron dioxygenase prolyl-4-hydroxylase six hydrogen atom abstraction barriers from substrate were investigated and the QM/MM predicted the correct regioselectivity and therefore the methods are expected to predict regio- and chemoselectivities well (Karamzadeh et al., 2010; Pratter et al., 2013; Timmins and de Visser, 2017; Timmins et al., 2017).

In agreement with the experimental studies on the iron(IV)-oxo species of the halogenase SyrB2, (Galonić Fujimori et al., 2007) we find the quintet spin state as the ground state for all chemical systems. Interestingly, the quintet spin state

is below the triplet spin state by a considerable margin of well over 13 kcal mol^{-1}; hence the triplet and singlet spin states of HctB will not play a major role during the reaction and are high in energy. Therefore, we will focus on the quintet spin state structures and mechanism only here. The **Supporting Information** gives all results of the alternative spin states investigated. Nevertheless, the reaction is expected to proceed through single-state-reactivity on the dominant quintet spin state surface in agreement with previous studies on pentacoordinated iron(IV)-oxo complexes reported before for nonheme iron enzymatic and model complexes (de Visser, 2006a,b, 2010; Hirao et al., 2006; Bernasconi and Baerends, 2008; Latifi et al., 2009; Ye and Neese, 2011; Ansari et al., 2013; Tang et al., 2013; Saouma and Mayer, 2014; Cantú Reinhard and de Visser, 2017a).

The quintet spin state of the iron(IV)-oxo(chloro) species in HctB has four unpaired electrons located in the π^*_{xy}, π^*_{xz}, π^*_{yz}, and σ^*_{x2-y2} molecular orbitals and a virtual σ^*_{z2} orbital. Thus, the π^*_{xz} and π^*_{yz} orbitals represent the antibonding interactions of the metal $3d_{xz}/3d_{yz}$ atomic orbitals with a $2p_x/2p_y$ orbital on the oxo group. These two orbitals are orthogonal and close in degeneracy in nonheme iron(IV)-oxo and located in the xz and yz molecular planes, whereby the z-axis is defined by the Fe–O axis. The π^*_{xy} and σ^*_{x2-y2} orbitals are both in the plane perpendicular to the Fe–O axis (xy-plane) and represent antibonding orbitals with ligands in the equatorial plane, namely the nitrogen of His$_{111}$, the carboxylate of succinate and the halide atom. The final metal-type 3d molecular orbital is the σ^*_{z2} orbital for the σ-type antibonding interaction along the Fe–O bond, which is virtual.

During the equilibration and MD simulation as well as the subsequent QM/MM geometry optimization the positions of the hexanoyl substrate chains have moved slightly. In particular this was the case for model **3** where the terminus moved away from the direction of the iron(IV)-oxo(chloro) group. As a result, the distance between substrate and iron(IV)-oxo(chloro) is quite large (>5 Å) in model **3**. **Figure 4** shows the

FIGURE 4 | QM/MM optimized geometries of the quintet spin iron(IV)-oxo reactant complexes as calculated at UB3LYP/BS1 level of theory in Turbomole:Charmm. Structures were optimized with either QM region A or QM region AB. Reactant structures for models **1–4** were calculated and are grouped by model: top row model **1** and model **2**, bottom row model **3** and **4**. Bond lengths are given in angstroms. Water molecules and aliphatic residues are shaded.

eight different optimized geometries of the iron(IV)-oxo(chloro) reactant species as calculated with QM/MM for models **1**, **2**, **3** and **4** with either QM region **A** or **AB**. As can be seen, all optimized structures, regardless of the substrate binding position, give very similar optimized geometries with Fe–O distances ranging from 1.621 to 1.641 Å and Fe–Cl distances between 2.278 and 2.326 Å. These values match previous calculations on nonheme iron(IV)-oxo(chloro) and nonheme iron(IV)-oxo complexes excellently (de Visser, 2006a,b, 2010; Hirao et al., 2006, 2011; Matthews et al., 2006; Noack and Siegbahn, 2007; Bernasconi and Baerends, 2008, 2013; de Visser and Latifi, 2009; Kulik et al., 2009; Latifi et al., 2009; Pandian et al., 2009; Dey, 2010; Ye and Neese, 2011; Quesne and de Visser, 2012; Ansari et al., 2013; Liu et al., 2013; Tang et al., 2013; Usharani et al., 2013; Kumar et al., 2014; Saouma and Mayer, 2014; Senn, 2014; Huang et al., 2016; Zhao et al., 2016; Cantú Reinhard and de Visser, 2017a; Timmins et al., 2018).

The QM/MM results from **Figure 4** show that the secondary environment and the substrate positioning have little or no effect on the Fe–O and Fe–Cl distances in the iron(IV)-oxo(chloro) species although it may affect the kinetics as shown below. In addition, no major electronic differences are seen for the eight optimized structures and all converge to a quintet spin ground state. To be specific, group spin densities of all structures are close with values of around $\rho_{Fe} = 3.12$ and $\rho_O = 0.63$. The orbital analysis and group spin densities implicate that most

spin density is located along the Fe–O bond and point to an orbital occupation of $\pi^{*1}_{xy} \, \pi^{*1}_{xz} \, \pi^{*1}_{yz} \, \sigma^{*1}_{x2-y2}$ for all optimized geometries. Our calculations are in agreement with experimental EPR and Mössbauer spectroscopy studies on the analogous structure in SyrB2 that also reported a high-spin ground state (Galonić Fujimori et al., 2007).

The only noticeable difference between the four substrate binding positions relates to the orientation of the active site Arg$_{245}$ residue, which is locked in hydrogen bonding interactions to the dangling terminal carboxylate group of succinate via a bridging water molecule in model **1** and **2**. By contrast, in model **3** the Arg$_{245}$ group forms a direct salt bridge with the carboxylate of the succinate moiety and also is in close proximity to the oxo group. Finally, in structure **4**, the Arg$_{245}$ forms a salt bridge with Glu$_{223}$, whereby it appears to close entrance channel II. It may very well be, therefore, that the Arg$_{245}$ residue is involved in α-ketoglutarate binding and/or succinate release from the active site. Indeed, Arg$_{245}$ is a conserved residue in most reported nonheme iron halogenases and therefore, is expected to play a key role in catalysis and/or substrate positioning. On the other hand, Glu$_{223}$ is not conserved in the majority of reported nonheme iron halogenases and only found in HctB. It would be interesting to see how mutation of either Glu$_{223}$ or Arg$_{245}$ affects the enzyme function, chemical catalysis and product distributions. But that will have to await a future experimental study.

Chemoselectivity Patterns

Next, the chemoselectivity of substrate halogenation vs. hydroxylation was investigated for the pathways described and defined as in **Scheme 1**. The reaction starts from the reactant complex of iron(IV)-oxo(chloro) with substrate (**Re**) and proceeds via a stepwise mechanism with an initial hydrogen atom abstraction transition state (**TS$_{HA}$**) leading to an iron(III)-hydroxo complex and a substrate radical (**I$_{HA}$**). In the next step the pathways diverge and either OH rebound (via transition state **TS$_{OH}$**) or Cl rebound (via transition state **TS$_{Cl}$**) occurs. These mechanisms then lead to the hydroxylated products (**P$_{OH}$**) and halogenated products (**P$_{Cl}$**), respectively. Technically, the reaction apart from being chemoselective is also enantioselective where only one of the two isomers is expected. Thus, due to tight substrate binding and positioning, the hydrogen atom abstraction will be selective from only one of the two hydrogen atoms from the ω-1 position of the substrate. Hence the hydrogen atom abstraction will guide the enantioselectivity as seen before on related enzymes (Karamzadeh et al., 2010; Pratter et al., 2013; Timmins and de Visser, 2017; Timmins et al., 2017). Nevertheless, in HctB the first halogenation of substrate is followed by a second halogen transfer through the binding of another molecule of O_2 and Cl^- to repeat the cycle and the formation of the dihalogenated product.

We tested the mechanisms for the first halogenation step in HctB on all low lying spin states (singlet, triplet, quintet), for each of the substrate binding positions (**1, 2, 3, 4**) and with different QM regions (**A** or **AB**). Previously, we found the septet spin state of nonheme iron(IV)-oxo to be at least 7 kcal mol^{-1} higher in energy than the quintet spin state and, therefore, we did not investigate this state further (Latifi et al., 2009). However, we tested the models and methods by using different optimization techniques (basis set, DFT method, snapshot).

Firstly, changing the basis set from SV(P) to TV(P) on all atoms had little effect on the optimized geometries and lowered the hydrogen atom abstraction barrier by only a small amount. In particular, geometry optimizations at UB3LYP/BS1 and UB3LYP/BS2 of ^{5}**Re$_1$** and ^{5}**TS$_{HA,1}$** using model **1** and the large QM region **AB** gave hydrogen atom abstraction barriers of 23.5 kcal mol^{-1} with UB3LYP/BS2 and 21.5 kcal mol^{-1} with UB3LYP/BS2//UB3LYP/BS1. In addition, the optimized

geometries obtained with UB3LYP/BS1 vs. UB3LYP/BS2 are very similar (vide infra). Therefore, we continued with geometry optimizations at UB3LYP/BS1 level of theory for the remainder of the project.

Secondly, choosing alternative snapshots from the MD simulation also had little effect on the optimized geometries of the reactant complexes of nonheme iron complexes and did not change spin state orderings and relative energies dramatically. This probably is as nonheme iron dioxygenases often have very rigid structures where the substrate is tightly bound (Aluri and de Visser, 2007; Kumar et al., 2011; Tchesnokov et al., 2016; Faponle et al., 2017). Previous QM/MM studies on analogous nonheme iron dioxygenases indeed reproduced experimental chemoselectivities excellently using the same methods and procedures described here (Aluri and de Visser, 2007; Karamzadeh et al., 2010; Kumar et al., 2011; Pratter et al., 2013; Quesne et al., 2014, 2016a; Hernández-Ortega et al., 2015; Tchesnokov et al., 2016; Faponle et al., 2017; Timmins and de Visser, 2017; Timmins et al., 2017; Hofer and de Visser, 2018). Moreover, changing the QM region, basis set or density functional method did not change the ordering of the transition states.

Let us first focus on the hydrogen atom abstraction step, which is rate-determining in the overall reaction mechanism and optimized geometries for all four substrate binding positions are shown in **Figure 5**. Hydrogen atom abstraction from a quintet spin iron(IV)-oxo species with configuration $\pi^{*1}_{xy}\ \pi^{*1}_{xz}\ \pi^{*1}_{yz}\ \sigma^{*1}_{x2-y2}$ leads to a sextet spin iron(III)-hydroxo with a nearby radical, or alternatively a quartet spin iron(III)-hydroxo with a nearby radical. Thus, the abstracted electron from the hydrogen atom either pairs up with the lowest π^*_{xy} electron to form a quartet spin iron(III)-hydroxo species or enters the virtual σ^*_{z2} orbital to give a sextet spin iron(III)-hydroxo species. Previously, (de Visser, 2006b; Bernasconi and Baerends, 2008; Ye and Neese, 2011; Ansari et al., 2013; Tang et al., 2013; Cantú Reinhard and de Visser, 2017a) it was shown that hydrogen atom abstraction by a quintet spin pentacoordinated nonheme iron(IV)-oxo species leads to electron transfer to the metal that preferentially fills the σ^*_{z2} orbital with one electron, designated the $^5\sigma$-pathway. Alternatively, the electron can move into one of the lower-lying and singly occupied π^* orbitals: the so-called $^5\pi$-pathway.

SCHEME 1 | Reaction mechanism investigated in this work.

FIGURE 5 | (A) Hydrogen atom abstraction barriers (structures and energies, in kcal mol^{-1}) for hydrogen atom abstraction from the C5 position of the hexanoyl substrate terminus by the iron(IV)-oxo(chloro) species of HctB as calculated with QM/MM at UB3LYP/BS2/ /UB3LYP/BS1:Charmm with Turbomole using QM region AB. (B) Optimized geometries of the hydrogen atom abstraction barriers with bond lengths in angstroms, the Fe–O–C^5 angle in degrees and the imaginary frequency in wave numbers.

A spin density and molecular orbital analysis for all hydrogen atom abstraction transition states revealed that all are of $^5\sigma$-type leading to a radical intermediate with five unpaired electrons in metal-type orbitals ($\pi^{*1}_{xy} \, \pi^{*1}_{xz} \, \pi^{*1}_{yz} \, \sigma^{*1}_{z2} \, \sigma^{*1}_{x2-y2}$) and a down-spin electron on the substrate. Thus, in all transition states there is negative spin density on the substrate (ranging from -0.27 to -0.47) while the spin density on iron increases to 4.01–4.14. These values are in agreement with hydrogen atom abstraction reactions calculated before for nonheme iron dioxygenases and synthetic model complexes (Karamzadeh et al., 2010; Hirao et al., 2011; Pratter et al., 2013; Kumar et al., 2014; Zhao et al., 2016; Timmins and de Visser, 2017; Timmins et al., 2017). We attempted to find a transition state for the $^5\pi$-pathway for several models, but found them considerably higher in energy and we did not manage to optimize those structures with QM/MM. Filling of the σ^{*}_{z2} orbital with one electron, as happens in the $^5\sigma$-pathway, usually means that the hydrogen atom abstraction is performed along the molecular z-axis, i.e., a linear Fe–O–C angle is seen in the transition state. However, due to constraints in substrate position and orientation and the location of amino acid residues in that region, in the protein the substrates are unable to approach the metal-oxo under this "ideal" angle. As a matter of fact, only in model **3** the substrate can approach the iron(IV)-oxo(chloro) from the top and an almost linear Fe–O–C^5 angle is found, while the angle is considerably less for the other structures. As a result, the barrier is the lowest for model **3**. Nevertheless, even with the transition states in the bent orientation they are still favored in the $^5\sigma$-pathway and no $^5\pi$-pathway is found. Thus, even though the transition states are not in the geometric ideal position, the lowest energy pathway is the $^5\sigma$-pathway. In a previous study on a biomimetic model complex, we investigated hydrogen atom abstraction from

dihydroanthracene by a pentacoordinated iron(IV)-oxo species, where the z-axis was blocked by ligands, so that the substrate had to approach from the side. For that system, analogously to what is seen here, the $^5\sigma$-pathway was still the lowest in energy (Latifi et al., 2013). Therefore, geometric constraints cannot overcome the orbital energy differences between the $^5\sigma$ − and $^5\pi$ pathways for the HctB active site.

As can be seen from **Figure 5**, the approach of the substrate on the iron(IV)-oxo center has a profound effect on the barrier height for hydrogen atom abstraction. In particular, pathway **3** appears to be catalytically most effective with a low hydrogen atom abstraction barrier of only 4.9 kcal mol^{-1} (UB3LYP/BS2//UB3LYP/BS1) for QM region **AB**. This is thanks to its geometric orientation, whereby substrate approach to the catalytic center is from the top and the donating hydrogen atom is aligned with the Fe–O axis. With substrates approaching through channels **1**, **2**, and **4**, however, the Fe–O–H angle is strongly bent and the hydrogen atom abstraction barriers are dramatically raised in energy, although they still should be accessible at room temperature.

All transition state geometries display short C–H and long O–H bonds as well as relatively long O–C distances. It may very well be that the enzyme has evolved to do this on purpose in order to guide the chemoselectivity of the reaction and prevent unwanted (more favorable) side reactions. Thus, alternative nonheme iron dioxygenases, such as taurine/αKG dioxygenase have the substrate located in a pocket nearby the iron(IV)-oxo species and a QM/MM study provided a substantially lower hydrogen atom abstraction barrier than found for HctB (Borowski et al., 2004; de Visser, 2006c; Nemukhin et al., 2006; Cicero et al., 2007; Sinnecker et al., 2007; Godfrey et al., 2008; Chen et al., 2011; Bushnell et al., 2012; Mai and Kim,

2016; Wójcik et al., 2016; Álvarez-Barcia and Kästner, 2017). All hydrogen atom abstraction transition states display a large imaginary frequency for the C–H–O stretch vibration; however, distinct differences in the magnitude are seen. In particular, the barriers $^5TS_{HA,2}$ and $^5TS_{HA,3}$ have imaginary frequencies well over i1,000 cm^{-1} and hence considerable amount of tunneling can be expected. Indeed, previous computational studies on hydrogen atom abstraction reactions by iron(IV)-oxo complexes predicted significant tunneling but also a huge change in barrier upon replacement of the transferring hydrogen atom by deuterium (Kumar et al., 2003, 2004; de Visser, 2006d; Quesne et al., 2016b; Cantú Reinhard et al., 2017b). In particular, the kinetic isotope effect was shown to linearly increase with the value of the imaginary frequency in the transition state.

Table 1 summarizes all hydrogen atom abstraction barriers for models 1, 2, and 4. In general, for model 2 and 4 the UB3LYP/BS1 and UB3LYP/BS2//UB3LYP/BS1 energies are very close, whereas a change from the small QM region to the larger one leads to a drop of about 4 kcal mol^{-1} in the transition state energy. More dramatic changes are seen for model 1. Thus, with the small QM region large variations in the transition state energy are found dependent on the calculation method and basis set. Moreover, the energies obtained for the small QM region are quite different from those for the larger QM region, where little fluctuation upon changing the basis set is observed. Therefore, the large QM region is the most suitable chemical system for our calculations and hence we used that for the rest of the work.

We even did a full geometry optimization at UB3LYP/BS2:Charmm with QM/MM for the same reaction mechanism using both QM region A and AB for model 1. The UB3LYP/BS2 QM/MM optimized geometry of $^5TS_{HA,1}$ is located at a value of 23.5 kcal mol^{-1} above the reactant complex, which is within 2 kcal mol^{-1} of the UB3LYP/BS1 and UB3LYP/BS2//UB3LYP/BS1 results. Moreover, the effect of the basis set on the structures of the optimized geometries is small as well and geometrically they are very close (Supporting Information). Thus, the C–H and H–O distances change from 1.22/1.45 Å calculated with UB3LYP/BS1:Charmm to 1.22/1.43 Å as obtained with UB3LYP/BS2:Charmm for model 1.

There is quite a bit of fluctuation in the hydrogen atom abstraction barrier depending on the position of the substrate (Figure 5). The lowest barrier is found for model 3 and is

4.9 kcal mol^{-1} in energy (UB3LYP/BS2//UB3LYP/BS1:Charmm with QM region AB), while the substrate approach via models 1, 2, and 4 are higher in energy by 16.6, 11.5, and 15.2 kcal mol^{-1}, respectively. The latter three, although somewhat higher in energy than model 3, have barrier heights that are still within range of what would be an accessible hydrogen atom abstraction barrier at room temperature. It is very interesting to see that models 1, 2, and 4 give hydrogen atom abstraction barriers to within a couple of kcal mol^{-1} of each other and hence, the substrate entrance and positioning appears to have a small effect on the ability of the iron(IV)-oxo(chloro) species to react with substrate. As the substrate is connected to an acyl-carrier protein that is latched on the surface of the protein, the hydrogen atom abstraction barrier is dependent on the distance the substrate terminus can be inserted into the protein. We show here that in all four orientations (models 1, 2, 3, and 4) the substrate entrance leads to a viable hydrogen atom abstraction channel; however, the ease of hydrogen atom abstraction is strongly influenced by its positioning. Nevertheless, a low hydrogen atom abstraction barrier not necessarily correlates with the correct chemoselectivity of reaction, which is determined in the subsequent reaction step. Therefore, even though the hydrogen atom abstraction is rate determining, it does not decide the product distributions.

The hydrogen atom abstraction reaction leads to an iron(III)-hydroxo(chloro) radical intermediate (I_{HA}) and proceeds to products through either OH or Cl rebound to the substrate radical to form the products P_{OH} and P_{Cl}. For all four substrate position orientations (1, 2, 3, and 4), we then investigated the mechanisms for either Cl rebound or OH rebound. As can be seen from Figure 6, dramatic changes in rebound as well as chemoselectivity of the reaction are obtained depending on the position of the substrate. The ordering and relative energies are not dependent on the choice of the basis set or the size of the QM region: all results point in the same direction and give the same conclusions (Supporting Information). With the substrate entering from a channel above the iron(IV)-oxo species, like through models 1 and 2, the chemoselectivity is in favor of halogenation over hydroxylation by more than 10 kcal mol^{-1}. In particular, the halide rebound step was found to have a negligible barrier in both cases; therefore, halide transfer will be fast in both cases.

A complete reversal of the chemoselectivity is seen in model 3, but now the OH rebound is barrierless. By contrast, almost identical OH and Cl rebound barriers are found in model 4. Clearly, the substrate position and orientation has a major effect in the bifurcation pathways and the chemoselectivity of the reaction. Moreover, the calculations presented in Figure 6 implicate that HctB has two viable substrate entrance channels (I and II in Figure 2) and models with the substrate in either of these entrance channels give low-energy hydrogen atom abstraction barriers and preferential halogenation over hydroxylation. However, substrate entrance through channel I in HctB can lead to the location of the substrate in various positions, whereby the one in model 4 probably will give a mixture of products and position 3 gives substrate hydroxylation. As such, substrate positioning and orientation is vital for efficient

TABLE 1 | Transition state energies of $^5TS_{HA}$ as calculated with different QM sizes and basis sets for model 1, 2, and 4[a].

QM region	QM method	$\Delta E^{\ddagger}_1$	$\Delta E^{\ddagger}_2$	$\Delta E^{\ddagger}_4$
A	UB3LYP/BS1	24.5	19.8	21.2
A	UB3LYP/BS2//UB3LYP/BS1	27.4	20.4	23.3
AB	UB3LYP/BS1	22.9	16.2	16.2
AB	UB3LYP/BS2//UB3LYP/BS1	21.5	16.4	20.1
AB	UB3LYP/BS2	23.5		

[a]In kcal mol^{-1}.

FIGURE 6 | Energy diagram (in kcal mol⁻¹) for the pathways leading to halogenation or hydroxylation products from the iron(III)-hydroxo(chloro) intermediate ($^5I_{HA}$) as calculated with QM/MM in Turbomole:Charmm. Energies (in kcal mol⁻¹) obtained with QM/MM at UB3LYP/BS2//UB3LYP/BS1:Charmm and are the result of a full geometry optimization of the local minima and transition states. Energies are taken relative to the value of the radical intermediate state.

halogenation in HctB and we predict the enzyme to latch the acyl carrier into substrate entrance channel II and position the substrate in the orientation as shown in model **1**. The structure and orientation of the substrate in model **1** leads to a reaction mechanism with thermochemically accessible hydrogen atom abstraction barrier at room temperature and chemoselectivity preference for substrate halogenation as expected from a halogenase. Most likely, therefore, the other channel located from our structure analysis (substrate entrance channel I) could be involved in guiding halide, dioxygen or α-ketoglutarate reactants into the active site and release of succinate and CO_2.

The hydrogen atom abstraction barriers reported above in **Figure 5** already give a clue on the potential chemoselectivity of the reaction. Thus, in structure $^5TS_{HA,1}$ and $^5TS_{HA,2}$ the Fe–Cl bond has elongated from about 2.28 Å in the reactant complexes to 2.42 and 2.33 Å, respectively. In both cases, the substrate radical is positioned in the quadrant in between the hydroxo and chloride ligands. On the other hand, in $^5TS_{HA,3}$ the radical is located above the hydroxo group and far away from chloride, thereby making chloride rebound unlikely. To understand the large chemoselectivity preference of halogenation over hydroxylation by model **2**, we show in **Figure 7** the optimized geometries of the radical intermediates $^5I_{HA}$ as obtained with QM region **A** and **AB**. The structures are very similar regarding of the size of the QM region chosen. The Fe–O bond has elongated from 1.62 Å in $^5Re_2$ to 1.90 (1.88) Å in $^5I_{HA}$ using QM region **A** (**AB**), respectively. This is due to occupation of the σ^*_{z2} orbital with one electron in the radical intermediate, which is antibonding along the Fe–O bond and hence leads to its elongation. Although the radical has dissipated from the metal center slightly, it is nearer to the halide than to the hydroxo group by almost 1 Å (**Figure 7**). Moreover, the hydrogen atom of the hydroxo group points in the direction of the radical. However, for hydroxo rebound it needs to move out of the way to make the C–O bond formation possible. In the gas-phase and in apolar environments this O–H rotation costs little energy and hydroxo rebound tends to be low in energy as seen in many aliphatic hydroxylation studies reported previously (Ogliaro et al., 2000; Dey, 2010; Hirao et al., 2011; Bernasconi and Baerends, 2013; Liu et al., 2013; Usharani et al., 2013; Kumar et al., 2014; Zhao et al.,

FIGURE 7 | QM/MM optimized geometries of the radical intermediates ($^5I_{HA}$) of model **2** as calculated with UB3LYP/BS1:Charmm in Turbomole:Charmm. Data are given for QM region A (AB) with bond lengths in angstroms.
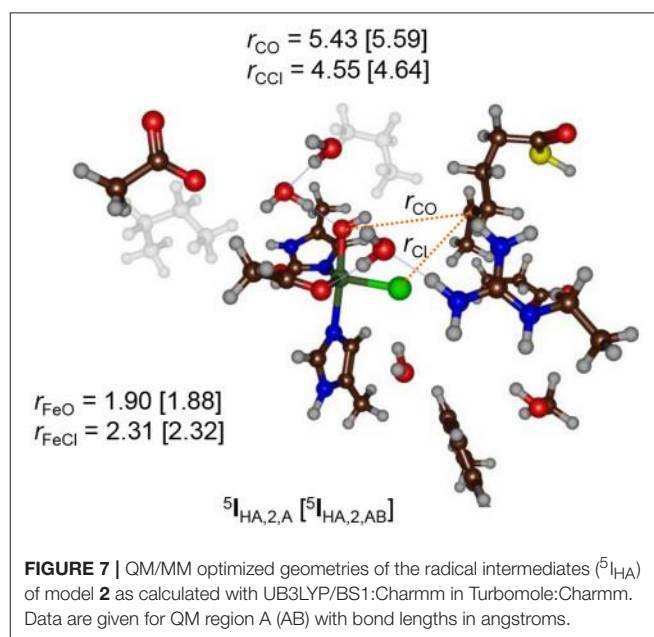
2016). By contrast, in the HctB models **1** and **2** the OH rebound barrier is high in energy due to the largely polar active site with lots of water molecules that prevent the OH group from rotation and keep it in a fixed orientation.

Recent QM/MM studies on the nonheme iron halogenase SyrB2 proposed a strong hydrogen bonding interaction between the hydroxo group in the radical intermediate with an active site Arg residue (Huang et al., 2016). In our optimized geometries of the radical intermediates ($^5I_{HA}$) no direct interaction was found for any substrate binding position, although in several cases a bridging hydrogen bonded water molecule was located in between, see **Figure 7**. The structure displayed in **Figure 7** shows that the water channel that enters the active site pocket serves as a drive to prevent OH rebound and trigger an alternative reaction pathway, namely halogenation. A recent QM/MM study on a cytochrome P450 peroxygenase also had an optimized geometry for an iron(IV)-hydroxo(heme) with an active site whereby

the hydroxo group underwent several hydrogen bonding interactions with crystal water molecules in the pocket (Faponle et al., 2016). Studies on the bifurcation pathways for substrate decarboxylation vs. hydroxylation also gave a preference for the energetically unfavorable decarboxylation mechanism similar to what is seen here.

Of course, it should be mentioned that the iron(II)-hydroxo complex with halogenated substrate P_{Cl} is not the final step in the catalytic cycle. One could envisage a proton transfer to generate iron(II)-water; however, our structure analysis did not characterize a proton-transfer channel, so that this pathway may not be feasible. Instead, we predict that after the formation of P_{Cl}, another chloride ion enters the substrate binding pocket and binds to the iron(II) center. A subsequent hydrogen atom abstraction by the iron(II)-hydroxo(chloro) and rebound of halide then gives the iron(II)-water resting state and a dihalogenated substrate as a product.

DISCUSSION

To gain further insight into the factors that determine the bifurcation patterns of hydroxylation vs. halogenation, we analyzed the chemoselectivity-determining transition states and investigated the role of the protein environment on their ordering and relative energies. In particular, we compare our QM/MM results with small model DFT calculations from previous work, which should give an indication how the protein constraints to the active site structure and how the long-range electrostatic interactions through the protein affect the chemoselectivity of the reaction. Specifically, small gas-phase model complexes of the nonheme iron halogenase reaction mechanism failed to find preferential halogenation over hydroxylation, (de Visser and Latifi, 2009; Kulik et al., 2009; Pandian et al., 2009; Huang et al., 2016) and only the inclusion of the full enzymatic structure and the QM/MM approach led to the correct chemoselectivity for SyrB2 (Borowski et al., 2010).
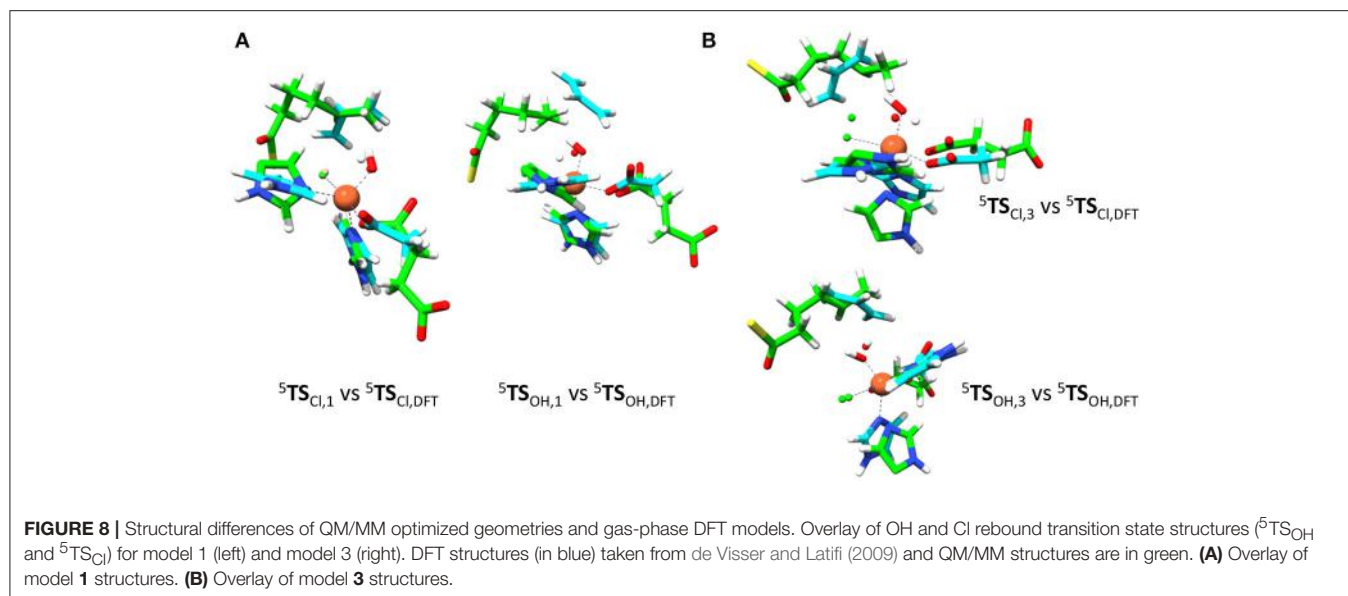
Therefore, we took the DFT optimized geometries from de Visser and Latifi (2009) calculated at UB3LYP/BS1 with a polarized continuum model with dielectric constant of $\varepsilon = 5.7$ included and perturb the small model complexes with external charges, electric fields and dipole moments and investigate the effects on the product distributions. As will become clear in the discussion, these external perturbations have a profound effect on the chemoselectivity of the reactions. Thus, we started with a detailed comparison of the QM/MM optimized geometries with the gas-phase DFT models from de Visser and Latifi (2009), see **Figure 8**. Of course, the gas-phase DFT model ignores the secondary coordination sphere of atoms and has large flexibility due to the absence of structural constraints. Indeed, we see from the optimized geometries that the substrate radical in the DFT model approaches the oxidant under an ideal angle (with little or no stereochemical repulsions of active site residues) to the iron(III)-hydroxide(halide) complex $^5I_{HA}$.

Figure 8 shows overlays of $^5TS_{Cl}/^5TS_{OH}$ structures of models **1** and **3** as compared to those obtained with the DFT model from (de Visser and Latifi, 2009). Overlays of models **2** and **4**

with these structures are given in the **Supporting Information** and show a similar pattern. As can be seen the optimized $^5TS_{Cl,1}$ structure overlaps the gas-phase DFT geometry almost perfectly with the substrate in the same location and under almost the same angle. However, the substrate location is far from the gas-phase orientation for $^5TS_{OH,1}$. Therefore, the substrate binding pocket appears to be designed to accommodate the substrate for efficient halogen transfer and not OH rebound as would be expected of a halogenase enzyme. Despite the good geometric agreement between $^5TS_{Cl,1}$ and the analogous DFT model, the barrier heights relative to their precursor complexes are quite different. Thus, an almost negligible Cl transfer barrier is found for model **1** (**Figure 6**), whereas the DFT model gave a much larger barrier of about 10 kcal mol^{-1} (de Visser and Latifi, 2009). Clearly, even though the $^5TS_{Cl,1}$ structures are almost the same, the QM/MM barrier is well lower in energy. Accordingly, the protein must stabilize the transition state for halogen transfer dramatically through long-range electrostatic interactions.

We then investigated the overlays of $^5TS_{Cl,3}/^5TS_{OH,3}$ with those obtained with the DFT model (**Figure 8B**). As can be seen, the OH rebound structure shows a reasonable match between the two models and hence rebound should stay low in energy. The overlay between the halogen transfer barriers of model **3** and the DFT model are much less good than seen for model **1** and explain why the chemoselectivity is reversed. The comparison of the gas-phase and QM/MM optimized structures, therefore, shows that the substrate should bind in a favorable position for halide rebound. In addition, as shown in **Figure 7**, water molecules prevent a low-energy OH-rotation pathway and lock the OH group in a non-rebound position. Consequently, substrate positioning in combination with a strong hydrogen bonding network around the hydroxo group destabilize OH rebound and drive the reaction toward halide transfer. A similar conclusion was reached by Mitchell et al. (2017) who reported experimental studies on protein engineering of the nonheme iron halogenase WelO5.

Subsequently, we explored the contribution of the protein environment to the chemoselectivity patterns. We initially investigated perturbations affecting the energetics of the $^5TS_{Cl}$ and $^5TS_{OH}$ barriers for the small model complex and particularly focused on electric field effects using previously described procedures (Shaik et al., 2004). We took the gas-phase DFT model and calculated the electronic energy under the influence of an applied electric field with magnitude 0.0050, 0.0100, and 0.0150 au along each individual coordinate axis in both positive and negative directions. Thus, the x-axis is along the Fe–Cl bond, the y-axis along the Fe–OH bond and the z-axis along the Fe–N(His) bond. Relative energies of the two bifurcation transition states $^5TS_{Cl}$ over $^5TS_{OH}$ were investigated with electric field effects for the small model and **Figure 9** shows the ordering and relative energies of $^5TS_{Cl}/^5TS_{OH}$ through perturbations along the x-axis, i.e., along the Fe–Cl bond. Similar calculations were also performed along the y- and z-axis, whereby the y-axis is defined along the Fe–OH bond and the z-axis is located along the Fe–N(His) bond. The results for the field along the y- and z-axis is given in the **Supporting Information**; however, in those cases the energies did not change by more than 3 kcal mol^{-1}

FIGURE 8 | Structural differences of QM/MM optimized geometries and gas-phase DFT models. Overlay of OH and Cl rebound transition state structures ($^5TS_{OH}$ and $^5TS_{Cl}$) for model 1 (left) and model 3 (right). DFT structures (in blue) taken from de Visser and Latifi (2009) and QM/MM structures are in green. **(A)** Overlay of model **1** structures. **(B)** Overlay of model **3** structures.

and in all cases $^5\mathbf{TS}_{OH}$ stayed well below $^5\mathbf{TS}_{Cl}$. The results in **Figure 9** show that without an electric field or a field along the negative x-axis, the hydroxylation pathway will be favored over the halogenation pathway by more than 7 kcal mol^{-1}. With an electric field pointing along the positive x-axis, by contrast, a dramatic change in relative energies between the two barriers is seen and a chemoselective halogenation over hydroxylation is observed.

To understand the chemoselectivity reversal through the addition of an electric field, we analyzed the unpaired group spin densities and charges of the individual complexes. The group charges under the addition of an electric field to the transition states are given at the bottom of **Figure 9**. A clear trend emerges from these calculations. Thus, in both **TS**'s charge density (Q) is removed from Cl upon increasing the electric field along the positive x-axis. In particular, the negative charge of the chloride atom decreases from $Q_{Cl} = -0.61$ with a field of $E_{0, X} = -0.015$ au to a value of $Q_{Cl} = -0.31$ with a field of $+0.015$ au, while an increase of the negative charge with 0.18 units is seen for the same field strengths for $^5\mathbf{TS}_{Cl}$. At the same time, the charge on the OH group stays virtually the same in $^5\mathbf{TS}_{OH}$, whereas it increases from -0.32 to -0.46 in $^5\mathbf{TS}_{Cl}$. Therefore, an applied electric field reduces the charge on chloride and puts more radical character on Cl and enables a Cl$^\bullet$ transfer. At the same time, charge is transferred to iron that is reduced from iron (III) to iron (II).

As a result the $^5\mathbf{TS}_{Cl}$ barrier drops below the $^5\mathbf{TS}_{OH}$ barrier when a positive electric field is applied. It also appears that the $^5\mathbf{TS}_{OH}$ barrier is more strongly affected than the $^5\mathbf{TS}_{Cl}$ barrier to an external electric field and hence the chemoselectivity switch can occur due to destabilization of the hydroxyl rebound barrier. The calculation with an electric field of 0.015 au along the positive x-axis shows that the substrate group loses most of its spin density with respect to the unperturbed system, namely the spin density on the substrate moiety drops from $\rho_{Sub} = -0.68$ for the system

without an electric field to $\rho_{Sub} = -0.12$ with an electric field of $E_{0, X} = +0.015$ au. At the same time also the Cl and OH groups lose radical character but to a much lesser extent than the substrate. These results are in line with our previous work on the chemoselectivity of aliphatic hydroxylation vs. epoxidation by a cytochrome P450 model Compound I complex, where we also showed a chemoselectivity switch upon addition of an electric field along a specific axis (Shaik et al., 2004). Further, DFT calculations on a Compound I model of cytochrome c peroxidase with an applied electric field identified radical character on the porphyrin in one direction, whereas with the electric field in the opposite direction a tryptophan radical was obtained (de Visser, 2005).

Our studies, therefore, highlight the importance of the shape and size of the protein of HctB that create an induced electric field on the active site and thereby favorably lead to substrate halogenation to substrate halogenation of the substrate while minimizing substrate hydroxylation. Evidently, an applied electric field from the dipole moment of the protein affects the charge distributions and spin densities of intermediates and transition states in the reaction mechanism and thereby affects the product distributions. In the HctB halogenase enzyme the protein perturbs the active site structure with an induced electric field in such a way that it stabilizes the halogenation pathway and lowers it below the hydroxylation pathway. This effect is the result of considerable destabilization of the OH rebound barrier and a smaller stabilization of the Cl rebound as based upon the changes of the charge distributions between no field and a positive electric field.

To find further evidence of environmental perturbations influencing the halogenase vs. hydroxylase activity of the protein, we searched for charged residues in the vicinity of the iron(IV)-oxo(chloride) complex in the pdb structures of $^5\mathbf{Re_1}$, $^5\mathbf{Re_2}$, $^5\mathbf{Re_3}$, and $^5\mathbf{Re_4}$. Nearby the halide atom we locate the Glu$_{223}$ and Arg$_{245}$ side chains, whereby the latter may be involved in α-ketoglutarate
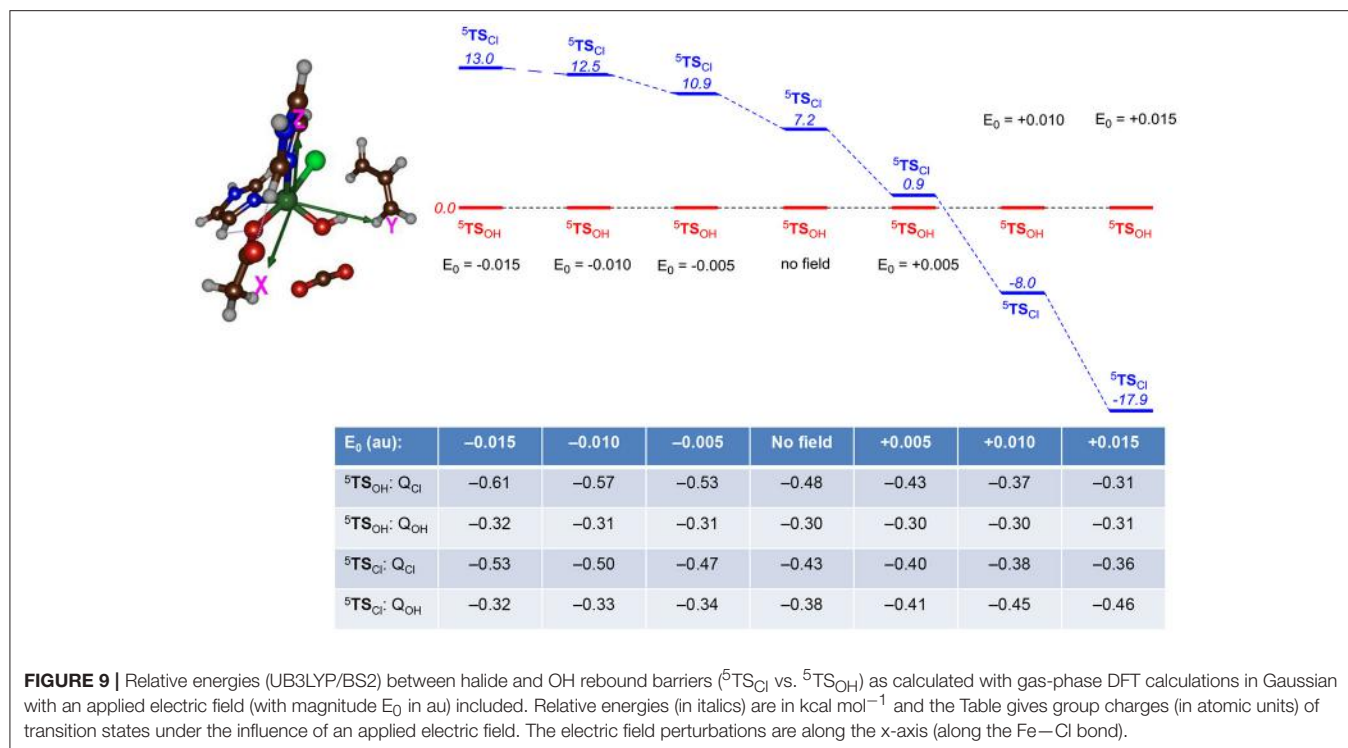
FIGURE 9 | Relative energies (UB3LYP/BS2) between halide and OH rebound barriers ($^5TS_{Cl}$ vs. $^5TS_{OH}$) as calculated with gas-phase DFT calculations in Gaussian with an applied electric field (with magnitude E_0 in au) included. Relative energies (in italics) are in kcal mol^{-1} and the Table gives group charges (in atomic units) of transition states under the influence of an applied electric field. The electric field perturbations are along the x-axis (along the Fe—Cl bond).

binding and succinate release. The top part of **Figure 10** gives the relative orientations of the Glu$_{223}$ and Arg$_{245}$ amino acids with respect to the iron(IV)-oxo(chloride) in the reactant complexes. As can be seen, the three enzyme models show considerable differences in the position of these amino acids that may incur Coulombic interactions with the active site a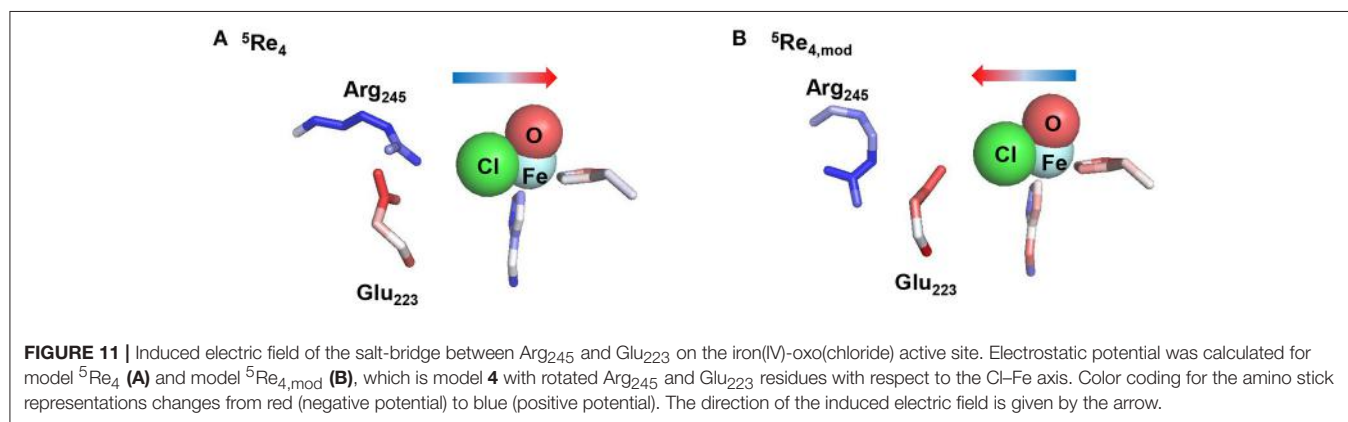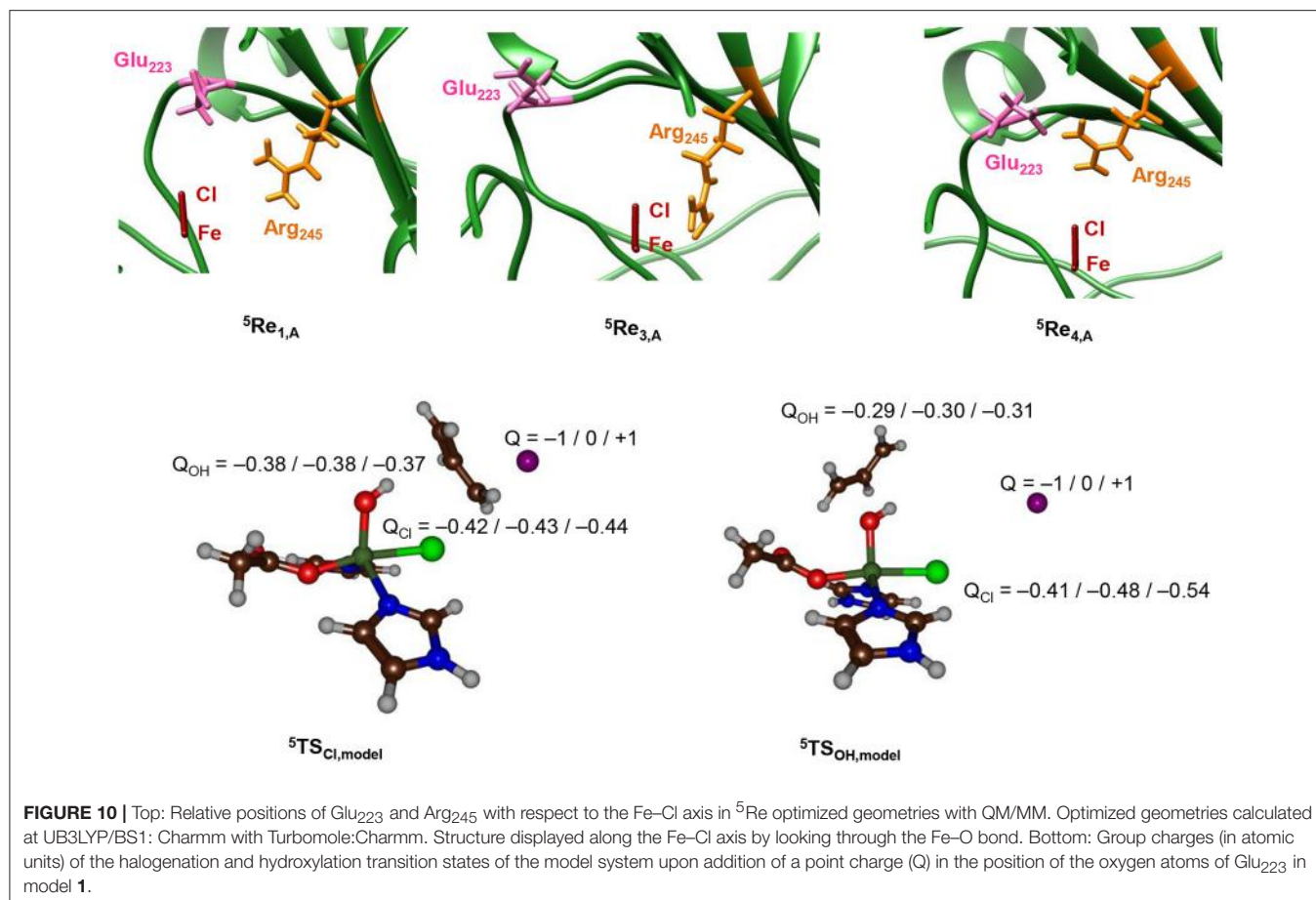toms resulting in chemoselectivity changes. Thus, in model **1** the carboxylate group of Glu$_{223}$ is aligned to the axis that goes through the Fe–Cl bond and appears to be set up to push electron density away from Cl and onto iron. By contrast, in **4** the Glu$_{223}$ and Arg$_{245}$ side chains have formed a salt bridge and now the positively charged Arg is on the axis of the Fe–Cl bond. Consequently, in model **4** charge density will be withdrawn from iron and move to chloride instead making it more difficult to transfer the Cl to substrate. In model **3** the Glu$_{223}$ is moved to the outside part of the protein and hence is located at a relatively large distance from the active center. In addition, the Arg$_{245}$ is positioned much closer to the halide but on the side rather than along the axis. Indeed, model **3** shows a regioselective hydroxylation and no halogenation, whereas in model **1** a preference for halogenation over hydroxylation is seen.

To find out whether the Glu$_{223}$ amino acid residue is near enough to the iron(IV)-oxo(chloride) to affect the charge distributions of the atoms, we took the DFT models of $^5TS_{OH}$ and $^5TS_{Cl}$ and included a point charge to the system in the position of the carboxylate oxygen atom of Glu$_{223}$ from model **1**. Single point calculations with a point charge of Q = +1 and −1 were performed and the results are given at the bottom of **Figure 10**. Surprisingly, only small electronic changes are observed to $^5TS_{Cl}$ upon addition of a point charge. By contrast, the hydroxyl rebound barrier sees major changes in the charge of the halogen atom under the influence of a point charge. Specifically, the

halogen charge increases from −0.48 to −0.41 with a point charge with magnitude −1 included. In addition, the barrier $^5TS_{Cl}$ drops below $^5TS_{OH}$ by 4.6 kcal mol^{-1}, whereas the energy difference is −20.0 kcal mol^{-1} with a point charge of magnitude +1 included. Therefore, a point charge located at a distance of over 5 Å from the reaction center can incur an electrostatic perturbation that affects the charge distributions in the catalytic reaction center and consequently the chemoselectivity of the reaction. This is an important caveat, given that contrary to the active site and core region of the halogenase domain, the structure of HctB is less well defined by the model, with the positioning of the two other domains and interactions of the halogenase domain not clear/not well defined.

To finally test the effect of the salt bridge between Arg$_{245}$ and Glu$_{223}$ we performed a series of electrostatic potential calculations on model **4** whereby in one case the salt bridge was rearranged so that either Arg$_{245}$ or Glu$_{223}$ is closer to the chloride. Calculations of electrostatic potential fields were made with the Protein-sol server software package (Hebditch et al., 2017) and the results are given in **Figure 11**, displayed using PyMOL (Schrödinger, LLC). Thus, the salt-bridge between Arg$_{245}$ and Glu$_{223}$ in model **4** (lower panel of **Figure 11**) induces a change in negative charge on the halide that is pushed toward the iron. Swapping the salt-bridge and changing the position of the Arg and Glu residues has a dramatic effect on the charge donation to the active site and now electron density is withdrawn from iron and moved toward the chloride group. Therefore, the salt-bridge is essential for creating the right charge distribution between the iron and chloride group and induces an electric field effect on the active site that enables efficient substrate halogenation as predicted by the DFT studies reported above.

FIGURE 10 | Top: Relative positions of Glu$_{223}$ and Arg$_{245}$ with respect to the Fe–Cl axis in ^{5}Re optimized geometries with QM/MM. Optimized geometries calculated at UB3LYP/BS1: Charmm with Turbomole:Charmm. Structure displayed along the Fe–Cl axis by looking through the Fe–O bond. Bottom: Group charges (in atomic units) of the halogenation and hydroxylation transition states of the model system upon addition of a point charge (Q) in the position of the oxygen atoms of Glu$_{223}$ in model **1**.

FIGURE 11 | Induced electric field of the salt-bridge between Arg$_{245}$ and Glu$_{223}$ on the iron(IV)-oxo(chloride) active site. Electrostatic potential was calculated for model ^{5}Re$_4$ **(A)** and model ^{5}Re$_{4,mod}$ **(B)**, which is model **4** with rotated Arg$_{245}$ and Glu$_{223}$ residues with respect to the Cl–Fe axis. Color coding for the amino stick representations changes from red (negative potential) to blue (positive potential). The direction of the induced electric field is given by the arrow.

In summary, the QM/MM and DFT calculations presented here highlight that for effective substrate halogenation, radical character on the halogen atom is needed that can pair up with the substrate radical. Perturbations from the protein, i.e. from nearby located anionic amino acid residues, e.g., Glu$_{223}$, can lead to a push-effect of charge density from the halogen to the metal to achieve this. At the same time, the active site Arg$_{245}$ should not be aligned with the Fe–Cl bond. Our observations are in excellent agreement with recent computational studies on SyrB2 using model complexes (Wong et al., 2013; Srnec et al., 2016; Srnec

and Solomon, 2017). In particular, it was proposed that ideal orbital overlap between the substrate radical and halogen atom would lower the halogen transfer transition state below the OH rebound barrier and thereby overcome the large hydroxylation driving force.

CONCLUSIONS

In this work we describe a detailed MM, MD, QM/MM, and DFT study on the nonheme iron halogenase HctB for the first time.

Our initial analysis of a structure from the literature revealed several substrate entrance channels. We, therefore, created four models with the substrate in different positions and ran QM/MM calculations on snapshots from these MD simulations. We confirm a reaction mechanism starting from an iron(IV)-oxo species that reacts via hydrogen atom abstraction followed by halogen rebound. Our calculated hydrogen atom abstraction barriers vary dramatically with substrate binding position and a low barrier is found for model **3**, while much higher barriers are seen for models **1** and **2**. Interestingly, two models give preferential halogenation, one model preferential hydroxylation and the fourth one is expected to give a mixture of products. The two models that give preferential halogenation in our analyses use distinct substrate access channels. We analyzed the individual structures and identify an active site Glu residue (Glu$_{223}$), which is unique for HctB as other halogenases typically have a Ser or Lys in that position. Finally, we compared the QM/MM calculations with small gas-phase model complexes and find that even though the gas-phase structure would imply preferential hydroxylation over halogenation, a complete chemoselectivity reversal can be achieved through external perturbations. Calculations on model systems using either an applied electric field along the positive x-axis or with a point charge with magnitude Q = −1 give dominant halogenation, whereas in all other cases, hydroxylation is predicted. It follows, therefore, that the HctB enzyme structure is designed in such a way, so as to destabilize the hydroxylation pathway and give favorable halogenation products. Overall, our calculations show that substrate binding position is essential for an optimal halogenation reaction. This may come at a cost through higher hydrogen atom abstraction barriers, but the enlarged hydrogen atom abstraction barriers may hamper optimal hydroxo rebound and the formation of alcohol products.

METHODS

QM/MM calculations were performed on the mechanism of substrate hydroxylation vs. halogenation in the HctB halogenase using methods and procedures utilized previously on analogous enzymatic systems (Aluri and de Visser, 2007; Kumar et al., 2011; Hernández-Ortega et al., 2015; Quesne et al., 2016a; Tchesnokov et al., 2016; Faponle et al., 2017; Hofer and de Visser, 2018).

Model Building

The work started from the halogenase domain reported by Pratter et al. (2014a,b) which contains an *in silico* docked hexanoyl phosphopentatheinyl (PPT) moiety (channel II, **Figure 2**). Upon our analysis of the structure; however, another substrate entrance channel was discovered similar to that identified in SyrB2 (channel I, **Figure 2**) (Blasiak et al., 2006). We manually docked the substrate into each substrate entrance channel and replaced the iron(II)-water group by iron(IV)-oxo and shortened the Fe–O distance to 1.63 Å. In addition, α-ketoglutarate was replaced by succinate (Succ) and the acyl protein carrier was removed. The substrate in channel I was positioned in three different orientations relative to the iron(IV)-oxo. The substrates were positioned in such a way that stereochemical clashes with protein residues along the

substrate entrance channels were minimized and after insertion of the substrate the system was solvated. During solvation and molecular dynamics simulations no additional water molecules entered the active site pocket. The structures were charge neutralized with 18 chloride and 13 magnesium ions on the surface of the protein. For example, model **1** contained a total of 32,458 atoms including 9,157 water molecules.

QM/MM Set-Up

Hydrogen atoms were added to models **1**–**4** using the PDB2PQR software package at pH = 7, (Dolinsky et al., 2007) whereby all Glu/Asp side chains were in their basic form and all Arg/Lys residues in the acidic forms. The protonation states of histidine residues was decided upon visual inspection of their local environment and His$_{67}$, His$_{111}$, His$_{157}$, His$_{185}$, His$_{227}$, and His$_{256}$ were singly protonated at the N$^{\delta}$ position, whereas His$_{64}$, His$_{163}$, His$_{268}$, and His$_{273}$ were protonated at the N$^{\epsilon}$ position. Histidine residue His$_{8}$ was chosen as being doubly protonated at both the N$^{\delta}$- and N$^{\epsilon}$-positions.

A constraint geometry optimization with all protein backbone atoms fixed was performed in Charmm, (Brooks et al., 1983) and subsequently the system was solvated in a sphere with a radius of 40 Å. An iterative solvation procedure (**Supporting Information Figure S1**) with fixed protein backbone was followed by an equilibration and heating procedure to 298 K. For all chemical systems described here, a full molecular dynamics simulation with the Charmm forcefield was performed without constraints for a minimum of 10 ns (**Supporting Information Figure S2**). All systems relaxed to stable structures and a random low-energy snapshot was chosen as the starting geometry for the QM/MM calculations.

QM Region

We selected a QM region containing the iron(IV)-oxo(chloro) group and included the imidazole groups of His$_{111}$ and His$_{227}$, the acetate terminus of succinate (Succ) and the thiohexanoic acid arm of the substrate as our minimal QM region **A**. In addition, a larger QM region was tested that included the full or partial amino acid R-groups within 6 Å of the iron(IV)-oxo(chloro) structure: QM region **AB**. For model **1**, the large QM region **AB** contained the amino acid side chains of Ile$_{108}$, Val$_{113}$, Asp$_{200}$, Asp$_{202}$, Glu$_{223}$, Val$_{225}$, Met$_{226}$, Arg$_{245}$, and six water molecules. QM region **AB** for Model **2** is described with the same amino acids except Asp$_{202}$, Val$_{225}$, and Met$_{245}$ and in addition also includes Ser$_{141}$ and Phe$_{221}$ and four water molecules. Model **3** in its most elaborate form covers Ile$_{108}$, Val$_{113}$, Ser$_{141}$, Asp$_{200}$, Phe$_{221}$, Asn$_{243}$, Arg$_{245}$, and four water molecules. Finally, the large QM region for Model **4** includes Ile$_{108}$, Val$_{113}$, Phe$_{221}$, Arg$_{245}$, and five water molecules.

QM/MM Calculations

Subsequent QM/MM calculations of the different substrate bound iron(IV)-oxo(chloro) enzyme intermediates were performed in Turbomole:Charmm and linked via ChemShell (Ahlrichs et al., 1989; Smith and Forester, 1996; Sherwood et al., 2003). Density functional theory methods at the unrestricted B3LYP level of theory (Lee et al., 1988; Becke, 1993) were

applied to the QM region in Turbomole with an SV(P) basis set on all atoms: basis set BS1 (Schafer et al., 1992). Single point calculations on the optimized structures in QM/MM were performed with an all-electron Wachters-type basis set on iron and def2-TZV(P) on the rest of the atoms: basis set BS2 (Wachters, 1970; Hay, 1977; Krishnan et al., 1980; Bauschlicher et al., 1989). For several systems we tested a full geometry optimization of using basis set BS2 but found similar structures and relative energies to those obtained with BS1; hence BS1 was used for all systems. The MM region was described with the Charmm forcefield (Brooks et al., 1983). A detailed benchmark study comparing computational free energies of activation with experimental data from Eyring plots gave good agreement between experiment and theory for the methods described here (Cantú Reinhard et al., 2016a). Furthermore, these methods were also seen to excellently reproduce reduction potentials of copper proteins, (Fowler et al., 2017) and regio- and chemoselectivities of reaction mechanisms (Jastrzebski et al., 2014; Barman et al., 2016b; Yang et al., 2016).

Geometries were optimized at UB3LYP/BS1 in Turbomole:Charmm and a frequency calculation on the QM region only was done to test whether it was a local minimum or first order saddle point. Geometry scans with one degree of freedom fixed were performed along a specified reaction coordinate. The maximum of these scans were subjected to a full geometry optimization of a first-order saddle point and the starting and end points of the scans established the proposed reaction paths.

AUTHOR CONTRIBUTIONS

AT and SdV devised the project. AT ran the QM/MM calculations. NF and JW performed the electrostatic calculations. AT, NF, GS, and SdV wrote the paper.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2018.00513/full#supplementary-material

REFERENCES

Abu-Omar, M. M., Loaiza, A., and Hontzeas, N. (2005). Reaction mechanisms of mononuclear non-heme iron oxygenases. *Chem. Rev.* 105, 2227–2252. doi: 10.1021/cr040653o

Agarwal, V., Miles, Z. D., Winter, J. M., Eustáquio, A. S., El Gamal, A. A., and Moore, B. S. (2017). Enzymatic halogenation and dehalogenation reactions: pervasive and mechanistically diverse. *Chem. Rev.* 117, 5619–5674. doi: 10.1021/acs.chemrev.6b00571

Ahlrichs, R., Bär, M., Häser, M., Horn, H., and Kölmel, C. (1989). Electronic structure calculations on workstation computers: the program system turbomole. *Chem. Phys. Lett.* 162, 165–169. doi: 10.1016/0009-2614(89)85118-8

Aluri, S., and de Visser, S. P. (2007). The mechanism of cysteine oxygenation by cysteine dioxygenase enzymes. *J. Am. Chem. Soc.* 129, 14846–14847. doi: 10.1021/ja0758178

Álvarez-Barcia, S., and Kästner, J. (2017). Atom tunneling in the hydroxylation process of taurine/α-ketoglutarate dioxygenase identified by quantum mechanics/molecular mechanics simulations. *J. Phys. Chem. B.* 121, 5347–5354. doi: 10.1021/acs.jpcb.7b03477

Ansari, A., Kaushik, A., and Rajaraman, G. (2013). Mechanistic insights on the ortho-hydroxylation of aromatic compounds by non-heme iron complex: a computational case study on the comparative oxidative ability of ferric-hydroperoxo and high-valent FeIV-O and FeV-O intermediates. *J. Am. Chem. Soc.* 135, 4235–4249. doi: 10.1021/ja307077f

Barman, P., Faponle, A. S., Vardhaman, A. K., Angelone, D., A.-,Löhr, M., Browne, W. R., et al. (2016b). Influence of ligand architecture in tuning reaction bifurcation pathways for chlorite oxidation by non-heme iron complexes. *Inorg. Chem.* 55, 10170–10181. doi: 10.1021/acs.inorgchem.6b01384

Barman, P., Upadhyay, P., Faponle, A. S., Kumar, J., Nag, S. S., Kumar, D., et al. (2016a). Deformylation reaction by a nonheme manganese(III)–peroxo complex via initial hydrogen-atom abstraction. *Angew. Chem. Int. Ed.* 55, 11091–11095. doi: 10.1002/anie.201604412

Bauschlicher C. W. Jr., Langhoff, S. R., Partridge, H., and Barnes, L. A. (1989). Theoretical studies of the first- and second-row transition-metal methyls and their positive ions. *J. Chem. Phys.* 91, 2399–2411. doi: 10.1063/1.456998

Becke, A. D. (1993). Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* 98, 5648–5652. doi: 10.1063/1.464913

Bernasconi, L., and Baerends, E. J. (2008). The EDTA complex of oxidoiron(IV) as realisation of an optimal ligand environment for high activity of FeO$_{2+}$. *Eur. J. Inorg. Chem.* 2008, 1672–1681. doi: 10.1002/ejic.200701135

Bernasconi, L., and Baerends, J. (2013). A frontier orbital study with ab initio molecular dynamics of the effects of solvation on chemical reactivity: solvent-induced orbital control in FeO-activated hydroxylation reactions. *J. Am. Chem. Soc.* 135, 8857–8867. doi: 10.1021/ja311144d

Blasiak, L. C., Vaillancourt, F. H., Walsh, C. T., and Drennan, C. L. (2006). Crystal structure of the non-haem iron halogenase SyrB2 in syringomycin biosynthesis. *Nature* 440, 368–371. doi: 10.1038/nature04544

Bollinger J., M. Jr., Price, J., C., Hoffart, L. M., Barr, E. W., and Krebs, C. (2005). Mechanism of taurine: α-ketoglutarate dioxygenase (TauD) from *Escherichia coli. Eur. J. Inorg. Chem.* 2005, 4245–4254. doi: 10.1002/ejic.200500476

Borowski, T., Bassan, A., and Siegbahn, P. E. M. (2004). Mechanism of dioxygen activation in 2-oxoglutarate-dependent enzymes: a hybrid DFT study. *Chem. Eur. J.* 10, 1031–1041. doi: 10.1002/chem.200305306

Borowski, T., Noack, H., Radon, M., Zych, K., and Siegbahn, P. E. M. (2010). Mechanism of selective halogenation by SyrB2: a computational study. *J. Am. Chem. Soc.* 132, 12887–12898. doi: 10.1021/ja101877a

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983). CHARMM: a program for macromolecular energy, minimization, and dynamics calculation. *J. Comput. Chem.* 4, 187–217. doi: 10.1002/jcc.540040211

Bruijnincx, P. C. A., van Koten, G., and Klein Gebbink, R. J. M. (2008). Mononuclear non-heme iron enzymes with the 2-His-1-carboxylate facial triad: recent developments in enzymology and modeling studies. *Chem. Soc. Rev.* 37, 2716–2744. doi: 10.1039/b707179p

Bugg, T. D. H. (2001). Oxygenases: mechanisms and structural motifs for O(2) activation. *Curr. Opin. Chem. Biol.* 5, 550–555. doi: 10.1016/S1367-5931(00)00236-2

Buongiorno, D., and Straganz, G. D. (2013). Structure and function of atypically coordinated enzymatic mononuclear non-heme-Fe(II) centers. *Coord. Chem. Rev.* 257, 541–563. doi: 10.1016/j.ccr.2012.04.028

Bushnell, E. A. C., Fortowsky, G. B., and Gauld, J. W. (2012). Model iron–oxo species and the oxidation of imidazole: insights into the mechanism of OvoA and EgtB? *Inorg. Chem.* 51, 13351–13356. doi: 10.1021/ic3021172

Butler, A., and Sandy, M. (2009). Mechanistic considerations of halogenating enzymes. *Nature* 460, 848–854. doi: 10.1038/nature08303

Cantú Reinhard, F. G., Barman, P., Mukherjee, G., Kumar, J., Kumar, D., Kumar, D., et al. (2017b). Keto–enol tautomerization triggers an electrophilic aldehyde deformylation reaction by a nonheme manganese(III)-peroxo complex. *J. Am. Chem. Soc.* 139, 18328–18338. doi: 10.1021/jacs.7b10033

Cantú Reinhard, F. G., and de Visser, S. P. (2017a). Oxygen atom transfer using an iron(IV)-oxo embedded in a tetracyclic N-heterocyclic carbene system: how does the reactivity compare to cytochrome P450 compound I? *Chem. Eur. J.* 23, 2935–2944. doi: 10.1002/chem.201605505

Cantú Reinhard, F. G., Faponle, A. S., and de Visser, S. P. (2016a). Substrate Sulfoxidation by an Iron(IV)-Oxo complex: benchmarking computationally calculated barrier heights to experiment. *J. Phys. Chem. A.* 120, 9805–9814. doi: 10.1021/acs.jpca.6b09765

Cantú Reinhard, F. G., Sainna, M. A., Upadhyay, P., Balan, G. A., Kumar, D., Fornarini, S., et al. (2016b). A systematic account on aromatic hydroxylation by a cytochrome P450 model compound i: a low-pressure mass spectrometry and computational study. *Chem. Eur. J.* 22, 18608–18619. doi: 10.1002/chem.201604361

Chatterjee, S., and Paine, T. K. (2016). Hydroxylation versus halogenation of aliphatic C-H bonds by a dioxygen-derived iron-oxygen oxidant: functional mimicking of iron halogenases. *Angew. Chem. Int. Ed.* 55, 7717–7722. doi: 10.1002/anie.201509914

Chen, H., Lai, W., Yao, J., and Shaik, S. (2011). Perferryl FeV–Oxo nonheme complexes: do they have high-spin or low-spin ground states? *J. Chem. Theory Comput.* 7, 3049–3053. doi: 10.1021/ct200614g

Chen, X., and van Pée, H. (2008). Catalytic mechanisms, basic roles, and biotechnological and environmental significance of halogenating enzymes. *Acta. Biochim. Biophys. Sin.* 40, 183–193. doi: 10.1111/j.1745-7270.2008.00390.x

Cicero, G., Carbonera, C., Valegård, K., Hajdu, J., Andersson, I., and Ranghino, G. (2007). Study of the oxidative half-reaction catalyzed by a non-heme ferrous catalytic center by means of structural and computational methodologies. *Int. J. Quant. Chem.* 107, 1514–1522. doi: 10.1002/qua.21275

Comba, P., and Wunderlich, S. (2010). Iron-catalyzed halogenation of alkanes: modeling of nonheme halogenases by experiment and DFT calculations. *Chem. Eur. J.* 16, 7293–7299. doi: 10.1002/chem.201000092

Costas, M., Mehn, M. P., Jensen, M. P., and Que Jr, L. (2004). Dioxygen activation at mononuclear nonheme iron active sites: enzymes, models, and intermediates. *Chem. Rev.* 104, 939–986. doi: 10.1021/cr020628n

de Visser, S. P. (2005). What Affects the Quartet–Doublet Energy Splitting in Peroxidase Enzymes? *J. Phys. Chem. A.* 109, 11050–11057. doi: 10.1021/jp053873u

de Visser, S. P. (2006a). Differences in and comparison of the catalytic properties of heme and non-heme enzymes with a central Oxo–iron group. *Angew. Chem. Int. Ed.* 45, 1790–1793. doi: 10.1002/anie.200503841

de Visser, S. P. (2006b). What factors influence the ratio of CH hydroxylation versus CC epoxidation by a nonheme cytochrome P450 biomimetic? *J. Am. Chem. Soc.* 128, 15809–15818. doi: 10.1021/ja065365j

de Visser, S. P. (2006c). Propene activation by the Oxo-iron active species of taurine/α-ketoglutarate dioxygenase (TauD) enzyme. How Does the Catalysis Compare to Heme-Enzymes? J. Am. Chem. Soc. 128, 9813–9824. doi: 10.1021/ja061581g

de Visser, S. P. (2006d). Substitution of Hydrogen by deuterium changes the regioselectivity of ethylbenzene hydroxylation by an oxo–iron–porphyrin catalyst. *Chem. Eur. J.* 12, 8168–8177. doi: 10.1002/chem.200600376

de Visser, S. P. (2010). Trends in substrate hydroxylation reactions by Heme and nonheme iron(IV)-oxo oxidants give correlations between intrinsic properties of the oxidant with barrier height. *J. Am. Chem. Soc.* 132, 1087–1097. doi: 10.1021/ja908340j

de Visser, S. P., and Latifi, R. (2009). Carbon dioxide: a waste Product in the catalytic cycle of α-ketoglutarate dependent halogenases prevents the formation of hydroxylated by-products. *J. Phys. Chem. B.* 113, 12–14. doi: 10.1021/jp8097632

de Visser, S. P., Quesne, M. G., Martin, B., Comba, P., and Ryde, U. (2014). Computational modelling of oxygenation processes in enzymes and biomimetic model complexes. *Chem. Commun.* 50, 262–282. doi: 10.1039/C3CC47148A

Denisov, I. G., Makris, T. M., Sligar, S. G., and Schlichting, I. (2005). Structure and Chemistry of Cytochrome P450. *Chem. Rev.* 105, 2253–2277. doi: 10.1021/cr0307143

Dey, A. (2010). Density functional theory calculations on the mononuclear non-heme iron active site of hmd hydrogenase: role of the internal ligands in tuning external ligand binding and driving H2 heterolysis. *J. Am. Chem. Soc.* 132, 13892–13901. doi: 10.1021/ja1041918

Dolinsky, T. J., Czodrowski, P., Li, H., Nielsen, J. E., Jensen, J. H., Klebe, G., et al. (2007). PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucl. Acids Res.* 35, 522–525. doi: 10.1093/nar/gkm276

Faponle, A. S., Quesne, M. G., and de Visser, S. P. (2016). Origin of the regioselective fatty-acid hydroxylation versus decarboxylation by a cytochrome p450 peroxygenase: what drives the reaction to biofuel production? *Chem. Eur. J.* 22, 5478–5483. doi: 10.1002/chem.201600739

Faponle, A. S., Seebeck, F. P., and de Visser, S. P. (2017). Sulfoxide synthase versus cysteine dioxygenase reactivity in a nonheme iron enzyme. *J. Am. Chem. Soc.* 139, 9259–9270. doi: 10.1021/jacs.7b04251

Fowler, N. J., Blanford, C. F., Warwicker, J., and de Visser, S. P. (2017). Prediction of reduction potentials of copper proteins with continuum electrostatics and density functional theory. *Chem. Eur. J.* 23, 15436–15445. doi: 10.1002/chem.201702901

Galonić Fujimori, D., Barr, E. W., Matthews, M. L., Koch, G. M., Yonce, J. R., Walsh, C. T., et al. (2007). Spectroscopic evidence for a high-spin Br-Fe(IV)-Oxo intermediate in the α-ketoglutarate-dependent halogenase CytC3 from Streptomyces. *J. Am. Chem. Soc.* 129, 13408–13409. doi: 10.1021/ja076454e

Godfrey, E., Porro, C. S., and de Visser, S. P. (2008). Comparative quantum mechanics/molecular mechanics (QM/MM) and density functional theory calculations on the Oxo–iron species of taurine/α-ketoglutarate dioxygenase. *J. Phys. Chem. A.* 112, 2464–2468. doi: 10.1021/jp710999v

Green, M. T., Dawson, J. H., and Gray, H. B. (2004). Oxoiron(IV) in chloroperoxidase compound II is basic: implications for P450 chemistry. *Science* 304, 1653–1656. doi: 10.1126/science.1096897

Gribble, G. W. (2003). The diversity of naturally produced organohalogens. *Chemosphere* 52, 289–297. doi: 10.1016/S0045-6535(03)00207-8

Hay, P. J. (1977). Gaussian basis sets for molecular calculations. The representation of 3d orbitals in transition-metal atoms. *J. Chem. Phys.* 66, 4377–4384. doi: 10.1063/1.433731

Hebditch, M., Carballo-Amador, M. A., Charonis, S., Curtis, R., and Warwicker, J. (2017). Protein–Sol: a web tool for predicting protein solubility from sequence. *Bioinformatics* 33, 3098–3100. doi: 10.1093/bioinformatics/btx345

Hernández-Ortega, A., Quesne, M. G., Bui, S., Heyes, D. J., Steiner, R. A., Scrutton, N. S., et al. (2015). Catalytic mechanism of cofactor-free dioxygenases and how they circumvent spin-forbidden oxygenation of their substrates. *J. Am. Chem. Soc.* 137, 7474–7487. doi: 10.1021/jacs.5b03836

Hirao, H., Kumar, D., Que Jr, L., and Shaik, S. (2006). Two-state reactivity in alkane hydroxylation by non-heme iron–oxo complexes. *J. Am. Chem. Soc.* 128, 8590–8606. doi: 10.1021/ja061609o

Hirao, H., Li, F., Que Jr, L., and Morokuma, K. (2011). Theoretical study of the mechanism of oxoiron(IV) formation from H2O2 and a nonheme iron(II) complex: O–O cleavage involving proton-coupled electron transfer. *Inorg. Chem.* 50, 6637–6648. doi: 10.1021/ic200522r

Hofer, T. S., and de Visser, S. P. (2018). Editorial: quantum mechanical/molecular mechanical approaches for the investigation of chemical systems – recent developments and advanced applications. *Front. Chem.* 6:357. doi: 10.3389/fchem.2018.00357

Huang, J., Li, C., Wang, B., Sharon, D. A., Wu, W., and Shaik, S. (2016). Selective chlorination of substrates by the halogenase SyrB2 is controlled by the protein according to a combined quantum mechanics/molecular mechanics and molecular dynamics study. *ACS Catal.* 6, 2694–2704. doi: 10.1021/acscatal.5b02825

Huang, X., and Groves, J. T. (2017). Beyond ferryl-mediated hydroxylation: 40 years of the rebound mechanism and C–H activation. *J. Biol. Inorg. Chem.* 22, 185–207. doi: 10.1007/s00775-016-1414-3

Jastrzebski, R., Quesne, M. G., Weckhuysen, B. M., de Visser, S. P., and Bruijnincx, P. C. A. (2014). Experimental and computational evidence for the mechanism of intradiol catechol dioxygenation by non-heme iron(III) complexes. *Chem. Eur. J.* 20, 15686–15691. doi: 10.1002/chem.201404988

Ji, L., Faponle, A. S., Quesne, M. G., Sainna, M. A., Zhang, J., Franke, A., et al. (2015). Drug metabolism by cytochrome P450 enzymes: what distinguishes the pathways leading to substrate hydroxylation over desaturation? *Chem. Eur. J.* 21, 9083–9092. doi: 10.1002/chem.201500329

Kaczmarek, M. A., Malhotra, A., Balan, G. A., Timmins, A., and de Visser, S. P. (2018). Nitrogen reduction to ammonia on a biomimetic mononuclear iron centre: insights into the nitrogenase enzyme. *Chem. Eur. J.* 24, 5293–5302. doi: 10.1002/chem.201704688

Karamzadeh, B., Kumar, D., Sastry, G. N., and de Visser, S. P. (2010). Steric factors override thermodynamic driving force in regioselectivity of proline Hydroxylation by Prolyl-4-hydroxylase enzymes. *J. Phys. Chem. A.* 114, 13234–13243. doi: 10.1021/jp1089855

Kim, S. H., Perera, R., Hager, L. P., Dawson, J. H., and Hoffman, B. M. (2006). Rapid freeze-quench ENDOR study of chloroperoxidase compound I: the site of the radical. *J. Am. Chem. Soc.* 128, 5598–5599. doi: 10.1021/ja060776l

Krishnan, R., Binkley, J. S., Seeger, R., and Pople, J. A. (1980). Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *J. Chem. Phys.* 72, 650–654. doi: 10.1063/1.438955

Kulik, H. J., Blasiak, L. C., Marzari, N., and Drennan, C. L. (2009). First-principles study of non-heme Fe(II) halogenase SyrB2 reactivity. *J. Am. Chem. Soc.* 131, 14426–14433. doi: 10.1021/ja905206k

Kumar, D., de Visser, S. P., and Shaik, S. (2003). How does product isotope effect prove the operation of a two-state "rebound" mechanism in C–H hydroxylation by cytochrome P450? *J. Am. Chem. Soc.* 125, 13024–13025. doi: 10.1021/ja036906x

Kumar, D., de Visser, S. P., Sharma, P. K., Cohen, S., and Shaik, S. (2004). Radical clock substrates, their C–H hydroxylation mechanism by cytochrome P450, and other reactivity patterns: what does theory reveal about the clocks' behavior? *J. Am. Chem. Soc.* 126, 1907–1920. doi: 10.1021/ja039439s

Kumar, D., Thiel, W., and de Visser, S. P. (2011). Theoretical study on the mechanism of the oxygen activation process in cysteine dioxygenase enzymes. *J. Am. Chem. Soc.* 133, 3869–3882. doi: 10.1021/ja107514f

Kumar, S., Faponle, A. S., Barman, P., Vardhaman, A. K., Sastri, C. V., Kumar, D., et al. (2014). Long-range electron transfer triggers mechanistic differences between Iron(IV)-Oxo and Iron(IV)-imido oxidants. *J. Am. Chem. Soc.* 136, 17102–17115. doi: 10.1021/ja508403w

Latifi, R., Bagherzadeh, M., and de Visser, S. P. (2009). Origin of the correlation of the rate constant of substrate hydroxylation by nonheme iron(IV)–oxo complexes with the bond-dissociation energy of the C-H bond of the substrate. *Chem. Eur. J.* 15, 6651–6662. doi: 10.1002/chem.200900211

Latifi, R., Sainna, M. A., Rybak-Akimova, E. V., and de Visser, S. P. (2013). Does hydrogen-bonding donation to manganese(IV)–Oxo and iron(IV)–Oxo oxidants affect the oxygen-atom transfer ability? A computational study. *chem. Eur. J.* 19, 4058–4068. doi: 10.1002/chem.201202811

Lee, C., Yang, W., and Parr, R. G. (1988). Density-functional thermochemistry. III. The role of exact exchange. *Phys. Rev. B* 37, 785–789. doi: 10.1103/PhysRevB.37.785

Liu, W., and Groves, J. T. (2010). Manganese porphyrins catalyze selective C–H bond halogenations. *J. Am. Chem. Soc.* 132, 12847–12849. doi: 10.1021/ja105548x

Liu, Y., Guan, X. E. L.,Wong, M., Liu, P. J., Huang, S. C., and Che, M. (2013). Nonheme Iron-mediated amination of C(sp3)–H bonds. Quinquepyridine-supported iron-imide/nitrene intermediates by experimental studies and DFT calculations. *J. Am. Chem. Soc.* 135, 7194–7204. doi: 10.1021/ja3122526

Mai, B. K., and Kim, Y. (2016). Is it Fe(III)-oxyl radical that abstracts hydrogen in the C–H activation of TauD? A theoretical study based on the DFT potential energy surfaces. *Inorg. Chem.* 55, 3844–3852. doi: 10.1021/acs.inorgchem.5b02939

Martinez, J. S., Carroll, G. L., Tschirret-Guth, R. A., Altenhoff, G., Little, D. R., and Butler, A. (2001). On the Regiospecificity of Vanadium Bromoperoxidase. *J. Am. Chem. Soc.* 123, 3289–3294. doi: 10.1021/ja004176c

Matthews, M. L., Krest, C. M., Barr, E. W., Vaillancourt, F. H., Walsh, C. T., Green, M. T., et al. (2009). Substrate-triggered formation and remarkable stability of the C–H bond-cleaving chloroferryl intermediate in the aliphatic halogenase, SyrB2. *Biochemistry* 48, 4331–4343. doi: 10.1021/bi900109z

Matthews, M. L., Neumann, C. S., Miles, L. A., Grove, T. L., Booker, S. J., Krebs, C., et al. (2006). Substrate positioning controls the partition between halogenation and hydroxylation in the aliphatic halogenase, SyrB2. *Proc. Natl. Acad. Sci. USA.* 106, 17723–17728. doi: 10.1073/pnas.0909649106

Messerschmidt, A., and Wever, R. (1996). X-ray structure of a vanadium-containing enzyme: chloroperoxidase from the fungus Curvularia inaequalis. *Proc. Natl. Acad. Sci. U.S.A.* 93, 392–396. doi: 10.1073/pnas.93.1.392

Meunier, B., de Visser, S. P., and Shaik, S. (2004). Mechanism of oxidation reactions catalyzed by cytochrome P450 enzymes. *Chem. Rev.* 104, 3947–3980. doi: 10.1021/cr020443g

Mitchell, A. J., Dunham, N. P., Bergman, J. A., Wang, B., Zhu, Q., Chen, W., et al. (2017). Structure-guided reprogramming of a hydroxylase to halogenate its small molecule substrate. *Biochemistry* 56, 441–444. doi: 10.1021/acs.biochem.6b01173

Neidig, M. L., Brown, C. D., Light, K. M., Galonić Fujimori, D., Nolan, E. M., Price, J. C., et al. (2007). CD and MCD of CytC3 and taurine dioxygenase: role of the facial triad in α-KG-dependent oxygenases. *J. Am. Chem. Soc.* 129, 14224–14231. doi: 10.1021/ja074557r

Nemukhin, A. V., Topol, I. A., Cachau, R. E., and Burt, S. K. (2006). On the nature of oxoiron (IV) intermediate in dioxygen activation by non-heme enzymes. *Theor. Chem. Acc.* 115, 348–353. doi: 10.1007/s00214-005-0047-3

Noack, H., and Siegbahn, P. E. M. (2007). Theoretical investigation on the oxidative chlorination performed by a biomimetic non-heme iron catalyst. *J. Biol. Inorg. Chem.* 12, 1151–1162. doi: 10.1007/s00775-007-0284-0

Ogliaro, F., Harris, N., Cohen, S., Filatov, M., de Visser, S. P., and Shaik, S. (2000). A model "rebound" mechanism of hydroxylation by cytochrome P450: stepwise and effectively concerted pathways, and their reactivity patterns. *J. Am. Chem. Soc.* 122, 8977–8989. doi: 10.1021/ja991878x

Pandian, S., Vincent, M. A., Hillier, I. H., and Burton, N. A. (2009). Why does the enzyme SyrB2 chlorinate, but does not hydroxylate, saturated hydrocarbons? A density functional theory (DFT) study. *Dalton Trans.* 2009, 6201–6207. doi: 10.1039/b906866j

Podgoršek, A., Zupan, M., and Iskra, J. (2009). Oxidative halogenation with "green" oxidants: oxygen and hydrogen peroxide. *Angew. Chem. Int. Ed.* 48, 8424–8450. doi: 10.1002/anie.200901223

Pratter, S. M., Ivkovic, J., Birner-Gruenberger, R., Breinbauer, R., Zangger, K., and Straganz, G. D. (2014a). More than just a halogenase: modification of fatty acyl moieties by a trifunctional metal enzyme. *ChemBioChem.* 15, 567–574. doi: 10.1002/cbic.201300345

Pratter, S. M., Konstantinovics, C., DiGiuro, C., L. M., Leitner, E., Kumar, D., de Visser, S. P., et al. (2013). Inversion of enantioselectivity of a mononuclear non-heme iron(II)-dependent hydroxylase by tuning the interplay of metal-center geometry and protein structure. *Angew. Chem. Int. Ed.* 52, 9677–9681. doi: 10.1002/anie.201304633

Pratter, S. M., Light, K. M., Solomon, E. I., and Straganz, G. D. (2014b). The role of chloride in the mechanism of O$_2$ activation at the mononuclear nonheme Fe(II) center of the halogenase HctB. *J. Am. Chem. Soc.* 136, 9385–9395. doi: 10.1021/ja503179m

Puri, M., Biswas, A. N., Fan, R., Guo, Y., and Que Jr, L. (2016). Modeling non-heme iron halogenases: high-spin oxoiron(IV)–halide complexes that halogenate C–H Bonds. *J. Am. Chem. Soc.* 138, 2484–2487. doi: 10.1021/jacs.5b11511

Quesne, M. G., Borowski, T., and de Visser, S. P. (2016a). Quantum mechanics/molecular mechanics modeling of enzymatic processes: caveats and breakthroughs. *Chem. Eur. J.* 22, 2562–2581. doi: 10.1002/chem.201503802

Quesne, M. G., and de Visser, S. P. (2012). Regioselectivity of substrate hydroxylation versus halogenation by a nonheme iron(IV)-oxo complex: possibility of rearrangement pathways. *J. Biol. Inorg. Chem.* 17, 841–852. doi: 10.1007/s00775-012-0901-4

Quesne, M. G., Latifi, R., Gonzalez-Ovalle, L. E., Kumar, D., and de Visser, S. P. (2014). Quantum mechanics/molecular mechanics study on the oxygen binding and substrate hydroxylation step in AlkB repair enzymes. *Chem. Eur. J.* 20, 435–446. doi: 10.1002/chem.201303282

Quesne, M. G., Senthilnathan, D., Singh, D., Kumar, D., Maldivi, P., Sorokin, A. B., et al. (2016b). Origin of the enhanced reactivity of μ-nitrido-bridged diiron(IV)-oxo porphyrinoid complexes over cytochrome P450 compound, I. *ACS Catal.* 6, 2230–2243. doi: 10.1021/acscatal.5b02720

Ramaswamy, A. V., Sorrels, C. M., and Gerwick, W. H. (2007). Cloning and biochemical characterization of the hectochlorin biosynthetic gene cluster from the marine cyanobacterium lyngbya majuscule. *J. Nat. Prod.* 70, 1977–1986. doi: 10.1021/np0704250

Rittle, J., and Green, M. T. (2010). Cytochrome P450 compound I: capture, characterization, and C-H bond activation kinetics. *Science* 330, 933–937. doi: 10.1126/science.1193478

Ryle, M. J., and Hausinger, R. P. (2002). Non-heme iron oxygenases. *Curr. Opin. Chem. Biol.* 6, 193–201. doi: 10.1016/S1367-5931(02)00302-2

Sainna, M. A., Kumar, S., Kumar, D., Fornarini, S., Crestoni, M. E., and de Visser, S. P. (2015). A comprehensive test set of epoxidation rate constants for iron(IV)–oxo porphyrin cation radical complexes. *Chem. Sci.* 1516–1529. doi: 10.1039/C4SC02717E

Saouma, C. T., and Mayer, J. M. (2014). Do spin state and spin density affect hydrogen atom transfer reactivity? *Chem. Sci.* 5, 21–31. doi: 10.1039/C3SC52664J

Schafer, A., Horn, H., and Ahlrichs, R. (1992). Fully optimized contracted Gaussian basis sets for atoms Li to Kr. *J. Chem. Phys.* 97, 2571–2577. doi: 10.1063/1.463096

Schnepel, C., and Sewald, N. (2017). Enzymatic halogenation: a timely strategy for regioselective C–H activation. *Chem. Eur. J.* 23, 12064–12086. doi: 10.1002/chem.201701209

Schofield, C. J., and Zhang, Z. (1999). Structural and mechanistic studies on 2-oxoglutarate-dependent oxygenases and related enzymes. *Curr. Opin. Struc. Biol.* 9, 722–731. doi: 10.1016/S0959-440X(99)00036-6

Senn, H. M. (2014). Insights into enzymatic halogenation from computational studies. *Front. Chem.* 2:98. doi: 10.3389/fchem.2014.00098

Shaik, S., de Visser, S. P., and Kumar, D. (2004). external electric field will control the selectivity of enzymatic-like bond activations. *J. Am. Chem. Soc.* 126, 11746–11749. doi: 10.1021/ja047432k

Shaik, S., Kumar, D., de Visser, S. P., Altun, A., and Thiel, W. (2005). Theoretical perspective on the structure and mechanism of cytochrome P450 enzymes. *Chem. Rev.* 105, 2279–2328. doi: 10.1021/cr030722j

Sherwood, P., de Vries, A. H., Guest, M. F., Schreckenbach, G., Catlow, C. R. A., French, S. A., et al. (2003). QUASI: A general purpose implementation of the QM/MM approach and its application to problems in catalysis. *J. Mol. Struct.* 632, 1–28. doi: 10.1016/S0166-1280(03)00285-9

Sinnecker, S., Svensen, N., Barr, E. W., Ye, S., Bollinger Jr, J. M., Neese, F., et al. (2007). Spectroscopic and Computational evaluation of the structure of the high-Spin Fe(IV)-Oxo intermediates in taurine: α-ketoglutarate dioxygenase from *Escherichia coli* and its His99Ala ligand variant. *J. Am. Chem. Soc.* 129, 6168–6179. doi: 10.1021/ja067899q

Smith, W., and Forester, T. R. (1996). DL_POLY_2.0: a general-purpose parallel molecular dynamics simulation package. *J. Mol. Graphics.* 14, 136–141. doi: 10.1016/S0263-7855(96)00043-4

Solomon, E. I., Brunold, T. C., Davis, M. I., Kemsley, J. N., Lee, K., Lehnert, K., et al. (2000). Geometric and electronic structure/function correlations in non-heme iron enzymes. *Chem. Rev.* 100, 235–349. doi: 10.1021/cr9900275

Srnec, M., and Solomon, E. I. (2017). Frontier molecular orbital contributions to chlorination versus hydroxylation selectivity in the non-heme iron halogenase SyrB2. *J. Am. Chem. Soc.* 139, 2396–2407. doi: 10.1021/jacs.6b11995

Srnec, M., Wong, S. D., Matthews, M. L., Krebs, C., Bollinger J. M. Jr., and Solomon, E. I. (2016). Electronic structure of the ferryl intermediate in the α-ketoglutarate dependent non-heme iron halogenase SyrB2: contributions to H atom abstraction reactivity. *J. Am. Chem. Soc.* 138, 5110–5122. doi: 10.1021/jacs.6b01151

Sundaramoorthy, M., Terner, J., and Poulos, T. L. (1995). The crystal structure of chloroperoxidase: a heme peroxidase–cytochrome P450 functional hybrid. *Structure.* 3, 1367–1377. doi: 10.1016/S.0969-2126(01)00274-X.

Tang, H., Guan, J., Liu, H., and Huang, X. (2013). Analysis of an alternative to the H-atom abstraction mechanism in methane C–H bond activation by nonheme iron(IV)-oxo oxidants. *Dalton Trans.* 42, 10260–10270. doi: 10.1039/c3dt50866h

Tchesnokov, E. P., Faponle, A. S., Davies, C. G., Quesne, M. G., Turner, R., Fellner, M., et al. (2016). An iron–oxygen intermediate formed during the catalytic cycle of cysteine dioxygenase. *Chem. Commun.* 52, 8814–8817. doi: 10.1039/C6CC03904A

Timmins, A., and de Visser, S. P. (2015). Enzymatic halogenases and haloperoxidases: computational studies on mechanism and function. *Adv. Protein Chem. Struct. Biol.* 100, 113–151. doi: 10.1016/bs.apcsb.2015.06.001

Timmins, A., and de Visser, S. P. (2017). How are substrate binding and catalysis affected by mutating Glu127 and Arg161 in Prolyl-4-hydroxylase? A QM/MM and MD Study. *Front.Chem.* 5:94. doi: 10.3389/fchem.2017.00094

Timmins, A., and de Visser, S. P. (2018). A comparative review on the catalytic mechanism of nonheme iron hydroxylases and halogenase. *Catalysts* 8:314. doi: 10.3390/catal8080314

Timmins, A., Quesne, M. G., Borowski, T., and de Visser, S. P. (2018). Group transfer to an aliphatic bond: a biomimetic study inspired by nonheme iron halogenases. *ACS Catal.* 8, 8685–8698. doi: 10.1021/acscatal.8b01673

Timmins, A., Saint-André, M., and de Visser, S. P. (2017). Understanding how prolyl-4-hydroxylase structure steers a ferryl oxidant toward scission of a Strong C–H bond. *J. Am. Chem. Soc.* 139, 9855–9866. doi: 10.1021/jacs.7b02839

Usharani, D., Lacy, D. C., Borovik, A. S., and Shaik, S. (2013). Dichotomous hydrogen atom transfer vs. proton-coupled electron transfer during activation of X–H bonds (X = C, N, O) by nonheme iron–oxo complexes of variable basicity. *J. Am. Chem. Soc.* 135, 17090–17104. doi: 10.1021/ja408073m

Vaillancourt, F. H., Yeh, E., Vosburg, D. A., Garneaur-Tsodikova, S., and Walsh, C. T. (2006). nature's inventory of halogenation catalysts: oxidative strategies predominate. *Chem. Rev.* 106, 3364–3378. doi: 10.1021/cr050313i,

van Pée, H., Dong, C., Flecks, S., Naismith, J., Patallo, E. P., and Wage, T. (2006). Biological halogenation has moved far beyond haloperoxidases. *Adv. Appl. Microbiol.* 59, 127–157. doi: 10.1016/S0065-2164(06)59005-7

Vardhaman, A. K., Barman, P., Kumar, S., Sastri, C. V., Kumar, D., and de Visser, S. P. (2013). Comparison of the reactivity of nonheme iron(IV)-oxo versus iron(IV)-imido complexes: which is the better oxidant? *Angew. Chem. Int. Ed.* 52, 12288–12292. doi: 10.1002/anie.201305370

Vardhaman, A. K., Sastri, C. V., Kumar, D., and de Visser, S. P. (2011). Nonheme ferric hydroperoxo intermediates are efficient oxidants of bromide oxidation. *Chem. Commun.* 47, 11044–11046. doi: 10.1039/c1cc13775a

Wachters, A. J. H. (1970). Gaussian basis set for molecular wavefunctions containing third-row atoms. *J. Chem. Phys.* 52, 1033–1036. doi: 10.1063/1.1673095

Wagenknecht, A. H., and Woggon, D. W. (1997). Identification of intermediates in the catalytic cycle of chloroperoxidase. *Chem. Biol.* 4, 367–372. doi: 10.1016/S1074-5521(97)90127-7

Wagner, C., Omari, M. E., and Konig, G. M. (2009). Biohalogenation: nature's way to synthesize halogenated metabolites. *J. Nat. Prod.* 72, 540–553. doi: 10.1021/np800651m

Wang, Y., Li, X., Yang, G., He, G., and Chen, G. (2016). A visible-light-promoted radical reaction system for azidation and halogenation of tertiary aliphatic C–H bonds. *Chem. Sci.* 7, 2679–2683. doi: 10.1039/C5SC04169D

Weichold, V., Milbredt, D., and van Pée, K. H. (2016). Specific enzymatic halogenation—from the discovery of halogenated enzymes to their applications *in vitro* and *in vivo*. *Angew. Chem. Int. Ed.* 55, 6374–6389. doi: 10.1002/anie.201509573

Wójcik, A., Radon, M., and Borowski, T. (2016). Mechanism of O_2 activation by α-ketoglutarate dependent oxygenases revisited. A Quantum Chemical Study. *J. Phys. Chem. A* 120, 1261–1274. doi: 10.1021/acs.jpca.5b12311

Wong, S. D., Srnec, M., Matthews, M. L., Liu, L. V., Kwak, Y., Park, K., et al. Solomon, E. I. (2013). Elucidation of the iron(IV)–oxo intermediate in the non-haem iron halogenase SyrB2. *Nature* 499, 320–323. doi: 10.1038/nature12304

Yang, T., Quesne, M. G., Neu, H. M., Cantú Reinhard, F. G., Goldberg, D. P., and de Visser, S. P. (2016). Singlet versus triplet reactivity in an mn(v)–oxo species: testing theoretical predictions against experimental evidence. *J. Am. Chem. Soc.* 138, 12375–12386. doi: 10.1021/jacs.6b05027

Ye, S., and Neese, F. (2011). Nonheme oxo-iron(IV) intermediates form an oxyl radical upon approaching the C–H bond activation transition state. *Proc. Natl. Acad. Sci. U.S.A.* 108, 1228–1233. doi: 10.1073/pnas.1008411108

Zhao, J., Sun, X., Huang, X., and Li, J. (2016). Reaction mechanisms of methanol oxidation by FeIVO biomimetic complex. *Int. J. Quantum Chem.* 116, 692–701. doi: 10.1002/qua.25090

On the Inhibition Mechanism of Glutathione Transferase P1 by Piperlongumine. Insight From Theory

Mario Prejanò, Tiziana Marino* and Nino Russo

Dipartimento di Chimica e Tecnologie Chimiche, Università della Calabria, Arcavacata di Rende, Italy

Piperlongumine (PL) is an anticancer compound whose activity is related to the inhibition of human glutathione transferase of pi class (GSTP1) overexpressed in cancerous tumors and implicated in the metabolism of electrophilic compounds. In the present work, the inhibition mechanism of hydrolyzed piperlongumine (hPL) has been investigated employing QM and QM/MM levels of theory. The potential energy surfaces (PESs) underline the contributions of Tyr residue close to *G site* in the catalytic pocket of the enzyme. The proposed mechanism occurs through a one-step process represented by the nucleophilic addition of the glutathione thiol to electrophilic species giving rise to the simultaneous C-S and H-C bonds formation. Both the used methods give barrier heights (19.8 and 21.5 kcal mol^{-1} at QM/MM and QM, respectively) close to that experimentally measured for the C-S bond formations (23.8 kcal mol^{-1}).

Keywords: glutathione S-transferase, piperlongumine, hydrolysis mechanism, inhibition mechanism, MD DFT, QM, QMMM

INTRODUCTION

Glutathione-S-transferases (GSTs) is a ubiquitous family of multifunctional enzymes of phase II detoxification system that conjugate reactive substrates with reduced tripeptide glutathione (GSH) in most cells, especially those in the liver and kidney (Hayes et al., 2005; Oakley, 2005; Stoddard et al., 2017). In particular, they catalyze the nucleophilic attack of the thiol group arising from cysteine residue (Cys) of the GSH on electrophilic substrates leading to formation of conjugates, that are less toxic and more water-soluble than the parent species, facilitating their elimination from cells (Broxterman et al., 1995; Townsend and Tew, 2003; Wang et al., 2017). Their role in protecting the cells from oxidative attack, in association with their overexpression in many cancer cells, makes them good candidates as cancer biomarkers (McIlwain et al., 2006; Lo and Ali-Osman, 2007). Furthermore, glutathione-S-transferases are associated with multidrug resistance of tumor cells and are involved in drug detoxification and in apoptosis control (Townsend and Tew, 2003; Mejerman et al., 2008). Mammalian cytosolic GSTs isoenzymes belong to different families or classes (alpha, mu, pi, theta, kappa, sigma, zeta, and omega) (Wilce and Parker, 1994; Armstrong, 1997; Sheehan et al., 2001) based on their molecular masses, isoelectric points and other properties. Every isoenzyme subunit contains an active site entailing a binding site for the cofactor GSH (*G-site*) and one for the electrophilic substrate (*H-site*) (Dirr et al., 1994; Wilce and Parker, 1994). In particular, the Glutathione S-transferase P1 (GSTP1) is overexpressed in different human malignancies affecting important organs as lung, colon, stomach, kidney, ovary, mouth, and testis (Green et al., 1993; Katagiri et al., 1993; Grignon et al., 1994; Okuyama et al., 1994; Zhang et al., 1994; Inoue et al., 1995; Ruiz-Gomez et al., 2000). This overexpression has been linked to

acquire multidrug resistance to chemotherapeutic agents (cisplatin, chlorambucil, and ethacrynic acid) (Ban et al., 1996; Oakley et al., 1997; Mejerman et al., 2008; Karpusas et al., 2013; Pei et al., 2013; Perperopoulou et al., 2018). GSTP1 has additional role in maintaining the cellular redox state (Tew, 2007) and "nonenzymatic" antiapoptotic activity through its interaction with the *c-Jun* NH$_2$-terminal kinase (JNK), a key enzyme implicated into the apoptotic cascade (Adler et al., 1999; Wang et al., 2001). For these reasons, GSTP1 is considered as a promising target for inactivation in cancer treatment and numerous researchers have spent considerable effort to propose potent inhibitors of this enzyme (Bezerra et al., 2007; Federici et al., 2009; Raj et al., 2011; Adams et al., 2012; Boskovic et al., 2013; Liao et al., 2016; Harshbarger et al., 2017; Zou et al., 2018). Among these, piperlongumine (PL) is a natural alkaloid isolated from *Piper longum* L. characterized by the presence of two α, β- unsaturated functionalities (see **Figure 1**) and recently has been reported as a promising anticancer molecule by targeting the stress response to ROS, inducing apoptosis (Adams et al., 2012; Boskovic et al., 2013; Liao et al., 2016; Harshbarger et al., 2017).

This molecule also represents a promising lead compound in the developing potent GSTP1 inhibitors stimulating the synthesis of a huge number of its structural analogs (Bezerra et al., 2007; Adams et al., 2012; Boskovic et al., 2013; Liao et al., 2016; Harshbarger et al., 2017; Stoddard et al., 2017). PL acts as Michael acceptor because can undergo heteroconjugate addition with the peptide-like molecules including nucleophilic thiols of cysteine residues in irreversible or reversible fashion. From stable isotope labeling (Raj et al., 2011) the anti-cancer effects of PL were related to the promotion of reactive oxygen species (ROS) and to the reduction of GSH cellular levels (Harshbarger et al., 2017). PL contains a trimethoxyphenyl head and two reactive olefins moieties (C2-C3 and C7-C8) that revealed to be essential for differentiating the cellular activity (Adams et al., 2012). The C2-C3 bond is critical for toxicity, ROS elevation and protein S-glutathionylation while C7-C8 is not necessary for these activities and is believed to enhance the toxicity (Adams et al., 2012; Harshbarger et al., 2017). This means that the two present olefins can be identified as the minimum pharmacophore of PL so that their modifications can originate analogs with different biological response (Adams et al., 2012). Furthermore, it can act as GSTP1-cosubstrate in both displacement and addition reactions. In this case, GSH bound in the G site of GSTP1 is the target of the inhibitor (Adams et al., 2012; Harshbarger et al., 2017). Recently, the high resolution X-ray crystal structure of GSTP1 (PDB code 5J41) in complex with PL and GSH proposed as the inhibition occurs without GSTP1 covalent modification by PL but, rather unexpectedly, PL results to be hydrolyzed to a trimethoxycinnamic acid (TMCA) deprived of the C2-C3 olefin (Harshbarger et al., 2017). This finding does not completely fit the behavior of PL toward other cysteine-containing peptides that react with the C2-C3 reactive bond *in vitro* conditions (Adams et al., 2012). Harshbarger et al. provided the first structural model for the interactions between PL, GSH and GSTP1 (Harshbarger et al., 2017). From this study emerged that PL acts as a prodrug. In fact, after entrance in the cell it undergoes hydrolysis giving rise to the TMCA that in turns reacts with GSH, located in the

G site of GSTP1, affording the hPL:GSH conjugate as product of the addition reaction and confirming that no covalent bond formation occurs between PL and GSTP1. Although the presence in the literature of many scientific works (Bezerra et al., 2007; Federici et al., 2009; Adams et al., 2012; Boskovic et al., 2013; Peng et al., 2015; Liao et al., 2016; Harshbarger et al., 2017; Zou et al., 2018) on the piperlongumine selective inhibition of tumor growth in different types of cancers, the molecular mechanism involved in PL mediated cancer cell death remains still poorly understood. With the aim to contribute to a better knowledge, at atomistic level, of the inhibition mechanism of GSH by the hydrolyzed product of PL into the GSTP1 enzyme, a theoretical investigation in the framework of density functional theory (DFT) was undertaken. In addition, a MD simulation of initial enzyme-inhibitor (EI) complex has been also performed.

METHODOLOGY

Active Site

The enzyme structure includes two identical homodimers, with a total mass of 48 kDa. The active sites are located in the interfaces between the two domains. Each active site in turn consists of two sub-sites: the *G site*, where GSH is located, in proximity to the outer side of protein surface and in direct contact with solvent molecules, and the *H site*, where the electrophilic inhibitor can be accommodated (Harshbarger et al., 2017). The interactions between GSH and hPL with different residues of the cavity of both sites were treated at quantum mechanical level in both QM and QM/MM calculations. In particular, the QM region includes: the Arg13 which is engaged in hydrogen bonds with the N-terminal portion of GSH and the carboxyl group of inhibitor, the Lys44 which is anchored to C-terminal part of GSH, the Tyr7 with its OH moiety oriented toward the S atom of GSH-cysteine in such a manner to establish H bond between them, and Tyr108 which is involved in π-π interaction with inhibitor aromatic ring. Finally, the QM portion, contains also the Ile104 since its crucial role in correctly orienting hPL (in *H site*) during the conjugation phase with GSH (Harshbarger et al., 2017). Due to the closeness of active site to the protein surface, several water molecules present in the catalytic cavity were considered in the QM/MM model. Starting from the available crystallographic structure of GSTP1 by *Homo Sapiens* (Harshbarger et al., 2017; PDB code 5J41, 1.19 Å resolution), the preparation of the models (see **Figure 2**) is described by the following procedure.

MD Calculations

As first step of the work it was necessary to perform the C8$_{hPL}$-S$_{GSH}$ bond cleavage and then to relax the enzyme:GSH:inhibitor supramolecular system at the molecular mechanics (MM) level of theory before starting the MD simulation because the used X-ray structure was related to the final product of the inhibition process. Furthermore, the presence of the inhibitor molecule of non-protein nature implied its optimization at HF/6-31G(d) level of theory in order to derive the parameters by Antechamber tool, as implemented in AMBER 16 package (AMBER 16, 2016). Intramolecular Lennard-Jones parameters and atomic charges were obtained using, respectively, General Amber Force Field (GAFF) (Wang et al., 2004)
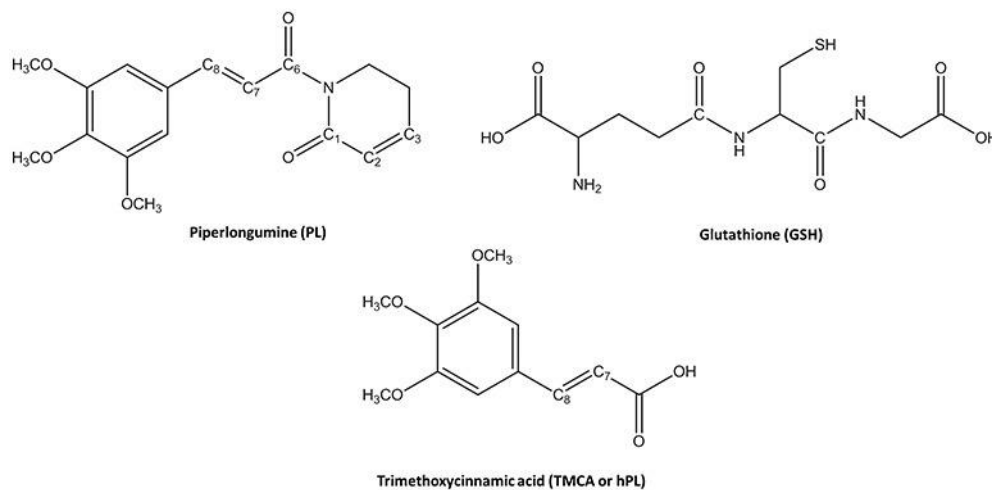
FIGURE 1 | Chemical Species involved in the inhibition reaction of GSTP1.
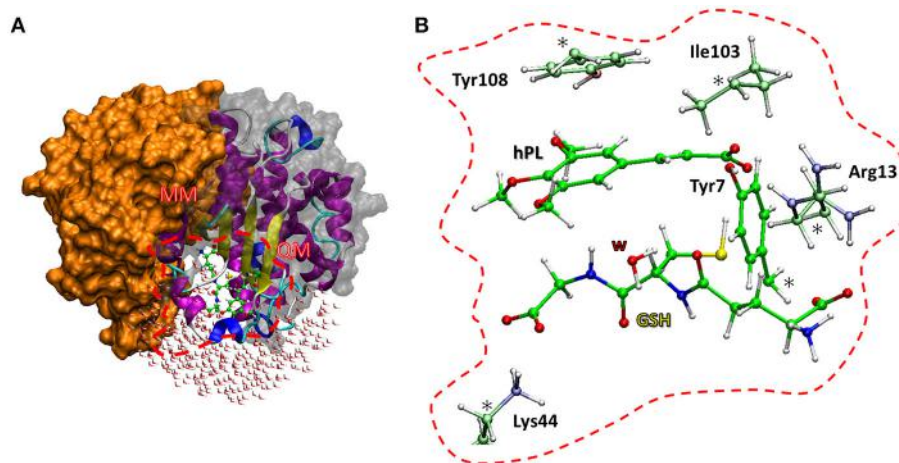
FIGURE 2 | The two models used. The QM/MM model **(A)** includes the whole protein with the water molecules, the red box defines the QM region used in both cluster **(B)** and QM/MM calculations. The amino acid residues of the QM portion are shown with ball and sticks representation.

and Restrained Electrostatic potential (RESP) method (Bayly et al., 1993). The obtained parameters of hPL are collected in **Table S1**.

The amber ff14SB (Maier et al., 2015) force field was applied using the Xleap module and hydrogen atoms were added to the whole system. The protonation state of each amino acid has been assigned using the H++ web server (Gordon et al., 2005; Myers et al., 2006; Anandakrishnan et al., 2012; H++ vesion 3.2, 2016). A rectangular box ($85 \times 70 \times 80$Å) was filled with TIP3P (Jorgensen et al., 1983) water molecules within 12.0 Å from the surface of the enzyme. The classical MD simulation was applied for 100 ps in NVT ensemble with a progressive heating phase, from 0 to 310 K. A final MD production of 20 ns was obtained in NPT ensemble (1 bar and 310 K). During the simulations, a cutoff radius for non-bonded interactions was fixed at 12 Å and Particle Mesh Ewald summation method

(PME) (Ewald, 1921) and SHAKE algorithm (Ryckaert et al., 1977) were employed to constrain the motion in H-including bonds, in order to use a 2 fs integration step The root-mean-square deviation (RMSD) analysis of the whole protein and the *H* and *G* active sites residues was performed to verify the stability of the system during the MD simulation (**Figure S1**). To better examine the conformational behavior of the inhibitor-protein system, a MD simulation has been also performed on the alone enzyme. The obtained root-mean-square fluctuation (RMSF) is shown in **Figure S2**. Furthermore, in order to verify conformational homogeneity for inhibitor binding modes in to the catalytic pocket, 20 structures were selected along MD simulation (**Figure S3**). Clustering results confirmed that the last frame, obtained at 20 ns, is a good representative configuration as to be adopted as starting configuration for creating QM cluster and QM/MM model.

FIGURE 3 | Hydrolysis mechanism of PL in neutral (bold line) and acidic conditions (dashed line).

FIGURE 4 | B3LYP-D3/6-311+G(2d,2p)//B3LYP/6-31+G(d,p) (SMD, $\varepsilon = 78$) energy profiles of PL hydrolysis in neutral conditions (blue line) and acid conditions (red line). The final energies contain ZPE and D3 corrections.
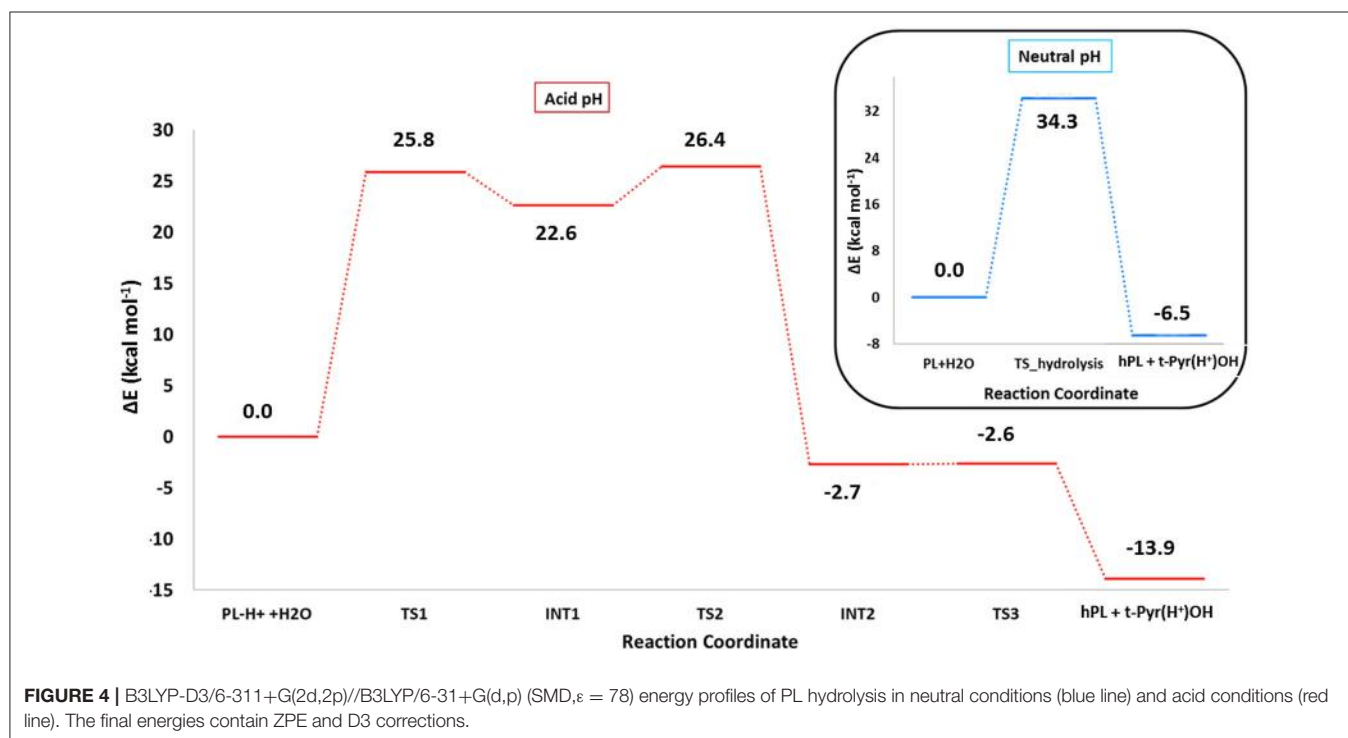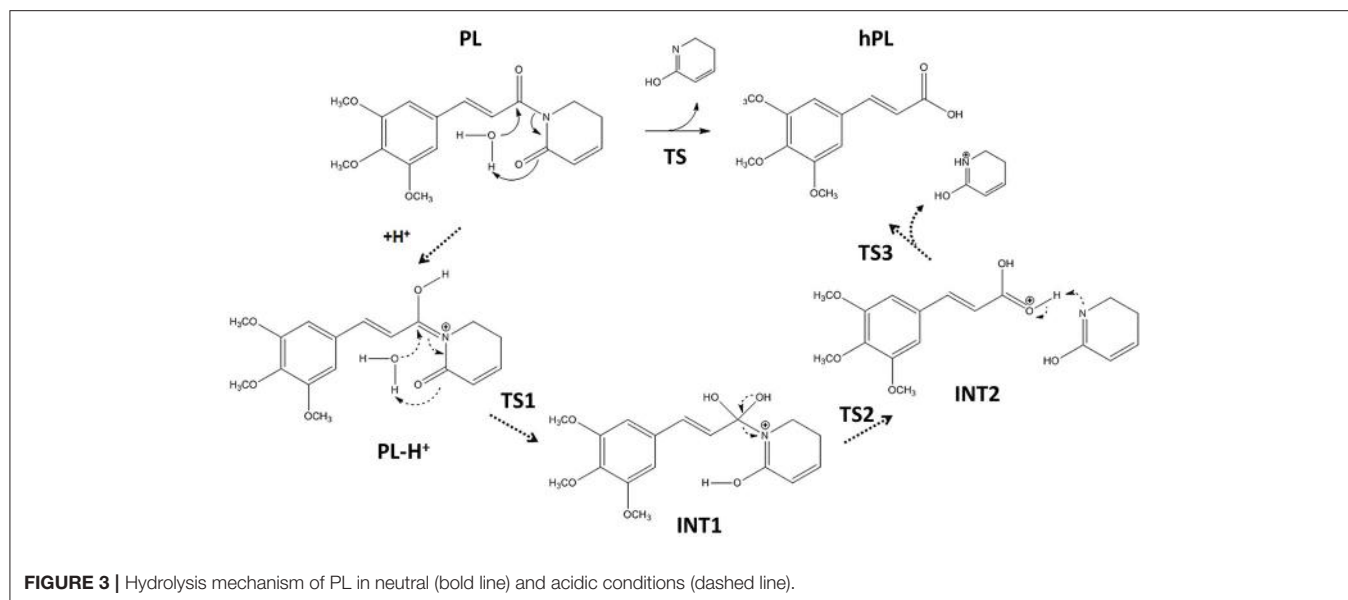
QM Cluster and QMMM Models

The amino acids considered in the QM region (Tyr7, Arg13, Tyr103, Ile104, Lys44) were truncated as depicted in **Figure 2**. The missing hydrogens were added manually and one water molecule (lying at 3.601 Å from the GSH) was explicitly considered, being implicated in direct interaction with nucleophilic agent while the other waters are located away than 4 Å. The C atoms labeled with "*" were kept fixed during geometry optimizations, applying the locking scheme, to prevent artificial

movements (Siegbahn and Himo, 2011; Piazzetta et al., 2015; Himo, 2017). The QM cluster model was found to be adequate in the elucidation of the catalytic mechanism followed by other enzymes (Amata et al., 2011; Lan and Chen, 2016; Prejanò et al., 2017). The obtained model consists of 136 atoms with a total charge equal to zero.

The QM/MM model was obtained applying the two layers ONIOM formalism (Svensson et al., 1996) as implemented in Gaussian09 code (Frisch et al., 2013), maintaining the same

FIGURE 5 | Proposed mechanisms for GSTP1 inhibition by hPL.

FIGURE 6 | Superposition of the ternary adduct (EI) from MD (violet) with the X-ray structure (yellow) related to the P S-conjugate product.
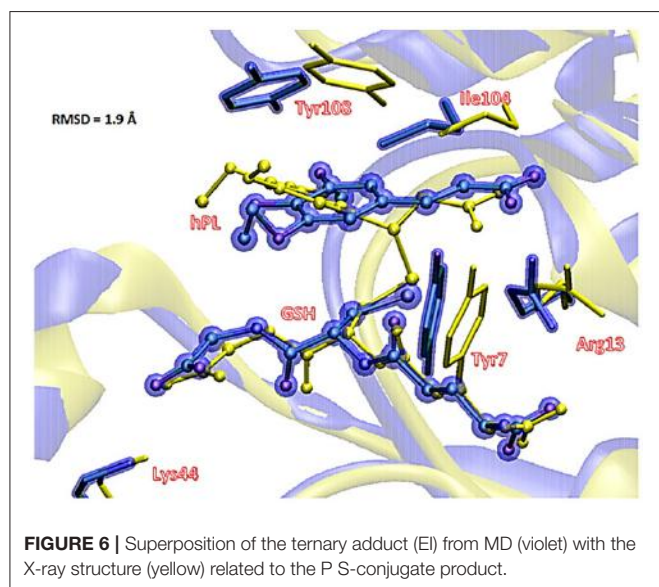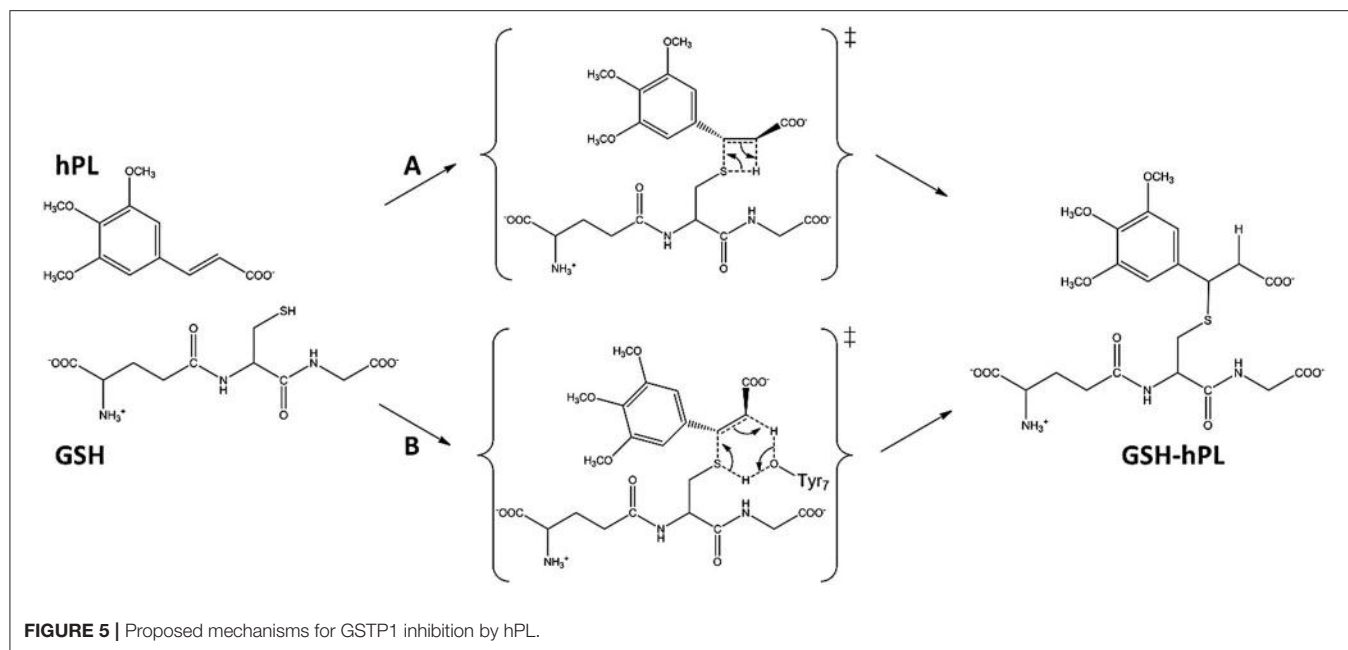
atoms mentioned in QM cluster model setup. The entire enzyme and a number of water molecules within 5 Å from the catalytic site were considered (**Figure 2**). During the optimization, all the water molecules and residues out of 18 Å sphere from the active site were kept frozen, applying the standard procedure for single conformation PES studies (Sousa et al., 2017). The final model contains 7811 atoms.

Technical Details

Gaussian 09 (Frisch et al., 2013) software package was used to perform calculations using B3LYP (Lee et al., 1988; Becke, 1993) hybrid functional in QM region of both used models.

For S, N, H, O and C atoms, 6-31+G(d,p) basis set was used during the optimizations. Linear transit scans were performed, in order to detect stationary points along reaction coordinates. To confirm the nature of intermediates or transition states, frequencies calculation was performed at the same level of theory, for each stationary point intercepted along potential energy surface (PES). To obtain more accurate electronic energies single point calculations with 6-311+G(2d,2p) larger basis set were performed. The final energy profiles include the zero point energy (ZPE) and dispersion corrections (evaluated using the DFT-D3 procedure; Grimme et al., 2011) and solvation energy.

The electrostatic embedding as implemented in Gaussian 09 was employed to evaluate the Coulomb interactions between MM and QM regions in all calculations (Vreven et al., 2006). For the QM cluster calculations, single point calculations adopting the SCRF-SMD solvation model with a dielectric constant $\varepsilon = 4$, simulating the enzyme environment, was used (Marenich et al., 2009). The same level of theory was adopted during the optimizations of species involved in hydrolysis of PL, considering the dielectric constant of 78.0, as successfully proposed in other studies (Ritacco et al., 2015; Marino et al., 2016). NBO (NBO version 3.1, 2001) and non-covalent interaction (NCI) (NCIPLOT, version 3.0, 2011) analyses were performed on all the stationary points of the investigated PESs at both QM and QM/MM levels.

As far as the proton affinity calculations for establishing the oxygen carbonyl to be considered in the hydrolysis mechanism at acidic conditions, the proton affinities as binding energies (BE) have been estimated as indicated by the following expression:

$$BE = -(\Delta H_{hPL-H+} - \Delta H_{hPL})$$

The BE is calculated as the difference between the enthalpy of the protonated system and that of the neutral one. In the calculations,

the H$^+$ contribution does not appear since we evaluated the energetic difference, therefore the obtained binding energies represent the energy involved in the formation of the protonated systems.

RESULTS AND DISCUSSION

Hydrolysis of Piperlongumine

Following the recent experimental indications that demonstrate as the PL suffers hydrolysis out of the enzyme pocket, (Harshbarger et al., 2017) we firstly study this process in aqueous media. The considered reaction mechanism is illustrated in

Figure 3. As from the experimental evidence, (Harshbarger et al., 2017) we considered the hydrolysis mechanism of PL to occur on the oxygen of the carbonyl (C6) functionality next to C7-C8 olefin, under both neutral and acidic conditions to take into account the different intracellular pH conditions, since the acid pH values are observed in cancer cells (Townsend and Tew, 2003; Wang et al., 2017).

On the contrary, our computed BE shows that the carbonyl moiety next to C2-C3 olefin has minor proton affinity (about 4 kcal mol^{-1}) with respect to that next to C7-C8 one, indicating as under the same conditions the favored protonation site is the oxygen of C6.

FIGURE 7 | /B3LYP-D3/6-311+G(2d,2p)//B3LYP/6-31+G(d,p) (SMD,ε = 4) (top) and B3LYP-D3/6-311+G(2d,2p):ff99SB//B3LYP/6-31+G(d,p):ff99SB (bottom) energy profiles of GSTP1 inhibition process by hPL for (A,B) mechanisms. In the black window are depicted the energy profile related to the reaction unassisted by the enzyme. The final energies contain ZPE and D3 corrections.
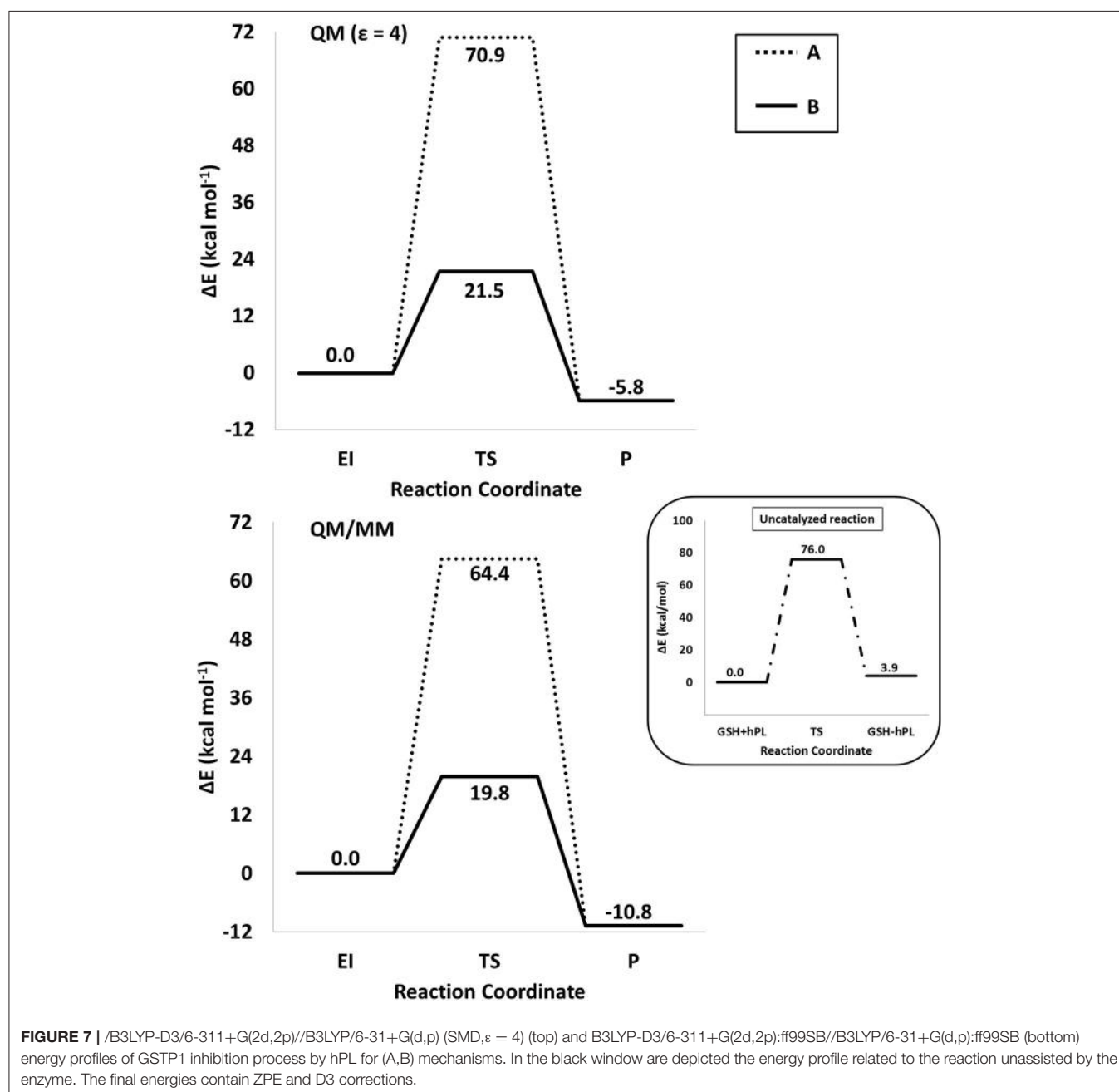
The optimized geometries of the stationary points are reported in (**Figures S4**, **S5**), while the calculated energy profiles are depicted in **Figure 4**. As shown from **Figure 3**, we propose at acid pH a mechanism occurring in a multistep process contrary to that at neutral conditions occurring in only one step. In both cases the product is the hPL, while the leaving group is 1,2,5,6-tetra-hydro-pyridin-2-ol (t-PyrOH) neutral and protonated, respectively. For clarity, in the text the remaining double bond in the hPL upon hydrolysis will retain the same numeration than in PL, (C7-C8). The processes are exothermic although at pH acid the exoergonicity is more pronounced (**Figure 4**). From our results, the acidic hydrolysis is strongly favored as suggested by lower activation barriers (by about 10 kcal mol^{-1}) than that found at neutral pH (see **Figure 4**). The calculated barrier in acidic environment well-agrees with those characterizing other anticancer molecules acting as prodrug (Alberto and Russo, 2011; Ritacco et al., 2015; Marino et al., 2016). Once the hydrolyzed product is formed, the process of GSH-conjugation favored by GSTP1 starts through the attack to the C7-C8 double bond.

GSTP1 Inhibition

To underline the role of GSTP1 during the inhibition process by hPL, we have considered, at both QM and QM/MM levels, two different reaction mechanisms (**A**, and **B**) as presented in **Figure 5**. In particular: (**A**) describes the nucleophile addition to the double bond of inhibitor by -SH group of GSH without involving any amino acid residue while path (**B**) takes into account the participation of the Tyr7 residue in the formation of the covalent adduct. In all the cases, the inhibition reaction occurs in a one-step process by the Michael addition of the thiol from GSH at the C7-C8 olefin of hPL. In all the considered mechanisms, the starting species is the ternary enzyme-hydrolyzed inhibitor-GSH complex (**EI**) obtained after the geometry optimization of the frame isolated by the previous MD cycle at 20 ns. From **Figure 6**, that illustrates the superposition of the crystallographic structure with the last MD snapshot, it is possible to note that hPL interacts with the binding cleft of the *H site* and no water molecules are close to the reaction center, in agreement with the hydrophobic nature of site (Tyr7, Tyr108, and Ile104 residues).
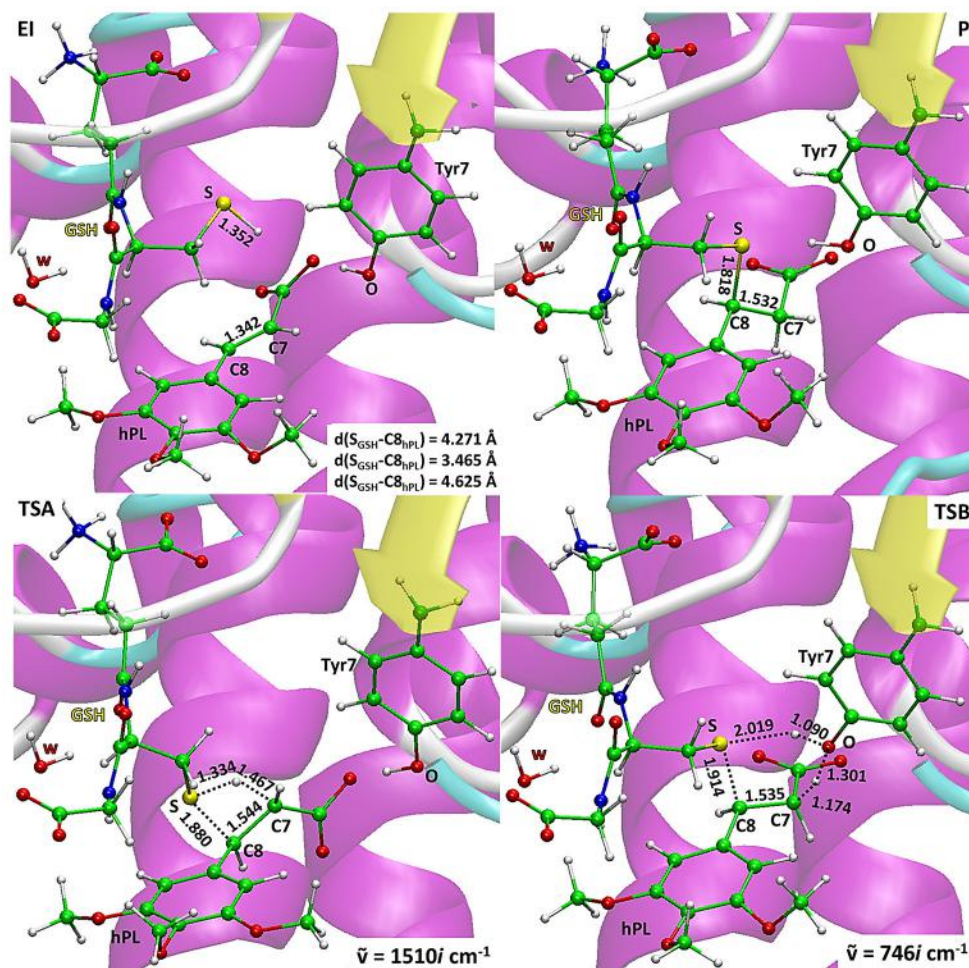

FIGURE 8 | B3LYP/6-31+G(d,p):ff99SB optimized geometries of EI, P, TSA, TSB. For the transition states the imaginary frequencies are reported.

As expected, the crystallographic pose (obtained with the reaction product) deviates in this moiety (see **Figure 6**). At the contrary, the GSH region is well-superimposed confirming that this molecule is well-positioned with a correct orientation of the thiol.
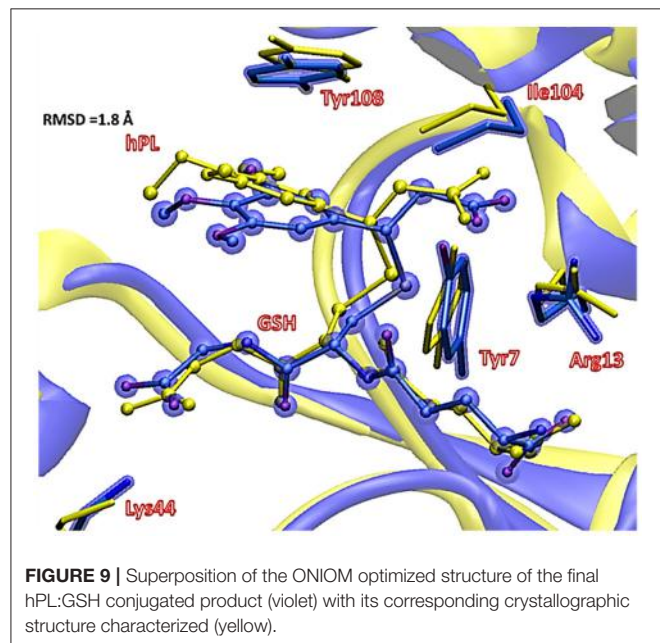
The energy profiles obtained employing QM and QM/MM tools for the two considered mechanisms, are reported in **Figure 7**. The reported QMMM energy values do not include the entropic contribution. In order to quantify this the Grimme procedure has been employed (Grimme, 2012). Results (see **Table S2**) evidence that the TΔS terms slightly affect the previously obtained energy values. QM/MM structures of the initial complex (**EI**), the final S-conjugated product (**P**) and that of the transition states are reported in **Figure 8**. All the QM optimized geometries are given in (**Figure S6**).
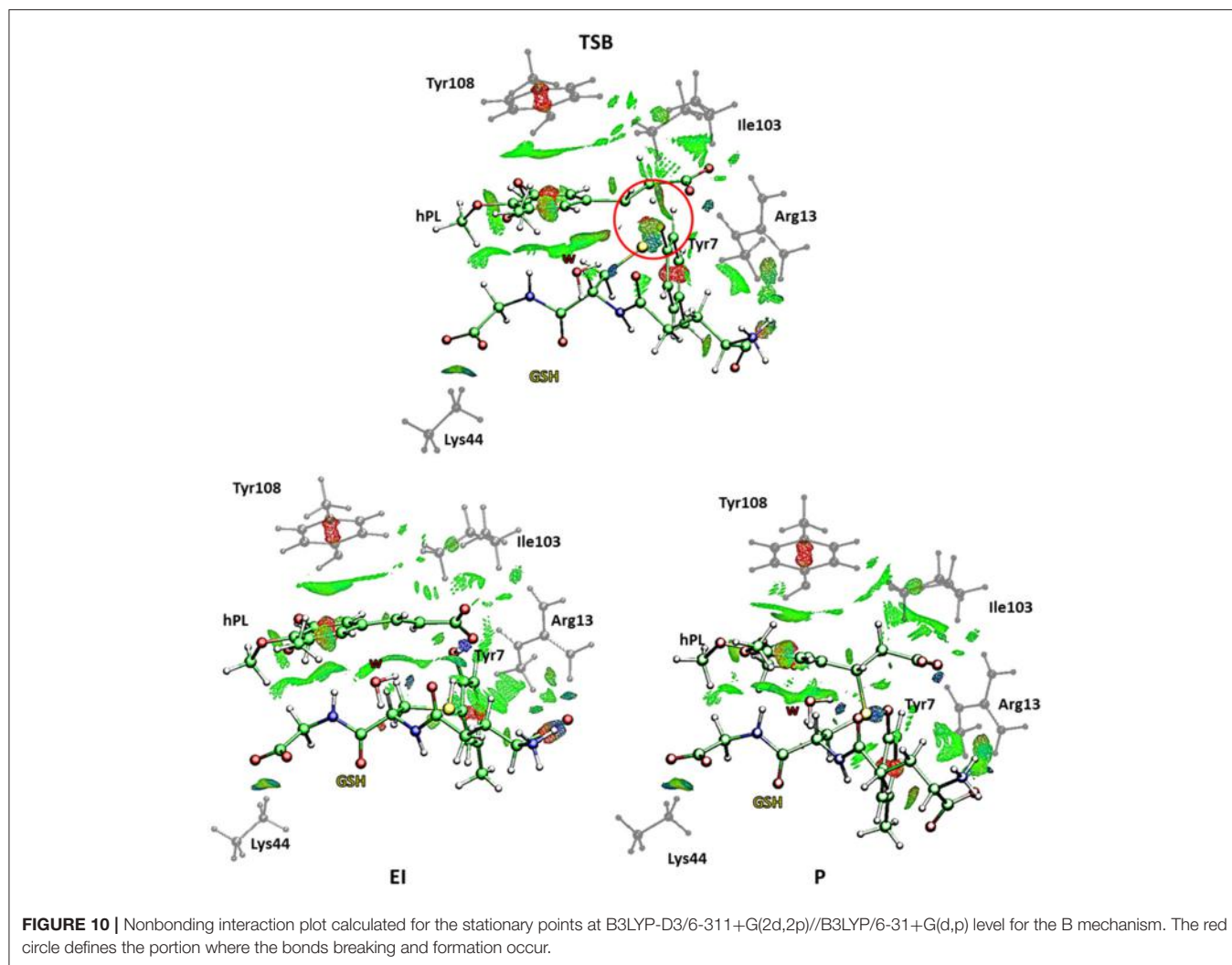
In **EI** the carboxylate moiety of the hPL is oriented in such a way to establish hydrogen bonds with Tyr7 (1.599 Å) and SH group of GSH (2.138 Å). Furthermore, van der Waals and hydrophobic interactions, such as those between the inhibitor and the Ile104 and Arg14 residues (see **Figure 8**) contribute to optimally accommodate the inhibitor into the H binding site. In fact, now the key reacting atoms, C8 of the hPL and SH nucleophile species, are placed in suitable way (at 4.271 Å) to allow the deactivation reaction. In path **A** the intercepted transition state (**TSA**) represents a four-centered structure where the sulfur addition to the C8 (1.880 Å) and the C7-H bond formation (1.467 Å) simultaneously occur. The corresponding frequency analysis confirms a first-order saddle point with an imaginary frequency (1510i cm^{-1}) which corresponds to a vibrational mode involving a strong C7–H coupling and a relatively weaker C8–S one. The C8-S bond is already established and the forming C-H one can be evinced by the elongation of the C7-C8 bond (1.544 Å). This barrier results to be 64.4

kcal mol^{-1} at QM/MM level and 70.9 kcal mol^{-1} at QM one. Both values are very close to that computed for the reaction unassisted by the catalyst (76.0 kcal mol^{-1}, see **Figure 7**). The resulting product (**P**), shown in **Figure 8,** evidences that the C-S bond is formed (1.818 Å) and the C8-C7 is elongated (1.532 Å) confirming the occurred sp^3 hybridization of the two involved carbon atoms. The exothermicity is evaluated to be 10.8 kcal mol^{-1} (5.8 kcal mol^{-1} in the QM cluster). The mechanism **B** (**Figure 5**) involves the participation of the Tyr7 residue. In **Figure 8** is reported the optimized structure of the **TSB** connecting the **EI** and the covalent final complex (**P**). The nucleophilic attack to C8 occurs by GSH-thiolate (1.914 Å) since the hydrogen of the S-H$_{GSH}$ group (2.019 Å) has been delivered to oxygen (O$_{Tyr}$) of the side chain of Tyr7 (1.090 Å). In fact, the OH group of Tyr7, oriented via hydrogen bonding to carboxylate moiety of the inhibitor (1.599 Å), in the TS becomes 1.310 Å and points toward C7 atom for delivering its hydrogen atom (O-H and H-C7 distances are found to be 1.310 and 1.174 Å, respectively) while the C7-C8 bond is elongated (1.535 Å) (see **Figure 8**). The TS located along the mechanism **B** lies at 21.5 kcal mol^{-1} (QM) and 19.8 kcal mol^{-1} (QM/MM) above the **EI**. Both values are very close to the available experimental one (23.8 kcal mol^{-1}) concerning the C-S bond formation (Huskey et al., 1991). The superposition of our optimized glutathionil-conjugated product **P** with the corresponding crystallographic structure (Harshbarger et al., 2017; see **Figure 9**) reveals a good RMSD value in both GSH and *H site* regions. The exothermicity 10.8 kcal mol^{-1} means that the reverse reaction can be accessible but much slower also for the high barrier required in the reverse process **P** EI (30.6 kcal mol^{-1}). **TSB** evidence as the formation of the S-C8 bond is strictly related to the deprotonation of the SH moiety of GSH at the expense of the Tyr7 acting as proton shuttle with a consequent reduction of barrier (19.8 kcal mol^{-1}).

This is in agreement with previous works on other GST enzymes (Zheng and Ornstein, 1997; Angelucci et al., 2005; Dourado et al., 2008) revealing the importance of the acidic properties of a Tyr during the catalysis of glutathione-S-Transferase. Furthermore, our findings corroborate the hypothesis advanced by the previous structural analysis (Harshbarger et al., 2017) revealing as no covalent bond formation between hPL and GSPT1 was observed and that PL acts as prodrug. With the aim to evaluate the nature of the interactions present inside the catalytic pocket during the process, in **Figure 10** we have reported the density of isosurfaces arising from NCI analysis, indicating the different contributions of the residues retained in the QM region.

In every characterized stationary point, it can be noted the salt bridges occurring between the side-chains of Lys44 and Arg13 with the carboxyl moiety of carboxyl- and amino-terminal of GSH (blue region indicates strong attractive interactions while the red isosurfaces account for the repulsive interactions related to the center of π systems of Tyr7 and Tyr108 and the inhibitor molecule, as usually for aromatic systems strong non-bonded overlap is indicated (Johnson et al., 2010). Further information arises from the green regions indicative of the van der Waals forces characterizing the cavity containing the inhibitor molecule

FIGURE 9 | Superposition of the ONIOM optimized structure of the final hPL:GSH conjugated product (violet) with its corresponding crystallographic structure characterized (yellow).

FIGURE 10 | Nonbonding interaction plot calculated for the stationary points at B3LYP-D3/6-311+G(2d,2p)//B3LYP/6-31+G(d,p) level for the B mechanism. The red circle defines the portion where the bonds breaking and formation occur.

and identified by the hydrophobic residues Tyr108 and Ile104. It is interesting to underline as the interaction involving the Ile104 becomes more intense as the reaction proceeds. At the contrary no relevant contributions arise from the NBO analysis (see **Table S3**) except for a little bit increased nucleophile nature of the sulfur atom of GSH in the enzyme and a decreased negative charge on the C7-C8 bond of the hPL species with the respect to the corresponding values obtained for the process unassisted by enzyme. In the **TSB** species, a more attractive interaction appears in proximity of the region interested in the chemical events (circled in red in **Figure 10**) symptomatic of the occurring S-C8 bond formation. Furthermore, the interatomic distances, during the mechanism, between the residues of the QM region and the GSH and hPL species, reported in **Table S4**, highlight how no significant change occurs in the catalytic pocket.

CONCLUSION

This study focuses on the inhibition mechanism of the glutathione-S-transferase Pi 1 by the hydrolyzed product of

piperlongumine. We propose the mechanism following the most recent experimental evidences taking into account in particular the role of Tyr7 on the complex formation between the glutathione and the inhibitor inside the catalytic pocket of enzyme.

The hydrolysis of PL for giving hPL has been considered in neutral and acid conditions. The last **one** provided the better energetic path.

The agreement between cluster QM and the more computational demanding hybrid QM/MM methods is quite good. Structural and energetic computed properties are in line with the available experimental data.

The lowest energy reaction mechanism for reaction of hPL with GSH corresponds to that in which the Tyr7 residue is involved in the inhibition reaction deprotonating the GSH and donates the proton, in a concerted fashion, to the C7 substrate atom. The computed barrier heights result 19.8 and 21.5 kcal mol^{-1} in both QM/MM and QM models, respectively. Both computations clearly indicate the same reaction mechanism by TSB as the preferred one with difference in the barrier eight is of only 1.7 kcal mol^{-1}and propose an exergonic reaction.

We hope that the obtained new insights on the reaction mechanism of human GSTP1 inhibition with natural piperlongumine substrate can be useful in the design of new selective and more potent inhibitors.

AUTHOR CONTRIBUTIONS

MP, TM, and NR have analyzed the results, edit, and reviewed equally the manuscript. MP, TM, and NR approved it for publications.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2018.00606/full#supplementary-material

REFERENCES

Adams, D. J., Dai, M. G., Pellegrino, G., Wagner, B. K., Stern, A. M., Shamji, A. F., et al. (2012). Synthesis, cellular evaluation, and mechanism of action of piperlongumine analogs. *PNAS* 109, 15115–15120. doi: 10.1073/pnas.1212802109

Adler, V., Yin, Z., Fuchs, S. Y., Benezra, M., Rosario, L., Tew, K. D., et al. (1999). Regulation of JNK signaling by GSTp. *EMBO J.* 18, 1321–1334. doi: 10.1093/emboj/18.5.1321

Alberto, M. E., and Russo, N. (2011). Methionine ligand selectively promotes monofunctional adducts between trans-EE platinum anticancer drug and guanine DNA base. *Chem. Commun.* 47, 887–889. doi: 10.1039/C0CC03605F

Amata, O., Marino, T., Russo, N., and Toscano, M. (2011). Catalytic activity of a ζ-class zinc and cadmium containing carbonic anhydrase. *Compared work mechanisms. Phys. Chem. Chem. Phys.* 13, 3468–3477. doi: 10.1039/C0CP01053G

AMBER 16 (2016). University of California, San Francisco, CA.

Anandakrishnan, R., Aguilar, B., and Onufriev, A. V. (2012). H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* 40, W537–541. doi: 10.1093/nar/gks375

Angelucci, F., Baiocco, P., Brunori, M., Gourlay, L., Morea, and, V., and Bellelli, A. (2005). Insights into the catalytic mechanism of glutathione s-transferase: the lesson from schistosoma haematobium. *Structure* 13, 1241–1246. doi: 10.1016/j.str.2005.06.007

Armstrong, R. N. (1997). Structure, catalytic mechanism, and evolution of the glutathione transferases. *Chem. Res. Toxicol.* 10, 2–18. doi: 10.1021/tx960072x

Ban, N., Takahashi, Y., Takayama, T., Kura, T., Katahira, T., Sakamaki, S., et al. (1996). Transfection of Glutathione S-Transferase (GST)-π Antisense Complementary DNA Increases the Sensitivity of a Colon Cancer Cell Line to Adriamycin, Cisplatin, Melphalan, and Etoposide. *Cancer Res.* 56, 3577–3582.

Bayly, C. I., Cieplak, P., Cornell, W. D., and Kollman, P. A. (1993). A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* 97, 10269–10280. doi: 10.1021/j100142a004

Becke, A. D. (1993). Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* 98, 5648–5652. doi: 10.1063/1.464913

Bezerra, D. P., Militão, G. C., de Castro, F. O., Pessoa, C., de Moraes, M. O., Silveira, E. R., et al. (2007). Piplartine induces inhibition of leukemia cell proliferation triggering both apoptosis and necrosis pathways. *Toxicol. in vitro* 21, 1–8. doi: 10.1016/j.tiv.2006.07.007

Boskovic, Z. V., Hussain, M. M., Drew, J. A., Mingji, D., and Schreiber, S. L. (2013). Synthesis of piperlogs and analysis of their effects on cells. *Tetrahedron* 69, 7559–7567. doi: 10.1016/j.tet.2013.05.080

Broxterman, H. J., Giaccone, G., and Lankelma, J. (1995). Multidrug resistance proteins and other drug transport-related resistance to natural product agents. *Curr. Opin. Oncol.* 7, 532–540. doi: 10.1097/00001622-199511000-0001

Dirr, H., Reinemer, P., and Huber, R. (1994). X-ray crystal structures of cytosolic glutathione S-transferases. Implications for protein architecture, substrate recognition and catalytic function. *Eur. J. Biochem.* 220, 645–661. doi: 10.1111/j.1432-1033.1994.tb18666.x

Dourado, D. F. A. R., Fernandes, P. A., Mannervik, B., and Ramos, M. J. (2008). Glutathione transferase: new model for glutathione activation. *Chem. Eur. J.* 14, 9591–9598. doi: 10.1002/chem.200800946.

Ewald, P. P. (1921). Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys-Berlin* 64, 253–287.

Federici, L., Lo Sterzo, C., Pezzola, S., Di Matteo, A., Scaloni, F., Federici, G., et al. (2009). Structural basis for the binding of the anticancer compound 6-(7-nitro-2,1,3-benzoxadiazol-4-ylthio)hexanol to human glutathione s-transferases. *Cancer Res.* 69, 8025–8034. doi: 10.1158/0008-5472.CAN-09-1314

Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., et al. (2013). *Gaussian 09, Revision D.01*. Wallingford, CT: Gaussian, Inc.

Gordon, J. C., Myers, J. B., Folta, T. M., Shoja, V., Heath, L. S., and Onufriev, A. (2005). H++: a server for estimating p Ka s and adding missing hydrogens to macromolecules. *Nucleic Acids Res.* 33,W368–371. doi: 10.1093/nar/gki464

Green, J. A., Robertson, L. J., and Clark, A. H. (1993). S-transferase expression in benign and malignant ovarian tumours. *Br. J. Cancer* 68, 235–239.

Grignon, D. J., Abdel-Malak, M., Mertens, W. C., Sakr, W. A., and Shepherd, R. (1994). Glutathione S-transferase expression in renal cell carcinoma: a new marker of differentiation. *Mod. Pathol.* 7, 186–189.

Grimme, S. (2012), Supramolecular binding thermodynamics by dispersion-corrected density functional theory. *Chem. Eur. J.*18, 9955–9964. doi: 10.1002/chem.201200497

Grimme, S., Ehrlich, S., and Goerigk, L.,(2011). Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* 32, 1456–1465. doi: 10.1002/jcc.21759

H++ vesion 3.2 (2016). Available online at: http://biophysics.cs.vt.edu/H++

Harshbarger, W., Gondi, S., Ficarro, S. B., Hunter, J., Udayakumar, D., Gurbani, D., et al.(2017). Structural and biochemical analyses reveal the mechanism of glutathione S-transferase Pi 1 inhibition by the anti-cancer compound piperlongumine. *J. Biol. Chem.* 292, 112–120. doi: 10.1074/jbc.M116.750299

Hayes, J. D., Flanagan, J. U., and Jowsey, I. R. (2005). Glutathione transferases. *Annu. Rev. Pharmacol. Toxicol.* 45, 51–88. doi: 10.1146/annurev.pharmtox.45.120403.095857

Himo, F. (2017). Recent trends in quantum chemical modeling of enzymatic reactions. *J. Am. Chem. Soc.*139, 6780–6786. doi: 10.1021/jacs.7b02671

Huskey, S. E. W., Phillip Huskey, W., and Lu, A. Y. H. (1991). Contributions of thiolate "desolvation" to catalysis by glutathione S-transferase isozymes 1-1 and 2-2: evidence from kinetic solvent isotope effects. *J. Am. Chem. Soc.* 113, 2283–2290. doi: 10.1021/ja00006a056

Inoue, T., Ishida, T., Sugio, K., Maehara, Y., and Sugimachi, K. (1995). Glutathione S transferase Pi is a powerful indicator in chemotherapy of human lung squamous-cell carcinoma. *Respiration* 62, 223–227. doi: 10.1159/000196451

Johnson, E. R., Keinan, S., Mori-Sanchez, P., Contreras-Garcia, J., Cohen, A. J., and Yang, W. (2010). Revealing noncovalent interactions. *J. Am. Chem. Soc.*, 132, 6498–6506. doi: 10.1021/ja100936w

Jorgensen, W., Chandrasekhar, J., Madura, J., and Klein, M. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935. doi: 10.1063/1.445869

Karpusas, M., Axarli, I., Chiniadis, L., Papakyriakou, A., Bethanis, K., Scopelitou, K., et al. (2013). The interaction of the chemotherapeutic drug chlorambucil with human glutathione transferase A1-1: kinetic and structural analysis. *PLoS ONE* 8:e56337. doi: 10.1371/journal.pone.0056337

Katagiri, A., Tomita, Y., Nishiyama, T., Kimura, M., and Sato, S. (1993). Immunohistochemical detection of P-glycoprotein and GSTP1-1 in testis cancer. *Br. J. Cancer.* 68, 125–129. doi: 10.1038/bjc.1993.299

Lan, C.-L., and Chen, S.-L. (2016). The Decarboxylation of α,β-Unsaturated Acid Catalyzed by Prenylated FMN-Dependent Ferulic Acid Decarboxylase and the Enzyme Inhibition. *J. Org. Chem.* 81, 9289–9295. doi: 10.1021/acs.joc.6b01872

Lee, C. T., Yang, W. T., and Parr, R. G. (1988). Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* 37, 785–789. doi: 10.1103/PhysRevB.37.785

Liao, Y., Niu, X., Chen, B., Edwards, H., Xu, L., Xie, C., et al. (2016). Synthesis and antileukemic activities of piperlongumine and HDAC inhibitor hybrids against acute myeloid leukemia cells. *J. Med. Chem.* 59, 7974–7990. doi: 10.1021/acs.jmedchem.6b00772

Lo, H. W., and Ali-Osman, F. (2007). Genetic polymorphism and function of glutathione S-transferases in tumor drug resistance. *Curr. Opin. Pharmacol.* 7, 367–374. doi: 10.1016/j.coph.2007.06.009

Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simemrling, C. (2015). ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* 11, 3696–3713. doi: 10.1021/acs.jctc.5b00255

Marenich, A. V., Cramer, C. J., and Truhlar, D. G. (2009). Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B* 113, 6378–6396. doi: 10.1021/jp810292n

Marino, T., Parise, A., and Russo, N. (2016). The role of arsenic in the hydrolysis and DNA metalation processes in an arsenous acid–platinum(II) anticancer complex. *Phys. Chem. Chem. Phys.* 19, 1328-−1334. doi: 10.1039/C6CP06179F

McIlwain, C. C., Townsend, D. M., and Tew, K. D. (2006). Glutathione S-transferase polymorphisms: cancer incidence and therapy. *Oncogene* 25, 1639–1648. doi: 10.1038/sj.onc.1209373

Mejerman, I., Beijnen, J. H., and Schellens, J. H. (2008). Combined action and regulation of phase II enzymes and multidrug resistance proteins in multidrug resistance in cancer. *Cancer Treat. Rev.* 34, 505–520. doi: 10.1016/j.ctrv.2008.03.002

Myers, J., Grothaus, G., Narayanan, S., and Onufriev, A. (2006). A simple clustering algorithm can be accurate enough for use in calculations of pKs in macromolecules. *Proteins* 63, 928–938. doi: 10.1002/prot.20922

NBO, version 3.1 (2001).

NCIPLOT, version 3.0 (2011). Available online at: http://www.lct.jussieu.fr/pagesperso/contrera/nciplot.html

Oakley, A. J. (2005). Glutathione transferases: new functions. *Curr. Opin. Struct. Biol.* 15, 716–723. doi: 10.1016/j.sbi.2005.10.005

Oakley, A. J., Rossjohn, J., Lo Bello, M., Caccuri, A. M., Federici, G., and Parker, M. W. (1997). Multifunctional role of Tyr 108 in the catalytic mechanism of human glutathione transferase P1-1. Crystallographic and kinetic studies on the Y108F mutant enzyme. *Biochemistry* 36, 576–585. doi: 10.1021/bi962813z

Okuyama, T., Maehara, Y., Endo, K., Baba, H., Adachi, Y., Kuwano, M., et al.(1994). Expression of glutathione s-transferasepi and sensitivity of human gastric cancer cells to cisplatin. *Cancer* 74, 1230–1236. doi: 10.1002/1097-0142(19940815)74:4<1230::AID-CNCR2820740409>3.0.CO;2-0

Pei, S., Minhajuddin, M., Callahan, K. P., Balys, M., Ashton, J. M., Neering, S. J., et al. (2013). Targeting aberrant glutathione metabolism to eradicate human acute myelogenous leukemia cells. *J. Biol. Chem.* 288, 33542- 33558. doi: 10.1074/jbc.M113.511170

Peng, S., Zhang, B., Meng, X., Yao, Y., and Fang, J. (2015). Synthesis of piperlongumine analogues and discovery of nuclear factor erythroid 2-related factor 2 (Nrf2) activators as potential neuroprotective agents. *J. Med. Chem.* 58, 5242–5255. doi: 10.1021/acs.jmedchem.5b00410

Perperopoulou, F., Pouliou, F., and Labrou, N. E. (2018). Recent advances in protein engineering and biotechnological applications

of glutathione transferases. *Crit. Rev. Biotechnol.* 38, 511–528. doi: 10.1080/07388551.2017.1375890

Piazzetta, P., Marino, T., and Russo, N., and Salahub, D. R. (2015). Direct hydrogenation of carbon dioxide by an artificial reductase obtained by substituting rhodium for zinc in the carbonic anhydrase catalytic center. A mechanistic study. *ACS Catal.* 5, 5397–5409. doi: 10.1021/acscatal.5b00185

Prejanò, M., Marino, T., and Russo, N. (2017). How can methanol dehydrogenase from methylacidiphilum fumariolicum work with the alien Ce^{III} ion in the active center? A theoretical study. *Chem. Eur. J.*23, 8652–8657. doi: 10.1002/chem.201700381

Raj, L., Ide, T., Gurkar, A. U., Foley, M., Schenone, M., Li, X., et al. (2011). Selective killing of cancer cells by a small molecule targeting the stress response to ROS. *Nature* 475, 231–234. doi: 10.1038/nature10167

Ritacco, I., Russo, N., and Sicilia, E. (2015). DFT investigation of the mechanism of action of organoiridium(III) complexes as anticancer agents. *Inorg. Chem.* 54, 10801–10810. doi: 10.1021/acs.inorgchem.5b01832

Ruiz-Gomez, M. J., Souviron, A., Martinez-Morillo, M., and Gil, L. (2000). P-glycoprotein, glutathione and glutathione S-transferase increase in a colon carcinoma cell line by colchicine. *J. Physiol. Biochem.* 56, 307–312. doi: 10.1007/BF03179798

Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. C. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes *J. Comput. Phys.* 23, 327–341. doi: 10.1016/0021-9991(77)90098-5

Sheehan, D., Meade, G., Foley, V. M., and Dowd, C. A. (2001). Structures, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. *Biochem. J.* 360, 1–16. doi: 10.1042/bj3600001

Siegbahn, P. E. M., and Himo, F. (2011). The quantum chemical cluster approach for modeling enzyme reactions. *Wiley Interdiscip. Rev: Comput. Mol. Sci.*1, 323–336. doi: 10.1002/wcms.13

Sousa, S. F., Ribeiro, A. J., Neves, R. R. P., Brás, N. F., Cerqueira, N. M. F. S. A., et al. Ramos, M. J. (2017). Application of quantum mechanics/molecular mechanics methods in the study of enzymatic reaction mechanisms. *WIREs Comput. Mol. Sci.* 7, e1281–1311. doi: 10.1002/wcms.1281

Stoddard, E. G., Killinger, B. J., Nair, R. N., Sadler, N. C., Volk, R. F., Purvine, S. O., et al. (2017). Activity-based probes for isoenzyme- and site-specific functional characterization of glutathione S-transferases. *J. Am. Chem. Soc.* 139, 16032–16035. doi: 10.1021/jacs.7b07378

Svensson, M., Humbel, S., Froese, R. D. J., Matsubara, T., Sieber, S., and Morokuma, K. (1996). ONIOM: a multilayered integrated MO + MM method for geometry optimizations and single point energy predictions. a test for diels–alder reactions and Pt(P(t-Bu)$_3$)$_2$ + H$_2$ oxidative addition. *J. Phys. Chem.* 100, 19357–19363. doi: 10.1021/jp962071j

Tew, K. D. (2007). Redox in redux: Emergent roles for glutathione S-transferase P (GSTP) in regulation of cell signaling and S-glutathionylation. *Biochem. Pharmacol.* 73, 1257–1269. doi: 10.1016/j.bcp.2006.09.027

Townsend, D. M., and Tew, K. D. (2003). The role of glutathione-S-transferase in anti-cancer drug resistance. *Oncogene* 22, 7369–7375. doi: 10.1038/sj.onc.1206940

Vreven, T., Byun, K. S., Komáromi, I., Dapprich, S., Montgomery, J. A. Jr., Morokuma, K. et al. (2006). Combining quantum mechanics methods with molecular mechanics methods in ONIOM. *J. Chem. Theory Comput.* 2, 815–826. doi: 10.1021/ct050289g

Wang, H. B., Jin, X. L., Zheng, J. F., Wang, F., Dai, F., and Zhou, B. (2017). Developing piperlongumine-directed glutathione S-transferase inhibitors by an electrophilicity-based strategy. *Eur. J. Med. Chem.* 126, 517–525. doi: 10.1016/j.ejmech.2016.11.034

Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004). Development and testing of a general amber force field. *J. Comp. Chem.* 25, 1157–1173. doi: 10.1002/jcc.20035

Wang, T., Arifoglu, P., Ronai, Z., and Tew, K. D. (2001). Glutathione S-transferase P1–1 (GSTP1–1) inhibits c-Jun N-terminal kinase (JNK1) signaling through interaction with the C terminus. *J. Biol. Chem.* 276, 20999–21003. doi: 10.1074/jbc.M101355200

Wilce, M. C., and Parker, M. W. (1994). Structure and function of glutathione S-transferases. *Biochim. Biophys. Acta Protein Struct. Mol. Enzymol.* 1205, 1–18. doi: 10.1016/0167-4838(94)90086-8

Zhang, L., Xiao, Y., and Priddy, R. (1994). Increase in placental glutathione S-transferase in human oral epithelial dysplastic lesions and squamous cell carcinomas. *J. Oral. Pathol. Med.* 23, 75–79. doi: 10.1111/j.1600-0714.1994.tb00260.x

Zheng, Y.-J., and Ornstein, R.: L. (1997). Role of active site tyrosine in glutathione S-transferase: insights from a theoretical study on model systems. *J. Am. Chem. Soc.* 1997, 1523–1528. doi: 10.1021/ja961667h

Zou, Y., Zhao, D., Yan, C., Ji, Y., Liu, J., Xu, J., et al. (2018). Novel ligustrazine-based analogs of piperlongumine potently suppress proliferation and metastasis of colorectal cancer cells *in vitro* and *in vivo*. *J. Med. Chem.* 61, 1821–1832. doi: 10.1021/acs.jmedchem.7b01096

Computational Understanding of the Selectivities in Metalloenzymes

Wen-Jie Wei, Hui-Xia Qian, Wen-Juan Wang and Rong-Zhen Liao *

Key Laboratory of Material Chemistry for Energy Conversion and Storage, Ministry of Education, Hubei Key Laboratory of Bioinorganic Chemistry and Materia Medica, Hubei Key Laboratory of Materials Chemistry and Service Failure, School of Chemistry and Chemical Engineering, Huazhong University of Science and Technology, Wuhan, China

Metalloenzymes catalyze many different types of biological reactions with high efficiency and remarkable selectivity. The quantum chemical cluster approach and the combined quantum mechanics/molecular mechanics methods have proven very successful in the elucidation of the reaction mechanism and rationalization of selectivities in enzymes. In this review, recent progress in the computational understanding of various selectivities including chemoselectivity, regioselectivity, and stereoselectivity, in metalloenzymes, is discussed.

Keywords: QM, QM/MM, metalloenzyme, selectivity, reaction mechanism

INTRODUCTION

Enzymes play a vital role in many biological processes, such as cell growth, food metabolism, signaling, regulation, energy transduction, and genetic translation (Bugg, 2001). More than 6,000 biochemical reactions have been found to be catalyzed by enzymes, (Schomburg et al., 2017) about one-third of which are metalloenzymes (Ragsdale, 2006). A particular advantage of using enzymes for biotransformation's, is their ability of accelerating chemical reactions by 7 to 19 orders of magnitude in a mild reaction condition (Wolfenden and Snider, 2001). In addition, the enzyme catalysis is generally associated with exquisite selectivity, including chemoselectivity, regioselectivity, and stereoselectivity. However, some enzymes also accept unnatural substrates and show promiscuity, (Khersonsky and Tawfik, 2010), which has been proposed as directly connected to enzyme evolution (Jensen, 1976). An enzyme can also be engineered to accept the non-natural substrate with high efficiency and selectivity. For example, Kan et al. have applied the directed evolution methodology to expand the catalytic function of the cytochrome c, which elegantly mediates the C-Si bond formation with a broad substrate scope as well as with high chemoselectivity and enantioselectivity (Kan et al., 2016).

Understanding reaction mechanisms and selectivities in enzymes, is of fundamental and practical importance. Various experimental techniques have been developed to elucidate many aspects of these two important questions, including X-ray structure analysis, NMR, kinetic analysis, site-directed mutagenesis, isotope labeling, and spectroscopic methods. With the continuous advancement of computer power, the computational chemistry methods have been developed as a crucial complimentary method, to the current experimental methods in the study of enzyme catalysis (Martí et al., 2004; Bruice, 2006; Gao et al., 2006; Warshel et al., 2006; Dal Peraro et al., 2007; Senn and Thiel, 2007a, 2009; Ramos and Fernandes, 2008; Lonsdale et al., 2010, 2012; Siegbahn and Himo, 2011; Rovira, 2013; Blomberg et al., 2014; Merz, 2014; Swiderek et al., 2014; Brunk and Rothlisberger, 2015; Quesne et al., 2016; Sousa et al., 2017; Ahmadi et al., 2018; Cerqueira et al., 2018). From theoretical calculations, different reaction pathways can be analyzed and the structures of transition states and intermediates can be obtained, which are

very difficult when using experimental tools. A wealth of new insights can be gained, in particular, selectivities and other important mechanistic aspects, such as kinetic isotope effects (Gao and Truhlar, 2002; Truhlar et al., 2002; Hammes-Schiffer, 2006; Pu et al., 2006), nuclear tunneling (Gao and Truhlar, 2002; Truhlar et al., 2002; Hammes-Schiffer, 2006; Pu et al., 2006) and dynamic effects (Villa and Warshel, 2001; Gao and Truhlar, 2002; Truhlar et al., 2002; Antoniou et al., 2006; Hammes-Schiffer, 2006; Olsson et al., 2006; Pu et al., 2006), can be rationalized.

In general, there are two popular approaches in the modeling of enzymes. The first one is called the quantum chemical cluster method, developed by Siegbahn, Blomberg, Himo, et al. (Siegbahn and Blomberg, 1999, 2000; Himo and Siegbahn, 2003; Noodleman et al., 2004; Siegbahn and Borowski, 2006; Siegbahn and Himo, 2009, 2011; Blomberg et al., 2014; Himo, 2017). In this approach, with more than 20 years' experience, the small model of the active site is capable of capturing the key feature of the catalysis. At the beginning, cluster models of 30–50 atoms were commonly used and nowadays cluster models of more than 200 atoms are routinely handled. The second approach is the combined quantum mechanics/molecular mechanics (QM/MM) method, which was first proposed by Warshel and Levitt (1976). In this case, the whole solvated enzyme was typically used as the model. Two QM/MM protocols were commonly applied, namely electronic energy calculations from geometry optimization starting from the X-ray structure or certain snapshots from the trajectory of a classical molecular dynamics simulations, and free energy calculations from QM/MM molecular dynamics simulations (Martí et al., 2004; Bruice, 2006; Gao et al., 2006; Warshel et al., 2006; Dal Peraro et al., 2007; Senn and Thiel, 2007a, 2009; Ramos and Fernandes, 2008; Lonsdale et al., 2010, 2012; Rovira, 2013; Merz, 2014; Swiderek et al., 2014; Brunk and Rothlisberger, 2015; Quesne et al., 2016; Sousa et al., 2017; Ahmadi et al., 2018; Cerqueira et al., 2018). With a proper selection and an increase of the size of the QM region, Liao, and Thiel have shown that the cluster approach and the QM/MM geometry optimization protocol gave similar results and the same conclusion (Liao and Thiel, 2012, 2013a).

In this review, the main focus was on the selectivities of metalloenzymes, for which the elucidation of the enzymatic mechanism was a perquisite. The reproduction of the selectivity can be considered as further support for the suggested mechanism. It should be pointed out that not all computational studies from the literature can be discussed, therefore, a number of illustrative examples have been chosen for each type of metalloenzymes, covering Mn, Fe, Co, Ni, Zn, Mo, and W. Recently, de Visser presented an excellent summary of the substrate selectivity of non-heme iron dioxygenases (de Visser, 2018).

METHODS AND MODELS

In the quantum chemical cluster modeling of metalloenzymes, a model of the active site is designed, typically containing the metal ions with their first-shell ligands and some important second-shell ligands. The whole model can be treated at the highest possible level, mainly hybrid density functional methods with very large basis sets.

The protein environment has two major effects, sterically and electrostatically, which can be covered in a simple but effective manner. The steric effects imposed by the protein matrix are taken into account using the coordinate-locking scheme, in which certain atoms are fixed at their X-ray structure positions. The general experience is to select those atoms at the periphery of the model, typically the α-carbon atoms of the residues where the truncation is introduced. In some cases, one or two more hydrogen atoms attached to each α-carbon atom along the backbone are fixed to avoid unrealistic rotation of the residues. Due to the usage of the coordinate-locking procedure, the accuracy of the results from the cluster calculations may depend on the quality (or the coordinate error) of the crystal structure used. With acetylene hydratase as an example, Liao and Thiel have demonstrated that a coordinate error of 0.1 Å for the backbone atoms, corresponding to a resolution of about 2.0 Å for the crystal structures, result in tolerable errors (about 1 kcal/mol for relative energies and 0.01 Å for bond distances) for cluster models with about 100 atoms (Liao and Thiel, 2013b). Larger models with more flexibility may be needed to obtain reliable results when starting from a crystal structure with a lower resolution and larger coordinate error. For residues that participate directly in the catalyzed reaction, it is advisable to lock atoms further away than the α-carbon.

The electrostatic interactions between the model and the surroundings, are treated using continuum solvation model methods, such as the SMD solvation method (Marenich et al., 2009), typically with a dielectric constant of 4 (Blomberg et al., 2014). For most enzymatic reactions, in which the total charge of the model does not change during the reaction, the choice of the dielectric constant has been shown to be unimportant when the cluster model reaches a size of around 200 atoms. This kind of phenomenon has been observed for four different electrostatically challenging examples, namely 4-oxalocrotonate tautomerase (Sevastik and Himo, 2007), haloalcohol dehalogenase (Hopmann and Himo, 2008), histone lysine methyltransferase (Georgieva and Himo, 2010), and aspartate decarboxylase (Liao et al., 2011a). However, when the total charge of the model changes, such as in the calculations of an electron redox potential and pKa values, the solvation effects become very large and also sensitive to the choice of the dielectric constant (Liao et al., 2015, 2016). An error of 5–10 kcal/mol may be expected, which is too large to be ignored. It should be pointed out that an empirical method has been developed by Siegbahn in the study of the oxygen evolving complex in photosystem II, for which a single parameter, derived from fitting the available experimental data, was used to determine the redox potentials and pKas (Siegbahn, 2013). This approach has also been successfully applied in the study of Cytochrome c Oxidase (Blomberg and Siegbahn, 2010), nitric oxide reductase (Blomberg and Siegbahn, 2013), and nitrogenase (Siegbahn, 2017).

The most popular functional applied for metalloenzymes is the B3LYP (Becke, 1993) functional (with 20% HF exchange), for which the major improvement is the addition of the empirical D3 dispersion corrections as proposed by Grimme (termed

B3LYP-D3) (Grimme et al., 2010). Other functionals, such as B3LYP*-D3 (with 15% HF exchange; Reiher et al., 2001), TPSSh-D3 (Staroverov et al., 2003), and the M06 series (Zhao and Truhlar, 2008), can be applied to examine the sensitivity of the results to the choice of functional. Geometry optimizations are usually carried out by using a double-zeta quality basis set with effective pseudopotentials on the metals, while all electron basis sets may also be used. The final energies are then estimated by performing single-point calculations using larger basis sets. Analytic harmonic frequencies are then computed at the same level of theory as the geometry optimizations. Due to the coordinate-locking scheme used, a number of small imaginary frequencies are obtained, which makes the entropy effects difficult to predict. Therefore, only the zero-point energy corrections are included in the final energies.

In the alternative QM/MM methodology, the protein surrounding, with the water solvent and counter-ions are treated at the classical molecular mechanics level. Depending how the coupling between the QM and the MM part is performed, the method can be divided into two classes, additive methods and subtractive methods. In the additive QM/MM method, the total energy of the system is calculated as the sum of the individual energies of the QM and MM parts (two independent calculations), plus a QM/MM coupling term (Equation 1):

$$E_{QM/MM,fullsystem} = E_{QM,innerlayer} + E_{MM,outerlayer} + E_{QM/MM,coupling}$$ (1)

In the subtractive QM/MM method, different parts are subjected to independent calculations at different levels of theory, and the total energy of the system is then obtained by additions and subtractions. With a typical two-layered system as an example, the energy of the whole system is calculated at the MM level, followed by energy calculations of the inner layer at both the QM level and the MM level. The final total energy is then expressed as Equation 2:

The "our own n-layered integrated MO and MM" method (ONIOM) developed by Morokuma and co-workers is nowadays a very popular subtractive QM/MM method. (Chung et al., 2012).

The boundary problem, which occurs when covalent bonds are presented between the QM and MM parts, can be handled using two common strategies, namely, the frozen orbital approach (Thery et al., 1994) and the link atom (typically hydrogen) approach (Singh and Kollman, 1986).

$$E_{QM/MM,fullsystem} = E_{MM,fullsystem} - E_{MM,innerlayer} + E_{QM,innerlayer}$$ (2)

The interactions between the QM and MM regions are approximately categorized into two parts, namely, van der Waals interactions and electrostatic interactions. The van der Waals interactions are usually described by an empirical Lennard–Jones potential. The electrostatic interactions can be treated at three different levels of sophistication, namely, mechanical embedding, electrostatic embedding, and polarized embedding. In the mechanical-embedding scheme, the QM–MM electrostatic

interactions are treated as the MM-MM electrostatics at the classical-classical level, in which the electrostatic effect of the MM environment on the QM region is neglected and the QM density is not polarized. In the electrostatic embedding scheme, the MM point charges are incorporated as one-electron terms in the QM Hamiltonian. Therefore, the QM–MM electrostatic interactions are treated at the quantum-classical level, and the QM density is polarized by the MM environment. In the polarized embedding scheme, a flexible MM charge model is introduced, which is polarized by the QM charge distribution. The mutual polarization of both QM and MM regions can be treated using a self-consistent formulation. Electrostatic embedding is still the most popular choice due to its tradeoff between accuracy and efficiency. Due to the artificial truncation of the QM and MM regions, the potential charge transfer between these two regions cannot be handled using these schemes in the QM/MM calculations. A simple and straightforward way to minimize the error of using force fields for the MM region, is to increase the size of the QM region by proper selection and inclusion of residues around the active site, into the QM region. A simple procedure is to select residues using the distance criteria. Alternatively, the charge deletion analysis (Bash et al., 1991) has been suggested to identify those residues that have strong electrostatic effects on the relative energies along the reaction pathways (Liao and Thiel, 2012, 2013a); Two different stationary points with the largest charge relocalization can be selected for single-point calculations, in which the point charges of every residue close to the active site are set to zero. Special attention should be given to residues or groups with more than one positive/negative charge, for example, a diphosphate group (Liao and Thiel, 2012). The magnitude of the change of the relative energy, upon the removal of the MM point charges of the selected residue, can be used to quantify the electrostatic contribution of each residue. The general idea is that if the residue gives quite a large relative energy difference in the analysis, then this residue should be selected into the QM region to get reliable results (Liao and Thiel, 2012, 2013a); By using this protocol, one can systematically increase the size of the QM region to check the convergence behavior of the QM/MM energies. Various different properties of enzymes and proteins have been investigated with respect to the enlargement of the QM size, for example, NMR shielding (Flaig et al., 2012; Hartman et al., 2015), excitation energies (Isborn et al., 2012; Provorse et al., 2016; Milanese et al., 2017), barrier heights (Solt et al., 2009; Liao and Thiel, 2013a; Sadeghian et al., 2014; Jindal and Warshel, 2016; Kulik et al., 2016; Karelina and Kulik, 2017; Das et al., 2018); reaction energies (Fox et al., 2011; Hu et al., 2011, 2013; Roßbach and Ochsenfeld, 2017), and others (Karelina and Kulik, 2017; Morgenstern et al., 2017).

Two different strategies have been developed in the QM/MM calculations. One is to perform QM/MM geometry optimizations using the X-ray structure directly (single-conformation) or using selected snapshots (multi potential energy surfaces) from a classical molecular dynamics trajectory. In this case, the QM/MM calculations must ensure that all stationary points are in the same local minima along each reaction path. To solve the conformational complexity problem and also to speed up the

QM/MM calculations, an active region of about 1,000–1,500 atoms around the QM region can be selected while the outer part remains fixed during the geometry optimizations (Shaik et al., 2010). When a large conformational change is presented during the reaction, especially for substrate binding and product release, the QM/MM molecular dynamics and Monte Carlo simulations can be used, along with the standard free-energy methods, such as free-energy perturbation, umbrella sampling, and thermodynamic integration (Senn and Thiel, 2007b). The QM/MM free energy calculations, which often require sufficiently long simulation time scales to achieve convergence, are much more time-demanding than the QM/MM geometry optimization calculations. As a compromise, a relatively smaller QM region and cheaper method, such as the semi-empirical methods and DFT functionals with double-zeta quality basis sets, are commonly used. Importantly, Senn et al. have shown that the differences between QM/MM electronic energy and free energy profiles are quite small in the study of local chemical events (Senn et al., 2009).

When performing QM/MM calculations on enzymes, a considerable amount of work has to be invested into the setup of the system before the actual QM/MM calculations. First, MM parameters are needed for the whole system. These parameters may not be available for substrates, cofactors and metals, which could be restrained to their X-ray positions in order to avoid the need to develop bonded parameters. However, non-bonded parameters, in particular atomic charges, should be developed at a reasonable level. The point charges are commonly derived from QM calculations on the selected molecules. The starting X-ray structure needs to be checked and revised prior to classical molecular mechanics simulations. Missing hydrogen atoms should be added, the protonation states of all titrable residues (e.g., His, Asp, and Glu) should be assigned according to their pKa values and local hydrogen bonding networks, the orientations of amide groups of Asn and Gln, as well as imidazole group of His, should be checked and flipped if needed. For the calculations of pKas, the empirical algorithms of PROPKA (Olsson et al., 2011) is a suitable choice, but a more rigorous solution can be obtained using Poisson-Boltzmann or QM/MM methods. The enzyme is then solvated into a water box under periodic boundary conditions or a spherical water droplet with a sufficiently large radius under spherical boundary potential. In the following step, the whole system is neutralized by either adding counter ions or by (de)protonation of charged residues at the surface of the enzyme. After these initial preparations, constrained energy minimizations, and classical molecular dynamics simulations can be performed, from which a number of snapshots can be selected as the starting geometries for the following QM/MM calculations. In the QM/MM calculations, the choice of QM method is essentially the same as that in the cluster approach, and the standard biomolecular force fields, such as Charmm (MacKerell et al., 1998), Amber (Duan et al., 2003), and Gromos (Schmid et al., 2011) are often used for the MM treatment. With the development of the very efficient domain-based local pair natural orbital coupled cluster method with single, double, and perturbatively included triple excitations (DLPNO-CCSD(T)) from the Neese group (Neese et al., 2009), it is advisable to perform QM/MM single-point calculations using DLPNO-CCSD(T) with large basis sets to obtain more accurate energies (Bistoni et al., 2018).

Various sources of errors can be envisioned from the QM/MM calculations, the setup and the starting conformation of the model system, the choice of the QM region, the choice of the QM method and MM method, the treatment of the QM/MM boundary, the entropy effects, and so on. Due to the different choices of QM/MM calculations, it is likely that different research groups obtain different results, but in general the same conclusion.

For the investigation of selectivities in enzymes, an important question is to uncover the origin of the selectivity. For metalloenzymes, both the metals and the active site residues could be involved in controlling the selectivity. A number of analytic tools from calculations could be used to find the factors that dictate the selectivity, such as frontier molecular orbitals, Fukui functions, distortion/interaction analysis (Ess and Houk, 2008; Fernández and Bickelhaupt, 2014; Bickelhaupt and Houk, 2017), and residue contribution analysis (Bash et al., 1991; Karelina and Kulik, 2017).

MODELING SELECTIVITIES IN METALLOENZYMES

Mn-Dependent Enzymes

Two different groups have investigated two different Mn-dependent enzymes that showed selectivities. One is a non-redox enzyme FosA with regioselectivity and chemoselectivity (Rife et al., 2002), and the other is a redox-active enzyme QueD with regioselectivity (Gopal et al., 2005).

FosA (Rife et al., 2002) is a manganese-dependent Fosfomycin resistance protein that catalyzes the inactivation of the Fosfomycin antibiotic by the nucleophilic addition of glutathione (GSH) on the epoxide ring (**Figure 1**) with high regioselectivity, in which the attack exclusively takes place at the C1 position. Interestingly, the uncatalyzed reaction in water solution was shown to be essentially unselective (yield of 40% on C1 and 60% on C2) (Bernat et al., 1999). There is another Fosfomycin resistance protein named FosX that mediates a nucleophilic water attack on the epoxide ring (Fillgrove et al., 2003). This also raises the question as to why FosA does not catalyze the addition of a water molecule on Fosfomycin, even though the enzyme has an open active site pocket for water binding. Consequently, FosA is both regioselectivity and chemoselective (Bernat et al., 1998).

Liao and Thiel have performed QM/MM calculations with quite a large QM region of 170 atoms, to elucidate the reaction mechanism and selectivities of FosA (Liao and Thiel, 2013c). The uncatalyzed reaction was first considered using a model consisting of a Fosfomycin and methanethiol as a model substrate for GSH, and two water molecules. Importantly, the calculations showed that the attacks on both C1 and C2 take place in a concerted step with a very similar barrier, of around 30 kcal/mol, which explains the experimental observation for the uncatalyzed reaction (Bernat et al., 1999). In FosA,

FIGURE 1 | Reaction pathways considered for FosA. QM/MM (B3LYP/def2-TZVPP:Charmm) relative energies are given in kcal/mol (Liao and Thiel, 2013c).
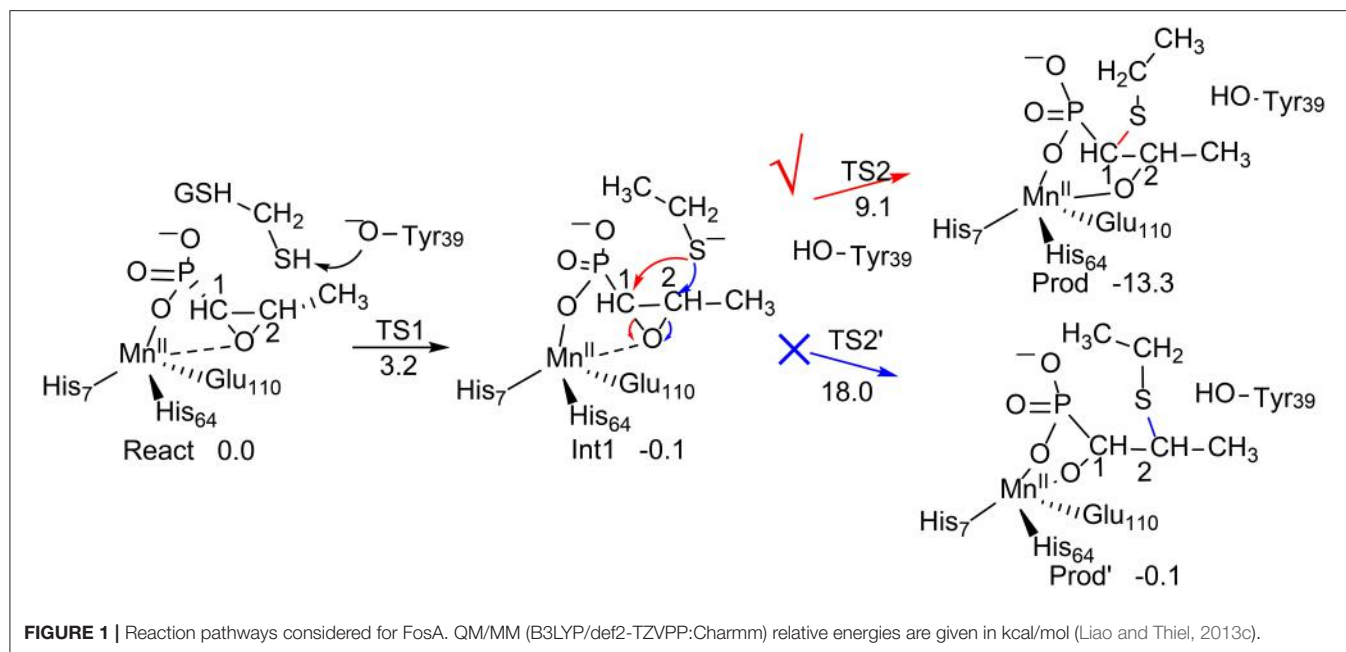
the reaction starts from a proton transfer from the GSH thiol group to a second-shell anion residue Tyr39, which is followed by the nucleophilic attack of the GSH thiolate on the epoxide, leading to the opening of the epoxide ring. The second step was suggested to be the rate-limiting step, with a barrier of 9.1 kcal/mol for the attack on C1, while it was 18.0 kcal/mol for the attack on C2. A distortion/interaction analysis (Ess and Houk, 2008; Fernández and Bickelhaupt, 2014; Bickelhaupt and Houk, 2017) was applied to understand the origin of the regioselectivity. It was shown that the distortion energy for $TS2_{C2}$ (27.1 kcal/mol) is much larger than that for $TS2_{C1}$ (20.0 kcal/mol), while the interaction energies were quite similar (−16.7 kcal/mol for $TS2_{C2}$ vs. −14.4 kcal/mol for $TS2_{C1}$). These results suggested that the regioselectivity for the GSH attack is mainly distortion-controlled.
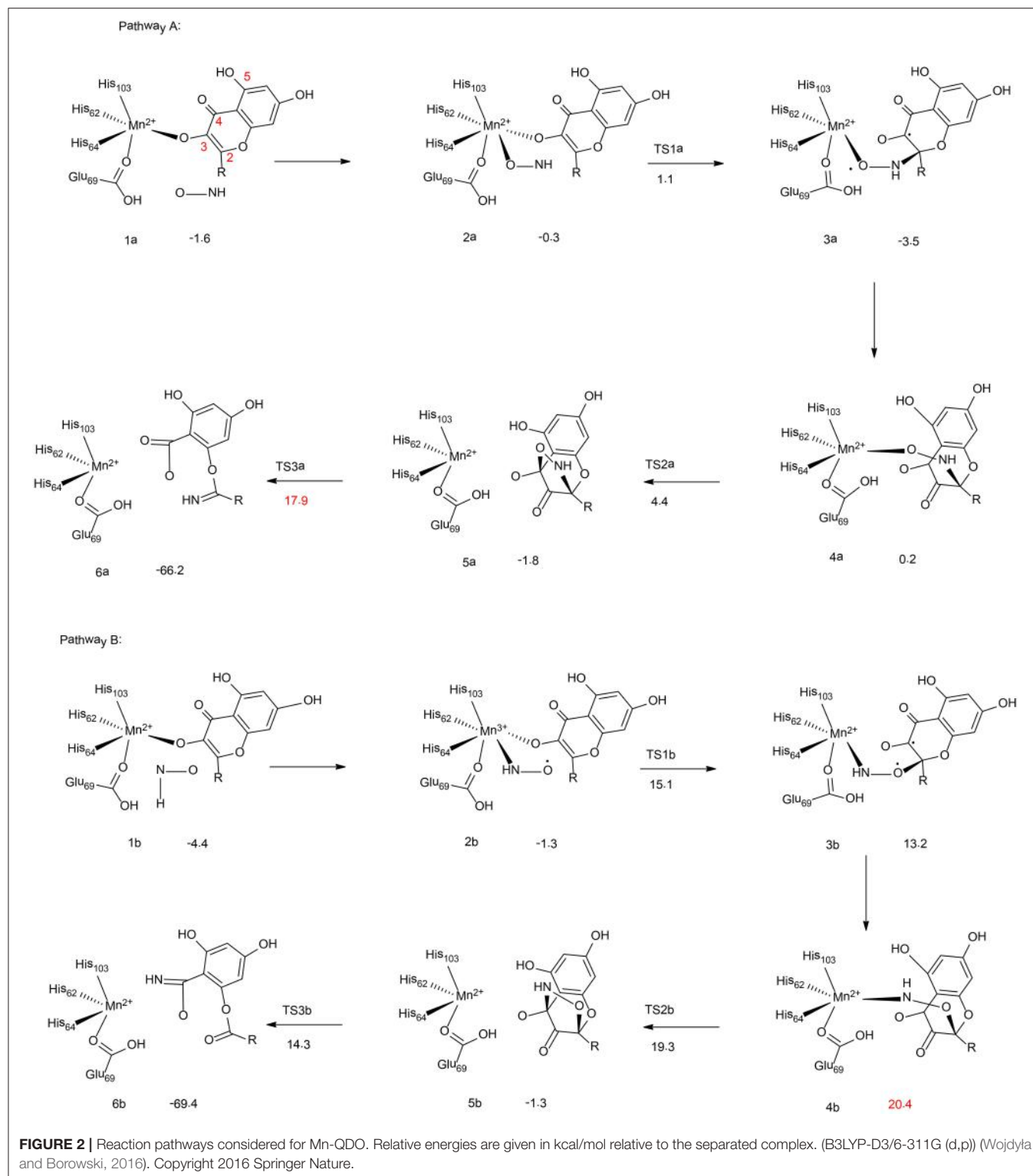
To understand the chemoselectivity, the water attack pathway was also taken into consideration. Different from the GSH attack, the deprotonation of the water substrate, the nucleophilic attack, and the ring-opening of the epoxide proceed in a single concerted step. The energy barrier was found to be 8.3 kcal/mol higher than that of the GSH attack. In addition, the attack on C2 is now preferred. However, the calculated barrier of 17.2 kcal/mol seems to be underestimated.

Quercetin 2,3-dioxygenase (QDO) from *Bacillus subtilis* is a metalloenzyme *that* is capable of using different transition metal ions (Mn2+, Co2+, Fe2+, and Cu2+) to catalyze the dioxygenation of quercetin to generate 2-protocatechuoylphloroglucinol carboxylic acid and carbon monoxide (Gopal et al., 2005). Interestingly, the Mn-QDO shows a nitroxygenase activity in the presence of HNO with high regioselectivity, and the sole product is 2-((3,4-dihydroxyphenyl)(imino)methoxy)-4,6-dihydroxybenzoate

(**Figure 2**). However, the Fe-QDO and the Co-QDO cannot catalyze the nitroxygenation reaction, implying a unique reactivity of the Mn-QDO (Kumar et al., 2011).

Wojdyla and Borowski have performed quantum chemical cluster calculations to elucidate the reactivity and regiospecificity of the Mn-QDO (Wojdyła and Borowski, 2016). The uncatalyzed reaction was first considered with a model containing a quercetin anion and a HNO molecule. The reaction started with the formation of a van der Waals complex. Then, both the nitrogen atom (pathway I) and the oxygen atom (pathway II) of the HNO molecule could perform an electrophilic attack on the C2 atom of the quercetin anion. This is followed by the formation of a five-membered ring intermediate, from which the ring cleavage takes place, coupled with the release of a CO molecule. The calculation showed that in pathway I the final step is rate-limiting step, associated with a barrier of 14.7 kcal/mol. While for pathway II, the rate-limiting step is the five-membered ring formation, with a barrier of 22.6 kcal/mol. Additionally, the Fukui function has been used to analyze the regioselectivity, and pathway I is more favored.

In the Mn-QDO enzyme, the HNO molecule could bind to the Mn^{2+} ion either via the oxygen atom (2a, **Figure 2**, pathway A) or the nitrogen atom (2b, **Figure 2**, pathway B). In pathway A, the nitrogen atom attacks the C2 atom of quercetin, (2a→3a), which is followed by a conformation change (3a→4a) and the five-membered ring formation (4a→5a). The last step is the cleavage of the five-membered ring (5a→6a), which is the rate-limiting step for the whole reaction, with a barrier of 17.9 kcal/mol. For pathway B, the barrier for the first C2-O bond formation is more than 10 kcal/mol higher than that of the first C2-N bond formation. In addition, the rate-limiting step was found to be the hydrogen bond shift process (3b→4b), with a barrier of 20.4 kcal/mol. The total barrier for pathway B is

FIGURE 2 | Reaction pathways considered for Mn-QDO. Relative energies are given in kcal/mol relative to the separated complex. (B3LYP-D3/6-311G (d,p)) (Wojdyła and Borowski, 2016). Copyright 2016 Springer Nature.

thus 2.5 kcal/mol higher than that of pathway A, which explains the regioselectivity in the Mn-QDO catalyzed nitroxygenation reaction.

To understand the metal selectivity, the substitution of Mn^{2+} by Co^{2+} and Fe^{2+} in the enzyme active site has also been taken into consideration. The calculations indicated that for both
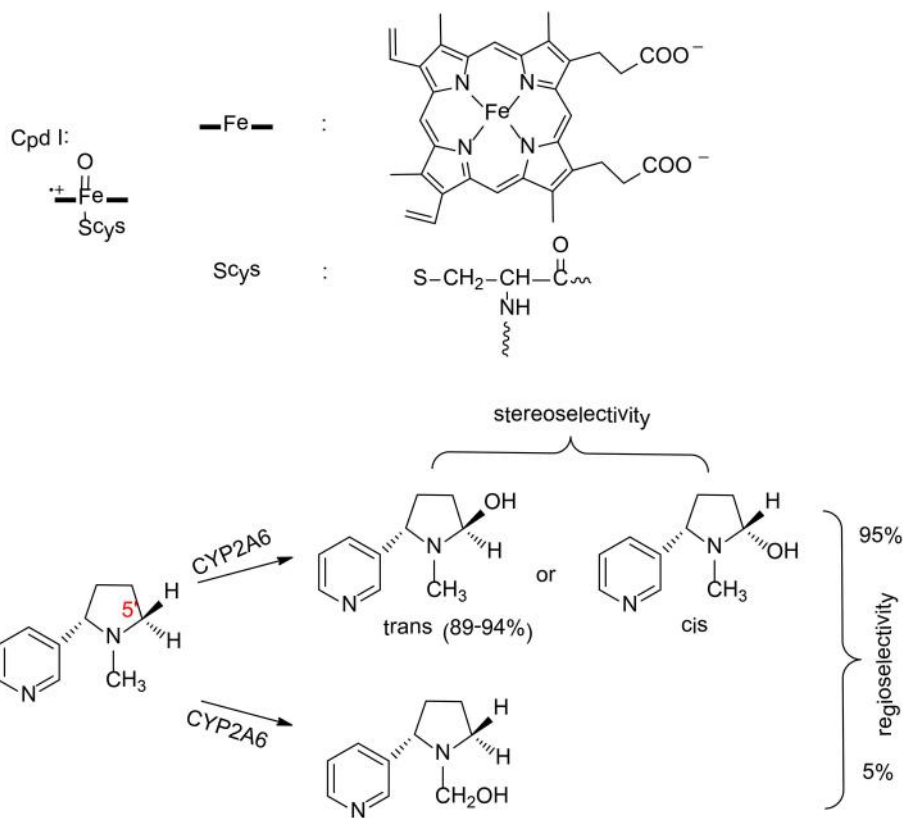
FIGURE 3 | The structure of compound I and reaction catalyzed by P450 2A6 (Li et al., 2011, 2013). Copyright 2013 Royal Society of Chemistry.

Co-QDO and Fe-QDO, the HNO prefered to be coordinated to the metal via its nitrogen atom. For both cases, the energies of the oxygen-coordinated complexes were more than 6 kcal/mol higher than those of the nitrogen-coordinated complexes. The strong preference for N-coordination over O-coordination explains that the Co-QDO and the Fe-QDO cannot catalyze the nitroxygenation reaction.

Heme-Dependent Enzymes

Cytochrome P450 is a large superfamily of enzymes that are capable of catalyzing many different types of oxidation reactions. In P450, the key part of the active site is a heme group, and the catalytic active species is a high–valent Fe^{IV}-oxo complex with a porphyrin radical cation, termed compound I (**Cpd I**, **Figure 3**). Due to the great importance of P450, extraordinary efforts have been dedicated to mechanistic studies of these enzymes using computational tools (de Visser and Shaik, 2003; Shaik et al., 2010; Li et al., 2011, 2013; Oláh et al., 2011; Krámos et al., 2012; Blomberg et al., 2014; Dubey et al., 2016; Faponle et al., 2016). Some representative examples of selectivities in P450 investigated theoretically are shown here.

P450 2A6 plays an important role in the metabolism of nicotine in humans (Yamazaki et al., 1999). Two principal pathways were observed for the oxidation of nicotine by P450 2A6, namely hydroxylation at C5′ position and at C_α position

(**Figure 3**; Jone et al., 1993). Experimentally, 95% of the product was the cotinine (C5′ hydroxylation; Murphy et al., 2005), which suggested a regioselectivity in this enzyme. In addition, C5′ hydroxylation can lead to either a *trans* or *cis* product, and the major product is the *trans*-5'-hydroxynicotine (Peterson et al., 1987), which implies a stereoselectivity in P450 2A6.

Zhan et al. have performed MD simulations and QM/MM calculations to uncover the origin of the regioselectivity and stereoselectivity of the oxidation of nicotine by P450 2A6 (Li et al., 2011, 2013). From MD simulations, six representative substrate binding models can be located. However, only three of them are suitable for the 5'-hydroxylation, which are labeled as CYP2A6-SR$_t$, CYP2A6-SR$_c$, and CYP2A6-SR$_H$. The MD simulations showed that the binding free energies of these three models are −6.86, −5.42, and −2.04 kcal/mol, respectively. When the distribution of the conformation of the free substrate in solution was taken into consideration, the CYP2A6-SR$_t$ complex is the major structure (95.4%), while 4.4% for the CYP2A6-SR$_c$ complex.

These two complexes were chosen to rationalize the regioselectivity and stereoselectivity by performing QM/MM calculations. The oxidation mechanism is a typical Cpd I mediated hydroxylation, namely C-H activation via hydrogen transfer from the substrate to the oxyl group, followed by a rebound of the hydroxyl to the substrate. For the C5′ oxidation,

FIGURE 4 | The transition state structures and the barriers (kcal/mol) in the reaction catalyzed by P450 2D6 (Oláh et al., 2011). Copyright (2011) National Academy of Sciences.
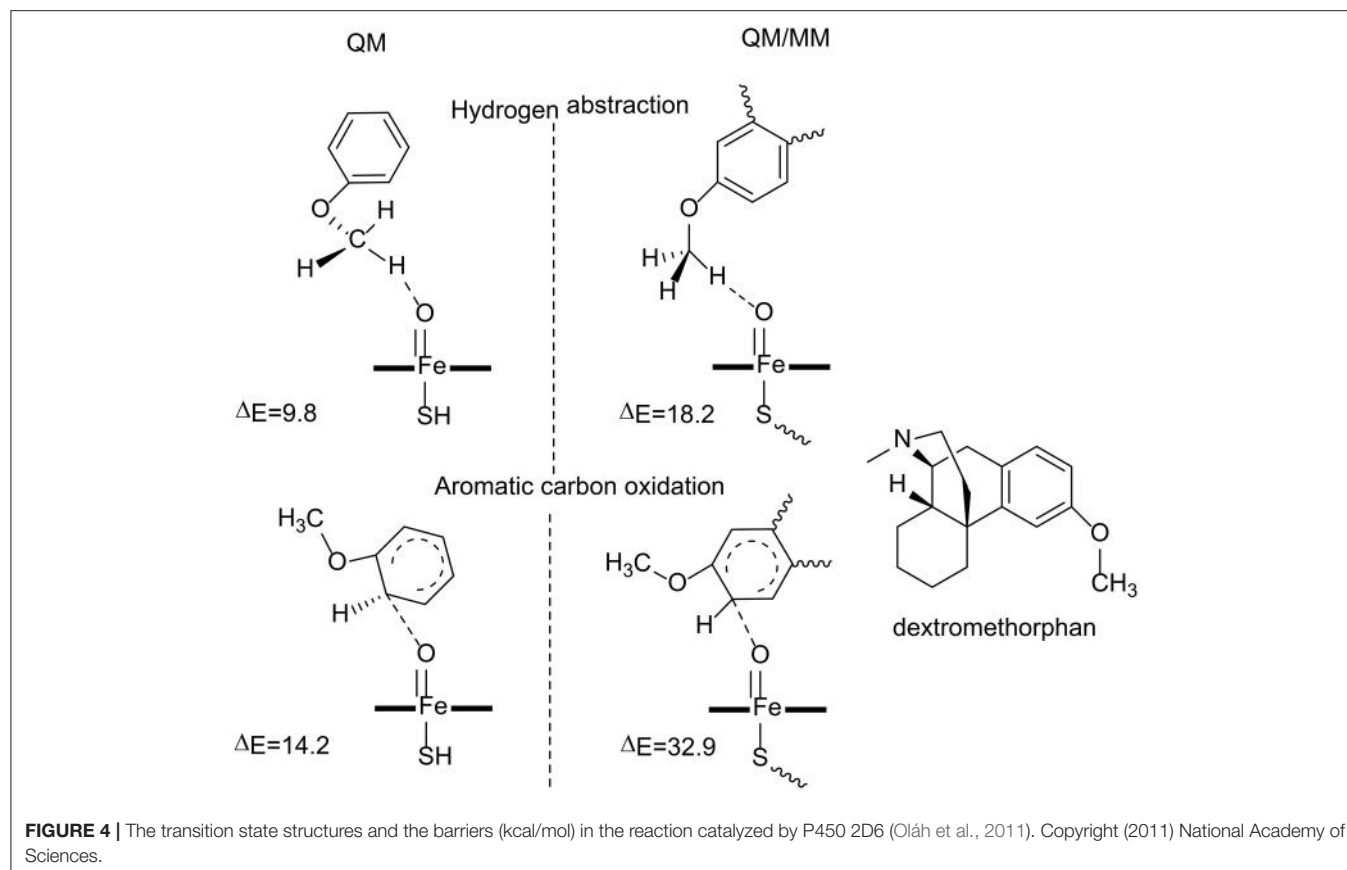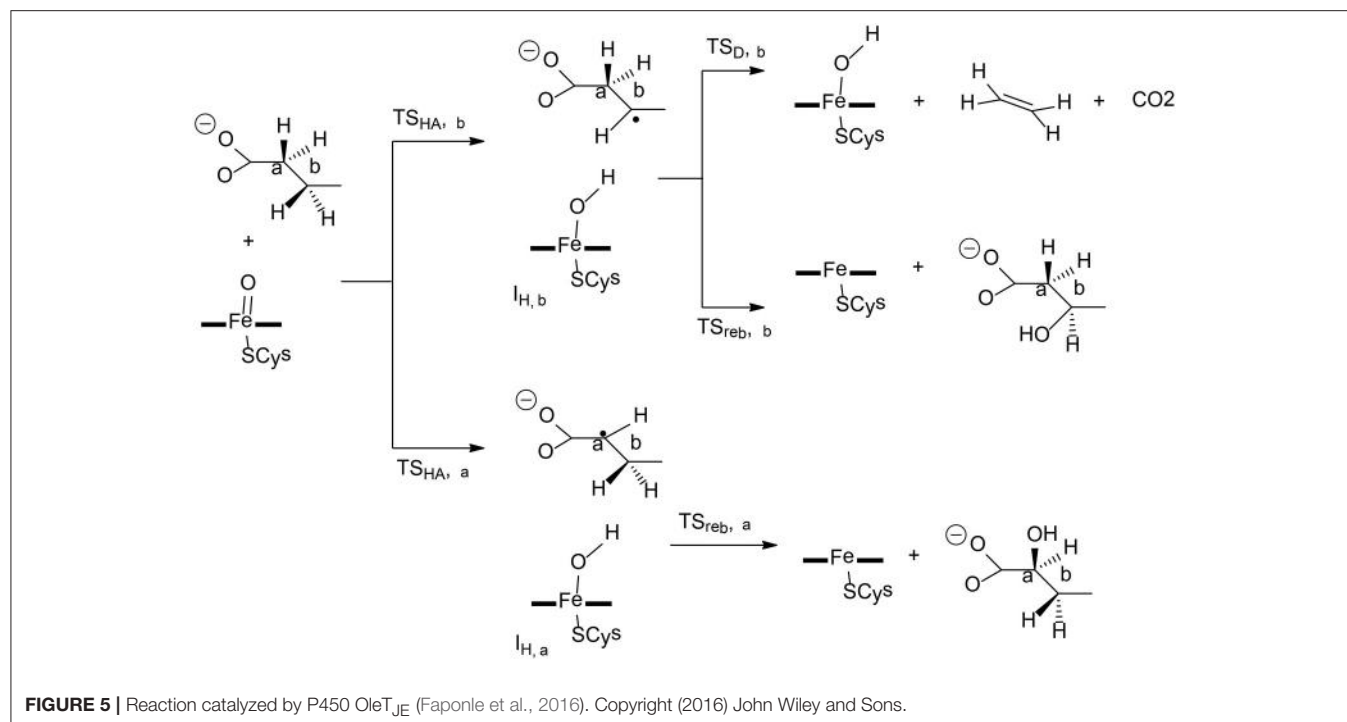
the first hydrogen abstraction is the rate-limiting step, and the barriers for the C5′-*trans* and the C5′-*cis* hydroxylations are 14.1 and 14.4 kcal/mol, respectively. Very close barriers indicate a competition between these two pathways. Thus, the free energy barrier alone cannot explain the stereoselectivity of P450 2A6. Instead, the distribution of CYP2A6-SR$_t$ and CYP2A6-SR$_c$ plays an important role in controlling the stereoselectivity. When this factor was considered, 97% of the product was the *trans*-5'hydroxylation product, which is in excellent agreement with the experimental result of 89–94% (Peterson et al., 1987). These results suggested that the stereoselectivity of P450 2A6 is primarily controlled by the substrate binding mode in the active site.

For the N-methylhydroxylation, the reaction mechanism is very similar. The first hydrogen transfer was found to be the rate–limiting step, with a barrier of 15.5 kcal/mol for the CYP2A6-SR$_t$ complex and 18.0 kcal/mol for the CYP2A6-SR$_c$ complex. On the basis of the conventional transition-state theory, the final phenomenological free energy barriers for the 5'-hydroxylation and the N-methylhydroxylation were estimated to be 14.1 kcal/mol and 15.6 kcal/mol, respectively. Consequently, the 5'-hydroxylation was kinetically more favorable, corresponding to a regioselectivity of 93%. This is very close to the experimental result of 95% (Murphy et al., 2005).

P450 2D6 is a human cytochrome P450 enzyme that metabolizes the cough suppressant drug dextromethorphan.

Two possible pathways can be envisioned for the oxidation of dextromethorphan, namely O-demethylation and aromatic carbon hydroxylation. (**Figure 4**) Interestingly, only the former reaction is observed in P450 2D6 catalyzed dextromethorphan oxidation, despite the fact that the aromatic carbon hydroxylation is a major route in the metabolism of other anisole derivatives (Ohi et al., 1992). Therefore, a chemoselectivity is presented in P450 2D6.

Oláh et al. have carried out both QM and QM/MM calculations to understand why the aromatic carbon oxidation route does not occur in the oxidation of dextromethorphan catalyzed by P450 2D6 (Oláh et al., 2011). Both reaction pathways were considered in their study (**Figure 4**). For the QM calculations, a small model was used, in which the dextromethorphan was represented by an anisole molecule and Cpd I was represented by an un-substituted heme ring with a ferryl oxo and a hydrogen sulfide ligand. In the O-demethylation pathway, the reaction started with a hydrogen atom transfer from the methyl group of the substrate to the oxygen atom of Cpd I, followed by the rebound of the hydroxyl group to the substrate. After the release of the hemiacetal species, a hydrolysis reaction took place with the production of the phenolic product. The first H-abstraction was found to be the rate-limiting step, with a barrier of only 9.8 kcal/mol. In the alternative aromatic carbon oxidation pathway, the oxyl group attacked the aromatic ring, followed by a rearrangement of the tetrahedral intermediate

FIGURE 5 | Reaction catalyzed by P450 OleT$_{JE}$ (Faponle et al., 2016). Copyright (2016) John Wiley and Sons.

to generate the phenol product. The rate-limiting step was the first C-O bond formation, with a barrier of 14.2 kcal/mol. The QM calculations suggested that both pathways are viable options despite the fact that the O-demethylation route is more preferred.

The QM calculations with a small model does not consider the protein environment in a proper manner, which seems to be very important to rationalize the chemoselectivity. QM/MM calculations were performed using the full enzyme model, and three snapshots were taken from MD simulations and considered. For the O-demethylation route, the energy barriers for the first H-abstraction step for the three snapshots were 18.2, 19.0, and 19.7 kcal/mol, respectively. While for the aromatic carbon oxidation route, the energy barriers for the first C-O bond formation step were 32.9, 33.5, 36.7 kcal/mol, respectively. The high energy barrier for the aromatic carbon oxidation route can safely rule out this pathway, which is in line with the experimental observation (Schmider et al., 1997). Compared with the QM calculations, the QM/MM calculations suggested that the interactions between the dextromethorphan substrate and some key residues (Ser304, Ala305, Val308, and Thr309) in the active site imposed constrains on the movement of the aromatic ring, which increased the energy barrier for the C-O bond formation. The enzyme environment therefore plays an important role in dictating the chemoselectivity of P450 2D6.

P450 OleT$_{JE}$ is a peroxygenase catalyzing the conversion of long-chain fatty acids to terminal olefin, which can be used for biofuels (Dennig et al., 2015; Grant et al., 2015). In the oxidation of fatty-acid, three different pathways can be located, namely, α-hydroxylation, β-hydroxylation, and decarboxylation

(**Figure 5**). Faponle et al. have performed QM/MM calculations to understand the regioselectivity and chemoselectivity in this enzyme (Faponle et al., 2016).

The first hydrogen abstraction takes place either at the C$_\alpha$ or the C$_\beta$ position of the substrate, and the calculations gave a barrier of 6.4 kcal/mol for TS$_{HA,\alpha}$ and a barrier of 6.3 kcal/mol for TS$_{HA,\beta}$. During the reaction, a substrate radical and an iron (IV)-hydroxo species are formed, which are labeled as I$_{H,\alpha}$ for α-hydroxylation and I$_{H,\beta}$ for β-hydroxylation. From I$_{H,\alpha}$, the rebound pathway to form a C-OH bond is very facile, with a barrier of only 7.3 kcal/mol. The decarboxylation from I$_{H,\alpha}$ was found to be associated with a barrier of over 40 kcal/mol, which is too high to be a viable option. However, from I$_{H,\beta}$, both the rebound process and the decarboxylation reaction can take place, and the corresponding barriers were calculated to be 9.3 and 6.7 kcal/mol, respectively. A valence bond model was used to analyze these two different pathways. It was demonstrated that electron-withdrawing substitutions at the C$_\alpha$ or C$_\beta$ position increased the barriers for the rebound process, in turn, making the decarboxylation easier.

P450 BM3 catalyzes the C-H hydroxylation of the fatty acid derivative N-palmitoylglycine (NPG) at the ω-1, ω-2, and ω-3 positions of the long chain, but not the terminal ω position, even though the terminal ω-carbon is closer to the heme iron in all of the crystal structures of NPG-bound P450 BM3 (**Figure 6**; Haines et al., 2001). Shaik et al. have used MD and QM/MM approaches to investigate the regio- and stereo-selectivity in P450 BM3 (Dubey et al., 2016). They first performed MD simulations on three different structures, namely, the open form (the substrate-free enzyme), the closed form (the substrate-bound enzyme), and the enzyme with substrate docked into the open form, to

FIGURE 6 | Structure of the NPG substrate.
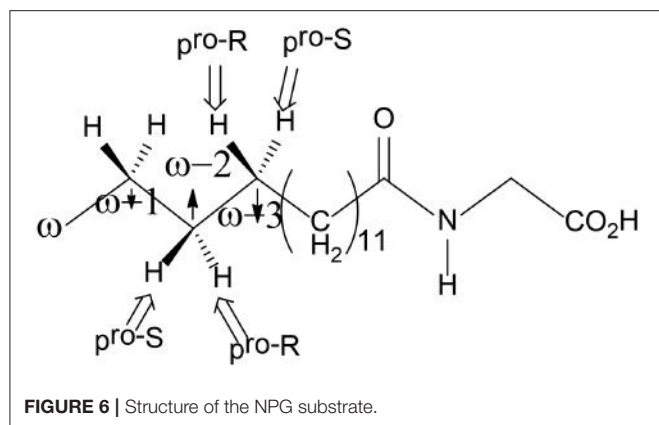
investigate the substrate binding process and to understand how the enzyme is prepared for catalysis. The MD simulations showed that during the binding of the substrate into the active site, the active site undergoes large conformational changes and some key residues play an important role in orientating the substrate to become exposed to the active oxidant.

To explain the regioselectivity, they performed MD simulations for two different states, namely the resting state and the Cpd I state. For the resting state, the calculations showed that in the beginning the ω carbon is the nearest to the heme iron, with a distance of 4–5 Å, while they are in the range of 6 to 8 Å for the ω-1, ω-2, and ω-3 carbons. After 275 ns, the terminal ω carbon becomes about 8 to 12 Å far away from the heme iron. In the case of the Cpd I state, a similar trend was observed. As a consequence, the MD simulations indicated that the favorable positions for C-H hydroxylation are the ω-1/ ω-2 /ω-3 sites, which are in good agreement with the experiment observation (Cryle and De Voss, 2006). They also performed another MD simulation, in which the Phe87 residue was replaced by a smaller Alanine residue. After 200 ns, the terminal ω position becomes the closest to Cpd I, while the ω-1, ω-2, and ω-3 positions become far away from the oxidant. These results are in line with the experiment, in which more than 90% ω hydroxylation product is formed upon Phe87Ala mutation (Dubey et al., 2016). Therefore, Phe87 imposes steric hindrance to the terminal ω-carbon and controls the regioselectivity of P450 BM3.

For the ω-1, ω-2, and ω-3 hydroxylations, the first H-abstraction is the rate-limiting step. In all three positions, either R or S product can be produced. However, the experiment showed that the R configuration was the major product (Dubey et al., 2016). For both the ω-1 and ω-2 positions, the MD simulations indicated that the pro-S C-H bond is closer to the ferryl-oxo moiety than the pro-S C-H bond, implying a preference of R product. In addition, the QM/MM calculations showed that for the ω-1 position, the barrier for the pro-S H abstraction is 6.3 kcal/mol higher than that for the pro-R H abstraction. In the case of the ω-3 position, even though the pro-S C-H bond is closer to the ferryl-oxo moiety than the pro-R C-H bond, the calculations using two different conformational basins (the major conformational basin and the minor conformational basin) showed that the barrier for the pro-R H-abstraction is lower than

that for the pro-S H-abstraction (19.5 kcal/mol for $TS_{major-R}$ vs. 26.2 kcal/mol for $TS_{major-S}$, 19.2 kcal/mol for $TS_{minor-R}$ vs. 21.7 kcal/mol for $TS_{minor-S}$). These results suggested that despite that the initial proximity of the pro-S ω-3 C-H bond, the R product is still the dominant one.
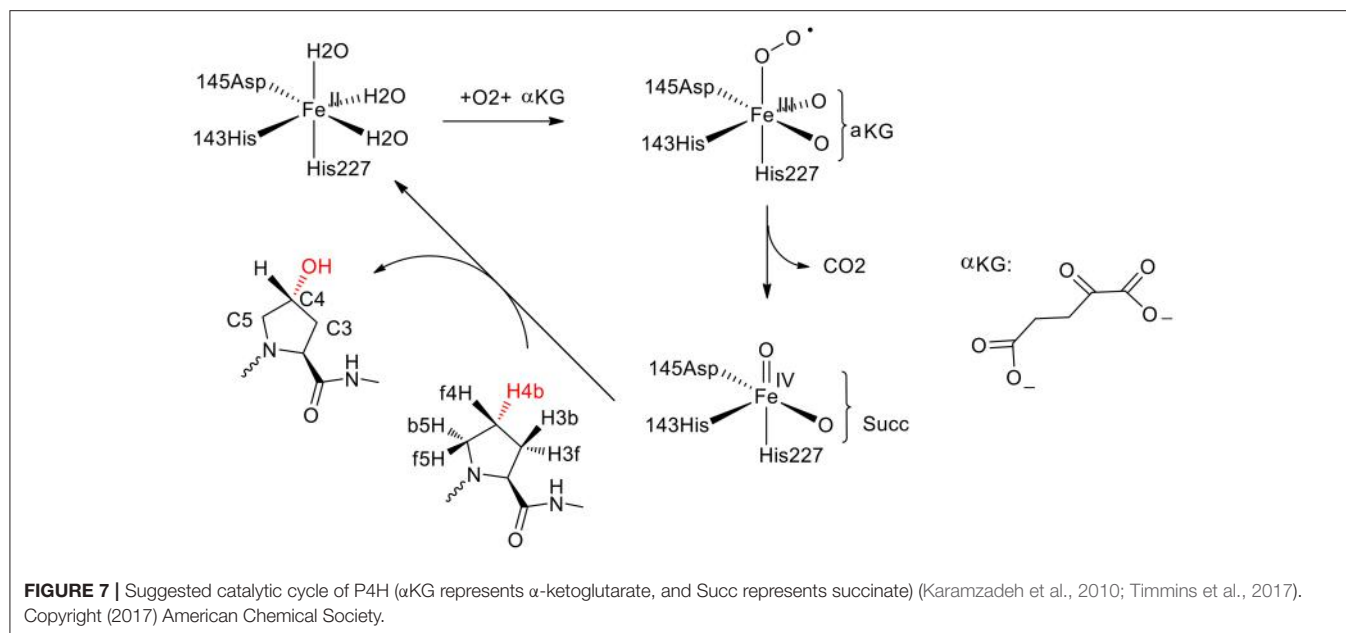
Non-heme-Dependent Enzymes

Non-heme-dependent enzymes are another superfamily of enzymes that participate in a large number of biological processes. Different from cytochrome P450, the active sites of these enzymes do not contain the porphyrin ligand. Much computational work has also been done on the selectivities of these enzymes (Karamzadeh et al., 2010; Suardíaz et al., 2013, 2014; Saura et al., 2014; Liao and Siegbahn, 2015; Christian et al., 2016; Roy and Kästner, 2017; Timmins et al., 2017; Wojdyla and Borowski, 2018), and some representative cases are shown here.

Prolyl-4-hydroxylase (P4H) is a non-heme iron hydroxylase that mediates the hydroxylation of a proline residue in a peptide chain to R-4-hydroxyproline with regioselectivity and stereoselectivity (**Figure 7**; Winter and Page, 2000). To rationalize the selectivities of this enzyme, the De Visser group has performed both the QM cluster and the QM/MM calculations (Karamzadeh et al., 2010; Timmins et al., 2017).

In the QM cluster calculations (Karamzadeh et al., 2010), quite a small model was used, consisting of the Fe with its first-shell ligands and the substrate. The calculations showed that the most thermochemical favorable hydroxylation takes place at the C5 position, which is inconsistent with the experiment observation (Winter and Page, 2000). The cluster model was then enlarged by including some important second-shell residues. It was found that the steric interactions imposed by Tyr140 and Trp243 increase the energy barriers for the hydroxylation at the C3 and C5 positions, leading to the preference of hydroxylation at the C4 position.

Recently, they performed molecular dynamics simulations and QM/MM calculations to find the origin of the regioselectivity and stereoselectivity in this enzyme (Timmins et al., 2017). Since their previous cluster calculations suggested that Tyr140 and Trp243 in the active site play an important role in the regioselectivity, both the wild-type and the mutant structures were taken into consideration. The rate-limiting step was found to be the hydrogen atom abstraction by the oxygen atom of the iron (IV)-oxo species. The hydroxylation at the C4 position ($TS_{HA, C4b}$) has the lowest energy barrier at 20.7 kcal/mol, while they are 21.7, 50.0 kcal/mol for $TS_{HA,C5b}$ and $TS_{HA,C3b}$, respectively. For the Tyr140Phe mutant, the hydrogen bond between Tyr140 and the ferryl-oxo moiety vanishes, and the barrier for the hydroxylation at C4 position becomes over 50 kcal/mol, which suggested that the mutant is not active. When Tyr140 is replaced by a Gly group, the most favorable place becomes H4f and the C5 position is also feasible. It was demonstrated that the hydrogen bond between Tyr140 and the ferryl-oxo moiety dictates the regio- and stereoselectivity of the enzyme. For the Trp243Phe mutant, the QM/MM calculations showed that the most favorable hydroxylation takes place at the H5b position. While for the Trp243Gly mutant, the energy barriers for the hydroxylation at both C4 position and C5

FIGURE 7 | Suggested catalytic cycle of P4H (αKG represents α-ketoglutarate, and Succ represents succinate) (Karamzadeh et al., 2010; Timmins et al., 2017). Copyright (2017) American Chemical Society.

position become over 30 kcal/mol. It was suggested that Trp243 along with Glu127 and Arg161 play an important role in orientating the substrate to a proper configuration for the hydroxylation.

Homoprotocatechuate 2, 3-dioxygeanse (HCPD) is a non-heme iron extradiol dioxygenase that catalyzes C-C bond cleavage and ring opening of catecholates with high regioselectivity (**Figure 8**; Kovaleva and Lipscomb, 2007).
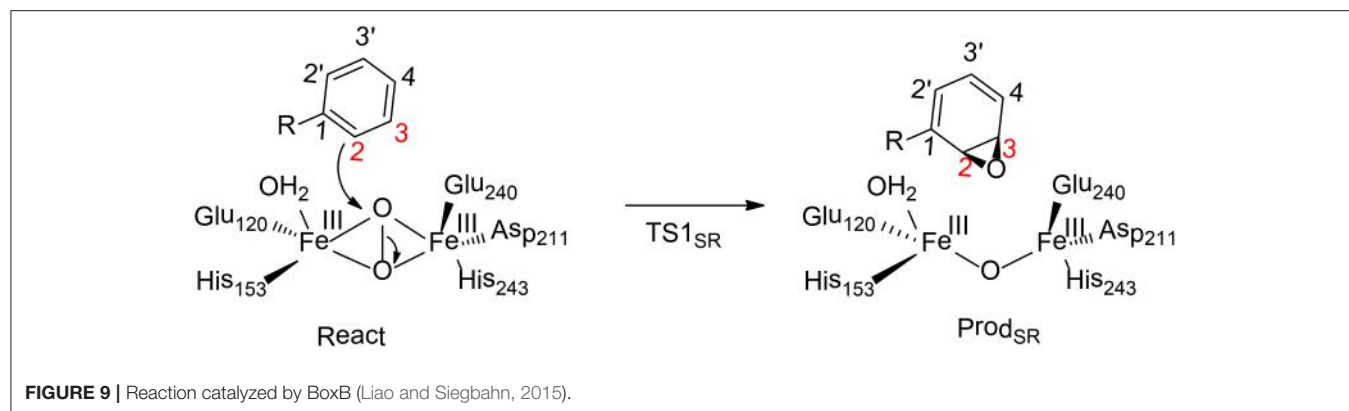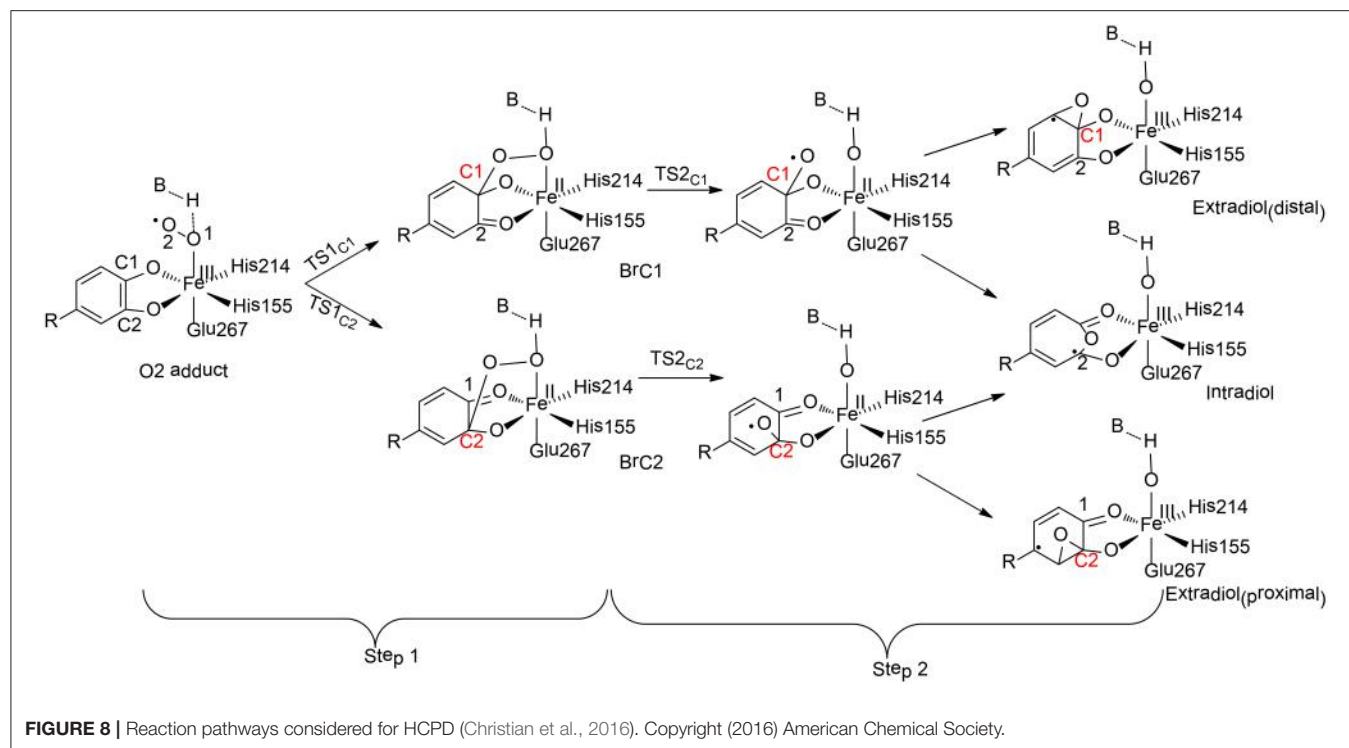
Christian and Ye have performed QM cluster calculations to elucidate the regiospecificity of HCPD (Christian et al., 2016). The calculations showed that the reaction started from the attack of the superoxide on the substrate with two possibilities (Step 1), namely at C1 and C2 positions, the barriers for which were calculated to be 31.3 kJ/mol and 28.3 kJ/mol, respectively. The resulting intermediate Br_{C1} is 28.4 kJ/mol higher in energy than Br_{C2}. These results indicated that the selectivity for Step 1 is thermodynamically controlled. To further understand various factors that control the selectivity, different types of substrates have first been tested. When the native substrate Homoprotocatechuate (HPCA) was replaced by 2, 3-dihydroxybenzoate (2, 3-DHB), the energy difference between the attack at C1 and C2 is only 3.6 kJ/mol for the uncatalyzed reaction. Then, the influence of the first-shell coordination was investigated. The calculations showed that the energy of Br_{C1} is only 1.4 kJ/mol higher than Br_{C2} when using a model only containing the first-shell ligands. Thus, it is neither the substrate itself nor the first-shell coordination that controls the selectivity. Indeed, the selectivity of Step 1 is mainly controlled by a second residue Tyr257, which lowers the energies of the electron accepting orbitals (C2=O2 π* orbitals) and facilitates the C2-O2 bond formation. In addition, His200 was also found to facilitate the C2-O2 bond formation through geometric effect.

For Step 2, the influences of both the metal center and the coordination sphere were taken into consideration. Since the formation of the C1-O bond is unfavorable, the production of the distal extradiol can be ruled out. The calculations showed that three key second-shell residues (Tyr257, His200, and His248) dictate the selectivity of the proximal extradiol formation vs. the intradiol formation. In particular, Tyr257 has the greatest electronic effect, while His200 and His248 both have steric and electronic effects, and make the extradiol pathway more favorable.

Benzoyl-CoA epoxidase (BoxB) is a dinuclear iron enzyme that catalyzes the epoxidation reaction of the aromatic ring of benzoyl-CoA (Rather et al., 2011). In principle, there are two different types of aromatic oxidation, namely hydroxylation and epoxidation. For BoxB, only the epoxidation reaction was observed experimentally, even though it is thermodynamically less favorable compared to hydroxylation. In addition, the epoxidation can take place at three possible positions, namely 1, 2-position, 2, 3-position or 3, 4-position (**Figure 9**). However, only 2, 3-epoxide was obtained. Furthermore, the epoxide product has a configuration of (2S, 3R), suggesting a stereoselectivity (Rather et al., 2010).

To clarify the mechanism and the various selectivities of BoxB, Liao, and Siegbahn have performed density functional calculations on this enzyme with quite a large cluster model of 208 atoms (Liao and Siegbahn, 2015). The calculations suggested that during the binding of the dioxygen molecule, each ferrous ion delivers an electron to the dioxygen moiety to generate a superoxide bridging the two ferric ions in a side-on symmetric fashion. The two high-spin ferric ions interact in an antiferromagnetic fashion to form a broken-symmetry singlet species. From the reactant, the cleavage of the O-O bond turned out to lead to a simultaneous attack of one of the oxygen atoms on the substrate aromatic carbon. Interestingly,
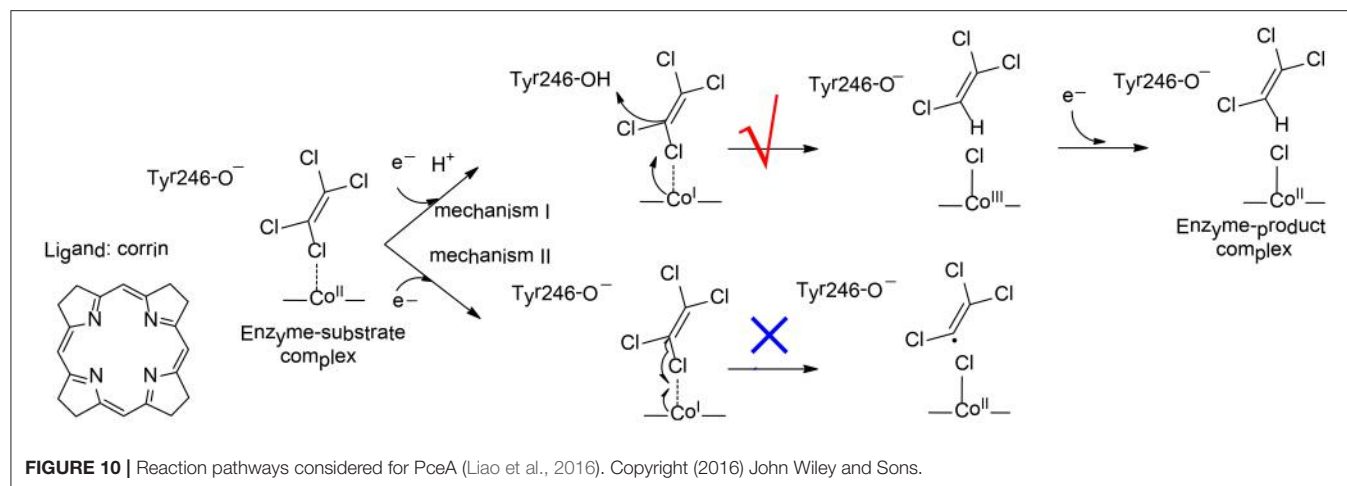
FIGURE 8 | Reaction pathways considered for HCPD (Christian et al., 2016). Copyright (2016) American Chemical Society.

FIGURE 9 | Reaction catalyzed by BoxB (Liao and Siegbahn, 2015).

only epoxidation at the 2,3-position and $2',3'$-position can be located. The energy barriers for $TS1_{2S,3R}$ and $TS1_{2R,3S}$ are 17.6 and 20.4 kcal/mol, respectively. Based on the classical transition state theory, the energy difference of 2.8 kcal/mol corresponds to a selectivity of more than 99%:1%, which agrees with the experimental observation (Rather et al., 2010). In addition, a distortion/interaction analysis (Ess and Houk, 2008; Fernández and Bickelhaupt, 2014; Bickelhaupt and Houk, 2017) was used to understand the origin of selectivity. The calculations showed that the distortion energies are quite similar (24.3 kcal/mol for $TS1_{2S,3R}$ and 24.8 kcal/mol for $TS1_{2R,3S}$), while the interaction energy for $TS1_{2S,3R}$ (18.3 kcal/mol) is somewhat smaller than $TS1_{2R,3S}$ (21.2 kcal/mol). Thus, the selectivity of BoxB was suggested to be mainly interaction-controlled.

To understand the chemoselectivity, the isomerization of epoxide to phenol was also taken into consideration. The C-O bond cleavage and deprotonation was found to proceed concertedly, associated with a barrier of 19.2 kcal/mol. This process was suggested to be slower than the release of the epoxide product from the enzyme active site.

Later, Rokob used the ONIOM (B3LYP/BP86/Amber) method to re-investigate this enzyme (Rokob, 2016). Four different pathways were considered for the aromatic ring oxidation, namely an electrophilic attached by a bis(μ-oxo)-diiron(IV) species, electrophilic attack via the σ^* orbital of a μ-η^2:η^2-peroxo-diiron(III) intermediate, radical attach via the π^*-orbital of a superoxo-diiron(II,III) species, and radical attach of a partially quenched bis(μ-oxo)-diiron(IV) intermediate (Rokob, 2016). Importantly, the most favorable pathway (barrier of 20.9

FIGURE 10 | Reaction pathways considered for PceA (Liao et al., 2016). Copyright (2016) John Wiley and Sons.

kcal/mol) was found to be very similar as those obtained by Liao and Siegbahn (Liao and Siegbahn, 2015).

Cobalt-Dependent Enzymes

PceA is a cobalamin-dependent reductive dehalogenase that catalyzes the dechlorination of perchloroethylene (PCE) to trichloroethylene (TCE), to cis-dichloroethylene (cis-DCE), and further to monochloroethylene (MCE) (Bommer et al., 2014). Three possible products can be envisioned for the dechlorination of TCE, namely, 1,1-DCE, cis-DCE, and trans-DCE. However, the cis-DCE was found to be the sole product, implying that PceA is regioselective.

To elucidate the reaction mechanism and regioselectivity of PceA, Liao et al. performed density functional calculations on this enzyme with quite a large model of 215 atoms (Liao et al., 2016). The dechlorination of PCE was first considered, and two different pathways have been analyzed (**Figure 10**). For pathway I, one electron reduction of Co^{II} to generate Co^I is coupled with the protonation of the second-shell anionic Tyr246 residue. This proton-coupled electron transfer step has a potential of -0.91 V. Then, a heterolytic C-Cl bond cleavage takes place, in concomitant with the proton transfer from Tyr246 to C1. During the reaction, the Co^I delivers two electrons to the Cl^+, leading to the formation of a Co^{III}-chloride complex. This step is rate-limiting, with a total barrier of 12.5 kcal/mol. Finally, the other one electron transfer to the metal center results in the formation of the Co^{II} product complex, and the whole reaction is exergonic by 36.0 kcal/mol. For pathway II, Tyr246 keeps anionic during the first one electron reduction. As a consequence, the following C-Cl bond cleavage is a homolytic process, and a Co^{II}-chloride substrate radical complex is produced. This step is associated with a barrier of 15.1 kcal/mol, which is slightly higher than that for pathway I. Thus, mechanism I is kinetically more favorable, which is in line with the suggested mechanism for the other reductive dehalogenase NpRdhA (Liao et al., 2015).

To rationalize the regioselectivity in the dechlorination of TCE, all six possible substrate binding modes inside the active have been considered. There are two substrate orientations for the formation of each product, cis-DCE, trans-DCE, and

1,1-DCE. The lowest energy barriers for the formation of cis-DCE, trans-DCE, and 1,1-DCE were calculated to be 13.8, 17.6, 18.4 kcal/mol, respectively. It was suggested that the amide group of cobalamin and other important second-shell residues form a pocket that favors the formation of the cis-DCE product but imposes larger steric repulsion in the formation of the other products.

The dechlorinations of cis-DCE and MCE have also been considered, and the corresponding barriers were calculated to be 20.8, 27.6 kcal/mol, respectively. The calculations showed that the energies of the HOMOs for PCE, TCE, cis-DCE, and MCE were -7.12, -7.11, -7.07, -7.15 eV, respectively, while the energies of the LUMOs were -1.11, -0.82, -0.42, -0.07 eV, respectively. During the dechlorination, an electron is transferred from Co^I to the LUMO of the substrate. Therefore, the substrate with lower LUMO was suggested to be more reactive, and with lower barrier.

Recently, Ji et al. investigated the cobalamin-mediated reductive dehalogenation reaction by DFT calculations (Ji et al., 2017). Both the inner-sphere and the out-sphere pathways have been considered. The comparison of the calculated kinetic isotope effect and the experimental kinetic isotope effect was used as a probe for diagnosing the reaction mechanism. The reaction in water solution was suggested to proceed via the formation of an organometallic intermediate with a Co-C bond (Ji et al., 2017). This is different from the case in enzyme, in which a second-shell residue was suggested to protonate the substrate during the reductive dehalogenation reaction (Liao et al., 2016).

Johannissen et al. have performed molecular docking and DFT calculations to compare the interactions between the substrate and the cobalamin in both NpRdhA and PceA (Johannissen et al., 2017). A [Co●●●X●●●R] adduct was suggested to be formed at the Co^I state, which weakens the substrate carbon-halide bond. This is evidenced by the elongation of the carbon-halide bond during the reduction of Co^{II} to Co^I, and similar results have been found by Liao et al. (2016).

Nickel-Dependent Enzymes

Wang et al. (2018) have investigated the mechanism and chemoselectivity of a nickel-dependent quercetin 2,4-dioxygenase (Ni-QueD) (Jeoung et al., 2016) by performing

QM/MM calculations at the B3LYP-D3/def2-TZVPP:Charmm level (**Figure 11**). In the dioxygenation of quercetin, two possible pathways can be envisioned, the 2,4-dioxygenolytic cleavage to harvest 2-protocatechuoylphloroglucinol carboxylic acid and carbon monoxide, and the 2,3-dioxygenolytic cleavage to produce α-keto acid. The formal pathway was observed exclusively by experiment, suggesting a chemoselectivity of this enzyme.

In their QM/MM calculations, the critical first-shell ligand Glu74 has been considered to be both in the neutral and in the ionized form. It was found that Glu74 must be deprotonated to favor the 2,4-dioxygenolytic pathway and to rationalize the chemoselectivity. The binding of a dioxygen molecule to the Ni^{II} ion results in the formation of an open-shell broken-symmetry singlet species, in which a triplet Ni^{II} is antiferromagnetically-coupled with the triplet dioxygen moiety, with partial electron transfer from the anionic quercetin substrate to the dioxygen moiety. However, the following reaction takes place preferentially in the triplet state, and a spin-crossing has to take place during the first C-O bond formation, which leads to the generation of a Ni^{II}-peroxide intermediate. Prior to the second C-O bond formation, a conformation change take place, which is required for the second C-O bond formation. Subsequently, the peroxide could attack either C3 or C4. The barrier for the attack on C4 is 3.6 kcal/mol lower than that on C3. The attack on C4 results in the formation of a pentabasic cyclic intermediate, from which the ring opening takes place with the release of a carbon monoxide molecule. This step was calculated to be rate-limiting, with a barrier of 17.4 kcal/mol. Alternatively, the opening of the four-membered ring after the attack on C3 is associated with a barrier of 30.6 kcal/mol. The calculations showed that the 2,4-dioxygenolytic pathway is much more favored than the 2,3-dioxygenolytic pathway. In addition, QM gas phase calculations and QM/MM calculations using mechanic embedding scheme also favor the 2,4-dioxygenolytic pathway significantly. The calculated barrier of 17.4 kcal/mol agrees quite well with the experimental kinetic constant of 40.1 s^{-1} (Merkens et al., 2008), which corresponds to barrier of about 15 kcal/mol.

Liu and co-workers have also investigated the reaction mechanism of this enzyme using the QM/MM method. However, only the 2,4-dioxygenolytic pathway was considered, and a similar mechanism was suggested (Li et al., 2018).

Zinc-Dependent Enzymes

Moa and Himo have used the quantum chemical cluster approach to investigate the stereoselectivity of the zinc-dependent secondary alcohol dehydrogenase (Moa and Himo, 2017). In order to rationalize the opposite enantioselectivity in the dehydrogenation of 2-butanol (R-selective) and 3-hexanol (S-selective), an active site model of more than 300 atoms were designed from the crystal structure.
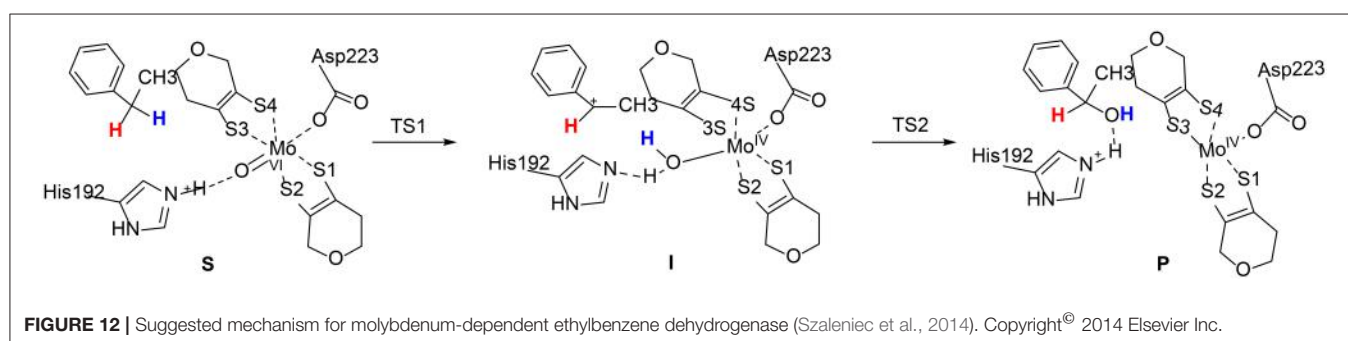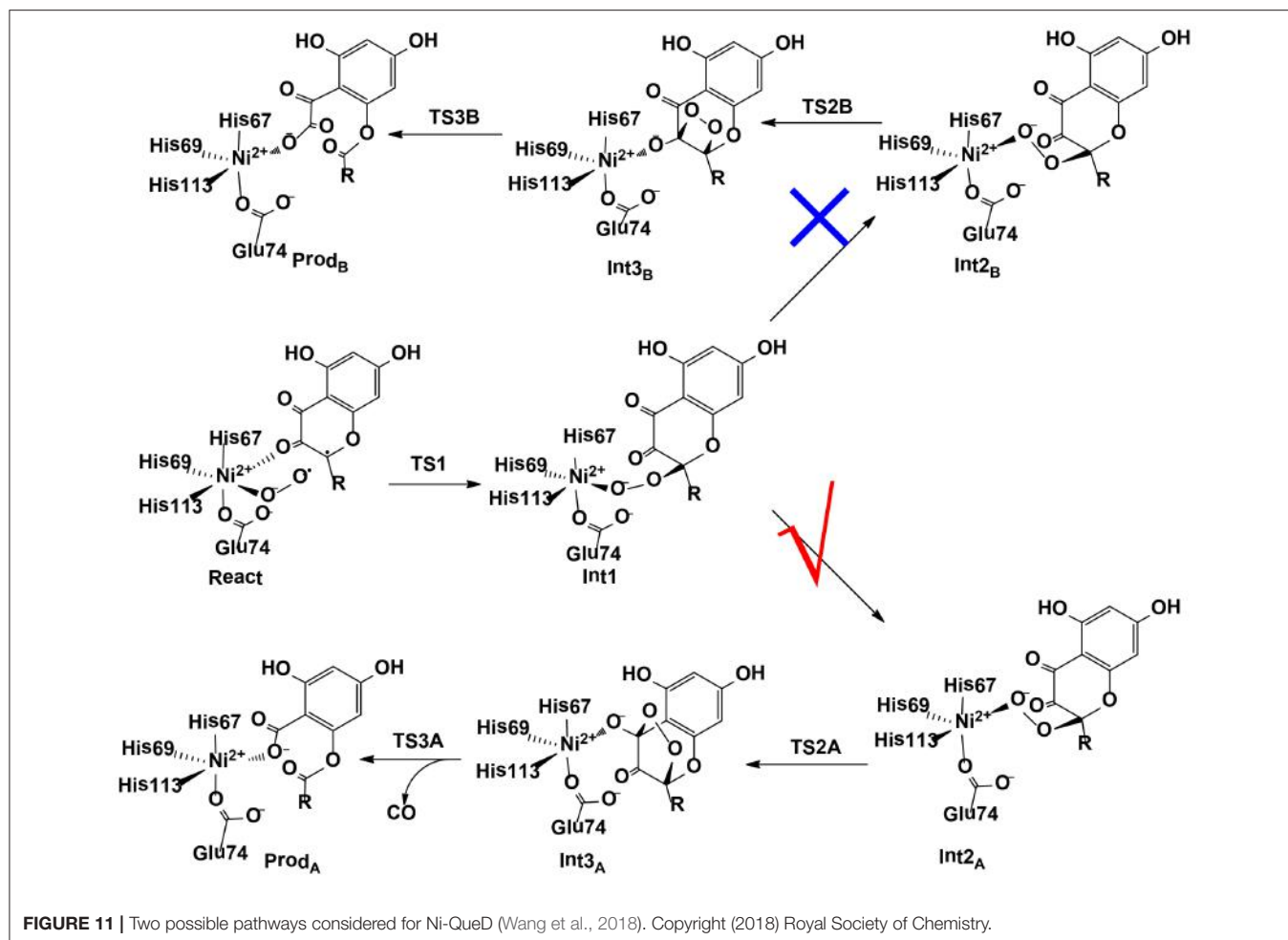
The calculations support the general mechanism as proposed for other alcohol dehydrogenase (Cui et al., 2002), in which a proton is first transferred from the substrate alcohol to a neutral histidine on the enzyme surface, followed by a hydride transfer from the substrate alkoxide to the NADP$^+$ cofactor. The zinc ion

functions as a Lewis acid to stabilize the alkoxide intermediate. For 2-butanol, two different substrate orientations have been found for both the (R)- and (S)-enantiomers. Interestingly, the orientation with lower energy is always non-productive. In addition, the small cavity in the active site prefers to bind the ethyl group for both 2-butanol and 3-hexanol. This is quite unexpected for 2-butanol, for which the large cavity prefers to bind the smaller methyl group. The energy decomposition analysis (Kitaura and Morokuma, 1976) showed that larger attractive dispersion interaction is presented when the ethyl group is in the small cavity. This also explains the preference of the (R)-TS compared with the (S)-TS (energy difference of 1.3 kcal/mol) for 2-butanol. Similar analysis have been performed for 3-hexanol, for which the (S)-TS is now preferred by 4.2 kcal/mol compared with the (R)-TS. Compared with the experimental data, the barrier difference between the (R)- and (S)-enantiomers were slightly overestimated, which was suggested to originate from the constrain of the model during the geometry optimization. More flexibility of the binding pockets by using even larger models is needed to rationalize the enantioselectivity of substrates with even larger substituents.

It should be pointed out that the reproduction of stereoselectivity in enzymes has been considered to be very challenging, as it requires accurate calculations of very small energy differences between different transition states. However, the Himo group showed that the quantum chemical cluster methodology is capable of solving this kind of important question, evidenced by four additional examples, namely limonene epoxide hydrolase (Lind and Himo, 2013), arylmalonate decarboxylase (Lind and Himo, 2014), soluble epoxide hydrolase (Lind and Himo, 2016), and phenolic acid decarboxylase (Sheng and Himo, 2017).

Molybdenum-Dependent Enzymes

Szaleniec et al. reported a QM/MM studies on the enantioselectivity of the molybdenum-dependent ethylbenzene dehydrogenase (Szaleniec et al., 2014; **Figure 12**). A two layered ONIOM method (B3LYP/Lacv3p**: Amber) was used for the calculations. A small QM region of 52 atoms (QM1) was used for the initial calculations, followed by calculations using a large QM region of 168 atoms (QM2). The oxidation of the ethylbenzene substrate was suggested to be divided into two phases, namely C-H activation and OH-rebound. The steric effects imposed by the enzyme active site enforce an almost planar conformation of ethylbenzene, with the pro(S) H-atom pointing to the Mo(VI)=O moiety. The first C-H activation step turns out to be a proton-coupled two-electron transfer process. A radical substrate is formed at the transition state (TS1) and it becomes a carbon cation at the intermediate. During the reaction, the second-shell residue His192 delivers a proton to the Mo(VI)=O moiety. In the second step, the Mo(IV)-bound water molecule performs a nucleophilic attack on the carbon cation intermediate (TS2), coupled with a proton transfer from this water to His192. The first step was calculated to be rate-limiting, with a barrier of 84.4 kJ/mol using QM2, and the second OH-rebound step was found to be barrierless.

FIGURE 11 | Two possible pathways considered for Ni-QueD (Wang et al., 2018). Copyright (2018) Royal Society of Chemistry.

FIGURE 12 | Suggested mechanism for molybdenum-dependent ethylbenzene dehydrogenase (Szaleniec et al., 2014). Copyright© 2014 Elsevier Inc.

On the basis of the proposed mechanism, the enantioselectivity was then analyzed by comparing the relative energies of all stationary points for the formation of both (S)- and (R)-phenylethanol. The energy difference between $TS1_{pro(S)}$ and $TS1_{pro(R)}$ was calculated to be 17.2 kJ/mol favoring the (S)-pathway, while the energy difference decreases to 8.6 kJ/mol for the second transition state. The interaction energies between the substrate and the nearby active site residues were used to analyze the source of the enantioselectivity. The corresponding values

were calculated to be −63.7 and −48.6 kJ/mol for $TS1_{pro(S)}$ and $TS1_{pro(R)}$, respectively.

Tungsten-Dependent Enzymes

The mechanism and selectivity of three different tungstoenzymes have been subjected to QM and QM/MM calculations, namely, acetylene hydratase (Liao et al., 2010; Liao and Himo, 2011), formaldehyde oxidoreductase (Liao et al., 2011b; Liao, 2013) and benzoyl CoA reductase (Culka et al., 2017; Qian and Liao, 2018).
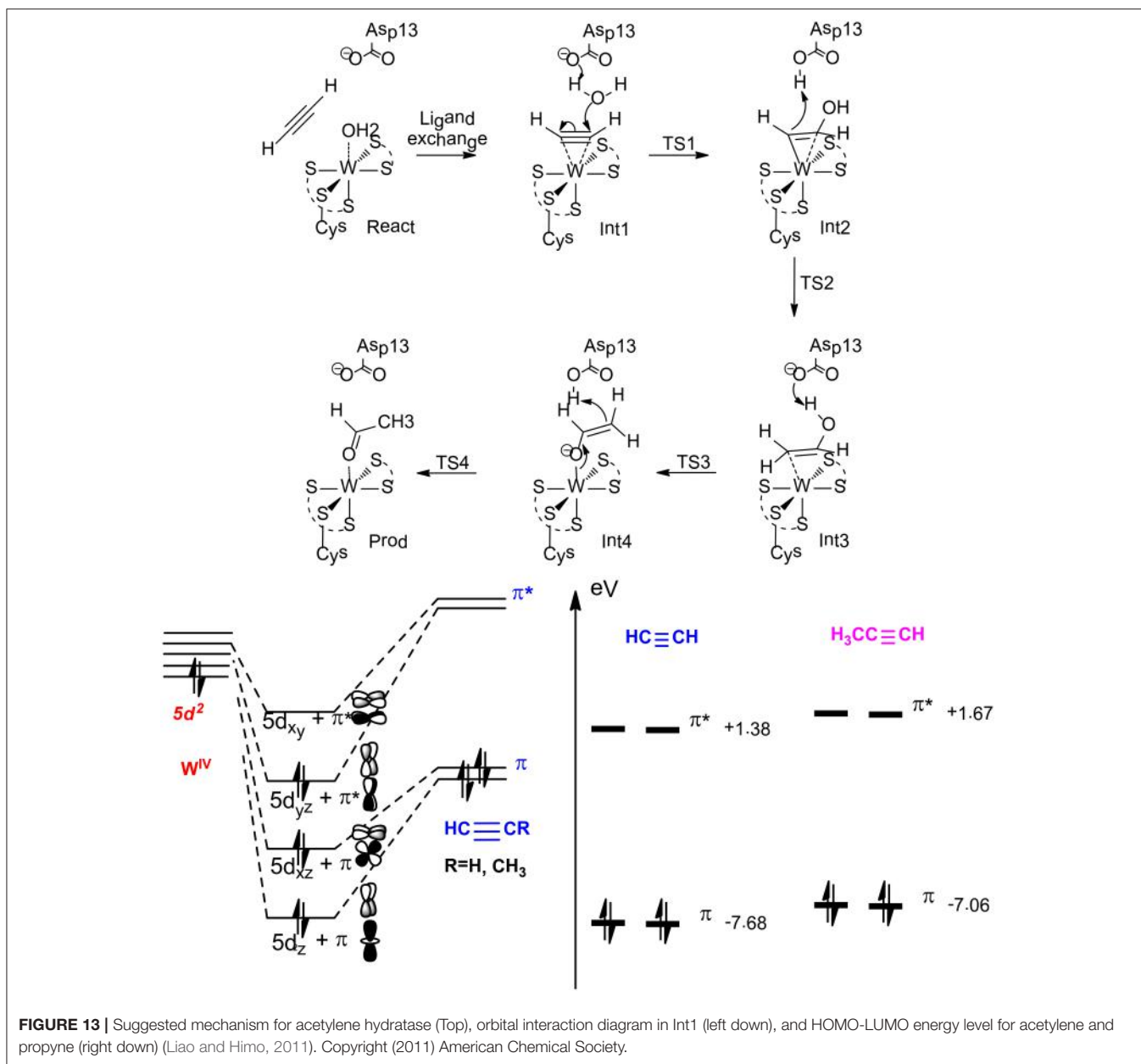
FIGURE 13 | Suggested mechanism for acetylene hydratase (Top), orbital interaction diagram in Int1 (left down), and HOMO-LUMO energy level for acetylene and propyne (right down) (Liao and Himo, 2011). Copyright (2011) American Chemical Society.
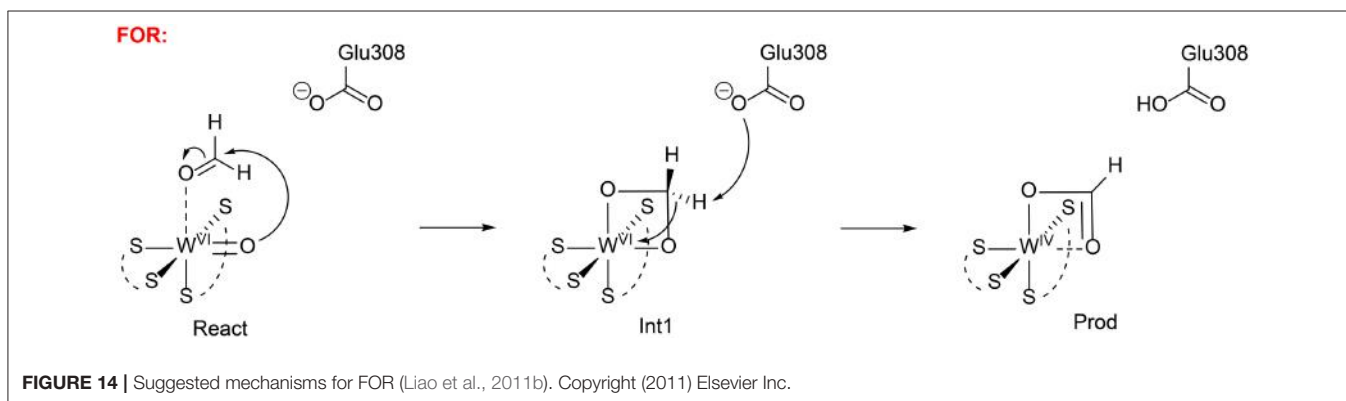
FIGURE 14 | Suggested mechanisms for FOR (Liao et al., 2011b). Copyright (2011) Elsevier Inc.

On the basis of QM calculations with a model of 116 atoms, Liao et al. (2010) suggested a first-shell mechanism that gave reasonable barrier (**Figure 13**). The reaction starts with a ligand exchange of a W(IV)-bound water molecule by the acetylene substrate. In the next step, the liberated water molecule, which gets deprotonated by the second-shell anionic Asp13 residue, performs a nucleophilic attack on the W (IV)-bound acetylene. This leads to the formation of a W(IV)-vinyl anion intermediate, which undergoes protonation by the Asp13 to produce a W(IV)-vinyl alcohol complex. This step was calculated to be rate-limiting with a barrier of 23.0 kcal/mol at the B3LYP/ LANL2TZ(f)-6-311+G(2d,2p) level. The isomerization of the vinyl alcohol to acetaldehyde catalyzed by the enzyme by two sequential proton transfer steps was found to be quite facile.

This new mechanism was then used to rationalize the chemoselectivity of this enzyme (Liao and Himo, 2011), which does not catalyze the hydration of either propyne or ethylene. The calculations showed that the ligand exchange of water by propyne is more exothermic than that by acetylene, and the following hydration has a much higher barrier for propyne, being over 30 kcal/mol. This explains that propyne is a competitive inhibitor as observed from experiment. The different reactivity of acetylene vs. propyne has been explained by analyzing the orbital interactions between the W(IV) ion and the substrate. W(IV) has an electronic configuration of $5d^2$, the two degenerate π orbitals of the substrate (acetylene and propyne) interact with two empty 5d orbitals of W(IV), and the occupied 5d orbital of W(IV) interacts with one of the two unoccupied π* orbitals via back-donation. The other unoccupied π* orbital of the substrate interacts with one of the remaining unoccupied 5d orbitals of W(IV) to generate a δ-like orbital, which facilitates the electron transfer to the empty π* orbital during the following nucleophilic attack of the water molecule on the substrate. Upon methyl substitution, the HOMO energy raises by 0.6 eV from acetylene to propyne, and this results in better π donation and larger binding energy for propyne. In addition, the LUMO energy also raises by 0.3 eV from acetylene to propyne, and this makes the electron transfer to the δ-like orbital more difficult during the nucleophilic attack. Steric repulsion has been suggested to be another plausible reason for the different reactivity of acetylene and propyne.

For the hydration of ethylene, the ligand exchange of water by ethylene becomes slightly endothermic, and the barrier for the following water attack has a barrier of more than 30 kcal/mol. The orbital interaction analysis showed that the two electron donor character of ethylene makes its π donation much weaker than that of acetylene. In addition, the back-donation of the occupied 5d orbital of W(IV) to the only unoccupied π* orbital of ethylene results in the involvement of an even higher π*-like orbital to accept the electron during the following nucleophilic attack, which is much less favorable.

Liao et al. has also investigated the mechanism of formaldehyde oxidoreductase (FOR) and the W vs. Mo selectivity for this enzyme (Liao et al., 2011b). QM calculations showed that this enzyme uses a W(VI)=O as the key oxidant and the formaldehyde coordinates to W(VI) directly via its oxygen atom. The W(VI)=O moiety then performs a nucleophilic attack on the formaldehyde carbon, generating a tetrahedral intermediate (**Figure 14**). Subsequently, a second-shell anionic residue Glu308

abstracts a proton from the tetrahedral intermediate, coupled with two-electron transfer from the intermediate to W(VI), which becomes reduced to W(IV). Other possible pathways have also been considered but were found to have much higher barriers. The suggested mechanism was then used to explain why the molybdenum substituted enzyme is not active, even though some other enzymes are able to use both molybdenum and tungsten for catalysis (Liao, 2013). The whole catalytic cycle including the formation of the active oxidant M(VI)=O (M = Mo and W) and the oxidation of formaldehyde were considered. The resting state of the enzyme is M(IV)-OH$_2$ and its oxidation to M(VI)=O involves two sequential proton-coupled electron transfer steps. For W-FOR, the energetics for the two oxidation steps can be estimated from the experimental redox potentials of the external electron acceptor $[Fe_4S_4]^{2+/+}$ (−350 mV) and the W(IV)/W(V) and W(V)/W(VI) couples. The formation of the active reactant complex is exothermic by 1.7 kcal/mol. For Mo-FOR, the relative redox potentials, which can be obtained with very good accuracy as the metal has the same environments before and after the oxidation, were calculated and used to set up the energetics. The formation of the active reactant complex now becomes endothermic by 11.7 kcal/mol. This large difference originates from the different redox potentials of W and Mo, which have also been observed experimentally for the redox potentials of pairs of molybdenum and tungsten complexes. The oxidation of formaldehyde by Mo-FOR proceeds via a similar pathway as that by W-FOR. However, due to the energetic penalty for the formation of the active Mo(VI)=O, the total barrier increases to 28.2 kcal/mol (17.6 kcal/mol for W-FOR).

Very recently, Qian and Liao performed QM/MM calculations (B3LYP-D3/def2-TZVPP:Charmm) to rationalize the regioselectivity of the benzoyl CoA reductase (Qian and Liao, 2018). They found a similar mechanism (**Figure 15**) as suggested by Culka et al. (2017) on the basis of QM/MM calculations. In the reactant complex, a water molecule is coordinated to the W(IV) ion. The reduction of the benzene ring proceeds via two sequential steps. First, the W(IV)-bound water molecule delivers a proton to the *para*-carbon (C4) of the benzene ring, in concomitant with an electron transfer from the W(IV) center to the substrate. Consequently, a W(V)-OH cyclohexadienyl radical intermediate is generated. The second step involves a proton transfer from a second-shell residue His260 to the *meta*-carbon (C3) of the cyclohexadienyl radical and an electron transfer from the pyranopterin cofactor to the substrate. The first step was calculated to be rate-limiting, with a barrier of 23.2 kcal/mol in the broken-symmetry singlet state.
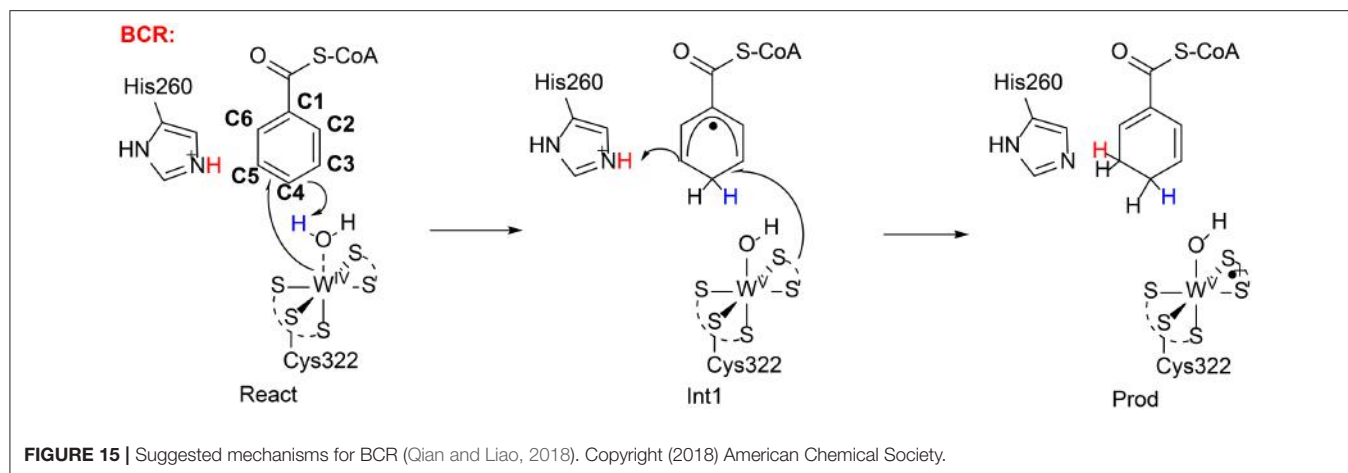
The reduction of the aromatic ring at other positions was then considered. Geometry constrains dictates that only the two other metal-carbons (C3 and C5) could accept the proton from the water molecule during the first reduction step. The barriers for these two pathways were found to be over 8 kcal/mol higher than that for the proton transfer to C4. The preference for the *para* attack was partially explained by the extra spin delocalization to the adjacent carbonyl group in the cyclohexadienyl radical intermediate, and the extra spin delocalization is absent for the *meta* attack. In addition, the substrate orientation also favors the attack on C4. For the reduction at both 2,3 and 5,6-positions, the second step turns out to be rate-limiting, with barriers of

FIGURE 15 | Suggested mechanisms for BCR (Qian and Liao, 2018). Copyright (2018) American Chemical Society.

more than 30 kcal/mol. The calculations thus reproduce the regioselectivity.

They have also investigated the substitution of tungsten by molybdenum to predict if the Mo-BCR is active or not. The reaction mechanism is generally the same, however, the barrier increases to 31.1 kcal/mol in the open-shell singlet for Mo-BCR. In addition, the reduction becomes endothermic by more than 15 kcal/mol. The main reason was suggested to originate from the different redox potentials of the M(VI)/M(V) and M(V)/M(IV) couples (M = Mo and W), and tungsten has more negative potentials than molybdenum, as found previously from both experimental and theoretical studies.

SUMMARY AND OUTLOOK

In this review, we have presented the progress of the computational modeling of selectivities in metalloenzymes. Both the quantum chemical cluster and the QM/MM approaches have shown to be successfully applied in the rationalization of selectivities and identification of factors that control the selectivity.

One of the most important questions in these enzymes, is the origin of the various selectivities, namely chemoselectivity, regioselectivity, stereoselectivity, metal oxidation preference (Fe^{2+} vs. Fe^{3+}), and metal selectivity (W vs. Mo). Mn, Fe, and Ni are typically used in the activation of dioxygen for oxidative transformations. The selectivity is typically controlled by the protein environment, and a proper description of the substrate surroundings is crucial for the rationalization of the selectivity. In Fe-dependent aldoxime dehydratase, the redox nature of the mechanism dictates the use of a lower oxidation state Fe^{2+} for electron delivering, rather than the more Lewis acidic Fe^{3+}. In the case of Mn-dependent FosA and Zn-dependent secondary alcohol dehydrogenase, both metal ions mainly function as a Lewis acid to stabilize the oxy anion during the reaction. Co is used for reductive dehalogenation, and the regioselectivity is controlled by the active site pocket. The stereoselectivity in the Mo-dependent ethylbenzene dehydrogenase is also controlled by the active site. For tungsten-dependent enzymes, the tungsten ion in acetylene hydratase enables orbital interactions with the

substrate, which illuminates the chemoselectivity. While for both formaldehyde oxidoreductase and benzoyl CoA reductase, the different oxidation potential for the W^{6+}/W^{4+} couple and the Mo^{6+}/Mo^{4+} couple, explains the metal preference in these two specific examples.

There are a number of challenging issues that need to be considered during the modeling of selectivities. First, a correction mechanism that fits all available experimental observations has to be suggested first. This is a known challenge for reactions where electrons and protons are penetrated or liberated, as the methods currently available are not accurate enough to determine the absolute pK_as and redox potentials. Second, the reproduction of selectivity, especially stereoselectivity, requires an accuracy of <1 kcal/mol. For many cases, error cancellation may be present, which makes rationalization possible. The standard DFT/MM method is a suitable choice for such applications. In other very few cases, one may need to push the accuracy limit by minimizing all possible errors and considering various issues, such as the effect of entropy and conformation, the use of high-level *ab initio* methods, the use of proper and larger QM region, the use of better force field, et al. in the QM/MM calculations. In practice, this may be very difficult to achieve due to the tradeoff between accuracy and speed. The most challenging issue is to make a prediction that can be confirmed by an experiment. This is extremely important in the field of directed evolution, which is capable of manipulating the selectivity using active site mutations. This area is likely where extraordinary efforts will be dedicated in the future.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

REFERENCES

Ahmadi, S., Herrera, L. B., Chehelamirani, M., Hostas, J., Jalife, S., and Salahub, D. R. (2018). Multiscale modeling of enzymes: QM-cluster, QM/MM, and QM/MM/MD: a tutorial review. *Int. J. Quant. Chem.* 118:e25558. doi: 10.1002/qua.25558

Antoniou, D., Basner, J., Núñez, S., and Schwartz, S. D. (2006). Computational and theoretical methods to explore the relation between enzyme dynamics and catalysis. *Chem. Rev.* 106, 3170–3187. doi: 10.1021/cr0503052

Bash, P. A., Field, M. J., Davenport, R. C., Petsko, G. A., Ringe, D., and Karplus, M. (1991). Computer simulation and analysis of the reaction pathway of triosephosphate isomerase. *Biochemistry* 30, 5826–5832. doi: 10.1021/bi00238a003

Becke, A. D. (1993). Density-functional thermochemistry. III. the role of exact exchange. *J. Chem. Phys.* 98, 5648–5652. doi: 10.1063/1.464913

Bernat, B. A., Laughlin, L. T., and Armstrong, R. N. (1998). Regiochemical and stereochemical course of the reaction catalyzed by the fosfomycin resistance protein, FosA. *J. Org. Chem.* 63, 3778–3780. doi: 10.1021/jo980014b

Bernat, B. A., Laughlin, L. T., and Armstrong, R. N. (1999). Elucidation of a monovalent cation dependence and characterization of the divalent cation binding site of the fosfomycin resistance protein (FosA). *Biochemistry* 38, 7462–7469. doi: 10.1021/bi990391y

Bickelhaupt, F. M., and Houk, K. N. (2017). Analyzing reaction rates with the distortion/interaction-activation strain model. *Angew. Chem. Int. Ed.* 56, 10070–10086. doi: 10.1002/anie.201701486

Bistoni, G., Polyak, I., Sparta, M., Thiel, W., and Neese, F. (2018). Toward accurate QM/MM reaction barriers with large QM regions using domain based pair natural orbital coupled cluster theory. *J. Chem. Theory Comput.* 14, 3524–3531. doi: 10.1021/acs.jctc.8b00348

Blomberg, M. R., Borowski, T., Himo, F., Liao, R.-Z., and Siegbahn, P. E. (2014). Quantum chemical studies of mechanisms for metalloenzymes. *Chem. Rev.* 114, 3601–3658. doi: 10.1021/cr400388t

Blomberg, M. R., and Siegbahn, P. E. (2010). Quantum chemistry as a tool in bioenergetics. *Biochim. Biophys. Acta* 1797, 129–142. doi: 10.1016/j.bbabio.2009.10.004

Blomberg, M. R. A., and Siegbahn, P. E. M. (2013). Why is the reduction of NO in cytochrome c dependent nitric oxide reductase (cNOR) not electrogenic? *Biochim. Biophys. Acta* 1827, 826–833. doi: 10.1016/j.bbabio.2013.04.005

Bommer, M., Kunze, C., Fesseler, J., Schubert, T., Diekert, G., and Dobbek, H. (2014). Structural basis for organohalide respiration. *Science* 346, 455–458. doi: 10.1126/science.1258118

Bruice, T. C. (2006). Computational approaches: reaction trajectories, structures, and atomic motions. enzyme reactions and proficiency. *Chem. Rev.* 106, 3119–3139. doi: 10.1021/cr050283j

Brunk, E., and Rothlisberger, U. (2015). Mixed quantum mechanical/molecular mechanical molecular dynamics simulations of biological systems in ground and electronically excited states. *Chem. Rev.* 115, 6217–6263. doi: 10.1021/cr500628b

Bugg, T. D. (2001). The development of mechanistic enzymology in the 20th century. *Nat. Prod. Rep.* 18, 465–493. doi: 10.1039/b009205n

Cerqueira, N. M. F. S. A., Fernandes, P. A., and Ramos, M. J. (2018). Protocol for computational enzymatic reactivity based on geometry optimisation. *ChemPhysChem* 19, 669–689. doi: 10.1002/cphc.201700339

Christian, G. J., Neese, F., and Ye, S. (2016). Unravelling the molecular origin of the regiospecificity in extradiol catechol dioxygenases. *Inorg. Chem.* 55, 3853–3864. doi: 10.1021/acs.inorgchem.5b02978

Chung, L. W., Hirao, H., Li, X., and Morokuma, K. (2012). The ONIOM method: its foundation and applications to metalloenzymes and photobiology. *WIREs Comput. Mol. Sci.* 2, 327–350. doi: 10.1002/wcms.85

Cryle, M. J., and De Voss, J. J. (2006). Is the ferric hydroperoxy species responsible for sulfur oxidation in cytochrome P450s? *Angew. Chem. Int. Ed.* 45, 8221–8223. doi: 10.1002/anie.200603411

Cui, Q., Elstner, M., and Karplus, M. (2002). A theoretical analysis of the proton and hydride transfer in liver alcohol dehydrogenase (LADH). *J. Phys. Chem. B* 106, 2721–2740. doi: 10.1021/jp013012v

Culka, M., Huwiler, S. G., Boll, M., and Ullmann, G. M. (2017). Breaking benzene aromaticity-computational insights into the mechanism of the tungsten-containing benzoyl-CoA reductase. *J. Am. Chem. Soc.* 139, 14488–14500. doi: 10.1021/jacs.7b07012

Dal Peraro, M., Ruggerone, P., Raugei, S., Gervasio, F. L., and Carloni, P. (2007). Investigating biological systems using first principles Car-Parrinello molecular dynamics simulations. *Curr. Opin. Struct. Biol.* 17, 149–156. doi: 10.1016/j.sbi.2007.03.018

Das, S., Nam, K., and Major, D. T. (2018). Rapid convergence of energy and free energy profiles with quantum mechanical size in quantum mechanical–molecular mechanical simulations of proton transfer in DNA. *J. Chem. Theory Comput.* 14, 1695–1705. doi: 10.1021/acs.jctc.7b00964

de Visser, S. P. (2018). Mechanistic insight on the activity and substrate selectivity of nonheme iron dioxygenases. *Chem. Rec.* 18, 1–17. doi: 10.1002/tcr.201800033

de Visser, S. P., and Shaik, S. (2003). A proton-shuttle mechanism mediated by the porphyrin in benzene hydroxylation by cytochrome P450 enzymes. *J. Am. Chem. Soc.* 125, 7413–7424. doi: 10.1021/ja034142f

Dennig, A., Kuhn, M., Tassoti, S., Thiessenhusen, A., Gilch, S., Bülter, T., et al. (2015). Oxidative decarboxylation of short-chain fatty acids to 1-alkenes. *Angew. Chem. Int. Ed.* 54, 8819–8822. doi: 10.1002/anie.201502925

Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W., et al. (2003). A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* 24, 1999–2012. doi: 10.1002/jcc.10349

Dubey, K. D., Wang, B., and Shaik, S. (2016). Molecular dynamics and QM/MM calculations predict the substrate-induced gating of cytochrome P450 BM3 and the regio- and stereoselectivity of fatty acid hydroxylation. *J. AM. Chem. Soc.* 138, 837–845. doi: 10.1021/jacs.5b08737

Ess, D. H., and Houk, K. N. (2008). Theory of 1,3-dipolar cycloadditions: distortion/interaction and frontier molecular orbital models. *J. Am. Chem. Soc.* 130, 10187–10198. doi: 10.1021/ja800009z

Faponle, A. S., Quesne, M. G., and de Visser, S. P. (2016). Origin of the regioselective fatty-acid hydroxylation versus decarboxylation by a cytochrome P450 peroxygenase: what drives the reaction to biofuel production? *Chem. Eur. J.* 22, 5478–5483. doi: 10.1002/chem.201600739

Fernández, I., and Bickelhaupt, F. M. (2014). The activation strain model and molecular orbital theory: understanding and designing chemical reactions. *Chem. Soc. Rev.* 43, 4953–4967. doi: 10.1039/C4CS00055B

Fillgrove, K. L., Pakhomova, S., Newcomer, M. E., and Armstrong, R. N. (2003). Mechanistic diversity of fosfomycin resistance in pathogenic microorganisms. *J. Am. Chem. Soc.* 125, 15730–15731. doi: 10.1021/ja039307z

Flaig, D., Beer, M., and Ochsenfeld, C. (2012). Convergence of electronic structure with the size of the QM region: example of QM/MM NMR shieldings. *J. Chem. Theory Comput.* 8, 2260–2271. doi: 10.1021/ct300036s

Fox, S. J., Pittock, C., Fox, T., Tautermann, C. S., Malcolm, N., and Skylaris, C.-K. (2011). Electrostatic embedding in large-scale first principles quantum mechanical calculations on biomolecules. *J. Chem. Phys.* 135:224107. doi: 10.1063/1.3665893

Gao, J., Ma, S., Major, D. T., Nam, K., Pu, J., and Truhlar, D. G. (2006). Mechanisms and free energies of enzymatic reactions. *Chem. Rev.* 106, 3188–3209. doi: 10.1021/cr050293k

Gao, J., and Truhlar, D. G. (2002). Quantum mechanical methods for enzyme kinetics. *Annu. Rev. Phys. Chem.* 53, 467–505. doi: 10.1146/annurev.physchem.53.091301.150114

Georgieva, P., and Himo, F. (2010). Quantum chemical modeling of enzymatic reactions: the case of histone lysine methyltransferase. *J. Comput. Chem.* 31, 1707–1714. doi: 10.1002/jcc.21458

Gopal, B., Madan, L. L., Betz, S. F., and Kossiakoff, A. A. (2005). The crystal structure of a quercetin 2,3-dioxygenase from *Bacillus subtilis* suggests modulation of enzyme activity by a change in the metal ion at the active site(s). *Biochemistry* 44, 193–201. doi: 10.1021/bi0484421

Grant, J. L., Hsieh, C. H., and Makris, T. M. (2015). Decarboxylation of fatty acids to terminal alkenes by cytochrome P450 compound I. *J. Am. Chem. Soc.* 137, 4940–4943. doi: 10.1021/jacs.5b01965

Grimme, S., Antony, J., Ehrlich, S., and Krieg, H. (2010). 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. doi: 10.1063/1.3382344

Haines, D. C., Tomchick, D. R., Machius, M., and Peterson, J. A. (2001). Pivotal role of water in the mechanism of P450BM-3. *Biochemistry* 40, 13456–13465. doi: 10.1021/bi011197q

Hammes-Schiffer, S. (2006). Hydrogen tunneling and protein motion in enzyme reactions. *Acc. Chem. Res.* 39, 93–100. doi: 10.1021/ar040199a

Hartman, J. D., Neubauer, T. J., Caulkins, B. G., Mueller, L. J., and Beran, G. J. O. (2015). Converging nuclear magnetic shielding calculations with respect to basis and system size in protein systems. *J. Biomol. NMR* 62, 327–340. doi: 10.1007/s10858-015-9947-2

Himo, F. (2017). Recent trends in quantum chemical modeling of enzymatic reactions. *J. Am. Chem. Soc.* 139, 6780–6786. doi: 10.1021/jacs.7b02671

Himo, F., and Siegbahn, P. E. (2003). Quantum chemical studies of radical-containing enzymes. *Chem. Rev.* 103, 2421–2456. doi: 10.1021/cr020436s

Hopmann, K. H., and Himo, F. (2008). Quantum chemical modeling of the dehalogenation reaction of haloalcohol dehalogenase. *J. Chem. Theory Comput.* 4, 1129–1137. doi: 10.1021/ct8000443

Hu, L., Söderhjelm, P., and Ryde, U. (2011). On the convergence of QM/MM energies. *J. Chem. Theory Comput.* 7, 761–777. doi: 10.1021/ct100530r

Hu, L., Söderhjelm, P., and Ryde, U. (2013). Accurate reaction energies in proteins obtained by combining QM/MM and large QM calculations. *J. Chem. Theory Comput.* 9, 640–649. doi: 10.1021/ct3005003

Isborn, C. M., Götz, A. W., Clark, M. A., Walker, R. C., and Martínez, T. J. (2012). Electronic absorption spectra from MM and ab initio QM/MM molecular dynamics: environmental effects on the absorption spectrum of photoactive yellow protein. *J. Chem. Theory Comput.* 8, 5092–5106. doi: 10.1021/ct3006826

Jensen, R. A. (1976). Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol.* 30, 409–425. doi: 10.1146/annurev.mi.30.100176.002205

Jeoung, J.-H., Nianios, D., Fetzner, S., and Dobbek, H. (2016). Quercetin 2,4-dioxygenase activates dioxygen in a side-On O₂-Ni complex. *Angew. Chem. Int. Ed.* 55, 3281–3284. doi: 10.1002/anie.201510741

Ji, L., Wang, C. C., Ji, S. C., Kepp, K. P., and Paneth, P. (2017). Mechanism of cobalamin-mediated reductive dehalogenation of chloroethylenes. *ACS Catal.* 7, 5294–5307. doi: 10.1021/acscatal.7b00540

Jindal, G., and Warshel, A. (2016). Exploring the dependence of QM/MM calculations of enzyme catalysis on the size of the QM region. *J. Phys. Chem. B* 120, 9913–9921. doi: 10.1021/acs.jpcb.6b07203

Johannissen, L. O., Leysab, D., and Hay, S. (2017). A common mechanism for coenzyme cobalamin-dependent reductive dehalogenases. *Phys. Chem. Chem. Phys.* 19, 6090–6094. doi: 10.1039/c6cp08659d

Jone, J. P., Trager, W. F., and Carlson, T. J. (1993). The binding and regioselectivity of reaction of (R)-and (S')-nicotine with cytochrome P-450cam: parallel experimental and theoretical studies. *J. Am. Chem. Soc.* 115, 381–387. doi: 10.1021/ja00055a002

Kan, S. B. J., Lewis, R. D., Chen, K., and Arnold, F. H. (2016). Directed evolution of cytochrome c for carbon–silicon bond formation: bringing silicon to life. *Science* 354, 1048–1051. doi: 10.1126/science.aah6219

Karamzadeh, B., Kumar, D., Sastry, G. N., and de Visser, S. P. (2010). Steric factors override thermodynamic driving force in regioselectivity of proline hydroxylation by Prolyl-4-hydroxylase Enzymes. *J. Phys. Chem. A* 114, 13234–13243. doi: 10.1021/jp1089855

Karelina, M., and Kulik, H. J. (2017). Systematic quantum mechanical region determination in QM/MM simulation. *J. Chem. Theory Comput.* 13, 563–576. doi: 10.1021/acs.jctc.6b01049

Khersonsky, O., and Tawfik, D. S. (2010). Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annu. Rev. Biochem.* 79, 471–505. doi: 10.1146/annurev-biochem-030409-143718

Kitaura, K., and Morokuma, k. (1976). New energy decomposition scheme for molecular interactions within the hartree-fock approximation. *Int. J. Quant. Chem.* 10, 325–340. doi: 10.1002/qua.560100211

Kovaleva, E. G., and Lipscomb, J. D. (2007). Crystal structures of Fe²⁺ dioxygenase superoxo, alkylperoxo, and bound product intermediates. *Science* 316, 453–457. doi: 10.1126/science.1134697

Krámos, B., Menyhárd, D. K., and Oláh, J. (2012). Direct hydride shift mechanism and stereoselectivity of P450nor confirmed by QM/MM calculations. *J. Phys. Chem. B* 116, 872–885. doi: 10.1021/jp2080918

Kulik, H. J., Zhang, J., Klinman, J. P., and Martínez, T. J. (2016). How large should the QM region be in QM/MM calculations? The case

of Catechol O-Methyltransferase. *J. Phys. Chem. B* 120, 11381–11394. doi: 10.1021/acs.jpcb.6b07814

Kumar, M. R., Zapata, A., Ramirez, A. J., Bowen, S. K., Francisco, W. A., and Farmer, P. J. (2011). Nitrosyl hydride (HNO) replaces dioxygen in nitroxygenase activity of manganese quercetin dioxygenase. *Proc. Natl. Acad. Sci. U.S.A.* 108, 18926–18931. doi: 10.1073/pnas.1111488108

Li, D., Huang, X., Han, K., and Zhan, C.-G. (2011). Catalytic mechanism of cytochrome P450 for 5-hydroxylation of nicotine: fundamental reaction pathways and stereoselectivity. *J. Am. Chem. Soc.* 133, 7416–7742. doi: 10.1021/ja111657j

Li, D., Huang, X., Lin, J., and Zhan, C.-G. (2013). Catalytic mechanism of cytochrome P450 for N-methylhydroxylation of nicotine: reaction pathways and regioselectivity of the enzymatic nicotine oxidation. *Dalton Trans.* 42, 3812–3820. doi: 10.1039/C2DT32106H

Li, H., Wang, X., Tian, G., and Liu, Y. G. (2018). Insights into the dioxygen activation and catalytic. *Catal. Sci. Technol.* 8, 2340–2351. doi: 10.1039/c8cy00187a

Liao, R.-Z. (2013). Why is the molybdenum-substituted tungsten-dependent formaldehyde ferredoxin oxidoreductase not active? A quantum chemical study. *J. Biol. Inorg. Chem.* 18, 175–181. doi: 10.1007/s00775-012-0961-5

Liao, R.-Z., Chen, S.-L., and Siegbahn, P. E. (2015). Which oxidation state initiates dehalogenation in the B12-dependent enzyme NpRdhA: CoII, CoI, or Co0? *ACS Catal.* 5, 7350–7358. doi: 10.1021/acscatal.5b01502

Liao, R.-Z., Chen, S.-L., and Siegbahn, P. E. M. (2016). Unraveling the mechanism and regioselectivity of the B12-dependent reductive dehalogenase PceA. *Chem. Eur. J.* 22, 12391–12399. doi: 10.1002/chem.201601575

Liao, R.-Z., and Himo, F. (2011). Theoretical study of the chemoselectivity of tungsten-dependent acetylene hydratase. *ACS Catal.* 1, 937–944. doi: 10.1021/cs200242m

Liao, R.-Z., and Siegbahn, P. E. M. (2015). Mechanism and selectivity of the dinuclear iron benzoyl-coenzyme A epoxidase BoxB. *Chem. Sci.* 6, 2754–2764. doi: 10.1039/C5SC00313J

Liao, R.-Z., and Thiel, W. (2012). Comparison of QM-only and QM/MM models for the mechanism of tungsten-dependent acetylene hydratase. *J. Chem. Theory Comput.* 8, 3793–3803. doi: 10.1021/ct3000684

Liao, R.-Z., and Thiel, W. (2013a). Convergence in the QM-only and QM/MM modeling of enzymatic reactions: a case study for acetylene hydratase. *J. Comput. Chem.* 34, 2389–2397. doi: 10.1002/jcc.23403

Liao, R.-Z., and Thiel, W. (2013b). On the effect of varying constraints in the quantum mechanics only modeling of enzymatic reactions: the case of acetylene hydratase. *J. Phys. Chem. B* 117, 3954–3961. doi: 10.1021/jp311705s

Liao, R.-Z., and Thiel, W. (2013c). Determinants of regioselectivity and chemoselectivity in fosfomycin resistance protein FosA from QM/MM calculations. *J. Phys. Chem. B.* 117, 1326–1336. doi: 10.1021/jp4002719

Liao, R.-Z., Yu, J.-G., and Himo, F. (2011a). Quantum chemical modeling of enzymatic reactions: the case of decarboxylation. *J. Chem. Theory Comput.* 7, 1494–1501. doi: 10.1021/ct200031t

Liao, R.-Z., Yu, J. G., and Himo, F. (2010). Mechanism of tungsten-dependent acetylene hydratase from quantum chemical calculations. *Proc. Natl. Acad. Sci. U.S.A.* 107, 22523–22527. doi: 10.1073/pnas.1014060108

Liao, R.-Z., Yu, J. G., and Himo, F. (2011b). Tungsten-dependent formaldehyde ferredoxin oxidoreductase: reaction mechanism from quantum chemical calculations. *J. Inorg. Biochem.* 105, 927–936. doi: 10.1016/j.jinorgbio.2011.03.020

Lind, M. E., and Himo, F. (2013). Quantum chemistry as a tool in asymmetric biocatalysis: limonene epoxide hydrolase test case. *Angew. Chem. Int. Ed.* 52, 4563–4567. doi: 10.1002/anie.201300594

Lind, M. E. S., and Himo, F. (2014). Theoretical study of reaction mechanism and stereoselectivity of arylmalonate decarboxylase. *ACS Catal.* 4, 4153–4160. doi: 10.1021/cs5009738

Lind, M. E. S., and Himo, F. (2016). Quantum chemical modeling of enantioconvergency in soluble epoxide hydrolase. *ACS Catal.* 6, 8145–8155. doi: 10.1021/acscatal.6b01562

Lonsdale, R., Harvey, J. N., and Mulholland, A. J. (2012). A practical guide to modelling enzyme-catalysed reactions. *Chem. Soc. Rev.* 41, 3025–3038. doi: 10.1039/c2cs15297e

Lonsdale, R., Ranaghan, K. E., and Mulholland, A. J. (2010). Computational enzymology. *Chem. Commun.* 46, 2354–2372. doi: 10.1039/B925647D

MacKerell, A. D. Jr., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., et al. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* 102, 3586–3616. doi: 10.1021/jp973084f

Marenich, A. V., Cramer, C. J., and Truhlar, D. G. (2009). Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B.* 113, 6378–6396. doi: 10.1021/jp810292n.

Martí, S., Roca, M., Andrés, J., Moliner, V., Silla, E., Tuñón, I., et al. (2004). Theoretical insights in enzyme catalysis. *Chem. Soc. Rev.* 33, 98–107. doi: 10.1039/B301875J

Merkens, H., Kappl, R., Jakob, R. P., Schmid, F. X., and Fetzner, S. (2008). Quercetinase QueD of Streptomyces sp. *FLA*, a monocupin dioxygenase with a preference for nickel and cobalt. *Biochemistry* 47, 12185–12196. doi: 10.1021/bi801398x

Merz, K. M. (2014). Using quantum mechanical approaches to study biological systems. *Acc. Chem. Res.* 47, 2804–2811. doi: 10.1021/ar5001023

Milanese, J. M., Provorse, M. R., Alameda, E. Jr., and Isborn, C. M. (2017). Convergence of computed aqueous absorption spectra with explicit quantum mechanical solvent. *J. Chem. Theory Comput.* 13, 2159–2171. doi: 10.1021/acs.jctc.7b00159.

Moa, S., and Himo, F. (2017). Quantum chemical study of mechanism and stereoselectivity of secondary alcohol dehydrogenase. *J. Inorg. Biochem.* 175, 259–266. doi: 10.1016/j.jinorgbio.2017.07.022

Morgenstern, A., Jaszai, M., Eberhart, M. E., and Alexandrova, A. N. (2017). Quantified electrostatic preorganization in enzymes using the geometry of the electron charge density. *Chem. Sci.* 8, 5010–5018. doi: 10.1039/c7sc01301a

Murphy, S. E., Raulinaitis, V., and Brown, K. M. (2005). Nicotine 5'-oxidation and methyl oxidation by P450 2A enzymes. *Drug Metab Dispos.* 33, 1166–1173. doi: 10.1124/dmd.105.004549

Neese, F., Wennmohs, F., and Hansen, A. (2009). Efficient and accurate local approximations to coupled-electron pair approaches: an attempt to revive the pair natural orbital method. *J. Chem. Phys.* 130:114108. doi: 10.1063/1.3086717

Noodleman, L., Lovell, T., Han, W. G., Li, J., and Himo, F. (2004). Quantum chemical studies of intermediates and reaction pathways in selected enzymes and catalytic synthetic systems. *Chem. Rev.* 104, 459–508. doi: 10.1021/cr020625a

Ohi, H., Takahara, E., Ohta, S., and Hirobe, M. (1992). Effects of oxygen concentration on the metabolism of anisole homologues by rat liver microsomes. *Xenobiotica* 22, 1329–1337. doi: 10.3109/00498259209053161

Oláh, J., Mulholland, A. J., and Harvey, J. N. (2011). Understanding the determinants of selectivity in drug metabolism through modeling of dextromethorphan oxidation by cytochrome P450. *Proc. Natl. Acad. Sci. U.S.A.* 108, 6050–6055. doi: 10.1073/pnas.1010194108

Olsson, M. H., Parson, W. W., and Warshel, A. (2006). Dynamical contributions to enzyme catalysis: critical tests of a popular hypothesis. *Chem. Rev.* 106, 1737–1756. doi: 10.1021/cr040427e

Olsson, M. H., Søndergaard, C. R., Rostkowski, M., and Jensen, J. H. (2011). PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. *J. Chem. Theory. Comput.* 7, 525–537. doi: 10.1021/ct100578z

Peterson, L. A., Trevor, A. and Castagnoli, N. Jr. (1987). Stereochemical studies on the cytochrome P-450 catalyzed oxidation of (S)-nicotine to the (S)-nicotme. DELTA. 1 '(5')-iminium Species. *J. Med. Chem.* 30, 249–254. doi: 10.1021/jm00385a004

Provorse, M. R., Peev, T., Xiong, C., and Isborn, C. M. (2016). Convergence of excitation energies in mixed quantum and classical solvent: comparison of continuum and point charge models. *J. Phys. Chem. B* 120, 12148–12159. doi: 10.1021/acs.jpcb.6b09176

Pu, J., Gao, J., and Truhlar, D. G. (2006). Multidimensional tunneling, recrossing, and the transmission coefficient for enzymatic reactions. *Chem. Rev.* 106, 3140–3169. doi: 10.1021/cr050308e

Qian, H. X., and Liao, R.-Z. (2018). QM/MM study of tungsten-dependent benzoyl-coenzyme a reductase: rationalization of regioselectivity and predication of W vs Mo selectivity. *Inorg. Chem.* 57, 10667–10678. doi: 10.1021/acs.inorgchem.8b01328

Quesne, M. G., Borowski, T., and de Visser, S. P. (2016). Quantum mechanics/molecular mechanics modeling of enzymatic processes: caveats and breakthroughs. *Chem. Eur. J.* 22, 2562–2581. doi: 10.1002/chem.201503802

Ragsdale, S. W. (2006). Metals and their scaffolds to promote difficult enzymatic reactions. *Chem. Rev.* 106, 3317–3337. doi: 10.1021/cr0503153

Ramos, M. J., and Fernandes, P. A. (2008). Computational enzymatic catalysis. *Acc. Chem. Res.* 41, 689–698. doi: 10.1021/ar7001045

Rather, L. J., Knapp, B., Haehnel, W., and Fuchs, G. (2010). Coenzyme A-dependent aerobic metabolism of benzoate via epoxide formation. *J. Biol. Chem.* 285, 20615–20624. doi: 10.1074/jbc.M110.124156

Rather, L. J., Weinert, T., Demmer, U., Bill, E., Ismail, W., Fuchs, G., et al. (2011). Structure and mechanism of the diiron benzoyl-coenzyme A epoxidase BoxB. *J. Biol. Chem.* 286, 29241–29248. doi: 10.1074/jbc.M111.236893

Reiher, M., Salomon, O., and Hess, B. A. (2001). Reparameterization of hybrid functionals based on energy differences of states of different multiplicity. *Theor. Chem. Acc.* 107, 48–55. doi: 10.1007/s00214-001-0300-3

Rife, C. L., Pharris, R. E., Newcomer, M. E., and Armstrong, R. N. (2002). Crystal structure of a genomically encoded fosfomycin resistance protein (FosA) at 1.19 Å resolution by MAD phasing off the L-III edge of Tl^+. *J. Am. Chem. Soc.* 124, 11001–11103. doi: 10.1021/ja026879v

Rokob, T. A. (2016). Pathways for arene oxidation in non-heme diiron enzymes: lessons from computational studies on benzoyl coenzyme A epoxidase. *J. Am. Chem. Soc.* 138, 14623–14638. doi: 10.1021/jacs.6b06987

Roßbach, S., and Ochsenfeld, C. (2017). Influence of coupling and embedding schemes on QM size convergence in QM/MM approaches for the example of a proton transfer in DNA. *J. Chem. Theory Comput.* 13, 1102–1107. doi: 10.1021/acs.jctc.6b00727

Rovira, C. (2013). The description of electronic processes inside proteins from Car-Parrinello molecular dynamics: chemical transformations. *WIREs Comput. Mol. Sci.* 3, 393–407. doi: 10.1002/WCMS.1153

Roy, S., and Kästner, J. (2017). Catalytic mechanism of salicylate dioxygenase: QM/MM simulations reveal the origin of unexpected regioselectivity of the ring cleavage. *Chem. Eur. J.* 23, 8949–8962. doi: 10.1002/chem.201701286

Sadeghian, K., Flaig, D., Blank, I. D., Schneider, S., Strasser, R., Stathis, D., et al. (2014). Ribose-protonated DNA base excision repair: a combined theoretical and experimental study. *Angew. Chem. Int. Ed.* 53, 10044–10048. doi: 10.1002/anie.201403334

Saura, P., Suardíaz, R., Masgrau, L., Lluch, J. M., and González-Lafont, À. (2014). Unraveling how enzymes can use bulky residues to drive site-selective c–h activation: the case of mammalian lipoxygenases catalyzing arachidonic acid oxidation. *ACS Catal.* 4, 4351–4363. doi: 10.1021/cs5006103

Schmid, N., Eichenberger, A. P., Choutko, A., Riniker, S., Winger, M., Mark, A. E., et al. (2011). Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* 40, 843–856. doi: 10.1007/s00249-011-0700-9

Schmider, J., Greenblatt, D. J., Fogelman, S. M., Von Moltke, L. L., and Shader, R. I. (1997). Metabolism of dextromethorphan *in vitro*: involvement of cytochromes P450 2D6 and 3A3/4, with a possible role of 2E1. *Biopharm. Drug Dispos.* 18, 227–240. doi: 10.1002/(SICI)1099-081X(199704)18:3<227::AID-BDD18>3.0.CO;2-L

Schomburg, I., Jeske, L., Ulbrich, M., Chang, P. A., and Schomburg, D. (2017). The BRENDA enzyme information system–From a database to an expert system. *J. Biotechnol.* 261, 194–206. doi: 10.1016/j.jbiotec.2017.04.020

Senn, H. M., Ka¨stner, J., Breidung, J., and Thiel, W. (2009). Finite-temperature effects in enzymatic reactions-Insights from QM/MM free-energy simulations. *Can. J. Chem.* 87, 1322–1337. doi: 10.1139/V09-092

Senn, H. M., and Thiel, W. (2007a). QM/MM studies of enzymes. *Curr. Opin. Chem. Biol.* 11, 182–187. doi: 10.1016/j.cbpa.2007.01.684

Senn, H. M., and Thiel, W. (2007b). QM/MM methods for biological systems. *Top. Curr. Chem.* 268, 173–290. doi: 10.1007/128_2006_084

Senn, H. M., and Thiel, W. (2009). QM/MM methods for biomolecular systems. *Angew. Chem. Int. Ed.* 48, 1198–1229. doi: 10.1002/anie.200802019

Sevastik, R., and Himo, F. (2007). Quantum chemical modeling of enzymatic reactions: the case of 4-oxalocrotonate tautomerase. *Bioorg. Chem.* 35, 444–457. doi: 10.1016/j.bioorg.2007.08.003

Shaik, S., Cohen, S., Wang, Y., Chen, H., Kumar, D., and Thiel, W. (2010). P450 enzymes: their structure, reactivity, and selectivity-modeled by QM/MM calculations. *Chem. Rev.* 110, 949–1017. doi: 10.1021/cr900121s

Sheng, X., and Himo, F. (2017). Theoretical study of enzyme promiscuity: mechanisms of hydration and carboxylation activities of phenolic acid decarboxylase. *ACS Catal.* 7, 1733–1741. doi: 10.1021/acscatal.6b03249

Siegbahn, P. E. (2013). Water oxidation mechanism in photosystem II, including oxidations, proton release pathways, O-O bond formation and O₂ release. *Biochim. Biophys. Acta* 1827, 1003–1019. doi: 10.1016/j.bbabio.2012.10.006

Siegbahn, P. E. (2017). Model calculations suggest that the central carbon in the FeMo-cofactor of nitrogenase becomes protonated in the process of nitrogen fixation. *J. Am. Chem. Soc.* 138, 10485–10495. doi: 10.1021/jacs.6b03846

Siegbahn, P. E., and Blomberg, M. R. (1999). Density functional theory of biologically relevant metal centers. *Annu. Rev. Phys. Chem.* 50, 221–249. doi: 10.1146/annurev.physchem.50.1.221

Siegbahn, P. E., and Blomberg, M. R. (2000). Transition-metal systems in biochemistry studied by high-accuracy quantum chemical methods. *Chem. Rev.* 100, 421–437. doi: 10.1021/cr980390w.

Siegbahn, P. E., and Borowski, T. (2006). Modeling enzymatic reactions involving transition metals. *Acc. Chem. Res.* 39, 729–738. doi: 10.1021/ar050123u

Siegbahn, P. E., and Himo, F. (2009). Recent developments of the quantum chemical cluster approach for modeling enzyme reactions. *J. Biol. Inorg. Chem.* 14, 643–651. doi: 10.1007/s00775-009-0511-y

Siegbahn, P. E. M., and Himo, F. (2011). The quantum chemical cluster approach for modeling enzyme reactions. *WIREs Comput. Mol. Sci.* 1, 323–336. doi: 10.1002/WCMS.13

Singh, U. C., and Kollman, P. A. (1986). A combined ab initio quantum-mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: applications to the CH3Cl + Cl⁻ exchange reaction and gas-phase protonation of polyethers. *J. Comput. Chem.* 7, 718–730. doi: 10.1002/jcc.540070604

Solt, I., Kulhánek, P., Simon, I., Winfield, S., Payne, M. C., Csányi, G., et al. (2009). Evaluating boundary dependent errors in QM/MM simulations. *J. Phys. Chem. B* 113, 5728–5735. doi: 10.1021/jp807277r

Sousa, S. F., Ribeiro, A. J. M., Neves, R. P. P., Bras, N. F., Cerqueira, N. M. F. S. A., Fernandes, P. A., et al. (2017). Application of quantum mechanics/molecular mechanics methods in the study of enzymatic reaction mechanisms. *WIREs Comput. Mol. Sci.* 7:e1281. doi: 10.1002/wcms.1281

Staroverov, V. N., Scuseria, G. E., Tao, J. M., and Perdew, J. P. (2003). Comparative assessment of a new nonempirical density functional: molecules and hydrogen-bonded complexes. *J. Chem. Phys.* 119, 12129–12137. doi: 10.1063/1.1626543

Suardíaz, R., Masgrau, L., Lluch, J. M., and González-Lafont, À. (2013). On the regio- and stereospecificity of arachidonic acid peroxidation catalyzed by mammalian 15-lypoxygenases: a combined molecular dynamics and QM/MM study. *Chemphyschem* 14, 3777–3787. doi: 10.1002/cphc.201300629

Suardíaz, R., Masgrau, L., Lluch, J. M., and González-Lafont, À. (2014). Regio- and stereospecificity in the oxygenation of arachidonic acid catalyzed by Leu597 mutants of rabbit 15-Lipoxygenase: a QM/MM study. *Chemphyschem* 15, 2303–2310. doi: 10.1002/cphc.201402045

Swiderek, K., Tunon, I., and Moliner, V. (2014). Predicting enzymatic reactivity: from theory to design. *WIREs Comput. Mol. Sci.* 4, 407–421. doi: 10.1002/WCMS.1173

Szaleniec, M., Dudzik, A., Kozik, B., Borowski, T., Heider, J., and Witko, M. (2014). Mechanistic basis for the enantioselectivity of the anaerobic hydroxylation of alkylaromatic compounds by ethylbenzene dehydrogenase. *J. Inorg. Biochem.* 139, 9–20. doi: 10.1016/j.jinorgbio.2014.05.006

Thery, V., Rinaldi, D., Rivail, J. L., Maigret, B., and Ferenczy, G. G. (1994). Quantum-mechanical computations on very large molecular systems – the local self-consistent-field method. *J. Comput. Chem.* 15, 269–282. doi: 10.1002/jcc.540150303

Timmins, A., Saint-Andr,é, M., and de Visser, S. P. (2017). Understanding how prolyl-4-hydroxylase structure steers a ferryl oxidant toward scission of a strong C–H bond. *J. Am. Chem. Soc.* 139, 9855–9866. doi: 10.1021/jacs.7b02839

Truhlar, D. G., Gao, J., Alhambra, C., Garcia-Viloca, M., Corchado, J., Sánchez, M. L., et al. (2002). The incorporation of quantum effects in enzyme kinetics modeling. *Acc. Chem. Res.* 35, 341–349. doi: 10.1021/ar0100226

Villa, J., and Warshel, A. (2001). Energetics and dynamics of enzymatic reactions. *J. Phys. Chem. B* 105, 7887–7907. doi: 10.1021/jp011048h

Wang, W.-J., Wei, W.-J., and Liao, R.-Z. (2018). Deciphering chemoselectivity of nickel-dependent quercetin 2, 4-dioxygenase. *Phys. Chem. Chem. Phys.* 20, 15784–15794. doi: 10.1039/c8cp02683a

Warshel, A., and Levitt, M. (1976). Theoretical studies of enzymatic reactions – dielectric, electrostatic, and steric stabilization of carbonium-ion in reaction of lysozyme. *J. Mol. Biol.* 103, 227–249. doi: 10.1016/0022-2836(76)90311-9

Warshel, A., Sharma, P. K., Kato, M., Xiang, Y., Lui, H., and Olsson, M. H. (2006). Electrostatic basis for enzyme catalysis. *Chem. Rev.* 106, 3210–3235. doi: 10.1021/cr0503106

Winter, A. D., and Page, A. P. (2000). Prolyl 4-hydroxylase is an essential procollagen-modifying enzyme required for exoskeleton formation and the maintenance of body shape in the nematode *Caenorhabditis elegans. Mol. Cell. Biol.* 20, 4084–4093. doi: 10.1128/mcb.20.11.4084-4093.2000

Wojdyła, Z., and Borowski, T. (2016). DFT study of the mechanism of manganese quercetin 2, 3-dioxygenase: quest for origins of enzyme unique nitroxygenase activity and regioselectivity. *J. Biol. Inorg. Chem.* 21, 475–489. doi: 10.1007/s00775-016-1356-9

Wojdyla, Z., and Borowski, T. (2018). On how the binding cavity of AsqJ dioxygenase controls the desaturation reaction regioselectivity: a QM/MM study. *J. Biol. Inorg. Chem.* 23, 795–808. doi: 10.1007/s00775-018-1575-3

Wolfenden, R., and Snider, M. J. (2001). The depth of chemical time and the power of enzymes as catalysts. *Acc. Chem. Res.* 34, 938–945. doi: 10.1021/ar000058i

Yamazaki, H., Inoue, K., Hashimoto, M., and Shimada, T. (1999). Roles of CYP2A6 and CYP2B6 in nicotine C-oxidation by human liver microsomes. *Arch. Toxicol.* 73, 65–70. doi: 10.1007/s002040050588

Zhao, Y., and Truhlar, D. G. (2008). Density functionals with broad applicability in chemistry. *Acc. Chem. Res.* 41, 157–167. doi: 10.1021/ar700111a

Check for updates

A Systematic DFT Approach for Studying Mechanisms of Redox Active Enzymes

Per E. M. Siegbahn* and Margareta R. A. Blomberg

Arrhenius Laboratory, Department of Organic Chemistry, Stockholm University, Stockholm, Sweden

When DFT has been applied to study mechanisms of redox processes a common procedure has been to study the results for many different functionals. For redox reactions involving the first row transition metals, this approach has given very different results for different functionals. The conclusion has been that DFT cannot be used for these reactions. In the meantime, results with strong predictability have been generated, most noteworthy for photosystem II, where all DFT predictions have been verified by experiments performed later. In order to obtain these predictive results using DFT, an alternative, systematic approach has been used, where the key differences between the results for different functionals can be rationalized by using a single parameter, rather than using the very large number of differences in the functionals.

I. INTRODUCTION

The results using density functional theory for molecules, have been continuously criticized the past decades. In particular, the results using b3lyp (Becke, 1993) have been shown to have severe errors in several cases. For main group elements, most of these cases were found to be corrected by the addition of dispersion corrections (Schwabe and Grimme, 2007). The cases where dispersion was needed could relatively easily be identified, and the majority of calculations made using b3lyp were not affected by dispersion in a significant way. However, molecules containing transition metals have always been considered particularly difficult to treat, and the use of DFT has continued to be criticized in these cases. DFT using different functionals were shown to give very different results for many redox reactions, in particular for molecules containing first row transition metals. The selection of different functionals for different reactions has become popular, but this approach obviously suffers from user bias, and lack of predictability.

When different functionals are compared, it is important to do this in a systematic way. During the past decade it has been possible to identify the most sensitive parameter for redox enzymes, and this is the amount of exact exchange. By identifying this single parameter, it is possible to get away from the non-systematic approach of testing a large number of functionals, which differ in many irregular ways. The best procedure found is to start with the usual b3lyp functional, which has 20% exact exchange, and decrease this percentage in steps.

A breakthrough for that approach came in the studies of photosystem II (Siegbahn, 2009, 2013). The catalyst is an Mn_4Ca complex, which shows most of the typical problems of using DFT. For example, the Mn(III) to Mn(IV) redox energy is very sensitive to the fraction of exact exchange used in hybrid methods like b3lyp. For every change of one percent of the exact exchange, the redox energy changes by one kcal/mol (Siegbahn and Blomberg, 2014). This means that for b3lyp using

10% exact exchange, the redox potential may change by 10 kcal/mol compared to the one using the usual 20%. Since there are three oxidations of Mn(III) in the catalytic cycle the difference becomes 30 kcal/mol from the start to the end. Still, it was possible to make new and certain predictions for water oxidation in PSII, that were confirmed in detail by experiments performed later.

II. COMPUTATIONAL DETAILS

The general computational approach used here has been applied and evaluated for a large number of metallo-enzymes (Blomberg et al., 2014). Density functional calculations are performed on large cluster models of the active sites (170–300 atoms). The active sites in the enzymes discussed here have two or more transition metal ions, and the models include these atoms, plus the first and sometimes also the second shell ligands. As mentioned in the introduction, the b3lyp functional (Becke, 1993) with a varying fraction of exact exchange is used in the present study. A strong argument for using b3lyp as a reference functional, is that this is the hybrid DFT functional with the smallest number of parameters. In fact, the sensitivity of the results for b3lyp is essentially dependent on only one parameter, the amount of exact exchange. Geometries are optimized using a double zeta basis with polarization functions on all second row atoms, and with a few atoms close to the truncations fixed to the X-ray coordinates. The reason not to use a larger basis set for the geometry optimization is based on a large amount of experience gathered during the past three decades, see for example Siegbahn (2001). In fact, an even smaller basis set would probably be accurate enough. More accurate energies for the optimized structures are obtained from single point calculations using a larger basis set, the lacv3p+ basis for the metal ions (Jaguar, 2009), and the large cc-pvtz(-f) basis set for the rest of the atoms. In the recommended use of b3lyp, the fraction of exact exchange has generally been set to 15% (Reiher et al., 2001) in previous studies. Empirical dispersion corrections according to Grimme (Schwabe and Grimme, 2007; Grimme et al., 2010), and solvent effects from the surrounding protein using the self-consistent reaction field (SCRF) approach are included in the energetic results described below. When H_2 is released and N_2 becomes bound in nitrogenase, the gain (loss) of translational entropy of about 10 kcal/mol is also included. The inclusion of entropy is very important in these cases but not elsewhere. Zero-point corrections are taken from the Hessians, calculated at the same level as the geometry optimizations. For nitrogenase, these effects were taken from smaller models. The Jaguar 7.9 program (Jaguar, 2009) is used for the nitrogenase geometry optimization and for all the calculations with the larger basis set, and the Gaussian 09 package (Gaussian, 2010) is used for the cytochrome c oxidase geometry optimizations and for the Hessian calculations.

The computational procedure described above has been kept as similar as possible to what has been used the past decades. The reason is that a large experience has been gained of this approach, and this knowledge is very useful when the accuracy of the predictions obtained is judged.
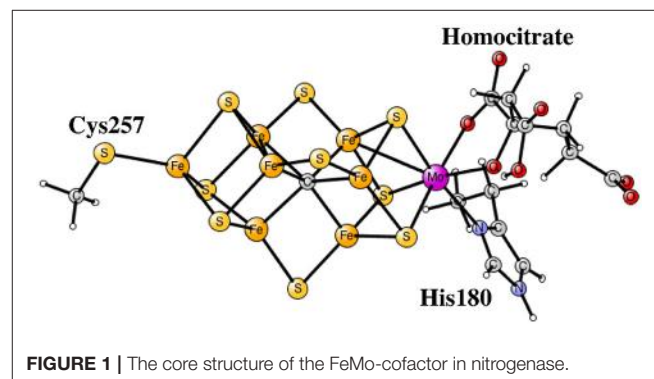
III. NITROGENASE

Nitrogenase is the main enzyme in nature that catalyzes nitrogen reduction from the air. The core of its catalytic cofactor is shown **Figure 1** (Kim and Rees, 1992; Spatzal et al., 2011). It has seven irons and one molybdenum connected by sulfide bridges. Species containing vanadium and with all-iron exist. A redox potential of −1.6 V, the lowest in nature, is used for the reduction. Quite surprisingly, the oxidation state of the cofactor is not particularly low, which would normally have been expected for a strongly reducing complex. Instead there are four Fe(III) and three Fe(II). Molybdenum is in the Mo(III) state.

Almost all experimental information concerns the ground state before reduction. However, there is one notable exception which is a key for finding the mechanism by DFT model calculations. It has been shown by EPR that the reduced state that activates N_2, termed E_4, has two bridging hydrides (Hoffman et al., 2013). These two hydrides were found to be removed as H_2 in a reductive elimination process, which is directly followed by the binding of N_2. The process was found to be easily reversible by changing the pressure of hydrogen and nitrogen. This means that the states involved must be nearly isoenergetic.

The experimental suggestion for the structure of E_4 the central carbon remains unprotonated and has two bridging hydrides (Lukoyanov et al., 2016). Since E_4 was suggested to appear after four reductions of the experimentally characterized ground state in the catalytic cycle, see **Figure 1**, there should be two remaining protonations. They were suggested to be on the sulfides. This meant that the redox state for E_4 should be the same as for the ground state with four Fe(III), a very surprising situation.

The results obtained for nitrogenase after four reductions of the ground state in **Figure 1** are shown in **Table 1**. The corresponding structures are shown in **Figures 2**, **3**. For each functional, the results for six different structures are listed. These structures were taken from the best ones obtained previously, where several hundred structures were compared. They were obtained with 20% exact exchange at the lacvp* level, and were then compared with a large cc-pvtz(-f) basis set at the 15% level. For each functional the geometries were optimized, which

FIGURE 1 | The core structure of the FeMo-cofactor in nitrogenase.

TABLE 1 | Relative energies (kcal/mol) for states obtained after four reductions from the ground state of nitrogenase, using density functionals with different fractions of exact exchange.

Structure	0%	10%	15%	20%
C,H⁻	0.0	0.0	0.0	0.0
C,2H⁻	−12.0	+7.2	+14.7	+22.6
CH	+0.1	−9.9	−10.8	−16.1
CH$_2$,H⁻	−23.4	−31.7	−34.0	−37.4
CH$_3$	−18.1	−30.7	−47.9	−59.2
C - H$_2$	−46.0	−36.1	−33.3	−32.0

FIGURE 2 | Optimized core structures for the C,H⁻, C,2H⁻, and CH structures.
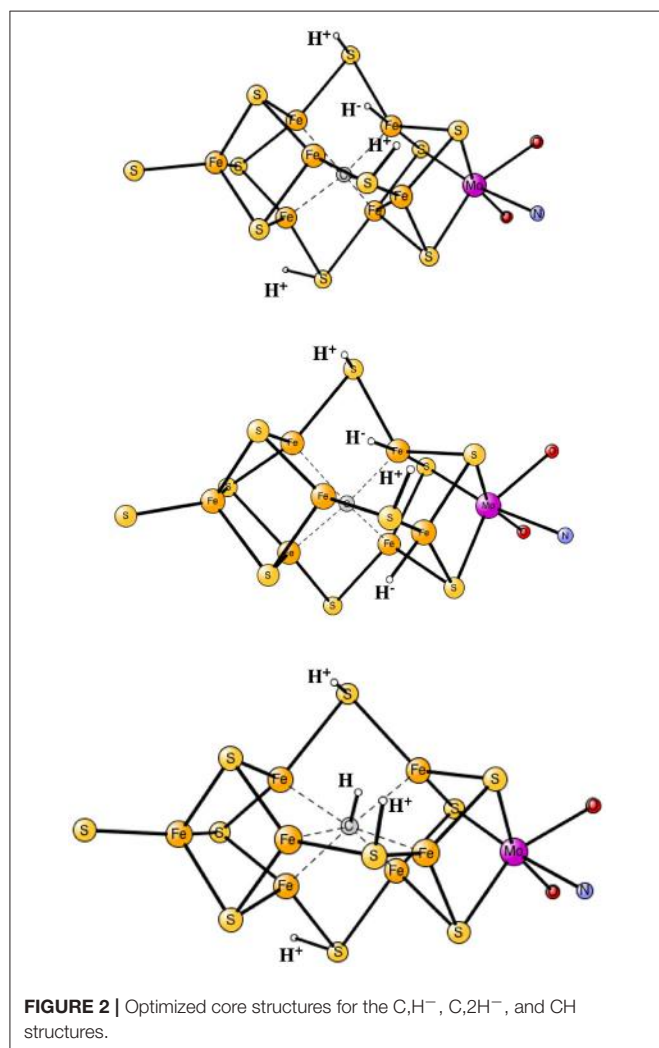
turned out to be important. The first structure, termed C,H⁻, has an unprotonated carbon and one hydride. There are three protonations of the sulfides. In the second structure, termed C,2H⁻, one of the protons on the sulfides has moved to a hydride position, so there are two hydrides and two protonated sulfides. This is the one of the six structures that corresponds to the experimentally suggested structure. In the third structure, termed CH, the hydride in the first structure has moved to carbon, and

there are still three protonated sulfides remaining. The fourth structure, termed CH$_2$,H⁻, has a doubly protonated carbon, one hydride and one protonated sulfide. In the fifth structure, the hydride has moved to carbon to form a CH$_3$ group, and there is one protonated sulfide. Finally, in the sixth state, the two hydrides in the second structure has been removed to form a free H$_2$ molecule.

The main conclusion that can be drawn from the results in **Table 1** is that the structure, termed C,2H⁻, does not fit the experimental observations. This is true for all functionals in the table. Already at this point, the important conclusion can be made, that the E$_4$ state does not look as suggested experimentally. However, the by far strongest argument against the experimental suggestion comes from the results for the sixth structure. Experimentally, it is known that two hydrides should be removed as H$_2$ essentially thermoneutrally. As can be seen in the table, the removal of H$_2$ is much too exergonic for all functionals to fit that observation. With 20%, the removal is exergonic by 54.6 (22.6 + 32.0) kcal/mol for 15% exergonic by 48.0 (14.7 + 33.3) kcal/mol, for 10% exergonic by 43.3 kcal/mol and for 0 % exergonic by 34.0 kcal/mol.

It is also found that all functionals prefer a protonated carbon, suggested in previous DFT studies (Siegbahn, 2016). For 20, 15, and 10%, this preference is very large. For 0%, the preference is less pronounced, but it is still there by 11.4 (23.4–12.0) kcal/mol.

It should be added that results with three hydrides are not included in **Table 1**, but have been done. The reason three hydrides are not in the table is that they can be excluded as possible E$_4$ states already from the beginning. It is chemically unreasonable that the first four reductions of the cofactor should lead to an oxidation of the metals rather than a reduction, in particular, since the lowest redox potential of –1.6 V is used. The already high oxidation state of the ground state, with four Fe(III), would be six Fe(III) with three hydrides. The possibility that there should be five Fe(III) and a Mo(IV) is equally unreasonable chemically. Furthermore, for 20, 15 and 10% exchange the structure with three hydrides is very high in energy compared to the other structures. Even for the functional with 0%, the best solution obtained is higher in energy than the CH$_2$H⁻ structure by +6.1 kcal/mol. However, that value is somewhat uncertain since the spin-coupling with three hydrides is very different from the other cases, and includes several low-spin-coupled Fe(III). This leads to a very large number of possibilities for the spin-state. Five spin-couplings were tried, all higher in energy than the CH$_2$H⁻ structure. They were selected to have the sums of the spins on Fe2-Fe4 opposite to the ones for Fe5-Fe7. Two of them were the ones optimal for the ground state and for the reduced states found previously (Siegbahn, 2016). The last argument against structures with three terminal hydrides is that they are incompatible with the EPR experiments (Hoffman et al., 2013), which clearly show two bridging hydrides.

There are other points of interest in **Table 1**. For 15 and 20% exchange, the structure with two hydrides are quite high in energy compared to the first structure, by +22.6 and by +14.7 kcal/mol, respectively. Furthermore, they are also terminal, not bridging, hydrides in contradiction to experiments. The energy of
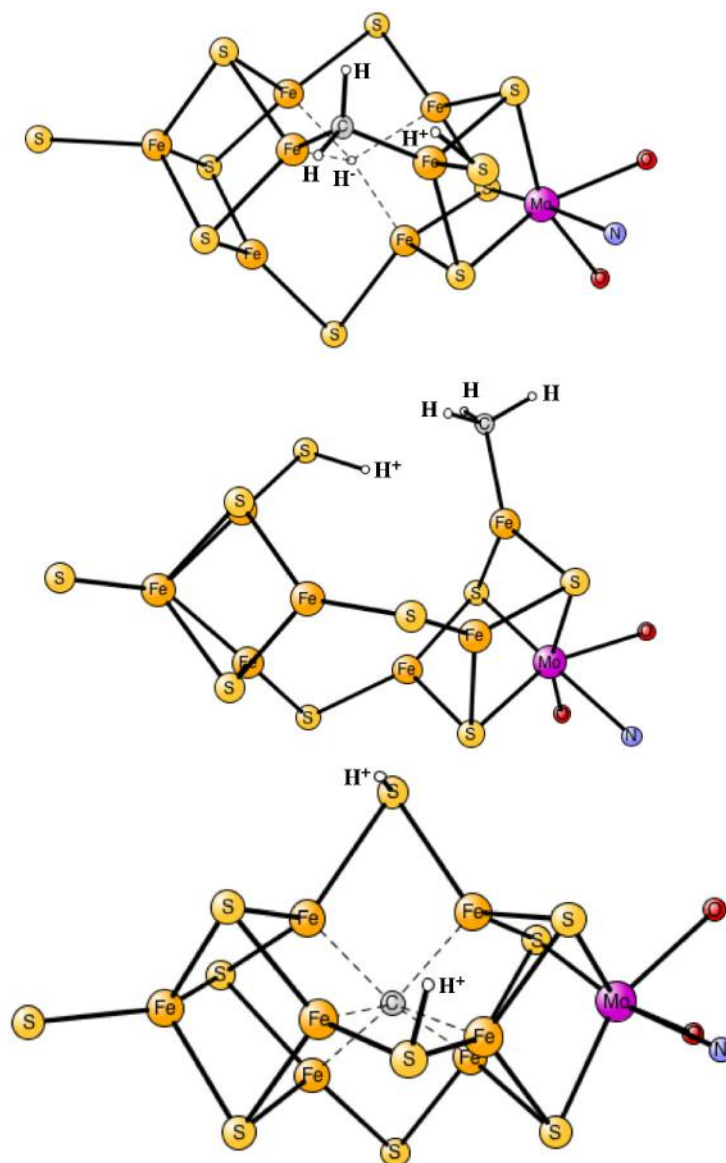
FIGURE 3 | Optimized core structures for the CH$_2$,H$^-$, CH$_3$, and C-2H structures.

this structure is somewhat lower for 10%, but still +7.2 kcal/mol higher than for the starting structure. For 0%, the situation is different. Here the structure with two hydrides is lower than for the first structure by -12.0 kcal/mol, but the two hydrides can still be removed to form H$_2$ with a gain of 34.0 kcal/mol, as mentioned above. The lowest state for 20 and 15% is the CH$_3$ structure, but for 10 and 0% it is the CH$_2$,H$^-$ structure, with a margin of −1.0 kcal/mol for 10 % and with −5.3 kcal/mol for 0%.

In summary, there are a few clear conclusions. First, none of the functionals prefers an unprotonated carbon. Also, all functionals bind the hydrides very poorly. To remove the hydrides as H$_2$ is strongly exergonic in all cases, by more than 30 kcal/mol, when experiments show that this should be almost thermoneutral. This means that the experimentally suggested

structure for E$_4$, does not agree with the results for any functional and can be ruled out by a large margin by DFT.

In a very recent paper, a new mechanism for H$_2$ release and N$_2$ binding in nitrogenase has been suggested (Raugei et al., 2018). A non-hybrid method was used. A mechanism was suggested in which the two hydrides in E$_4$ endergonically form a locally bound H$_2$ molecule. To avoid the problem with the very large computed exergonicity when H$_2$ is released, the key to their mechanism is that this bound H$_2$ molecule could only be released with a very high barrier. If the barrier is lower than 18 kcal/mol, there would be no protonation of N$_2$, but a high barrier should prevent H$_2$ from being released. In their mechanism, the endergonic binding of N$_2$ was found to reduce the barrier for releasing H$_2$ by a significant amount, and H$_2$ can then be released. There are

many bound H_2 complexes in the literature, but none of them behaves like suggested in Ref. (Raugei et al., 2018). In all the published cases there is at most a weakly bound H_2, which can be released without additional barriers apart from the endergonicity. A search for a bound H_2 was initiated in the present study using 0% exchange (non-hybrid), and a picture is obtained which is very similar to the ones previously published in the literature. A weak local minimum for a bound H_2 is obtained and releasing it from that minimum goes over at most a very small barrier, less than 5 kcal/mol. The release is quite exergonic including a gain of entropy of about 10 kcal/mol.

Since DFT rules out an unprotonated carbon structure, another structure has to be found for E_4. If no other structure can be found, the conclusion must be that no version of DFT can handle nitrogenase, a very unlikely scenario. It has been suggested in previous DFT studies that the lowest energy structure in **Table 1**, is merely a starting structure for catalysis (Siegbahn, 2016). The first four reductions of the ground state structure in **Figure 1** would then be just an initial activation process, done only once, before catalysis starts leading to a new E_0 state. Following that suggestion, four additional reductions would lead to the actual E_4 structures. The results for the E_4 structures, determined in this way, are shown in **Table 2** for the same four functionals as discussed above. The results are also displayed in **Figure 4**. The results for six structures are shown, following the conclusions of the EPR experiments. The first entry, termed $2H^-$, has two hydrides, a CH_3 ligand and three protonated sulfurs, altogether eight protonations of the cofactor. The second entry, termed "H-H re TS" shows the barrier for the re mechanism (reductive elimination) of two hydrides to form H_2. The third entry, termed "H-H hp TS," shows the barrier for the hp (hydride, proton) mechanism, where one hydride and one proton form H_2. In the fourth entry, termed "- H_2^a," the two hydrides have been removed to form a free hydrogen molecule. The fifth entry, termed "- H_2^b," differs from the second one by a rotation of the homocitrate ligand. This rotation was found in earlier studies (Siegbahn, 2018) to be required for binding N_2. In the final sixth entry, termed "+ $N_2–H_2^b$," N_2 binds to the fifth structure.

The results for the functional with 20% agree very well with the EPR experiments. H_2 can be removed with a very small energy gain of –2.1 kcal/mol. The barrier for H_2 elimination using the re mechanism is 13.1 kcal/mol, well within the range required by experiments. Even more importantly, the barrier using the hp mechanism of 17.3 kcal/mol is 4.2 kcal/mol higher than for the re mechanism, leading to the required N_2 activation rather than H_2 production. The following homocitrate rotation is only uphill by +1.4 kcal/mol. In the next step, N_2 binds exergonically to the third structure by –0.1 kcal/mol. This means that the release of H_2 and the binding of N_2 should be easily reversible as observed experimentally. There is a minor discrepancy, since there should be some driving force (exergonicity) for this process, but the error is small for such a complicated process. Going to 15%, the discrepancy is somewhat larger. The release of H_2 is exergonic by -6.6 kcal/mol and N_2 binding is endergonic by +4.3 kcal/mol. However, the discrepancy to experiments is not alarming. Again, the barrier eliminating H_2 by the re mechanism is preferred

TABLE 2 | Relative energies (kcal/mol) for states obtained after eight reductions from the ground state of nitrogenase, using density functionals with different fractions of exact exchange.

Structure	0%	10%	15%	20%
$2H^-$	0.0	0.0	0.0	0.0
H–H TS re	+16.8	+12.4	+9.6	+13.1
H–H TS hp	+13.9	+16.0	+15.2	+17.3
$- H_2^a$	–7.1	–3.6	–6.6	–2.1
$- H_2^b$	–3.5	–1.9	–4.8	–0.7
$+ N_2 - H_2^b$	+10.7	+9.3	–0.5	–0.6

^{a} H_2 is removed with homocitrate non-rotated. ^{b} H_2 is removed with homocitrate rotated.
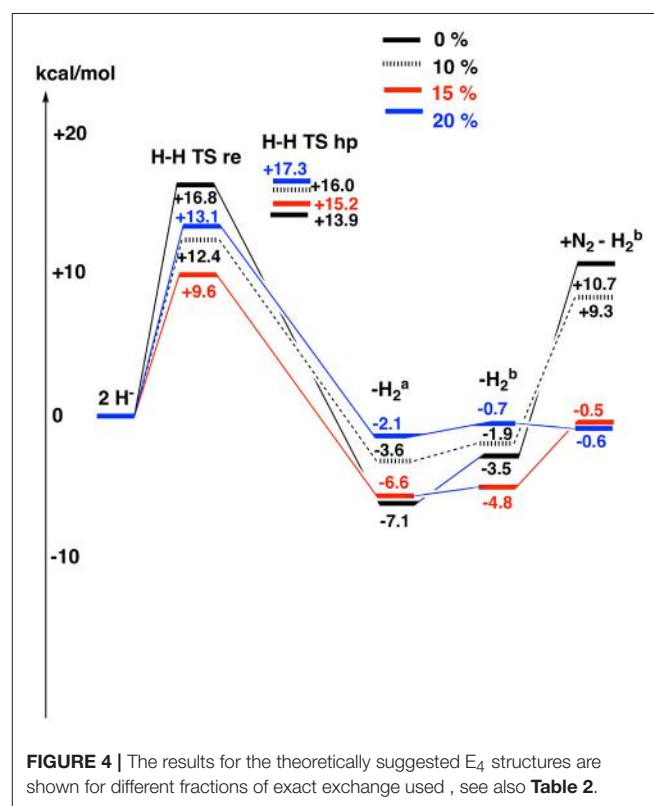
FIGURE 4 | The results for the theoretically suggested E_4 structures are shown for different fractions of exact exchange used , see also **Table 2**.

over the one using the hp mechanism, now by 6.2 kcal/mol. For the functional with 10%, the discrepancy to experiments is significantly increased. Most noteworthy, the binding of N_2 is now endergonic by + 11.2 (9.3 + 1.9) kcal/mol. However, the barrier for the re mechanism is still lower than the one for the hp mechanism. Finally, for the functional with 0%, the discrepancy to experiments increases further. For example, the binding of N_2 is now endergonic by +14.2 kcal/mol. Furthermore, the barrier for the hp mechanism is now lower than for the re mechanism, leading to production of H_2 rather than protonation of N_2.

IV. CYTOCHROME c OXIDASE

The membrane bound enzyme cytochrome c oxidase (CcO) catalyzes the reduction of molecular oxygen to water as the last
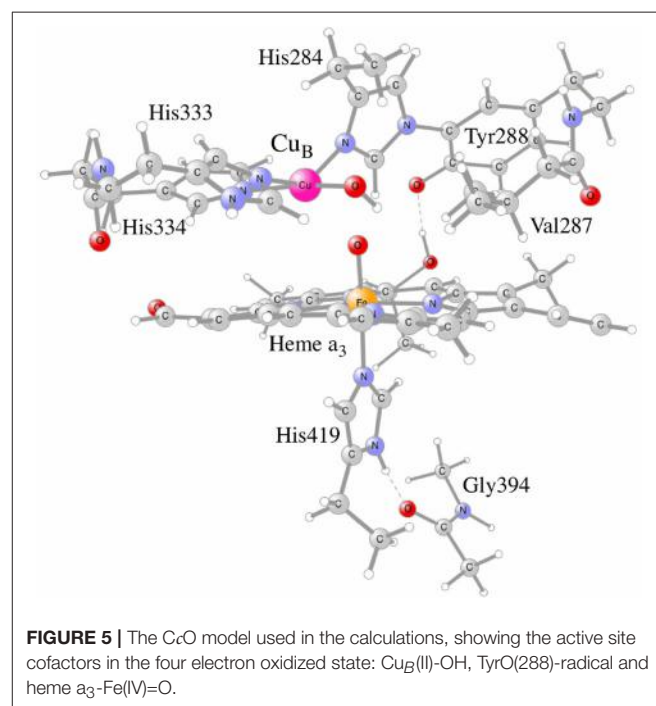
step in the respiratory chain in aerobic organisms. The chemistry occurs in an active site consisting of a high-spin heme group, a copper complex, labeled Cu_B, and a redox active tyrosine, referred to as the binuclear center (BNC). The electrons are delivered to the BNC from a reduced cytochrome c, located on the positive side of the membrane. The protons are transferred to the BNC from bulk water on the negative side of the membrane. Molecular oxygen binds to the reduced BNC, with heme-Fe(II), Cu_B(I) and TyrOH. The O-O bond is cleaved in a single reaction step, yielding a four electron oxidized BNC, with heme-Fe(IV)=O, Cu_B(II)OH and a neutral TyrO-radical (Proshlyakov et al., 1998; Fabian et al., 1999). The rest of the catalytic cycle consists of four reduction steps, each taking up one electron and one proton to the BNC, leading back to the reduced state with two new water molecules. The overall energetics of the reduction process is obtained from the difference in reduction potential between the electron donor and the acceptor, molecular oxygen. With cytochrome c as the electron donor, reduction of one oxygen molecule becomes exergonic by 50.7 kcal/mol (2.2 eV) (Brzezinski, 2004). A significant part of this free energy is conserved as an electrochemical gradient across the membrane, which in turn is used by another enzyme, ATP-synthase, to produce ATP, the energy currency of the cells. Two processes contribute to the gradient buildup, one is the electrogenic chemistry (taking the electrons and the protons from opposite sides of the membrane), and the other is the so called proton pumping, which means that the chemistry is coupled to proton transfer across the entire membrane. The largest group of C*c*O:s are known to pump one proton per electron, i.e., four protons per oxygen molecule (Brzezinski and Gennis, 2008; Kaila et al., 2010). The mechanism for the proton pumping, i.e., how to couple the transfer of one electron to the active site with the uptake of two protons from the negative side of the membrane, is still under debate (Rich, 2017).

The process of oxygen reduction in C*c*O has been studied in detail using density functional theory. A mechanism for the O-O bond cleavage step was suggested based on computational results at an early stage (Blomberg and Siegbahn, 2006, 2010), and it was later confirmed in a combined experimental and computational study (Poiana et al., 2017). Another result from the computational studies concerns the mechanism of proton pumping, for which it has been suggested that the redox active tyrosine in the active site plays an essential role (Blomberg, 2016). Experimental support for the suggested pumping mechanism is that an active site tyrosine is conserved in all families of C*c*O's (Hemp et al., 2006). To understand the proton pumping in C*c*O it is essential to know the energetics of the individual reduction steps in the catalytic cycle, which depends on the reduction potential of the active site cofactor that is reduced in each particular step. Experimental investigations have indicated that the four different active site reduction potentials involved are quite different, and that only two of them seem to be large enough (about 0.8 V) to afford proton pumping (Wikström and Morgan, 1992; Kaila et al., 2010). Also the BNC potentials have been studied computationally, with result that partly differ from the experimental measurements (Blomberg and Siegbahn, 2015a), see further below. In all these computational studies the b3lyp

type of functional was used as described above in Computational details.

To further test the reliability of the results obtained for C*c*O, it was decided to systematically investigate the calculated proton coupled reduction potentials for the active site cofactors by varying the fraction of exact exchange in the b3lyp potential, using the model shown in **Figure 5**. It is noted that each of the proton coupled reduction potentials corresponds to the formation of a new O-H bond in the active site. The individual reduction potentials can therefore be estimated from the strengths of the different O-H bonds. The strength of an O-H bond is not sensitive to a distant surrounding, since the charge is not changed by the addition of a (H^+, e^-)-couple, which means that reasonably sized models (150–200 atoms) can be used in the calculations. For each functional (different percentage of exact exchange) new structures are optimized, and since the most recently used procedure for C*c*O has been to optimize the structures with dispersion, the D3 dispersion correction with parameters from the original b3lyp-D3 functional (Grimme et al., 2010) was used in both geometry optimizations and energy calculations. Since results for the functional with 15% were also obtained using structures optimized with 20% (not reported) it could be noted that the final energetics differs only slightly (<2 kcal/mol) from the ones reported below.

To estimate the energetics of the reduction steps of the C*c*O catalytic cycle, the sum of the energy of an electron transferred from cytochrome c plus a proton transferred from bulk water is needed. This energy is parameterized to reproduce the experimental results for the overall reaction (50.7 kcal/mol, see above). By subtracting the parameterized energy from each O-H bond strength, the exergonicity relative to the cytochrome c donor is obtained, and by comparing to the midpoint potential

FIGURE 5 | The C*c*O model used in the calculations, showing the active site cofactors in the four electron oxidized state: Cu_B(II)-OH, TyrO(288)-radical and heme a_3-Fe(IV)=O.

TABLE 3 | Calculated energetics as a function of the amount of exact exchange for the reduction of Cu_B (taking up one electron and one proton) in the catalytic cycle of cytochrome c oxidase.

Reduction step	0%	10%	15%	20%	exp[a]
$Cu_B(II)OH \rightarrow Cu_B(I)OH_2$					
$\Delta H(O-H)$ (kcal/mol)	89.0	85.2	84.3	86.7	
ΔG^b (kcal/mol)	−21.7	−17.5	−16.5	−18.8	
$E_{m,7}$ (V)	1.19	1.01	0.97	1.07	0.4+

[a] From Kaila et al. (2010). Although the experimental measurements have given low values, below 0.4 V, it has been suggested based on experiment that during turnover the Cu_B potential actually is higher than the measured values (Kaila et al., 2010).
[b] Exergonicity relative to cytochrome c, with a reduction potential of 0.25 V.

TABLE 4 | Calculated energetics as a function of the amount of exact exchange for two of the four reduction steps (taking up one electron and one proton) in the catalytic cycle of cytochrome c oxidase.

Reduction step	0%	10%	15%	20%	exp[a]
$TyrO^\bullet \rightarrow TyrOH$					
$\Delta H(O-H)$ (kcal/mol)	75.4	78.7	81.1	82.0	
ΔG^b (kcal/mol)	−8.1	−11.0	−13.3	−14.1	
$E_{m,7}$ (V)	0.60	0.73	0.83	0.86	0.82
$heme\text{-}Fe(IV)=O \rightarrow heme\text{-}Fe(III)OH$					
$\Delta H(O-H)$ (kcal/mol)	56.4	72.4	77.7	85.5	
ΔG^b (kcal/mol)	+10.9	−4.7	−9.9	−17.6	
$E_{m,7}$ (V)	−0.22	0.45	0.68	1.01	0.76

[a] From Kaila et al. (2010).
[b] Exergonicity relative cytochrome c, with a reduction potential of 0.25 V.

of cytochrome c (0.25 V) the midpoint potential can be estimated for each cofactor. The first result to be discussed concerns the proton coupled reduction potential for Cu_B, for which the experimental measurements have given low values, 0.2–0.4 V (Jancura et al., 2006; Brand et al., 2007; Vilhjámsdóttir et al., 2018). In contrast, the previous calculations indicate a much higher value (0.9–1.0 V) during catalytic turnover (Blomberg and Siegbahn, 2015a), and an explanation for the low experimental values has been suggested based on the computational results (Blomberg and Siegbahn, 2015b). As can be seen in **Table 3**, all functionals support a large reduction potential for Cu_B, with results that vary only slightly and in a somewhat irregular way with the fraction of exact exchange in the functional, 0.97-1.19 V. Interestingly, using a small model complex it was also shown that the b3lyp* (15%) results for the O-H bond strength in $Cu_B(I)OH_2$ agrees to within one kcal/mol with CCSD(T) results (Blomberg and Siegbahn, 2015a).

As mentioned above, experiments have shown that two of the active site reduction potentials are about 0.8 V (Kaila et al., 2010), which is large enough for proton pumping in the corresponding reduction steps. These two steps concern reduction of the tyrosyl radical and heme-Fe(IV)=O, respectively (Kaila et al., 2010). The calculated results for these two cofactors using the different functionals are reported in **Table 4**. For the tyrosyl radical the

experimental value is 0.82 V, and all functionals with a fraction exact exchange (10–20%) are in reasonable agreement with experiment, 0.73–0.86 V. Again, the reduction potential does not vary very much with the exact exchange, and even the functional without exact exchange (0%) is not too far off, with a value of 0.60V. For the heme-Fe(IV)=O reduction the situation is quite different. The experimental value is 0.76 V (Kaila et al., 2010), and here the functional with 15% gives the best agreement, with a value of 0.68 V. The result with 20%, 1.01 V, is also in reasonable agreement with experiment. The result without exact exchange (0%) disagrees qualitatively with experiment, the calculated value is −0.22 V, and also with only 10% exchange the calculated value, 0.45 V, is quite far from the experimental value. An important factor for the strong variation of this reduction potential is that there is a change of spin coupling in the heme-Fe(IV)=O → heme-Fe(III)OH transition, where iron is low-spin coupled in heme-Fe(IV)=O and high-spin coupled in heme-Fe(III)OH.

V. CONCLUSIONS

By using the present systematic approach of DFT, a few major conclusions can be drawn for the nitrogenase mechanism. In **Table 1**, the results are given for different E_4 structures obtained after four reductions of the ground state, as has been suggested experimentally. The most important result here is that no functional gives results consistent with the conclusions of the experimental EPR study (Hoffman et al., 2013). In particular all functionals give a very large exergonicity for releasing H_2. This means that the experimentally suggested structure can not be supported by any DFT functional. The discrepancy to experiments is very large, independent of fraction of exact exchange used. Even the non-hybrid method (0%), gives very poorly bound hydrides, which can be removed with a gain of 34 kcal/mol, when experiments indicate that the release should be nearly thermoneutral. In contrast, the theoretically suggested structure for E_4, obtained after eight reductions of the ground state, shows a much more reasonable agreement with the experimental EPR information, in particular, with 20 and 15% exchange. Since DFT indeed found a structure that agrees well with what is known for the E_4 state, a conclusion that DFT totally fails for nitrogenase is very farfetched. If that would have been the case, it would be the first example among the many redox reactions studied, that would show that behavior (Blomberg et al., 2014). If no structure would have been found that agreed with experimental information for E_4, the conclusion could have been different. A very recent study with many different functionals did not find any functional that gave a preference for the experimental structure (Cao et al., 2018), in agreement with the prediction made here.

For CcO, the calculations show that the active site cofactor Cu_B has a large midpoint potential, about 1 V, a result obtained regardless of the fraction of exact exchange in the functional. This is in contrast to the much lower potentials obtained in experimental measurements, but in agreement with experimental observations on proton pumping. For another active site cofactor,

heme-Fe(IV)=O, the functional with 15% exact exchange gives the best agreement with experimental observations for the midpoint potential, and the functional without exact exchange gives qualitatively wrong results.

In summary, the results for the different functionals show that the best agreement with experiments, for both nitrogenase and cytochrome *c* oxidase, is obtained with 15–20% exact exchange in the functional. 10% exact exchange gives results slightly worse, and the use of a non-hybrid functional (0% exact exchange) gives qualitatively wrong potential surfaces in both cases. These results are in line with experience gathered during the past two decades (Blomberg et al., 2014). This experience can be used in two ways. It could either be used to calibrate the exact exchange fraction based on some well-known experimental fact for that system. Or,

the difference between the results for 15 and 20% could be used as an estimate of the error in the calculations. The latter approach has mainly been used in our previous studies.

AUTHOR CONTRIBUTIONS

PS wrote the paper and did nitrogenase calculations. MB wrote the paper and did cytochrome *c* oxidase calculations.

FUNDING

REFERENCES

Becke, A. D. (1993). Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* 98, 5648–5652. doi: 10.1063/1.464913

Blomberg, M. R .A., and Siegbahn, P. E. M. (2015a) How cytochrome *c* oxidase can pump four protons per oxygen molecule at high electrochemical gradient *Biochim. Biophys. Acta.* 1847, 364–376. doi: 10.1016/j.bbabio.2014.12.005

Blomberg, M. R. A. (2016). The mechanism of oxygen reduction in cytochrome *c* oxidase and the role of the active site tyrosine. *Biochemistry* 55, 489–500. doi: 10.1021/acs.biochem.5b01205

Blomberg, M. R. A., Borowski, T., Himo, F., Liao, R.-Z., and Siegbahn, P. E. M. (2014). Quantum chemical studies of mechanisms for metalloenzymes. *Chem. Rev.* 114, 3601–3658. doi: 10.1021/cr400388t

Blomberg, M. R. A., and Siegbahn, P. E. M. (2006). Quantum chemistry applied to the mechanisms of transition metal containing enzymes-cytochrome c oxidase a particularly challenging case. *J. Comp. Chem.* 27, 1373–1384. doi: 10.1002/jcc.20448

Blomberg, M. R. A., and Siegbahn, P. E. M. (2010). Quantum chemistry as a tool in bioenergetics. *Biochim. Biophys. Acta* 1797, 129–142. doi: 10.1016/j.bbabio.2009.10.004

Blomberg, M. R. A., and Siegbahn, P. E. M. (2015b). Protonation of the binuclear active site in cytochrome *c* oxidase decreases the reduction potential of Cu$_B$. *Biochim. Biophys. Acta.* 1847, 1173–1180. doi: 10.1016/j.bbabio.2015.06.008

Brand, S. E., Rajagukguk, S., Ganesan, K., Geren, L., Fabian, M., Han, D., et al. (2007). F. A new ruthenium complex to study single-electron reduction of the pulsed OH state of detergent-solubilized cytochrome oxidase. *Biochemistry* 46, 14610–14618. doi: 10.1021/bi701424d

Brzezinski, P. (2004). Redox-driven membrane-bound proton pumps. *Trends Biochem. Sci.* 29, 380–387. doi: 10.1016/j.tibs.2004.05.008

Brzezinski, P., and Gennis, R. B. (2008). Cytochrome c oxidase: exciting progress and remaining mysteries. *J. Bioenerg. Biomembr.* 40, 521–531. doi: 10.1007/s10863-008-9181-7

Cao, L., Caldararu, O., and Ryde, U. (2018). Protonation and reduction of the FeMo cluster in nitrogenase studied by quantum mechanics/molecular mechanics (QM/MM) calculations. *J. Chem. Theory Comput.* 14, 6653–6678. doi: 10.1021/acs.jctc.8b00778

Fabian, M., Wong, W. W., Gennis, R. B., and Palmer, G. (1999). Mass spectrometric determination of dioxygen bond splitting in the "peroxy" intermediate of cytochrome c oxidase. *Proc. Natl. Acad. Sci. U.S.A.* 96, 13114–13117.

Gaussian 09, (2010). *Revision C.01.* Wallingford, CT: Gaussian Inc.

Grimme, S., Anthony, J., Ehrlich, S., and Krieg, H. (2010). 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. doi: 10.1063/1.3382344

Hemp, J., Robinson, D. E., Ganesan, K. B., Martinez, T. J., Kelleher, N. L., and Gennis, R.B. (2006). Evolutionary migration of a post-translationally modified active-site residue in the proton-pumping heme-copper oxygen reductases. *Biochemistry* 45, 15405–15410. doi: 10.1021/bi062026u

Hoffman, B. M., Lukoyanov, D., Dean, D. R., and Seefeldt, L. C. (2013). Nitrogenase: a draft mechanism. *Acc. Chem. Res.* 46, 587–595. doi: 10.1021/ar300267m

Jaguar 7.6 (2009). *Schrödinger.* New York, NY: LLC.

Jancura, D., Berka, V., Antalik, M., Bagelova, J., Gennis, R. B., Palmer, G., et al. (2006). Spectral and kinetic equivalence of oxidized cytochrome c oxidase as isolated and "activated" by reoxidation. *J. Biol. Chem.* 281, 30319–30325. doi: 10.1074/jbc.M605955200

Kaila, V. R. I., Verkhovsky, M. I., and Wikström, M. (2010). Proton-coupled electron transfer in cytochrome oxidase. *Chem. Rev.* 110, 7062–7081. doi: 10.1021/cr1002003

Kim, J., and Rees, D. C. (1992). Structural models for the metal centers in the nitrogenase molybdenum-iron protein. *Science* 257, 1677–1682. doi: 10.1126/science.1529354

Lukoyanov, D., Khadka, N., Dean, D. R., Raugei, S., Seefeldt, L. C., and Hoffman, B. M. (2016). Photoinduced reductive elimination of H$_2$ from the nitrogenase dihydride (Janus) State Involves a FeMo-cofactor-H$_2$ intermediate. *Inorg. Chem.* 56, 2233–2240. doi: 10.1021/acs.inorgchem.6b02899

Poiana, F., von Ballmoos, C., Gonska, N., Blomberg, M. R. A., Ädelroth, P., and Brzezinski, P. (2017). Splitting of the O-O bond at the heme-copper catalytic site of respiratory oxidases. *Sci. Adv.* 3:e1700279. doi: 10.1126/sciadv.1700279

Proshlyakov, D. A., Pressler, M. A., and Babcock, G. T. (1998). Dioxygen activation and bond cleavage by mixed-valence cytochrome c oxidase. *Proc. Natl. Acad. Sci. U.S.A.* 95, 8020–8025. doi: 10.1073/pnas.95.14.8020

Raugei, S., Seefeldt, L. C., and Hoffman, B. M. (2018). Critical computational analysis illuminates the reductive-elimination mechanism that activates nitrogenase for N$_2$ reduction. *Proc. Natl. Acad. Sci. U.S.A.* 115, E10521–E10530. doi: 10.1073/pnas.1810211115

Reiher, M., Salomon, O., and Hess, B. A. (2001). Reparameterization of hybrid functionals based on energy differences of states of different multiplicity. *Theor. Chem. Acc.* 107, 48–55. doi: 10.1007/s00214-001-0300-3

Rich, P. R. (2017). Mitochondrial cytochrome c oxidase: catalysis, coupling and controversies. *Biochem. Soc. Trans.* 45, 813–829. doi: 10.1042/BST20160139

Schwabe, T., and Grimme, S. (2007). Double-hybrid density functionals with long-range dispersion corrections: higher accuracy and extended applicability. *Phys. Chem. Chem. Phys.* 9, 3397–3406. doi: 10.1039/b704725h

Siegbahn, P. E. M. (2001). Modeling aspects of mechanisms for reactions catalyzed by metalloenzymes. *J. Comp. Chem.* 22, 1634–1645. doi: 10.1002/jcc1119

Siegbahn, P. E. M. (2009). Structures and energetics for O$_2$ formation in photosystem II, *Acc. Chem. Res.*, 42, 1871–1880. doi: 10.1021/ar900117k

Siegbahn, P. E. M. (2013). Water oxidation mechanism in photosystem II, including oxidations, proton release pathways, O-O bond formation and O$_2$ release. *Biochim. Biophys. Acta* 1827, 1003–1019. doi: 10.1016/j.bbabio.2012.10.006

Siegbahn, P. E. M. (2016). Model calculations suggest that the central carbon in the FeMo-cofactor of nitrogenase becomes protonated in the process of nitrogen fixation. *J. Am. Chem. Soc.* 138, 10485–10495. doi: 10.1021/jacs.6b03846

Siegbahn, P. E. M. (2018). A major structural change of the homocitrate ligand of probable importance for the nitrogenase mechanism, *Inorg. Chem.* 57, 1090–1095. doi: 10.1021/acs.inorgchem.7b02493

Siegbahn, P. E. M., and Blomberg, M. R. A. (2014). Energy diagrams for water oxidation in photosystem II using different density functionals. *J. Chem. Theory Comput.* 10, 268–272. doi: 10.1021/ct401039h

Spatzal, T., Aksoyoglu, M., Zhang, L., Andrade, S. L. A., Schleicher, E., Weber, S., et al. (2011). Evidence for interstitial carbon in nitrogenase FeMo cofactor. *Science* 334, 940. doi: 10.1126/science.1214025

Vilhjámsdóttir, J., Gennis, R.B., and Brzezinski, P. (2018). The electron distribution in the "activated" state of cytochrome *c* oxidase. *Sci. Rep.* 8:7502. doi: 10.1038/s41598-018-25779-w

Wikström, M., and Morgan, J. E. (1992). The dioxygen cycle. Spectral, kinetic, and thermodynamic characteristics of ferryl and peroxy intermediates observed by reversal of the cytochrome oxidase reaction. *J. Biol. Chem.* 267, 10266–10273.

Computational Study of Protein-Ligand Unbinding for Enzyme Engineering

Sérgio M. Marques[1,2], David Bednar[1,2] and Jiri Damborsky[1,2]*

[1] Loschmidt Laboratories, Department of Experimental Biology and Research Centre for Toxic Compounds in the Environment RECETOX, Faculty of Science, Masaryk University, Brno, Czechia, [2] International Clinical Research Center, St. Anne's University Hospital Brno, Brno, Czechia

The computational prediction of unbinding rate constants is presently an emerging topic in drug design. However, the importance of predicting kinetic rates is not restricted to pharmaceutical applications. Many biotechnologically relevant enzymes have their efficiency limited by the binding of the substrates or the release of products. While aiming at improving the ability of our model enzyme haloalkane dehalogenase DhaA to degrade the persistent anthropogenic pollutant 1,2,3-trichloropropane (TCP), the DhaA31 mutant was discovered. This variant had a 32-fold improvement of the catalytic rate toward TCP, but the catalysis became rate-limited by the release of the 2,3-dichloropropan-1-ol (DCP) product from its buried active site. Here we present a computational study to estimate the unbinding rates of the products from DhaA and DhaA31. The metadynamics and adaptive sampling methods were used to predict the relative order of kinetic rates in the different systems, while the absolute values depended significantly on the conditions used (method, force field, and water model). Free energy calculations provided the energetic landscape of the unbinding process. A detailed analysis of the structural and energetic bottlenecks allowed the identification of the residues playing a key role during the release of DCP from DhaA31 via the main access tunnel. Some of these hot-spots could also be identified by the fast CaverDock tool for predicting the transport of ligands through tunnels. Targeting those hot-spots by mutagenesis should improve the unbinding rates of the DCP product and the overall catalytic efficiency with TCP.

Keywords: unbinding kinetics, protein engineering, molecular dynamics, metadynamics, adaptive sampling, CaverDock

INTRODUCTION

Until recently, the modern methods of structure-based drug design relied primarily on the high binding affinity toward the targets to predict their biological performance. However, that paradigm has been changed once it was realized that the half-life of a drug is equally important to define its *in vivo* efficacy, and hence both thermodynamics and kinetics profiles must be taken into account (Lu and Tonge, 2010). For this reason, we have recently witnessed a boom of different methods for the computational prediction of receptor-ligand (un)binding kinetics (Chiu and Xie, 2016; Ferruz and De Fabritiis, 2016; Dickson et al., 2017; Rydzewski and Nowak, 2017; Bruce et al., 2018; Kokh et al., 2018). The importance of determining association and dissociation rates (k_{on} and k_{off}, respectively), however, is not restricted to the field of drug design. In structural biology and biocatalysis, the

study of the thermodynamics and kinetics of binding and unbinding can be very important to attain a deep understanding of the biological processes of interest. There are well-known cases where the substrate binding or the product release are the rate-limiting steps in the catalytic cycle (Wang et al., 2001; Bosma et al., 2003; Yao et al., 2005). Interestingly, it has been shown that the substrate unbinding, under certain condition, may also have a positive impact on the enzymatic turnover (Reuveni et al., 2014).Therefore, the computational study of the (un)binding processes might reveal their kinetic and/or thermodynamics bottlenecks, and, in some cases, lead to finding improved biocatalysts for biotechnological applications.

The haloalkane dehalogenases (HLDs, E.C.3.8.1.5) are one of such cases. These bacterial enzymes can perform the hydrolytic conversion of halogenated aliphatic compounds into the respective alcohols (**Scheme 1A**). They have several practical applications, namely in the synthesis of enantiopure chemical compounds, recycling of by-products, bioremediation, and biosensing (Koudelakova et al., 2013). As several other haloalkanes, 1,2,3-trichloropropane (TCP) is an anthropogenic compound which sometimes ends up contaminating the groundwater as a recalcitrant toxic pollutant. Therefore, biodegradation would be a possible solution for the remediation of the contaminated sites (Samin and Janssen, 2012). The HLD from *R. rhodochrous*, DhaA, can only moderately hydrolyze TCP into 2,3-dichloropropan-1-ol (DCP). However, the 5-point mutant DhaA31 (**Figure 1**) has been reported to display a turnover number enhanced by 32-fold, resulting in a turnover number (k_{cat}) of 1.26 s^{-1} (Pavlova et al., 2009). DhaA31 is currently one of the best known HLDs in hydrolyzing TCP, and it has been included in the biodegradation pathway to stepwise convert the toxic TCP into glycerol (Dvorak et al., 2014; Kurumbang et al., 2014).

The HLDs have a buried active site connected to the surface by molecular tunnels (**Figure 1**). Their catalytic cycle (**Scheme 1B**) consists of: the substrate binding to the enzyme (1), rearrangement of the substrate in the catalytic site to form a reactive configuration (2), a multi-stage chemical step (3), and the release of the alcohol and halide products to regenerate the free enzyme (4). The chemical step involves an S_N2 attack of the D106 nucleophile on the electrophilic carbon atom of the substrate (DhaA numeration according to UniProt ID P0A3G2). The halide ion and the alkyl-enzyme intermediate are formed, the latter is attacked by a water molecule, activated by the catalytic base H272, which ultimately leads to the final products (Verschueren et al., 1993; Kutý et al., 1998; Marques et al., 2017). It is known that the hydrolysis of TCP by the wild-type DhaA (DhaAwt) is rate-limited by the S_N2 reaction, while in DhaA31 the slowest step is the release of DCP. This knowledge was attained from comparison of steady-state kinetic rates with

SCHEME 1 | Catalytic cycle of the HLDs. **(A)** Hydrolytic dehalogenation reaction catalyzed by the HLDs converting TCP to DCP. **(B)** Illustration of the respective catalytic cycle: (1) binding of TCP (orange and green balls) to the enzyme tunnels that connect the buried active site (pink cavity) to the bulk solvent; (2) formation of the pre-reactive Michaelis complex (the nucleophile D106 and the halide-stabilizing residues N41-W107 are represented in magenta and gray, respectively); (3) chemical steps converting TCP into DCP (cyan, green, and red balls) and Cl$^-$ (dark green ball), (4) release of the products to regenerate the free enzyme. The unbinding of DCP (step 4, highlighted with the blue arrows) is the kinetic bottleneck that limits the rate of DhaA31, and is the focus of this work. Adapted with permission from Marques et al. (2017). Copyright 2017 American Chemical Society.
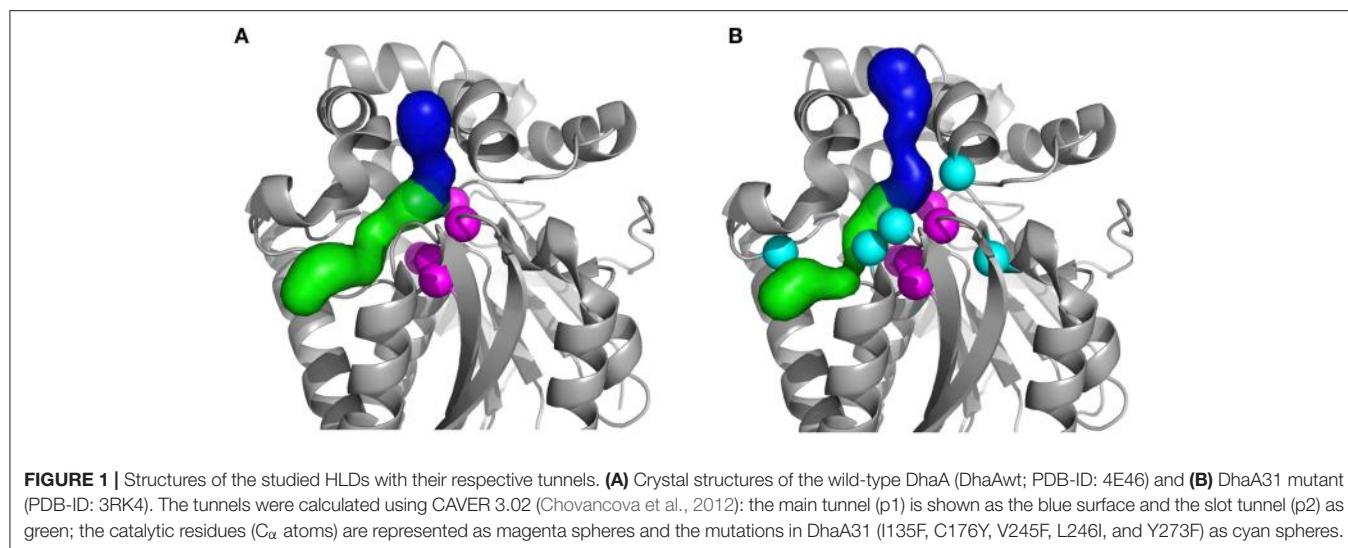
pre-steady-state rates (Pavlova et al., 2009; Marques et al., 2017). Moreover, the mutations C176F and V245F in DhaA31 contributed the most to the improvement of the S_N2 step toward TCP, whilst most of its bulky mutations—including C176F—narrow down the molecular tunnels and thus hinder the release of the alcohol product (Marques et al., 2017).

By accelerating the DCP unbinding from DhaA31 with mutagenesis, without hampering other steps in the catalytic cycle, we might improve the efficiency of DhaA31 to degrade TCP even further, which is desirable for biotechnological applications. We have recently targeted the geometric bottleneck of DhaA31's main tunnel, with mutations introduced to position 176, that had a high impact on the catalysis with different substrates (Kaushik et al., 2018). However, only minor improvements in the activity toward TCP were attained. The present work allows us to tackle the challenge from a new perspective.

Here we report a thorough computational study of the unbinding of DCP from DhaA31 and DhaAwt. Initially, we calculated the kinetic rates using metadynamics (MTD), and

FIGURE 1 | Structures of the studied HLDs with their respective tunnels. **(A)** Crystal structures of the wild-type DhaA (DhaAwt; PDB-ID: 4E46) and **(B)** DhaA31 mutant (PDB-ID: 3RK4). The tunnels were calculated using CAVER 3.02 (Chovancova et al., 2012): the main tunnel (p1) is shown as the blue surface and the slot tunnel (p2) as green; the catalytic residues (C$_\alpha$ atoms) are represented as magenta spheres and the mutations in DhaA31 (I135F, C176Y, V245F, L246I, and Y273F) as cyan spheres.

adaptive sampling under different simulation conditions. This helped us to assess the best procedures for predicting absolute and relative unbinding rates. Next, we performed free energy calculations using funnel-metadynamics (funnel-MTD), and calculated the energetic profiles of the product unbinding. This allowed us to compare the energy barriers, identify the thermodynamic bottlenecks, and thus predict several hot-spots for mutagenesis that could potentially improve the release of the DCP product and thus enhance the conversion of TCP by DhaA31.

MATERIALS AND METHODS

Metadynamics Unbinding Kinetics

System Preparation

The complexes of DhaA31 and DhaAwt bound with DCP and chloride products in their active site were prepared using the positions of DCP (only the R-enantiomer was studied here) docked into the corresponding crystal structures (PDB entries 3RK4 and 4E46, respectively), protonated and treated as previously described (Marques et al., 2017). The positions of the Cl$^-$ ion were taken from the respective crystal structures. The PREPI parameters for DCP were prepared using the Antechamber module of AmberTools 14 (Case et al., 2014), from the MOL2 structure containing the partial atomic charges, as previously calculated (Marques et al., 2017), and compiled using the atom types of GAFF force field. The topology and coordinates of the complexes hydrated only with the crystallographic water molecules were generated with tLEaP module of AmberTools 14, with the protein, and ions described by the AMBER ff12SB force field (Maier et al., 2015), and converted to the GROMACS format using the ACPYPE script (Sousa da Silva and Vranken, 2012). Each system was solvated with a cubic box of TIP3P water molecules (Jorgensen et al., 1983) with the edges at least 8 Å away from the protein atoms and then neutralized with Na$^+$ ions using the *editconf* module of GROMACS 5.0 package (Abraham et al., 2015).

System Equilibration

Energy minimization was performed with GROMACS 5.0.7 (Abraham et al., 2015) without restraints to relax the whole system, using the steepest descent method until the maximum force converged to values below 1 kJ/mol·nm with a maximum of 500 steps. The Particle Mesh Ewald method was used for the treatment of the long range non-bonding interactions beyond the 10 Å cut-off (Darden et al., 1993), and the periodic boundary condition was applied. Equilibration dynamics was run in two steps: a first equilibration of 500 ps in the isothermal-isobaric ensemble (*NPT*), at 1 atm, with the isotropic Berendsen barostat (Berendsen et al., 1984), and coupling constant 0.2 ps, and a second one of 1 ns in the isothermal-isochoric ensemble (*NVT*). Both steps were conducted at 300 K with the velocity-rescaled Berendsen thermostat, to ensure the proper canonical ensemble (Bussi et al., 2007), with constant for coupling of 0.1 ps. All simulations were performed with the periodic boundary conditions in all directions, the Verlet pair-lists scheme (Verlet, 1967) with cut-off values of 10 Å for both short-range coulombic and van der Waals potentials, and the LINear Constraint Solver (LINCS) (Hess et al., 1997) algorithm to constrain the bonds and eliminate drifts. The integration time step was 2 fs and the energy and coordinates of the system were recorded every 1 ps.

Setup of the Collective Variable

A path-based collective variable (path CV) was defined to describe the release process of DCP along the p1 tunnel, according to the formalism as previously described (Branduardi et al., 2007; Bonomi et al., 2008). It involves a distance *s* along a reference path that leads from state A (the fully bound state, the docked conformation in the active site) to B (fully unbound state, with DCP in the bulk solvent). The path was constructed based on several snapshots selected from previous accelerated molecular dynamics (aMD) simulations with DhaA31 and DhaAwt (Marques et al., 2017) to have DCP at different distances and orientations between states A and B. In total 9 frames were chosen for each system, and only

the ligand and the residues of p1 tunnel in contact with DCP during the release were selected as the path reference (**Figure S1** and **Table S1**). The path CV (hereafter termed *p3*) was then defined by the root mean square deviation (RMSD) space. From a further analysis of a set of unbinding metadynamics simulations, it was found that the direct variable of the path *s* (named *p3.sss*) was degenerated and hence was not suitable to be used alone in this study. The degeneracy was lifted using a second CV, which was the distance (*d1*) between the center of mass of DCP to the active site cavity, defined by the center of mass of the atoms Y176-C_β, F205-C_α, L209-C_α, and H272-C_α for DhaA31, and C176-C_β, F205-C_α, L209-C_α, and H272-C_α for DhaAwt. The λ parameter was set to 92 for the DhaA31/DCP and 100 for DhaAwt/DCP. The values of λ were obtained from the analysis of the RMSD matrix obtained from the frames.

Infrequent Metadynamics Simulations

All metadynamics (MTD) simulations were performed using PLUMED (Tribello et al., 2014) plugin, version 2.2.3 with the GROMACS 5.0.7 (Abraham et al., 2015) code. The *NVT* ensemble at 300 K was used as in the equilibration, with further position restraints on the atoms Leu36-C_α, Ile104-C_α and Leu237-C_α with harmonic constant of 2.38 kcal/mol·Å^2 (1,000 kJ/mol·nm^2) in each dimension, to prevent drifting of the protein across the periodic cell. The potential biases were added to the path CV *s* dimension and the distance *d* variables, deposited every 50 ps, with initial height of 0.60 kcal/mol (2.50 kJ/mol) for both variables. The Gaussian widths (σ) for *s* and *d1* were set, respectively, as 0.05 and 0.014 Å for DhaA31/DCP and 0.07 and 0.013 Å for DhaAwt/DCP, and a decay corresponding to a bias factor of 10. In total 25 independent infrequent MTD simulations were run until the ligand was released to reach distances *d1* > 22 Å from the active site without immediate rebinding. These times corresponded to the biased release times, *t*$_{biased}$. The trajectories were visualized using VMD 1.9.1 (Humphrey et al., 1996) and PyMOL 1.7.4 (The PyMO L Molecular Graphics System, 2014).

To obtain the unbiased release time *t*$_{unbiased}$, the acceleration factor α was used as describe by Equations 1, 2 (Tiwary and Parrinello, 2013; Tiwary et al., 2015):

$$\alpha = \langle e^{\frac{V(r,t)}{k_B T}} \rangle \tag{1}$$

$$t_{unbiased} = t_{biased} \times \alpha \tag{2}$$

where $\langle\ \rangle$ denotes the running average accumulated through the course of the simulation at the biased time, *t* or *t*$_{biased}$, *V*(*r,t*) is the time-dependent metadynamics bias, *r* the set of CV descriptors, and $k_B T$ is the temperature in energy units, which has the value of 2.50 kJ·mol^{-1} at 300 K.

Calculation of k_{off} From Metadynamics

The estimation of the unbinding rates k_{off} involves the calculation of the characteristic transition time τ of a Poisson process, obtained through a least squares fitting of the empirical cumulative distribution function (ECDF) obtained with the metadynamics unbiased times, *t*$_{unbiased}$, with the theoretical expression of a cumulative distribution function (TCDF), which in the case of a homogeneous Poisson process is given by Equation (3) (Tiwary et al., 2015):

$$TCDF = 1 - e^{-\frac{t}{\tau}} \tag{3}$$

The theoretical (TCDF) and empirical (ECDF) distributions are compared by a Kolmogorov-Smirnoff test, which estimates an associated *p-value*, which represents the probability that the distribution of times extracted from metadynamics is obtained from the theoretical exponential distribution, and describes the quality of the data (Salvalaglio et al., 2014; Tiwary et al., 2015). Acceptable distributions should always present *p-value* >0.05, otherwise the set of results is discarded. To perform the fitting of those distributions, the Kolmogorov-Smirnov test, and calculate the dissociation transition time τ_{off}, the **STPtest.m** Matlab script was used as provided (Salvalaglio et al., 2014). The dissociation rate k_{off} was then calculated from τ_{off} by the Equation 4:

$$k_{off} = \frac{1}{\tau_{off}} \tag{4}$$

The error associated with the calculated k_{off} value was estimated by a bootstrap analysis of the data set of unbiased release times obtained for each system. This was performed by re-analyzing 500 sub-samples extracted randomly from the original ensemble of release times.

Adaptive Sampling Kinetics
System Preparation

The complexes of DhaA31 and DhaAwt bound with DCP and chloride products in their active site, hydrated with the crystallographic waters, were prepared as described for the metadynamics. Na$^+$ and Cl$^-$ ions were added in order to achieve ionic strength of 0.1 M, and a TIP3P (Jorgensen et al., 1983) cubic box of water molecules with the edges 10 Å distant from the original system was added. The topology and coordinates of the hydrated complexes were generated with tLEaP module of AmberTools 14 (Case et al., 2014), with the protein and ions described with the ff12SB AMBER force field (Hornak et al., 2006; Joung and Cheatham, 2008, 2009; Nguyen et al., 2014). For comparison testing different simulation conditions, the systems were also prepared with force field ff14SB (Maier et al., 2015) and OPC3 water model (Izadi and Onufriev, 2016).

System Equilibration

The systems were equilibrated using the *Equilibration_v2* module of high-throughput molecular dynamics (HTMD) (Doerr et al., 2016). The system was first minimized using conjugate-gradient method for 500 steps. Then the system was heated and minimized as follows: (I) 500 steps (2 ps) of *NVT* equilibration with the Berendsen barostat to 298 K, with constraints on all heavy atoms of the protein, (II) 625 000 steps (2.5 ns) of NPT equilibration with Langevin thermostat with 1 kcal·mol^{-1}·Å^{-2} constraints on all heavy atoms of the protein, and (III) 625 000 steps (2.5 ns) of NPT equilibration with the Langevin thermostat without constraints. During the equilibration simulations, holonomic

constraints were applied on all hydrogen-heavy atom bond terms and the mass of the hydrogen atoms was scaled with factor 4, enabling the simulations to run with 4 fs time steps (Feenstra et al., 1999; Harvey and De Fabritiis, 2009; Harvey et al., 2009; Hopkins et al., 2015). The simulations employed periodic boundary conditions, using the particle mesh Ewald method for treatment of interactions beyond 9 Å cut-off, electrostatic interactions suppressed for more than 4 bond terms away from each other and the smoothing and switching van der Waals and electrostatic interaction cut-off at 7.5 Å (Harvey and De Fabritiis, 2009).

Adaptive Sampling

The HTMD was used to perform adaptive sampling of the RMSD of the Cα atoms. The 20 ns production molecular dynamics (MD) runs were started with the system resulting from the equilibration cycle and employed the same settings as the last step of the equilibration. The trajectories were saved every 0.1 ns. Adaptive sampling was performed using the distance between the central C-2 atom of DCP and the Cγ atom of the catalytic nucleophile D106 as the reaction coordinate, and a time-based independent component analysis (TICA) (Naritomi and Fuchigami, 2011) in 1 dimension. Unless stated otherwise, 40 epochs of 10 MDs each were performed for DhaA31 and 30 epochs for DhaAwt, corresponding to cumulative simulation times of 8 and 6 μs, respectively.

Markov State Model Construction

The simulations were made into a simulation list using HTMD method and water was filtered out, and unsuccessful simulations with length <20 ns were omitted. This resulted in 8 μs of simulation time (400 × 20 ns) for DhaA31 and 6 μs of simulation time (300 × 20 ns) for DhaAwt (Doerr et al., 2016). The DCP dynamics was studied by the distance between the C-2 atom of DCP and the Cγ atom of the catalytic nucleophile D106. The data was clustered using MiniBatchKmeans algorithm to 200 clusters. 15 ns lag time was used in the models to construct 3 Markov states, and the Chapman-Kolmogorov test was performed to assess the quality of the constructed states. A bootstrapping calculation was performed with 80% of the data and repeated 500 times to estimate the errors in the estimated kinetic parameters.

Funnel Metadynamics

The MTD simulations were performed using the GROMACS 5.0.7 (Abraham et al., 2015) patched with the PLUMED plugin (Tribello et al., 2014), version 2.2.3, modified to include the funnel metadynamics (funnel-MTD) algorithm and used as provided by the authors of the method (Limongelli et al., 2013). The NVT ensemble at 300 K was used as previously, with the further position restraints on the Leu36-C$_α$, Ile104-C$_α$, and Leu237-C$_α$ atoms with a harmonic force constant of 59.8 kcal/mol·Å^2 (25,000 kJ/mol·nm^2) in each dimension to prevent the protein from drifting across the periodic cell. These atoms were chosen for being buried and having some of the lowest B-factors in the respective crystal structures. The potential biases were added to the path CV s variable, deposited every 1 ps, with initial height of 0.60 kcal/mol (2.50 kJ/mol). The Gaussian

width (σ) was 0.05 Å for DhaA31/DCP and 0.07 and 0.013 Å for DhaAwt/DCP, as previously, and a decay corresponding to a bias factor of 10. A funnel-shaped restraint with 83.6 kcal/mol.A^2 (35,000 kJ/mol.nm^2) force constant, was defined by the Z axis passing through the points A—the coordinates of the D106-C$_α$ atom—and B—the geometric center of the F144-C$_α$, F152-C$_α$, A167-C$_α$, and K175-C$_α$ atoms –, the $α$ angle of 0.55 rad, Z_{cc} 20.0 Å, and R_{cyl} 5.0 Å. To prevent the ligand from crossing the periodic cell, an upper distance restraint with 12.0 kcal/mol.A^2 (5,000 kJ/mol.nm^2) force constant was imposed at 23 Å from point A. The free energy surface (FES) was computed using the SUM_HILLS module of PLUMED, from the histogram distribution reweighted from the biases added by the metadynamics (Bonomi et al., 2009; Tiwary and Parrinello, 2015). The FES was reanalyzed for the variable $d1$, defined above, using the DRIVER module of PLUMED. The histogram reweighting was performed by taking into account all the biases from the metadynamics and the restraints. The relevant states were selected from the FES, and the simulation snapshots corresponding to $d1$ values within ± 0.01 Å were extracted using GROMACS 5.0.7 (Abraham et al., 2015), in PDB format, after the trajectory was aligned by the C$_α$ atoms.

Binding Free Energy

The free energy of binding was calculated by the molecular mechanics/generalized Born solvent accessible surface area (MM/GBSA) method to determine the interaction energy of DCP with the protein residues in each one of the ensembles obtained from the selected states of the FES. For that, the topology of the systems in the GROMACS format were converted to the AMBER format using the ParmEd program (Swails, 2010). The *ante-MMPBSA.py* (Miller et al., 2012) module of AmberTools 14 (Case et al., 2014) was used to remove the solvent and ions from the resulting topology files and define the Born radii as *mbondi2*, and generate the corresponding topologies of the *complex*, *receptor*, and *ligand*, to be used in the MM/GBSA calculations. The state ensembles, in PDB format, were manually stripped from any ions and solvent. The *MMPBSA.py.MPI* (Miller et al., 2012) module of AmberTools 14 was used to calculate, in parallel, the mean free energy of binding considering every snapshot of the PDB ensemble. The generalized Born method was used (*&gb* namelist) with implicit generalized Born solvent model (*igb=8*) and 0.1 M ionic strength (*saltcon=0.1*). Decomposition of the pairwise interactions were generated (*&decomp* namelist) with discrimination of all types of energy contributions (*idecomp=4*) for the whole residue (*dec_verbose=0*).

CaverDock Simulations
Tunnel Calculations

CAVER 3.02 (Chovancova et al., 2012) was used to calculate the tunnels in the protein structure of the DhaA31 and DhaAwt, as previously prepared and treated prior to the metadynamics. The tunnels were calculated using a probe radius of 0.7 Å, a shell radius of 3 Å, and shell depth 4 Å. The starting point for the tunnel calculation was the same point in the active site as previously used to calculate the distance $d1$ (center of mass of the

atoms Y176-C$_\beta$, F205-C$_\alpha$, L209-C$_\alpha$, and H272-C$_\alpha$ for DhaA31, and C176-C$_\beta$, F205-C$_\alpha$, L209-C$_\alpha$, and H272-C$_\alpha$ for DhaAwt).

CaverDock Calculations

CaverDock package ("CaverDock," 2018; Filipovič et al., 2018) was used to calculate the trajectories of DCP through the p1 tunnel of DhaA31 and DhaAwt, as calculated by CAVER. The input files previously prepared for the receptors and ligand, respectively, in PDB or MOL2 format, were converted to the AutoDock Vina-compatible PDBQT format using the MGLTools v1-5-7rc1 (Morris et al., 2009), preserving the previously calculated partial charges for the ligand. The tunnels were extended by 6 Å and discretized with 0.3 Å increments. The ligand started in the active site and it was moved toward the protein surface. The side chain flexibility was iteratively introduced using the default settings, with two automatically chosen tunnel residues made flexible per iteration.

RESULTS

Metadynamics Unbinding Kinetics

The unbinding of the alcohol product has become the rate-limiting step in the catalytic conversion of TCP by DhaA31, and that step is expected to be slower than with the wild-type DhaAwt (Pavlova et al., 2009; Marques et al., 2017). Aiming at verifying this computationally, we have used a state-of-the-art method for the calculation of the kinetic rates of ligand unbinding, which is based on metadynamics (Tiwary et al., 2015).

The metadynamics (MTD) relies on a set of collective variables (CVs) to describe the system and the process under study. Here we have made use of the knowledge acquired from a previous computational work (Marques et al., 2017) to define the unbinding path of DCP. In that study, the release of DCP from the active site was observed with both DhaA31 and DhaAwt, always through the main tunnel (p1 tunnel; see **Figure 1**), and these simulations were used to provide several frames for describing the path CV for those two systems. A discussion on the optimization of the CV is presented in **Discussion S1** in **Supplementary Material**.

Twenty five MTD simulations were performed with each protein containing DCP in their active site, and run until DCP was fully released to the bulk solvent (distance $d1 > 22$ Å from the active site) without immediate rebinding to the tunnel. The release times obtained from the MTD simulations were converted to unbiased times, and the results showed a large dispersion of the release times for each system. But, as expected, in average DhaAwt released DCP significantly faster than DhaA31 (**Figure 2** and **Figure S2**). The unbiased release times were fitted to determine the distribution and probabilities of the transitions (the ligand unbinding; see **Figure S3**). These analyses resulted in *p-values* above the minimum confidence threshold of 0.05 (**Table S2**), and the average dissociation times (τ_{off}) and kinetic constants (k_{off}) were obtained (**Table 1**). The evolution analysis of the τ_{off} with the number of simulation runs showed that the estimation of the dissociation time was well-converged even using lower number of simulations (**Figure S4**). With transition times in the range of microsecond timescales,

the predicted k_{off} value was 36-fold faster for DhaAwt than for DhaA31 (**Table 1**). Such trend is in agreement with the previous findings that DCP was significantly more prone to be released from DhaAwt than from DhaA31 (Marques et al., 2017).

Adaptive Sampling Kinetics

At this point we wanted to validate the kinetic rates previously calculated with the MDT approach by using another advanced and independent method. Thereby we can compare and assess the reproducibility of the kinetic predictions using different methods. So, we decided to perform high-throughput molecular dynamics (HTMD) using the adaptive sampling technique in combination with Markov state models (MSMs). This method allows us to obtain the transition matrix between the states and thus predict the kinetic rates of unbinding (Doerr et al., 2016).

Initially, the adaptive simulations were performed using the same force field and water model as the MTD simulations (ff12SB and TIP3P). We found that the distance between the ligand and the catalytic nucleophile D106 (defined as described in the methods) was a good metrics for calculating the Markov state models for describing the events that we wanted to survey (the release of DCP from DhaA31 and DhaAwt). We defined 3 Markov states which were satisfying when visually inspected: one state corresponded to DCP located in the active site, an intermediate state with DCP in the main access tunnel (p1), and the unbound state with DCP outside the protein (**Figure 3**). The Chapman-Kolmogorov test was performed to assess the quality of the Markov state models, which revealed satisfying for the parameters used (**Figures S5–S7**). The kinetic rates between the fully bound and fully unbound states were calculated (**Table 1**). The default simulation time was 8 μs for DhaA31 and 6 μs for DhaAwt, which generally proved satisfactory according to the errors obtained by bootstrapping. The release rates of DCP (k_{off}) obtained from the current adaptive sampling method (**Table 1**, column "ff12SB+TIP3P") were 1-2 orders of magnitude higher than the ones previously calculated with the MTD method. Regarding the relative values of k_{off} values between DhaAwt and DhaA31, the order is maintained but with smaller difference between the two enzymes. DCP was released 2.5 times faster from DhaAwt than from DhaA31, which is only moderately in agreement with the MTD results (36 times faster for DhaAwt) and previous computational evidence (Marques et al., 2017). Regarding the remaining kinetic parameters, DhaA31 seemed to be more prone for rebinding DCP than DhaAwt, showing higher k_{on} and lower K_d values than DhaAwt. This is not in agreement with the experimental data, which showed higher K_d for DhaA31 than for DhaAwt (**Table 1**). The free energy of the bound state, compared to the unbound state, was -1.76 ± 0.19 kcal/mol for DhaA31, and -0.54 ± 0.34 kcal/mol for DhaAwt. This means that DCP's bound state in DhaA31 is thermodynamically more stable than that of DhaAwt by 1.22 kcal/mol. Furthermore, the proteins remained stable throughout the simulations, as can be inferred from the RMSD plots (**Figure S8**).

Due to the dissimilarity between the values and the kinetic rates' ratios obtained from MTD and adaptive sampling, we decided to perform a similar study under different simulation condition. For that, we varied the force field (ff14SB instead of
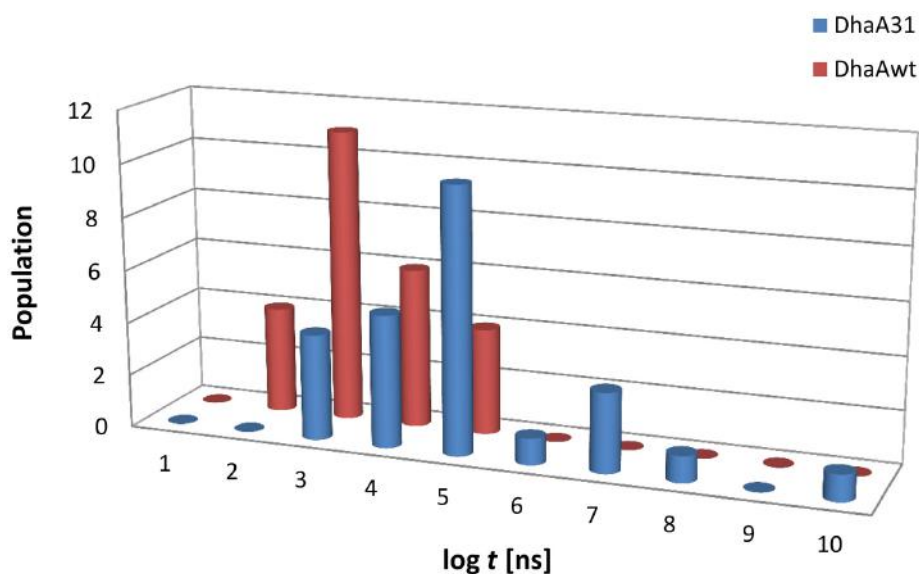
FIGURE 2 | Histogram distribution of the unbiased release times (log t) of DCP from the buried active sites of **(A)** DhaA31 and **(B)** DhaAwt. The release times were obtained from the infrequent metadynamics simulations (25 runs were performed for each system).

TABLE 1 | Experimental and theoretical kinetic parameters obtained from the metadynamics and adaptive sampling simulations of DCP with DhaA31 and DhaAwt for all the tested force fields and water models.

	Parameter[a]	Experimental	Metadynamics	Adaptive sampling		
			ff12SB+TIP3P[b]	ff12SB+TIP3P[b]	ff14SB+TIP3P[b]	ff14SB+OPC3[b]
DhaA31	τ_{off} [ns]		$3.5 \pm 1.4 \times 10^4$	188 ± 28	350 ± 49	$2.2 \pm 1.1 \times 10^3$
	k_{off} [s^{-1}]		$2.8 \pm 1.2 \times 10^4$	$5.32 \pm 0.72 \times 10^6$	$2.86 \pm 0.38 \times 10^6$	$4.5 \pm 1.3 \times 10^5$
	τ_{on} [ns]			$1.74 \pm 0.45 \times 10^3$	$5.1 \pm 3.9 \times 10^3$	$5.7 \pm 2.6 \times 10^4$
	k_{on} [M^{-1}s^{-1}]			$1.28 \pm 0.34 \times 10^8$	$4.4 \pm 1.4 \times 10^7$	$3.9 \pm 1.5 \times 10^6$
	k_{off}/k_{on} [M]			0.042 ± 0.014	0.065 ± 0.053	0.116 ± 0.073
	K_d [M]	$> 0.020^c$		0.057 ± 0.017	0.12 ± 0.10	0.77 ± 0.48
	ΔG_{eq}^0 [kcal/mol]			-1.76 ± 0.19	-1.33 ± 0.3	-0.16 ± 0.33
DhaAwt	τ_{off} [ns]		$9.9 \pm 5.6 \times 10^2$	75 ± 11	73.0 ± 5.3	163 ± 29
	k_{off} [s^{-1}]	$>>10^{3d}$	$1.01 \pm 0.56 \times 10^6$	$1.33 \pm 0.18 \times 10^7$	$1.37 \pm 0.09 \times 10^7$	$6.2 \pm 1.2 \times 10^6$
	τ_{on} [ns]			$9.3 \pm 7.1 \times 10^3$	$1.72 \pm 0.54 \times 10^3$	337 ± 127
	k_{on} [M^{-1}s^{-1}]			$2.4 \pm 1.4 \times 10^7$	$1.33 \pm 0.27 \times 10^8$	$6.7 \pm 2.9 \times 10^8$
	k_{off}/k_{on} [M]			0.54 ± 0.42	0.103 ± 0.035	$9.1 \pm 4.0 \times 10^{-3}$
	K_d [M]	$0.95 \pm 0.34 \times 10^{-3}$		0.42 ± 0.38	0.132 ± 0.030	$8.3 \pm 2.5 \times 10^{-3}$
	ΔG_{eq}^0 [kcal/mol]			-0.54 ± 0.34	-1.25 ± 0.13	-2.95 ± 0.18
DhaAwt/31	Rel. k_{off}		36	2.5	4.8	14

[a]τ_{off}, mean dissociation transition time; k_{off}, dissociation rate; τ_{on}, mean association transition time; k_{on}, association rate; K_d, equilibrium dissociation constant; ΔG_{eq}^0, free energy difference between bound and unbound states; rel. k_{off}, DhaAwt/DhaA31 ratio of k_{off} rates; [b]force field and water model used; [c]solubility concentration; [d]detection limit of the instrument. The variability of the parameters is the SD obtained from a bootstrap analysis.

ff12SB), and the water model (OPC3 instead of TIP3P). The k_{off} values obtained with ff14SB were lower than with ff12SB for both systems (**Table 1**), suggesting that the dynamic properties with ff14SB are slower than with ff12SB. When the OPC3 water model was used, slower unbinding rates were observed, as compared to those obtained with TIP3P. Other kinetic parameters were also significantly affected by the force field and solvent model, in some cases differing by orders of magnitude (namely k_{on}, K_d, and

ΔG_{eq}^0). The K_d values predicted with ff14SB+OPC3 approached the experimental ones more than the other conditions (8.3 ± 2.5 mM prediction vs. 0.95 mM experimental for DhaAwt). In some cases, the dynamic behavior of the systems changed so significantly that the initial simulation times (8 μs for DhaA31 and 6 μs for DhaAwt) did not provide enough sampling to produce precise estimations. This occurred for DhaA31 with ff14SB+OPC3 and for DhaAwt with ff14SB+TIP3P. In these
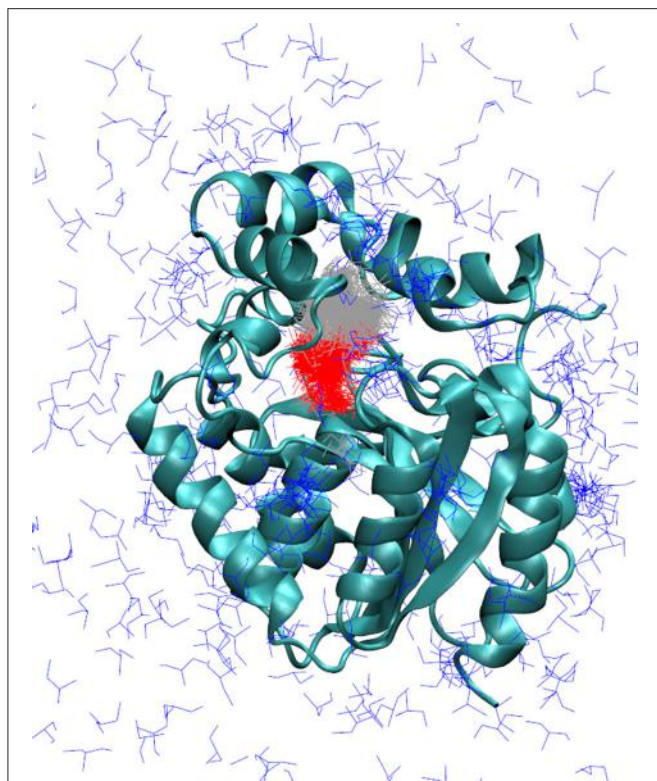
FIGURE 3 | Markov state models describing the release of DCP in the simulations with adaptive sampling. Representative models of DCP with DhaA31: DCP fully bound in the active site (red), intermediate state with DCP bound in the access tunnel (gray), and unbound state (blue). The enzyme is represented as cyan cartoon. The Markov models obtained with DhaAwt were similar and are not shown here.

cases, the simulations were run twice as long (16 μs for DhaA31 and 12 μs for DhaAwt), although for DhaA31 even a longer simulation time might be required due to the timescales of the observed events. Overall, the ff14SB+OPC3 combination seemed to produce more accurate results, with predicted K_d values closer to the experimental ones.

Pre-steady State Kinetics

To validate the findings of the present computational study, the calculated kinetic properties were compared with the previously reported results from transient kinetic measurements.

The basis for the current work is that the rate-limiting step in the catalytic conversion of TCP by the DhaA31 mutant is the product release, i.e., unbinding of DCP from the protein to the bulk solvent. This conclusion has been made based on comparison of steady-state kinetic rates with results from the transient kinetics measurements (Pavlova et al., 2009). In this study we observed, that after a rapid mixing of DhaA31 with excess of TCP, there was a burst of both DCP and chloride, followed by a linear steady-state phase with the rate constant 1.36 $\pm$ 0.18 s^{-1} for DCP. This rate is in a very good agreement with the value from the steady-state kinetics k_{cat} 1.26 $\pm$ 0.07 s^{-1}. Further studies showed that the release of the halide was a fast process,
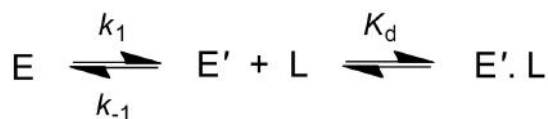
SCHEME 2 | Kinetic scheme showing the binding of DCP (L) to the enzyme (E). The transition of E to E' represents a conformational change in the enzyme.

$$k_{obs} = k_1 + \frac{k_{-1} \cdot K_d}{[L] + K_d} \tag{5}$$

which allowed us to conclude that the unbinding of DCP is rate-limiting for DhaA31, while DhaAwt is limited by the catalytic step (Marques et al., 2017).

Binding experiments of DCP were also carried out with DhaA31 and DhaAwt using stopped-flow fluorescence. Unfortunately, these experiments proved unsuccessful with DhaA31 due to the very low affinity of this enzyme for DCP, for which no binding was observed at concentrations near the solubility limit. This also implies a high dissociation constant, with value $K_d > 20$ mM. For DhaAwt, the fluorescence curves revealed a slow kinetics profile upon mixing with DCP that could be associated with a single exponential. This was an indication that a slow conformational change of the enzyme preceded the fast binding of DCP (**Scheme 2** and Equation 5). This fact, together with the time scale limitations of the instruments, disallowed the calculation of the binding, and unbinding rates of DCP, k_{on} and k_{off}, but the equilibrium dissociation constant was obtained as $K_d = 0.95 \pm 0.34$ mM. Moreover, the ratio between the enzyme conformations in equilibrium, E and E', favor the nonbinding form (E) by 2:1, with $k_1 = 3.31 \pm 0.27$ s^{-1} and $k_{-1} = 6.16 \pm 0.42$ s^{-1} (Marques et al., 2017).

Free Energy Calculations

The free energy profiles allow a deep understanding of the thermodynamic and kinetic determinants of individual steps of the catalytic cycle, since they reveal the energy of the different states and the energy barriers to the required transitions. Therefore, their study can be important to asses not only the differences between the unbinding of DCP from DhaA31 and DhaAwt, but also to understand how the current bottlenecks can be overcome in the scope of protein engineering.

The funnel-metadynamics (funnel-MTD) is a method that allows the efficient calculation of the free energy surface (FES) of the process of ligand (un)binding (Limongelli et al., 2013). In this method, a funnel-shaped restraint prevents the ligand from drifting away and thus allows sampling several forward and reverse events (unbinding/binding), needed for a correct estimation of the free energy profile associated with the process. The funnel restraints used in this study (**Figure S9**) were defined iteratively in order to allow the free motion of the ligand within the active site, main tunnel (p1), and respective tunnel mouth. The funnel-MTD simulations were performed using the same path CV as previously used in the MTD unbinding kinetics. The simulations were run until the free energy achieved convergence.

In these simulations DCP was released (reaching distances $d1 > 15$ Å) and rebound to the active site (reaching $d1 < 5$ Å) several times with each system, as desirable for the free-energy calculations (**Figure S10**). In both cases, the proteins remained stable throughout most of the simulations, with only a steep increase in the RMSD of DhaAwt backbone between 100 and 110 ns, which was reversible to levels below 1.2 Å (**Figure S11**).

The free energy was primarily calculated with respect to the original CV used in the MTD simulations, and computed by reweighting the histogram distribution with the biasing Gaussians added to the system. this resulted in the respective one-dimensional FES (Bonomi et al., 2009; Tiwary and Parrinello, 2015). To analyze the convergence, the FES was calculated cumulatively by taking into account an increasing number of snapshots. Two energy basins were integrated and the difference between those basins was plotted as a function of the increasing simulation time (**Figure S12**). Similarly, the energy barrier between the same two energy basins was measured and plotted as a function of the simulation time (**Figure S13**). Altogether, this analysis allowed us to conclude that the simulations were well-converged after the respective running times (500 ns for DhaA31 and 400 ns for DhaAwt).

The FES was projected against the distance of DCP to the active site (**Figure 4A**). For simplicity, the global minimum of each FES was adjusted to 0 kcal/mol. In both systems, the global minimum corresponded to some region in the middle of the p1 tunnel ($d1 \approx 5$–6 Å), and not in the active site. From a previous study (Marques et al., 2017), we know that the length of p1 tunnel varies between ≈ 9 and 13 Å (associated with $d1$), which can roughly define the limits of the tunnel mouth (**Figure 1**). This corresponds well with the local minima at ≈ 12 Å for DhaA31, and ≈ 10 and 13 Å for DhaAwt (**Figure 4**). Two-dimensional FES profiles can be projected for any set of parameters, which might be useful to assess the potential degeneracy of the CV used (e.g., path CV and $d1$, **Figure S14**). DhaA31 presented one very steep energy barrier of 4.81 kcal/mol between the global minimum (at 5.94 Å) and a second minimum at the tunnel mouth (12.14 Å), while DhaAwt seems to have a rather smoother and stepwise transport process with two lower energy barriers (2.26 and 1.46 kcal/mol) to reach the tunnel mouth. This fact can have a strong impact on the product release kinetics (k_{off}), since the transitions between states with lower barriers can occur exponentially faster. This result is consistent with the kinetic studies described above that showed slower DCP unbinding rates with DhaA31 than with DhaAwt.

## Structural Analysis of the Free Energy Landscapes

Here we aimed at understanding structurally the meaning of the different states if the calculated FES and identifying potential hot-spots for improving the unbinding rate of DCP from DhaA31. The FES calculated for the release of DCP through the main tunnel of DhaA31 and DhaAwt have in common quite similar locations of the global minima and their shapes for longer distances. However, they differ significantly in the number of local minima and the heights of the different energy barriers

along the pathway. Several relevant stages of the FES have been identified, and the respective simulation frames extracted (**Table S3**). The respective ensembles can be considered as representative structures of the main states of the systems along the process of DCP unbinding through the main tunnels of DhaA31 and DhaAwt (**Figure 4**). One first observation reveals that DCP was more confined within DhaA31 than DhaAwt, where it was more especially at the first transition state TS1 (**Figure 4**). Moreover, the states with $d1 \geq 13$ Å contain DCP outside of the tunnel, where it forms interactions with the residues at the tunnel mouth, before it can be fully released to the bulk solvent (last state, for $d1 \geq 20$ Å). A closer look at the enzymes' structures during the simulations revealed that some of the tunnel-lining residues were highly flexible and presented diverse conformations, which allow the ligand transport through the tunnel. One of such residues is F149, which clearly had two states, observed in both DhaA31 and DhaAwt: (i) the aromatic ring either pointed toward the middle of tunnel, or (ii) it pointed toward the side of the structure under the α4 helix (**Figure S15**). Because of these two conformations, F149 may play the role of gatekeeper to the transport of ligands through the p1 tunnel.

The measurement of the interactions formed by DCP with each residue may provide a quantitative assessment of what was discussed before and confirm the pivotal role of some residues during the unbinding process. Therefore, the free energy of binding (ΔG_{bind}) of DCP was calculated for the structural ensembles. The average interaction energies were calculated for the global minimum energy (**Figure S16**) and for the TS1 clusters (**Figure S17**). As expected, at the minimum energy the residues at the tunnel bottleneck dominated the interactions with DCP. The high standard deviations (SD) found for several residues reflect how diverse the structures within the same cluster were. We also tried to assess which residues contributed the most to prevent the transition from the global minima to TS1 due to the strength of their interactions with DCP. For that, we calculated the difference in binding energy between those two states. For DhaA31, the residues with the strongest influence (most negative $\Delta\Delta G_{bind}$) were F152 > F168 > F149 > F245 (**Figures 5**, **S18**). These are potential hot-spots for decreasing the energy barrier in DhaA31 and thus improve the unbinding rate of DCP. The residues F144, T148, and K175 form strong interactions in TS1, and might also reveal interesting hot-spots for mutagenesis (**Figure 5**). The hypothesis here is that the energy barrier may be lowered by increasing the interactions at TS1.

CaverDock Calculations

Here we wanted to test the ability of a computationally cheap method for predicting the release of DCP from the buried active sites of DhaA31 and DhaAwt, and compare the results with those obtained from the robust methods used above. CaverDock (CaverDock, 2018; Filipovič et al., 2018) was selected for that task. This is a computer program developed for the rapid prediction of the trajectory and energy profile of a ligand being transported through a molecular tunnel. This tool, based on molecular docking, can be used for the fast assessment or high-throughput screening of potential substrates, drugs or metabolites that are
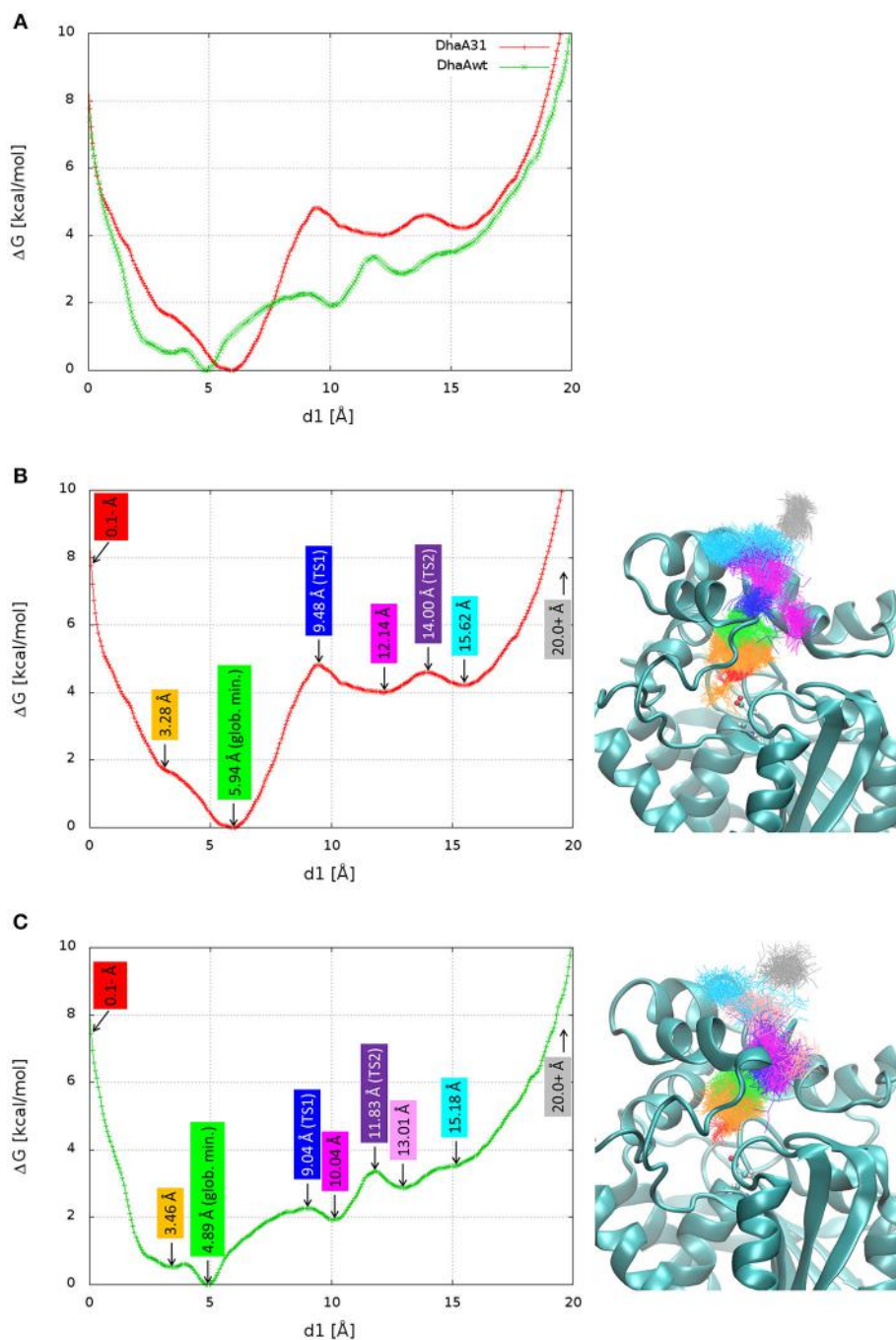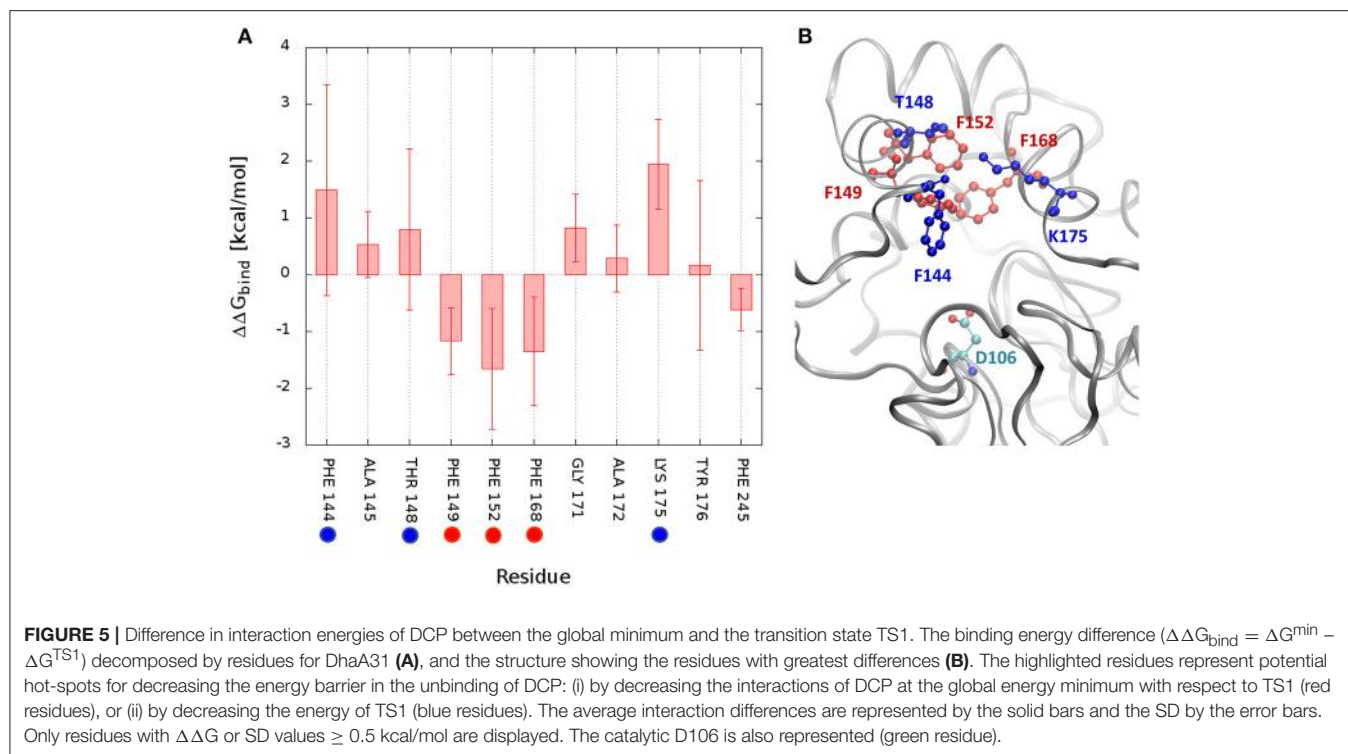
FIGURE 4 | Free energy surface (FES) and structures of the relevant stages in the release of DCP. The FES projected on the distance of DCP to the active site (*d1*) superimposed for DhaA31 and DhaAwt **(A)**, and dissected with the relevant states for DhaA31 **(B)** and DhaAwt **(C)**. The relevant states are labeled with the corresponding *d1* values (left), and the respective clusters are shown by the superimposed DCP structures (right), represented as lines of the same colors. The protein is represented as cyan cartoons and the catalytic D106 as ball-and-sticks. Glob. min. global minimum; TS1, transition state 1; TS2, transition state 2.

expected to bind, or be transported through the tunnels of biomolecules (Pinto et al., 2018; Vávra et al., 2018).

CaverDock was used here to predict the trajectories and energy profiles of DCP through p1 tunnel of DhaA31 and DhaAwt, which were compared with the ones obtained from the robust MTD method. The results showed that the two enzymes have energy minima with similar binding energy ($\Delta E_{bind} = -4.1$ kcal/mol), located at the active site instead of the middle of the tunnels (**Figure 6** and **Table S4**), which is in contrast with the FES obtained from the funnel-MTD. When the calculations
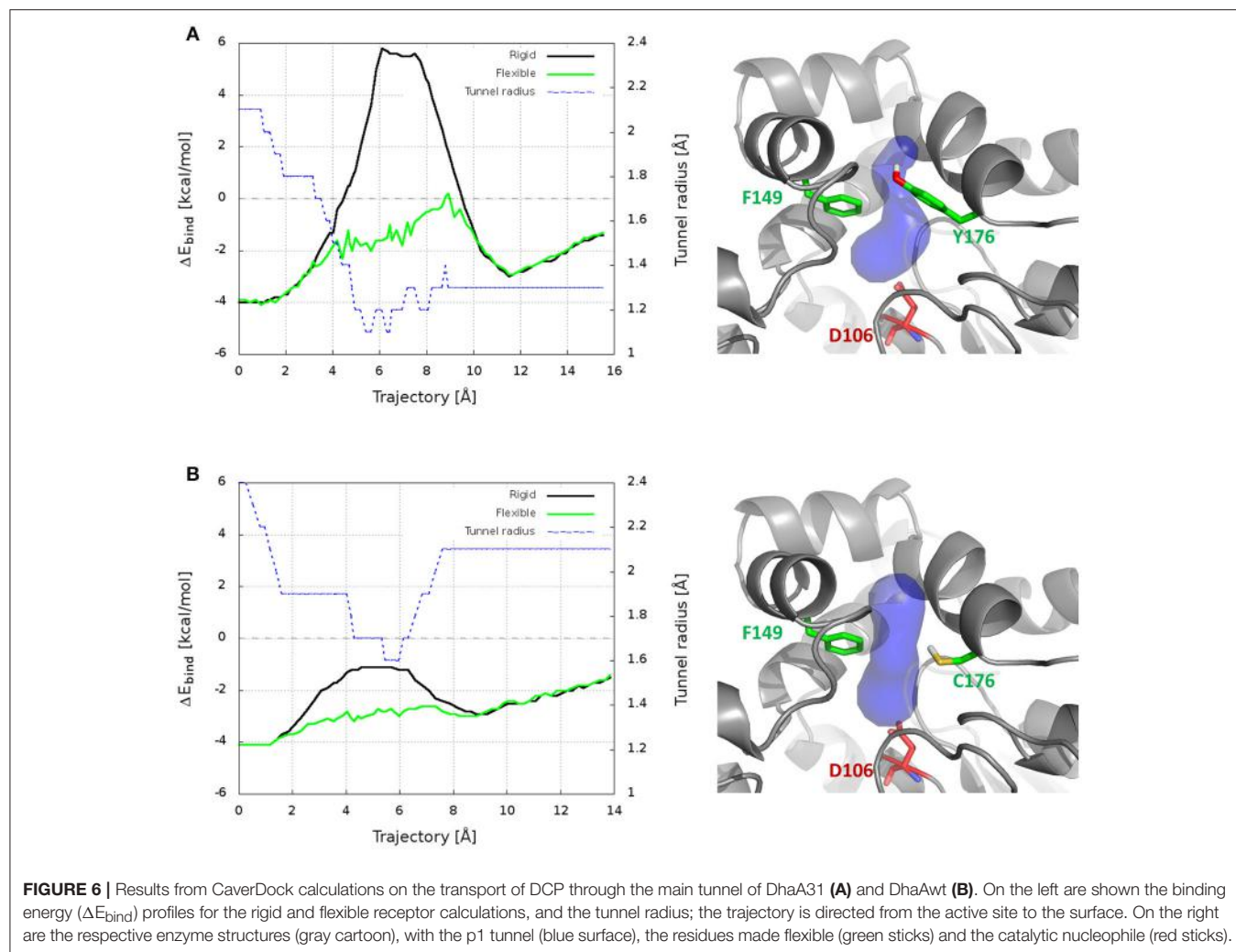
FIGURE 5 | Difference in interaction energies of DCP between the global minimum and the transition state TS1. The binding energy difference ($\Delta\Delta G_{bind} = \Delta G^{min} - \Delta G^{TS1}$) decomposed by residues for DhaA31 **(A)**, and the structure showing the residues with greatest differences **(B)**. The highlighted residues represent potential hot-spots for decreasing the energy barrier in the unbinding of DCP: (i) by decreasing the interactions of DCP at the global energy minimum with respect to TS1 (red residues), or (ii) by decreasing the energy of TS1 (blue residues). The average interaction differences are represented by the solid bars and the SD by the error bars. Only residues with $\Delta\Delta G$ or SD values ≥ 0.5 kcal/mol are displayed. The catalytic D106 is also represented (green residue).

were performed with static receptors, DCP showed very high and repulsive energy barrier (with $\Delta E_{bind} = +9.9$ kcal/mol) for DhaA31, which was in great contrast with DhaAwt that had lower barrier (3.1 kcal/mol) and always favorable energies ($\Delta E_{bind} <$ 0). This was due to the clashes of DCP with the protein residues passing through the narrower tunnel of DhaA31. The energy barrier was much higher for DhaA31 than for DhaAwt, which is qualitatively in agreement with the FES profiles (**Table S4**). When the CaverDock calculations were performed with flexibility (a feature still under development), the energy profiles became smoother and the energy barriers for the unbinding of DCP dropped significantly, to 4.3 kcal/mol for DhaA31, and 2.7 kcal/mol for DhaAwt. The residues that are made flexible are selected by the extent of clashes during the rigid docking. In this case, they were F149 and Y176 for DhaA31, and F149, and C176 for DhaAwt. These residues are located at the tunnel bottleneck, and they were shown to form strong interactions with DCP at the global energy minima identified in the FES, thus confirming their importance (**Figure S16**). This result is remarkable, especially when considering the dramatic difference in the calculation costs for CaverDock (hours) and the free energy methods (weeks).

DISCUSSION

The calculation of the kinetics and thermodynamics of ligand (un)binding has recently shown to be pivotal in drug design, but it can also be important in structural biology and biocatalysis. This is the case of the mutant enzyme DhaA31, which is currently the best known HLD for hydrolyzing the genotoxic compound

TCP, but whose catalytic turnover number is limited by the release of the DCP product. For this system, the assessment of the kinetic and thermodynamics bottlenecks in the unbinding of DCP may pave the way to the design of improved biocatalysts.

In this study we have calculated the unbinding kinetic rates (k_{off}) of DCP from the active sites of two enzymes, DhaA31, and DhaAwt, using two different methods—metadynamics (MTD) and adaptive sampling. Both methods predicted faster unbinding rates from DhaAwt than from DhaA31 (**Table 1**), which is in good agreement with previous evidence (Marques et al., 2017). However, there were considerable differences in the results from those two methods. For each system, the k_{off} values differed by 1-2 orders of magnitude, being slower with the MTD method than with the adaptive sampling. Regarding the relative values of the k_{off} rates, we obtained DhaAwt/DhaA31 ratio of 36 with MTD and only 2.5 with adaptive sampling, meaning that the latter method predicted faster rates for the two systems, but also closer values for those enzymes. On the other hand, the precision of the k_{off} values obtained with adaptive sampling was higher (the relative errors were lower) than with the MTD method. Previous studies have attributed the differences between the predicted and experimental k_{off} values to the errors in the force fields, the lack of polarizability, or the existence of tautomers, among other factors (Tiwary et al., 2015; Ferruz and De Fabritiis, 2016; Bruce et al., 2018). The simulations were performed under the same force field and water model (ff12SB and TIP3P, respectively), and therefore these differences are probably due to the intrinsic differences in the two methods. Since both methods have different types of biases—the MTD relies on a bias of repulsive potential energy deposited based on the CV

FIGURE 6 | Results from CaverDock calculations on the transport of DCP through the main tunnel of DhaA31 **(A)** and DhaAwt **(B)**. On the left are shown the binding energy (ΔE_{bind}) profiles for the rigid and flexible receptor calculations, and the tunnel radius; the trajectory is directed from the active site to the surface. On the right are the respective enzyme structures (gray cartoon), with the p1 tunnel (blue surface), the residues made flexible (green sticks) and the catalytic nucleophile (red sticks).

defined by the user, while the adaptive sampling uses Markov state models, calculated on-the-fly upon a user-defined metrics to start new epochs of MDs—and there are no experimental values available, it is difficult to assess which results are more accurate. Moreover, it is known that MDs performed with the same force fields but using different software packages may produce different conformational ensembles, and consequently different results (Childers and Daggett, 2018). The effects of the force field and solvent on the predicted kinetic rates were tested by additional adaptive sampling simulations performed with the ff14SB force field and OPC3 water model (**Table 1**). When compared to the available experimental data, the K_d value predicted for DhaAwt using the combination ff14SB+OPC3 (8.3 $\pm$ 2.5 mM) was the closest to the experimental value (K_d = 0.95 $\pm$ 0.34 mM). This seems to suggest that these conditions can better represent the physical properties of that system. A deeper discussion on this topic is presented in **Discussion S2** in **Supplementary Material**. We have demonstrated that the choice of the method, force field and water model can have a high impact on the prediction of kinetic properties. However, important conclusions could consistently be inferred from the comparative study of the two

systems, namely the higher propensity of DhaAwt to release DCP as compared to DhaA31. This strongly supports the value of comparative studies with similar systems, namely for the design of new enzyme variants in protein engineering.

The funnel-MTD simulations provided the free energy profiles for the unbinding of DCP from DhaA31 and DhaAwt, which allowed us to derive some conclusions about the respective energetic barriers and bottlenecks. The global energy minima in both enzymes were found in the middle of the tunnel (for *d1* ≈ 5-6 Å). After the global minimum, DhaA31 presented one steep energy barrier of 4.81 kcal/mol before DCP could reach the tunnel mouth, while DhaAwt had two steps with considerably reduced barriers (with 2.26 and 1.46 kcal/mol, respectively). This implies faster unbinding kinetic rates for DhaAwt, which is in good agreement with the kinetic calculations. The first transition state intermediate (TS1) in DhaA31 also corresponds to the geometric bottleneck and presents much higher steric constraints than the one observed in DhaAwt (**Figure 4**). The structural clusters, corresponding to the significant state along the FES, allowed the assessment of the respective binding energy of DCP with the protein's residues. From this analysis it was

possible to identify the residues that interact with DCP at the energy minima and transition states, and thus contribute to the stabilization of these states. We hypothesize that the difference in binding energy between the global energy minimum and the transition states may help identify the residues that contribute the most to retain DCP in that minimum and prevent the enzyme-product complex to proceed further to the full unbinding. Therefore, the residues with more negative $\Delta\Delta G_{bind}$ are the most likely hot-spots for improving the unbinding rates. In DhaA31, there residues correspond to F152, F168, F149, and F245 (**Figure 5**). It is known, however, that residues F149 and F245 are important to stabilize and orient TCP toward the S_N2 step (Marques et al., 2017), and therefore they should not be mutated to avoid undesirable disruption of the chemical steps. Previous studies performed with this system have identified several of these bottleneck residues as highly interacting with DCP, e.g., F149, F152, and F168 (Marques et al., 2017). However, their role was not so clear, and the current analysis came to confirm their pivotal importance in preventing the transition from the energy minimum toward the release. The residues with strongest interactions at the transition state, F144, T148, and K175, also represent interesting hot-spots to decrease the energy of TS1 and promote the transition of DCP along the unbinding process. However, the results from this approach are more unpredictable since the entropy can also be highly affected. CaverDock provided very interesting insights into the transport of DCP through the tunnels, especially considering that it is fast and has very low hardware requirements. With these simple calculations it could be concluded that DhaA31 has higher energy barrier to the unbinding of DCP as compared with DhaAwt, and we identified some of the residues that may hinder the transport the most.

Overall, we have shown that the computation of the kinetics and thermodynamics of protein-ligand unbinding can be a powerful tool for protein engineering when the goal is to improve the unbinding rates of a ligand from a biomolecule. Similar methods used can also be applied when the aim is to improve the ligand (substrate) binding. We illustrated that even highly sophisticated methods cannot precisely estimate kinetic values due to the computational limitation, and the results may highly depend on the selected parameters. However, they can be very useful for comparative purposes, which are typically needed in protein engineering projects. The free energy computation with funnel-MTD, or other enhanced-sampling free energy methods, can provide a deep insight into the binding/unbinding process, allow identification of the critical stages energetic and disclose the key residues for the unbinding. On the other hand, CaverDock is very fast and user-friendly, yet it may provide significant information about the ligand transport and enable the identification of key residues to improve the ligand transport. Different strategies can be followed for engineering new enzymes with improved ligand unbinding kinetic rates. The potential hot-spots for mutagenesis can be selected based on: (i) the residues showing the highest interaction differences between the energy minimum and transition state—aiming to decrease the energy barrier; (ii) residues interacting at the transition state—aiming to decrease the transition state energy; (iii) tunnel-lining residues—aiming to change the shape and geometric bottleneck of the tunnel; and (iv) residues in contact with the tunnel-lining residues—aiming to change the flexibility and dynamic properties of the tunnel residues. The selected hot-spot residues can be targeted by site-directed mutagenesis, smart libraries or saturation mutagenesis. The effects of particular mutations on the unbinding rates can be anticipated with *in silico* calculations, either with the thorough but costly approaches (MTD or adaptive sampling), or using the cheaper CaverDock for a faster screening.

CONCLUSIONS

Here we reported the application of metadynamics and adaptive sampling for computationally estimating the unbinding rates of the DCP product from two enzymes, DhaA31 and DhaAwt, and for aiding the design of improved biocatalysts. The unbinding of DCP is the rate-limiting step in the catalytic conversion of the toxic TCP with DhaA31, and improving this rate has biotechnological importance. Free energy calculation confirmed the different energetic profiles in the release of DCP by the two enzymes, and provided insights into the energetic bottlenecks in the unbinding process. By analyzing the interactions of DCP with DhaA31 at the critical stages we have identified several hot-spot residues that can be targeted by mutagenesis. Strikingly, some of these hot-spots were identified by the far less demanding CaverDock tool based on molecular docking. Site-directed mutagenesis or directed evolution applied on those hot-spots may result in new enzyme variants with the ability to release the DCP product at faster rates and thus present enhanced catalytic properties.

AUTHOR CONTRIBUTIONS

SM carried out the computational work and wrote the manuscript. All authors contributed to the design of the study, interpretation of the data, and have given approval of the final version of the manuscript.

FUNDING

ACKNOWLEDGMENTS

Dr. Stefano Raniolo (ETH Zurich) for his help with setting up the funnel-MTD method, to Dr. Piia Kokkonen (Masaryk University) for her help with the adaptive sampling simulations, and to Dr. Zbynek Prokop (Masaryk University) for the valuable discussions on the transient kinetic studies.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2018.00650/full#supplementary-material

REFERENCES

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., et al. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2, 19–25. doi: 10.1016/j.softx.2015.06.001

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak,. J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81, 3684–3690. doi: 10.1063/1.448118

Bonomi, M., Barducci, A., and Parrinello, M. (2009). Reconstructing the equilibrium boltzmann distribution from well-tempered metadynamics. *J. Comput. Chem.* 30, 1615–1621. doi: 10.1002/jcc.21305

Bonomi, M., Branduardi, D., Gervasio, F. L., and Parrinello, M. (2008). The unfolded ensemble and folding mechanism of the C-terminal GB1 β-hairpin. *J. Am. Chem. Soc.* 130, 13938–13944. doi: 10.1021/ja803652f

Bosma, T., Pikkemaat, M. G., Kingma, J., Dijk, J., and Janssen, D. B. (2003). Steady-state and pre-steady-state kinetic analysis of halopropane conversion by a rhodococcus haloalkane dehalogenase. *Biochemistry* 42, 8047–8053. doi: 10.1021/bi026907m

Branduardi, D., Gervasio, F. L., and Parrinello, M. (2007). From A to B in free energy space. *J. Chem. Phys.* 126:054103. doi: 10.1063/1.2432340

Bruce, N. J., Ganotra, G. K., Kokh, D. B., Sadiq, S. K., and Wade, R. C. (2018). New approaches for computing ligand-receptor binding kinetics. *Curr. Opin. Struct. Biol.* 49, 1–10. doi: 10.1016/j.sbi.2017.10.001

Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126:014101. doi: 10.1063/1.2408420

Case, D. A., Babin, V., Berryman, J. T., Betz, R. M., Cai, Q., Cerutti, S., Cheatham, T. E. III, et al. (2014). *AMBER 14 (version 14)*. San Francisco, CA: University of California.

CaverDock (2018). Available online at: https://loschmidt.chemi.muni.cz/caverdock/

Childers, M. C., and Daggett, V. (2018). Validating molecular dynamics simulations against experimental observables in light of underlying conformational ensembles. *J. Phys. Chem. B* 122, 6673–6689. doi: 10.1021/acs.jpcb.8b02144

Chiu, S. H., and Xie, L. (2016). Toward high-throughput predictive modeling of protein binding/unbinding kinetics. *J. Chem. Inform. Model.* 56, 1164–1174. doi: 10.1021/acs.jcim.5b00632

Chovancova, E., Pavelka, A., Benes, P., Strnad, O., Brezovsky, J., Kozlikova, B., et al. (2012). CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. Edited by Andreas Prlic. *PLoS Comput. Biol.* 8:e1002708. doi: 10.1371/journal.pcbi.1002708

Darden, T., York, D., and Pedersen, L. (1993). Particle mesh ewald: an N·log(N) method for ewald sums in large systems. *J. Chem. Phys.* 98, 10089–10092. doi: 10.1063/1.464397

Dickson, A., Tiwary, P., and Vashisth, H. (2017). Kinetics of ligand binding through advanced computational approaches: a review. *Curr. Top. Med. Chem.* 17, 2626–2641. doi: 10.2174/1568026617666170414142908

Doerr, S., Harvey, M. J., Noé, F., and De Fabritiis, G. (2016). HTMD: high-throughput molecular dynamics for molecular discovery. *J. Chem. Theory Comput.* 12, 1845–1852. doi: 10.1021/acs.jctc.6b00049

Dvorak, P., Bidmanova, S., Damborsky, J., and Prokop, Z. (2014). Immobilized synthetic pathway for biodegradation of toxic recalcitrant pollutant 1,2,3-trichloropropane. *Environ. Sci. Tech.* 48, 6859–6866. doi: 10.1021/es500396r

Feenstra, K. A., Hess, B., and Berendsen, H. J. C. (1999). Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J. Comput. Chem.* 20, 786–798. doi: 10.1002/(SICI)1096-987X(199906)20:8<786::AID-JCC5>3.0.CO;2-B

Ferruz, N., and De Fabritiis, G. (2016). Binding kinetics in drug discovery. *Molecul. Inform.* 35, 216–226. doi: 10.1002/minf.201501018

Filipovič, J., Vávra, O., Plhák, J., Bednár, D., Marques, S. M., Brezovský, J., et al. (2018). CaverDock: *A Novel Method for the Fast Analysis of Ligand Transport*. Available online at: https://arxiv.org/abs/1809.03453

Harvey, M. J., and De Fabritiis, G. (2009). An implementation of the smooth particle mesh ewald method on GPU hardware. *J. Chem. Theory Comput.* 5, 2371–2377. doi: 10.1021/ct900275y

Harvey, M. J., Giupponi, G., and Fabritiis, G. D. (2009). ACEMD: accelerating biomolecular dynamics in the microsecond time scale. *J. Chem. Theory Comput.* 5, 1632–1639. doi: 10.1021/ct9000685

Hess, B., Bekker, H., Berendsen, H., and Fraaije, J. G. E. M. (1997). LINCS: a linear constraint solver for molecular simulations. *J. Computat. Chem.* 18, 1463–1472. doi: 10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H

Hopkins, C. W., Le Grand, S., Walker, R. C., and Roitberg, A. E. (2015). Long-time-step molecular dynamics through hydrogen mass repartitioning. *J. Chem. Theory Comput.* 11, 1864–1874. doi: 10.1021/ct5010406

Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006). Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins* 65, 712–725. doi: 10.1002/prot.21123

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graphics* 14, 33–38. doi: 10.1016/0263-7855(96)00018-5

Izadi, S., and Onufriev, A. V. (2016). Accuracy limit of rigid 3-point water models. *J. Chem. Phys.* 145:074501. doi: 10.1063/1.4960175

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein,. M., L (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935. doi: 10.1063/1.445869

Joung, I. S., and Cheatham, T. E. (2008). Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* 112, 9020–9041. doi: 10.1021/jp8001614

Joung, I. S., and Cheatham, T. E. (2009). Molecular dynamics simulations of the dynamic and energetic properties of alkali and halide ions using water-model-specific ion parameters. *J. Phys. Chem. B* 113, 13279–13290. doi: 10.1021/jp902584c

Kaushik, S., Marques, S. M., Khirsariya, P., Paruch, K., Libichova, L., Brezovsky, J., et al. (2018). Impact of the access tunnel engineering on catalysis is strictly ligand-specific. *FEBS J.* 285, 1456–1476. doi: 10.1111/febs.14418

Kokh, D. B., Amaral, M., Bomke, J., Grädler, U., Musil, D., Buchstaller, H. P., et al. (2018). Estimation of drug-target residence times by τ-random acceleration molecular dynamics simulations. *J. Chem. Theory Comput.* 14, 3859–3869. doi: 10.1021/acs.jctc.8b00230

Koudelakova, T., Bidmanova, S., Dvorak, P., Pavelka, A., Chaloupkova, R., Prokop, Z., et al. (2013). Haloalkane dehalogenases: biotechnological applications. *Biotech. J.* 8, 32–45. doi: 10.1002/biot.201100486

Kurumbang, N. P., Dvorak, P., Bendl, J., Brezovsky, J., Prokop, Z., and Damborsky, J. (2014). Computer-assisted engineering of the synthetic pathway for biodegradation of a toxic persistent pollutant. *ACS Synth. Biol.* 3, 172–181. doi: 10.1021/sb400147n

Kutý, M., Damborský, J., Prokop, M., and Koča, J. (1998). A molecular modeling study of the catalytic mechanism of haloalkane dehalogenase. 2. quantum chemical study of complete reaction mechanism. *J. Chem. Inform. Comp. Sci.* 38, 736–741. doi: 10.1021/ci970290b

Limongelli, V., Bonomi, M., and Parrinello, M. (2013). Funnel metadynamics as accurate binding free-energy method. *Proc. Natl. Acad. Sci. U.S.A.* 110, 6358–6363. doi: 10.1073/pnas.1303186110

Lu, H., and Tonge, P. J. (2010). Drug-target residence time: critical information for lead optimization. *Curr. Opin. Chem. Biol.* 14, 467–474. doi: 10.1016/j.cbpa.2010.06.176

Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015). Ff14SB: improving the accuracy of protein side chain and backbone parameters from Ff99SB. *J. Chem. Theory Comput.* 11, 3696–3713. doi: 10.1021/acs.jctc.5b00255

Marques, S. M., Dunajova, Z., Prokop, Z., Chaloupkova, R., Brezovsky, J., and Damborsky, J. (2017). Catalytic cycle of haloalkane dehalogenases toward unnatural substrates explored by computational modeling. *J. Chem. Inform. Model.* 57, 1970–1989. doi: 10.1021/acs.jcim.7b00070

Miller, B. R., McGee, T. D., Swails, J. M., Homeyer, N., Gohlke, H., and Roitberg, A. E. (2012). MMPBSA.Py: an efficient program for end-state free energy calculations. *J. Chem. Theory Comput.* 8, 3314–3321. doi: 10.1021/ct300418h

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., et al. (2009). AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* 30, 2785–2791. doi: 10.1002/jcc.21256

Naritomi, Y., and Fuchigami, S. (2011). Slow dynamics in protein fluctuations revealed by time-structure based independent component analysis: the case of domain motions. *J. Chem. Phys.* 134:065101. doi: 10.1063/1.3554380

Nguyen, H., Maier, J., Huang, H., Perrone, V., and Simmerling, C. (2014). Folding simulations for proteins with diverse topologies are accessible in days with a physics-based force field and implicit solvent. *J. Am. Chem. Soc.* 136, 13959–13962. doi: 10.1021/ja5032776

Pavlova, M., Klvana, M., Prokop, Z., Chaloupkova, R., Banas, P., Otyepka, M., et al. (2009). Redesigning dehalogenase access tunnels as a strategy for degrading an anthropogenic substrate. *Nat. Chem. Biol.* 5, 727–733. doi: 10.1038/nchembio.205

Pinto, G., Vávra, O., Filipovi,č, J., Bednar, D., and Damborsky, J. (2018). *Fast Screening of Binding and Unbinding of Inhibitors Using Novel Software Tool CaverDock*. Under Publication, September.

Reuveni, S., Urbakh, M., and Klafter, J. (2014). Role of substrate unbinding in michaelis–menten enzymatic reactions. *Proc. Natl. Acad. Sci.U.S.A.* 111, 4391–4396. doi: 10.1073/pnas.1318122111

Rydzewski, J., and Nowak, W. (2017). Ligand diffusion in proteins via enhanced sampling in molecular dynamics. *Physics Life Rev.* 22–23, 58–74. doi: 10.1016/j.plrev.2017.03.003

Salvalaglio, M., Tiwary, P., and Parrinello, M. (2014). Assessing the reliability of the dynamics reconstructed from metadynamics. *J. Chem. Theory Comput.* 10, 1420–1425. doi: 10.1021/ct500040r

Samin, G., and Janssen, D. B. (2012). Transformation and Biodegradation of 1,2,3-Trichloropropane (TCP). *Environ. Sci. Pollut. Res. Int.* 19, 3067–3078. doi: 10.1007/s11356-012-0859-3

Sousa da Silva, A. W., and Vranken, W. F. (2012). ACPYPE - antechamber python parser interfacE. *BMC Res.* 5:367. doi: 10.1186/1756-0500-5-367

Swails, J. (2010). *ParmEd*. Available online at: https://github.com/ParmEd/ParmEd.

The PyMO L Molecular Graphics System (2014). *(version 1.7.4)*. Schrödinger, LLC.

Tiwary, P., Limongelli, V., Salvalaglio, M., and Parrinello, M. (2015). Kinetics of protein-ligand unbinding: predicting pathways, rates, and rate-limiting steps. *Proc. Natl. Acad. Sci. U.S.A.* 112, E386–E391. doi: 10.1073/pnas.1424461112

Tiwary, P., and Parrinello, M. (2013). From metadynamics to dynamics. *Phys. Rev. Lett.* 111:230602. doi: 10.1103/PhysRevLett.111.230602

Tiwary, P., and Parrinello, M. (2015). A time-independent free energy estimator for metadynamics. *J. Phys. Chem. B* 119, 736–742. doi: 10.1021/jp504920s

Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C., and Bussi, D. (2014). PLUMED 2: new feathers for an old bird. *Comput. Phys. Commun.* 185, 604–613. doi: 10.1016/j.cpc.2013.09.018

Vávra, O. J., Filipovič J., Plhák, J., Bednár, D., Marques, S. M., Brezovský, J., et al. (2018). *CaverDock: Ligand Transport Analysis Based on Molecular Docking*. Under Publication.

Verlet, L. (1967). Computer 'Experiments' on classical fluids. I. thermodynamical properties of lennard-jones molecules. *Phys. Rev.* 159, 98–103.

Verschueren, K. H., Seljée, F., Rozeboom, H. J., Kalk, K. H., and Dijkstra, B. W. (1993). Crystallographic analysis of the catalytic mechanism of haloalkane dehalogenase. *Nature* 363, 693–698. doi: 10.1038/363693a0

Wang, L. H., Tsai, A. L., and Hsu, P. Y. (2001). Substrate binding is the rate-limiting step in thromboxane synthase catalysis. *J. Biol. Chem.* 276, 14737–14743. doi: 10.1074/jbc.M009177200

Yao, L., Li, Y., Wu, Y., Liu, A., and Yan, H. (2005). Product release is rate-limiting in the activation of the prodrug 5-fluorocytosine by yeast cytosine deaminase. *Biochemistry* 44, 5940–5947. doi: 10.1021/bi050095n

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The Challenging *in silico* Description of Carbon Monoxide Oxidation as Catalyzed by Molybdenum-Copper CO Dehydrogenase

Anna Rovaletti[1], Maurizio Bruschi[1], Giorgio Moro[2], Ugo Cosentino[1] and Claudio Greco[1*]

[1] Dipartimento di Scienze dell'Ambiente e della Terra, Università Degli Studi di Milano-Bicocca, Milan, Italy, [2] Dipartimento di Biotecnologie e Bioscienze, Università Degli Studi di Milano-Bicocca, Milan, Italy

Carbon monoxide (CO) is a highly toxic gas to many living organisms. However, some microorganisms are able to use this molecule as the sole source of carbon and energy. Soil bacteria such as the aerobic *Oligotropha carboxidovorans* are responsible for the annual removal of about $2x10^8$ tons of CO from the atmosphere. Detoxification through oxidation of CO to CO_2 is enabled by the MoCu-dependent CO-dehydrogenase enzyme (MoCu-CODH) which—differently from other enzyme classes with similar function—retains its catalytic activity in the presence of atmospheric O_2. In the last few years, targeted advancements have been described in the field of bioengineering and biomimetics, which is functional for future technological exploitation of the catalytic properties of MoCu-CODH and for the reproduction of its reactivity in synthetic complexes. Notably, a growing interest for the quantum chemical investigation of this enzyme has recently also emerged. This mini-review compiles the current knowledge of the MoCu-CODH catalytic cycle, with a specific focus on the outcomes of theoretical studies on this enzyme class. Rather controversial aspects from different theoretical studies will be highlighted, thus illustrating the challenges posed by this system as far as the application of density functional theory and hybrid quantum-classical methods are concerned.

Keywords: molybdenum, copper, dehydrogenase, DFT, carbon monoxide

Carbon monoxide (CO) is a fatal gas to many living organisms as well as an indirect greenhouse gas in the atmosphere (Liu et al., 2018). Global CO emissions derive both from anthropogenic and natural sources (Choi et al., 2017). One of the main sinks of atmospheric CO is constituted by the soil, in which it is consumed in large amounts by microbial oxidation (Liu et al., 2018). One example of these important soil microorganisms is represented by the aerobic bacteria *Oligotropha carboxidovorans*. The latter is able to grow using CO as its sole source of carbon and energy (Hille et al., 2015). This metabolism is ascribed to the air-stable Mo/Cu-dependent CO dehydrogenase (MoCu-CODH) enzyme that catalyzes the oxidation of CO to CO_2 (Zhang et al., 2010).

This enzyme contains a unique active site composed by a molybdenum/copper bimetallic center (see **Figure 1A**). The molybdenum ion is found in a square-pyramidal geometry with one apical oxo ligand, a dithiolene ligand from the molybdopterin cytosine dinucleotide (MCD) cofactor, an equatorial oxo ligand and a sulfido ligand. The latter bridges to the copper center, which links the active site to the protein matrix by coordinating the sulfur atom of Cys388. Moreover, Cu is also

coordinated to a weakly bound water molecule (Gnida et al., 2003; Rokhsana et al., 2016), and can coordinate not only CO (i.e., the physiologic substrate) but also H_2. In fact, the MoCu-CODH enzyme has the ability to catalyze dihydrogen oxidation, even though such hydrogenase activity is rather low (Santiago and Meyer, 1996; Wilcoxen and Hille, 2013).

The protonation state of the active site has been object of debate. In fact, experimental X-ray diffraction (XRD) results were interpreted as indicative of a Mo(=O)OH state both for the oxidized and the reduced forms of the enzyme (Dobbek et al., 2002). Differently, extended X-ray absorption fine structure (EXAFS) spectroscopy suggested the presence of a $Mo^{VI}(=O)O$ unit and a $Mo^{IV}(=O)OH_2$ unit for the oxidized and reduced states, respectively (Gnida et al., 2003). Recent data in support of the EXAFS-based hypothesis came from electron paramagnetic resonance (EPR) analysis of the Mo^V analog and from theoretical
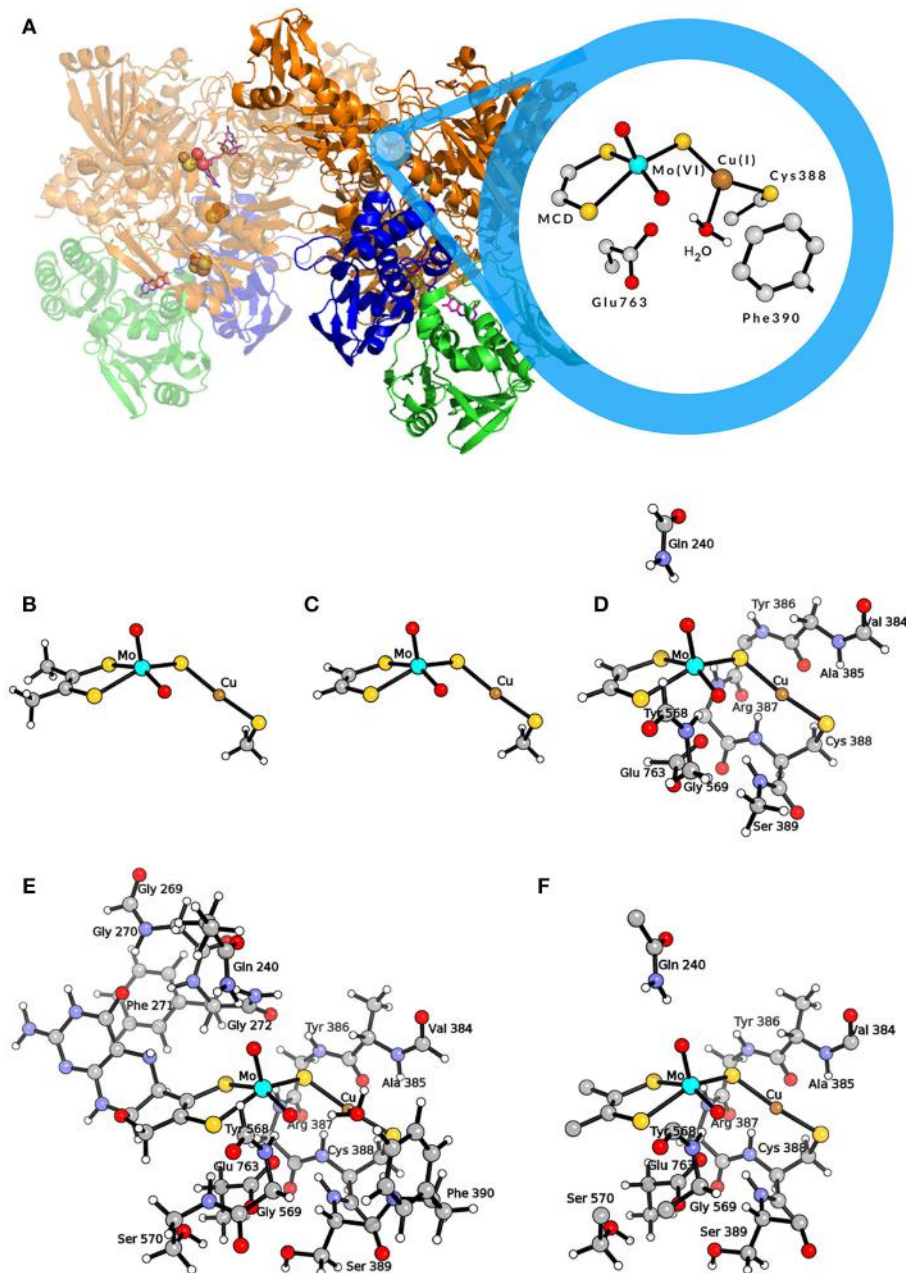
FIGURE 1 | (A) Representation of the MoCu-CODH enzyme and of the active site in the Mo(VI)Cu(I) resting state. (B) QM model used by Hofmann et al. (2005) and by Stein and Kirk (2014); (C) QM model used by Siegbahn and Shestakov (2005) and by Breglia et al. (2017); (D) larger QM model used by Siegbahn and Shestakov (2005); (E) QM model used by Rokhsana et al. (2016); (F) QM region of the hybrid QM/MM model used by Xu and Hirao (2018). Color code of atoms in the active site: cyan, molybdenum; orange, copper; yellow, sulfur; red, oxygen; gray, carbon; white, hydrogen.

calculations (Zhang et al., 2010; Rokhsana et al., 2016). As far as the oxidation state of the Cu ion is concerned, it maintains the +1 state throughout the enzymatic catalytic cycle (Dobbek et al., 2002; Gnida et al., 2003). In fact, CO oxidation occurs directly at the Cu center (Shanmugam et al., 2013), and the two-electron transfer to Mo at each catalytic cycle is allowed by the highly delocalized nature of the Mo(μ-S)Cu unit (Gourlay et al., 2006). Key second-sphere amino acid residues are conserved in MoCu-CODH enzymes and in homologues with different activity, i.e., xanthine oxidases (Hille, 2013). In particular, a conserved glutamate residue (Glu763) is found in proximity to the equatorial oxo ligand of Mo and is considered to act as a base to facilitate deprotonation events (Wilcoxen and Hille, 2013; Hille et al., 2015). Moreover, the aromatic ring of a phenylalanine residue (Phe390), located in front of the Cu^I ion, is thought to have relevance for the correct positioning of the substrate within the active site (Rokhsana et al., 2016).

Carbon monoxide ($|C{\equiv}O$) is expected to show similar reactivity with respect to the isoelectronic isocyanide ($|C{\equiv}N$-R) species. They share the presence of non-bonding electron pairs in the sp orbital of the terminal carbon atom and a triple bond between the latter and a more electronegative atom. Indeed, Dobbek et al. reported the inhibitory activity of n-butylisocyanide, $|C{\equiv}N$-$(CH_2)_3CH_3$, toward the oxidized MoCu-CODH (Dobbek et al., 2002). In the same study, the corresponding crystallographic structure of the inhibited enzyme was determined (PDB ID: 1N62). The resulting inactive complex is characterized by a thiocarbamate geometry in which the isocyanide group forms covalent bonds with the μ-sulfido ligand, the equatorial oxygen of Mo and the Cu atom, while the alkyl chain of n-butylisocyanide extends into the hydrophobic interior of the substrate channel.

The features of the crystal structure of the n-butylisocyanide-bound state prompted Dobbek and coworkers to advance the first hypothesis ever proposed for the MoCu-CODH catalytic mechanism (see **Figure 2A**) (Dobbek et al., 2002)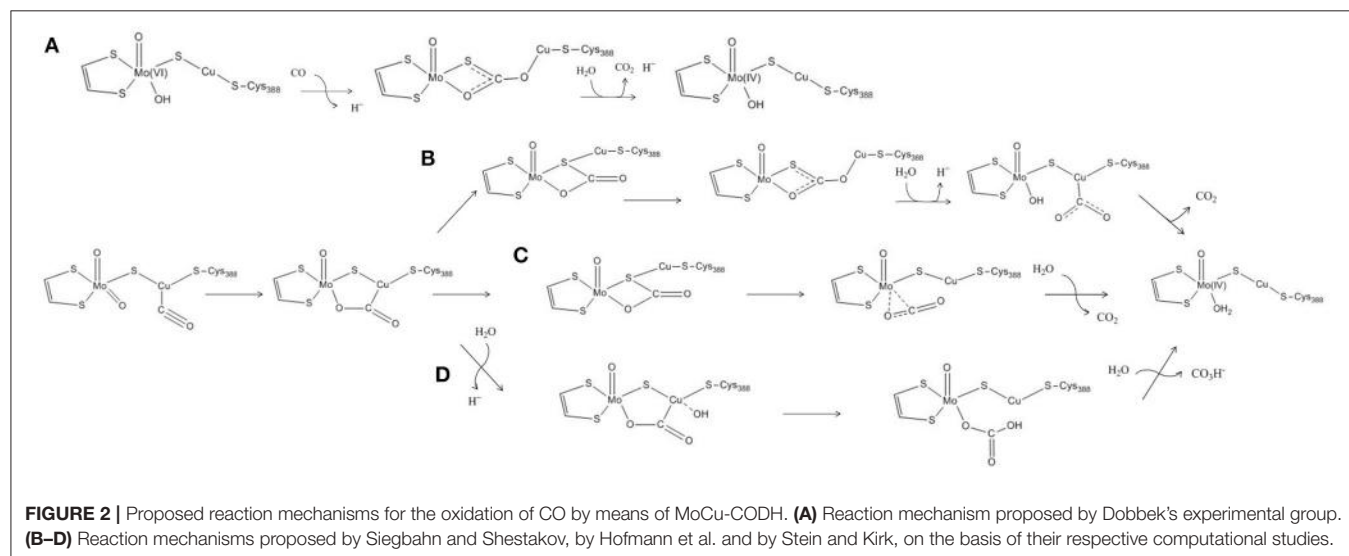. The latter involves the formation of a thiocarbonate-like intermediate—analogous to the thiocarbamate derivative formed during the aforementioned inhibition—after the CO substrate accesses the oxidized active site. Such thiocarbonate species would be characterized by the insertion of CO between copper, the μ-sulfido ligand and the equatorial oxo ligand of the Mo atom. Taking inspiration from such a mechanistic hypothesis for CO-oxidation catalysis, the enzymatic mechanism has been subsequently studied with computational methods by several groups, thus giving birth to a debate that is still ongoing (*vide infra*).

Recently, it has been reported that rather bulky thiol molecules—e.g., L-cysteine, coenzyme A or glutathione—can reach the bimetallic active site (Kreß et al., 2014). They cause a reversible inactivation of the enzymatic activity, competing with the substrates for the same position at the Cu^I center.

In the following two sections of this manuscript we aim at reviewing the theoretical studies that have been published on MoCu-CODH using quantum mechanical (QM) and hybrid quantum mechanical/molecular mechanical (QM/MM) approaches. In doing so, we will pay special attention to the controversies and the challenges that have emerged. Moreover, promising future developments in the theoretical description of this system will be proposed in the concluding section.

1. SMALL- AND MEDIUM-SIZED QM-CLUSTER MODELS

The first theoretical investigations of the catalytic activity of MoCu-CODH were carried out independently by two theoretical groups in 2005 (Hofmann et al., 2005; Siegbahn and Shestakov, 2005). To model the enzymatic active site, Hofmann et al. used a small cluster of 24 atoms representing the two transition metals and their first coordination spheres (see **Figure 1B**) (Hofmann et al., 2005), whereas Siegbahn and Shestakov performed calculations with two different models, a small one

FIGURE 2 | Proposed reaction mechanisms for the oxidation of CO by means of MoCu-CODH. **(A)** Reaction mechanism proposed by Dobbek's experimental group. **(B–D)** Reaction mechanisms proposed by Siegbahn and Shestakov, by Hofmann et al. and by Stein and Kirk, on the basis of their respective computational studies.

of about 20 atoms and a bigger one composed of about 70 atoms (see **Figures 1C,D**). In the latter, some residues belonging to the second coordination spheres of the metals were explicitly included (Siegbahn and Shestakov, 2005). The hybrid density functional B3LYP (Lee et al., 1988; Becke, 1993) was employed in both cases to optimize the geometries and compute relative energies of intermediates along the putative catalytic cycles. For geometry optimizations, Hofmann and coworkers adopted the Lanl2DZ (Dunning and Hay, 1976; Hay and Wadt, 1985a,b; Wadt and Hay, 1985) effective core basis set with additional d-type functions on S atoms. This was followed by single point energy calculations using the SDD (Dunning, 1970; Dunning and Hay, 1977; Dolg et al., 1987) basis set augmented by d-type polarization functions on all non-hydrogen, non-metal atoms. The basis sets employed by Siegbahn and Shestakov were lacvp and lacv3p*—with ECP for Mo, Cu, and S atoms—for geometry optimizations and energy calculations, respectively. In both studies, the protein matrix surrounding the active site was modeled by a continuum dielectric with $\epsilon = 4$ (Eckert and Klamt, 2002; Cossi et al., 2003).

A comparative analysis of the results coming from two such early investigations evidences a significant variability of the results obtained as a function of the adopted level of theory, within models of the same size. In particular, the calculated energy differences showed significant dependency on the basis sets used, with deviations up to 21 kJ/mol in the case of intermediates involved in formation of the new C–O bond, a key step in the catalytic process. Interestingly, previous studies of oxygen-atom transfer (OAT) reactions involving Mo-complexes (Li et al., 2013) and other transition-metal-containing systems (Hu and Chen, 2015; Li et al., 2015) also evidenced the strong basis set effect on computed energy differences along the reactive paths.

Notwithstanding the shortcomings deriving from the choice of basis sets, both Siegbahn, Shestakov and their respective coworkers evidenced a surprisingly high stability for the thiocarbonate intermediate. The presence of such a deep minimum on the energy landscape pertaining to the reaction mechanism was interpreted differently by the two groups. Siegbahn and Shestakov in particular proposed that the thiocarbonate derivative represents an intermediate of the CO-oxidation mechanism. However, in the proposed mechanism the barrier for the release of the CO_2 product was estimated to be rather high, as it would require the insertion of a water molecule which was reported not to be a facile step (see **Figure 2B**). Differently, Hofmann and coworkers raised the possibility that the thiocarbonate adduct lies outside the catalytic cycle, in a deep potential energy well that would effectively slow down enzymatic activity (see **Figure 2C**). These authors further proposed that the constrains imposed by the protein matrix could prevent formation of such a stable off-path adduct, a hypothesis that— however—was later discarded as a result of a theoretical study focused on this topic (Siegbahn, 2011).

In a more recent theoretical study, a different mechanism for the oxidation of CO by the MoCu-CODH enzyme was proposed (see **Figure 2D**) (Stein and Kirk, 2014). Using a cluster model analogous to the one previously employed by Hofmann and coworkers (see **Figure 1B**), at the PBE/TZP (Perdew et al., 1996; Ernzerhof and Scuseria, 1999) level of theory and including continuum dielectric contributions (ϵ = hexane) (Klamt and Schüürmann, 1993), Stein and Kirk proposed that the stable thiocarbonate intermediate formation could be bypassed by evolving bicarbonate as a final product rather than CO_2. Bicarbonate formation would proceed via nucleophilic attack of a copper-activated water molecule on the C atom of the metal-bound CO_2. However, such a picture is at odds with recent experimental studies, which appear to exclude the possibility of forming a bicarbonate complex during catalysis (Dingwall et al., 2016).

Breglia and coworkers published the most recent theoretical study of MoCu-CODH, in which only the first shell coordination spheres were included in a QM model (see **Figure 1C**) (Breglia et al., 2017). Such a study mainly regards the hydrogenase activity of the enzyme and includes a comparative analysis of the binding reactions of the physiologic substrate—i.e., CO—and of dihydrogen to the Cu ion. Similarly to previous studies on H_2- and CO-binding enzyme models in which a pure functional was used in conjuction with triple-zeta bases (Greco et al., 2015; Rovaletti and Greco, 2018), geometry optimizations and energy calculations were carried out in vacuo at BP86/def2-TZVP level (Perdew, 1986; Becke, 1988; Weigend and Ahlrichs, 2005). As far as the energetics of CO binding is concerned, the computed ΔE was as negative as −64 kJ/mol. A comparison of the latter value with the results of corresponding calculations previously performed at different levels of theory for the same reaction by Siegbahn, Hofmann and their respective coworkers, evidences significant discrepancies ($\Delta\Delta E$ = 50 and 31 kJ/mol, respectively). Actually, the occurrence of such large differences comes with little surprise, given the well-known shortcomings in binding energy calculations using quantum chemical models of coordination complexes and their ligands (Husch et al., 2018).

2. LARGE QM-CLUSTER MODELS AND HYBRID MODELS

The importance of extending the dimension of the bimetallic active site model and systematically accounting for the effects of the second-sphere residues on energetics was evidenced in a recent theoretical work by Rokhsana and coworkers (Rokhsana et al., 2016). In fact, a large-size cluster model of around 180 atoms (see **Figure 1E**) turned out to be required for a fully satisfactory reproduction of experimentally determined structural parameters. The same held true for the evaluation of the most plausible protonation states of the Mo/Cu core, which was done taking into account key geometric features of the enzyme crystal structure and redox potential measurements available in literature. The authors were able to assess in particular the protonation state of the equatorial oxo ligand of Mo, at the different redox states attained during catalysis. As for the adopted level of theory, Rokhsana et al. employed the def2-TZVP basis set for all elements, apart from H and C atoms, for which the def2-SVP basis set was used (Weigend and Ahlrichs, 2005). The protein environment was modeled by a continuum dielectric with $\epsilon = 4$

(Klamt and Schüürmann, 1993). The BP86 and B3LYP density functionals were used for geometry optimizations and energy evaluation, respectively.

In a subsequent study, explicit consideration of the whole protein environment was achieved by means of a hybrid quantum mechanics/molecular mechanics (QM/MM) approach (Xu and Hirao, 2018). In the work of Xu and Hirao, the active site was described by a QM region of 89 atoms (see **Figure 1F**) using the B3LYP functional. During geometry optimization, the SDD effective core potential basis set was employed to represent the transition metal ions, whereas the 6-31G* basis set was adopted for all the other atoms (Dolg et al., 1987; Andrae et al., 1990). Single-point energy calculations were carried out using the larger def2-TZVP basis set. For the molecular mechanics calculations, the AMBER03 force field was employed (Duan et al., 2003). Moreover, the Grimme's D3-correction with Becke—Johnson damping [D3(BJ)] was taken into account in the calculations (Grimme et al., 2011). The previously proposed catalytic mechanisms that involve the formation of the thiocarbonate species were re-investigated at such a level of theory. According to Xu and Hirao's study, the S-C-bound adduct would be formed along the reaction pathway as previously suggested (Dobbek et al., 2002; Siegbahn and Shestakov, 2005). However, based on the novel QM/MM results, the thiocarbonate species thus formed would not be as stable as previously proposed. It has to be remarked that the thiocarbonate intermediate was not found to be directly linked to the transition state for CO_2 releasing. In fact, it was proposed that— after thiocarbonate formation— the reaction needs to follow a reverse process to productively proceed toward CO_2 evolution. Notably, the overall barrier for the proposed catalytic mechanism was found to be low (in the order of 50 kJ/mol). In the same study, Xu and Hirao also carried out purely QM calculations with a QM-cluster of the same size of the quantum-mechanical region of their hybrid model, and compared the obtained results with those coming from QM/MM modeling. Such a comparison evidenced that the protein environment is not involved in modulating the kinetic barrier associated with the investigated catalytic mechanism. However, it was found that the protein matrix plays an important role in the stabilization of the CO_2-released state. Finally, it was also reported that the inclusion of dispersive corrections lowers by 15 kJ/mol the activation barrier of the product-releasing step, in line with what was expected for the modeling of a bimolecular reaction step.

CONCLUDING REMARKS

Over the last fifteen years, the theoretical investigation of the CO oxidation mechanism by MoCu-CODH has given rise to a debate, the essentials of which are centered on the possible occurrence and on the role of a thiocarbonate catalytic intermediate. In the above sections, we have reported key details of the various computational studies published to date, and we are now in the condition to present a more general outlook on the state of the art regarding MoCu-CODH.

The early studies by Siegbahn, Hofmann, and their respective coworkers evidenced that the thiocarbonate

intermediate would occupy a deep well in the energy profiles pertaining to the investigated reaction mechanisms. However, the kinetic barriers they computed for CO_2 evolution were at least 30 kJ/mol higher than the recently determined experimental counterpart (Zhang et al., 2010). In part, this picture depends on the neglect of dispersive corrections: their inclusion became a standard possibility only after the publication of the mentioned study (Siegbahn, 2011).

Results more compatible with the experimental evidence of a kinetic barrier of around 50 kJ/mol were obtained by Xu and Hirao, who exploited a larger QM-cluster model with the explicit inclusion of most of the second-sphere coordination environment, along with employment of large basis sets and dispersion corrections (Xu and Hirao, 2018). It is noticeable that, according to Xu and Hirao's results, the thiocarbonate species still appears to behave as a thermodynamic sink. Even though not a very deep one, such a sink would effectively hamper the advancement along the proposed path toward products, a rather unusual role for a species formed during an enzymatic process.

All the catalytic mechanisms proposed in the theoretical studies reviewed here focus on the possibility that the Mo-bound equatorial oxo ligand performs the nucleophilic attack on the activated CO substrate bound to the Cu ion. This is in line with what has been suggested in the case of the catalytic mechanism of the homologous xanthine oxidases enzymes. However, a variant with respect to such a picture has been recently proposed, in which an activated water molecule would play the role of nucleophile (Hille et al., 2015). Notably, to the best of our knowledge this alternative mechanism has not yet been investigated by QM or hybrid QM/MM studies.

Future theoretical studies on this (and other) putative catalytic mechanism will possibly face the challenge associated with rather pronounced fluctuations of computed energy differences as a function of the adopted level of theory. As far as the application of density functional theory is concerned, extensive benchmarking could in principle help to improve the theoretical predictions. In this regard, the available experimental data on enzyme inhibition— by thiols in particular—could represent a useful dataset, keeping in mind that the reproduction of binding energies in the case of bulky thiols might require extensive protein matrix phase space sampling in the case of QM/MM studies. High level *ab initio* methods are a viable—though still challenging—alternative for providing reliable results. In fact, thanks to recent methodological developments, the treatment of relatively large bimetallic systems has been shown to be computationally affordable with multiconfigurational post-Hartree-Fock approaches (Phung et al., 2016, 2018; Dong et al., 2017).

The relevance of MoCu-CODH as an inspiring system for future biomimetic and bioengineering applications is currently growing. This is due not only to the relevance of the reactions it catalyzes, but also to its resistance to atmospheric O_2 exposure— a rare feature in the case of enzymes expressing carbon monoxide dehydrogenase and hydrogenase activities (Choi et al., 2017; Gourlay et al., 2018; Groysman et al., 2018). Notably, the recent establishment of a functional heterologous expression system for

the MoCu-CODH enzyme (Kaufmann et al., 2018) together with developments in the computational chemistry field will hopefully boost the positive feedback among biochemical, biomimetic and quantum chemical studies, opening new perspectives for a deeper understanding of this interesting metalloenzyme.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

REFERENCES

Andrae, D., Haeussermann, U., Dolg, M., Stoll, H., and Preuss, H. (1990). Energy-adjusted *ab initio* pseudopotentials for the second and third row transition elements. *Theor. Chim. Acta* 77, 123–141.

Becke, A. D. (1988). Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* 38:3098.

Becke, A. D. (1993). A new mixing of Hartree–Fock and local density-functional theories. *J. Chem. Phys.* 98, 1372–1377.

Breglia, R., Bruschi, M., Cosentino, U., De Gioia, L., Greco, C., Miyake, T., et al. (2017). A theoretical study on the reactivity of the Mo/Cu-containing carbon monoxide dehydrogenase with dihydrogen. *Protein Eng. Design Select.* 30, 169–174. doi: 10.1093/protein/gzw071

Choi, E. S., Min, K., Kim, G.-J., Kwon, I., and Kim, Y. H. (2017). Expression and characterization of *Pantoea* CO dehydrogenase to utilize CO-containing industrial waste gas for expanding the versatility of CO dehydrogenase. *Sci. Rep.* 7:44323. doi: 10.1038/srep44323

Cossi, M., Rega, N., Scalmani, G., and Barone, V. (2003). Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J. Comput. Chem.* 24, 669–681. doi: 10.1002/jcc.10189

Dingwall, S., Wilcoxen, J., Niks, D., and Hille, R. (2016). Studies of carbon monoxide dehydrogenase from *Oligotropha carboxidovorans*. *J. Mol. Catal. B Enzymat.* 134, 317–322. doi: 10.1016/j.molcatb.2016.10.007

Dobbek, H., Gremer, L., Kiefersauer, R., Huber, R., and Meyer, O. (2002). Catalysis at a dinuclear [CuSMo(=O)OH] cluster in a CO dehydrogenase resolved at 1.1-Å resolution. *Proc. Natl. Acad. Sci. U.S.A.* 99, 15971–15976. doi: 10.1073/pnas.212640899

Dolg, M., Wedig, U., Stoll, H., and Preuss, H. (1987). Energy-adjusted *ab initio* pseudopotentials for the first row transition elements. *J. Chem. Phys.* 86, 866–872.

Dong, G., Phung, Q. M., Hallaert, S. D., Pierloot, K., and Ryde, U. (2017). H₂ binding to the active site of [NiFe] hydrogenase studied by multiconfigurational and coupled-cluster methods. *Phys. Chem. Chem. Phys.* 19, 10590–10601. doi: 10.1039/C7CP01331K

Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W., et al. (2003). A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* 24, 1999–2012. doi: 10.1002/jcc.10349

Dunning, T. and Hay, P. (1976). *Modern Theoretical Chemistry*. Plenum, NY: HF Schaefer III.

Dunning, T. H., and Hay, P. J. (1977). "Gaussian basis sets for molecular calculations," in *Methods of Electronic Structure Theory*, ed H. F. Schaefer (Boston, MA: Springer), 1–27.

Dunning, T. H. Jr. (1970). Gaussian basis functions for use in molecular calculations. I. contraction of (9s5p) atomic basis sets for the first-row atoms. *J. Chem. Phys.* 53, 2823–2833.

Eckert, F., and Klamt, A. (2002). Fast solvent screening via quantum chemistry: COSMO-RS approach. *AIChE J.* 48, 369–385. doi: 10.1002/aic.690480220

Ernzerhof, M., and Scuseria, G. E. (1999). Assessment of the Perdew–Burke–Ernzerhof exchange-correlation functional. *J. Chem. Phys.* 110, 5029–5036.

Gnida, M., Ferner, R., Gremer, L., Meyer, O., and Meyer-Klaucke, W. (2003). A novel binuclear [CuSMo] cluster at the active site of carbon monoxide dehydrogenase: characterization by X-ray absorption spectroscopy. *Biochemistry* 42, 222–230. doi: 10.1021/bi026514n

Gourlay, C., Nielsen, D. J., Evans, D. J., White, J. M., and Young, C. G. (2018). Models for aerobic carbon monoxide dehydrogenase: synthesis,

characterization and reactivity of paramagnetic Mo^V O(μ-S)Cu^I complexes. *Chem. Sci.* 9, 876–888. doi: 10.1039/c7sc04239f

Gourlay, C., Nielsen, D. J., White, J. M., Knottenbelt, S. Z., Kirk, M. L., and Young, C. G. (2006). Paramagnetic active site models for the molybdenum-copper carbon monoxide dehydrogenase. *J. Am. Chem. Soc.* 128, 2164–2165. doi: 10.1021/ja056500f

Greco, C., Ciancetta, A., Bruschi, M., Kulesza, A., Moro, G., and Cosentino, U. (2015). Influence of key amino acid mutation on the active site structure and on folding in acetyl-CoA synthase: a theoretical perspective. *Chem. Commun.* 51, 8551–8554. doi: 10.1039/c5cc01575h

Grimme, S., Ehrlich, S., and Goerigk, L. (2011). Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* 32, 1456–1465. doi: 10.1002/jcc.21759

Groysman, S., Hollingsworth, T., Hollingsworth, R. L., and Lord, R. L. (2018). Cooperative bimetallic reactivity of a heterodinuclear molybdenum-copper model of Mo-Cu CODH. *Dalton Trans.* 47, 10017–10024. doi: 10.1039/c8dt02323a

Hay, P. J., and Wadt, W. R. (1985a). *Ab initio* effective core potentials for molecular calculations. Potentials for K to Au including the outermost core orbitals. *J. Chem. Phys.* 82, 299–310. doi: 10.1063/1.448975

Hay, P. J., and Wadt, W. R. (1985b). *Ab initio* effective core potentials for molecular calculations. Potentials for the transition metal atoms Sc to Hg. *J. Chem. Phys.* 82, 270–283. doi: 10.1063/1.448799

Hille, R. (2013). The molybdenum oxotransferases and related enzymes. *Dalton Trans.* 42, 3029–3042. doi: 10.1039/c2dt32376a

Hille, R., Dingwall, S., and Wilcoxen, J. (2015). The aerobic CO dehydrogenase from Oligotropha carboxidovorans. *J. Biol. Inorgan. Chem.* 20, 243–251. doi: 10.1007/s00775-014-1188-4

Hofmann, M., Kassube, J. K., and Graf, T. (2005). The mechanism of Mo/Cu dependent CO dehydrogenase. *J. Biol. Inorgan. Chem.* 10, 490–495. doi: 10.1007/s00775-005-0661-5

Hu, L., and Chen, H. (2015). Assessment of DFT methods for computing activation energies of Mo/W-mediated reactions. *J. Chem. Theor. Comput.* 11, 4601–4614. doi: 10.1021/acs.jctc.5b00373

Husch, T., Freitag, L., and Reiher, M. (2018). Calculation of ligand dissociation energies in large transition-metal complexes. *J. Chem. Theor. Comput.* 14, 2456–2468. doi: 10.1021/acs.jctc.8b00061

Kaufmann, P., Duffus, B. R., Teutloff, C., and Leimkuhler, S. (2018). Functional studies on oligotropha carboxidovorans molybdenum–copper CO dehydrogenase produced in *Escherichia coli*. *Biochemistry* 57, 2889–2901. doi: 10.1021/acs.biochem.8b00128

Klamt, A., and Schüürmann, G. (1993). COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc. Perkin Trans.* 2, 799–805. doi: 10.1039/P29930000799

Kreß, O., Gnida, M., Pelzmann, A. M., Marx, C., Meyer-Klaucke, W., and Meyer, O. (2014). Reversible inactivation of CO dehydrogenase with thiol compounds. *Biochem. Biophys. Res. Commun.* 447, 413–418. doi: 10.1016/j.bbrc.2014.03.147

Lee, C., Yang, W., and Parr, R. G. (1988). Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* 37:785. doi: 10.1103/PhysRevB.37.785

Li, J., Andrejić, M., Mata, R. A., and Ryde, U. (2015). A computational comparison of oxygen atom transfer catalyzed by dimethyl sulfoxide reductase with Mo and W. *Eur. J. Inorganic Chem.* 2015, 3580–3589. doi: 10.1002/ejic.201500209

Li, J.-L., Mata, R. A., and Ryde, U. (2013). Large density-functional and basis-set effects for the DMSO reductase catalyzed oxo-transfer reaction. *J. Chem. Theor. Comput.* 9, 1799–1807. doi: 10.1021/ct301094r

Liu, L., Zhuang, Q., Zhu, Q., Liu, S., van Asperen, H., and Pihlatie, M. (2018). Global soil consumption of atmospheric carbon monoxide: an analysis

using a process-based biogeochemistry model. *Atmos. Chem. Phys.* 18:7913. doi: 10.5194/acp-18-7913-2018

Perdew, J. P. (1986). Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B* 33:8822. doi: 10.1103/PhysRevB.33.8822

Perdew, J. P., Burke, K., and Ernzerhof, M. (1996). Generalized gradient approximation made simple. *Phys. Rev. Lett.* 77:3865. doi: 10.1103/PhysRevLett.77.3865

Phung, Q. M., Domingo, A., and Pierloot, K. (2018). Dinuclear iron(II) spin-crossover Compounds: a theoretical study. *Chem. A Eur. J.* 24, 5183–5190. doi: 10.1002/chem.201704441

Phung, Q. M., Wouters, S., and Pierloot, K. (2016). Cumulant approximated second-Order perturbation theory based on the density matrix renormalization group for transition metal complexes: a benchmark study. *J. Chem. Theor. Comput.* 12, 4352–4361. doi: 10.1021/acs.jctc.6b00714

Rokhsana, D., Large, T. A., Dienst, M. C., Retegan, M., and Neese, F. (2016). A realistic *in silico* model for structure/function studies of molybdenum–copper CO dehydrogenase. *J. Biol. Inorgan. Chem.* 21, 491–499. doi: 10.1007/s00775-016-1359-6

Rovaletti, A., and Greco, C. (2018). Organophosphorous ligands in hydrogenase-inspired iron-based catalysts: a DFT study on the energetics of metal protonation as a function of P-atom substitution. *J. Phys. Org. Chem.* 31:e3748. doi: 10.1002/poc.3748

Santiago, B., and Meyer, O. (1996). Characterization of hydrogenase activities associated with the molybdenum CO dehydrogenase from Oligotropha carboxidovorans. *FEMS Microbiol. Lett.* 136, 157–162. doi: 10.1111/j.1574-6968.1996.tb08042.x

Shanmugam, M., Wilcoxen, J., Habel-Rodriguez, D., Cutsail III, G. E., Kirk, M. L., Hoffman, B. M., et al. (2013). 13C and 63,65Cu ENDOR studies of CO Dehydrogenase from Oligotropha carboxidovorans. Experimental evidence in support of a copper–carbonyl intermediate. *J. Am. Chem. Soc.* 135, 17775–17782. doi: 10.1021/ja406136f

Siegbahn, P. E. (2011). The effect of backbone constraints: the case of water oxidation by the oxygen-evolving complex in PSII. *ChemPhysChem* 12, 3274–3280. doi: 10.1002/cphc.201100475

Siegbahn, P. E. M., and Shestakov, A. F. (2005). Quantum chemical modeling of CO oxidation by the active site of molybdenum CO dehydrogenase. *J. Comput. Chem.* 26, 888–898. doi: 10.1002/jcc.20230

Stein, B. W., and Kirk, M. L. (2014). Orbital contributions to CO oxidation in Mo–Cu carbon monoxide dehydrogenase. *Chem. Commun.* 50, 1104–1106. doi: 10.1039/C3CC47705C

Wadt, W. R., and Hay, P. J. (1985). *Ab initio* effective core potentials for molecular calculations. Potentials for main group elements Na to Bi. *J. Chem. Phys.* 82, 284–298. doi: 10.1063/1.448800

Weigend, F., and Ahlrichs, R. (2005). Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for h to rn: design and assessment of accuracy. *Phys. Chem. Chem. Phys.* 7, 3297–3305. doi: 10.1039/b508541a

Wilcoxen, J., and Hille, R. (2013). The hydrogenase activity of the molybdenum/copper-containing carbon monoxide dehydrogenase of oligotropha carboxidovorans. *J. Biol. Chem.* 288, 36052–36060. doi: 10.1074/jbc.M113.522441

Xu, K., and Hirao, H. (2018). Revisiting the catalytic mechanism of Mo–Cu carbon monoxide dehydrogenase using QM/MM and DFT calculations. *Phys. Chem. Chem. Phys.* 20, 18938–18948. doi: 10.1039/C8CP00858B

Zhang, B., Hemann, C. F., and Hille, R. (2010). Kinetic and spectroscopic studies of the molybdenum-copper CO dehydrogenase from Oligotropha carboxidovorans. *J. Biol. Chem.* 285, 12571–12578. doi: 10.1074/jbc.M109.076851

Computational Protocol to Understand P450 Mechanisms and Design of Efficient and Selective Biocatalysts

Kersti Caddell Haatveit, Marc Garcia-Borràs and Kendall N. Houk *

Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA, United States

Cytochrome P450 enzymes have gained significant interest as selective oxidants in late-stage chemical synthesis. Their broad substrate scope enables them to be good candidates for their use in non-natural reactivity. Directed evolution evolves new enzyme biocatalysts that promote alternative reactivity for chemical synthesis. While directed evolution has proven useful in developing biocatalysts for specific purposes, this process is very time and labor intensive, and therefore not easily repurposed. Computational analysis of these P450 enzymes provides great insights into the broad substrate scope, the variety of reactions catalyzed, the binding specificity and the study of novel biosynthetic reaction mechanisms. By discovering new P450s and studying their reactivities, we uncover new insights into how this reactivity can be harnessed. We discuss a standard protocol using both DFT calculations and MD simulations to study a variety of cytochrome P450 enzymes. The approach entails theozyme models to study the mechanism and transition states via DFT calculations and subsequent MD simulations to understand the conformational poses and binding mechanisms within the enzyme. We discuss a few examples done in collaboration with the Tang and Sherman/Montgomery groups toward elucidating enzyme mechanisms and rationally designing new enzyme mutants as tools for selective C–H functionalization methods.

Keywords: density functional theory (DFT), theozymes, MD simulations, biocatalysis, cytochrome P450s, oxidations

INTRODUCTION

Cytochrome P450s are a highly-conserved class of enzymes that contain an Fe-heme cofactor with an axial cysteine ligand that can perform different oxidation reactions on a large variety of native substrates. (Ortiz de Montellano, 2005) P450 mechanisms have been extensively studied (Shaik et al., 2010) and require the oxidative generation of the reactive cofactor species, an iron(IV)-oxo ($Fe^{IV} = O$) radical cation. With this reactive species, the oxidation step in the hydroxylation reaction occurs via a hydrogen abstraction step followed by a radical rebound step between the cofactor and substrate (Meunier et al., 2004). Due to their broad substrate scope (Durairaj et al., 2016) and the growing interest developing selective methods for late-stage C–H oxidations, P450s have been utilized extensively as biocatalysts for many applications (Kumar, 2010), particularly

those that are difficult to perform with chemocatalysts. Cytochrome P450s are an excellent starting point for developing biocatalysts, and many great successes (Loskot et al., 2017) in modifying and engineering P450s have come from experimental directed evolution (non-rational) approaches (Romero and Arnold, 2009), pioneered by Nobel Laureate, Frances Arnold. However, these require many rounds of screening to arrive at a biocatalyst tailored for a specific purpose. This process can be sped up by developing a better understanding about how these P450s function, particularly by using computational methods. Rational approaches utilizing computational methods to study these enzymes can facilitate quicker access understanding their reactivity and predicting better variants.

While enzymes are complex systems, and their study can require elaborate computational methods, the Houk group has developed a simpler standard approach to understanding the mechanisms of complex (metallo)enzymes (particularly P450s) with great success. We utilize quantum mechanics (QM) calculations to study the transition states and mechanisms of theozyme ("theoretical enzyme" Tantillo et al., 1998) models (**Figure 1A**), a truncated portion of the enzyme that includes catalytically relevant active site residues and cofactors along with the substrate. We investigate the intrinsic mechanistic preferences of such transformations to understand which type of catalysis by the enzyme is most likely. Others have also utilized more rigorous approaches to study the enzyme catalysis, such as the cluster model (CM) approach used by Siegbahn, Thiel, and Himo (Siegbahn and Crabtree, 1997; Siegbahn and Himo, 2009, 2011; Blomberg et al., 2014) and QM/MM (Warshel and Levitt, 1976; Senn and Thiel, 2009; Karasulu and Thiel, 2015; Aranda et al., 2016) methods. In the beginning, CM were quite comparable to theozymes, but with the advance of computational power, CM has become more complex, and expanded to include systems of 200 atoms with other constraints on the amino acids residues to more accurately reflect the shape of the active site. QM/MM has been utilized to study the effects that mutations have on substrate binding. While these techniques are often more accurate and include the entire enzyme system, our approach is a more rapid method to study various enzyme systems (particularly P450s) and to understand their intrinsic preferences.

We combine DFT calculations with molecular dynamics (MD) simulations to explore the selectivities of enzymes and mutants. MD simulations allow us to study the substrate binding poses and positioning relative to the Fe = O active species. We can compare these geometries from MD simulations to the ideal transition state geometries determined from DFT to establish the enzyme's selectivity control. This method has allowed us to propose mutations to improve catalytic activity of enzymes. This review will focus on a few recent examples from the Houk group, first using QM theozymes to understand the feasibility of various mechanistic pathways within an enzyme, and subsequently, a combination of both QM and MD simulations to assess how the enzyme controls the reactivity and selectivity in a variety of P450 enzymes. Ultimately, we describe predictions to develop more efficient and selective enzyme mutants.

USE OF QUANTUM MECHANICS (QM) THEOZYME MODELS TO UNDERSTAND NOVEL ENZYMATIC MECHANISMS

The Houk group has used quantum mechanical DFT calculations to elucidate the mechanisms of novel P450 enzymes discovered in biosynthetic pathways. These demonstrate promiscuous, yet selective, reactivity or new oxidation reactivity that has great potential for biocatalysis. By modeling the reaction using a theozyme that includes the substrate and a truncated Fe-oxo heme active species, this method leads to the intrinsically preferred mechanism. Here, we discuss an example of the mechanism for a novel C–O bond formation performed by a P450 enzyme, GsfF, to generate the natural product, griseofulvin.
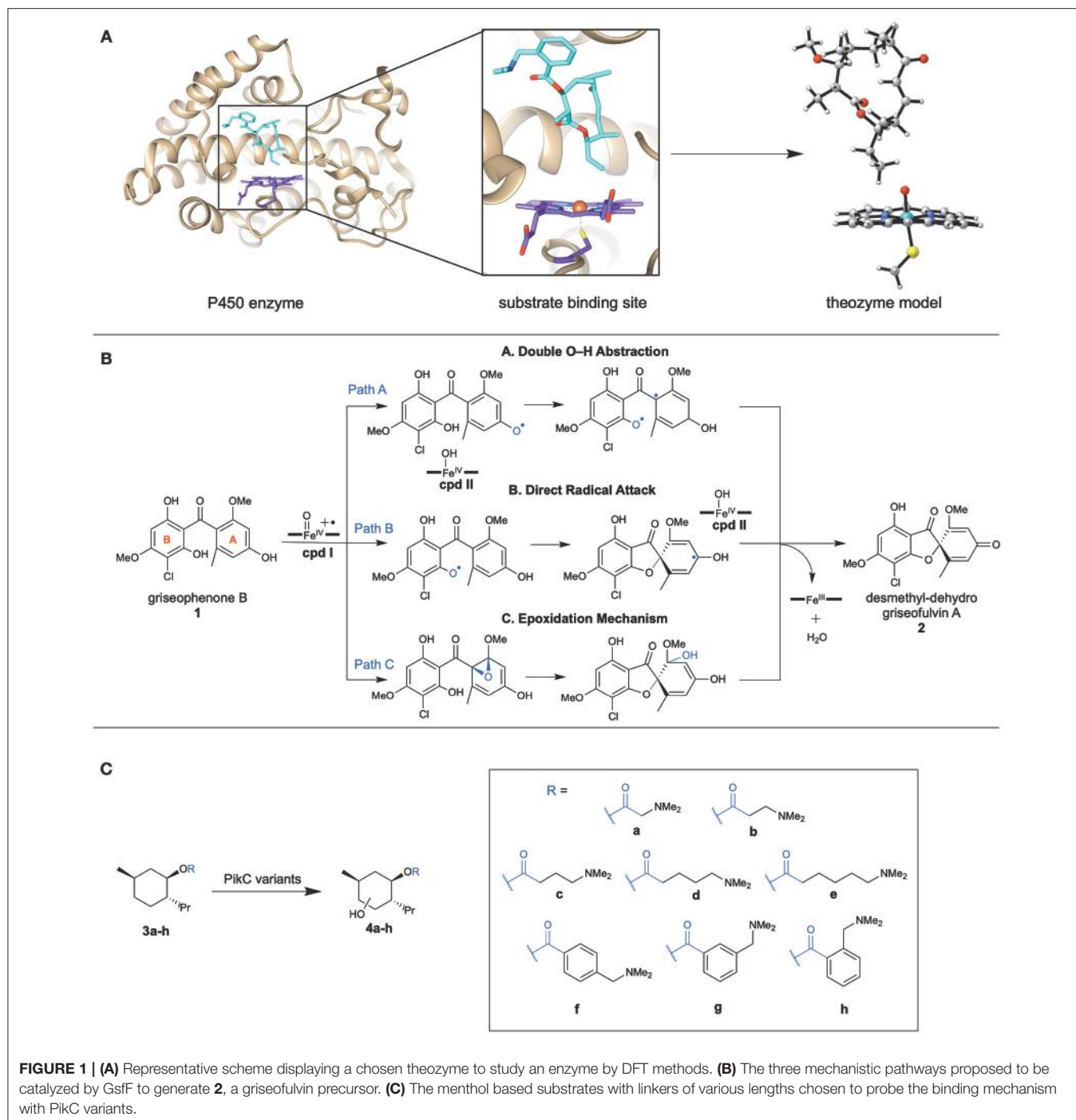
In 2016, the Tang group discovered GsfF (Chooi et al., 2010; Cacho et al., 2013) from the *gsf* gene cluster responsible for the synthesis of the natural product, griseofulvin. GsfF catalyzes an oxidative cyclization that generates the spirocyclic core within the final product. Initially, two different mechanisms, a double O–H abstraction (Pathway A) and epoxidation (Pathway C), were proposed (**Figure 1B**). Because it was difficult to determine a plausible mechanism experimentally, the various mechanisms were studied computationally (Grandner et al., 2016). An alternative direct radical attack C–O bond formation mechanism (Pathway B) was proposed within this paper to avoid diradical formation (**Figure 1B**). These mechanisms were studied computationally using DFT methods [at (U)B3LYP-D3(BJ)/LanL2DZ(Fe)/6-311+G(d,p)// (U)B3LYP/LanL2DZ(Fe)/6-31G(d) levels] with a theozyme compromised of a truncated Fe porphyrin model with the substrate (**Figure 1A**).

Pathway C was quickly discarded as a possibility due to the high barrier (18.7 kcal/mol) as compared to the double O–H abstraction and direct radical attack mechanism that are 0.5 and 0.0 kcal/mol, respectively. The differences between pathways A and B are small (0.5 kcal/mol), and therefore further analysis was considered to distinguish the most plausible mechanism. Pathway A requires a diradical intermediate, which is highly reactive. Due to the lack of crystal structure of GsfF, homology modeling was used to evaluate the possible binding positions of the substrate in the active site. Pathway A requires the substrate to reorient its binding pose within the enzyme between the mechanistic steps, which is likely not feasible without deleterious side reactions involving the reactive diradical intermediate. Consequently, Pathway B, which doesn't require substrate reorientations, is proposed to be the most plausible pathway.

COMBINATION OF THEOZYMES AND MOLECULAR DYNAMICS SIMULATIONS AS A TOOL TO UNDERSTAND P450S WITHIN BIOSYNTHETIC PATHWAYS

While the use of QM and theozyme models provides significant information on the mechanisms of novel enzymatic catalysis

FIGURE 1 | (A) Representative scheme displaying a chosen theozyme to study an enzyme by DFT methods. **(B)** The three mechanistic pathways proposed to be catalyzed by GsfF to generate **2**, a griseofulvin precursor. **(C)** The menthol based substrates with linkers of various lengths chosen to probe the binding mechanism with PikC variants.

of P450s, we usually combine QM and MD to obtain more knowledge of enzyme reactivity. DFT calculations of model theozymes demonstrates the intrinsic preferences for oxidation sites with P450 enzymes, while the MD simulations illustrate the various conformations and binding poses that the substrate can explore within the active site. We describe several collaborations with the Sherman and Montgomery groups to understand various P450 intrinsic reactivity for oxidations, how the enzyme controls binding orientations to overcome that innate

selectivity, and how new mutations can enhance selectivity and activity.

Prediction of Mutations to Improve WT Catalytic Activity in a P450 Monooxygenase

The Sherman group discovered a cytochrome P450 monooxygenase, PikC, a promiscuous enzyme that hydroxylates

12- and 14- membered macrolides, YC-17 and narbomycin (Xue et al., 1998). Co-crystal structures with both YC-17 and narbomycin revealed that the source of this selective, yet promiscuous, scope is due to a salt bridge interaction between the desosamine sugar anchor on the macrolides and E94 residue, enabling different substrates to react similarly if they maintain the desosamine sugar (Sherman et al., 2006). This interesting discovery led us to use PikC as the starting point for engineering a P450 biocatalyst that performs selective hydroxylations on a broad set of unnatural cyclic substrates. While the co-crystal structure provided great insights into one critical aspect of the binding mechanism, a dynamic view of PikC by utilizing both MD and QM enabled a rapid design of substrates and improved mutants (Narayan et al., 2015).

Menthol was chosen as a model non-native substrate due to the variety of C–H bonds and the vast amount of information on previous C–H functionalization attempts. MD simulations were performed on several menthol-based substrates that contain various lengths of synthetic anchors (**Figure 1C**). From those simulations, it was discovered that when the linker is too short (**3a**), the salt bridge interactions with E94/E85 are lost, and alternatively it develops a new interaction with E246, which unproductively orients the substrate such that unreactive C–H bonds are aimed at the iron-oxo reactive site (**Figure 2A**). Simulations also showed that the carboxylate group in D176 can unproductively interact with the substrate, forcing it to lose its close proximity to the Fe = O active species. It was also discovered that the loss of E94/E85 interactions destabilize the tertiary structure forcing PikC to adopt its open conformation, the preferred conformation in the apo state. However, when the linkers are longer (**3e**) and contain more rigid groups such as phenyls (**3f**), salt bridge interactions with E94/E85 give a new stabilizing interaction with E48 while avoiding the harmful E246 interaction. Experimentally, a PikC D50 N mutant demonstrated improved reactivity and the MD simulations confirmed this. Furthermore, it was predicted that PikC had greater selectivity for (−) menthol over (+) menthol as the latter substrate left the active site after 100 ns of MD simulations or oscillated in and out of the pocket, depending on the synthetic anchor group. Experimental results corroborated these predictions.

From these findings, we predicted two key mutation points, E246 and D176, that would improve the efficiency of the enzymes. From the mutagenesis experiments, the triple mutant PikC$_{D50N/E176Q/E246A}$, showed an increase in total turnover number (TTN) for all the unnatural substrates tested. The MD simulations demonstrated a higher frequency of closed conformations and persistent favorable salt bridge interactions in the triple mutant (**Figure 2B**) versus PikC$_{wt}$ or single mutant PikC$_{D50N}$.

While MD allowed us to predict essential mutations to improve catalysis in PikC, QM theozyme calculations were utilized to develop a model to predict the site selectivity of hydroxylation of these unnatural substrates. C–H abstraction transition states were computed for all the possible C–H bonds on menthol and C4 position was determined as the most reactive bond. In addition, the stereoselectivity at C4 was analyzed and a minimal difference in preference was found as the axial hydrogen abstraction barrier is lower by 0.7 kcal/mol. The

chiral environment of the enzyme is responsible for biasing for one stereoisomer, thus overriding the intrinsic reactivity. Key geometric parameters, H–O distance and O–H–C angles, from snapshots of 0.5 μs MD simulations were compared to the ideal geometry from DFT calculations (**Figures 2C,D**). This analysis indicated that the enzyme exposes the equatorial hydrogen in the appropriate geometry to the iron-oxo species more than axial hydrogen. Experimentally, hydroxylations of these substrates gave results consistent with the computational predictions.

Model for Origin of Selectivity for Salt Bridge Anchor Bound Substrate in a P450 Mutant

With this triple mutant PikC, the scope of the biocatalyst for oxidation of more complex substrates was explored. Anchor groups were simplified to have greater control of the selectivity, and we probed the origins of selectivity controlled by various anchors.

Our computational approach involved QM theozymes and MD simulations to predict the selectivities of substrate, **5**, modified with two linkers, **a** and **b** (**Figure 2E**). DFT calculations of the truncated system were utilized to analyze the intrinsic preference for oxidation for the various C–H bonds within the macrocycle (Gilbert et al., 2017). Positions C3 and C10 were determined to have the lowest C–H abstraction barriers in the two lowest energy conformers when using truncated model of macrolactone, **5**, where the linker was simplified as a methyl group (**Figure 2F**). These results are consistent with the two regioisomers observed experimentally. The stereochemistry at each of these positions was analyzed by comparing the C–H abstraction barrier for the axial and equatorial hydrogen at both C3 and C10. For both positions, the equatorial hydrogen abstraction had a lower energy barrier, due to better conjugation with the exocyclic olefin for C3 and less distortion to obtain conjugation for C10 (**Figure 2G**).

MD simulations were then performed on **5** attached with either linker **a** or linker **b**. These were docked into the crystal structure of the PikC mutant. MD simulation snapshots with linker **a** demonstrated that both C3 and C2 positions were maintained close to the reactive iron-oxo cofactor. Despite the favorable proximity of both positions, a C3–H reacts more rapidly than C2–H, due to its lower C–H abstraction barrier (5 kcal/mol lower) as shown by DFT calculations. The stereoselectivity of reaction at C3 was analyzed by comparing the geometry for C–H abstraction of both the axial and equatorial positions during MD trajectories to that of the ideal DFT transition state (**Figure 2H**). These results show that the PikC mutant active site bends the substrate so that the equatorial hydrogen has the appropriate orientation C–H abstraction. The same procedure was performed on the substrate with linker **b** which points the C10 close to the iron-oxo species. For C10, the axial hydrogen is in a geometry better for C–H abstraction (**Figure 2H**). However, the barrier for the equatorial C–H is much lower than that for the axial C–H. This process was repeated for many substrates, and experimental results were consistent with their predicted selectivity.
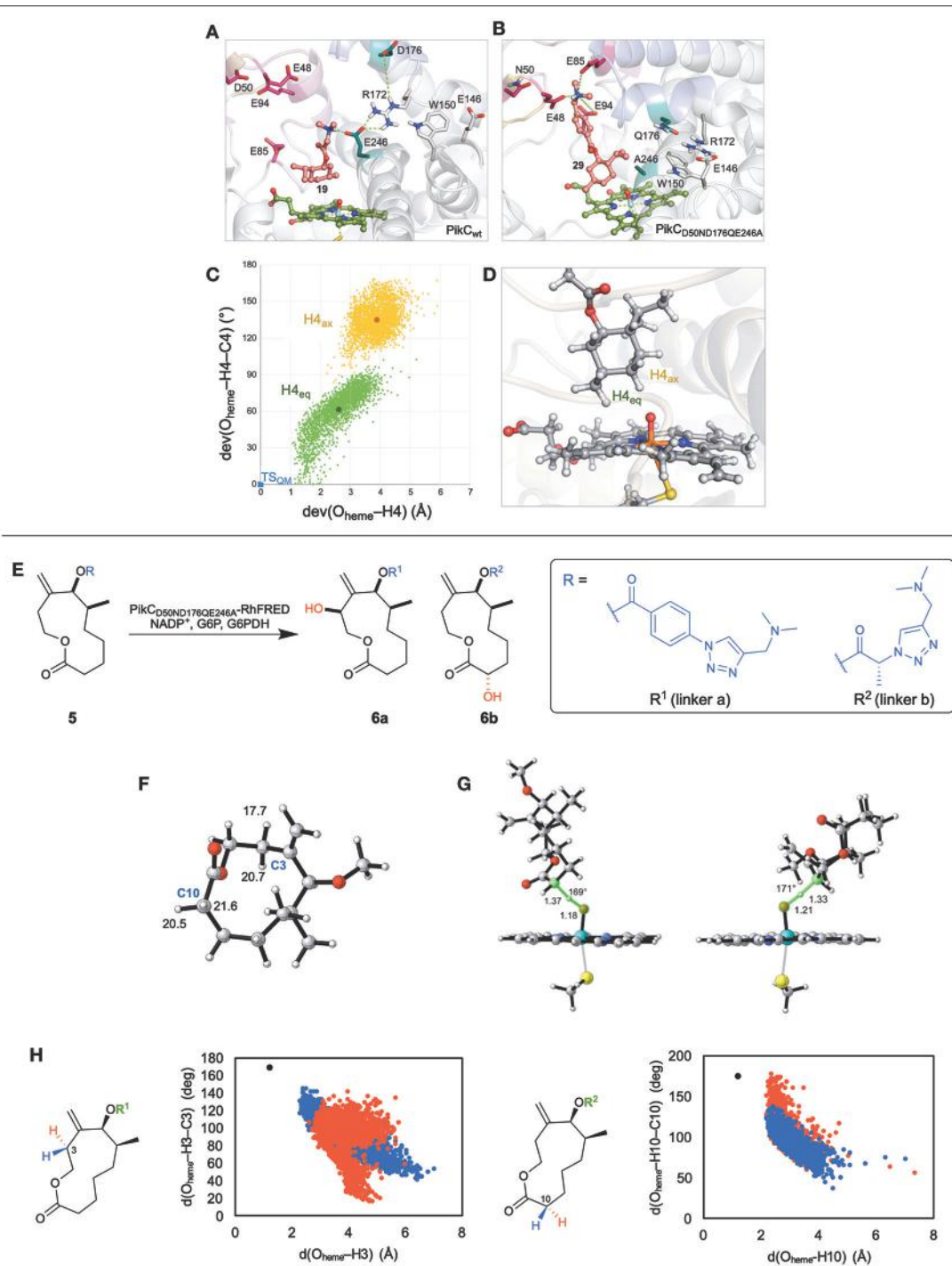
FIGURE 2 | (A) The detrimental binding interactions of menthol-based substrate with E246 in WT PikC enzyme. **(B)** The beneficial binding interactions of menthol-based substrate with E48, E85, and E94 in triple mutant PikC, D50N/D176Q/E246A. **(C)** Orientation of H4$_{eq}$ and H4$_{ax}$ relative to the reactive iron-oxo species from 0.5 µs MD trajectories. Each point represents a simulation snapshot. The x-axis captures deviations of O$_{heme}$−H distances and y-axis of O$_{heme}$−H−C angles from DFT optimized transition state (in blue). **(D)** Representative snapshot from MD simulations showing the closer proximity of equatorial hydrogen to active Fe-oxo species. **(E)** The regio and stereoselectivity of macrolactone substrate, **5**, with different linkers **a** and **b** in PikC triple mutant. **(F)** The lowest energy conformer of truncated **5** with C–H abstraction barriers for both axial and equatorial hydrogens at C3 and C10. **(G)** The geometry of the C–H abstraction transition state for the lowest energy hydrogen (equatorial) for C3 (left) and C10 (right). **(H)** Orientation of axial (red) and equatorial (blue) hydrogens on C3 in **5a** (left) and C10 in **5b** (right) relative to the reactive iron-oxo species from 500 ns MD trajectories. Each point represents a simulation snapshot. The x-axis captures the O$_{heme}$−H distances and y-axis of O$_{heme}$−H−C angles compared to that of DFT optimized transition state (in black). Adapted with permission from references Narayan et al. (2015) and Gilbert et al. (2017).

CONCLUSION

We have described several cases where we have combined QM theozyme models and MD simulations to understand enzymatic reactivities and selectivities. This computational protocol provides understanding of enzymatic mechanisms and substrate conformational space for various cytochrome P450s. We used this knowledge to predict mutations to establish better biocatalyst variants and predict the selectivities on natural and non-natural substrates. Although we showed some successful applications of this protocol, one of its limitations is that accurate energy barriers, that can be directly compared with experimental values, cannot be obtained. For these purposes, more computationally expensive methods and protocols are required. For instance, multiple QM/MM calculations can be performed to get more accurate free energies on different conformations sampled by the enzymatic system during MD simulations. This better accounts for the catalytically relevant conformations explored by the enzyme in solution. Although the accuracy of these methods is well-established, the large amount of computational time needed dramatically limit the speed and generality of those protocols.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

REFERENCES

Aranda, J., Zinovjev, K., Swiderek, K., Roca, M., and Tuñón, I. (2016). Unraveling the reaction mechanism of enzymatic c5-cytosine methylation of DNA. A combined molecular dynamics and QM/MM study of wild type and Gln199 variant. *ACS Catal.* 6, 3262–3276. doi: 10.1021/acscatal.6b00394

Blomberg, M. R., Borowski, T., Himo, F., Liao, R.-Z., and Siegbahn, P. E. (2014). Quantum chemical studies of mechanisms for metalloenzymes. *Chem. Rev.* 114, 3601–3658. doi: 10.1021/cr400388t

Cacho, R. A., Chooi, Y.-H., Zhou, H., and Tang, Y. (2013). Complexity generation in fungal polyketide biosynthesis: a spirocycle-forming P450 in the concise pathway to the antifungal drug griseofulvin. *ACS Chem. Biol.* 8, 2322–2330. doi: 10.1021/cb400541z

Chooi, Y.-H., Cacho, R., and Tang, Y. (2010). Identification of the viridicatumtoxin and griseofulvin gene clusters from *Penicillium aethiopicum*. *Chem. Biol.* 17, 483–494. doi: 10.1016/j.chembiol.2010.03.015

Durairaj, P., Hur, J.-S., and Yun, H. (2016). Versatile biocatalysis of fungal cytochrome P450 monooxygenases. *Microb. Cell Fact.* 15:125. doi: 10.1186/s12934-016-0523-6

Gilbert, M. M., DeMars, M. D., Yang, S., Grandner, J. M., Wang, S., Wang, H., et al. (2017). Synthesis of diverse 11- and 12-membered macrolactones from a common linear substrate using a single biocatalyst. *ACS Cent Sci.* 3, 1304–1310. doi: 10.1021/acscentsci.7b00450

Grandner, J. M., Cacho, R. A., Tang, Y., and Houk, K. N. (2016). Mechanism of the P450-catalyzed oxidative cyclization in the biosynthesis of griseofulvin. *ACS Catal.* 6, 4506–4511. doi: 10.1021/acscatal.6b01068

Karasulu, B., and Thiel, W. (2015). Amine oxidation mediated by N-methyltryptophan oxidase: computational insights into the mechanism, role of active-site Residues, and covalent flavin binding. *ACS Catal.* 5, 1227–1239. doi: 10.1021/cs501694q

Kumar, S. (2010). Engineering cytochrome P450 biocatalysts for biotechnology, medicine, and bioremediation. *Expert Opin. Drug Metab. Toxicol.* 6, 115–131. doi: 10.1517/17425250903431040

Loskot, S. A., Romney, D. K., Arnold, F. H., and Stoltz, B. M. (2017). Enantioselective total synthesis of nigelladine a via late-stage c–h oxidation enabled by an engineered P450 enzyme. *J. Am. Chem. Soc.* 139, 10196–10199. doi: 10.1021/jacs.7b05196

Meunier, B., de Visser, S. P., and Shaik, S. (2004). Mechanism of oxidation reactions catalyzed by cytochrome P450 enzymes. *Chem. Rev.* 104, 3947–3980. doi: 10.1021/cr020443g

Narayan, A. R., Jiménez-Osés, G., Liu, P., Negretti, S., Zhao, W., and Gilbert, M. M. (2015). Enzymatic hydroxylation of an unactivated methylene C–H bond guided by molecular dynamics simulations. *Nat. Chem.* 7, 653–660. doi: 10.1038/nchem.2285

Ortiz de Montellano, P. (2005). *Cytochrome P450: Structure, Mechanism, and Biochemistry, 3e.* New York, NY: Kluwer Academic/Plenum Publishers.

Romero, P. A., and Arnold, F. N. (2009). Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* 10, 866–876. doi: 10.1038/nrm2805

Senn, H. M., and Thiel, W. (2009). QM/MM methods for biomolecular systems. *Angew. Chem. Int. Ed.* 48, 1198–1229. doi: 10.1002/anie.200802019

Shaik, S., Cohen, S., Wang, Y., Chen, H., Kumar, D., and Thiel, W. (2010). P450 enzymes: their structure, reactivity, and selectivity-modeled by QM/MM calculations. *Chem. Rev.* 110, 949–1017. doi: 10.1021/cr900121s

Sherman, D. H., Li, S., Yermalitskaya, L. V., Kim, Y., Smith, J. A., Waterman, M. R., et al. (2006). The structural basis for substrate anchoring, active site selectivity, and product formation by P450 PikC from *Streptomyces venezuelae*. *J. Biol. Chem.* 281, 26289–26297. doi: 10.1074/jbc.M605478200

Siegbahn, P. E., and Himo, F. (2009). Recent developments of the quantum chemical cluster approach for modeling enzyme reactions. *J. Biol. Inorg. Chem.* 14, 643–651. doi: 10.1007/s00775-009-0511-y

Siegbahn, P. E. M., and Crabtree, R. (1997). Mechanism of C–H activation by diiron methane monooxygenases: quantum chemical studies. *J. Am. Chem. Soc.* 119, 3103–3113. doi: 10.1021/ja963939m

Siegbahn, P. E. M., and Himo, F. (2011). The quantum chemical cluster approach for modeling enzyme reactions. *Rev. Comput. Mol. Sci.* 1, 323–336. doi: 10.1002/wcms.13

Tantillo, D. J., Chen, J., and Houk, K. N. (1998). Theozymes and compuzymes: theoretical models for biological catalysis. *Curr. Opin. Chem. Biol.* 2, 743–750. doi: 10.1016/S1367-5931(98)80112-9

Warshel, A., and Levitt, M. (1976). 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. doi: 10.1016/0022-2836(76)90311-9

Xue, Y., Wilson, D., Zhao, L., Liu, H. W., and Sherman, D. H. (1998). Hydroxylation of macrolactones YC-17 and narbomycin is mediated by the PikC-encoded cytochrome P450 in *Streptomyces venezuelae*. *Chem. Biol.* 5, 661–667. doi: 10.1016/S1074-5521(98)90293-9

Theoretical Studies on Mechanism of Inactivation of Kanamycin A by 4′-O-Nucleotidyltransferase

Sergio Martí[1], Agatha Bastida[2] and Katarzyna Świderek[1]*

[1] Departament de Química Física i Analítica, Universitat Jaume I, Castelló de La Plana, Spain, [2] Departamento de Química Bio-orgánica, Instituto de Química Orgánica General (CSIC), Madrid, Spain

This work is focused on mechanistic studies of the transfer of an adenylyl group (Adenoside-5′-monophosfate) from adenosine 5′-triphosphate (ATP) to a OH-4′ hydroxyl group of an antibiotic. Using hybrid quantum mechanics/molecular mechanics (QM/MM) techniques, we study the substrate and base-assisted mechanisms of the inactivation process of kanamycin A (KAN) catalyzed by 4′-O-Nucleotidyltransferase [ANT(4′)], an active enzyme against almost all aminoglycoside antibiotics. Free energy surfaces, obtained with Free Energy Perturbation methods at the M06-2X/MM level of theory, show that the most favorable reaction path presents a barrier of 12.2 kcal·mol^{-1} that corresponds to the concerted activation of O4′ from KAN by Glu145. In addition, the primary and secondary ^{18}O kinetic isotope effects (KIEs) have been computed for bridge O3α, and non-bridge O1α, O2α, and O5′ atoms of ATP. The observed normal 1°-KIE of 1.2% and 2°-KIE of 0.07% for the Glu145-assisted mechanism are in very good agreement with experimentally measured data. Additionally, based on the obtained results, the role of electrostatic and compression effects in enzymatic catalysis is discussed.

Keywords: kanamycin, antibiotic, QM/MM, aminoglycosides, kinetic isotope effects, O-Nucleotidyltransferase, electrostatic effects, compression effects

INTRODUCTION

Aminoglycoside antibiotics (AGAs) belong to the class of agents used to treat serious infections caused by bacteria that either multiply very quickly or are difficult to treat. Their role is to stop bacteria from producing proteins needed for their survival (Shaw et al., 1993). Unfortunately, due to widespread usage of these antibiotics in clinical treatment, bacterial strains have appeared that make these compounds ineffective. In fact antibiotic-resistant bacterial infection has become a concerning global threat to human health, according to World Health Organization (WHO) reports (Nature, 2013; Berendonk et al., 2015).

Several mechanisms of resistance to AGAs have been proposed including (a) the presence of AGA-modifying enzymes (Ramirez and Tolmasky, 2010), (b) the decrease of bacteria membrane permeability toward AGA uptake into the bacteria and extrusion of AGAs from the cell by efflux pumps (Kumar and Schweizer, 2005), and (c) the modification of the drug target as a result of mutations in the ribosome (Pfister et al., 2003) or methylations by 16S rRNA methyltransferases influencing the binding of AGAs (Doi and Arakawa, 2007; Leban et al., 2017). Nevertheless, the major mechanism of bacterial resistance in clinical isolates of gram-negative and gram-positive bacteria is assigned to enzymatic modification of the amino or hydroxyl groups of AGAs (Vakulenko and Mobashery, 2003).

FIGURE 1 | Sites of possible modification on kanamycin A by various AGAs-modifying enzymes. The arrows point to the sites of modification by acetyltransferases (AACs), phosphotransferases (APHs), and nucleotidyltransferases (ANTs).
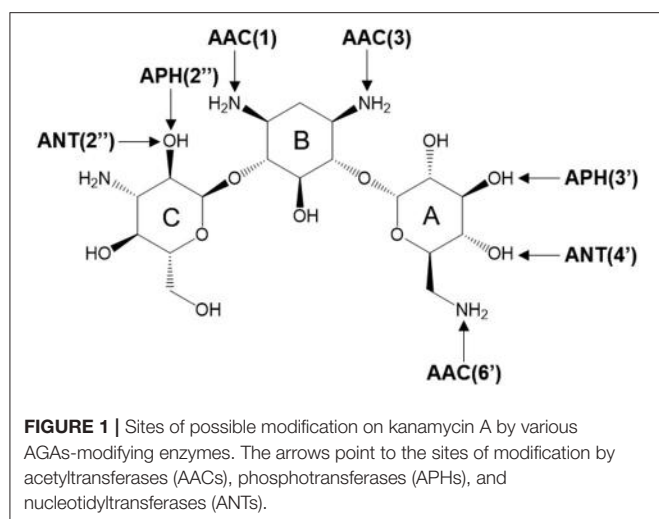
There are three different families of AGA-modifying enzymes: the acetyl-CoA-dependent aminoglycoside acetyltransferases (AACs), the ATP-dependent aminoglycoside phosphotransferases (APHs), and the ATP-dependent aminoglycoside nucleotidyltransferases (ANTs) (Becker and Matthew, 2013). Sites of modification in kanamycin A (KAN) by different AGA-modifying enzymes are indicated on **Figure 1**. Each of these families catalyzes a different type of reaction. Thus, APHs catalyze regiospecific transfer of γ-phosphoryl group of ATP to one of the hydroxyl substituents presented in AGAs, AACs facilitate N/O-acetylation groups in AGAs, and finally ANTs promote reaction between MgATP and AGAs allowing to form O-adenylylated aminoglycoside and the magnesium chelate of inorganic pyrophosphate (MgPP$_i$) formed after the reaction, as presented in **Figure 2**.

Herein, we focus on mechanistic studies of reaction catalyzed by 4′-O-Nucleotidyltransferase from *S. aureus*, which modifies KAN antibiotic in the 4′-OH [ANT(4′)] position. KAN is a second-line injectable drug used in the treatment of tuberculosis (TB) (Ventola, 2015). TB is classified among 18 of the most serious drug-resistant threats[1], since in many cases, as reported by WHO, TB can be resistant to first-line drugs (600,000 cases in year 2016) and/or to any other treatment (490,000 with multidrug-resistant TB). The role of KAN is to interact with the 30S subunit of prokaryotic ribosomes (exactly to bind in 16S rRNA, at the tRNA acceptor A site) in order to substantially rise the amounts of mistranslation (incorrect alignment with the mRNA), and it indirectly inhibits translocation that provokes insertion of the wrong amino acid into the peptide during protein synthesis (Pestka, 1974; Misumi and Tanaka, 1980; Carter et al., 2000). KAN belongs to an AGA family containing a 4,6-disubstituted 2-deoxystreptamine core (B ring) glycosidically linked at the 4 position to a glucosaminopyranose (ring A) and at position 6 to a N-acetylglucopyranose (ring C), as shown on **Figure 1**.

---

[1]Center of Disease Control and Prevention. Available online at: https://www.cdc.gov/drugresistance/biggest_threats.html
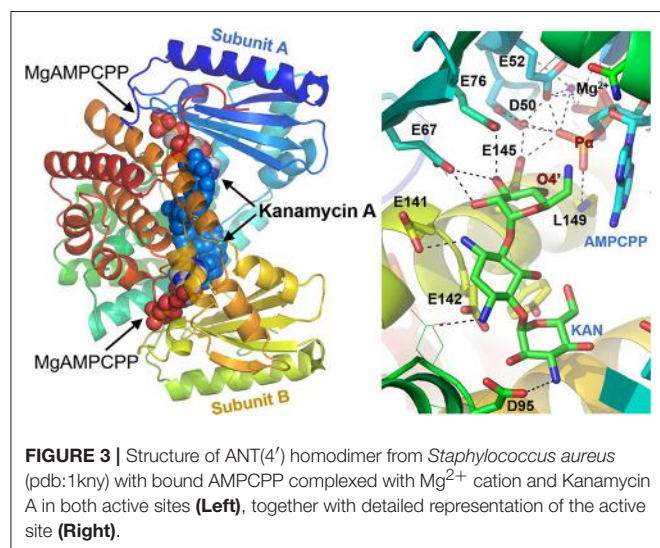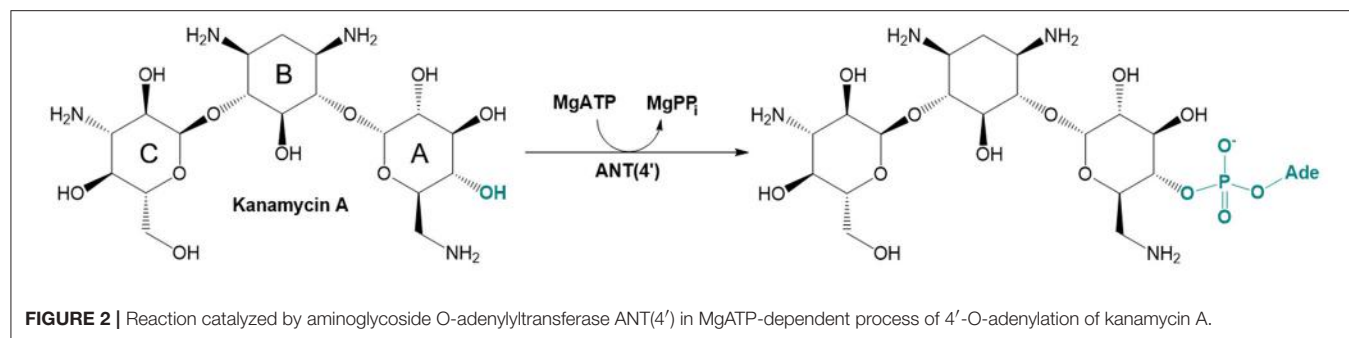
The crystal structures of the D80Y variant of ANT(4′) with bound KAN and AMPCPP (adenosince-5′-[(α,β)-methylene]triophosphate) indicate that the active form of the enzyme is a homodimer (Pedersen et al., 1995). Each monomer binds both substrates, nucleotide and aminoglycoside (KAN). However, and interestingly, the reaction takes place between substrates that are bound to different subunits. In general, as shown in **Figure 3**, the two nucleotides are far apart, and the two KAN molecules are at a distance of 3.5 Å from each other. The amino acid residues from the N-terminal domain interact mostly with the triphosphate moiety of the nucleotide, while those from C-terminal with the KAN (Revuelta et al., 2010).

Based on the analysis of the organization of the active site, it is believed that in order to access their binding pockets, both substrates must diffuse via the same cavity characterized by a strong negative electrostatic potential (Matesanz et al., 2012). Thus, it is assumed that KAN should bind before the nucleotide, reducing in this way possible repulsive interaction by an anionic specie (ATP) and highly charged regions of protein (Chen-Goodspeed et al., 1999). This is in contradiction to the studies previously done for ANT(2′), where an inverse order of binding was suggested (Lombardini and Cheng-Chu, 1980; Gates and Northrop, 1988). Nevertheless, the presence of both substrates is required for catalysis to take place.

The results of previous kinetic studies (Pelt et al., 1986) of *S. aureus* ANT(4′), with gentamicin as a substrate, indicate that the mechanism of antibiotic inactivation involves direct nucleotidyl transfer from ATP to the aminoglycoside. In fact, what was exactly observed experimentally, is inversion of stereochemistry on the α-phosphorus atom, suggesting that the chemical step involves an uneven number of phosphotransfers. Additionally, a concerted and slightly associative structure of transition state (TS) for the enzymatic 4′-adenylation of KAN was suggested from the measurements of ^{18}O kinetic isotope effects (KIEs) for key oxygen atoms (Gerratana et al., 2001).

The transferred group in this reaction is the nucleoside monophosphate (see **Figure 2**), which is a monosubstituted derivative of the phosphoryl group, and it can be assumed that its chemical nature along the reaction path is similar to the non-substituted relative. Thus, as it is well known, phosphoryl transfer reactions with low-lying d-orbitals on the phosphorus atoms that permit the existence of phosphorus pentavalent moiety as intermediate (Cleland and Hengge, 2006; Marcos et al., 2008; Kamerlin, 2011; Wymore et al., 2014) have been proposed to proceed by two alternative mechanisms (Kamerlin et al., 2008), i.e., associative, where nucleophilic attack takes place before the living group departures, and dissociative, in which withdrawal of the living group preludes the nucleophilic attack. The reaction can proceed via only one step in which asynchronous forming and breaking bonds can still be observed since associative-like or dissociative-like mechanisms can be found for concerted paths (Bordes et al., 2017).

The degree to which nucleotidyl transfer goes by an associative mechanism is difficult to determine, since experimental characterization of TS geometry is not feasible. The nature of the TS for the non-enzymatic (Zhang et al.,

FIGURE 2 | Reaction catalyzed by aminoglycoside O-adenylyltransferase ANT(4′) in MgATP-dependent process of 4′-O-adenylation of kanamycin A.

FIGURE 3 | Structure of ANT(4′) homodimer from *Staphylococcus aureus* (pdb:1kny) with bound AMPCPP complexed with Mg^{2+} cation and Kanamycin A in both active sites **(Left)**, together with detailed representation of the active site **(Right)**.

2014) and enzymatic (Sucato et al., 2007, 2008; Oertell et al., 2012, 2014) nucleotidyl transfer reaction has recently been studied in human DNA polymerase β using different modified dNTP substrates that differed in the chemical structure.

In this paper, insight into the nucleotidyl transfer mechanism for KAN inactivation by ANT(4′) at molecular level is done using a QM/MM approach (Ridder and Mulholland, 2003). Molecular details of reactions involving ATP-cofactor are essential for understanding the mechanism phosphoryl transfer reactions promoted by presence of metals in the living cell, responsible for a wide range of processes (Cherepanov et al., 2008). In order to obtain values which can be directly compared with experimental data, the free energy surfaces were computed and compared to experimentally measured rate constants using the transition state theory (TST). Moreover, the ^{18}O kinetic isotope effects (KIEs) for bridge and non-bridge oxygen of ATP were determined and compared with experimental values. Finally, electrostatic potential generated by ANT(4′) in the active site was computed and compared with non-catalyzed reaction in order to understand the role of electrostatic effects for this particular reaction.

COMPUTATIONAL METHODS

System Setup

A molecular model of aminoglycoside 4'-nucleotidyltransferase [ANT(4′)] was built based on biological assembly of homodimer structure with PDB ID 1KNY (Pedersen et al., 1995). The X-ray structure contains two KAN molecules bound into both active sites present in the enzyme, together with the non-hydrolizable ATP-cofactor, AMPCPP-Mg^{2+} ion. In order to modify AMPCPP to ATP, the carbon atom was changed by oxygen in the α3-position. Thus, the final enzymatic model consisted of two subunits (253 residues each), two Mg^{2+} chelate of ATP, and two KAN molecules.

Missing hydrogen atoms were added to the enzyme structure using the tLEAP (Schafmeister et al., 1995) program based on results of pKa computed for titratable residues at pH 7 using the PropKa 3.1 program (Olsson et al., 2011; Sondergaard et al., 2011). Results of pKa calculations are presented in **Table S1**. Interestingly, all glutamic acid residues for which the predicted values of pKa were over 7 are involved in the strong H-bond interactions with KAN (Glu67 ring A; Glu76 ring A; Glu142 ring C) and they should not be protonated. This misleading result is an effect of absence of hydrogen atoms in the substrate once the pKa values are computed. For the rest of the titratable residues, standard values of pKa were obtained. The geometrical analysis of specific interactions involving histidine residues were done and the following protonation states for these residues were assigned: for His17, 66, 181, and 241 the nitrogen atom of imidazole ring in δ-position was protonated, while in case of His100, 180, 181, and 207, hydrogen was added to the nitrogen atom in ε-position. Finally, no S-S bridge between Cys residues was detected. The kanamycin molecule was assumed to be a neutral species with the total charge set to 0. All hydroxyl groups were protonated, and all amine groups were assigned to be neutral.

Subsequently, neutralization of the system was achieved by adding 40 Na^+ counterions in the electrostatically most favorable positions. Finally, a full enzyme model was solved in a 10 × 10 × 10 nm^3 cubic box of TIP3P (Jorgensen et al., 1983) water. Then, 10 ns of NVT molecular dynamic (MD) simulations with a time step of 1 fs at 300 K were carried out after prior optimization, heating (from 0 to 300 K with 0.001 K temperature increment) and equilibration processes of 100 ps using an AMBER (Duan et al., 2003) force field implemented in NAMD (Phillips et al., 2005) software. The temperature during the

MD simulation was controlled using the Langevin thermostat (Grest and Kremer, 1986). In order to improve the time of simulations, cut-offs for non-bonding interactions were applied using a smooth switching function between 14.5 and 16 Å. Additionally, periodic boundary conditions (PBC) were used. The parameters developed previously by Carlson and co-workers were used to describe the ATP-cofactor at MM level (Meagher et al., 2003). Missing parameters for KAN were generated using the antechamber tool and GAFF (Wang et al., 2004), with charges computed at Austin Model 1 (AM1) (Dewar et al., 1985) level. The obtained parameters are presented in **Table S2**. Root-mean-square-deviation (RMSD) and more analysis of MD simulations are provided in **Supplementary Material**. According to the time-dependent evolution of the RMSD for the position of C-Cα-N, atoms of the protein backbone, the system can be considered equilibrated (see **Figure S1**).

Potential Energy Surfaces

The last structure of MD simulation was then employed for the QM/MM calculations using the M06-2X hybrid functional (Zhao and Truhlar, 2008a,b) with the standard 6-31+G(d,p) basis set, to treat the QM subset of 76 atoms (see **Figure S2**). The rest of the system was described applying the AMBER and TIP3P force fields, as implemented in the fDYNAMO library (Field et al., 2000; Krzemińska et al., 2015). The position of all atoms above 20 Å from KAN was frozen. In order to explore proposed mechanisms, the potential energy surfaces (PESs) at AM1, semiempirical level combined with MM force fields were computed. Based on chosen structures, a micro-macro iteration optimization algorithm (Turner et al., 1999; Martí et al., 2005) at M06-2X/MM level was used to localize, optimize, and characterize the TS structures using a Hessian matrix containing all the coordinates of the QM subsystem, whereas the gradient norm of the remaining movable atoms was maintained at <0.01 kcal·mol^{-1}·Å^{-1}. Intrinsic reaction coordinates (IRCs) were traced down from located TSs to the valleys of the reactants and products in mass-weighted Cartesian coordinates. Subsequently, last structures from IRC were used to localize, optimize, and characterize the ground states, i.e., reactant and product complexes.

Free Energy Calculations

To describe the mechanism of the reaction in condensed media, a free energy perturbation (FEP) (Świderek et al., 2013) method was used employing the M06-2X DFT functional to describe the QM sub-set of atoms. This method is usually second in choice, after umbrella sampling (Torrie and Valleau, 1977) (US), in terms of a potential mean force (PMF), and the selection is often dictated by a number of coordinates involved in approximate reaction coordinate (Świderek et al., 2015a). However, in these studies FEP employment originates in the limitations of semiempirical methods (SM) since, as it was shown recently by Otyepka and co-workers (Mlýnský et al., 2014) SM can easily fail and provide wrong conclusions about mechanistic paths in such theoretical models where the organic phosphorus atom is directly involved in chemical reactions.

Since FEP required the sampling of the environment along IRC traced previously from TS located at QM/MM level, the free energy pathway is obtained along a realistic reaction coordinate. Nevertheless, the limitation of this technique is assigned to lack of sampling on the chemical system, since just one characterized TS structure is used. Nevertheless, the FEP method opens the possibility of exploring the reaction path directly at a high level of theory. Thus, the QM wave function obtained at DFT level is polarized by the charges of the MM subset of atoms.

Free energy differences were estimated by mean of FEP methodology for the structures obtained along the IRC characterized by a single s coordinate:

$$s_i = s_{i-1}$$
$$+ \sqrt{\sum_{j \in QM} m_j \left(\left(x_{j,i} - x_{j,i-1} \right)^2 + \left(y_{j,i} - y_{j,i-1} \right)^2 + \left(z_{j,i} - z_{j,i-1} \right)^2 \right)}$$

(1)

where $x_{j,i}$, $y_{j,i}$, and $z_{j,i}$ are the coordinates of the i_{th} structure for the j_{th} QM atom belonging to the IRC traced from the transition state structure (x_{j0}, y_{j0}, and z_{j0} coordinates) and m_j represents the corresponding masses of the atoms. Therefore, the free energy relative to the reactant is expressed as a function of the s coordinate as explained elsewhere (Świderek et al., 2013; Viciano et al., 2015). The MD simulations for the FEP calculation were performed at 300 K, using the NVT ensemble for the each window. 20 ps of production, with a time step of 1 fs, were completed. The total amount of windows required to generate the complete free energy path was 66 for the ATP-assisted, 37 for the Glu145-assisted mechanism in enzyme and 101 for reaction in water.
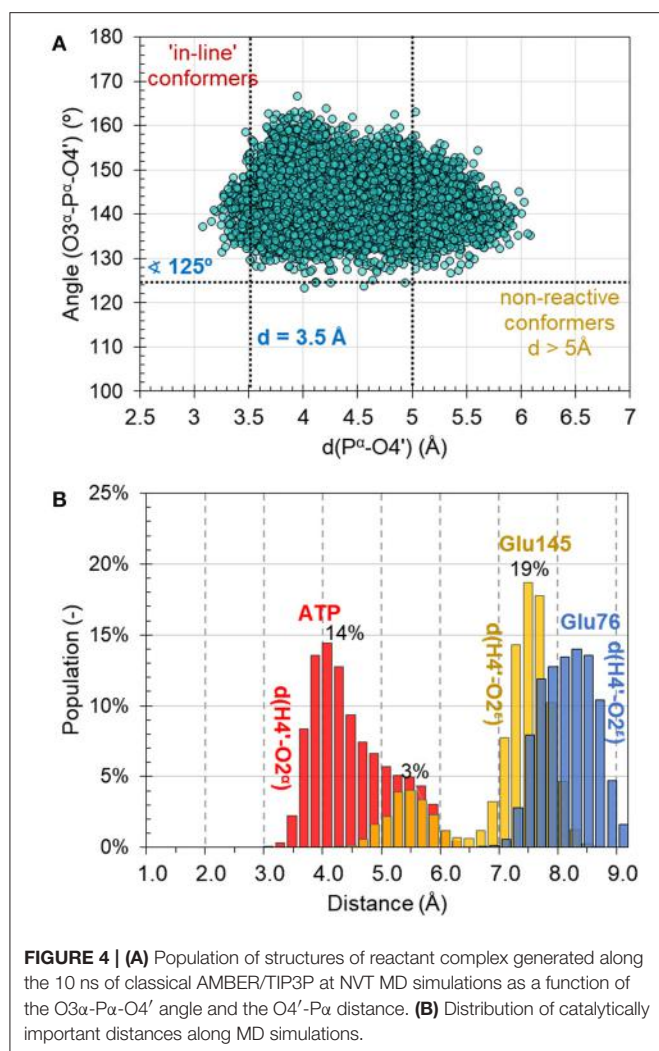
Kinetic Isotope Effects

Kinetic isotope effects (KIEs) have been computed for isotopic substitutions of key atoms, from the TSs and the reactant complexes localized at the M06-2X/MM level of theory. Detailed information about the method used herein for computing KIEs can be found elsewhere (Świderek et al., 2014, 2017a).

RESULTS AND DISCUSSION

Reaction Mechanism

MD simulations on the reactant complex at MM level shed some light onto the possible reaction mechanisms, to be later explored at QM/MM level. First, as can be seen in **Figure 4A**, the "in-line" position of the 4′-hydroxyl group of KAN and pyrophosphate of ATP required for ensuring direct attack at the α-phosphorus atom, exists in only 2% of cases of 10,000 overall structures saved along 10 ns of MD simulations [note: distance between O4′-Pα$^{(ATP)}$ not longer than 3.5 Å, and angle not smaller than 125° for O3α-O4′-Pα$^{(ATP)}$ were assigned as boundary conditions, similar to the definition used by York and co-workers (Heldenbrand et al., 2014)]. This result supports the statement that in-line conformations are often rare. Moreover, the free energy required to bring the nucleophile in-line has been predicted, at least in some cases, to be only modest and likely not a dominant factor

FIGURE 4 | (A) Population of structures of reactant complex generated along the 10 ns of classical AMBER/TIP3P at NVT MD simulations as a function of the O3α-Pα-O4′ angle and the O4′-Pα distance. (B) Distribution of catalytically important distances along MD simulations.

on the overall catalytic rate (Min et al., 2007). In the X-ray structure, the same distance between nucleophile and Pα(ATP) of ATP imitator, AMPCPP equal to 5.0 Å, was originally interpreted as corresponding to the inactive conformation. Based on analysis made for this distance it is observed that the change of AMPCPP to ATP reduces the presence of non-reactive conformations to 15% of the entire population of structures generated along the MD. Nevertheless, a rather large spread in the values of the distances and angles can still be observed, suggesting that both substrates are not completely immobilized in the binding pocket of ANT(4′).
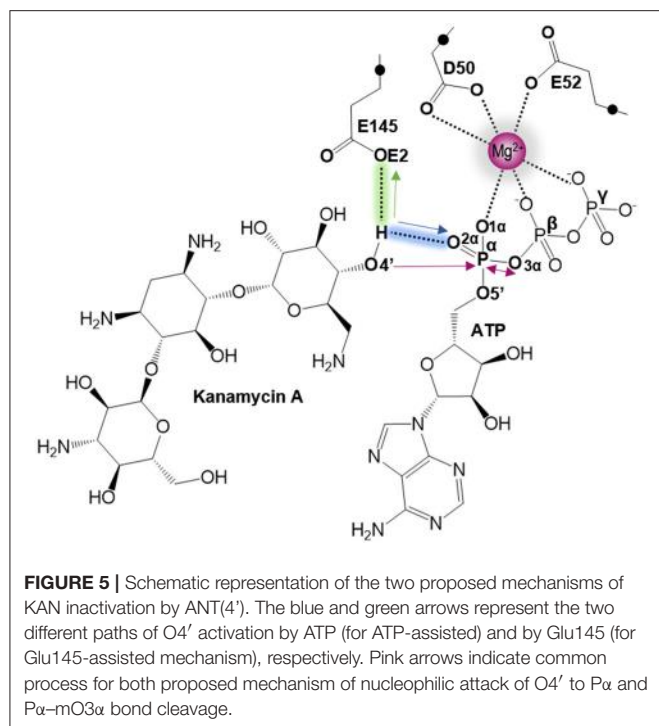
Secondly, the analysis of the distribution of distances indicates that the O4′ atom of KAN, which plays the role of a nucleophile in the studied reaction, can be activated by either O2α(ATP) of the ATP-cofactor, or OE2(Glu145) from Glu145, and rather not by Glu76 in the wild-type variant of ANT(4′). According to the population analysis, in c.a. 32% of the snapshots, the strongest interaction between H4′ and O2α was found with distance lower than 4 Å, as shown in **Figure 4B**.

Glu145 approaches the OH4′ group at a distance lower than 5 Å only in 5% of cases. However, Glu76 is located much farther and not closer than 7 Å. These observations are especially

interesting taking into account results of experimental kinetic studies done for a single Glu145Gln, or Glu76Gln, and double Glu76Gln/Glu145Gln mutated variants of ANT4′ (Matesanz et al., 2012). In these studies, a loss of catalytic activity was observed only in case of the double mutated enzyme, suggesting that ANT(4′) is equipped with duplicated basis catalyst. However, from results of MD obtained in this work it can be concluded that, in case of the wild-type variant, the main role of catalytic base is played by Glu145, but it is possible that its role is taken over by Glu76 once the Glu145 is absent in the active site.

As indicated by key interactions found along the MD trajectory, two possible mechanisms for the adenylyl group transfer should be considered, i.e., ATP-assisted and Glu145-assisted mechanism, as shown in **Figure 5**. In both proposed mechanisms, the common process of nucleophilic attack of O4′ to Pα is expected to take place together with the Pα-O3α bond cleavage, and the only difference is related to the origin of acceptor of transferred H4′ from O4′ of the hydroxyl group. In the case of the ATP-assisted mechanism, H4′ is transferred to O2α, the oxygen of phosphate group not involved in interaction with Mg^{2+} cation, while in the Glu145-assisted mechanism the same proton is transferred to the OE2 atom of the deprotonated carboxyl group of glutamic acid 145. Thus, in order to explore possible reaction pathways, two potential energy surfaces were computed at AM1/MM level (for details see Computational Method section), describing activation of O4′ of KAN by controlling antisymmetric combination of distances between the oxygen atom from the hydroxyl group of KAN (O4′) and its proton (H4′) and the same proton H4′ and its possible acceptor i.e., O2α or OE2(Glu145), for ATP-assisted or Glu145-assisted mechanism, respectively, together with nucleophilic attack of O4′ to Pα of ATP. PESs are presented in **Figure 6**. The position of the TS and the reaction pathway traced by IRC calculation computed at the M06-2X/MM level are projected on the surfaces. As can be noticed, the position of the localized M06-2X/MM TS structures are quite close to quadratic regions defined on PES computed at AM1/MM level. Based on the shape of the obtained surfaces and after analysis of geometries of localized TS structures,(for which geometrical coordinates are given in **Table S5** and structures are presented in **Figure 7**), it is confirmed that both reactions proceed *via* concerted mechanisms with only one TS formed along the chemical path. Structures of both optimized TSs have an associative character, [see (More O'Ferrall, 1970; Jencks, 1985) plot presented in **Figure S3** and **Table S3**] in which nucleophilic attack slightly precedes the living group departure.

Additionally, the free energy barriers for both mechanisms were computed at M06-2X/MM level using the FEP method. The resulting activation free energies are 53.4 and 12.2 kcal·mol^{-1} for the ATP-assisted and Glu145-assisted mechanisms, respectively (the computed profiles are presented in **Figure S4**). These results clearly indicate that the second mechanism is the more favorable one. Moreover, a stable (of −1.6 kcal·mol^{-1}) product complex is formed only in this mechanism. In contrary, in ATP-assisted mechanism the obtained product complex is energetically much more unstable, with energy of 40.3 kcal·mol^{-1} with respect to the reactant complex. Thus, it can be concluded that inactivation of KAN in the active site of ANT(4′) takes place
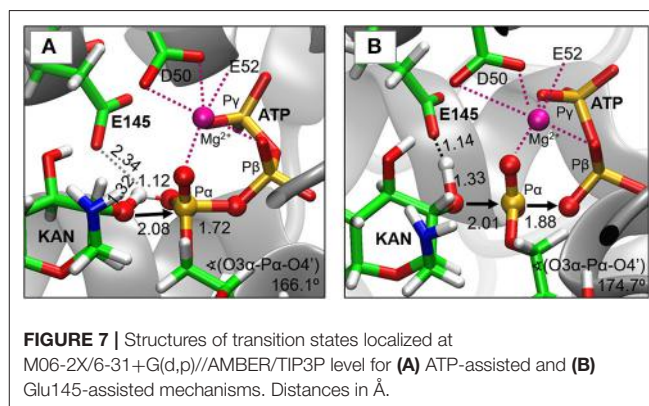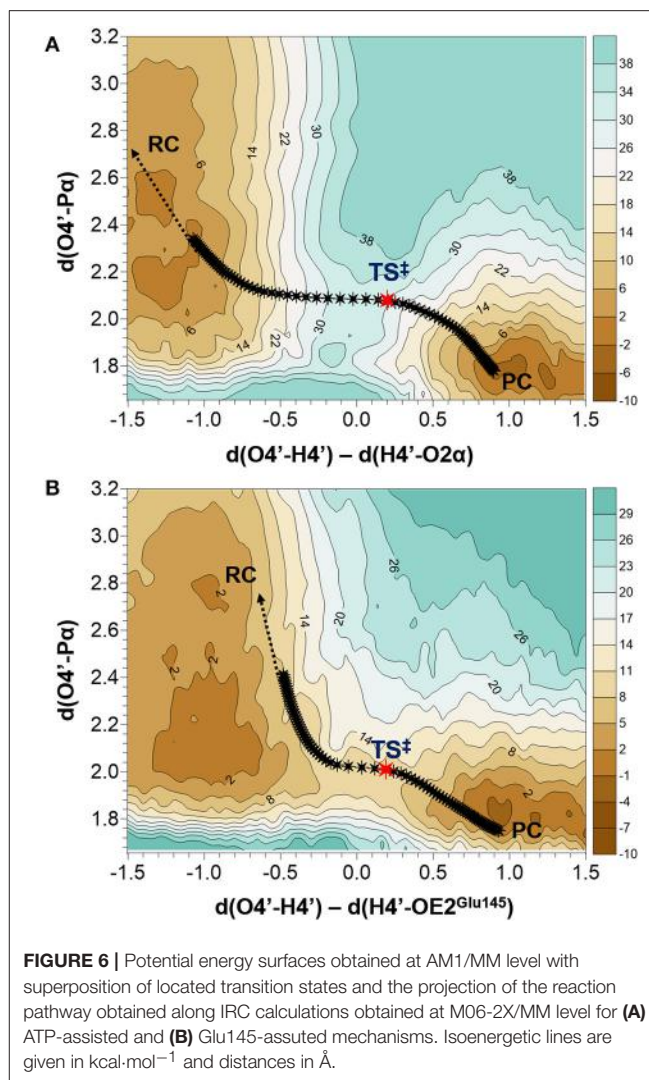
FIGURE 5 | Schematic representation of the two proposed mechanisms of KAN inactivation by ANT(4'). The blue and green arrows represent the two different paths of O4' activation by ATP (for ATP-assisted) and by Glu145 (for Glu145-assisted mechanism), respectively. Pink arrows indicate common process for both proposed mechanism of nucleophilic attack of O4' to Pα and Pα–mO3α bond cleavage.

with direct participation of the Glu145 residue in the chemical process.

It is worth mentioning that the computed free energy barrier of 12.2 kcal·mol^{-1} for the Glu145-assisted mechanism is much smaller than those deduced from experimentally measured rate constants of 19.7 kcal·mol^{-1} (for k_{cat} = 0.06 ± 0.01 measured in 35°C) (Revuelta et al., 2008), and 17.2 kcal·mol^{-1} (for k_{cat} = 1.3 ± 0.1 s^{-1} measured in 25°C) (Gerratana et al., 2001). This meaningful difference between experimental and theoretical values can be explained based on the fact that the rate-limiting step in this process is the release of the product rather than the chemical step, and thus experimental barriers can be treated, at most, as the upper limit value.

Kinetic Isotope Effects (KIEs)

In order to ensure our predictions, the reaction mechanisms described above can now be examined using such a sensitive tool as heavy-atom kinetic isotope effects (KIEs). KIEs are defined as the ratio of rate constants for the reactions involving the light and the heavy isotopically substituted reactants, and reflect changes in bond order between the ground and rate-limiting transition states. In this section, the interpretation of KIEs obtained theoretically for studied reaction is done, including comparison with experimentally measured intrinsic KIEs for bridge and non-bridge oxygen atoms with labeled slow substrate analog *m*-nitrobenzyl triphosphate (*m*NBTP) instead of ATP (Gerratana et al., 2001). The change of the substrate seems not to influence the reaction mechanism, which still has the same regiospecificity, however it slows down inactivation of KAN of 2 orders of magnitude (Gerratana et al., 2001).

FIGURE 6 | Potential energy surfaces obtained at AM1/MM level with superposition of located transition states and the projection of the reaction pathway obtained along IRC calculations obtained at M06-2X/MM level for (A) ATP-assisted and (B) Glu145-assuted mechanisms. Isoenergetic lines are given in kcal·mol^{-1} and distances in Å.

FIGURE 7 | Structures of transition states localized at M06-2X/6-31+G(d,p)//AMBER/TIP3P level for (A) ATP-assisted and (B) Glu145-assisted mechanisms. Distances in Å.

As shown in **Table 1**, in both ATP-assisted and Glu145-assisted mechanisms, normal primary KIEs (1°-KIE) of 0.9% and 1.2% for isotopically substituted oxygen in position 3α (oxygen of the leaving group), respectively, were computed. Both 1°-KIE values are in very good agreement with the experimentally

	Exp.*	Atoms	ATP-assisted	Glu145-assisted
1°-KIE ^{18}O-bridge (O3α)	1.014 ± 0.002	O3α	1.0092	1.0121
2°-KIE ^{18}O-non-bridge (O1α, O2α, O5′)	1.0024 ± 0.0002	O5′	1.0041	1.0026
		O1α	0.9980	1.0000
		O2α	0.9884	0.9981

*Provided experimental values correspond to KIEs measured at pH = 7.7, labeled slow substrate analog m-nitrobenzyl triphosphate (mNBTP) and are taken from (Gerratana et al., 2001).

measured value of 1.4% (Gerratana et al., 2001). In all cases, the existence of normal 1°-KIE reflects loss of bond order formed between phosphorus atom Pα and the bridge oxygen O3α in the TS with respect to reactant complex. This result can surely indicate that in the rate-limiting step, the P-O3α bond is cleaved. However, 1°-KIE values make it impossible to distinguish between ATP- and Glu145-assisted mechanisms and, moreover, to provide conclusive evidence if reaction proceed *via* concerted or step-wise mechanism. In such a case determination of 2°-KIE is crucial to dispel these doubts. Calculation of KIE for ^{18}O-substituted non-bridge oxygen such as O5′, O1α, and O2α has been done. As can be deduced from the results presented in **Table 1**, the total 2°-KIE for the ATP-assisted mechanism is slightly inverse, ca. −1% in contrast to the KIE computed for the Glu145-assisted mechanism with a very small normal KIE of 0.07%. The last one indicates clearly that loss of bond order for the non-bridge oxygens occurs simultaneously with P-O3α bond breaking and reveals a concerted mechanism. Existence of the small normal 2°-KIE is also observed experimentally (Gerratana et al., 2001), supporting our previous conclusion based on energetic analysis that the Glu145-assisted mechanism is the most favorable candidate and should correspond to the most realistic reaction pathway. Interestingly, the inverse 2°-KIE observed in the case of the ATP-assisted mechanism can be explained by the protonation of the non-bridge oxygen, O2α. Nevertheless, this result is in complete disagreement with obtained experimental results, which allows us once again to discard the ATP-assisted mechanism.

Electrostatic Effects

The most accepted hypothesis explaining the origin of the enzymatic catalysis is based on an assumption that stabilization of TS is achieved by means of better electrostatic interactions with the protein compared with its equivalent reaction in solution (Adamczyk et al., 2011; Moliner, 2011). Validity of this hypothesis was examined and strongly supported by our previous results obtained for systems such as protease HIV-1 (Krzemińska et al., 2016), glycine N-methyltransferase (GNMT) (Świderek et al., 2018) and *de novo* designed KEMP eliminase (Świderek et al., 2015b, 2017b). Thus, we decided to test this hypothesis also in the case of ANT(4′). For this purpose, one additional theoretical model was built, where KAN together with MgATP was solvated in a box of explicit water molecules. Additionally, in order to

obtain the same reaction mechanism as proposed for Glu145-assited pathway, the presence of a base was required, and this was fulfilled by adding the propionate molecule as imitator of Glu residue to the model. Afterwards, TS structure was localized at M06-2X/MM level (its geometrical coordinates are given in **Table S6**). Subsequently, identical procedure, as explained in the computational method section was applied in order to explore free energy surface. Free energy surface obtained for reaction in aqueous solution is shown in **Figure S5**. As it is well-known, non-catalyzed phosphoryl transfer reactions are extremely slow, and enzymes can provide rate enhancements of >10^{20}-fold (Lad et al., 2003).

Hence, our target herein was to understand the origin of enzyme catalytic power. In other words, the role of the enzyme in the KAN deactivation process is explored based on analysis of electrostatic potential generated by protein in a key position, i.e., on Pα atom, $V^{P\alpha}(r)$ the center of transferred group. Since the rate-limiting step of the reaction catalyzed by ANT(4′) is the release of the KAN-AMP complex from the active site, it can be assumed that this enzyme achieved the highest possible catalytic effect to enhance the speed of this reaction.

The obtained results of $V^{P\alpha}(r)$, generated by ANT(4′) and water solvent together with formal charge accumulated on the transferred group, are collected in **Table 2** and **Table S4**. A comparison of these magnitudes computed for TS structures for the reaction occurring through the base-assisted mechanism in enzyme and aqueous solution reveals the existence of a much larger positive $V^{P\alpha}(r)$ generated on the negatively charged transferred group in the active site of the enzyme (801.5 ± 17.7 kJ·mol^{-1}) than in the water (628.2 ± 49.9 kJ·mol^{-1}). According to the basic laws of physics, larger positive electrostatic potential should better stabilize the transition state in which accumulation of strong negative charge of −0.698 and −0.619 a.u., in water and enzyme, respectively, is observed in this atom. Therefore, as a result of this stabilization, the reduction of the free energy barrier in enzyme is expected. This expectation is fulfilled by theoretical results where a higher barrier (of 47.9 kcal·mol^{-1}) was computed for reactions taking place in water solvent than in enzyme (12.2 kcal·mol^{-1}). Thus, the reduction of the rate constant of ca. 10^{26}-fold can be considered, according to presented hypothesis, to arise from the stronger electrostatic potential created by ANT(4′). The possible contribution of each amino acid residue to the catalytic power was later explored, indicating their specific role in catalysis. It was found that the highest contribution to the overall positive value of the electrostatic potential comes from the positively charged residues located in the surroundings of the active site such as Lys-149 (~33.7%) from chain A, and Arg-42 (~13.5%) and Lys-74 (~10.7%) from chain B (see **Figure S7**). These residues can be considered to have the highest impact on enhancing the rate constant of the chemical reaction. On the other hand, nearby negatively charged Glu-141 and Glu-142 from chain A and Glu-76 from chain B produce negative electrostatic potential. Nevertheless, this unfavorable effect is too small to perturb the overall potential magnitude.

Finally, the analysis done for ATP-assisted and Glu145-assisted mechanisms occurring in the same active site shows, as expected, that values of $V^{P\alpha}(r)$ generated on the transferred group in TS and RC are very similar. However, the meaningful

TABLE 2 | Atomic charge (in a.u.) computed for structures localized at M06-2X/AMBER/TIP3P level of theory using the ChelpG method assigned to the transferred -PαO3R group, with electrostatic potential (in kJ·mol^{-1}) generated by ANT(4′) in Pα position in reactant complex and transition state, donor-acceptor distance (Å) and free energy barriers ($\Delta G^{\ddagger}$ in kcal·mol^{-1}).

Mechanism		Charge^a (a.u)	$V^{P\alpha}$(r) (kJ·mol^{-1})	DAD (Å)	$\Delta G^{\ddagger}$ (kcal·mol^{-1})
ATP-assisted	RC	−0.846^a	834.0 ± 17.5	4.45	53.4
	TS	−0.130^b	747.1 ± 20.4	3.77	
Glu145-assisted	RC	−0.582^a	824.6 ± 24.4	4.26	12.2
	TS	−0.619^a	801.5 ± 17.7	3.88	
Water	RC	−0.734^a	586.3 ± 38.8	4.30	47.9
	TS	−0.698^a	628.2 ± 49.9	3.81	

^{a} presented values of charges correspond to sum of formal charges computed for Pα, O1α, O2α, O5′ atoms ^{b} in case of TS for ATP-assisted the charge of transferred H4′ to O2α was added.

difference is found in the charge distribution on substrates. (see **Table 2**) For the most favorable mechanism (with lower free energy barrier), a slight increase of negative charge on the phosphate group in TS in respect to RC (−0.619 and −0.582 a.u., respectively) is observed. However, the largest change in charge distribution from RC to TS was found in the ATP-assisted mechanism. In this case the larger negative charge of −0.846 a.u. is observed in RC. Since the H4′ proton is transferred from the hydroxyl group of KAN to the adenylyl group in this mechanism, its presence dramatically decreases the negative charge accumulated on the adenylyl group up to −0.130 a.u in TS. Consequently, $V^{P\alpha}$(r) generated by ANT(4′) in the case of the ATP-assisted mechanism would rather over-stabilize the ground state in comparison to the TS structure, and this could explain the high free energy barrier obtained for this mechanism.

Compression Effects

The complementary hypothesis for enzymes catalyzing S_N2 reactions is the so-called "compression hypothesis," originally proposed by Showen (Hegazi et al., 1979) for enzymatic methyl transfer. Extrapolated to hydrogen transfer reactions (Roston et al., 2013), this hypothesis explains the origin of the enzymatic catalysis, suggesting that specific protein fluctuations might reduce the donor-acceptor distance (DAD), and as a consequence decrease the reaction barrier by increasing the number of reactive trajectories (Kohen, 2015). Thus, since mechanistic results of the transfer of the adenylyl group presented in this work have revealed its S_N2 character, we decided to use this reaction as an example that could shed some new light and contribute to the ongoing debate about the validity of the proposed hypothesis. For this purpose, the evolution of DAD vs. distances describing the −PO$_3$R group transfer was analyzed for aqueous and enzymatic reaction, as presented in **Figure S6**. As it was observed, the decrease of the DAD from RC to TS is 0.68 and 0.38 Å for the ATP and Glu145-assisted mechanisms, respectively. DAD values are provided in **Table 2**. Interestingly, the minimum value of DAD (DAD$_{min}$) is not achieved in TS, and in both cases it appears much beyond the highest energetic point along the reaction pathway, to be then elongated to allow for product formation. Decrease of a DAD is also observed in reaction in water, where ΔDAD$^{(RC-TS)}$ was found to be 0.49 Å. However, in contrast to enzymatic reactions, in this case the DAD value
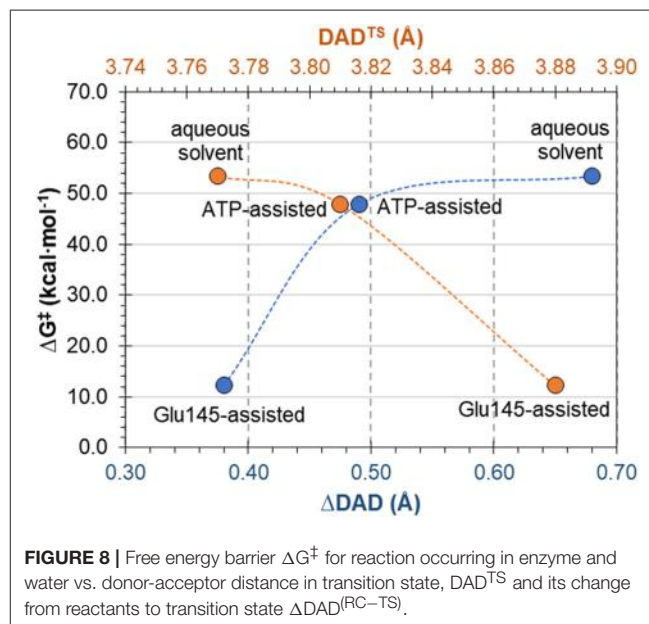
FIGURE 8 | Free energy barrier $\Delta G^{\ddagger}$ for reaction occurring in enzyme and water vs. donor-acceptor distance in transition state, DAD^{TS} and its change from reactants to transition state ΔDAD$^{(RC-TS)}$.

decreases constantly until a pentavalent intermediate is formed. Analysis of the influence of DAD distance in TS to the values of the free energy barriers, as plotted in **Figure 8**, reveals that short DAD^{TS} does not guarantee reduction in the reaction barrier. Nevertheless, an interesting relation was found indicating that the larger change in DAD from RC to TS, the slower the reaction is. This would suggest that large changes in DAD require more energy to be involved in the chemical process. Finally, it can be concluded that reduction of the reaction barrier is achieved only when the short DAD in the Michaelis complex is reached, i.e., 4.26 Å for the Glu145-assisted mechanism, vs. 4.45 Å for the ATP-assisted mechanism or 4.30 Å for reactions taking place in water.

CONCLUSIONS

In this work the molecular mechanism of the transfer of the adenylyl group from ATP to KAN, and formation of KAN-AMP-Mg ternary complex as a result of the reaction catalyzed by the ANT(4′) enzyme was investigated. Two proposed mechanisms

were studied, the ATP-assisted and Glu145-assisted mechanisms, which were shown to proceed in a concerted manner, in which the O4′ atom of KAN is activated by proton abstraction by either ATP or Glu-145 residue, respectively, and at the same time formation of the O4′-Pα bond and breaking of the Pα-O3α bond occur. Based on the obtained free energy barriers, it was deduced that the most favorable mechanism for KAN deactivation by ANT(4′) proceeds *via* the Glu145-assisted mechanism, with the barrier of 12.2 kcal·mol^{-1} much smaller than that obtained for the ATP/Mg-assisted process (53.4 kcal·mol^{-1}). This value is slightly underestimated when compared to the barriers obtained from experimentally measured rate constants, i.e., between 17.2 and 19.7 kcal·mol^{-1}. The lower barrier obtained in theoretical studies can be explained by the fact that the rate-limiting step in this reaction is, in fact, not the chemical conversion, but the release of the KAN-AMP complex. Then, the theoretical prediction of the most favorable mechanism was confirmed by determination of KIEs. This sensitive tool allowed us to discriminate between both mechanisms based on 2°-KIEs computed for isotopic substitution of non-bridge oxygen atoms. In the case of the ATP-assisted mechanism, the inverse (<1) value of 2°-KIE was obtained, being completely opposite to the experimentally measured data. However, very good agreement between the experimentally and theoretically determined primary and secondary KIEs was found in the case of the Glu145-assisted mechanism, confirming its existence.

Additionally, calculation of the electrostatic potential generated by enzymes on atoms involved in chemical reactions reveals that electrostatic effects can be correlated with the height of computed free energy barriers. Thus, the high barrier of 47.9 kcal·mol^{-1} observed for the reaction in the water solvent can be explained by the low positive potential (of c.a. 628.2 ± 49.9 kJ·mol^{-1}) generated on the negatively charged adenyl-group. In contrast, the higher electrostatic potential of ca. 800 kJ·mol^{-1} generated in the active site of the enzyme stabilizes a much better structure of TS and as a result reduces the reaction barrier ca. 4-fold.

Moreover, in the case of reactions with more than one possible mechanism as the one under study in the present paper, the obtained results allow us to conclude the important role of enzymes to direct chemical transformation through the reaction mechanism in which distribution of charges on key atoms in the structures of TS is the most compatible one to its electrostatic potential.

Finally, the "compression hypothesis" for reaction of S_N2 character was examined, indicating that a short donor-acceptor distance in TS does not guarantee reduction in the reaction barrier, and that this reduction is achieved only when the short DAD in the Michaelis complex is found.

AUTHOR CONTRIBUTIONS

SM implemented FEP method in fDynamo library and computed free energy surfaces. AB served as an expert in ANT(4′) enzyme, corrected the manuscript. KŚ designed the studies, ran calculations, analyzed results and finally wrote the manuscript.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2018.00660/full#supplementary-material

Supplementary Table 1 | Empirically predicted pKa values for titratable residues. Atom types, charges and parameters obtained for Kanamycin A used in MM simulations. Results of MD simulations: RMSD, RMSF, evolution of energy and thermostat control. Schematic representation of active site. O'Ferrall-Jencks plot for ATP-assisted and Glu145-assisted mechanism. Free energy profiles for ATP-assisted and Glu145-assisted mechanism catalyzed by ANT(4′) computed at M06-2X/6-31+G(d,p)//AMBER/TIP3P level. Free energy profile for base-assisted mechanism in aqueous solution computed at M06-2X/6-31+G(d,p)//AMBER/TIP3P level. Key distances and angles for reactant complex, transition state and product complex localized along ATP-assisted and Glu145-assisted mechanism at M06-2X/6-31+G(d,p)//AMBER/TIP3P level. Atomic charge computed for structures localized at M06-2X/AMBER/TIP3P level. Geometrical coordinates of QM atoms for transition state structures localized at M06-2X/6-31+G(d,p)//AMBER/TIP3P level for reaction catalyzed by ANT(4′). Evolution of donor-acceptor distance (DAD) along reaction path in catalyzed and uncatalyzed reaction. Contribution of key amino acid residues to overall value of electrostatic potential. Geometrical coordinates of QM atoms for transition state structure localized at M06-2X/6-31+G(d,p)//AMBER/TIP3P for reaction is aqueous solution.

REFERENCES

Adamczyk, A. J., Cao, J., Kamerlin, S. C. L., and Warshel, A. (2011). Catalysis by dihydrofolate reductase and other enzymes arises from electrostatic preorganization, not conformational motions. *Proc. Natl. Acad. Sci. US.A.* 108, 14115–14120. doi: 10.1073/pnas.1111252108

Becker, B., and Matthew, A. (2013). Cooper aminoglycoside antibiotics in the 21st century. *ACS Chem. Biol.* 8, 105–115. doi: 10.1021/cb3005116

Berendonk, T. U., Manaia, C. M., Merlin, C., Fatta-Kassinos, D., Cytryn, E., Walsh, F., et al. (2015). Tackling antibiotic resistance: the environmental framework. *Nat. Rev. Microbiol.* 13, 310–317. doi: 10.1038/nrmicro3439

Bordes, I., Castillo, R., and Moliner, V. (2017). Theoretical study of the phosphoryl transfer reaction from ATP to Dha catalyzed by DhaK from *Escherichia coli*. *J. Phys. Chem. B* 12, 8878–8892. doi: 10.1021/acs.jpcb.7b04862

Carter, A. P., Clemons, W. M., Brodersen, D. E., Morgan-Warren, R. J., Wimberly, B. T., and Ramakrishnan, V. (2000). Functional insights from the structure of

the 30S ribosomal subunit and its interactions with antibiotics. *Nature* 407, 340–348. doi: 10.1038/35030019

Chen-Goodspeed, M., Vanhooke, J. L., Holden, H. M., and Raushel, F. M. (1999). Kinetic mechanism of kanamycin nucleotidyltransferase from *Staphylococcus aureus*. *Bioorg. Chem.* 27, 395–408.

Cherepanov, A. V., Doroshenko, E. V., Matysik, J., de Vries, S., and de Groot, H. J. M. (2008). The associative nature of adenylyl transfer catalyzed by T4 DNA ligase. *Proc. Natl. Acad. Sci. U.S.A.* 105, 8563–8568. doi: 10.1073/pnas.0709140105

Cleland, W. W., and Hengge, A. C. (2006). Enzymatic mechanisms of phosphate and sulfate transfer. *Chem. Rev.* 106, 3252–3278. doi: 10.1021/cr050287o

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

Doi, Y., and Arakawa, Y. (2007). 16S ribosomal RNA methylation: emerging resistance mechanism against aminoglycosides. *Clin. Infect. Dis.* 45, 88–94. doi: 10.1086/518605

Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W., et al. Kollman, P. (2003). A point charge force field for molecular mechanics simulations of proteins based on condensed phase quantum mechanical calculations. *J. Comput. Chem.* 24, 1999–2012. doi: 10.1002/jcc.10349

Field, M. J., Albe, M., Bret, C., Proust-De Martin, F., and Thomas, A. (2000). The dynamo library for molecular simulations using hybrid quantum mechanical and molecular mechanical potentials. *J. Comp. Chem.* 21, 1088–1100. doi: 10. 1002/1096-987X(200009)21:123.3.CO;2-#

Gates, C. A., and Northrop, D. B. (1988). Substrate specificities and structure-activity relationships for the nucleotidylation of antibiotics catalyzed by aminoglycoside nucleotidyltransferase 2"-I. *Biochemistry* 27, 3820–3825. doi: 10.1021/bi00410a045

Gerratana, B., Cleland, W. W., and Reinhardt, L. A. (2001). Regiospecificity assignment for the reaction of kanamycin nucleotidyltransferase from *Staphylococcus aureus*. *Biochemistry* 40, 2964–2971. doi: 10.1021/bi00 25565

Gerratana, B., Frey, P. A., and Cleland, W. W. (2001). Characterization of the transition-state structure of the reaction of kanamycin nucleotidyltransferase by heavy-atom kinetic isotope effects. *Biochemistry* 40, 2972–2977. doi: 10.1021/bi002557x

Grest, G. S., and Kremer, K. (1986). Molecular-dynamics simulation for polymers in the presence of a heat bath. *Phys. Rev. A Gen. Phys.* 33, 3628–3631. doi: 10.1103/PhysRevA.33.3628

Hegazi, M. F., Borchardt, R. T., and Schowen, R. L. (1979). α-deuterium and carbon-13 isotope effects for methyl transfer catalyzed by catechol O-methyltransferase.SN2-like transition state. *J. Am. Chem. Soc.* 101, 4359–4365. doi: 10.1021/ja00509a052

Heldenbrand, H., Janowski, P. A., Giambaşu, G., Giese, T. J., Wedekind, J. E., and York, D. M. (2014). Evidence for the role of active site residues in the hairpin ribozyme from molecular simulations along the reaction path. *J. Am. Chem. Soc.* 136, 7789–7792. doi: 10.1021/ja500180q

Jencks, W. P. (1985). A primer for the Bema Hapothle. An empirical approach to the characterization of changing transition-state structures. *Chem. Rev.* 85, 511–527. doi: 10.1021/cr00070a001

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935. doi: 10.1063/1.445869

Kamerlin, S. C. L. (2011). Theoretical comparison of p-nitrophenyl phosphate and sulfate hydrolysis in aqueous solution: implications for enzyme-catalyzed sulfuryl transfer. *J. Org. Chem.* 76, 9228–9238. doi: 10.1021/jo201104v

Kamerlin, S. C. L., Florian, J., and Warshel, A. (2008). Associative versus dissociative mechanisms of phosphate monoester hydrolysis: on the interpretation of activation entropies. *ChemPhysChem* 9, 1767–1773. doi: 10.1002/cphc.200800356

Kohen, A. (2015). Role of dynamics in enzyme catalysis: substantial versus semantic controversies. *Acc. Chem. Res.* 48, 466–473. doi: 10.1021/ar50 0322s

Krzemińska, A., Moliner, V., and Świderek K. (2016). Dynamic and electrostatic effects on the reaction catalyzed by HIV-1 protease. *J. Am. Chem. Soc.* 138, 16283–16298. doi: 10.1021/jacs.6b06856

Krzemińska, A., Paneth, P., Moliner, V., and Świderek, K. (2015). Binding isotope effects as a tool for distinguishing hydrophobic and hydrophilic binding sites of HIV-1 RT. *J. Phys. Chem. B* 119, 917–927. doi: 10.1021/jp506119h

Kumar, A., and Schweizer, H. P. (2005). Bacterial resistance to antibiotics: active efflux and reduced uptake. *Adv. Drug Deliv. Rev.* 57, 1486–1513. doi: 10.1016/j.addr.2005.04.004

Lad, C., Williams, N. H., and Wolfenden, R. (2003). The rate of hydrolysis of phosphomonoester dianions and the exceptional catalytic proficiencies of protein and inositol phosphatases. *Proc. Natl. Acad. Sci. U.S.A.* 100, 5607–5610. doi: 10.1073/pnas.0631607100

Leban, N., Kaplana, E., Chaloin, L., Godreuil, S., and Lionne, C. (2017). Kinetic characterization and molecular docking of novel allosteric inhibitors of aminoglycoside phosphotransferases. *Biochim. Biophys. Acta Gen. Subj.* 1861, 3464–3473. doi: 10.1016/j.bbagen.2016.09.012

Lombardini, J. B., and Cheng-Chu, M. (1980). Properties of gentamicin adenylyl transferase obtained from R-facor-resistant Escherichia coli. *Int. J. Biochem.* 12, 427–431. doi: 10.1016/0020-711X(80)90123-8

Marcos, E., Crehuet, R., and Anglada, J. M. (2008). Inductive and external electric field effects in pentacoordinated phosphorus compounds. *J. Chem. Theory Comput.* 4, 49–63. doi: 10.1021/ct700220z

Martí, S., Moliner, V., and Tuñón, I. (2005). Improving he QM/MM description of chemical processes: a dual level strategy to explore the potential energy surface in very large systems. *J. Chem. Theory Comput.* 1, 1008–1016. doi: 10.1021/ct0501396

Matesanz, R., Diaz, J. F., Corzana, F., Santana, A. G., and Bastida, A., Asensio, J.L. (2012). Multiple keys for a single lock: the unusual structural plasticity of the nucleotidyltransferase (4')/kanamycin complex. *Chem. Eur. J.* 18, 2875–2889. doi: 10.1002/chem.201101888

Meagher, K. L., Redman, L. T., and Carlson, H. A. (2003). Development of polyphosphate parameters for use with the AMBER force field. *J. Comput. Chem.* 24, 1016–1025. doi: 10.1002/jcc.10262

Min, D., Xue, S., Li, H., and Yang, W. (2007). 'In-line attack' conformational effect plays a modest role in an enzyme-catalyzed RNA cleavage: a free energy simulation study. *Nucleic Acids Res.* 35, 4001–4006. doi: 10.1093/nar/gkm394

Misumi, M., and Tanaka, N. (1980). Mechanism of inhibition of translocation by kanamycin and viomycin: a comparative study with fusidic acid. *Biochem. Biophys. Res. Commun.* 92, 647–654. doi: 10.1016/0006-291X(80)90382-4

Mlýnský, V., Banáš, P.,Šponer, J., van der Kamp, M. W., Mulholland, A. J, and Otyepka, M. (2014). Comparison of ab Initio, DFT, and semiempirical QM/MM approaches for description of catalytic mechanism of hairpin ribozyme. *J. Chem. Theory Comput.* 10, 1608–1622. doi: 10.1021/ct401015e

Moliner, V. (2011). Eppur si muove (yet it moves). *Proc. Natl. Acad. Sci. U.S.A.* 108, 15013–15014. doi: 10.1073/pnas.1112014108

More O'Ferrall, R. A. (1970). Relationships between E2 and E1cB mechanisms of β-elimination. *J. Chem. Soc. B* 274–277.

Nature (2013). The antibiotic alarm. *Nature* 495:141. doi: 10.1038/495141a

Oertell, K., Chamberlain, B. T., Wu, Y., Ferri, E., Kashemirov, B. A., Beard, W. A., et al. (2014). The transition-state in DNA polymerase β catalysis: ratelimiting chemistry altered by base-pair configuration. *Biochemistry* 53, 1842–1848. doi: 10.1021/bi500101z

Oertell, K., Wu, Y., Zakharova, V. M., Kashemirov, B. A., Shock, D. D., Beard, W. A., et al. (2012). Effect of β, γ-CHF- and β, γ-CHCl-dGTP halogen atom stereochemistry on the transition state of DNA polymerase β. *Biochemistry* 51, 8491–8501. doi: 10.1021/bi3010335

Olsson, M. H. M., Sondergaard, C. R., Rostkowski, M., and Jensen, J. H. (2011). PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. *J. Chem. Theory Comput.* 7, 525–537. doi: 10.1021/ct 100578z

Pedersen, L. C., Benning, M. M., and Holden, H. M. (1995). Structural investigation of the antibiotic and ATP-binding sites in kanamycin nucleotidyltransferase. *Biochemistry* 34, 13305–13311.

Pelt, J. E., Iyengar, R., and Frey, P. A. (1986). Gentamicin nucleotidyltransferase. Stereochemical inversion at phosphorus in enzymatic 2'-deoxyadenylyl transfer to tobramycin. *J. Biol. Chem.* 261, 15995–15999.

Pestka, S. (1974). The use of inhibitors in studies on protein synthesis. *Methods Enzymol.* 30, 261–282. doi: 10.1016/0076-6879(74)30030-4

Pfister, P., Hobbie, S., Vicens, Q., Böttger, E. C., and Westhof, E. (2003). The molecular basis for A-site mutations conferring aminoglycoside resistance:

relationship between ribosomal susceptibility and X-ray crystal structures. *Chembiochem* 4, 1078–1088. doi: 10.1002/cbic.200300657

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., et al. (2005). Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26, 1781–1802. doi: 10.1002/jcc.20289

Ramirez, M. S., and Tolmasky, M. E. (2010). Aminoglycoside modifying enzymes. *Drug Resist. Updat.* 13, 151–171. doi: 10.1016/j.drup.2010.08.003

Revuelta, J., Corzana, F., Bastida, A., and Asensio, J. L. (2010). The unusual nucleotide recognition properties of the resistance enzyme ANT(4′): inorganic tri/polyphosphate as a substrate for aminoglycoside inactivation. *Chem. Eur. J.* 16, 8635–8640. doi: 10.1002/chem.201000641

Revuelta, J., Vacas, T., Torrado, M., Corzana, F., Gonzalez, C., Jiménez-Barbero, J., et al. (2008). NMR-based analysis of aminoglycoside recognition by the resistance enzyme ANT(4′): the pattern of OH/NH3+substitution determines the preferred antibiotic binding mode and is critical for drug inactivation. *J. Am. Chem. Soc.* 130, 5086–5103. doi: 10.1021/ja076835s

Ridder, L., and Mulholland, A. J. (2003). Modeling biotransformation reactions by combined quantum mechanical/molecular mechanical approaches: from structure to activity. *Curr. Top. Med. Chem.* 3, 1241–1256. doi: 10.2174/1568026033452005

Roston, D., Islam, Z., and Kohen, A. (2013). Isotope effects as probes for enzyme catalyzed hydrogen-transfer reactions. *Molecules* 18, 5543–5567. doi: 10.3390/molecules18055543

Schafmeister, C. E. A. F., Ross, W. S., and Romanovski, V. (1995). *LEAP.* University of California, San Francisco, CA.

Shaw, K. J., Rather, P. N., Hare, R. S., and Miller, G. H. (1993). Molecular genetics of aminoglycoside resistance genes and familial relationships of the aminoglycoside-modifying enzymes. *Microbiol. Rev.* 57, 138–163.

Sondergaard, C. R., Olsson, M. H. M., Rostkowski, M., and Jensen, J. H. (2011). Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pKa values. *J. Chem. Theory Comput.* 7, 2284–2295. doi: 10.1021/ct200133y

Sucato, C. A., Upton, T. G., Kashemirov, B. A., Osuna, J., Oertell, K., Beard, W. A., et al. (2007). Modifying the β, γ. Leaving group bridging oxygen alters nucleotide incorporation efficiency, fidelity, and the catalytic mechanism of DNA polymerase β. *Biochemistry* 46, 461–471. doi: 10.1021/bi061517b

Sucato, C. A., Upton, T. G., Kashemirov, B. A., Osuna, J., Oertell, K., Beard, W. A., et al. (2008). DNA polymerase β fidelity: halomethylene-modified leaving groups in pre-steady-state kinetic analysis reveal differences at the chemical transition state. *Biochemistry* 47, 870–879. doi: 10.1021/bi7014162

Świderek, K., Arafet, K., Kohen, A., and Moliner, V. (2017a). Benchmarking Quantum Mechanics/Molecular Mechanics (QM/MM) methods on the thymidylate synthase-catalyzed hydride transfer. *J. Chem. Theory Comput.* 13, 1375–1388. doi: 10.1021/acs.jctc.6b01032

Świderek, K., Martí, S., and Moliner, V. (2014). Theoretical study of primary reaction of Pseudozyma antarctica lipase B as the starting point to understand its promiscuity. *ACS Catal.* 4, 426–434. doi: 10.1021/cs401047k

Świderek, K., Martí, S., Tuñón, I., Moliner, V., and Bertran, J. (2015a). Peptide bond formation mechanism catalyzed by ribosome. *J. Am. Chem. Soc.* 137, 12024–12034. doi: 10.1021/jacs.5b05916

Świderek, K., Tuñón, I., Martí, S., Moliner, V., and Bertran, J. (2013). Role of solvent on non-enzymatic peptide bond formation mechanism and kinetic isotope effects. *J. Am. Chem. Soc.* 135, 8708–8719. doi: 10.1021/ja403038t

Świderek, K., Tuñón, I., Moliner, V., and Bertran, J. (2015b). Protein flexibility and preorganization in the design of enzymes. The Kemp elimination catalysed by HG3.17. *ACS Catal.* 5, 2587–2595. doi: 10.1021/cs501090w

Świderek, K., Tuñón, I., Moliner, V., and Bertran, J. (2017b). Revealing the origin of the efficiency of the *de novo* designed Kemp eliminase HG-3.17 by comparison with the former developed HG-3. *Chem. Eur. J.* 23, 7582–7589. doi: 10.1002/chem.201700807

Świderek, K., Tuñón, I., Williams, I. H., and Moliner, V. (2018). Insights on the origin of catalysis on glycine N-methyltransferase from computational modelling. *J. Am. Chem. Soc.* 140, 4327–4334. doi: 10.1021/jacs.7b13655

Torrie, G. M., and Valleau, J. P. (1977). Nonphysical sampling distributions in monte carlo free-energy estimation: umbrella sampling. *J. Comput. Phys.* 23, 187–199. doi: 10.1016/0021-9991(77)90121-8

Turner, A., Moliner, V., and Williams, I. H. (1999). Transition-state structural refinement with GRACE and CHARMM: flexible QM/MM modelling for lactate dehydrogenase. *Phys. Chem. Chem. Phys.* 1, 1323–1331.

Vakulenko, S. B., and Mobashery, S. (2003). Versatility of aminoglycosides and prospects for their future. *Clin. Microbiol. Rev.* 16, 430–450. doi: 10.1128/CMR.16.3.430-450.2003

Ventola, C. L. (2015). The antibiotic resistance crisis part 1: causes and threats. *P. T.* 40, 277–283.

Viciano, I., Castillo, R., and Martí, S. (2015). QM/MM modeling of the hydroxylation of the androstenedione substrate catalyzed by cytochrome P450 aromatase (CYP19A1). *J. Comput. Chem.* 36, 1736–1747. doi: 10.1002/jcc.23967

Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004). Development and testing of a general AMBER force field. *J. Comput. Chem.* 25, 1157–1174. doi: 10.1002/jcc.20035

Wymore, T., Field, M. J., Langan, P., Smith, J. C., and Parks, J. M. (2014). Hydrolysis of DFP and the nerve agent (S)-sarin by DFPase proceeds along two different reaction pathways: implications for engineering bioscavengers. *J. Phys. Chem. B* 118, 4479–4489. doi: 10.1021/jp410422c

Zhang, Z., Eloge, J., and Florian, J. (2014). Quantum mechanical analysis of nonenzymatic nucleotidyl transfer reactions: kinetic and thermodynamic effects of β-γ bridging groups of dNTP substrates. *Biochemistry* 53, 4180–4191. doi: 10.1021/bi5003713

Zhao, Y., and Truhlar, D. G. (2008a). Density functionals with broad applicability in chemistry. *Acc. Chem. Res.* 41, 157–167. doi: 10.1021/ar700111a

Zhao, Y., and Truhlar, D. G. (2008b). The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* 120:215–241. doi: 10.1007/s00214-007-0310-x

Comparing Hydrolysis and Transglycosylation Reactions Catalyzed by *Thermus thermophilus* β-Glycosidase. A Combined MD and QM/MM Study

Sonia Romero-Téllez [1,2], José M. Lluch [1,2], Àngels González-Lafont [1,2]* and Laura Masgrau [1,2]*

[1] Departament de Química, Universitat Autònoma de Barcelona, Barcelona, Spain, [2] Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Barcelona, Spain

The synthesis of oligosaccharides and other carbohydrate derivatives is of relevance for the advancement of glycosciences both at the fundamental and applied level. For many years, glycosyl hydrolases (GHs) have been explored to catalyze the synthesis of glycosidic bonds. In particular, retaining GHs can catalyze a transglycosylation (T) reaction that competes with hydrolysis (H). This has been done either employing controlled conditions in wild type GHs or by engineering new mutants. The goal, which is to increase the T/H ratio, has been achieved with moderate success in several cases despite the fact that the molecular basis for T/H modulation are unclear. Here we have used QM(DFT)/MM calculations to compare the glycosylation, hydrolysis and transglycosylation steps catalyzed by wild type *Thermus thermophilus* β-glycosidase (family GH1), a retaining glycosyl hydrolase for which a transglycosylation yield of 36% has been determined experimentally. The three transition states have a strong oxocarbenium character and ring conformations between $^4H_3$ and ^4E. The atomic charges at the transition states for hydrolysis and transglycosylation are very similar, except for the more negative charge of the oxygen atom of water when compared to that of the acceptor Glc. The glycosylation transition state has a stronger S_N2 character than the deglycosylation ones and the proton transfer is less advanced. At the QM(PBE0/TZVP)/MM level, the TS for transglycosylation has shorter $O4_{GLC}-C1_{FUC}$ (forming bond) distance and longer $OE2_{GLU338}-C1_{FUC}$ (breaking) distance than the hydrolysis one, although the H_{ACC} proton is closer to the Glu164 base in the hydrolysis TS. The QM(SCC-DFTB)/MM free energy maxima show the inverted situation, although the hydrolysis TS presents significant structural fluctuations. The 3-OH_{GLC} group of the acceptor Glc (transglycosylation) and WAT432 (neighbor water in hydrolysis) are identified to stabilize the oxocarbenium transition states through interaction with $O5_{FUC}$ and $O4_{FUC}$. The analysis of interaction suggests that perturbing the Glu392-Fuc interaction could increase the T/H ratio, either by direct mutation of this residue or indirectly as reported experimentally in the Asn390I

and Phe401S cases. The molecular understanding of similarities and differences between hydrolysis and transglycosylation steps may be of help in the design of new biocatalysts for glycan synthesis.

INTRODUCTION

Carbohydrates and their derivatives (glycoconjugates and glycosides) play important biological functions, including energy storage, structural roles and also the encoding of a molecular and cell recognition language that drives the specificity of such processes. Accordingly, glycans are vital for normal life development and are also key factors in the progression of many diseases (from pathogen infection to cancer) (Seeberger and Cummings, 2017). In order to develop fundamental and applied research in the glycosciences field, access to pure glycans and in sufficient amounts is needed. Glycans biosynthesis involves the action of a repertoire of enzymes, amongst them glycosyltransferases (GTs, which catalyze the synthesis of a new glycosidic bond) and glycoside hydrolases (GHs). These enzymes catalyze the corresponding reaction with two possible stereochemical outcomes, that is, retention or inversion of the configuration at the anomeric carbon of the transferred/hydrolyzed sugar.
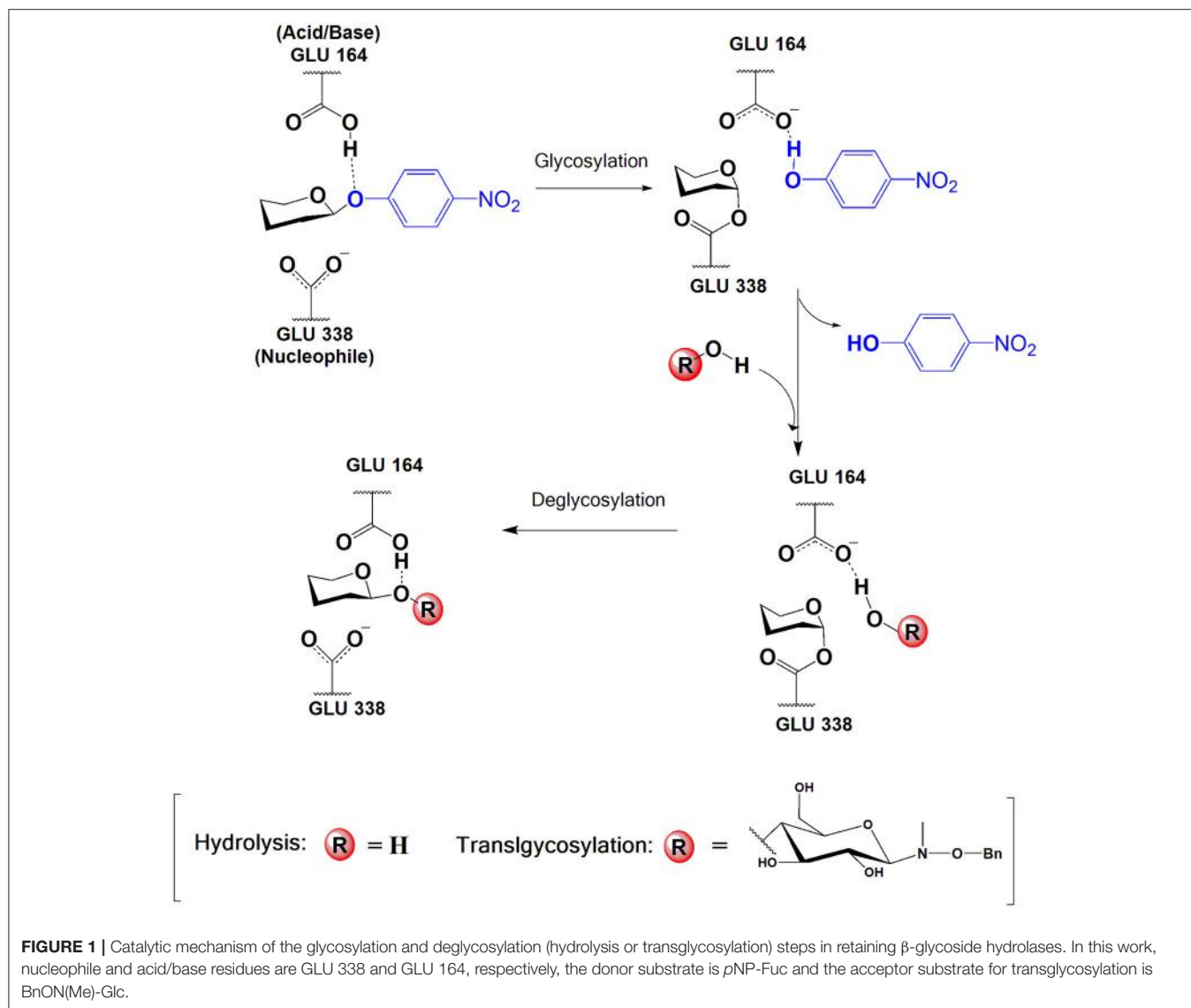
Besides being the target of drug design investigations, carbohydrate-active enzymes are of interest in glycans processing and synthesis, an important area of fundamental research and also for the preparation of commercially-valuable products. Glycoside hydrolytic enzymes (e.g., amylases, cellulases) are used in the e.g., food, textile, detergents, cosmetics, pulp, and paper industries, representing around one third of the global industrial enzyme market (Plou et al., 2007; Sajith et al., 2016). The enzymatic synthesis of glycosidic bonds by carbohydrate-active enzymes has also been a subject of study for over 60 years and different strategies have been developed that circumvent or complement chemical approaches, which usually require multiple protection, and deprotection steps to obtain the desired oligosaccharide (Kiessling and Splain, 2010; Bissaro et al., 2015b; Danby and Withers, 2016). In Nature, GTs are the main catalysts for the synthesis of glycosidic bonds. However, their broad application is hindered by the difficulties in their expression and purification and due to the economic costs of obtaining the nucleotide-phosphate sugars they use as donor substrates, despite progresses have been made in that direction (Field, 2011). As an alternative, GHs have been explored to catalyze glycosidic bond formation (Planas and Faijes, 2002; Cobucci-Ponzano et al., 2011; Bissaro et al., 2015b). GHs are more abundant than GTs, are much easier to obtain, cover a wide range of substrate specificities and their substrates are cheap.

In particular, retaining GHs can operate in a synthetic mode if the equilibrium is displaced toward glycoside bond formation (thermodynamically controlled processes) or by using activated glycosyl donors (kinetically controlled transglycosylation) (Planas and Faijes, 2002). Retaining GHs follow a double displacement mechanism in two subsequent steps with formation of a covalent glycosyl-enzyme intermediate. Two catalytic carboxyl groups (separated by ∼5Å) act as general acid/base and nucleophile in the reaction (as exemplified in **Figure 1** for the process studied in this work). In a first step (referred as glycosylation, G), the nucleophile attacks the anomeric carbon while the carboxylic acid protonates the glycosidic oxygen of the leaving group, and the glycosyl-enzyme intermediate is formed. In the final step (deglycosylation, D), the residue previously acting as an acid now deprotonates the incoming acceptor substrate that attacks the anomeric carbon forming the final product with net retention of the configuration. If the acceptor is a water molecule, hydrolysis (H) occurs. However, with the presence of a different suitable sugar acceptor, many retaining GHs are capable of catalyzing transglycosylation (T).

Different approaches and experimental conditions have been used to enhance the transglycosylation vs. hydrolysis (T/H) ratio, e.g., high acceptor concentrations, the use of highly reactive glycosyl donors (activated donors) like aryl glycosides or glycosyl fluorides, removal of the transglycosylation product or enzyme immobilization. With all, yields and regiospecificities are still low (Planas and Faijes, 2002). Engineering of GHs has also been investigated. A major breakthrough in the field was accomplished with the development of glycosynthases, in which hydrolysis is abolished by replacing the catalytic nucleophile by a non-nucleophilic residue and the use of activated substrates (Mackenzie et al., 1998; Malet and Planas, 1998; Moracci et al., 1998; Faijes and Planas, 2007; Cobucci-Ponzano and Moracci, 2012; Danby and Withers, 2016). Other studies have tried to understand the molecular determinants for having a transglycosylation vs. an hydrolytic enzyme activity by applying mutational studies (Bissaro et al., 2015b). Notably, Dion and coworkers (Feng et al., 2005) successfully applied directed evolution techniques to the β-glycosidase of *Thermus thermophilus* (Ttβ-gly), belonging to the CAZy family GH1 (Henrissat, 1991; Cantarel et al., 2009; Hart and Copeland, 2010), to increase its ability to synthesize oligosaccharides by transglycosylation.

Their findings were later "semi-rationalized" by realizing that mutations leading to an increase in the T/H ratio in Ttβ-gly were located at highly conserved positions at the−1 subsite (the one accommodating the hydrolyzed/transferred monosaccharide) (Teze et al., 2014). In this way, they created Ttβ-gly mutants (e.g., Tyr284Phe, Asn282Thr, Phe401Ser, Asn163Ala, Arg75Ala) that catalyzed transglycosylation between 4-nitrophenyl β-D-fucopyranoside (*p*NP-Fuc) (donor substrate) and *N*-methyl-*O*-benzyl-*N*-(β-*D*-glucopyranosyl)-hydroxylamine (BnON(Me)-Glc, acceptor), with increased yields of up to 82% against a 36% in wild type enzyme, although in all cases the catalytic efficiency

FIGURE 1 | Catalytic mechanism of the glycosylation and deglycosylation (hydrolysis or transglycosylation) steps in retaining β-glycoside hydrolases. In this work, nucleophile and acid/base residues are GLU 338 and GLU 164, respectively, the donor substrate is *p*NP-Fuc and the acceptor substrate for transglycosylation is BnON(Me)-Glc.

is significantly reduced when compared to WT Ttβ-gly. The approach has since then been applied to other GHs families (glycosyl hydrolases have been classified into different families according to sequence similarity) (Henrissat, 1991; Cantarel et al., 2009; Teze et al., 2015; Saumonneau et al., 2016). The authors postulated that mutation of these well-conserved residues around the−1 subsite may induce lower stabilization of the deglycosylation transition states (TS), being the effect larger for hydrolysis than for transglycosylation reaction, therefore resulting in the higher T/H ratio measured experimentally (Teze et al., 2014). This suggests that the H and T transition states present different characteristics and stabilization interactions. For another family GH1 β-glucosidase (NkBgl), mutation of the catalytic acid/base glutamate to aspartate was also found to generate new transglycosylation products not produced by the wild-type enzyme (Jeng et al., 2012). The authors also proposed that this catalytic residue is important in substrate entry and

product release and noticed the importance of aromatic rings in the aglycone site to facilitate acceptor (other than water) binding. Compiling evidences support the idea that the properties of the deglycosylation transition state, substrate-specific interactions involving the acceptor substrate and specific features to channel and retain water molecules close to the catalytic center, are key determinants for the T/H partition (Kempton and Withers, 1992; Bissaro et al., 2015a,b; David et al., 2017). Still, the molecular details for the basis of T/H modulation are unclear.

Computational studies on retaining GHs have mainly been focused on the glycosylation step, fewer to the hydrolysis one and very few to transglycosylation. For enzymes belonging to CAZy family GH1, the one investigated in this work, quantum mechanical molecular mechanical (QM/MM) studies on glycosylation and hydrolysis steps catalyzed by *Oryza sativa* (rice) β-glucosidase (Osβ-gly) acting on glucose disaccharides have been reported (Wang et al., 2011, 2013; Badieyan et al.,

2012). They corroborated the predicted oxocarbenium-like nature of both TSs, which present the corresponding breaking bond already broken and the forming one far from established. The interaction of the C2 hydroxyl group from the sugar with the catalytic residues was also confirmed to play an important role in the catalytic reaction by facilitating the sugar ring conformational change toward the TS and by contributing to TS stabilization. This is in agreement with experimental data that have for long identified this interaction as important for TS stabilization, with a contribution of up to 10 kcal/mol (Zechel and Withers, 2000). Recently, the effect of such interaction on lowering the free energy barrier for the glycosylation and transglycosylation reactions catalyzed by a transglycosidase of CAZy family 72, was calculated to be of 11 and 16 kcal/mol, respectively (Raich et al., 2016). The effect of using different QM/MM partitions on the barrier heights of glycosylation and hydrolysis in Osβ-gly has also been investigated (Badieyan et al., 2012). It was concluded that a Tyr residue equivalent to Ttβ-gly Tyr284 (thus interacting with the catalytic nucleophile and the sugar O5 atom), has an important contribution to the energy profile, being the effect more significant for deglycosylation than for glycosylation. Importantly, the study showed that this residue had to be included in the QM region to obtain reliable potential energy barriers and ring conformations. The effect of other residues from the −1 subsite was also analyzed.

Very few QM/MM studies have compared hydrolysis and transglycosylation reactions for retaining glycosidases and, to the best of our knowledge, none has focused on family GH1. BB1K:AMBER QM/MM calculations on a β-galactosidase belonging to family GH2 were used to study the catalytic mechanism of glycosylation, hydrolysis and transglycosylation reactions, the latter leading to different regioproducts (Bráa et al., 2010). All TSs were characterized as dissociative and with the proton still attached to the acidic residue, although, unfortunately, more detailed geometric information on the transglycosylation TS was not given. The 2-OH group H-bond with the nucleophile was also related to pyranosyl ring distortion and TS stabilization. The energy barriers for transglycosylation were found to be higher than for hydrolysis by ∼2–4 kcal/mol, and the origin of the observed regioselectivity was found to be thermodynamic more than kinetic. Recently, the hydrolysis and transglycosylation reactions catalyzed by a family GH3 β-glucosidase have been studied by umbrella sampling calculations at the SCC-DFTB/CHARMM level (Geronimo et al., 2018). Both reactions were found to have similar free energy barriers (∼18 kcal/mol) but quite different TSs; the predicted TS for hydrolysis also differed from the ones previously described. For hydrolysis, the TS had the glycosyl-water bond practically formed (1.50 ± 0.04 Å), resulting in a reduced ionic character of the sugar which, according to the Cremer-Pople polar coordinates (Cremer and Pople, 1975) that describe ring conformations, presented a $^4C_1$ conformation. For transglycosylation, an earlier TS was predicted (with the new bond only partially formed, 1.9 ± 0.1 Å), with the transferred sugar more positively charged and in a $^4H_3$ conformation. In both cases, proton transfer to the catalytic base occurred late, especially in the hydrolysis step, where after approaching the base catalyst at 1.6 Å.

Here we present a QM(DFT)/MM study of the glycosylation, hydrolysis and transglycosylation steps catalyzed by wild-type Ttβ-gly using pNP-Fuc as donor and BnON(Me)-Glc as acceptor. Ttβ-gly catalyzes the hydrolysis of β-D-galactoside, β-D-glucoside, and β-D-fucoside derivatives, showing the highest activity with the latter (Dion et al., 1999). As mentioned, in the presence of a suitable acceptor, it can also catalyze transglycosylation; BnON(Me)-Glc has been reported to produce a high transglycosylation yield in the native enzyme and a sole regioisomer product (N-methyl-O-benzyl-N-(β-D-fucopyranosyl(1→ 4)β-D-glucopyranosyl)-hydr oxylamine) (Teze et al., 2013). Our main goal is to gain understanding and compare the molecular mechanisms underlying hydrolysis and transglycosylation. It has been proposed that the main driving force for increasing the T/H ratio must be the relative destabilization of the hydrolysis TS as compared to the transglycosylation one, which implies that differential interactions are involved in each case.

METHODOLGY

System Setup

Initial coordinates for the wild-type Ttβ-gly were taken from the corresponding X-ray structure with PDB code 1UG6 (resolution 0.99 Å) (Lokanath et al., in press). For the incorporation of the substrates into the enzyme active site, the 1UG6 coordinates were overlaid with those of the Osβ-gly structure crystalized with non-hydrolyzed cellotetraose (CTT) substrate(PDB code: 3F5J, resolution 1.95 Å) (Chuenchor et al., 2011). In this way, we took advantage of the position of CTT and modified it with PyMol program[1], by building the fucose moiety from the CTT glucose ring closer to the catalytic residues (Glu338 and Glu164 of the 1UG6 wild-type structure). The second ring of the CTT substrate was used to build up the (p-nitrophenol (pNP)) leaving group of the pNP-Fuc substrate. Two crystallographic water molecules from the wild-type structure that clashed to the added substrate were removed; the remaining of crystallographic waters were kept in the model. Missing hydrogen atoms were added to the model and the ionizable residues were protonated using Propka3.0 (Olsson et al., 2011) at pH = 7.0, except for Glu338 and Glu392 that were modeled as deprotonated, Glu164 as protonated and His119 was protonated at Nδ. All molecular dynamics (MD) simulations were done using the ff14SB (Maier et al., 2015) Amber force field for the protein; GLYCAM06j (Kirschner et al., 2008) and GAFF (Wang et al., 2004) atom types and parameters were employed for the sugar moieties and the pNP group, respectively. In addition, several missing parameters corresponding to a bond, and several intramolecular angles and dihedrals of the pNP-Fuc substrate were completed searching for the corresponding atom types translation in GAFF and GLYCAM06j force field databases (those parameters are given in **Supplementary Scheme S1** and **Supplementary Table S1**). Three sodium ions were added to neutralize the system, which was finally solvated in a cubic box of preequilibrated TIP3P

---

[1]The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

water molecules, with no <15 Å from the edge of the water box to the nearest protein atom. The resulting system contained 70,227 atoms.

Molecular Dynamics (MD) Simulation Details

The solvated Ttβ-gly/pNP-Fuc model system was relaxed by performing energy minimization at the MM level with the steepest descent and conjugate gradient methods. The minimization protocol consisted in a three-step process, firstly restraining all the system but the ligand (pNP-Fuc), secondly restraining only the protein atoms (except hydrogen atoms) and a third step without restraints. Once the system was relaxed, an MD simulation was performed under periodic boundary conditions (PBC). The MD protocol consisted in a heating simulation (from 0 to 300 K) during 200 ps under NVT conditions using Langevin dynamics. Following, an equilibration period of 400 ps was performed. The first 200 ps were run under NPT conditions to reach a system density of around $1 \, g \, cm^{-3}$ and using a isotropic weak-coupling algorithm and the Berendsen barostat (Berendsen et al., 1984) at 1 atm. Then, the system volume was fixed and the next 200 ps were run under NVT conditions. An equilibration of 10 ns and the final production run of 100 ns were performed at 300 K using the NVT ensemble without any restraints. Along the MD, the covalent bonds containing hydrogen were constrained using the SHAKE algorithm (Ryckaert et al., 1977), and the particle-mesh Ewald method (Essmann et al., 1995) was used to treat long-range electrostatic interactions. A 1fs time step was used in all the MD trajectories. The analysis of the MD simulation was carried out using the standard tools of the AMBER 16 package (Case et al., 2005) [cpptraj program (Roe and Cheatham, 2013)] and VMD (Humphrey et al., 1996). Based on this analysis, we established the criteria for the selection of a few frames of the MD trajectory that served as starting point for the QM/MM calculations of the glycosylation (G) reaction.

To obtain the starting structure for the QM/MM calculations of the hydrolysis reaction (H), the product geometry of the glycosylation process was modified by deleting all the pNP coordinates except those of the just formed hydroxyl group that was converted into a water molecule using the PyMol program. The system was solvated with a cubic box of TIP3P water molecules, and their coordinates were relaxed at the MM level by constraining all protein atoms and the just built water molecule (acceptor). This was followed by a short MD simulation (210 ps) to allow the rearrangement of water molecules in the active site. From the last frame of that MD trajectory a QM/MM minimization was carried out to locate a reactant minimum to initiate the hydrolysis reaction.

For the transglycosylation reaction (T), the coordinates of the above mentioned CTT substrate were used as template to build those of the BnON(Me)-Glc acceptor substrate, enforcing the 4-OH hydroxyl group from the glucose moiety to overlay with the hydroxyl coordinates of the pNP leaving group from the previous glycosylation step. Three water molecules were deleted due to clashes with the acceptor molecule. The system was relaxed as done for the hydrolysis reactant.

All simulations were carried out using AMBER16 software (GPU (CUDA) version of the PMEMD (Götz et al., 2012; Salomon-Ferrer et al., 2013) package.
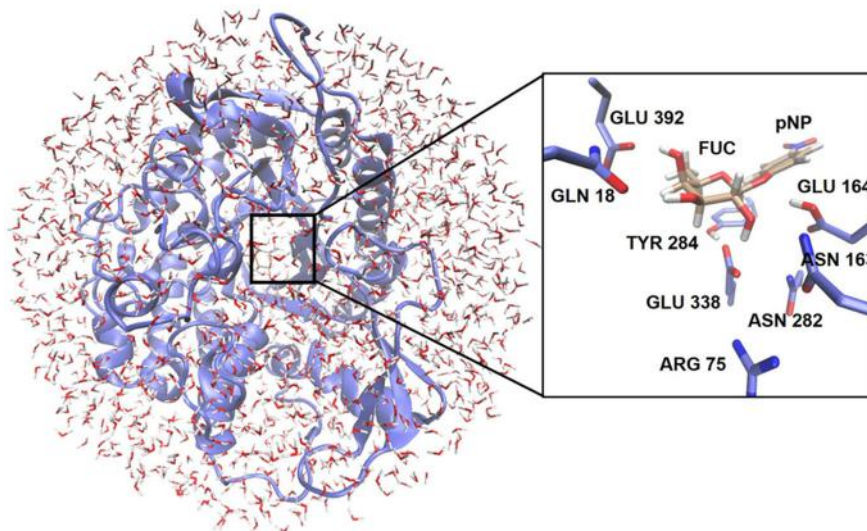
QM/MM Calculations

QM/MM calculations were performed with the modular program package ChemShell (Sherwood et al., 2003; Metz et al., 2014) using TURBOMOLE-V6.3 (Ahlrichs et al., 1989) to obtain the QM energies and gradients at the DFT level. The PBE0 functional (Adamo and Barone, 1999) was used, as it gave errors smaller than 0.5 kcal/mol in a recent benchmarking modeling glycosidic bond hydrolysis by glycosidases (Pereira et al., 2017) and inclusion of D3 correction did not affect significantly the calculated energies. MM energies and gradients were evaluated by DL_POLY (Smith and Forester, 1996), which was accessed through the ChemShell package, using the AMBER force field. The electrostatic embedding scheme (Bakowies and Thiel, 1996) was used within the QM/MM approach to let the MM point charges to polarize the electronic density of the QM region. No cutoffs were introduced for the nonbonding MM and QM/MM interactions. In all the QM/MM calculations, the cubic box of water molecules was simplified to a 30 Å sphere surrounding the full protein (**Figure 2**). All residues and water molecules within 15 Å from the anomeric center were included in the optimization process as active region (around 2,100 atoms) while the remaining atoms were kept fixed.

Two different QM regions were studied (**Figure 3**): the small QM region containing 84 atoms in the glycosylation step (the pNP-Fuc substrate, four nearby waters, and the side-chains of Glu164, Glu338, and Tyr284), and a large QM region with 138 atoms (including the small QM region and, in addition, the side-chains of Asn282, Asn163, Arg75, Gln18, and Glu392). For the hydrolysis and transglycosylation QM/MM descriptions the O-pNP group was substituted in the QM regions by a water molecule and the BnON(Me)-Glc molecule, respectively. Three water molecules were added (besides the nucleophilic water, WAT431) in the QM zones for the hydrolysis calculations whereas only one QM water molecule was left in the transglycosylation study. The number of QM atoms in the small/large regions for the hydrolysis and transglycosylation steps are 67/121 and 100/154, respectively. The methodology based on link atoms (three in the QM(small)/MM models and eight in the QM(large)/MM partitions) was used to define the QM/MM boundary with the charge-shift approach (Claeyssens et al., 2006). The total charge of both QM regions is−1 in the three catalytic reaction steps.

QM/MM optimizations were carried out employing the limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) (Liu and Nocedal, 1989) algorithm combined with the Hybrid Delocalized Internal Coordinate Scheme (Billeter et al., 2000) as implemented in Chemshell (Metz et al., 2014). Reaction paths were scanned by performing harmonically restrained optimizations along a properly defined reaction coordinate for each chemical process in steps of 0.2 Å. Reaction coordinates were defined as linear combinations of the three main distances involved in each step, a definition that we have successfully used when modeling other carbohydrate-active

FIGURE 2 | Representation of the enzymatic model used for the QM/MM calculations. The location of the QM region is highlighted. The inset shows as an example the arrangement of the QM atoms for the glycosylation reaction at the active site.
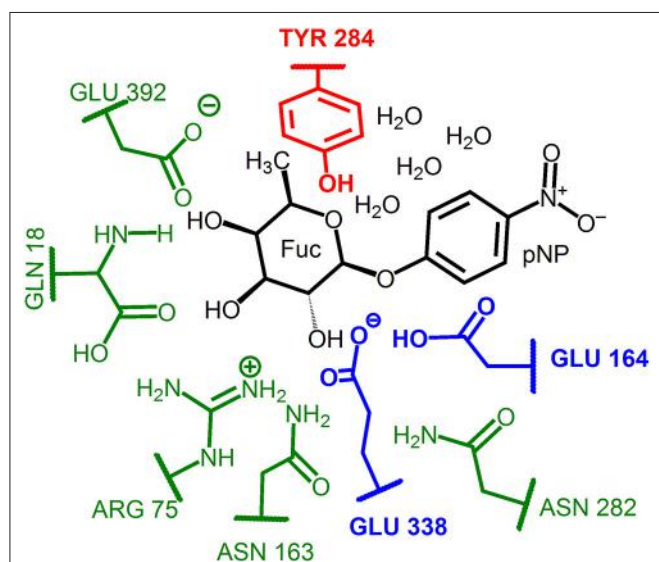
FIGURE 3 | Representation of the QM regions used in the QM/MM study of the glycosylation reaction. The small QM region includes: the substrate atoms and water molecules (in black), the catalytic residues (in blue) and the Tyr184 residue (in red). For the hydrolysis and transglycosylation the O-pNP group was substituted by a water molecule and the BnON(Me)-Glc acceptor substrate, respectively. Three additional waters were included for the hydrolysis and just one for the transglycosylation step. The large QM region comprises for the three reaction steps the corresponding small QM region plus the protein residues depicted in green.

enzymes (Gómez et al., 2015). Thus, RC = [d(C1$_{FUC}$-O4D$_{pNP}$)-d(C1$_{FUC}$-OE2$_{GLU338}$)-d(H$_{GLU164}$-O4D$_{pNP}$)] for glycosylation, RC = [d(C1$_{FUC}$-OE2$_{GLU338}$)-d(C1$_{FUC}$-O$_{WAT431}$)-d(H1$_{WAT431}$-OE2$_{GLU164}$)] for hydrolysis and RC = [d(C1$_{FUC}$-OE2$_{GLU338}$)-d(C1$_{FUC}$-O4$_{GLC}$)-d(H4O$_{GLC}$-OE2$_{GLU164}$)] for transglycosylation. For each reaction step, the structure

corresponding to the maximum of the potential energy profile was taken as indicative of the corresponding transition state. The PBE0 hybrid functional and two basis sets (SVP and TZVP) were used for the description of the small QM region, whereas the large QM region was always treated at the PBE0/TZVP level. In addition, QM(PBE0/TZVP)/MM single point energy calculations on the QM(PBE0/SVP)/MM geometries along the hydrolysis and transglycosylation potential energy profiles were also performed.

Natural population analysis (NPA) charges (Reed et al., 1985) were determined from QM/MM calculations with the QM region described at the PBE0/TZVP level. The contribution of the different residues to the QM/MM potential energy barriers of the hydrolysis and transglycosylation steps was examined by setting their point charges to zero in additional single point energy calculations at the QM(PBE0/TZVP)/MM level along the QM(PBE0/TZVP)/MM reaction paths.

Umbrella sampling at the QM(SCC-DFTB)/AMBER level was performed to compute the free energy profile for the hydrolysis and transglycosylation steps, using the dynamics module within ChemShell. The reaction coordinates, defined as before, were scanned at 0.1 Å intervals using a harmonic force constant of 300 kcal/mol·Å^2. The same force constant was used for restraining the distance between OE1$_{GLU338}$ and O2$_{FUC}$ to 2.4 Å in the hydrolysis step to correct for undesired proton transfers occurring during the simulations. Still, some of the hydrolysis data had to be discarded in the analyses presented below. Forty-five windows were used for hydrolysis and fifty for transglycosylation. Each simulation consisted of 20 ps of equilibration and 60 ps of production/data collection, under the NVT ensemble and using the Nosé-Hoover thermostat (Nosé, 1984; Hoover, 1985). All atoms from residues beyond 20 Å of C1$_{FUC}$ were frozen. The histograms show sufficient overlap of the windows for the chosen stepsize

and force constant. The umbrella integration analysis method (Kästner and Thiel, 2005, 2006) was used to compute the free energy profiles.

RESULTS AND DISCUSSION

Glycosylation Step

Molecular Dynamics Simulations

To obtain a first structural insight into the dynamics of the Ttβ-gly/pNP-Fuc Michaelis complex, the root-mean-square deviations (RMSDs) of the protein backbone and of the heavy atoms of the pNP-Fuc substrate were calculated along the 100 ns of production trajectory (**Supplementary Figure S1**). The protein structure is well equilibrated. The substrate RMSD (with an average value of 0.78 ± 0.37 Å) presents an oscillatory pattern because the pNP phenyl group is flipping between two conformations all along the trajectory. As it will be detailed below, the Fuc ring is more rigidly bound at the −1 subsite than the pNP group is at the larger +1 subsite, thus having this latter moiety of the substrate more room to move along the simulation.

To analyze in more detail the molecular interactions more likely to play a role in the catalytic mechanism, a search of the most populated hydrogen bonds between the substrate, protein residues and water molecules at the active site was carried out along the MD simulation. In **Table 1** a summary of these hydrogen bonds is presented (a more complete analysis is given in **Supplementary Tables S2–S5**). The fraction of frames the hydrogen bond is present (the H-bond occupancy), along with the corresponding donor⋯acceptor average distance, are given. An occupancy higher than 90% is obtained for the interaction between the carboxylate oxygen OE1 of the catalytic acid/base residue Glu164 and HD22 of Asn282. This interaction helps to maintain (with an occupancy of 90.6%) the hydrogen bond between the proton of Glu164 and the glycosidic oxygen (O4D$_{pNP}$). This is a very important interaction for the glycosylation process because in this way Glu164 remains well-positioned (average heavy atoms distance of 2.76 Å) for proton donation to the leaving group. The catalytic nucleophilic residue Glu338 mainly interacts through its carboxylate oxygen OE2 with Tyr284 (88.9% occupancy) so that the negative charge on the nucleophile becomes stabilized. In addition, this same carboxylate group of the nucleophilic residue Glu338 (through its OE1 oxygen) establishes nearly half of the time another hydrogen bond with the glycosyl 2-OH group of the substrate. Some oscillation of this 2-OH group between OE1$_{GLU338}$ (46.6% occupancy) and OE2$_{Glu338}$ (16.3% occupancy) is actually observed. As commented in the Introduction, this interaction is key for the stabilization of the glycosylation transition state, especially for retaining β-glycosidases because the interaction with the nucleophile will affect oxocarbenium cation formation. Moreover, OE1$_{Glu338}$ is also H-bonded with Arg75 side chain and a water molecule. On the other hand, the 2-OH group of the Fuc ring also interacts with Asn163 side-chain (with an occupancy of 46.8%). The 3-OH group of Fuc establishes a H-bond with HE2$_{HIS119}$ (34.4% occupancy) and with OE1$_{GLN18}$ (79.3%). The 4-OH and 5-OH of Fuc are hydrogen bonded as acceptors to the same water molecule (with an occupancy of around 20% for

TABLE 1 | Analysis over the MD simulation of the most populated hydrogen-bonds between pNP-Fuc, protein residues, and water molecules at the active site.

Acceptor	Donor	H-bond occupancy %	Average Distance (Å)
O$_{NO2pNP}$	Solvent	99.6	2.79
OE1$_{GLU164}$	ND2$_{ASN282}$	92.5	2.81
O4D$_{pNP}$	OE2$_{GLU164}$	90.6	2.76
OE2$_{GLU338}$	OH$_{TYR284}$	89.0	2.76
OE1$_{GLN18}$	O3$_{FUC}$	79.3	2.72
OE1$_{GLU392}$	O4$_{FUC}$	62.5	2.80
O2$_{FUC}$	ND2$_{ASN163}$	46.8	2.90
OE1$_{GLU338}$	O2$_{FUC}$	46.6	2.84
O3$_{FUC}$	NE2$_{HIE119}$	34.4	2.87
O5$_{FUC}$	Solvent	24.0	2.85
O4$_{FUC}$	Solvent	20.4	2.84
O3$_{FUC}$	NE1$_{TRP393}$	16.6	2.91
OE2$_{GLU338}$	O2$_{FUC}$	16.3	2.85
OE1$_{GLU338}$	NH1$_{ARG75}$	13.6	2.91

H-bond occupancies are defined as the fraction of frames the bond is present. Only H-bond occupancies above 10% are included. The average donor to acceptor heavy atoms distances of the bonds when present are also given.

each Fuc⋯water interaction), and the axial 4-OH of Fuc is H-bonded to OE1$_{GLU392}$ (62.5% occupancy). This latter interaction structurally differs from that reported by Badieyan et al. (2012) when cellobiose is hydrolyzed by Osβ-gly, as in Glc this hydroxyl group is in the equatorial position, which makes it interact with the other oxygen of the Glu392 carboxylate.

To summarize, the glycosyl moiety of the substrate bound at the −1 subsite establishes a complex H-bond network with the protein and some water molecules that contributes to its stabilization. It is worth highlighting that all the first- and second-shell protein residues around the −1 subsite identified in this H-bond analysis are highly conserved residues among the GH1 family (Bráa et al., 2010; Camargo et al., 2019). In contrast, the H-bond analysis did not identify any conserved residue around the +1 site so confirming the lack of specificity for the aglycone moiety within the GH1 family as previously observed (Bráa et al., 2010). The pNP molecule only establishes H-bonds with two water molecules all along the MD trajectory through its NO2 group. In addition to those H-bond interactions, the Fuc ring remains surrounded all along the trajectory by hydrophobic interactions established with conserved residues such as Trp385, Trp393, Trp120, and also Phe401, whereas the pNP leaving group basically interacts with Trp312.

To initiate the QM/MM study of the glycosylation catalytic mechanism, two MD snapshots of the Ttβ-gly/pNP-Fuc Michaelis complex were selected. Those structures were filtered by imposing the presence of the direct H-bonds (O4D$_{pNP}$-H$_{GLU164}$ and OE2$_{GLU338}$-H2O$_{FUC}$) between the two catalytic residues, Glu338 and Glu164, with the pNP-Fuc substrate. In addition, the presence of the H-bond between Tyr284 and Glu338 was also verified in the selected frames. The relevance of this interaction (and of describing it at the QM level) for the

energetics of the glycosylation step and hydrolysis steps has been recently highlighted by Badieyan et al. (2012) in their QM/MM study of Osβ-gly mentioned above. This is the reason why we have included Tyr284 in the QM(small)/MM partition. Finally, in the two frames selected it was also verified that the additional predominant H-bond interactions highlighted in **Table 1** were also present.

QM/MM Calculations

In this section the results of the QM/MM calculations on the glycosylation reaction will be presented. This catalytic step consists in the cleavage of the glycosidic bond of the pNP-Fuc molecule to form a covalent glycosyl-enzyme intermediate with Glu338. Despite pNP is a good leaving group, acid catalysis is required for this mechanistic step, a role that is carried out by the protonated Glu164 (**Figure 1**).

The two selected frames from the MD trajectory (denoted as G_I and G_{II} in this paper) were optimized using the two QM/MM partitions described above and the corresponding minima were located and characterized. From those minimum energy structures, the potential energy profiles along the glycosylation step were calculated as a function of the reaction coordinate RC $= [d(C1_{FUC}\text{-}O4D_{pNP})\text{-}d(C1_{FUC}\text{-}OE2_{GLU338})\text{-}d(H_{GLU164}\text{-}O4D_{pNP})]$. In **Table 2** the QM/MM potential energy barriers (obtained from the corresponding energy profiles) and the reaction energies (considering the optimized glycosyl-enzyme product) for the glycosylation step at the QM = PBE0/SVP and PBE0/TZVP levels using the QM(small)/MM and QM(large)/MM partitions are given for both frames. The corresponding energy values obtained by means of PBE0/TZVP single-point energy calculations on the PBE0/SVP(small)/MM geometries are also given. It can be observed that the potential energy barrier using the QM(small)/MM partition takes values in between 28 and 30 kcal/mol when the smaller basis set is used. The energy correction using the larger basis set tends to increase those energy barriers (by 0.4–2 kcal/mol) especially for the G_{II} frame. However, when the geometry optimization is carried out with the TZVP basis set, no significant differences are obtained with respect to the SVP energy values. These energy barriers are in the range of that calculated with a large QM region for the glycosylation reaction involving cellobiose and Osβ-gly (Badieyan et al., 2012), and also of those calculated for other retaining GHs (Petersen et al., 2009, 2010; Biarnés et al., 2011). However, a large effect is observed in the present system when the large QM region is used with the TZVP basis set. Lower potential energy barriers (by 5.9 and 9.8 kcal/mol for frames G_I and G_{II}, respectively) were obtained for both frames. Notice that, as pNP is a better leaving group than the Glc of cellobiose modeled in the glycosylation reaction catalyzed by Osβ-gly (Badieyan et al., 2012), a lower energy barrier should be expected. In particular, the potential energy barriers calculated at QM(large)/MM level and the TZVP basis set are 22.0 kcal/mol (for G_I) and 20.4 kcal/mol (G_{II}). These values are in better and in reasonable qualitative agreement with the value of 17.1 kcal/mol for the phenomenological free energy of activation derived from the experimental k_{cat} at 40°C (Teze, 2012). The significant differences between the calculated potential energy

TABLE 2 | QM/MM potential energy barriers ($\Delta V^{\neq}$) and reaction energies (ΔV_R) calculated for the glycosylation step for the two complexes studied (G_I and G_{II}).

| QM level	G_I		G_{II}	
	$\Delta V^{\neq}$	ΔV_R	$\Delta V^{\neq}$	ΔV_R
PBE0/SVP[a]	28.4	7.1	30.2	0.4
PBE0/TZVP//PBE0/SVP[a]	28.8	5.1	32.1	−0.1
PBE0/TZVP[a]	27.9	6.3	30.2	0.6
PBE0/TZVP[b]	22.0	4.8	20.4	6.6

[a] QM(small)/MM.
[b] QM(large)/MM.
Energies are given in kcal/mol.

barriers reflect that they are very sensitive to different QM/MM partitioning schemes and that non-catalytic residues (other than Tyr284) might play a role in the stabilization of the glycosylation transition state, as previously stated. As for the reaction energies, the glycosylation process turns out to be more endoergic for frame G_I than for frame G_{II} when the QM(small)/MM partition is used. However, both glycosylation reactions become clearly endoergic when the QM region is enlarged. Interestingly, the endothermic nature of the glycosylation reaction was recently observed in previous studies where their authors suggest that an endothermic glycosylation might be a prerequisite for efficient transglycosylation (Raich et al., 2016).

In **Figure 4** the evolution of the four main interatomic distances involved in the glycosylation reaction ($C1_{FUC}$ - $O4D_{pNP}$, $C1_{FUC}$ - $OE2_{GLU338}$, H_{GLU164} - $O4D_{pNP}$, and H_{GLU164} - $OE2_{GLU164}$) are plotted for the G_I frame along the reaction coordinate calculated with the QM(large)/MM partition and at the PBE0/TZVP level for the QM region. In **Figure 5** the molecular representation of the corresponding reactant, transition state and product are depicted. It can be observed that the $C1_{FUC}$ - $OE2_{GLU338}$ distance between the nucleophile residue and the anomeric carbon gradually diminishes from 3.33 Å at reactants (RC$= -3.90$ Å) to a value of 2.61 Å at the transition state (RC$= -2.10$ Å) (what corresponds to a bond not yet formed), and finally takes a value of 1.50 Å at the covalent glycosyl-enzyme intermediate (RC$= 0.70$ Å) in which this bond is already completely formed. The length of the glycosidic bond $C1_{FUC}$ - $O4D_{pNP}$ is 1.41 Å at the reactants, increases to 2.33 Å at the TS (what corresponds to the glycosylic bond already broken) and finally takes a value of 3.17 Å at the product structure in which the leaving group has been released. For the second frame (G_{II}) the distances are very similar (**Supplementary Table S6**). These structural results are coherent with the dissociative nature of the glycosylation transition state. The oxocarbenium character of this TS is also made apparent by a higher degree of double bond character of the $C1_{FUC}$-$O5_{FUC}$ bond (the $C1_{FUC}$-$O5_{FUC}$ distance diminishes 0.14 Å from the reactant to the TS structure for both frames). In addition, the acidic residue Glu164 gradually approaches the glycosidic bond along the two reaction paths (the $O4D_{pNP}$ - H_{GLU164} hydrogen bond diminishes from 1.99 Å (G_I) and 2.02 Å (G_{II}) at the reactants to 1.82 and 1.88 Å at the G_I and G_{II} TS, respectively) although the proton is not yet transferred. Proton transfer takes place when the
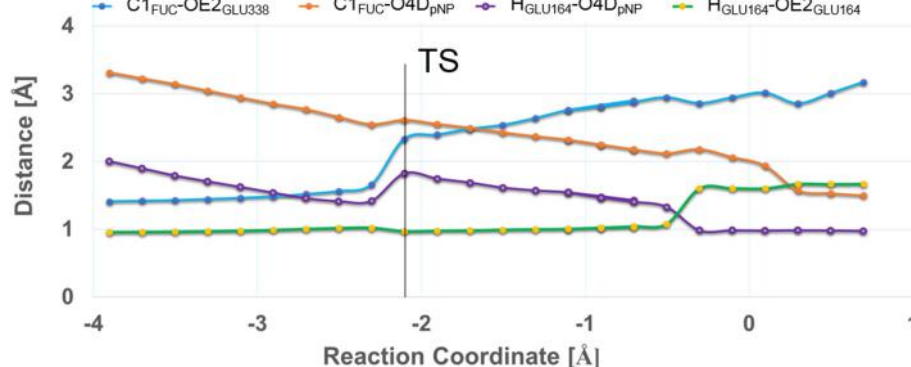
FIGURE 4 | Evolution of $C1_{FUC}$-$O4D_{pNP}$ (in blue), $C1_{FUC}$-$OE2_{GLU338}$ (in orange), H_{GLU164}-$O4D_{pNP}$ (in purple), and H_{GLU164}-$OE2_{GLU164}$ (in green) distances (in Å) along the glycosylation reaction. The results correspond to frame G_I calculated with the QM(large)/MM partition and at the QM(PBE0/TZVP) level for the QM region.
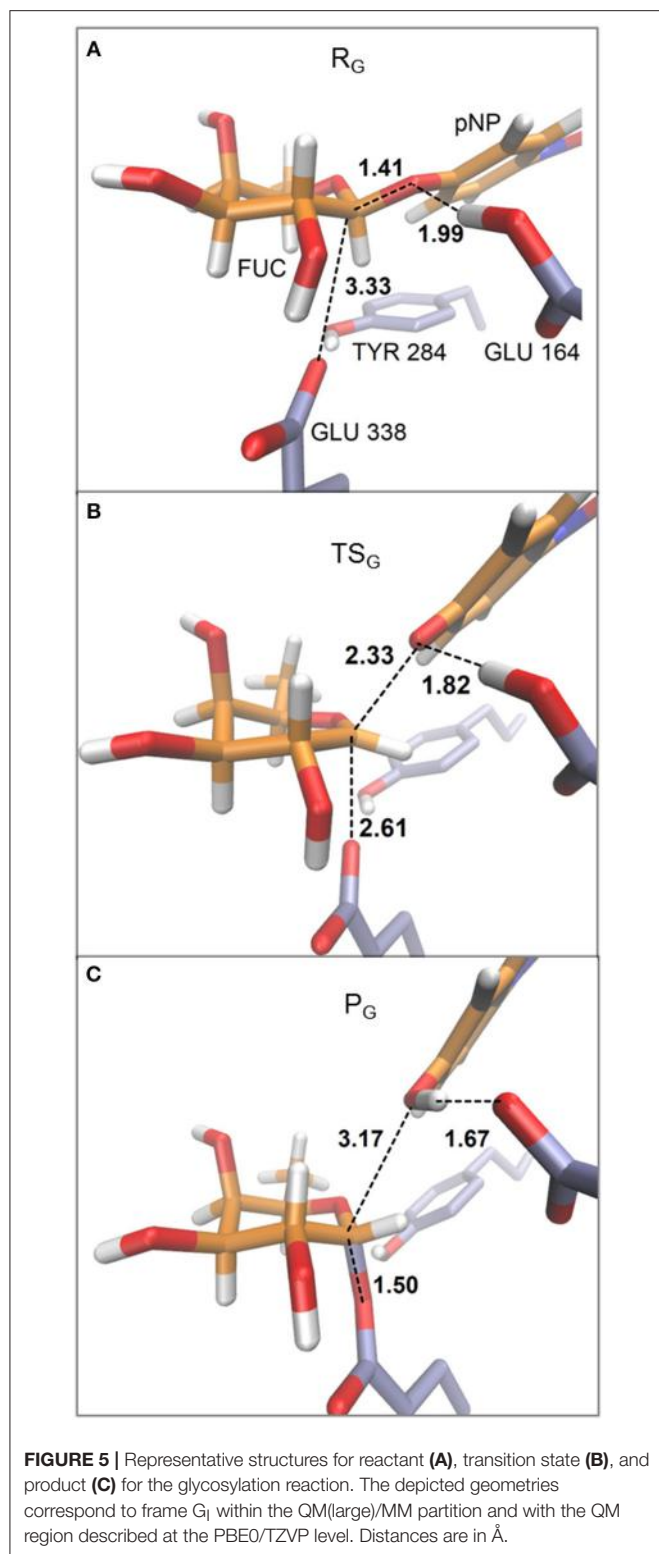
$OE2_{GLU338}$-$C1_{FUC}$ distance is ~2 Å, and is followed by complete $OE2_{GLU338}$-$C1_{FUC}$ bond formation. This is consistent with a change in pK_a of Glu164 that makes proton transfer easier as the negative charge of the leaving group gradually increases and that of the nucleophile gradually decreases with the progress of the glycosylation reaction. When a smaller QM region is used within the QM/MM approach, and independently of the basis set employed for the QM description, the transition state can also be described as dissociative (**Supplementary Table S7** and **Supplementary Figure S2**). Curiously, the $O4D_{pNP}$ - H_{GLU164} H-bond distance at the reactant varies by 0.42 Å between the two frames considered (G_I and G_{II}), and a significant difference in this distance is also observed at the corresponding TSs. When applying the QM(large)/MM partition, these differences between frames disappear. With the smaller QM region, the nucleophilic attack by Glu338 is completed at lower RC values than with the QM(large)/MM partition and quite before the proton is fully transferred. Thus, the description at the QM level of the H-bonds that Glu338 and Glu164 establish with their neighboring residues seems to affect significantly their nucleophilic and acid/base characteristics.

The inclusion of other protein residues besides the catalytic ones (Glu338, and Glu164) and Tyr184 in the zone treated quantum-mechanically (as in our QM(large)/MM model) also affects the stabilization by the enzyme environment of the oxocarbenium ion and the negatively charged glycosyl oxygen anion. Moreover, as mentioned in the Introduction, one of the most stabilizing protein-substrate interactions at the glycosylation TS is the H-bond between the 2-OH group of the sugar moiety and the nucleophilic residue Glu338. This stabilizing contribution correlates with a reduction of the $OE1_{GLU338}$ - $H2O_{FUC}$ H-bond distance, from 2.09 Å and 2.37 Å at reactants to 1.82 Å and 1.85 Å at the TS of G_I and G_{II}, respectively (**Supplementary Table S6**). The Fuc at the −1 subsite is also stabilized along the glycosylation pathway by the $HD21_{ASN163}$ - $O2_{FUC}$ interaction as the corresponding H-bond distance diminishes around 0.10 Å from reactants to the transition state structures. Tyr284 plays a relevant role along the glycosylation process. As the $OE2_{GLU338}$ atom loses negative charge upon

formation of the $C1_{FUC}$ - $OE2_{GLU338}$ bond, the Tyr284 hydroxyl group separates from Glu338 and clearly approaches (with a reduction in the bond distance of around 1.3 Å at the TSs and of around 1.9 Å at the product structures) the $O5_{FUC}$ atom. The expected oxocarbenium ion-like character of the TS is favored by the coplanarity of the C2, C1, O5, and C5 atoms of the fucose ring which adopts a slightly distorted chair conformation close to a half-chair ($^4H_3$) and with some ^4E character.

At this point it is interesting to comment that it has been shown that GHs usually bind the substrate in a subtle but critical distorted conformation (Davies et al., 2012). Such conformational change brings the glycosidic bond closer to the equatorial orientation and to a more planar ring structure (so closer to the oxocarbenium TS geometry) that facilitates in-line nucleophilic attack. A possible exception to this has been recently reported for family GH3 enzymes, for which the glucose bound at the −1 subsite seems to adopt a $^4C_1$ conformation (maybe slightly distorted to $^4H_5$) in the Michaelis complex according to QM/MM calculations and crystallographic data (Hrmova et al., 2001; Geronimo et al., 2018). Different conformational itineraries (from reactants to TS and to products) have been associated with different GHs (Ardèvol and Rovira, 2015). Thus, for retaining β-glucosidases acting on glucose substrates the $^1S_3 \rightarrow {^4H_3} \rightarrow {^4C_1}$ itinerary has been described for the glycosylation step. The QM/MM study by Badieyan et al. (2012) of the glycosylation step catalyzed by rice Osβ-gly using cellobiose as substrate agrees with this itinerary. However, in the present study, in which we have pNP-Fuc instead of Glc(β1-4)β-Glc as substrate, we were unable to characterize a Michaelis complex in a $^1S_3$ conformation. The QM(large)/MM optimized reactant has a slightly distorted $^4C_1$ conformation (with θ and φ angles values of 216 and 27°, respectively). Attempts to obtain a different conformation by maintaining the ring restrained previous to the full QM(large)/MM minimization only produced an alternative minimum close to a ^4E (slightly $^4H_3$) conformation. This structure presents a small increase in the $C1_{FUC}$ – $O4D_{pNP}$ distance (1.49 Å) and slightly shorter $C1_{FUC}$ – $OE2_{GLU338}$ (3.29 Å), $C1_{FUC}$ – $O5_{FUC}$ (1.36 Å), and H_{GLU164} – $O4D_{pNP}$ (1.77 Å) distances than the previous one. The potential energy

FIGURE 5 | Representative structures for reactant **(A)**, transition state **(B)**, and product **(C)** for the glycosylation reaction. The depicted geometries correspond to frame G_I within the QM(large)/MM partition and with the QM region described at the PBE0/TZVP level. Distances are in Å.

profile calculated from this minimum gives a barrier height of only 10 kcal/mol (**Supplementary Figure S3**), which is much smaller than the experimental value. This suggests that, although such conformation may be possible, it may not be the most

favorable conformation that the enzyme will bind and, thus, it has no catalytic significance. Badieyan et al. (2012) reported the importance of having Glu440 of Osβ-gly (that interacts with the 4-OH and 6-OH groups of Glc) in the QM region in order to characterize the $^1S_3$ conformation. In the present study, the QM(large)/MM partition includes the equivalent Glu392 residue, amongst many others. At the origin of these differences may be the fact that we are modeling Fuc instead of Glc (they differ in the position of the 4-OH group and Fuc has no 6-OH group but a methyl in this position); or that our substrate is not an oligosaccharide but a synthetic phenyl derivative that may introduce less steric requirements than a monosaccharide at the +1 subsite. Nevertheless, as glycosidic bond breakage occurs early in the reaction, the ring is distorted early along the calculated reaction coordinate with subsequent attack by the nucleophilic Glu338 occurring in the expected orientation.

Hydrolysis and Transglycosylation Steps (Deglycosylation)

Once the pNP molecule leaves the enzyme active site, a water molecule can compete as nucleophile in a hydrolysis reaction with the acceptor ligand (BnON(Me)-Glc), that also attacks as a nucleophile on the anomeric carbon of the fucose molecule in a transglycosylation reaction leading to the N-methyl-O-benzyl-N-($β$-D-fucopyranosyl(1-4)$β$-D-glucopyranosyl)-hydroxylamine product. The attacking water molecule as well as the BnON(Me)-Glc ligand are activated by a proton transfer to the Glu164 residue that acts as a base along those two nucleophilic processes.

Starting from the minimized MD structure (see Methodology section) in which a water molecule (WAT431) was located at the active site, close enough to the anomeric carbon of fucose ring, we calculated the QM/MM energy profile for the hydrolysis process using a reaction coordinate defined as $RC = [d(C1_{FUC}\text{-}OE2_{GLU338})\text{-}d(C1_{FUC}\text{-}O_{WAT431})\text{-}d(H1_{WAT431}\text{-}OE2_{GLU164})]$. As mentioned above, a H-bond network of three water molecules, in addition to the nucleophilic WAT431, were included in the QM region for the hydrolysis simulation. On the other hand, for the transglycosylation step, initial energy profiles starting from minimized MD structures gave highly overestimated energy barriers (>34 kcal/mol), suggesting some structural deficiencies in the model. With the aim of improving this model, we followed a strategy successfully used by our group in previous studies of enzyme catalyzed reactions involving sugar-transfer between donor and acceptor substrates (Albesa-Jové et al., 2015; Mendoza et al., 2016, 2017). The refinement protocol consists on running a QM(SCC-DFTB)/MM MD of, in this case, the transglycosylation product, followed by scan calculations of frames from this MD to generate new models for the transglycosylation reactant that is then minimized at the QM(PBE0)/MM level. By doing this, interactions between donor and acceptor (or also involving water molecules) that are difficult to capture in the MM MD seem to be better represented. In the present system, the structural changes observed upon refinement involve the reorientation of the 4-OH hydroxyl of the acceptor Glc toward the Fuc ring. This displaces a water molecule that was located between the two and that most likely hindered acceptor

approach during reaction (**Supplementary Figure S4**). In the new reactant structure, the Glc hydroxyl will be able to better interact with Fuc as transglycosylation takes place. Substrate-substrate interactions have been shown to have a very important contribution in the sugar transfer catalyzed by retaining glycosyltransferases (Gómez et al., 2012, 2013; Mendoza et al., 2017) and could, thus, be also relevant for transglycosylation. Two frames (called here T_I and T_{II}) were generated by this approach. From the two optimized QM/MM reactant structures, the transglycosylation potential energy profiles for the corresponding β(1-4) glycosidic linkages were calculated using as reaction coordinate the following expression RC = [d(C1$_{FUC}$-OE2$_{GLU338}$)-d(C1$_{FUC}$-O4$_{GLC}$)-d(H4O$_{GLC}$-OE2$_{GLU164}$)]. In this model only one water molecule remains in the QM region. We did not analyze other glycosidic linkages between the Fuc and BnON(Me)-Glc because it has been proven experimentally that the native enzyme does not produce other product regioisomers when these two substrates are used (Teze et al., 2013).

In **Table 3** the potential energy barriers and the reaction energies corresponding to the hydrolysis and the transglycosylation steps are given for the two QM/MM partitions used in this study and at the different QM levels of electronic structure theory analyzed. The first general observation is that the potential energy barrier for the hydrolysis step is very similar to that of the transglycosylation one, as expected because the experimental result for this reaction catalyzed by WT Ttβ-gly gives a transglycosylation yield of 36%, which implies differences in the energy barriers no larger than 1 kcal/mol. With the small QM region at the PBE0/SVP level, the hydrolysis potential energy barrier is 2.1 and 0.9 kcal/mol higher than for the transglycosylation frames T_I and T_{II}, respectively. According to this result, the transglycosylation reaction would be faster than the hydrolysis process. This trend is inverted when single-point energy calculations are carried out at a higher level (QM(PBE0/TZVP)), being the hydrolysis energy barrier now slightly lower (by 0.7 (T_I) and 0.9 (T_{II}) kcal/mol). The correct trend between the hydrolysis and transglycosylation barriers is confirmed when a bigger basis set is used in the QM/MM optimizations and also when the QM region is enlarged. As for the glycosylation step, the values of the energy barriers calculated with the smaller QM region are around 6 kcal/mol higher than those obtained with the larger QM region. Those energy differences reflect again that the two reaction processes are very sensitive to different QM/MM partitioning schemes. Despite this, the difference between transglycosylation and hydrolysis energy barriers is similar (1.2 kcal/mol for T_I and 2.6 kcal/mol for T_{II}). Thus, the energy trends might be reproduced at a lower computational cost (QM(PBE0/TZVP)/MM but QM(small)/MM. As for the reaction energies, the hydrolysis process is more exoergic than the transglycosylation reaction. This result is in agreement with previous results by Bráa et al. (2010) in their study of the catalytic mechanism of a β-galactosidase and with the known thermodynamically favored hydrolysis over synthesis reaction (Planas and Faijes, 2002).

In **Table 4** the most relevant interatomic distances (calculated with the QM(large)/MM at the QM(PBE0/TZVP)/MM level) along the hydrolysis and transglycosylation pathways are

presented and the corresponding molecular structures of the different reactants, transition states and products are depicted in **Figure 6** (for QM(small)/MM results see **Supplementary Table S8**). At reactants, with QM(large)/MM, the initial location of the water molecule in the hydrolysis process corresponds to a slightly more distant position from the anomeric carbon (3.45 Å) than in the two reactants of the transglycosylation process (3.27 and 3.35 Å). The hydrolysis values reported for Glc hydrolysis by rice β-glucosidase (Osβ-gly) are slightly larger (3.62–3.75 Å) (Badieyan et al., 2012; Wang et al., 2013), as occurs at the same level of theory when the QM(small)/MM partition is used (**Supplementary Table S8**). Also at TS structures the C1$_{FUC}$-O$_{Acc}$ bond, that is still forming, is slightly longer for hydrolysis (with distances of 2.27, 2.11, and 2.12 Å at H, T_I and T_{II} transition states, respectively), while the nucleophile Glu338 side chain is completely displaced from the sugar in both reactions but slightly closer from C1$_{FUC}$ in the case of hydrolysis (with C1$_{FUC}$-OE2$_{GLU338}$ distances of 3.12, 3.42, and 3.34 Å at H, T_I, and T_{II} transition states, respectively). For Glc hydrolysis by Osβ-gly, slightly shorter distances have been reported, especially for the C1$_{GLC}$-O$_{Nu}$ one (1.89–2.14 and 2.45–2.51 Å, for C1$_{GLC}$-O$_{Acc}$ and C1$_{GLC}$-O$_{Nu}$, respectively) (Badieyan et al., 2012; Wang et al., 2013). Along the nucleophilic attacks, the proton from the water molecule or the BnON(Me)-Glc acceptor comes closer to the general base residue Glu164 but it has not been transferred yet at the transition states (H$_{Acc}$-OE2$_{GLU164}$ distances between 1.42 and 1.48 Å). This is in agreement with the hydrolysis description given on previous works (Badieyan et al., 2012; Wang et al., 2013; Geronimo et al., 2018). Interestingly, this proton transfer is more advanced in the hydrolysis and transglycosylation TSs that it was the Glu164 to pNP one in the glycosylation TS (d(H$_{GLU164}$-O4D$_{pNP}$ ~1.8Å). Notice that this is the first time that the transglycosylation step is studied for wild-type family GH1, thus direct comparison for this step with previous works is not possible. For family GH2, potential energy scans followed by TS characterization identified a dissociative TS for transglycosylation and with the proton not yet transferred; however, detailed structural information (i.e., distances) was not provided (Bráa et al., 2010). Besides, for the GH2 enzyme a magnesium ion interacting with the acid/base residue participates in catalysis, which is likely to introduce some structural differences when compared to the present system (e.g., the hydrolysis TS presented the anomeric C1 atom equidistant from the nucleophile and the attacking water (2.25 Å), which differs from all the results commented above).

Despite the sensitivity observed for these TS geometries to the DFT functional, the basis set, the QM/MM partition, the characterization method, the substrates or the specific enzyme, it is clear that both the hydrolysis and the transglycosylation transition states are very loose structurally, and they have a dissociative nature. From the distances reported in **Table 4**, which are thus to be taken as orientative, some trends can be extracted at the QM(PBE0/TZVP)/MM with QM(large)/MM level. For example, the sum of the C1$_{FUC}$-O$_{Acc}$ and C1$_{FUC}$-OE2$_{GLU338}$ distances for the deglycosylation TSs (5.39, 5.53, and 5.46 Å for H, T_I, and T_{II} transition states, respectively)

TABLE 3 | Potential energy barriers ($\Delta V^{\neq}$) for the hydrolysis (H) and transglycosylation processes (T_I and T_{II}, corresponding to the two different frames studied) along with the corresponding reaction energies (ΔV_R) with the two QM/MM partitions and at the different QM levels employed.

	PBE0/SVP[a]		PBE0/TZVP//PBE0/SVP[a]		PBE0/TZVP[a]		PBE0/TZVP[b]	
	$\Delta V^{\neq}$	ΔV_R	$\Delta V^{\neq}$	ΔV_R	$\Delta V^{\neq}$	ΔV_R	$\Delta V^{\neq}$	ΔV_R
H	25.4	−16.7	22.7	−17.8	24.6	−15.8	18.3	−17.4
T_I	23.3	−9.7	23.4	−8.5	25.6	−8.3	19.5	−12.5
T_{II}	24.5	−8.9	23.6	−4.6	27.9	−5.8	20.9	−4.0

[a]QM(small)/MM.
[b]QM(large)/MM.
All energies are given in kcal/mol.

TABLE 4 | Distances (in Å) between selected atoms involved in each reaction step for the reactant, transition state (TS) and product of the hydrolysis (H) and transglycosylation (T_I and T_{II}, corresponding to the two different frames studied) steps.

	Reactant			TS			Product		
	H	T_I	T_{II}	H	T_I	T_{II}	H	T_I	T_{II}
$d(C1_{FUC}\text{-}OE2_{GLU338})$	1.49	1.51	1.49	3.12	3.42	3.34	3.32	3.37	3.31
$d(C1_{FUC}\text{-}O_{Acc})$	3.45	3.27	3.35	2.27	2.11	2.12	1.40	1.42	1.43
$d(C1_{FUC}\text{-}O5_{FUC})$	1.36	1.35	1.36	1.24	1.25	1.25	1.39	1.38	1.38
$d(H_{Acc}\text{-}OE2_{GLU164})$	1.81	1.80	1.74	1.48	1.48	1.42	0.96	0.96	0.96
$d(H_{TYR284}\text{-}O5_{FUC})$	2.08	2.19	2.13	3.00	3.22	3.19	3.94	3.98	3.81
$d(H_{TYR284}\text{-}OE2_{GLU338})$	2.46	2.46	2.48	1.90	1.86	1.87	1.82	1.83	1.82
$d(H_{ARG75}\text{-}OE1_{GLU338})$	2.19	2.25	2.25	2.09	2.27	2.10	2.05	2.11	2.08
$d(H2O_{FUC}\text{-}OE1_{GLU338})$	1.78	1.78	1.75	1.89	1.84	1.75	2.05	2.15	1.92
$d(H4O_{FUC}\text{-}O_{GLU392})$	2.29	2.73	4.13	2.16	2.56	4.10	1.93	2.00	4.00
$d(H_{ASN163}\text{-}O2_{FUC})$	1.97	1.94	1.93	1.96	1.94	1.93	1.96	2.06	1.98
$d(H3O_{GLC}\text{-}O4_{FUC})$	–	2.86	2.83	–	2.74	2.45	–	3.11	2.35
$d(H3O_{GLC}\text{-}O5_{FUC})$	–	3.96	4.06	–	3.43	3.55	–	1.88	2.35
$d(H3O_{GLC}\text{-}O4_{GLC})$	–	2.34	2.33	–	2.41	2.41	–	2.53	2.57
$d(O3_{GLC}\text{-}H2_{WAT433})$	–	2.00	2.23	–	2.03	2.25	–	2.01	2.25
$d(O_{WAT431}\text{-}H2_{WAT432})$	1.82	–	–	1.86	–	–	2.77	–	–
$d(H2_{WAT431}\text{-}O_{WAT433})$	2.02	–	–	2.19	–	–	2.33	–	–
$d(H2_{WAT432}\text{-}O5_{FUC})$	3.58	–	–	3.33	–	–	1.95	–	–
$d(H1_{WAT432}\text{-}O4_{FUC})$	2.28	–	–	2.25	–	–	2.59	–	–

The results correspond to QM(large)/MM and at the PBE0/TZVP level. The Acc subscript refers to the acceptor water and glucose moieties in hydrolysis and transglycosylation processes, respectively.

are all significantly larger than the sum of the analogous $C1_{FUC}\text{-}O4D_{PNP}$ and $C1_{FUC}\text{-}OE2_{GLU338}$ distances at the glycosylation TS (which is 4.97 Å on average). On the other hand, the differences between $C1_{FUC}\text{-}O_{Acc}$ and $C1_{FUC}\text{-}OE2_{GLU338}$ distances in the deglycosylation TSs (0.85, 1.31, and 1.22 Å for H, T_I, and T_{II} TS structures, respectively) are much larger than the corresponding glycosylation value (0.27 Å). Notice also that the hydrolysis value is significantly smaller than the transglycosylation ones, and of the same order as the values obtained for Glc hydrolysis by Osβ-gly (0.31–0.62 Å) (Badieyan et al., 2012; Wang et al., 2013). These results are in agreement with experimental observations indicating a more S_N2 character for the glycosylation step than for the deglycosylation one (Kempton and Withers, 1992). Moreover, transglycosylation is predicted to have a transition state more advanced on the reaction coordinate than the hydrolysis one

(RC values of −0.63, −0.17, and −0.20 Å for H, T_I, and T_{II} TS structures, respectively).

In an attempt to determine the robustness of the geometrical differences found between the hydrolysis and transglycosylation TSs, the corresponding free energy profiles at the QM(SCC-DFTB)/MM level were obtained by umbrella sampling molecular dynamics simulations (**Supplementary Figures S5, S6**). Both reactions show very similar free energy barriers (21.10 and 21.93 kcal/mol for H and T, respectively), being the transglycosylation one slightly higher. At the corresponding TSs, the average $C1_{FUC}\text{-}O_{Acc}$, $C1_{FUC}\text{-}OE2_{Glu338}$ and $H_{Acc}\text{-}OE2_{Glu164}$ distances are 1.95 ± 0.15, 2.60 ± 0.18, and 1.52 ± 0.26 Å for hydrolysis, and 2.11 ± 0.05, 2.31 ± 0.08, and 1.32 ± 0.03 Å for transglycosylation. Thus, the dissociative character of the TSs is maintained, as well as the late proton transfer, although at this level of theory the hydrolysis TS would be more advanced on the reaction coordinate (RC_H
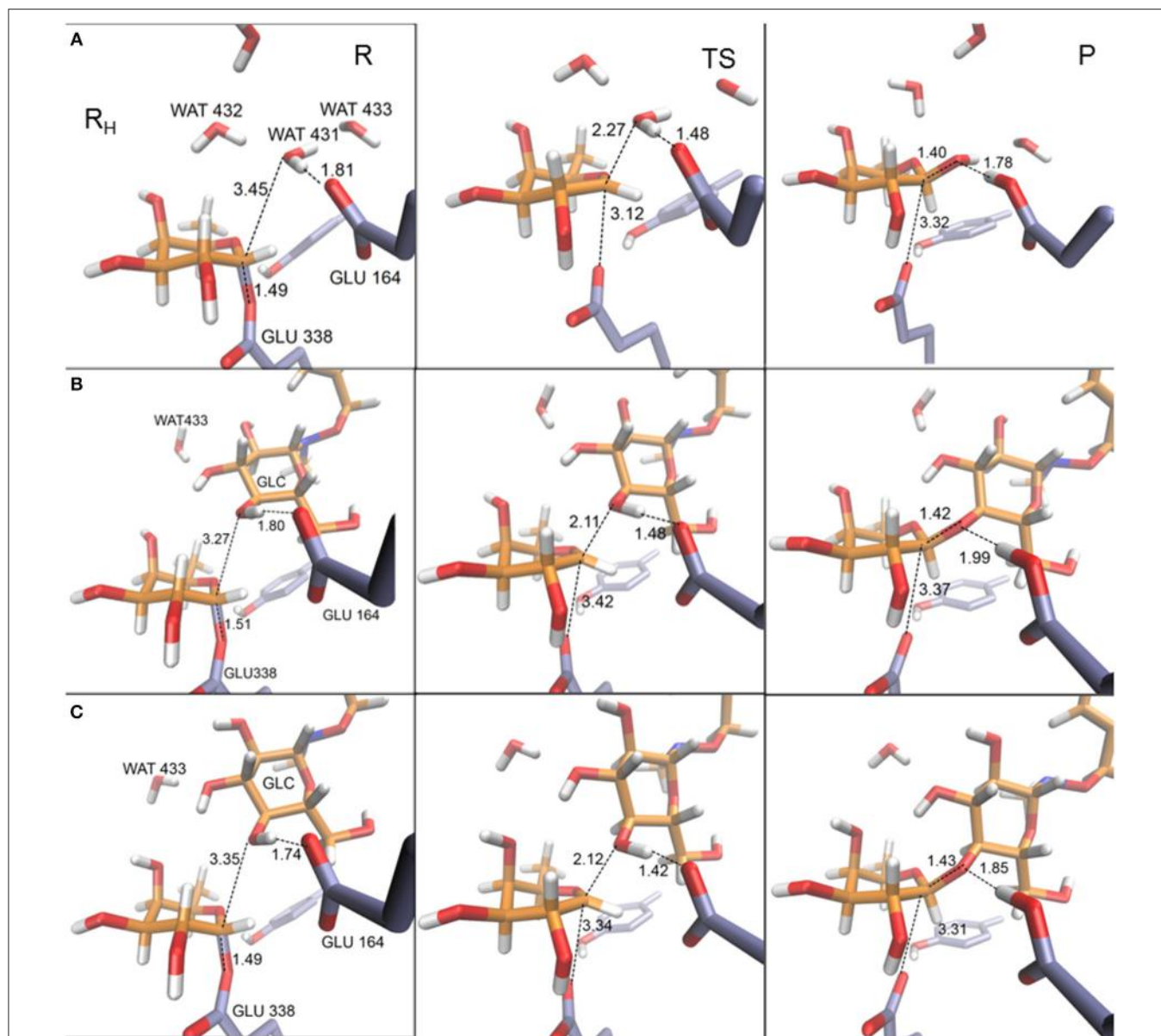
FIGURE 6 | Representative structures for reactant (R), transition state (TS), and product (P) for hydrolysis **(A)** and transglycosylation reactions (frame T_I **(B)** and frame T_{II} **(C)**. The depicted geometries were calculated within the QM(large)/MM partition and with the QM region described at the PBE0/TZVP level. Distances are in Å.

= −0.87 Å) than the transglycosylation one (RC_T = −1.04 Å). However, it is important to note the significant fluctuations on the distances measured for the hydrolysis TS, which may be due to the flatness of the free energy profile at this zone and are indicative of a wide ensemble of available TS configurations with different reaction coordinate values. For the transglycosylation reaction, these fluctuations are much smaller. Interestingly, the transglycosylation QM(SCC-DFTB)/MM potential energy maximum is more advanced than the hydrolysis one (as with QM=PBE0/TZVP), which suggests that the apparent switch in TS localization along the reaction coordinate is a consequence of the introduction of dynamics. On the other hand, the more compact TS geometries when compared to the QM=PBE0/TZVP

ones (sum of $C1_{FUC}$-O_{Acc} and $C1_{FUC}$-$OE2_{GLU338}$ distances) is already present at the potential energy maxima, indicating that it is likely a consequence of the choice of QM level and not to the umbrella sampling simulations. QM(SCC-DFTB)/MM umbrella sampling simulations have been reported for hydrolysis and transglycosylation of Glc substrates catalyzed by a family GH3 β-glucosidase. Transglycosylation TS presented $C1_{GLC}$-O_{Acc}, $C1_{GLC}$-O_{Nuc} and H_{Acc}-O_{Base} distances of 3.0 ± 0.2, 1.9 ± 0.1, and 1.6 ± 0.1 Å, which are relatively similar to those obtained here at the QM=PBE0/TZVP level. However, the hydrolysis TS had the $C1_{GLC}$-O_{Acc} bond practically formed (1.5 ± 0.04 Å) and the proton at a distance of 2.06 ± 0.08 Å from the catalytic base (although right before the TS this
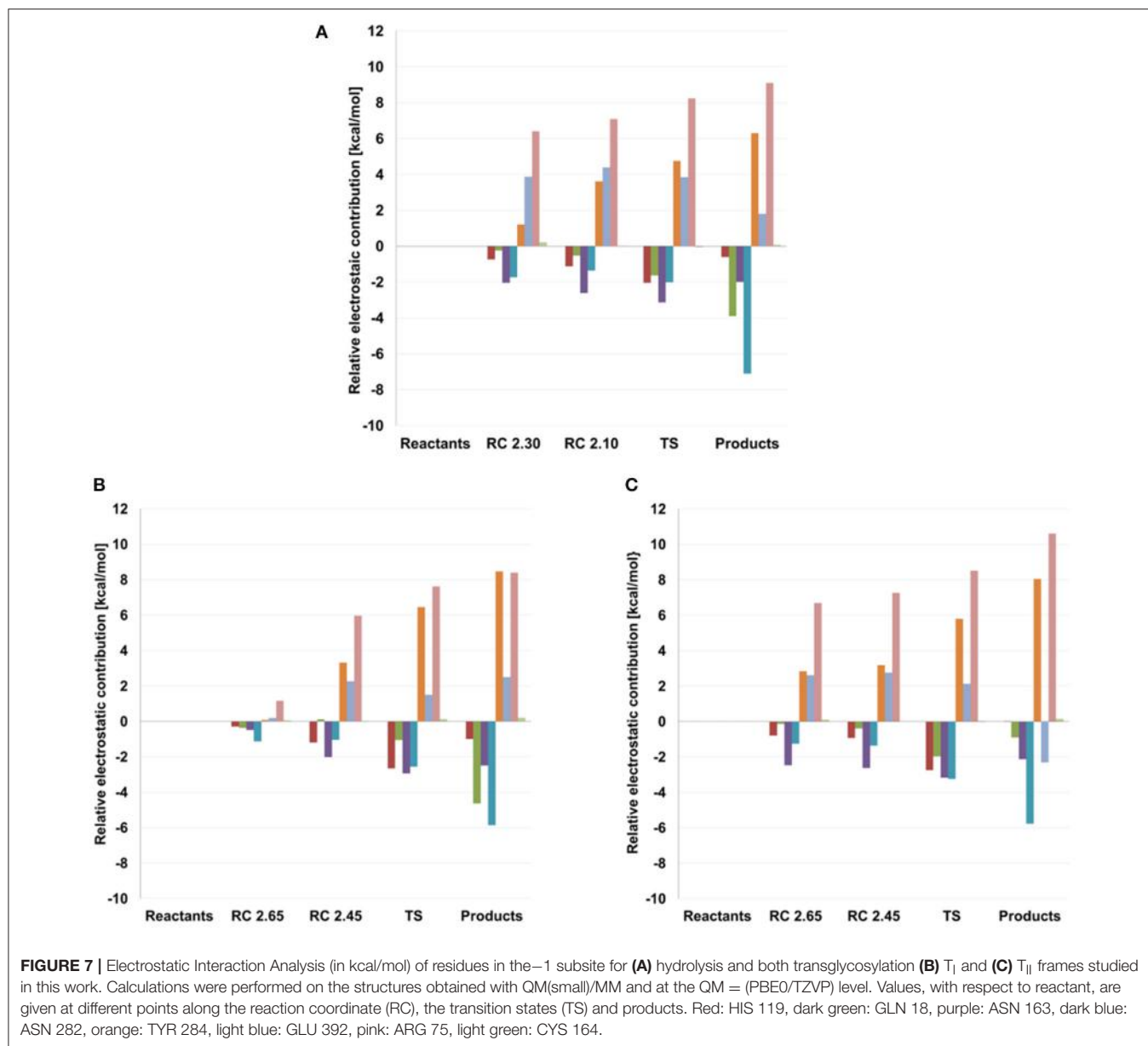
distance had shortened to 1.4–1.5 Å) (Geronimo et al., 2018). Therefore, water was predicted not to be activated by the basic residue, which is not what we observe in our calculations. The authors also related these geometric parameters to the lack of oxocarbenium ion-like character of this TS and a preference of the GH3 −1 subsite for a TS in a $^4C_1$ conformation. (Geronimo et al., 2018). In the present system, the $C1_{FUC}$-$O5_{FUC}$ bond becomes shorter from reactants to the corresponding transition states accordingly to its partial double bond character along the nucleophilic attack. Hence, the positive charge on $C1_{FUC}$ increases in between 0.24 and 0.27 au from reactants to the transition state structures (**Supplementary Table S9** for QM(PBE0/TZVP)/MM and QM(large)/MM) while $O5_{FUC}$ in the fucose ring becomes less negative (around 0.06 au) (with the QM(small)/MM partition the change on $C1_{FUC}$ is slightly bigger (0.29–0.30 au), **Supplementary Table S10**). Those charge variations confirm the oxocarbenium ion-like character observed in deglycosylation reactions (hydrolysis as well as transglycosylation) in the GH1 β-glucosidases family. The ring puckering is very similar in both reactions, with Cremer-Pople ϕ values of 41° (for H), 44 and 47° (for T_I and T_{II}) and ϕ angles of 224° (H), 224, and 221° (for T_I and T_{II}), which correspond to a conformation between $^4H_3$ and ^4E. For the transglycosylation TS in family GH3 β-glucosidase mentioned above, a $^4H_3$ conformation was predicted. (Geronimo et al., 2018).

It can be observed from the QM(PBE0/TZVP)/MM data in **Supplementary Table S9** that there are only very small differences on the atomic charges of the atoms involved in the hydrolysis and transglycosylation reactions, in accordance with the small differences obtained in the energy barriers for the two kinds of deglycosylation processes. The lower energy barrier for hydrolysis might be due then to the more negative charge on the oxygen atom of the nucleophilic water in comparison to the negative charge value on the O4 atom of the glucose acceptor molecule.

As can be seen in **Figure 6** and **Table 4**, some interactions between the substrates or with the surrounding network of water molecules also vary along the two reactions studied. For transglycosylation, the hydrogen of the 3-OH group of the Glc acceptor substrate (adjacent to the attacking 4-OH_{GLC}), in going from reactant to TS, approaches $O5_{FUC}$ (by 0.53 and 0.38 Å for T_I and T_{II}, respectively) and $O4_{FUC}$ (by 0.12 and 0.51 Å for T_I and T_{II}, respectively). In the hydrolysis reaction, a water molecule (WAT432) establishes a bridge between the oxygen atom of the attacking water (WAT431) and that of the 4-OH_{FUC} group. The O_{WAT431}-$H2_{WAT432}$ and the $H1_{WAT432}$-$O4_{FUC}$ distances show little variation from reactant to TS, but at the product the O_{WAT431}-$H2_{WAT432}$ distance increases by 0.91 Å due to a rotation of WAT432 (as O_{WAT431} has formed the new bond with $C1_{FUC}$), which now points its $H2_{WAT432}$ toward $O5_{FUC}$ (d($H2_{WAT432}$-$O5_{FUC}$) is 3.58 Å at reactants, 3.33 Å at TS and 1.95 Å at products). Thus, WAT432 and 3-OH_{GLC} seem to share a similar role in stabilizing the hydrolysis and transglycosylation TS, respectively, *via* interaction with the oxocarbenium ion. This kind of interactions have been identified in retaining glycosyltransferases to help the reaction, with the difference that due to the different disposition of the substrates they serve leaving group departure (Gómez

et al., 2015). It is interesting to notice that the hydrogen atom of this 4-OH_{FUC} group is in turn interacting with Glu392; the $H4O_{FUC}$-$OE1_{GLU392}$ interaction becomes shorter along the hydrolysis and the T_I paths (it is not present along the T_{II} energy profile because a water molecule is located as a H-bonding bridge between $O4_{FUC}$ and $OE1_{GLU392}$). Moreover, the presence of Fuc instead of Glc as substrate (differing on the 4-hydroxyl configuration) has provoked a rotation of the Glu392 sidechain. The analysis of electrostatic contributions to the potential energy barrier of residues at the −1 subsite (**Figure 7**) show that Glu392 is stabilizing both the hydrolysis and transglycosylation TSs, being the effect larger for hydrolysis. In fact, this residue is the one showing the largest difference between the two competitive reactions, suggesting that altering this interaction could be a way to increase the T/H ratio. This observation could be related to the reported transglycosylation increase in the Asn390Ile and Phe401Ser mutants, both residues in contact with Glu392; Asn390 is H-bonding to Glu392 so that its mutation is likely to alter the Glu392-Fuc interaction. Three other residues interact with Fuc hydroxyl groups: In this case, an according to the electrostatic interactions analysis, they provide a small destabilization effect in this deglycosylation steps: Asn163, which interacts with the 2-OH_{FUC} group, with similar values for hydrolysis and transglycosylation; His119, that interacts with $O3_{FUC}$ and $O2_{FUC}$, and provides a slightly larger destabilization effect for transglycosylation; and Gln18, that interacts with $O3_{FUC}$ with similar effect for both TSs.

In the hydrolysis and transglycosylation steps here studied, the nucleophile Glu338 presents strong interactions with the 2-OH_{FUC} group, Tyr284, Arg75, and Asn282. The $H2O_{FUC}$-$OE1_{GLU338}$ distance increases as the Glu338 residue moves away with the breakage of the $C1_{FUC}$-$OE2_{GLU338}$, although the H-bond interaction is maintained along the three reaction pathways and the negative charge of Glu338 also increases (around 0.3 au from reactants to the transition states, **Supplementary Table S9**). Concomitantly with those molecular changes, Tyr384 side-chain moves away from $O5_{FUC}$ while it approaches $OE2_{GLU338}$ atom (which, as mentioned, is increasing its negative charge). It has been suggested that the short hydrogen bond with the hydroxyl group of the equivalent tyrosine residue of a β-galactosidase could facilitate the elimination of the nucleophile protein residue (Bráa et al., 2010). The analyses depicted in **Figure 7** show this increasing stabilization by Tyr284 as reaction proceeds (by 4.76 and 6.12 kcal/mol at the hydrolysis and transglycosylation TSs and 6.31 and 8.26 kcal/mol at their respective products). The electrostatic stabilization of the TSs as compared to the reactants provided by Arg75, that tends to slightly approach the $OE1_{GLU338}$ as reaction proceeds, is the largest one (by 8.24 and 8.06 kcal/mol for hydrolysis and transglycosylation, respectively). These results qualitatively agree with the much lower activity measured when these two residues are mutated. Finally, Asn282 is predicted to have a moderate destabilizing effect, especially for the products. This residue interacts with both catalytic Glu338 and Glu164, thus the results indicate that its interaction with Glu164, which will go from a total −1 au charge to neutral, is the predominant one.

FIGURE 7 | Electrostatic Interaction Analysis (in kcal/mol) of residues in the−1 subsite for **(A)** hydrolysis and both transglycosylation **(B)** T$_I$ and **(C)** T$_{II}$ frames studied in this work. Calculations were performed on the structures obtained with QM(small)/MM and at the QM = (PBE0/TZVP) level. Values, with respect to reactant, are given at different points along the reaction coordinate (RC), the transition states (TS) and products. Red: HIS 119, dark green: GLN 18, purple: ASN 163, dark blue: ASN 282, orange: TYR 284, light blue: GLU 392, pink: ARG 75, light green: CYS 164.

The residues that directly interact with the nucleophile (Tyr284, Asn282, Arg75) or with the acid/base catalyst (Asn163; which also interacts with O2$_{FUC}$ and Arg75) are part of the ones identified to increase the T/H ratio when properly mutated (Teze et al., 2014) Bissaro et al. recently proposed that mutations impairing the optimized electron displacement system that the enzyme has for catalyzing hydrolysis (which includes the nucleophile and the oxocarbenium TS), results in the destabilization of the deglycosylation TSs and leads to accumulation of the glycosyl-enzyme intermediate; such accumulation has been linked to an increased T/H ratio (Bissaro et al., 2015a). Hydrophobic interactions (e.g., C-H – π interactions) augmenting the affinity of the enzyme for glycoside-like acceptors is another factor that can favor transglycosylation.

(Tran et al., 2010; Bissaro et al., 2015a,b). In the present system the acceptor substrate can interact with Trp312. On the other hand, mutation of the catalytic acid/base Glu to Asp in NkBgl also catalyzed unforeseen transglycosylation (Jeng et al., 2012). The crystal structure of the mutant revealed a water molecule close to the shorter Asp chain and the anomeric carbon; thus, it was proposed that Asp would play its role through this nearby water molecule. A structurally conserved water molecule observed in a wild-type enzyme belonging to family GH101, was also proposed to enable a Grotthuss proton shuttle between the protein residue carboxylate and the glycosidic oxygen (Gregg et al., 2015). Still, the particular reasons why all these mutations in the −1 subsite favor transglycosylation over hydrolysis have not been fully unveiled.

The small differences in energy barriers these mutagenesis experiments imply and the high dimensionality of enzymatic systems, make it a great challenge to reproduce the effect of mutation computationally or to clearly identify its origin. For example, in the case of Tyr284 a transglycosylation yield of 76 % (but a lower rate) was measured for the Tyr284Phe mutant as compared to a 36 % for the WT enzyme (Teze et al., 2014). This was interpreted as a destabilizing effect produced by the mutation that affected more hydrolysis than transglycosylation. Our qualitative analysis of electrostatic interactions, though, estimates that in WT Ttβ-gly Tyr284 provides a slightly higher stabilization in the transglycosylation case. Attempts to reproduce the experimental observations by directly introducing the Tyr284Phe mutation on frames H, T_I, and T_{II} and recalculating the potential energy profiles also failed. Thus, further computational studies of mutants taking into account structural rearrangements upon mutation (even if small) would be needed to try to reproduce these trends. However, and more importantly, these analyses are not able to capture the subtle changes in energy barriers involved in these cases, which are below the current "chemical accuracy."

CONCLUSIONS

Here we have presented a QM(DFT)/MM study of the glycosylation, hydrolysis and transglycosylation steps catalyzed by wild type *Thermus thermophilus* β-glycosidase (family GH1), a retaining glycosyl hydrolase for which a transglycosylation yield of 36 % has been determined experimentally for pNP-Fuc and BnON(Me)-Glc substrates.

In the glycosylation step (first step of the double-displacement mechanism of retaining GHs), the pNP-Fuc substrate establishes a strong network of hydrogen bonds at the−1 subsite and is found to be bound in an only slightly distorted $^4C_1$ conformation, contrary to what has been reported when the substrate is a Glc disaccharide. At the transition state for this step, that has a strong oxocarbenium character (close to $^4H_3$ conformation but with some ^4E characteristics), proton transfer from Glu164 to the leaving group has not yet started but it takes place at a OE2$_{GLU338}$-C1$_{FUC}$ distance of ∼2 Å with the QM(large)/MM partition and the PBE0/TZVP QM level. The potential energy barrier at this level of theory (20.4–22.0 kcal/mol) is in qualitative agreement with the experimental barrier (17.1 kcal/mol). The use of a smaller QM region increases this value by more than 5 kcal/mol. The size of the QM region also changes the location of the TS along the reaction coordinate. Thus, it seems to affect the nucleophilic and acid/base characteristics of Glu338 and Glu164, respectively.

For the hydrolysis and transglycosylation reactions (competing reactions in the deglycosylation step), a significant reduction of the barrier heights is also observed with the QM(large)/MM partition, although the T vs. H trends are well reproduced with the QM(small)/MM one provided that the PBE0/TZVP level is used to obtain the energies (or also for geometry optimization). The transglycosylation potential energy barrier is predicted to be 1.2–2.6 kcal/mol higher than the hydrolysis one. Hydrolysis and transglycosylation transition states are both very loose, with strong oxocarbenium-like character and the proton not yet transferred. Their ring conformation falls between $^4H_3$ and ^4E and they present very similar atomic charges, except for the more negative charge of the oxygen atom of the attacking water when compared to that of the attacking 4-OH group of Glc. Structural differences appear between the glycosylation and the deglycosylation steps, with the former having a more pronounced S_N2 character. The proton transfer is also more advanced in the deglycosylation TSs that was the proton transfer in the glycosylation one. Differences between the hydrolysis and the transglycosylation TSs are also observed, but they are very dependent on the level of theory used. At the QM(PBE0/TZVP)/MM level, the potential energy maxima for transglycosylation is more advanced on the reaction coordinate than the hydrolysis one: has shorter O4$_{GLC}$-C1$_{FUC}$ (forming bond) distance and longer OE2$_{GLU338}$-C1$_{FUC}$ (breaking) distance, although the H$_{Acc}$ proton is closer to the Glu164 base in hydrolysis. QM(SCC-DFTB)/MM free energy maxima show the inverted situation, although large geometric fluctuations are seen for the hydrolysis TS, indicating a large ensemble of TS configurations covering a range of reaction coordinate values. Interestingly, some interactions between the substrates (3-OH$_{GLC}$ group) or with neighboring water molecules (WAT432) are identified that stabilize the oxocarbenium TSs through interaction with O5$_{FUC}$ and O4$_{FUC}$. An analysis of electrostatic interactions of the substrates with the residues at the−1 subsite has also been performed. However, correlation with experimental mutagenesis studies found to alter the T/H ratio is very challenging due to the small differences on energy barriers involved and potential structural (even if small) rearrangements upon mutation. The analysis suggests, though, that perturbing the Glu392-Fuc interaction could increase the T/H ratio, even by direct mutation of this residue or indirectly as reported experimentally in the N390I and F401S cases.

AUTHOR CONTRIBUTIONS

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2019.00200/full#supplementary-material

REFERENCES

Adamo, C., and Barone, V. (1999). Toward reliable density functional methods without adjustable parameters: the PBE0 model. *J. Chem. Phys.* 110, 6158–6170. doi: 10.1063/1.478522

Ahlrichs, R., Bär, M., Häser, M., Horn, H., and Kölmel, C. (1989). Electronic structure calculations on workstation computers: the program system turbomole. *Chem. Phys. Lett.* 162, 165–169. doi: 10.1016/0009-2614(89)85118-8

Albesa-Jové, D., Mendoza, F., Rodrigo-Unzueta, A., Gomollón-Bel, F., Cifuente, J. O., Urresti, S., et al. (2015). A native ternary complex trapped in a crystal reveals the catalytic mechanism of a retaining glycosyltransferase. *Angew. Chemie Int. Ed.* 54, 9898–9902. doi: 10.1002/anie.201504617

Ardèvol, A., and Rovira, C. (2015). Reaction mechanisms in carbohydrate-active enzymes: glycoside hydrolases and glycosyltransferases. Insights from ab Initio quantum mechanics/molecular mechanics dynamic simulations. *J. Am. Chem. Soc.* 137, 7528–7547. doi: 10.1021/jacs.5b01156

Badieyan, S., Bevan, D. R., and Zhang, C. (2012). Probing the active site chemistry of β-glucosidases along the hydrolysis reaction pathway. *Biochemistry* 51, 8907–8918. doi: 10.1021/bi300675x

Bakowies, D., and Thiel, W. (1996). Hybrid models for combined quantum mechanical and molecular mechanical approaches. *J. Phys. Chem.* 100, 10580–10594. doi: 10.1021/jp9536514

Berendsen, H. J. C., Postma, J. P. M., Van Gunsteren, W. F., Dinola, A., and Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81, 3684–3690. doi: 10.1063/1.448118

Biarnés, X., Ardèvol, A., Iglesias-Fernández, J., Planas, A., and Rovira, C. (2011). Catalytic itinerary in 1,3-1,4-β-glucanase unraveled by QM/MM metadynamics. Charge is not yet fully developed at the oxocarbenium ion-like transition state. *J. Am. Chem. Soc.* 133, 20301–20309. doi: 10.1021/ja207113e

Billeter, S. R., Turner, A. J., and Thiel, W. (2000). Linear scaling geometry optimisation and transition state search in hybrid delocalised internal coordinates. *Phys. Chem. Chem. Phys.* 2, 2177–2186. doi: 10.1039/a909486e

Bissaro, B., Durand, J., Biarnés, X., Planas, A., Monsan, P., O'Donohue, M. J., et al. (2015a). Molecular design of non-leloir furanose-transferring enzymes from an α- l -arabinofuranosidase: a rationale for the engineering of evolved transglycosylases. *ACS Catal.* 5, 4598–4611. doi: 10.1021/acscatal.5b00949

Bissaro, B., Monsan, P., Fauré, R., and O'Donohue, M. J. (2015b). Glycosynthesis in a waterworld: new insight into the molecular basis of transglycosylation in retaining glycoside hydrolases. *Biochem. J.* 467, 17–35. doi: 10.1042/BJ20141412

Brás, N. F., Fernandes, P. A., and Ramos, M. J. (2010). QM/MM studies on the β-galactosidase catalytic mechanism: hydrolysis and transglycosylation reactions. *J. Chem. Theory Comput.* 6, 421–433. doi: 10.1021/ct900530f

Camargo, H. A., Caputa, P., Das, D., Heller, M. P., and Jefferson, R. (2019). Complexity as a novel probe of quantum quenches: universal scalings and purifications. *Phys. Rev. Lett.* 122:081601. doi: 10.1103/PhysRevLett.122.081601

Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* 37, D233–D238. doi: 10.1093/nar/gkn663

Case, D. A., Cheatham, T. E. III., Darden, T., Gohlke, H., Luo, R., Merz, K. M. Jr, et al. (2005). The Amber biomolecular simulation programs. *J. Comput. Chem.* 26, 1668–1688. doi: 10.1002/jcc.20290

Chuenchor, W., Pengthaisong, S., Robinson, R. C., Yuvaniyama, J., Svasti, J., and Ketudat Cairns, J. R. (2011). The structural basis of oligosaccharide binding by rice BGlu1 beta-glucosidase. *J.Struct.Biol.* 173, 169–179. doi: 10.1016/j.jsb.2010.09.021

Claeyssens, F., Harvey, J. N., Manby, F. R., Mata, R. A., Mulholland, A. J., Ranaghan, K. E., et al. (2006). High-accuracy computation of reaction barriers in enzymes. *Angew. Chemie Int. Ed.* 45, 6856–6859. doi: 10.1002/anie.200602711

Cobucci-Ponzano, B., and Moracci, M. (2012). Glycosynthases as tools for the production of glycan analogs of natural products. *Nat. Prod. Rep.* 29:697. doi: 10.1039/c2np20032e

Cobucci-Ponzano, B., Strazzulli, A., Rossi, M., and Moracci, M. (2011). Glycosynthases in biocatalysis. *Adv. Synth. Catal.* 353, 2284–2300. doi: 10.1002/adsc.201100461

Cremer, D., and Pople, J. A. (1975). General definition of ring puckering coordinates. *J. Am. Chem. Soc.* 97, 1354–1358. doi: 10.1021/ja00839a011

Danby, P. M., and Withers, S. G. (2016). Advances in enzymatic glycoside synthesis. *ACS Chem. Biol.* 11, 1784–1794. doi: 10.1021/acschembio.6b00340

David, B., Irague, R., Jouanneau, D., Daligault, F., Czjzek, M., Sanejouand, Y.-H., et al. (2017). Internal water dynamics control the transglycosylation/hydrolysis balance in the agarase (AgaD) of *Zobellia galactanivorans*. *ACS Catal.* 7, 3357–3367. doi: 10.1021/acscatal.7b00348

Davies, G. J., Planas, A., and Rovira, C. (2012). Conformational analyses of the reaction coordinate of glycosidases. *Acc. Chem. Res.* 45, 308–316. doi: 10.1021/ar2001765

Dion, M., Fourage, L., Hallet, J.-N., and Colas, B. (1999). Cloning and expression of a β-glycosidase gene from Thermus thermophilus. Sequence and biochemical characterization of the encoded enzyme. *Glycoconj. J.* 16, 27–37. doi: 10.1023/A:1006997602727

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995). A smooth particle mesh Ewald method. *J. Chem. Phys.* 103, 8577–8593. doi: 10.1063/1.470117

Faijes, M., and Planas, A. (2007). *In vitro* synthesis of artificial polysaccharides by glycosidases and glycosynthases. *Carbohydr. Res.* 342, 1581–1594. doi: 10.1016/j.carres.2007.06.015

Feng, H.-Y., Drone, J., Hoffmann, L., Tran, V., Tellier, C., Rabiller, C., et al. (2005). Converting a β-Glycosidase into a β-Transglycosidase by directed evolution. *J. Biol. Chem.* 280, 37088–37097. doi: 10.1074/jbc.M502873200

Field, R. A. (2011). Challenging reaction equilibria. *Nat. Chem. Biol.* 7, 658–659. doi: 10.1038/nchembio.668

Geronimo, I., Payne, C. M., and Sandgren, M. (2018). Hydrolysis and transglycosylation transition states of glycoside hydrolase family 3 β-glucosidases differ in charge and puckering conformation. *J. Phys. Chem. B* 122, 9452–9459. doi: 10.1021/acs.jpcb.8b07118

Gómez, H., Lluch, J. M., and Masgrau, L. (2013). Substrate-assisted and nucleophilically assisted catalysis in bovine α1,3-galactosyltransferase. mechanistic implications for retaining glycosyltransferases. *J. Am. Chem. Soc.* 135, 7053–7063. doi: 10.1021/ja4024447

Gómez, H., Mendoza, F., Lluch, J. M., and Masgrau, L. (2015). QM/MM studies reveal how substrate–substrate and enzyme–substrate interactions modulate retaining glycosyltransferases catalysis and mechanism. *Adv. Protein Chem. Struct. Biol.* 100, 225–254. doi: 10.1016/bs.apcsb.2015.06.004

Gómez, H., Polyak, I., Thiel, W., Lluch, J. M., and Masgrau, L. (2012). Retaining glycosyltransferase mechanism studied by QM/MM methods: lipopolysaccharyl-α-1,4-galactosyltransferase C transfers α-galactose via an oxocarbenium ion-like transition state. *J. Am. Chem. Soc.* 134, 4743–4752. doi: 10.1021/ja210490f

Götz, A. W., Williamson, M. J., Xu, D., Poole, D., Le Grand, S., and Walker, R. C. (2012). Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. generalized born. *J. Chem. Theory Comput.* 8, 1542–1555. doi: 10.1021/ct200909j

Gregg, K. J., Suits, M. D. L., Deng, L., Vocadlo, D. J., and Boraston, A. B. (2015). Structural analysis of a family 101 glycoside hydrolase in complex with carbohydrates reveals insights into its mechanism. *J. Biol. Chem.* 290, 25657–25669. doi: 10.1074/jbc.M115.680470

Hart, G. W., and Copeland, R. J. (2010). Glycomics hits the big time. *Cell* 143, 672–676. doi: 10.1016/j.cell.2010.11.008

Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* 280(Pt. 2), 309–316. doi: 10.1042/bj2800309

Hoover, W. G. (1985). Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A* 31, 1695–1697. doi: 10.1103/PhysRevA.31.1695

Hrmova, M., Varghese, J. N., De Gori, R., Smith, B. J., Driguez, H., and Fincher, G. B. (2001). Catalytic mechanisms and reaction intermediates along the hydrolytic pathway of a plant β-D-glucan glucohydrolase. *Structure* 9, 1005–1016. doi: 10.1016/S0969-2126(01)00673-6

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* 14, 33–38. doi: 10.1016/0263-7855(96)00018-5

Jeng, W.-Y., Wang, N.-C., Lin, C.-T., Chang, W.-J., Liu, C.-I., and Wang, A. H.-J. (2012). High-resolution structures of Neotermes koshunensis β-glucosidase mutants provide insights into the catalytic mechanism and the synthesis of glucoconjugates. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 68, 829–838. doi: 10.1107/S0907444912013224

Kästner, J., and Thiel, W. (2005). Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: "Umbrella integration." *J. Chem. Phys.* 123:144104. doi: 10.1063/1.2052648

Kästner, J., and Thiel, W. (2006). Analysis of the statistical error in umbrella sampling simulations by umbrella integration. *J. Chem. Phys.* 124:234106. doi: 10.1063/1.2206775

Kempton, J. B., and Withers, S. G. (1992). Mechanism of Agrobacterium.beta.-glucosidase: kinetic studies. *Biochemistry* 31, 9961–9969. doi: 10.1021/bi00156a015

Kiessling, L. L., and Splain, R. A. (2010). Chemical approaches to glycobiology. *Annu. Rev. Biochem.* 79, 619–653. doi: 10.1146/annurev.biochem.77.070606.100917

Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González-Outeiriño, J., Daniels, C. R., Foley, B. L., et al. (2008). GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.* 29, 622–655. doi: 10.1002/jcc.20820

Liu, D. C., and Nocedal, J. (1989). On the limited memory BFGS method for large scale optimization. *Math. Program.* 45, 503–528. doi: 10.1007/BF01589116

Lokanath, N. K., Shiromizu, I., Miyano, M., Yokoyama, S., Kuramitsu, S., and Kunishima, N. (in press). Structure of Beta-Glucosidase at Atomic Resolution from Thermus Thermophilus *Hb8*.

Mackenzie, L. F., Wang, Q., Warren, R. A. J., and Withers, S. G. (1998). Glycosynthases: mutant glycosidases for oligosaccharide synthesis. *J. Am. Chem. Soc.* 120, 5583–5584. doi: 10.1021/ja980833d

Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015). ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* 11, 3696–3713. doi: 10.1021/acs.jctc.5b00255

Malet, C., and Planas, A. (1998). From β-glucanase to β-glucansynthase: glycosyl transfer to α-glycosyl fluorides catalyzed by a mutant endoglucanase lacking its catalytic nucleophile. *FEBS Lett.* 440, 208–212. doi: 10.1016/S0014-5793(98)01448-3

Mendoza, F., Gómez, H., Lluch, J. M., and Masgrau, L. (2016). α1,4-N-Acetylhexosaminyltransferase EXTL2: the missing link for understanding glycosidic bond biosynthesis with retention of configuration. *ACS Catal.* 6, 2577–2589. doi: 10.1021/acscatal.5b02945

Mendoza, F., Lluch, J. M., and Masgrau, L. (2017). Computational insights into active site shaping for substrate and regiospecificity in the EXTL2 retaining glycosyltransferase. *Organic Biomol. Chem.* doi: 10.1039/C7OB01937H

Metz, S., Kästner, J., Sokol, A. A., Keal, T. W., and Sherwood, P. (2014). ChemShell—a modular software package for QM/MM simulations. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 4, 101–110. doi: 10.1002/wcms.1163

Moracci, M., Trincone, A., Perugino, G., Ciaramella, M., and Rossi, M. (1998). Restoration of the activity of active-site mutants of the hyperthermophilic β-glycosidase from sulfolobus solfataricus: dependence on the mechanism on the action of external nucleophiles. *Biochemistry* 37, 17262–17270. doi: 10.1021/bi981855f

Nosé, S. (1984). A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* 81, 511–519. doi: 10.1063/1.447334

Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M., and Jensen, J. H. (2011). PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. *J. Chem. Theory Comput.* 7, 525–537. doi: 10.1021/ct100578z

Pereira, A. T., Ribeiro, A. J. M. M., Fernandes, P. A., and Ramos, M. J. (2017). Benchmarking of density functionals for the kinetics and thermodynamics of the hydrolysis of glycosidic bonds catalyzed by glycosidases. *Int. J. Quantum Chem.* 117:e25409. doi: 10.1002/qua.25409

Petersen, L., Arde, A., Rovira, C., and Reilly, P. J. (2010). Molecular mechanism of the glycosylation step catalyzed by golgi. *J. Am. Chem. Soc.* 132, 8291–8300. doi: 10.1021/ja909249u

Petersen, L., Ardèvol, A., Rovira, C., and Reilly, P. J. (2009). Mechanism of cellulose hydrolysis by inverting GH8 endoglucanases: A QM/MM Metadynamics Study. *J. Phys. Chem. B* 113, 7331–7339. doi: 10.1021/jp811470d

Planas, A., and Faijes, M. (2002). Glycosidases and Glycosynthases in enzymatic synthesis of oligosaccharides. An overview. *Afinidad* 59, 295–313.

Plou, F. J., de Segura, A. G., and Ballesteros, A. (2007). "Application of Glycosidases and Transglycosidases in the Synthesis of Oligosaccharides," in *Industrial Enzymes* (Dordrecht: Springer), 141–157.

Raich, L., Borodkin, V., Fang, W., Castro-López, J., Van Aalten, D. M. F., Hurtado-Guerrero, R., et al. (2016). A trapped covalent intermediate of a glycoside hydrolase on the pathway to transglycosylation. insights from experiments and quantum mechanics/molecular mechanics simulations. *J. Am. Chem. Soc.* 138, 3325–3332. doi: 10.1021/jacs.5b10092

Reed, A. E., Weinstock, R. B., and Weinhold, F. (1985). Natural population analysis. *J. Chem. Phys.* 83, 735–746. doi: 10.1063/1.449486

Roe, D. R., and Cheatham, T. E. (2013). PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* 9, 3084–3095. doi: 10.1021/ct400341p

Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* 23, 327–341. doi: 10.1016/0021-9991(77)90098-5

Sajith, S., Priji, P., Sreedevi, S., and Benjamin, S. (2016). An overview on fungal cellulases with an industrial perspective. *J. Nutr. Food Sci.* 6:461. doi: 10.4172/2155-9600.1000461

Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S., and Walker, R. C. (2013). Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh ewald. *J. Chem. Theory Comput.* 9, 3878–3888. doi: 10.1021/ct400314y

Saumonneau, A., Champion, E., Peltier-Pain, P., Molnar-Gabor, D., Hendrickx, J., Tran, V., et al. (2016). Design of an α-l-transfucosidase for the synthesis of fucosylated HMOs. *Glycobiology* 26, 261–269. doi: 10.1093/glycob/cwv099

Seeberger, P. H., and Cummings, R. D. (2017). "Glycans in Biotechnology and the Pharmaceutical Industry," in *Essentials of Glycobiology,* eds E. A. Varki, Cummings, R. D., Esko, J. D., et al. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Sherwood, P., de Vries, A. H., Guest, M. F., Schreckenbach, G., Catlow, C. R. A., French, S. A., et al. (2003). QUASI: A general purpose implementation of the QM/MM approach and its application to problems in catalysis. *J. Mol. Struct. Theochem* 632, 1–28. doi: 10.1016/S0166-1280(03)00285-9

Smith, W., and Forester, T. R. (1996). DL_POLY_2.0: a general-purpose parallel molecular dynamics simulation package. *J. Mol. Graph.* 14, 136–141. doi: 10.1016/S0263-7855(96)00043-4

Teze, D. (2012). *Recherche des Bases Moléculaires de L'équilibre Transglycosylation / Hydrolyse d'une Glycoside Hydrolase de la Famille* 1. Université de Nantes, Faculté des Sciences et des Techniques. 1–198.

Teze, D., Daligault, F., Ferrières, V., Sanejouand, Y.-H., and Tellier, C. (2015). Semi-rational approach for converting a GH36 α-glycosidase into an α-transglycosidase. *Glycobiology* 25, 420–427. doi: 10.1093/glycob/cwu124

Teze, D., Dion, M., Daligault, F., Tran, V., André-Miral, C., and Tellier, C. (2013). Alkoxyamino glycoside acceptors for the regioselective synthesis of oligosaccharides using glycosynthases and transglycosidases. *Bioorg. Med. Chem. Lett.* 23, 448-−451. doi: 10.1016/j.bmcl.2012.11.065

Teze, D., Hendrickx, J., Czjzek, M., Ropartz, D., Sanejouand, Y.-H., Tran, V., et al. (2014). Semi-rational approach for converting a GH1 β-glycosidase into a β-transglycosidase. *Protein Eng. Des. Sel.* 27, 13–19. doi: 10.1093/protein/gzt057

Tran, V., Hoffmann, L., Rabiller, C., Tellier, C., and Dion, M. (2010). Rational design of a GH1 -glycosidase to prevent self-condensation during the transglycosylation reaction. *Protein Eng. Des. Sel.* 23, 43–49. doi: 10.1093/protein/gzp068

Wang, J., Hou, Q., Dong, L., Liu, Y., and Liu, C. (2011). QM/MM studies on the glycosylation mechanism of rice BGlu1 β-glucosidase. *J. Mol. Graph. Model.* 30, 148–152. doi: 10.1016/j.jmgm.2011.06.012

Wang, J., Hou, Q., Sheng, X., Gao, J., Liu, Y., and Liu, C. (2013). Theoretical study on the deglycosylation mechanism of rice BGlu1 β-glucosidase. *Int. J. Quantum Chem.* 113, 1071–1075. doi: 10.1002/qua.24131

Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004). Development and testing of a general amber force field. *J. Comput. Chem.* 25, 1157–1174. doi: 10.1002/jcc.20035

Zechel, D. L., and Withers, S. G. (2000). Glycosidase mechanisms : anatomy of a finely tuned catalyst. *Acc. Chem. Res.* 33, 11–18. doi: 10.1021/ar970172+

The Effect of Cofactor Binding on the Conformational Plasticity of the Biological Receptors in Artificial Metalloenzymes: The Case Study of LmrR

Lur Alonso-Cotchico[1,2], Jaime Rodríguez-Guerra Pedregal[1], Agustí Lledós[1] and Jean-Didier Maréchal[1]*

[1] Departament de Química, Universitat Autònoma de Barcelona, Barcelona, Spain, [2] Stratingh Institute for Chemistry, University of Groningen, Groningen, Netherlands

The design of Artificial Metalloenzymes (ArMs), which result from the incorporation of organometallic cofactors into biological structures, has grown steadily in the last two decades and important new-to-Nature reactions have been reached. These type of exercises could greatly benefit from an understanding of the structural impact that the inclusion of organometallic moieties may have on the biological host. To date though, our understanding of this phenomenon is highly partial. This lack of knowledge is one of the elements that condition that first-generation ArMs generally display relatively poor catalytic profiles. In this work, we approach this matter by assessing the dynamics and stability of a series of ArMs resulting from the inclusion, via different anchoring strategies, of a variety of organometallic cofactors into the Lactococcal multidrug resistance regulator (LmrR) protein. To this aim, we coupled standard force field-based techniques such as Protein-Ligand Docking and Molecular Dynamics simulations with a variety of trajectory convergence analyses, capable of assessing both the stability and flexibility of the different systems under study upon the binding of cofactors. Together with the experimental evidence obtained in other studies, we provide an overview on how these changes can affect the catalytic outcomes obtained from the different ArMs. Fundamentally, our results show that the convergence analysis used in this work can assess how the inclusion of synthetic metallic cofactors in proteins can condition different structural modulations of their host. Those conformational modifications are key to the success of the desired catalytic activity and their proper identification can be wisely used to improve the quality and the rate of success of the ArMs.

Keywords: molecular modeling, artificial metalloenzymes, molecular dynamics, interactive analysis, cofactor binding, molecular plasticity

INTRODUCTION

Incorporating homogenous catalysts into biological scaffolds (e.g., protein, DNA, or peptides) has become a common strategy to expand the scope of the biological space and produce biocompatible man-made biocatalysts (Schwizer et al., 2017; Diéguez et al., 2018). These *de novo* enzymes, also referred as Artificial Metalloenzymes (ArMs), can be generated under numerous strategies including post-translational approaches i.e., supramolecular (Ohashi et al., 2003; Mahammed and Gross, 2005; Reetz and Jiao, 2006; Bos et al., 2015), covalent (Reetz et al., 2002; Bos et al., 2013), or dative (Kokubo et al., 1983; Van De Velde et al., 2000) interactions, or eventually the incorporation of unnatural amino acids (UAA) (Drienovská et al., 2015, 2017) via sequencing approaches or by the direct expression through cellular vectors. The ArM design process can be divided in two different stages: (1) discovery, when a first catalytically efficient biohybrid shows some activity for a given reaction, and (2) optimization, when the initial candidates are chemically and/or genetically altered to reach improved activity in terms of yield, substrate selectivity, or regiospecificity. Whatever the stage at which the ArM design stands, the most important and complex molecular variable that needs to be controlled by designers is the stability of the interactions between the host and the artificial cofactor: a *sine qua non* condition which reaches pre-reactive resting states and catalytically competent geometries after the binding of the substrate.

Foreseeing the quality of the host-cofactor complementarity requires extensive molecular knowledge and remains a challenging exercise in the path of achieving experimentally efficient candidates. In fact, experimentalists base most of their design on trial-and-error strategies until they reach a first hit. In a way, designers are engaged in an unfair battle against evolution since they try to find good enough affinities between two moieties that never occur in Nature. One of the predominant variables for defining the quality of the interaction between the two entities, is the conformational adaptation of the receptor upon binding of the cofactor. In fact, for any *de novo* design, one of the major weaknesses is the poor consideration of protein dynamics along the designing exercises (Hammes-Schiffer and Benkovic, 2006; Henzler-Wildman and Kern, 2007; Nagel and Klinman, 2009; Hammes et al., 2011; Callender and Dyer, 2014; Maria-Solano et al., 2018). This is probably the reason why the catalytic efficiency of the new candidates is frequently many orders of a magnitude lower than that achieved by naturally-occurring enzymes (Jiang et al., 2008; Röthlisberger et al., 2008; Siegel et al., 2010). Despite the increasing number of ArMs over the last decades, little has been done to estimate the sensitivity of the biological host to the insertion of non-natural cofactors and how this, in turn, conditions the nature of the resting state of the ArM prior to any catalytic step. In this matter, *in silico* methods can be very helpful.

Molecular Modeling has been widely used to decode the nature of the dynamical events involved in structure-function relationship of naturally-occurring biological macromolecules. From short to large scale motions, theoretical (e.g., Molecular Dynamics (MD) simulations and Normal Modes Analysis)
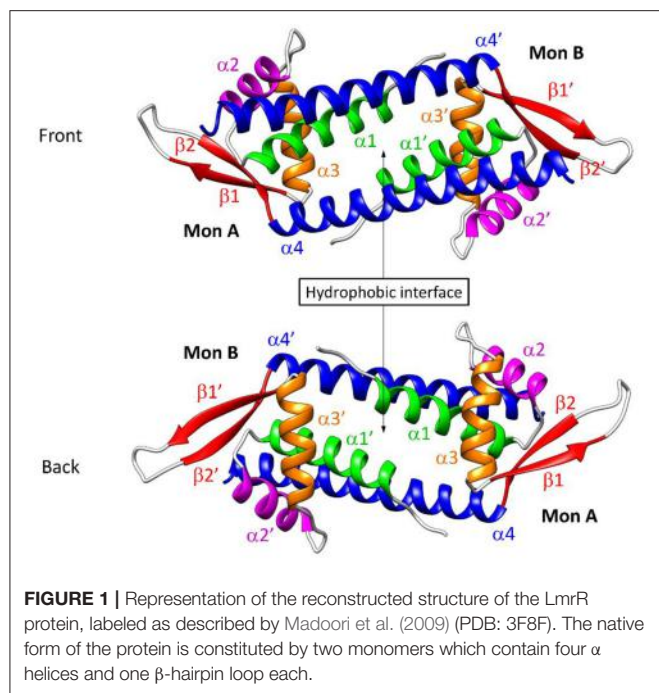
studies constantly provide evidence on how the dynamics of the protein host is influenced by the presence or absence of substrates and/or inhibitors as well as the tight relationship between these changes and the accessibility to catalytically efficient configurations (Dutta and Mishra, 2017; Sen et al., 2017; Sharma et al., 2017; Wilson and Wetmore, 2017; Wilson et al., 2017; Kamariah et al., 2018; Luirink et al., 2018; Rout et al., 2018; Schlee et al., 2018). It is therefore a legitimate question to use computation to assess to what extent bridging chemical and biological entities disturb the natural conformational space of the biomolecules and how damaging/beneficial this can be for catalysis.

Our group entered the field of ArM about a decade ago and focused both on understanding the mechanism of non-natural enzymes as well as providing protocols for the design of new systems. Our strategies are mostly based on integrated protocols, where physical models could range from Quantum Mechanics (QM) to Molecular Mechanics (MM) approaches (Muñoz Robles et al., 2015). One of the questions we wanted to solve is the magnitude of the conformational rearrangement experienced by the biological host under cofactor inclusion, a phenomenon that requires substantial computational improvements of MM methodologies in order to simulate metal-mediated recognition processes. In previous works we decoded the electronic origin of the control of the binding site motions and helix re-arrangements in the ArM constructed by the insertion of salophen into heme-oxygenase apo-enzyme (Muñoz Robles et al., 2011). Another interesting case came from describing how the inclusion of organometallic complexes into a protein scaffold can alter its structure leading to significant variations in the catalytic outcomes (Drienovská et al., 2017; Villarino et al., 2018). These studies provide us with some clues about the structural sensitivity of the receptor upon inclusion of the organometallic moiety, but no clear tendencies could be drawn as they were comprised of a case-specific analysis. To shed light on this matter we will perform a structural assessment benchmark, focusing on a unique receptor loaded with different homogeneous catalysts.

Over the past few years, Roelfes et al. have focused on the Lactococcal multidrug resistance Regulator (LmrR) protein–a transcriptional repressor from the *Lactococcus lactis* organism–as a biological host for a variety of organometallic cofactors, leading to a set of enantioselective artificial metalloenzymes, including hydratases (Bos et al., 2013; Drienovská et al., 2015, 2017), cyclopropanases (Villarino et al., 2018) and Diels-Alderases (Bos et al., 2012). LmrR is a homodimeric protein with a particularly flat and hydrophobic dimer interface, capable of packing foreign aromatic molecules at the patch constituted by the tryptophan's W96/W96' of chains α4/α4', respectively, which are located at the center of the cavity (**Figure 1**).

Most of the new LmrR-based designs are the result of (a) the post-translational inclusion of a phenanthroline moiety (**Figure 2A**) or (b) the expression of the (2,2'-bipyridin-5yl)alanine (BpyA) unnatural amino acid (**Figure 2B**), at positions 89/89' of the dimeric LmrR protein. Both of them are nitrogen-based ligands of copper(II) ions and presented interesting activity for the hydration of ketones as well as the Friedel-Crafts alkylation reaction for either LmrR or DNA based

FIGURE 1 | Representation of the reconstructed structure of the LmrR protein, labeled as described by Madoori et al. (2009) (PDB: 3F8F). The native form of the protein is constituted by two monomers which contain four α helices and one β-hairpin loop each.

catalysis (Arnold, 2009; Boersma et al., 2010; Bos et al., 2013; Drienovská et al., 2015, 2017). Separately, the same protein was used as a scaffold for the supramolecular recognition of hemin, resulting in an ArM with acquired cyclopropanase activity (**Figure 2C**) (Villarino et al., 2018). Recently, it has been employed to covalently attach a Rh(I) complex, forming a biohybrid capable of hydrogenating CO_2 (Laureanti et al., 2019). Those works therefore present a unique opportunity to test how a given host could be sensitive to the insertion of different cofactors into their binding site.

From a computational point of view, dealing with bioorganometallic systems represents one of the most challenging modeling tasks, due to the necessity of describing biometallic interactions under standard force fields (an area still in strong development, Riccardi et al., 2018), while also accounting for possible wide structural variations resulting from embedding net charges (the metal center) in a natural hydrophobic environment. As a result, the identification of a set of *in silico* modeling techniques sensitive enough to predict structural variations that arise from the incorporation of inorganic moieties into the protein host, is fundamental to speed up the success rate of ArM designs.

Here, we focus on the dynamical implications of inserting non-natural organometallic cofactors in the core of the LmrR protein. The computational approach consists of the combination of Protein-Ligand Docking and MD simulations followed by diverse trajectory analyses including all-to-all Root-mean-square-deviation (RMSD), Principal Component Analysis (PCA), cluster counting, and Root-mean-square-fluctuation (RMSF) approaches (see **Figure 3**). Together with the visual inspection of the trajectories, these analyses provide valuable information on two main areas: (1) the MD simulation time-scale required to ensure a proper conformational exploration

by the hybrid systems, and (2) how the inclusion of different organometallic external moieties can promote conformational variations of the same biological host.
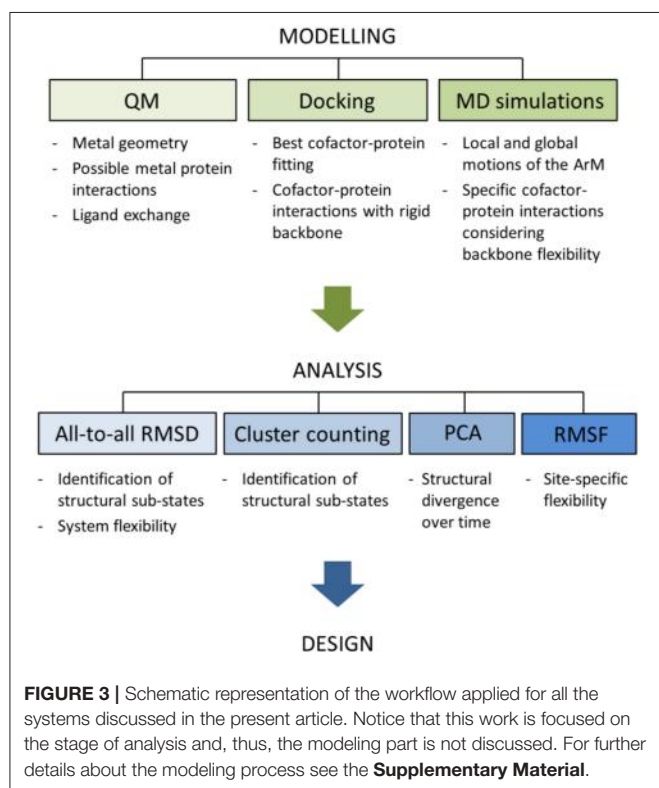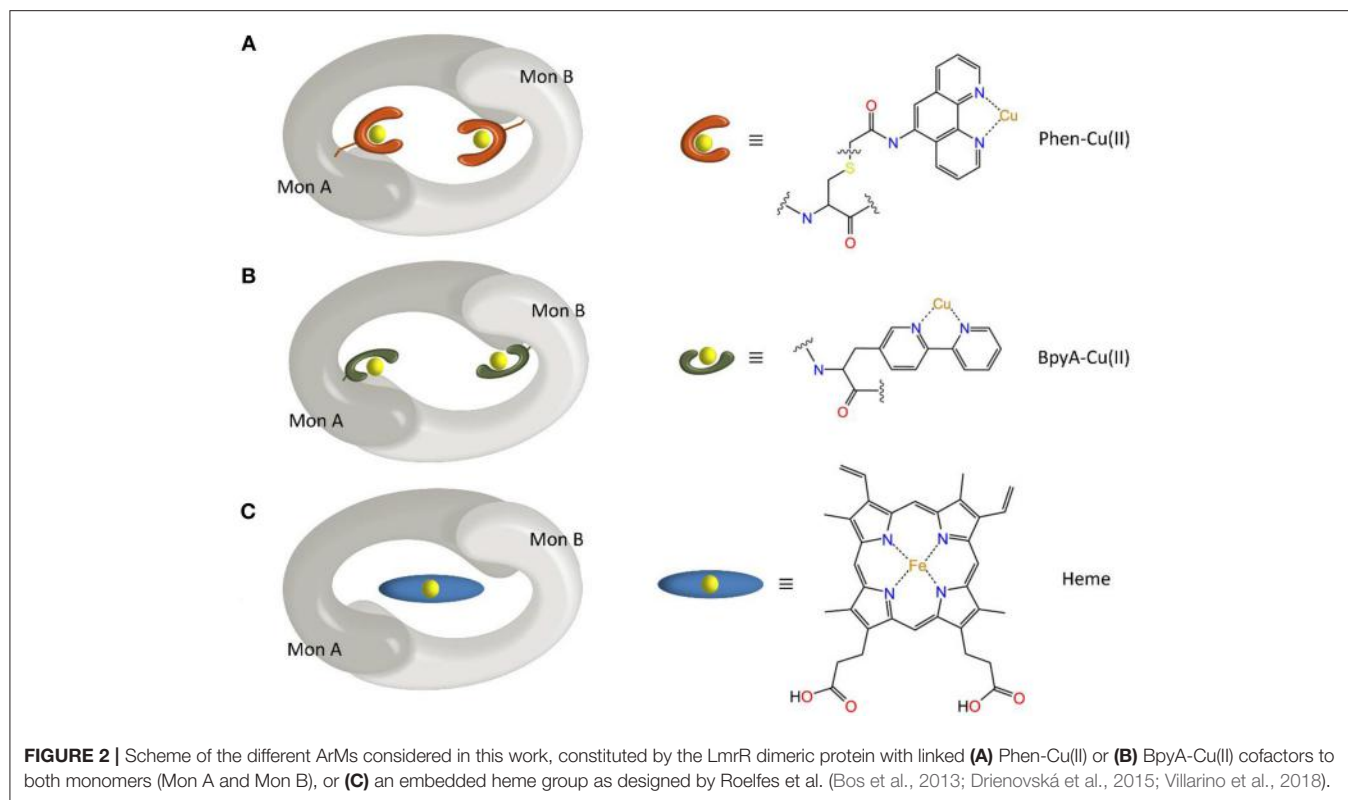
MATERIALS AND METHODS

Construction of the Dataset

Together with two natural forms of the LmrR protein found in the Protein Data Bank (PDB), the models include: the *apo* form of the LmrR protein (1), the LmrR protein bound to its inhibitor daunomycin (2), the ArMs resulting from the supramolecular interaction between the heme group and LmrR (3) and the linking of the biaqua form of two Phen-Cu(II) (4) or BpyA-Cu(II) (5) cofactors to the positions M89C/M89C' of the LmrR protein, as illustrated in **Table 1**.

Initially, quantum calculations were performed to obtain the optimized structures of the different organometallic cofactors. All the complexes were optimized with Gaussian 09 (Frisch et al., 2009) at the DFT level using the B3LYP-D3 (Becke, 1993; Stephens et al., 1994; Grimme et al., 2010) functional. For all the non-metallic atoms the 6-31G(d,p) basis set was used. For the iron and copper atoms the SDD effective core potential and its associated basis set (Dolg et al., 1987) including *f* functions for (Ehlers et al., 1993) was employed. For systems 4 and 5, for which no X-ray data was available, the organometallic complexes were incorporated into the LmrR protein scaffold via a covalent Protein-Ligand Docking approach, considering flexible side-chains for all the residues pointing toward the active site. For systems 1, 2, and 3 the X-ray structures of the *apo* LmrR (PDB code: 3F8B), the LmrR bound to daunomycin drug (PDB code: 3F8F), or the LmrR bound to the heme group (PDB code: 6FUU), respectively, were used as a starting point for the MD simulations. For system 3, the crystallographic data shows several orientations of the heme group corresponding with a rotated porphyrin with respect to the perpendicular axis passing across the iron ion. Thus, in this case, the selection of the starting point was based on a consensus between the X-ray and docked structures, resulting in the orientation corresponding to the best scored pose. Daunomycin (for system 1) and crystallographic water molecules were manually removed. Systems 4 and 5 were constructed via a covalent Protein-Ligand Docking approach by linking the aqua bound form of both BpyA-Cu(II) and Phen-Cu(II) cofactors, respectively, to the position 89/89' of the LmrR, using the X-ray structure bound to its inhibitor daunomycin (PDB code: 3F8F) as a scaffold. All docking runs were performed with GOLD 5.2 (Verdonk et al., 2003). The best scored structures, together with the X-ray data for systems 1, 2 and 3, were embedded in boxes of around 37,000 water molecules and were used as the starting point for 300 ns MD simulations. The simulations were run with the OpenMM 7.0 engine (Eastman and Pande, 2010) as wrapped in the OMMProtocol program (Pedregal et al., 2018).

The resulting trajectories were processed with an in house-automated procedure that computes a series of complementary analysis to assess the convergence: (1) the changes in the RMSD of the coordinates with respect to the starting structure, (2) the RMSD evolution with respect to all the frames in
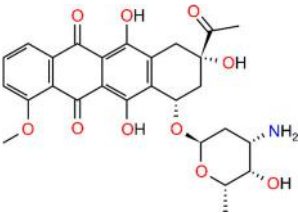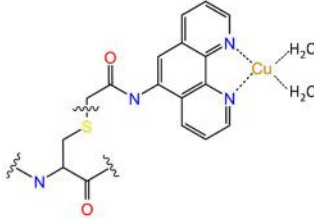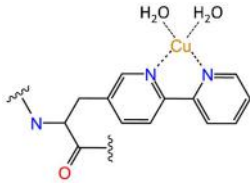
FIGURE 2 | Scheme of the different ArMs considered in this work, constituted by the LmrR dimeric protein with linked **(A)** Phen-Cu(II) or **(B)** BpyA-Cu(II) cofactors to both monomers (Mon A and Mon B), or **(C)** an embedded heme group as designed by Roelfes et al. (Bos et al., 2013; Drienovská et al., 2015; Villarino et al., 2018).

FIGURE 3 | Schematic representation of the workflow applied for all the systems discussed in the present article. Notice that this work is focused on the stage of analysis and, thus, the modeling part is not discussed. For further details about the modeling process see the **Supplementary Material**.

the simulation, (3) a PCA on the structural variability of the backbone (Balsera et al., 1996; David and Jacobs, 2014), and (4) a

cluster counting method (Daura et al., 1999; Smith et al., 2002), as described in the following section. Root-mean square fluctuation (RMSF) plots were also calculated for all systems to identify the regions with a wider range of movement in each trajectory. For depiction purposes, the dihedral angles of each trajectory were analyzed with time-structure Independent Components Analysis (tICA) (Naritomi and Fuchigami, 2013) and clustered by k-means with MSMBuilder (Harrigan et al., 2017). The resulting structures were superposed with UCSF Chimera's *matchmaker* command.

A Few Words About the Analysis of MD Trajectories

The most common strategy to analyze trajectories for assessing the structural stability of biological macromolecules is based on aligning the structure along the simulation and computing the root mean square deviation (RMSD) against a reference structure, which helps to determine whether the simulation has stabilized around a given average conformation. While easy to understand and calculate, RMSD analyses fail to show the nature of the conformational states that are sampled and only provide partial information about the structural steadiness of the system. As a result, several authors have proposed additional procedures to obtain deeper information about the convergence status of MD simulations (Daura et al., 1999; Smith et al., 2002; Grossfield and Zuckerman, 2009; Knapp et al., 2011). One of them is the all-to-all RMSD, which, instead of calculating the RMSD against a single structure, considers all the possible conformations along the simulation, resulting in a two-dimensional matrix of

TABLE 1 | Definition of the different systems considered in this work.

System	Description	Binding	Model	Bound molecule
1	*apo* LmrR	–	3F8B	–
2	LmrR + daunomycin	Supramolecular	3F8F	
				Daunomycin
3	LmrR + heme	Supramolecular	6FUU	
				Heme
4	LmrR + 2Phen-Cu(II)	Covalent (post-translational)	Docking	
				Phen-Cu(II)-(H$_2$O)$_2$
5	LmrR + 2BpyA-Cu(II)	Covalent (UAA)	Docking	
				BpyA-Cu(II)-(H$_2$O)$_2$

RMSD measurements. The resulting plot helps to depict the different visited regions along the MD trajectory, which might be challenging to detect with a standard RMSD analysis. Another noteworthy analytical strategy is the cluster counting tool which, in consensus with the above method, is useful to identify the rate of the appearance of new sub-states along the MD trajectory. Last, the Principal Component Analysis (PCA) is a standard statistical procedure used to study the correlation within a dataset (Abdi and Williams, 2010). Concerning the scope of this work, it allows the visual inspection of the conformational space explored

along the simulation timescale and whether the system continues to visit new regions or not, which enables the user to see at a glance how stable the system is and therefore whether the trajectory is approaching convergence or not. As recommended by Grossfield and Zuckerman (2009), the integration of these four complementary analytical tools can provide an accurate assessment of the conformational space explored as well as the degree of convergence of a MD trajectory.

To compute the four analyses of the trajectories, we set up several scripts using Python 3.6 through the Jupyter Notebook interface (Kluyver et al., 2016). CPPTraj (Roe and Cheatham, 2013) was first used to remove the waters and reimage the system within the same periodic box. MDTraj (McGibbon et al., 2015) was then used to load and align the trajectories. This library also provided the RMSD calculations. PCA was calculated with the routines available in scikit-learn (Pedregosa et al., 2011) set to generate two components. Cluster counting followed the algorithm proposed in Smith et al. (2002), with a RMSD cut-off of 2.0 Å. Figures were plotted with matplotlib 2.0 (Hunter, 2007). In all cases, only the α-carbons belonging to the α-helix segments were considered for distance calculations, thus discarding highly flexible regions of the protein scaffold as evidenced in the per-residue root mean square fluctuation (RMSF) plots (**Figure S6** RMSF).

The methods described above are able to depict the magnitude of the changes along a MD simulation but not what the nature of these changes are; i.e., which regions of the protein are involved in those structural variations. For this reason, their use in consensus with both the visual inspection of the trajectories as well as tools able to identify local and global motions is of great importance. In this regard, the structures obtained from tICA and k-means clustering can be very helpful. Since tICA tends to group the structure changes explained by slower motions in the first components (Naritomi and Fuchigami, 2013), it can be used to distinguish between rapid and slow motions as the system evolves over time, since these are normally related to local and global structural changes of the system, respectively. This allows the identification of the structural features that lead to the variability/stability detected by the above methods and, thus, a clearer assessment of the structural impact promoted by the incorporation of the different external moieties into the protein scaffold.

For further details about the computational procedure the reader is referred to the **Supplementary Material**.

Non-metallic Cofactor Bound LmrR

To provide a reference to assess the impact of the incorporation of artificial metallic cofactors in the LmrR protein scaffold, we first studied two experimental structures of LmrR available at the PDB without any metallic moieties bound: an *apo* form of the LmrR (PDB code: 3F8B) (**Table 1**, system 1) and a LmrR form bound to the daunomycin inhibitor (PDB code: 3F8F) (**Table 1**, system 2). This inhibitor is a substantially large and hydrophobic molecule and consequently quite reminiscent of the cofactors that have been studied in this work.

The 300 ns-long Molecular Dynamics simulations of the *apo* (system 1) and daunomycin (system 2) form of LmrR revealed

local conformational changes that mainly involve the ends of helixes α4 and α4' and the β hairpin loops of both monomers (**Figure S1**). For system 1, in addition, collective motions, not found in system 2, related to the opening and closing of the LmrR interdimeric binding site were elucidated as illustrated in **Figures S1**, **S2**. It appears that the presence of the daunomycin inhibitor, which is sandwiched between tryptophan's W96/W96' of helix α4/α4' at the center of the cavity, does not significantly influence the global motility of the protein scaffold in contrast to the *apo* form of LmrR (**Figures S1**, **S3**, systems 1 and 2). The lack of the hydrophobic inhibitor at the dimer interface seems to promote the closing of the pore by bringing the helices α4/α4' closer: around 8 Å between the alpha carbons of tryptophan's W96/W96', which are located in the center of helix α4/α4' (from now this parameter will be used as reference to assess the opening/closing of the dimer interface; see **Figure S2**, system 1). In contrast, the drug-bound form maintains an opened arrangement of helices α4/α4' (around 12.5 Å; see **Figure S2**, system 2).

The combination of cluster counting, all-to-all RMSD and PCA analyses (**Figures S3–S5**, systems 1 and 2) provides more data about the conformational sampling for systems 1 and 2. The appearance of new structural clusters reaches its plateau at 150 ns of the MD trajectory for both systems, which suggests a converged structural sampling for the simulation time scale. This is consistent with the all-to-all RMSD analysis, which indicates that both systems fluctuate between well-defined sub-states, which is also associated with a converged trajectory (**Figure S4**, systems 1 and 2). These sub-states are related to the structures which present certain structural divergences with respect to the X-ray structure, consistent with the local/global motions described above, associated mostly to the flexibility of the β hairpin loops and the ends of the α4/α4' helixes and, only for system 1, the closing of the dimer interface (**Figure S5**, systems 1 and 2).

Altogether, these analyses show that the natural motions of LmrR involve mostly the interdominial interface between monomers A and B, as well as their relative rotation, which is smoothed in the inhibitor-bound form of the protein. Additionally, this state is associated with a wide active cavity resulting from direct hydrophobic interactions of W96/W96', A92/A92', and V15/V15' with the daunomycin inhibitor. From these results we can conclude, on one hand, that a time scale of 150 ns is enough to reach a proper conformational sampling in the MD simulations for the *apo* and the daunomycin-bound forms of LmrR and, on the other hand, that, overall, the accommodation of hydrophobic moieties at the hydrophobic interface would reduce the global conformational plasticity of LmrR.

The Heme-Based Artificial Metalloenzyme

Our study then focused on the Artificial Metalloenzyme systems. We started with the ArM resulting from the supramolecular interaction between the heme group and the LmrR protein. Recent studies showed that the LmrR-heme system is able to reach efficient cyclopropanation profiles (Villarino et al., 2018). The X-ray structure of the LmrR bound to heme (PDB: 6FUU)
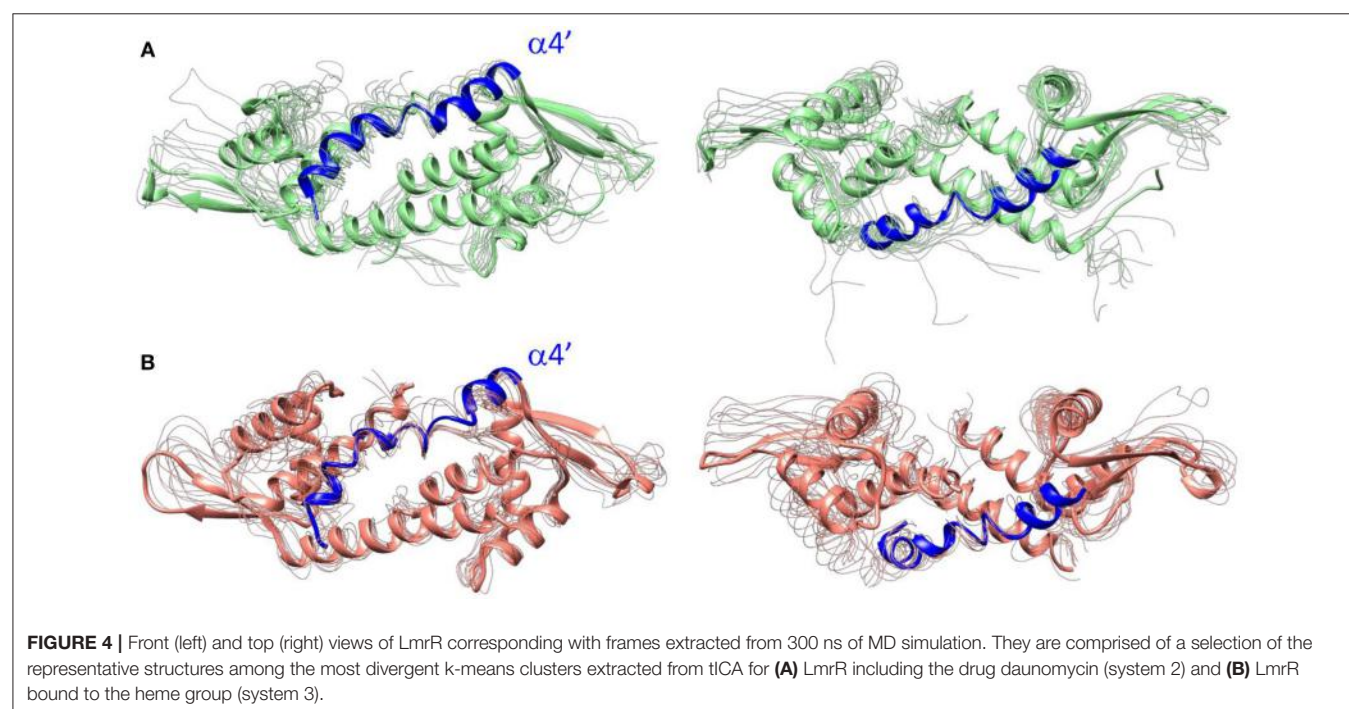
displays the prosthetic group sandwiched in between the two α4/α4' helices of the homodimers with major hydrophobic interactions between the macrocycle and the side chains of the two tryptophan's W96/W96', as well as polar interactions between N19/N19' and N14/14' and the carboxylate groups of the heme moiety. Due to the similar size and configuration between the heme group and the inhibitor daunomycin in the LmrR active site, we expected a similar impact on the protein dynamics between systems 2 (daunomycin-bound) and 3 (heme-bound). Interestingly, results showed that the dynamics of both systems followed the same trend: in both cases, the presence of the planar hydrophobic molecule at the dimer interface promotes an increase in the flexibility of the α4' helix, which for the LmrR⊂heme ArM occurs with more plasticity, comprising mainly the distortion of the same helix (**Figure 4**). This effect promotes the closing of the hydrophobic interface with respect to system 2 (**Figure S2**, systems 2 and 3), which is broader than for system 1 but narrower than for system 2, and is reflected in the increased amount of sub-states found for this system: cluster counting shows that, in this case, MD simulation reaches convergence with a maximum of 12 clusters of structures, in contrast to systems 1 and 2, which reach convergence in a maximum of eight clusters of structures (**Figure S3**). In addition, all-to-all RMSD and PCA analyses show not only a higher amount of sub-states related with a major structural exploration of the system (dark zones covering the diagonal of the all-to-all RMSD plot), but also a greater flexibility (light areas in the background of the all-to-all RMSD plot), and divergence on the nature of these sub-states (no overlapping spots in the PCA plot).

These results show that the presence of an organometallic complex at the hydrophobic interface does not change the dynamical tendency of the natural form of LmrR in system 2 but promotes an increase on the plasticity of the system, that could only be deciphered by the use of cluster counting and PCA analysis, and was particularly related to the motion of the helix α4'. More likely, this effect can be explained by the presence of the porphyrin metal center, absent in the daunomyicin-bound system 2, at the LmrR active site. Interestingly, the observed motion of the helix α4' was associated with the catalytic activity of the LmrR⊂heme ArM, in which the displacement of the tryptophan W96', located in this helix, was key for vacating one of the axial faces on the porphyrin and, thus, making the metal center accessible to the substrates to reach pre-catalytic states for the cyclopropanase activity (Villarino et al., 2018).

The Copper-Based Artificial Metalloenzymes

Next, we wanted to assess the effect of incorporating organometallic cofactors but, in this case, covalently linked to the protein scaffold at positions 89/89'. These were the copper-bound nitrogenated compounds, Ala-bipyridine (BpyA) and phenanthroline (Phen), which are covalently linked to the protein scaffold at positions 89/89'. Interestingly, despite sharing same chemical properties and a similar size and overall planarity, previous reports have shown greatly different catalytic outcomes for the enantioselective addition of water to conjugated ketones (Bos et al., 2013; Drienovská et al., 2017). These differences have been associated to the different lengths of the linkers that bind the aromatic rings of the nitrogenated moieties with the backbone (**Table 1**, systems 4 and 5), which seems to affect the overall behavior of the artificial systems (Drienovská et al., 2017). Thus, we decided to include these ArMs in our study to identify the key elements that promote such a differential effect.

FIGURE 4 | Front (left) and top (right) views of LmrR corresponding with frames extracted from 300 ns of MD simulation. They are comprised of a selection of the representative structures among the most divergent k-means clusters extracted from tICA for **(A)** LmrR including the drug daunomycin (system 2) and **(B)** LmrR bound to the heme group (system 3).

For that purpose, we studied the conformational variability of the LmrR scaffold loaded with Phen-Cu(II) and BpyA-Cu(II) cofactors simulating their most likely biaqua resting state (**Table 1**, systems 4 and 5). These were included into both monomers of the LmrR protein scaffold via a covalent Protein-Ligand Docking procedure, showing a good affinity with the protein binding site (**Table S1**) in both cases, being slightly better for Phen-Cu(II)-(H$_2$O)$_2$ (38.89 ChemScore units) than for BpyA-Cu(II)-(H$_2$O)$_2$ (34.85 ChemScore units). Consistently with previous systems, the best score poses were selected and used as the starting point for 300 ns of Molecular Dynamics simulation.

Due to their structural similarities, it would be expected that the inclusion of the Phen and BpyA cofactors into LmrR promotes a similar impact on the scaffold than the heme organometallic cofactor (system 3). However, both systems 4 and 5 present strongly different dynamic tendencies, which are particularly noteworthy for the Phen-containing ArM. Results showed a strong decrease of the global motions and the plasticity of the system with respect to the systems 1, 2, and 3 (**Figures S4–S6**, systems 1–4), the main motions being reduced to the ends of the α4/α4' helices. The appearance of new clusters of sub-states converges at the very beginning of the simulation, without identification of changes along the MD time-scale, which is consistent with the very low plasticity of the system as shown in **Figure S3** for system 4 (only 3 clusters of structures were identified after 300 ns of MD simulation). In this configuration, the biaqua form of the Phen-Cu(II) cofactors appeared stabilized by interactions comprising mainly π-stacking with phenylalanine residues F93/F93' and hydrophobic contacts with I103/I103' and the side chains of the R90/R90' residues. In addition, hydrogen bonds between the tail of the cofactor with N19/N19' residues as well as between the waters bound to the metal center and the aspartates D100/D100' seemed to further stabilize the location of the cofactor at the entrance of the dimer interface (**Figure 5A**). This binding mode is accompanied by a broader dimer interface (around 14 Å), resulting, for the daunomycin bound system 2, from the hydrophobic interactions between the active site residues with the aromatic cofactor. Additionally, all-to-all RMSD and PCA analysis evidenced a significantly reduced plasticity of the LmrR⊂Phen system (**Figures S3**, **S4**, system 4). The former analysis showed few sub-states, especially well-defined and with strong presence after 50 ns of MD simulation (the big dark area which center lies at the diagonal of the plot). Furthermore, the later analysis showed that between the few sub-states identified there is a very low divergence (PCA data appears superimposed). These results evidence that the LmrR system loaded with the Phen-Cu(II) cofactors presents major stability than that of the natural form of LmrR (system 1). It is of great relevance to note that, in contrast to the flexible nature of the catalytically active Lmr⊂heme system (system 3), this ArM, which presents the highest stability among all those considered in this study, showed a very good catalytic activity, in this case for the enantioselective hydration of ketones (Bos et al., 2013). This observation suggests that one of the factors driving the promiscuity of the LmrR based artificial enzymes is also related to the flexibility of the protein backbone,

which needs to be controlled in order to perform the different catalytic transformations.

Regarding system 5, the results showed a totally different scenario. The MD simulation evidenced the poor capability of the BpyA-Cu(II)-(H$_2$O)$_2$ cofactors to reach the center of the cavity. Instead, they appeared pointing toward the solvent during most of the trajectory and at any time along 300 ns of MD simulation the complexes appeared embedded at the dimer interface. This is the result of a lack of hydrophobic and polar interactions between the active site residues and the bipyridine cofactors (**Figure 5B**). The bipyridine cofactor has a shorter tail linking the bipyridine rings and the protein backbone (the beta carbon) than phenanthroline (five atoms, see **Figure 2**), which makes the complexes lie out of the hydrophobic cavity (**Figure 5B**). Consequently, this lack of aromatic ligands at the dimer interface promotes a narrower arrangement between helixes α4/α4' (around 8 Å between the alpha carbons of W96/W96' residues, see **Figure S2** system 5), the same as the distance observed for the *apo* form of LmrR (system 1). Additionally, the lack of stability in the positioning of the bipyridine cofactors is extended to the rest of the protein backbone, promoting an increase in its plasticity as well as distorting the global protein structure (**Figure 6**). This behavior is well-captured by the cluster counting, all-to-all RMSD and PCA analysis (**Figures S3–S5**). As expected, the number of identified sub-states highly increased for this system in contrast to system 4 (LmrR⊂Phen), being grouped in a total of nine clusters of structures along 300 ns of MD simulation, in contrast to the three clusters identified for system 4. In addition, they showed a more flexible system (light areas in the background of the all-to-all RMSD plot) as well as a strong divergence of the identified sub-states (the spots in the PCA plot do not superimpose along 300 ns of MD simulation). RMSF analysis also revealed this difference: system 5 showed higher average fluctuation values for all chains (see **Figure S6**). Interestingly, this ArM was not able to efficiently perform the enantioselective hydration of ketones, presenting much lower levels of both conversion and enantioselectivity in contrast to the LmrR⊂Phen ArM (Bos et al., 2013; Drienovská et al., 2017). After deciphering the mechanism of system 5 (Drienovská et al., 2017), it was revealed that an aspartate located at helix α4' was responsible for boosting the nucleophilic attack, the first step of the hydration reaction. For this to occur in an enantioselective manner, the cofactor-substrate complex needed to be not only located inside the active site but also stabilized in only one orientation. For this reason, the stability of the interactions at the active site seems key for the desired reaction to proceed in the LmrR-based artificial hydratases (systems 4 and 5), in contrast to the LmrR⊂heme ArM, which requires the flexibility of helix α4' to reach pre-catalytic structures.

In summary, the results showed that the stabilizing interactions that occur between the active site residues and the copper cofactors seem crucial in providing stability to the global motions of the LmrR scaffold. In this regard, the hydrophobic and polar interactions and the length of the cofactor linkers play a critical role in reducing or increasing the flexibility of the system. Additionally, it is reasonable to think that the presence of non-stabilized metal moieties close
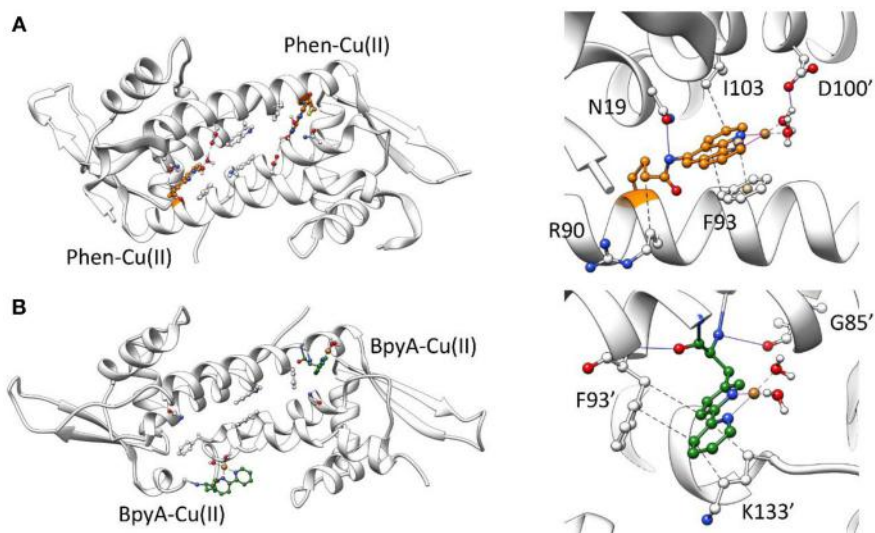
FIGURE 5 | Representative structures along 300 ns of MD simulation for the LmrR protein containing **(A)** the Phen-Cu(II)-$(H_2O)_2$ cofactor and **(B)** the BpyA-Cu(II)-$(H_2O)_2$ cofactor. A general view of the system (left) as well as specific interactions between the copper cofactors and surrounding residues analyzed with Plip (Salentin et al., 2015) (right) is illustrated.
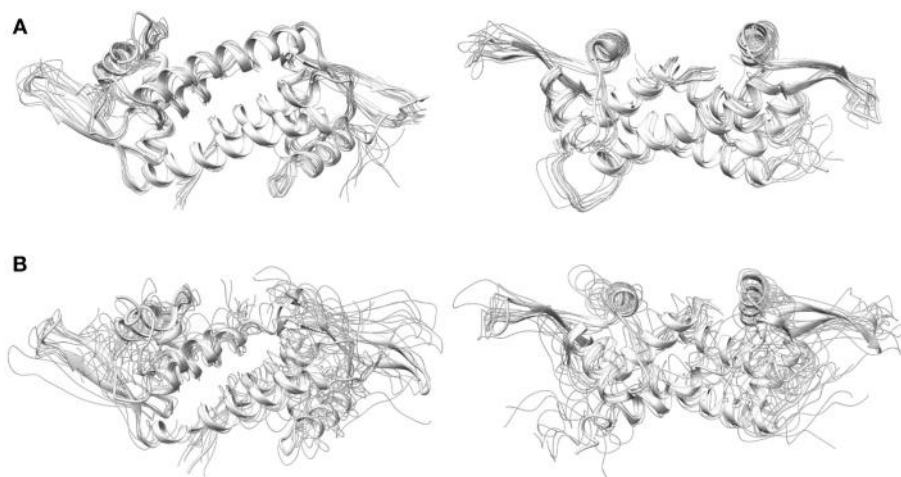
FIGURE 6 | Front (left) and top (right) view of LmrR corresponding with frames extracted from 300 ns of MD simulation comprising the 10 most representative structures extracted from k-means clustering in the tICA space for **(A)** LmrR bound to the Phen-Cu(II)-$(H_2O)_2$ cofactor (system 4) and **(B)** LmrR bound to the BpyA-Cu(II)-$(H_2O)_2$ cofactor (system 5).

to the protein backbone is one of the factors that may promote the strong distortion of the protein backbone as observed for the LmrR⊂BpyA (system 4). As a result, only the LmrR⊂Phen ArM is able to maintain an asymmetric environment around the copper cofactor during the entire simulation time scale, which is reflected in the catalytic efficiency of this artificial hydratase (Bos et al., 2013).

CONCLUSIONS

This study aims to shed light on questions concerning the structure-function relationship in proteins, especially focusing on organometallic-containing systems such as ArMs. Dealing with organometallic moieties adds an extra layer of difficulty to modeling and simulation exercises due to the challenging task of describing non-natural metal-containing moieties as part of the standard force fields, as well as their structural effect when interacting with the biological scaffold. Thus, the identification of a computational modeling workflow, together with an analytical protocol sensitive enough to decipher the changes that result from the incorporation of external moieties into proteins, appears to be of great relevance, especially in the enzyme design field.

To guarantee the quality of the systems used in this study, we made use of a set of well-characterized ArMs (LmrR loaded with heme, copper-bound phenanthroline, or copper-bound

bipyridine) designed by Roelfes et al. whose models have been experimentally validated in previous works (Drienovská et al., 2017; Villarino et al., 2018) to perform a comparative analysis of simulations resulting from the combination of Quantum Mechanics, Protein-Ligand Docking, and Molecular Dynamics techniques. Our results provide evidence that the convergence analysis used in this work can help explain the structural trends of the different systems under study. They show how the insertion of different non-natural metallic cofactors into the same biological scaffold may condition different structural modulations that, in addition, are key to the success of the desired catalytic activity. Put into the context of ArM design and *in silico* exercises, it is therefore crucial to first assess (or, at least, consider this magnitude as a variable to control sooner or later in the designing pipeline) the degree of rigidity/flexibility of the receptor-cofactor partner throughout MD simulations to understand how this can affect the reaction mechanism of interest.

Aiming at including dynamical notions in computer-aided design of ArMs, here we show the strength of combining an integrative strategy (docking + MD simulations) with convergence-based analysis, including all-to-all RMSD, PCA, RMSF, and cluster counting, to characterize the structural behavior of these complex organometallic systems. Our results also contribute to the debate on the benefit of accounting for stable vs. flexible protein scaffolds to drive the designs of the first generations of ArMs. This work makes clear that, due to the high amount of degrees of freedom controlling the different catalytic mechanisms occurring in ArMs, each of them must be considered as a separate system with its own particular patterns and

features (like flexibility/rigidity): ideally, their specific structural requirements would need to be evaluated on a one-by-one basis. Unfortunately, this means that we are still far from establishing a universal metric to guide the design of any ArM. Thus, at present, we find it essential, at least, to account for a proper protocol that establishes which modeling and analytical tools, such as the ones selected for this work, will ensure the gain of enough structural knowledge before investing in further efforts.

AUTHOR CONTRIBUTIONS

JR-G and LA-C performed the calculations and analysis. J-DM designed the project. JR-G, LA-C, and J-DM discussed the analysis and wrote the manuscript. AL participated in the discussion and the writing of the manuscript.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem. 2019.00211/full#supplementary-material

REFERENCES

Abdi, H., and Williams, L. J. (2010). Principal component analysis. *Wiley Interdiscip. Rev. Comput. Stat.* 2, 433–459. doi: 10.1002/wics.101

Arnold, F. H. (2009). How proteins adapt: lessons from directed evolution. *Trends Biochem. Sci.* 74, 41–46. doi: 10.1101/sqb.2009.74.046

Balsera, M. A., Wriggers, W., Oono, Y., and Schulten, K. (1996). Principal component analysis and long time protein dynamics. *J. Phys. Chem.* 100, 2567–2572. doi: 10.1021/jp9536920

Becke, A. D. (1993). Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* 98, 5648–5652. doi: 10.1063/1.464913

Boersma, A. J., Coquière, D., Geerdink, D., Rosati, F., Feringa, B. L., and Roelfes, G. (2010). Catalytic enantioselective syn hydration of enones in water using a DNA-based catalyst. *Nat. Chem.* 2, 991–995. doi: 10.1038/nchem.819

Bos, J., Browne, W. R., Driessen, A. J. M., and Roelfes, G. (2015). Supramolecular assembly of artificial metalloenzymes based on the dimeric protein LmrR as promiscuous scaffold. *J. Am. Chem. Soc.* 137, 9796–9799. doi: 10.1021/jacs.5b05790

Bos, J., Fusetti, F., Driessen, A. J. M., and Roelfes, G. (2012). Enantioselective artificial metalloenzymes by creation of a novel active site at the protein dimer interface. *Angew. Chem. Int. Ed.* 51, 7472–7475. doi: 10.1002/anie.201202070

Bos, J., García-Herraiz, A., and Roelfes, G. (2013). An enantioselective artificial metallo-hydratase. *Chem. Sci.* 4, 3578–3582. doi: 10.1039/c3sc51449h

Callender, R., and Dyer, R. B. (2014). The dynamical nature of enzymatic catalysis. *Acc. Chem. Res.* 48, 407–413. doi: 10.1021/ar5002928

Daura, X., Gunsteren, W. F., Van, and Mark, A. E. (1999). Folding-unfolding thermodynamics of a β-heptapeptide from equilibrium simulations. *Proteins Struct. Funct. Genet.* 34, 269–280. doi: 10.1002/(SICI)1097-0134(19990215)34:3<269::AID-PROT1>3. 0.CO;2-3

David, C. C., and Jacobs, D. J. (2014). Principal component analysis: a method for determining the essential dynamics of proteins. *Methods Mol. Biol.* 1084, 193–226. doi: 10.1007/978-1-62703-658-0_11

Diéguez, M., Bäckvall, J.-E., and Pàmies, O. (2018). *Artificial Metalloenzymes and MetalloDNAzymes in Catalysis: From Design to Applications.* Weinheim: John Wiley & Sons.

Dolg, M., Wedig, U., Stoll, H., and Preuss, H. (1987). Energy-adjusted ab initio pseudopotentials for the first row transition elements. *J. Chem. Phys.* 86, 866–872. doi: 10.1063/1.452288

Drienovská, I., Alonso-Cotchico, L., Vidossich, P., Lledós, A., Maréchal, J.-D., and Roelfes, G. (2017). Design of an enantioselective artificial metallo-hydratase enzyme containing an unnatural metal-binding amino acid. *Chem. Sci.* 8, 7228–7235. doi: 10.1039/C7SC03477F

Drienovská, I., Rioz-Martínez, A., Draksharapu, A., and Roelfes, G. (2015). Novel artificial metalloenzymes by *in vivo* incorporation of metal-binding unnatural amino acids. *Chem. Sci.* 6, 770–776. doi: 10.1039/C4SC01525H

Dutta, D., and Mishra, S. (2017). Active site dynamics in substrate hydrolysis catalyzed by DapE enzyme and its mutants from hybrid QM/MM-molecular dynamics simulation. *J. Phys. Chem. B* 121, 7075–7085. doi: 10.1021/acs.jpcb.7b04431

Eastman, P., and Pande, V. (2010). OpenMM: a hardware-independent framework for molecular simulations. *Comput. Sci. Eng.* 12, 34–39. doi: 10.1109/MCSE.2010.27

Ehlers, A. W., Böhme, M., Dapprich, S., Gobbi, A., Höllwarth, A., Jonas, V., et al. (1993). A set of f-polarization functions for pseudo-potential basis sets of the transition metals Sc-Cu, Y-Ag and La-Au. *Chem. Phys. Lett.* 208, 111–114. doi: 10.1016/0009-2614(93)80086-5

Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., et al. (2009). *Gaussian 09, Revision D.01.* Gaussian, Inc., Wallingford CT, 2016.

Grimme, S., Antony, J., Ehrlich, S., and Krieg, H. (2010). 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. doi: 10.1063/1.3382344

Grossfield, A., and Zuckerman, D. M. (2009). Quantifying uncertainty and sampling quality in biomolecular simulations. *Annu. Rep. Comput. Chem.* 1, 23–48. doi: 10.1016/S1574-1400(09)00502-7

Hammes, G. G., Benkovic, S. J., and Hammes-Schiffer, S. (2011). Flexibility, diversity, and cooperativity: pillars of enzyme catalysis. *Biochemistry* 50, 10422–10430. doi: 10.1021/bi201486f

Hammes-Schiffer, S., and Benkovic, S. J. (2006). Relating protein motion to catalysis. *Annu. Rev. Biochem.* 75, 519–541. doi: 10.1146/annurev.biochem.75.103004.142800

Harrigan, M. P., Sultan, M. M., Hernández, C. X., Husic, B. E., Eastman, P., Schwantes, C. R., et al. (2017). MSMBuilder: statistical models for biomolecular dynamics. *Biophys. J.* 112, 10–15. doi: 10.1016/j.bpj.2016.10.042

Henzler-Wildman, K., and Kern, D. (2007). Dynamic personalities of proteins. *Nature* 450, 964–972. doi: 10.1038/nature06522

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* 9, 90–95. doi: 10.1109/MCSE.2007.55

Jiang, L., Althoff, E. A., Clemente, F. R., Doyle, L., Rothlisberger, D., Zanghellini, A., et al. (2008). *De novo* computational design of retro-aldol enzymes. *Science* 319, 1387–1391. doi: 10.1126/science.1152692

Kamariah, N., Eisenhaber, B., Eisenhaber, F., and Grüber, G. (2018). Active site CP-loop dynamics modulate substrate binding, catalysis, oligomerization, stability, over-oxidation and recycling of 2-Cys peroxiredoxins. *Free Radic. Biol. Med.* 118, 59–70. doi: 10.1016/j.freeradbiomed.2018.02.027

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., et al. (2016). "Jupyter Notebooks–a publishing format for reproducible computational workflows," in *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, eds F. Loizides and B. Scmidt (Göttingen: IOS Press), 87–90.

Knapp, B., Frantal, S., Cibena, M., Schreiner, W., and Bauer, P. (2011). Is an intuitive convergence definition of molecular dynamics simulations solely based on the root mean square deviation possible? *J. Comput. Biol.* 18, 997–1005. doi: 10.1089/cmb.2010.0237

Kokubo, T., Sugimoto, T., Uchida, T., Tanimoto, S., and Okano, M. (1983). The bovine serum albumin—2-phenylpropane-1,2-diolatodioxo-osmium(VI) complex as an enantioselective catalyst for cis-hydroxylation of alkenes. *J. Chem. Soc. Chem. Commun.* 0, 769–770. doi: 10.1039/C39830000769

Laureanti, J. A., Buchko, G. W., Katipamula, S., Su, Q., Linehan, J. C., Oleg, A., et al. (2019). Protein scaffold activates catalytic CO_2 hydrogenation by a rhodium bis(diphosphine) complex. *ACS Catal.* 9, 620–625. doi: 10.1021/acscatal.8b02615

Luirink, R. A., Dekker, S. J., Capoferri, L., Janssen, L. F. H., Kuiper, C. L., Ari, M. E., et al. (2018). A combined computational and experimental study on selective flucloxacillin hydroxylation by cytochrome P450 BM3 variants. *J. Inorg. Biochem.* 184, 115–122. doi: 10.1016/j.jinorgbio.2018.04.013

Madoori, P. K., Agustiandari, H., Driessen, A. J., and Thunnissen, A. M. (2009). Structure of the transcriptional regulator lmrr and its mechanism of multidrug recognition. *EMBO J.* 28, 156–166. doi: 10.1038/emboj.2008.263

Mahammed, A., and Gross, Z. (2005). Albumin-conjugated corrole metal complexes: extremely simple yet very efficient biomimetic oxidation systems. *J. Am. Chem. Soc.* 127, 2883–2887. doi: 10.1021/ja045372c

Maria-Solano, M. A., Serrano-Hervás, E., Romero-Rivera, A., Iglesias-Fernández, J., and Osuna, S. (2018). Role of conformational dynamics in the evolution of novel enzyme function. *Chem. Commun.* 54, 6622–6634. doi: 10.1039/C8CC02426J

McGibbon, R. T., Beauchamp, K. A., Harrigan, M. P., Klein, C., Swails, J. M., Hernández, C. X., et al. (2015). MDTraj: a modern open library for the analysis of molecular dynamics trajectories. *Biophys. J.* 109, 1528–1532. doi: 10.1016/j.bpj.2015.08.015

Muñoz Robles, V., Ortega-Carrasco, E., Alonso-Cotchico, L., Rodriguez-Guerra, J., Lledós, A., and Maréchal, J.-D. (2015). Toward the computational design of artificial metalloenzymes: from protein-ligand docking to multiscale approaches. *ACS Catal.* 5, 2469–2480. doi: 10.1021/acscatal.5b00010

Muñoz Robles, V., Ortega-Carrasco, E., González Fuentes, E., Lledós, A., and Maréchal, J.-D. (2011). What can molecular modelling bring to the design of artificial inorganic cofactors? *Faraday Discuss.* 148, 137–228. doi: 10.1039/C004578K

Nagel, Z. D., and Klinman, J. P. (2009). A 21st century revisionist's view at a turning point in enzymology. *Nat. Chem. Biol.* 5, 543–550. doi: 10.1038/nchembio.204

Naritomi, Y., and Fuchigami, S. (2013). Slow dynamics of a protein backbone in molecular dynamics simulation revealed by time-structure based independent component analysis. *J. Chem. Phys.* 134:65101. doi: 10.1063/1.3554380

Ohashi, M., Koshiyama, T., Ueno, T., Yanase, M., Fujii, H., and Watanabe, Y. (2003). Preparation of artificial metalloenzymes by insertion of chromium(III) Schiff base complexes into apomyoglobin mutants. *Angew. Chem. Int. Ed.* 42, 1005–1008. doi: 10.1002/anie.200390256

Pedregal, J. R., Alonso-Cotchico, L., Velasco-Carneros, L., and Maréchal, J.-D. (2018). OMMProtocol: a command line application to launch molecular dynamics simulations with OpenMM. *ChemRxiv.* doi: 10.26434/chemrxiv.7059263

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830. arXiv:1201.0490v4.

Reetz, M. T., and Jiao, N. (2006). Copper-phthalocyanine conjugates of serum albumins as enantioselective catalysts in Diels-Alder reactions. *Angew. Chem. Int. Ed.* 45, 2416–2419. doi: 10.1002/anie.200504561

Reetz, M. T., Rentzsch, M., Pletsch, A., and Maywald, M. (2002). Towards the directed evolution of hybrid catalysts. *Chimia* 56, 721–723. doi: 10.2533/000942902777679920

Riccardi, L., Genna, V., and Vivo, M. D. (2018). Metal-ligand interactions in drug design. *Nat. Rev. Chem.* 2, 100–112. doi: 10.1038/s41570-018-0018-6

Roe, D. R., and Cheatham, T. E. (2013). PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* 9, 3084–3095. doi: 10.1021/ct400341p

Röthlisberger, D., Khersonsky, O., Wollacott, A. M., Jiang, L., DeChancie, J., Betker, J., et al. (2008). Kemp elimination catalysts by computational enzyme design. *Nature* 453, 190–195. doi: 10.1038/nature06879

Rout, M. K., Lee, B. L., Lin, A., Xiao, W., and Spyracopoulos, L. (2018). Active site gate dynamics modulate the catalytic activity of the ubiquitination enzyme E2-25K. *Sci. Rep.* 8:7002. doi: 10.1038/s41598-018-25476-8

Salentin, S., Schreiber, S., Haupt, V. J., Adasme, M. F., and Schroeder, M. (2015). PLIP: fully automated protein-ligand interaction profiler. *Nucleic Acids Res.* 43, 443–447. doi: 10.1093/nar/gkv315

Schlee, S., Klein, T., Schumacher, M., Nazet, J., Merkl, R., Steinhoff, H.-J., et al. (2018). Relationship of catalysis and active site loop dynamics in the (βα)8-barrel enzyme indole-3-glycerol phosphate synthase. *Biochemistry* 57, 3265–3277. doi: 10.1021/acs.biochem.8b00167

Schwizer, F., Okamoto, Y., Heinisch, T., Gu, Y., Pellizzoni, M. M., Lebrun, V., et al. (2017). Artificial metalloenzymes: reaction scope and optimization strategies. *Chem. Rev.* 118, 142–231. doi: 10.1021/acs.chemrev.7b00014

Sen, K., Horrell, S., Kekilli, D., Yong, C. W., Keal, T. W., Atakisi, H., et al. (2017). Active-site protein dynamics and solvent accessibility in native Achromobacter cycloclastes copper nitrite reductase. *IUCrJ* 4, 495–505. doi: 10.1107/S2052252517007527

Sharma, B., Jamdar, S. N., Ghosh, B., Yadav, P., Kumar, A., Kundu, S., et al. (2017). Active site gate of M32 carboxypeptidases illuminated by crystal structure and molecular dynamics simulations. *Biochim. Biophys. Acta Proteins Proteomics* 1865, 1406–1415. doi: 10.1016/j.bbapap.2017.07.023

Siegel, J. B., Zanghellini, A., Lovick, H. M., Kiss, G., Lambert, A. R., St. Clair, J. L., et al. (2010). Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science* 329, 309–313. doi: 10.1126/science.1190239

Smith, L. J., Daura, X., and Gunsteren, W. F., Van. (2002). Assessing equilibration and convergence in biomolecular simulations. *Proteins Struct. Funct. Genet.* 48, 487–496. doi: 10.1002/prot.10144

Stephens, P. J., Devlin, F. J., Chabalowski, C. F., and Frisch, M. J. (1994). Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem.* 98, 11623–11627. doi: 10.1021/j100096a001

Van De Velde, F., Arends, C. E., and Sheldon, R. A. (2000). Biocatalytic and biomimetic oxidations with vanadium. *J. Inorg. Biochem.* 80, 81–89. doi: 10.1016/S0162-0134(00)00043-X

Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W., and Taylor, R. D. (2003). Improved protein-ligand docking using GOLD. *Proteins* 52, 609–623. doi: 10.1002/prot.10465

Villarino, L., Splan, K., Reddem, E., Souza, C., de, Alonso-Cotchico, L., Lledós, A., et al. (2018). An artificial heme enzyme for cyclopropanation reactions. *Angew. Chem. Int. Ed.* 57, 7785–7789. doi: 10.1002/anie.201 802946

Wilson, K. A., and Wetmore, S. D. (2017). Conformational fexibility of the benzyl-guanine adduct in a bypass polymerase active site permits replication: insights from molecular dynamics simulations. *Chem. Res. Toxicol.* 30, 2013–2022. doi: 10.1021/acs.chemrestox.7b00179

Wilson, R. H., Zamfir, S., and Sumner, I. (2017). Molecular dynamics simulations reveal a new role for a conserved active site asparagine in a ubiquitin-conjugating enzyme. *J. Mol. Graph. Model.* 76, 403–411. doi: 10.1016/j.jmgm.2017. 07.006

Conformational Itinerary of Sucrose During Hydrolysis by Retaining Amylosucrase

Santiago Alonso-Gil[1], Joan Coines[1], Isabelle André[2*] and Carme Rovira[1,3*]

[1] Departament de Química Inorgànica i Orgànica (Secció de Química Orgànica) and Institut de Química Teòrica i Computacional, Universitat de Barcelona, Martí i Franquès 1, Barcelona, Spain, [2] Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés, LISBP, Université de Toulouse, CNRS, INRA, INSA, Toulouse, France, [3] Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

By means of QM(DFT)/MM metadynamics we have unraveled the hydrolytic reaction mechanism of *Neisseria polysaccharea* amylosucrase (*Np*AS), a member of GH13 family. Our results provide an atomistic picture of the active site reorganization along the catalytic double-displacement reaction, clarifying whether the glycosyl-enzyme reaction intermediate features an α-glucosyl unit in an undistorted $^4C_1$ conformation, as inferred from structural studies, or a distorted $^1S_3$-like conformation, as expected from mechanistic analysis of glycoside hydrolases (GHs). We show that, even though the first step of the reaction (glycosylation) results in a $^4C_1$ conformation, the α-glucosyl unit undergoes an easy conformational change toward a distorted conformation as the active site preorganizes for the forthcoming reaction step (deglycosylation), in which an acceptor molecule, i.e., a water molecule for the hydrolytic reaction, performs a nucleophilic attack on the anomeric carbon. The two conformations ($^4C_1$ ad E_3) can be viewed as two different states of the glycosyl-enzyme intermediate (GEI), but only the E_3 state is preactivated for catalysis. These results are consistent with the general conformational itinerary observed for α-glucosidases.

Keywords: glycoside hydrolases, quantum mechanics/molecular mechanics, double-displacement reaction, *ab initio* molecular dynamics, metadynamics, conformational analysis

INTRODUCTION

Glycoside hydrolases (GHs) are enzymes responsible for the degradation or hydrolysis of glycosidic bonds in carbohydrates. GHs are of fundamental interest in glycobiology and glycomics (Gamblin et al., 2009) since they are responsible for the modification of polysaccharides and glycoconjugates involved in numerous biological processes such as cell-cell recognition and polysaccharide degradation for biofuel processing (Pauly and Keegstra, 2008). These enzymes also provide paradigms for enzymatic catalysis that extend beyond the bounds of carbohydrate chemistry (Koshland, 1953; Davies et al., 2012). GHs are systematically classified in 156 families (by March 2019) according to their sequence similarity (http://www.cazy.org) (Cantarel et al., 2009). Enzymes from the same family
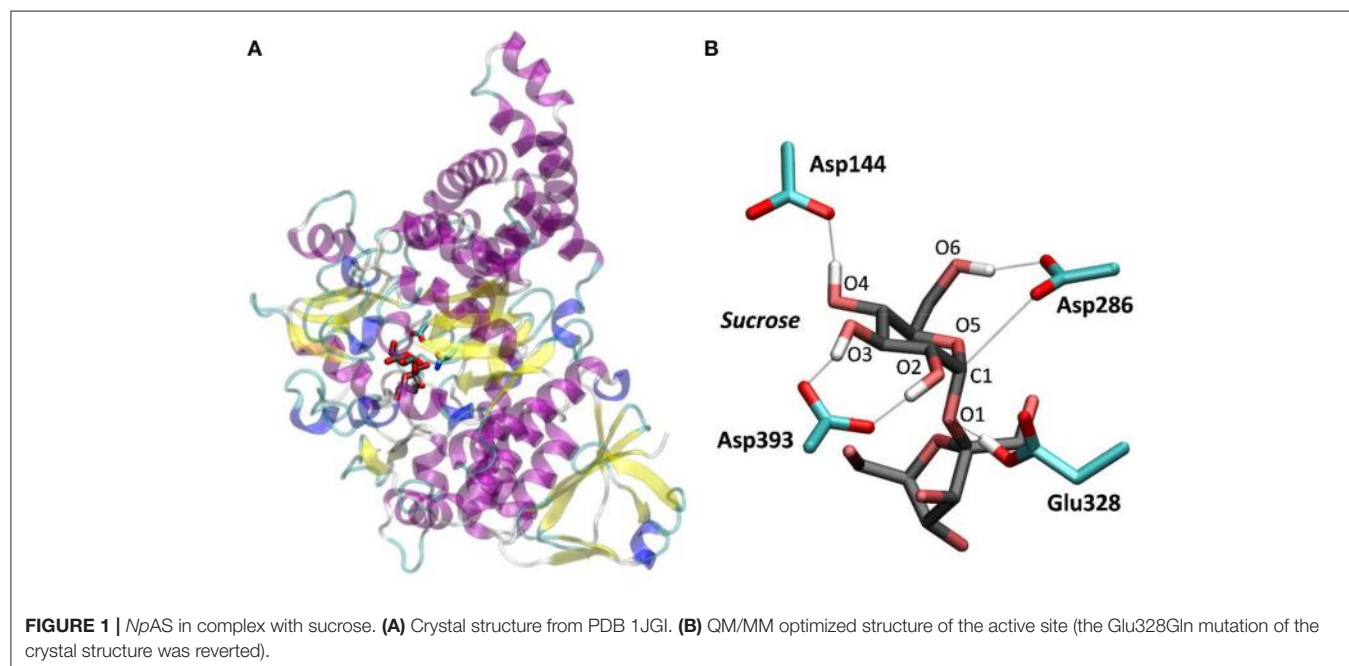
usually share the same catalytic mechanism, either with retention or inversion of the anomeric configuration (**Figure S1**), in which an oxocarbenium ion-like transition state is formed (Rye and Withers, 2000; Zechel and Withers, 2000). Enzymes from the same family also share the same *conformational itinerary* of the saccharide unit at the *-1* subsite (hereafter named *-1* sugar), in which this saccharide unit changes conformation following a straight line or a equatorial one in Stoddart's diagram (**Figure S2**), either the one representing the Northern of Southern projection of the puckering sphere (Iglesias-Fernandez et al., 2015). Because there are a few conformations that satisfy the stereoelectronic requirements for a stable oxocarbenium ion ($^4H_3$, $^3H_4$, $B_{2,5}$, $^{2,5}B$, ^3E, E_3, ^4E, and E_4) the number of itineraries is limited (Speciale et al., 2014; Ardèvol and Rovira, 2015). However, there are several GH families for which the itinerary has not been proven yet or it is controversial.

Members of GH13 family are retaining GHs that catalyze the cleavage of the α-glucosidic linkages, such as α-amylase, responsible for starch and amylose degradation. This family also gathers unique enzymes called amylosucrases that catalyze from sole sucrose the synthesis of α-D-glucopyranosyl homopolymers and oligomers accompanied with limited sucrose hydrolysis (Albenne et al., 2004). These enzymes are thus considered to be both glycoside hydrolases and transglucosylases. The substrate conformational itinerary from the Michaelis complex (MC, **Figure S1**) to the covalent glycosyl-enzyme intermediate (GEI), is expected to be identical.

Several X-Ray structures on GH13 enzymes have demonstrated that the conformation of the *-1* sugar ring in the MC is $^4C_1$ (Fujimoto et al., 1998; Mirza et al., 2001; Skov et al., 2002). The lack of distortion in the conformation, unlike many GHs, is because the α-stereochemistry at the anomeric carbon assures that the leaving group (*+1* sugar) is in a reactive orientation for the S_N2 reaction to take place, unlike e.g., β-glucosidases (Biarnés et al., 2006, 2011). In fact, α-glucosidases have been suggested to follow a $^4C_1 \rightarrow [^4H_3]^{\ddagger} \rightarrow {}^1S_3$ itinerary, which is the opposite as the $^1S_3 \rightarrow [^4H_3]^{\ddagger} \rightarrow {}^4C_1$ for β-glucosidases (Davies et al., 2012). These differences are being exploited in the field of inhibitor design (Beenakker et al., 2017).

Concerning the glycosyl-enzyme covalent intermediate (GEI) of the reaction, structural analyses of GH13 amylosucrase show that the *-1* sugar bears a $^4C_1$ conformation (Jensen et al., 2004). In the case of α-amylase, the GEI trapped using sugar analogs exhibits a similar $^4C_1$ conformation (Zhang et al., 2009; Caner et al., 2016). QM/MM calculations of human pancreatic α-amylase captured the relevant role of the catalytic residues during the hydrolysis mechanism (Pinto et al., 2015), but the conformational itinerary of the *-1* sugar was not investigated. On the other hand, the GEI of *Bifidobacterium adolescentis* sucrose phosphorylase, a GH13 enzyme, shows the *-1* sugar distorted in a $^1S_3$ conformation, suggesting a completely different conformational pathway ($^4C_1 \rightarrow$ TS $\rightarrow {}^1S_3$) (Mirza et al., 2006). Being both enzymes from the same family and acting over the same substrate, this discrepancy is puzzling. The $^4C_1$ conformation of the GEI is also in contrast with the irreversible cyclosulphate inhibitors of α-glucosidases from other GH families, which show an unambiguous $^1S_3$ conformation of the GEI analog (Artola et al., 2017). To solve this conundrum, we here uncover the conformational itinerary of the α-glucosyl unit of *Neisseria polysaccharea* amylosucrase (*Np*AS), a member of family GH13 for which a structure of the Michaelis complex with the natural substrate (sucrose) is available (**Figure 1**), during catalysis. Our simulations, performed by means of QM/MM metadynamics methods, show that the GEI can exhibit both a

FIGURE 1 | *Np*AS in complex with sucrose. **(A)** Crystal structure from PDB 1JGI. **(B)** QM/MM optimized structure of the active site (the Glu328Gln mutation of the crystal structure was reverted).

relaxed $^4C_1$ conformation or a distorted E_3-like conformation of the -1 sugar, depending on whether the catalytic water is properly placed and oriented for catalysis or it is on its way in.

RESULTS

Michaelis Complex Structure

To model the Michaelis complex of NpAS, we reverted the Glu328Gln mutation of the X-ray structure of the NpAS-sucrose complex and performed classical MD simulations. The conformation of the α-glucosyl ring at the -1 sugar changed from $^4C_1$ to $B_{3,O}$ during the first 6 ns of the MD simulation, due to an increase in the distance between the nucleophile and the anomeric carbon, but returned to $^4C_1$ during the remaining 8 ns, indicating that it is probably the most stable state. The $^4C_1$ conformation was maintained during the subsequent QM/MM MD equilibration.

To further ascertain which is the most favored conformation, we computed the conformational free energy landscape (FEL) of the α-glucosyl ring at the -1 subsite by QM/MM metadynamics, using the Cremer-Pople puckering coordinates θ/φ as collective variables. This is a well-tested approach that we have used with success to analyse the conformation of carbohydrates in isolation or in the active sites of GHs (Biarnés et al., 2007; Alonso-Gil et al., 2017). The computed FEL, shown in **Figure 2**, confirms that the most stable conformation is $^4C_1$, with a secondary minimum (4.7 kcal/mol above in free energy) corresponding to a $B_{3,O}$ distorted conformation. The most relevant hydrogen bond interactions involving the -1 sugar for the most stable $^4C_1$ conformer are listed in **Table 1** (see atom labeling in **Figure 1B**). This conformer shows a longer glycosidic bond distance (C1-O1 = 1.44 Å) compared to $B_{3,O}$, (1.42 Å) and the C1-O5 bond is shorter by 0.03 Å. Notably, the distance between the nucleophile

oxygen (O_{Asp393}) and the anomeric carbon (C1) is shorter when the sugar is not distorted, with values of 3.10 Å ($^4C_1$) and 3.30 Å ($B_{3,O}$), which would facilitate the nucleophilic attack by Asp286. Clearly, only the $^4C_1$ conformation of the substrate is preactivated for catalysis.

Glycosylation Reaction

The simulation of the first step of the double displacement reaction (glycosylation) was initiated from a snapshot of the global minimum ($^4C_1$ conformation) of the conformational FEL of the Michaelis complex. Three collective variables were used to drive the glycosylation reaction (**Figure S3** and Methods section), representing the proton transfer (CV1), the nucleophilic attack (CV2) and the glycosidic bond cleavage (CV3).

The free energy landscape reconstructed from the metadynamics simulation (reaction FEL) is shown in **Figure 3A** (for further detail, **Figure 3B** shows a two-dimensional projection). The FEL exhibits two clear minima in opposite regions: reactants (Michaelis complex, MC) and products (glycosyl-enzyme intermediate, GEI), separated by a transition state (TS1). The rate-limiting step of NpAS is not known but we can assume it is the glycosylation step, as in most GHs acting on substrates with a sugar aglycone (Li et al., 2001). Under this assumption, the reaction free energy barrier (17.2 kcal/mol) is in very good agreement with the experimental value of 17.9 kcal/mol estimated from the room temperature rate constant (Potocki de Montalk et al., 2000).
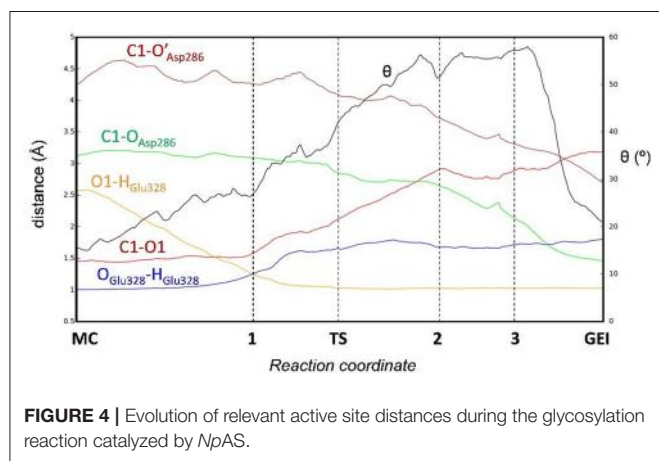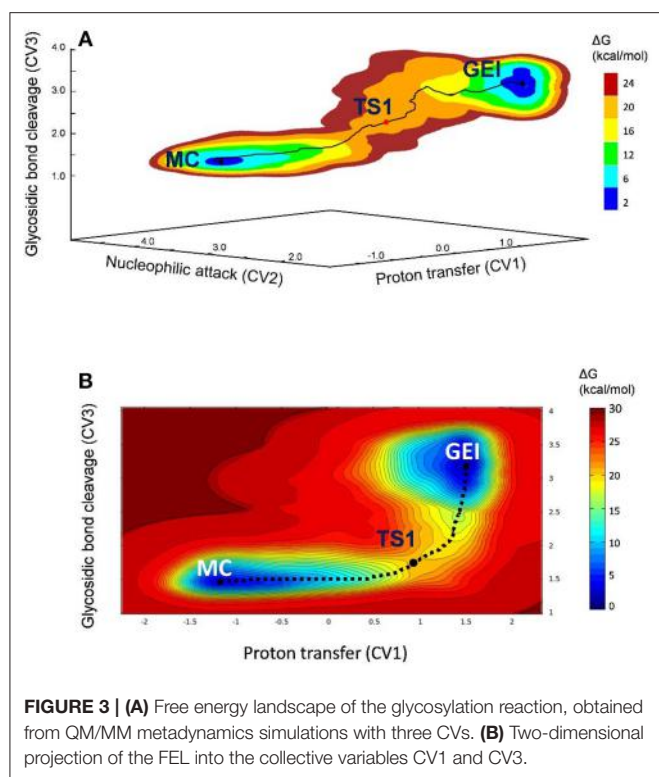
Figure 4 and **Table 2** show the evolution of the main catalytic distances along the minimum energy pathway and the structures of the active site at the stationary states are represented in **Figure 5** (top panels). The glycosylation reaction begins with the approach of Glu328 to the glycosidic oxygen. The distance between the Glu328 acid/base proton and the glycosidic oxygen ($O1-H_{Glu328}$) quickly decreases from 2.5 to 1.2 Å. Afterwards (**1** in **Figure 4**), the acid/base residue transfers the proton to the glycosidic oxygen and the glycosidic bond (C1-O1) starts to increase. Protonation of the glycosidic oxygen takes place in the *anti*-configuration with respect to the C1-O5 bond, as common in α-glycosidases (Alonso-Gil et al., 2017). At the reaction transition state (TS1), the glycosidic bond is partially
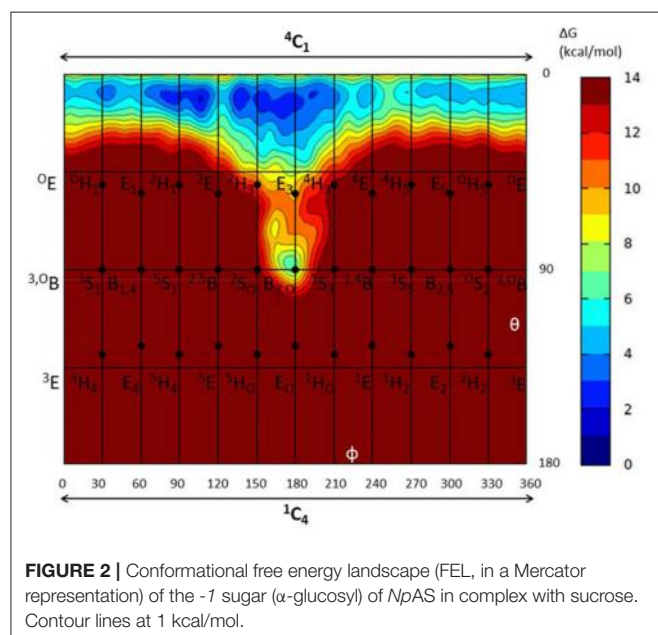
FIGURE 2 | Conformational free energy landscape (FEL, in a Mercator representation) of the -1 sugar (α-glucosyl) of NpAS in complex with sucrose. Contour lines at 1 kcal/mol.

TABLE 1 | Calculated values of the most important interactions between the -1 α-glucosyl residue and the enzyme in the Michaelis complex (energies in kcal/mol, distances in Å).

Conformation	$^4C_1$	$B_{3,O}$
φ, θ	136°, 10°	176°, 87°
ΔG	0.0	4.7
C1 – O1	1.44 ± 0.05	1.42 ± 0.04
C1 – O5	1.42 ± 0.04	1.45 ± 0.04
2-OH – O_{Asp393}	1.51 ± 0.10	1.54 ± 0.09
3-OH – O'_{Asp393}	1.60 ± 0.16	1.66 ± 0.16
4-OH – O_{Asp144}	1.49 ± 0.09	1.50 ± 0.09
6-OH – O'_{Asp286}	1.63 ± 0.24	1.66 ± 0.16
C1 – O_{Asp286}	3.10 ± 0.12	3.30 ± 0.16

FIGURE 3 | (A) Free energy landscape of the glycosylation reaction, obtained from QM/MM metadynamics simulations with three CVs. **(B)** Two-dimensional projection of the FEL into the collective variables CV1 and CV3.

FIGURE 4 | Evolution of relevant active site distances during the glycosylation reaction catalyzed by NpAS.

broken (1.98 Å) but the nucleophile has not yet started to attack the anomeric carbon (2.98 Å). Therefore, the reaction follows a dissociative mechanism and can be described as $D_N A_N$ (Guthrie and Jencks, 1989; Schramm and Shi, 2001). Afterwards, the distance between the nucleophilic oxygen of Asp286 and the anomeric carbon decreases and the glycosyl-enzyme covalent bond forms. As it was found for endo-β-glucanase (Biarnés et al., 2011), the maximum oxocarbenium ion character does not occur at the TS but later on the reaction pathway (**2** in **Figure 4**). At this point, both the nucleophile and the leaving group are well separated from the C1 atom and its charge is higher than the one at the TS (by 0.05 electrons, **Figure S4**). The calculations thus reveal that the glycosylation reaction features an early TS with respect to charge development. The glycosyl-enzyme intermediate is almost formed at **3** (C1-O_{Asp286} = 2.14 Å). Finally, the hydrogen bond interaction between the free oxygen of Asp286 (O') and the 6-OH breaks and Asp286 rotates around the covalent bond with the -1 sugar, which collapses to an undistorted $^4C_1$ conformation (GEI).

It is interesting to analyse in detail the conformation of the -1 sugar during the reaction, in relation with the hydrogen bond between the nucleophile residue (Asp286) and the 6-OH. The evolution of the sugar conformation during the glycosylation reaction, defined by the θ puckering coordinate, is shown in **Figure 6**. From the MC to the TS, the conformation evolves from $^4C_1$ to $^4H_3/^4C_1$. Afterwards, the sugar adopts an E_3 envelope conformation (**2**), followed by a $^4H_3$ half-chair (**3**). At this point, a sudden conformational change toward $^4C_1$ takes place (θ changes from 58° to 20°), concomitant with rotation of Asp286 around the glycosyl-enzyme bond, and the disruption of the 6-OH hydrogen bond interaction. The covalent intermediate is already formed before the rotation and the two configurations of the GEI (with and without hydrogen bond with the 6-OH) are close in energy. However, only the one without the hydrogen bond, and thus with a $^4C_1$ conformation, corresponds to a minimum on the FEL of **Figure 3**. Thus, the conformation of the GEI after the first step of the double displacement reaction is $^4C_1$. In other words, the conformational itinerary is in principle cyclic, i.e., it starts and ends up in the same conformation ($^4C_1 \rightarrow [^4H_3/E_3]^\ddagger \rightarrow {}^4C_1$). This is in agreement with the conformations of the experimental structures of the MC and GEI of NpAS (both in $^4C_1$ conformation) (Mirza et al., 2001; Skov et al., 2002; Jensen et al., 2004). However, the computed itinerary differs from the expected one for α-glucosidases, for which a distorted GEI is

TABLE 2 | Calculated values of the most relevant catalytic distances (in Å) and puckering coordinates (in degrees) and their standard deviations along the glycosylation minimum energy pathway.

	ΔG	C1 – O_{Asp286}	C1 – O1	O1 – H_{Glu328}	O_{Glu328} – H_{Glu328}	θ	Conformation
MC	0.0	3.13 ± 0.09	1.46 ± 0.11	2.58 ± 0.07	1.00 ± 0.03	16 ± 7	$^4C_1$
1	14.2	3.08 ± 0.10	1.58 ± 0.12	1.25 ± 0.08	1.25 ± 0.08	27 ± 8	$^4C_1$
TS1	17.3	2.98 ± 0.11	1.98 ± 0.10	1.05 ± 0.05	1.61 ± 0.12	37 ± 8	$^4H_3/^4C_1$
2	14.4	2.72 ± 0.14	2.79 ± 0.12	1.02 ± 0.04	1.72 ± 0.13	55 ± 8	E_3
3	11.7	2.14 ± 0.17	2.89 ± 0.12	1.02 ± 0.04	1.71 ± 0.12	57 ± 8	$^4H_3$
GEI	−1.5	1.46 ± 0.05	3.18 ± 0.11	1.02 ± 0.04	1.78 ± 0.13	21 ± 8	$^4C_1$

FIGURE 5 | Reaction mechanism of *Np*AS obtained in this work (MC, Michaelis complex; TS, transition state; GEI, glycosyl-enzyme intermediate; P, reaction products). Hydrogen atoms have been also omitted, except the ones attached to heteroatoms. For the sake of clarity, the leaving β-fructose molecule has been omitted from the deglycosylation reaction (bottom panels).

FIGURE 6 | Evolution of the conformation of the *-1* sugar (α-glucosyl residue) during the glycosylation reaction.

expected ($^4C_1 \rightarrow$ [$^4H_3$]$^{\ddagger} \rightarrow$ $^1S_3$) (Davies et al., 2012; Speciale et al., 2014). As we will see in the next section, both views can be reconciled if we consider the reorganization of the active site required to start the deglycosylation reaction.

Deglycosylation Reaction

As shown above, the glycosylation reaction leads to an active site configuration in which the *-1* sugar is in a $^4C_1$ conformation.

This configuration is not suitable for the second step of the double displacement reaction for two main reasons. First, the leaving group (C1-O$_{Asp286}$) is not in axially oriented, which makes the S$_N$2 reaction unfavorable. Second, there is no water molecule in the vicinity of the anomeric carbon. The nearest water molecule remains at ∼6 Å from the anomeric carbon in the simulation (**Figure 7**, left panel), forming hydrogen bond interactions with the acid/base residue (Glu328) and an aspartate residue (Asp393). This water molecule is the best candidate to act as a nucleophile in the deglycosylation process, but it is still too far and not well oriented to attack the anomeric carbon. Most likely, there is an energy barrier to bring the water molecule to a reactive configuration.

To obtain the reactive configuration for the deglycosylation reaction, we selected the above water molecule and approached it to the anomeric carbon with the metadynamics algorithm using the O$_{water}$-C1 distance as collective variable. In order to allow the C1-Asp286 bond to adopt an axial orientation (necessary for the water nucleophilic attack), a second collective variable that accounts for the hydrogen bond between Asp286 and the 6-OH was used. This simulation facilitated the identification of a configuration of the glycosyl-enzyme intermediate in which the water molecule is at ≈ 3.5 Å from C1 (**Figure 7**, right panel). The position of the water molecule is stabilized by hydrogen bond interactions with the 2-OH of the *-1* sugar, Asp393, and the Glu328 catalytic residue. The new configuration, hereafter

FIGURE 7 | Reorganization of the active site of *Np*AS in the GEI state as the catalytic water approaches the anomeric carbon. The leaving β-fructose (product of the glycosylation reaction) is not shown for clarity.

FIGURE 8 | (A) Free energy landscape (FEL) of the deglycosylation reaction, obtained from QM/MM metadynamics simulations with three CVs. **(B)** Two-dimensional projection of the FEL into the collective variables CV1 and CV2.

named GEI*, is separated from the initial GEI by a small free energy barrier (4.7 kcal/mol; **Figure S5**). Remarkably, the *-1* sugar at GEI* is preactivated for catalysis, as it exhibits a distorted conformation (E_3) with a pseudo-axial orientation of the leaving group. Moreover, the water molecule is well oriented for nucleophilic attack, as it has the oxygen atom lone pairs pointing toward the anomeric carbon. Therefore, this configuration is the one we chose to start the modeling of the deglycosylation reaction. Interestingly, a distorted conformation of the GEI of lysozyme, a family 22 β-GH, was recently described by the Mulholland group based on QM(SCC-DFTB)/MM calculations (Limb et al., 2019). However, the GEI* states found here differs from the one of (Limb et al., 2019) in that the nucleophile carboxylate group exhibits the usual syn configuration with respect to the C1-O_{Asp286} bond ($\tau_{C1-O-C-O} \approx 0°$). Differences in the substrate, active site and reaction stereochemistry might be the reason of the discrepancy.

The deglycosylation reaction was modeled using three collective variables that take into account the cleavage of the glycosyl-enzyme covalent bond (C1-O_{Asp286}), the attack of the water oxygen to the anomeric carbon (C1-O_{wat}), and the deprototonation of the water molecule by Glu328. The FEL reconstructed from the metadynamics simulation, shown in **Figure 8**, shows two minima corresponding to the activated glycosyl-enzyme intermediate (GEI*) and the hydrolysis products (P), which are 7.6 kcal/mol more stable. The transition state (TS2) is 13.3 kcal/mol higher in energy with respect to the GEI* state, consistent with deglycosylation not being rate limiting.

The structure of the active site and the *-1* sugar conformation along the deglycosylation reaction pathway is shown in **Figure 5** (bottom panels) and **Figure 9**, whereas **Table 3** lists the evolution of the most important distances. The deglycosylation reaction begins by cleavage of the glycosyl-Asp286 bond, followed by attack of the water molecule on the anomeric carbon, while the water forms a tight hydrogen bond with the acid/base residue (Glu328). The system overcomes the transition state (TS2) and the conformation of the *-1* sugar changes to an E_3 envelope. Afterwards, the Glu328 abstracts a proton from the catalytic water, the covalent bond between the water molecule and the anomeric carbon forms and the -1 sugar adopts a $^4C_1$

FIGURE 9 | Evolution of the conformation of the *-1* sugar (α-glucosyl residue) during the deglycosylation reaction.
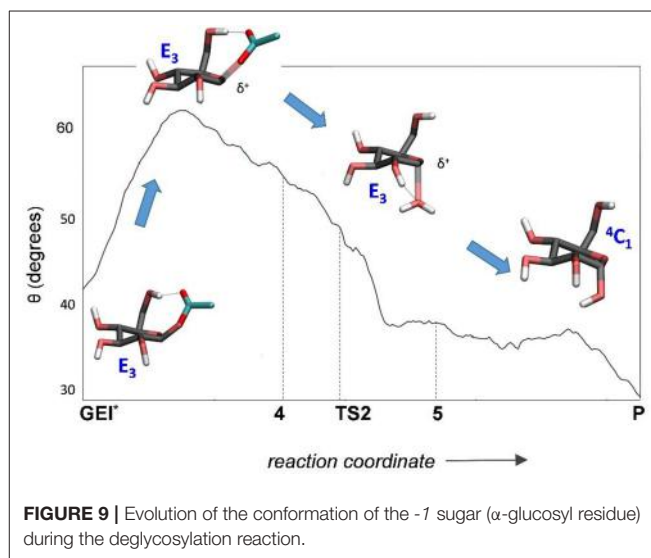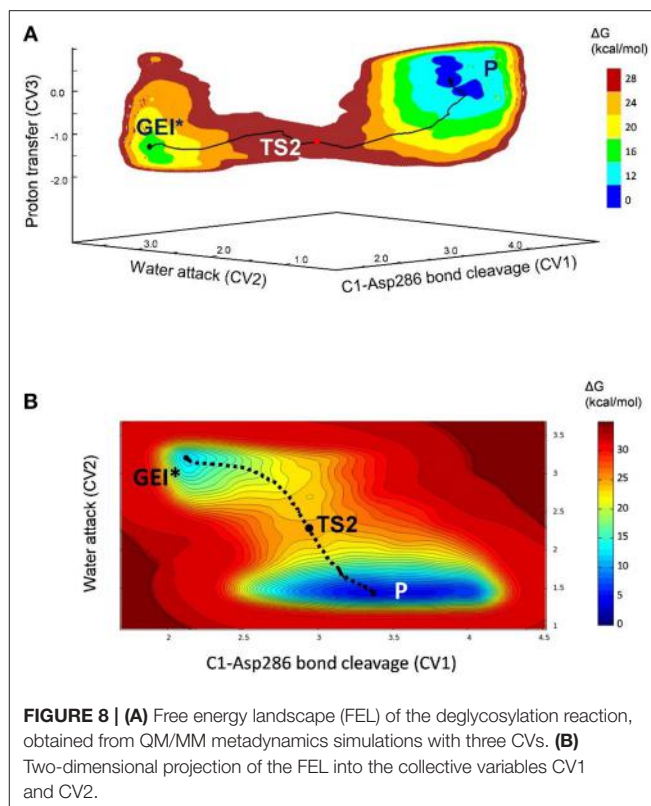
conformation. Therefore, the conformational itinerary of the deglycosylation reaction of *Np*AS can be described as $E_3 \rightarrow [E_3]^{\ddagger} \rightarrow {}^4C_1$.

Summary and Conclusions

By means of QM(DFT)/MM metadynamics we have unraveled the reaction mechanism of α-amylosucrase during hydrolysis, in particular the conformational dynamics of the glycosyl-enzyme

TABLE 3 | Calculated values of the most relevant catalytic distances (in Å) and their standard deviations along the deglycosylation minimum energy pathway.

	ΔG	$C1 - O_{Asp286}$	$C1 - O_{wat}$	$H-O_{wat}$	$H'_{wat}-O_{Glu328}$	θ	Conformation
GEI*	0.0	2.14 ± 0.09	3.19 ± 0.11	1.02 ± 0.04	1.65 ± 0.10	42 ± 11	E_3
4	11.0	2.81 ± 0.12	2.50 ± 0.12	1.06 ± 0.06	1.48 ± 0.10	52 ± 9	E_3
TS2	13.3	2.81 ± 0.12	2.23 ± 0.12	1.05 ± 0.05	1.53 ± 0.11	48 ± 7	E_3
5	3.2	3.11 ± 0.12	1.55 ± 0.12	1.11 ± 0.07	1.42 ± 0.08	38 ± 12	$E_3/{}^4C_1$
P	−7.6	3.38 ± 0.12	1.44 ± 0.10	2.19 ± 0.11	1.02 ± 0.04	30 ± 15	${}^4C_1$

intermediate (GEI), for which structural studies have assigned a ${}^4C_1$ conformation. Our results show that the glycosylation reaction, assisted by the catalytic residues Asp286 (general base) and Glu328 (acid/base) features a dissociative transition state, in which the α-glucosyl residue at the -1 subsite adopts a ${}^4H_3$ conformation. The GEI intermediate adopts a ${}^4C_1$ conformation, thus being not preactivated for catalysis. However, the active site is very dynamic and it can easily evolve toward another configuration, in which the -1 sugar adopts a reactive E_3 conformation, with a pseudo-axial orientation of the leaving group, as the catalytic water enters the active site. The two conformations can be viewed as two different states of the GEI intermediate, dry or wet. We conclude that the catalytic itinerary of amylosucrase for the Michaelis complex → TS → intermediate enzymatic half reaction could be either described indistinctly as ${}^4C_1 \rightarrow [{}^4H_3]^{\ddagger} \rightarrow {}^4C_1$ (cyclic itinerary) or ${}^4C_1 \rightarrow [{}^4H_3]^{\ddagger} \rightarrow E_3$ (linear itinerary). This reconciles the results obtained by X-ray, with an undistorted GEI (${}^4C_1$), with the expected conformation of α-glucosidases (distorted GEI with a conformation near ${}^1S_3$).

COMPUTATIONAL DETAILS

Model Building

The initial structure of the *Np*AS-sucrose complex was taken from the Michaelis complex structure of the Asp328Asn mutant (PDB 1JGI) (Mirza et al., 2001). The protonation states and hydrogen atom positions of all histidine residues were selected based on their hydrogen bond network (specifically, histidines 39, 173, 233, 332, 370, 392, 414, 463, 512, and 540 were considered neutral with the proton at N_ε, histidines 192, 306, 565 and 591 were neutral with a proton at N_δ and histidines 377, 382, and 601 were protonated). All Asp and Glu residues were taken as deprotonated (i.e., negatively charged) except Glu328 (the acid/base residue). The system was solvated with 20725 water molecules and 16 sodium ions were added to achieve neutrality of the protein structure, forming a rectangular box with dimensions of 89.1 Å × 106.4 Å × 81.1 Å.

Classical MD Simulations

Classical MD simulations of the *Np*AS-sucrose complex at room temperature were performed using the AMBER11 software (Case et al., 2010). The force-fields ff99SB (Hornak et al., 2006), GLYCAM06 (Kirschner et al., 2008) and TIP3P (Jorgensen et al., 1983) were used for the protein, sucrose substrate and solvent water, respectively. The system was equilibrated in several

steps. First, all water molecules were relaxed with a gradient minimizer, holding the protein, and substrate fixed. Next, the whole system was allowed to relax. To gradually reach the desired temperature of 300 K, spatial constraints were initially added to the interactions between the protein and the substrate, while water molecules and sodium ions were allowed to move freely. The constraints were then removed and the whole system was allowed to reach the desired temperature. During all the process, the acid/base residue (Glu328) was very mobile, with the proton position evolving from cis to trans conformations and vice versa, thus its relative position with respect to the sugar was controlled with a smooth restraint (force constant of 5 kcal/mol/Å^2). The simulation was pursued for 100 ps at constant pressure, allowing the cell volume to evolve, until the density stabilized (~1.06 g/cm^3). The MD simulation was extended to 15 ns at constant volume without restraints until the system had reached equilibrium. A timestep of 1 fs was used, increasing it to 2 fs during the last 14 ns (using SHAKE). The −1 sugar conformation evolved from ${}^4C_1$ to $B_{3,O}$ and returned to the original ${}^4C_1$ conformation during the last 7 ns of the simulation. Analysis of the trajectories was carried out by using standard tools of AMBER and VMD (Humphrey et al., 1996). The root-mean-square-deviation (RMSD) of the protein backbone atoms with respect to the crystal structure was stabilized around 1.4 Å in the equilibrated structure. One snapshot from the last 0.5 ns of simulation was taken as starting structure for the subsequent QM/MM MD simulations.

QM/MM MD Simulations

QM/MM MD simulations were performed with the CPMD software (Car and Parrinello, 1985), using the QM/MM interface developed by Laio et al. (2002). The QM region was considered as follows: (i) the whole sucrose molecule for the simulations of the conformational free energy landscape (FEL) of the -1 sugar (α-glucosyl); (ii) the sucrose, the acid/base residue (Glu328, capped at the C_β) and the nucleophile residue (Asp286, capped at the C_α) for the simulation of the glycosylation reaction; (iii) same as (ii) plus the catalytic water, for the simulation of the deglycosylation reaction. In all cases, the frontier atoms between QM and MM region were described using pseudopotential carbon link atoms. The fictitious electronic mass of the Car-Parrinello Lagrangian was taken as 600 au and the timestep was set at 0.12 fs in all CPMD simulations. All systems were enclosed in an isolated cubic box of 12.0 Å × 12.0 Å × 12.0 Å, using a fictitious electron mass of 700 au and a time step of 0.12 fs. The Kohn-Sham orbitals were expanded in a plane wave (PW) basis set

with a kinetic energy cutoff of 70 Ry. Ab initio pseudopotentials generated within the Troullier-Martins scheme were employed (Troullier and Martins, 1991). The Perdew, Burke and Ernzerhoff generalized gradient-corrected approximation (PBE) (Perdew et al., 1996) was selected in view of its good performance in previous work on isolated sugars (Biarnés et al., 2007; Marianski et al., 2016), glycosidases (Jin et al., 2016) and glycosyltransferases (Bilyard et al., 2018).

QM/MM Metadynamics Simulations

QM/MM metadynamics (Laio and Parrinello, 2002; Barducci et al., 2011) simulations were performed to characterize the conformational FEL of the α-glucosyl residue of sucrose in the active site of NpAS and to simulate the different steps of the enzymatic reaction. The following collective variables were used: (i) Conformational FEL: the Cremer-Pople puckering coordinates (Cremer and Pople, 1975) phi and theta (φ, θ) of the α-glucosyl unit, following the methodology previously used in our group to rationalize and predict catalytic itineraries of GHs (Biarnés et al., 2007; Ardèvol and Rovira, 2015; Iglesias-Fernandez et al., 2015): (ii) glycosylation reaction: three collective variables representing the proton transfer (CV1), the nucleophilic attack (CV2) and the glycosidic bond cleavage (CV3) (**Figure S3**); (iii) deglycosylation reaction: three collective variables representing the covalent enzyme-substrate interaction (CV1), the water attack (CV2) and the proton transfer (CV3) were used (**Figure S3**). The metadynamics algorithm (Laio and Parrinello, 2002; Barducci et al., 2011), provided by the Plumed 2 plugin (Tribello et al., 2014), was used to explore the conformational free energy landscape of the systems. The height/width of the Gaussian terms was tested according to the oscillations of the CVs in free dynamics. In the case of the conformational FEL simulation, the height/width of the Gaussian terms was set at 0.75 kcal/mol/0.10 Å and a new Gaussian-like potential was added every 400 MD steps. The simulation of the conformational FEL was stopped once energy differences between the two wells ($^4C_1$ and $B_{3,O}$) were maintained (1971 Gaussian terms were added), which was further tested by a time-independent free energy estimator (Tiwary and Parrinello, 2015). The error in the region of the two relevant minima and the pathway interconnecting them was $\leq$ 1.2 kcal/mol (**Figure S6**).

In the case of the glycosylation reaction, the height/width of the Gaussian terms was set at 1 kcal·mol^{-1}/0.20 Å (CV1) and 1 kcal/mol/0.10 Å (CV2 and CV3) and a new Gaussian-like potential was added every 300 MD steps. Walls for each CV at appropriate distances were used to reduce the FEL

space to the chemical event. For the deglycosylation reaction, values of 1 kcal/mol/ 0.10 Å (height/width) and 250 MD steps (deposition time) were used. Metadynamics simulations were stopped after one crossing over the transition state (**Figure S7**), as recommended for chemical reactions (Ensing et al., 2005). Previous work on carbohydrate-active enzymes shows that the error associated to the metadynamics is <1 kcal/mol using this criteria (Raich et al., 2016). The total number of Gaussian terms added was 2284 (glycosylation reaction) and 4947 (deglycosylation reaction). The reaction coordinate was taken from the minimum free energy pathway, computed according to the intrinsic reaction coordinate method (Fukui, 1981). Structures at a given point along the reaction coordinate were taken from averages over a small region defined by CV1 ± 0.2, CV2 ± 0.2, CV3 ± 0.2 Å and were used for analysis.

REFERENCES

Albenne, C., Skov, L. K., Mirza, O., Gajhede, M., Feller, G., D'amico, S., et al. (2004). Molecular basis of the amylose-like polymer formation catalyzed by Neisseria polysaccharea amylosucrase. J. Biol. Chem. 279, 726–734. doi: 10.1074/jbc.M309891200

Alonso-Gil, S., Males, A., Fernandes, P. Z., Williams, S. J., Davies, G. J., and Rovira, C. (2017). Computational design of experiment unveils the conformational

reaction coordinate of GH125 alpha-mannosidases. J. Am. Chem. Soc. 139, 1085–1088. doi: 10.1021/jacs.6b11247

Ardèvol, A., and Rovira, C. (2015). Reaction mechanisms in carbohydrate-active enzymes: glycoside hydrolases and glycosyltransferases. Insights from ab initio quantum mechanics/molecular mechanics dynamic simulations. J. Am. Chem. Soc. 137, 7528–7547. doi: 10.1021/jacs.5b01156

Artola, M., Wu, L., Ferraz, M. J., Kuo, C. L., Raich, L., Breen, I. Z., et al. (2017). 1,6-cyclophellitol cyclosulfates: a new class of irreversible

glycosidase inhibitor. *ACS Cent. Sci.* 3, 784–793. doi: 10.1021/acscentsci.7b00214

Barducci, A., Bonomi, M., and Parrinello, M. (2011). Metadynamics. *Wiley Interdisc. Rev. Comput. Mol. Sci.* 1, 826–843. doi: 10.1002/wcms.31

Beenakker, T. J. M., Wander, D. P. A., Offen, W. A., Artola, M., Raich, L., Ferraz, M. J., et al. (2017). Carba-cyclophellitols are neutral retaining-glucosidase inhibitors. *J. Am. Chem. Soc.* 139, 6534–6537. doi: 10.1021/jacs.7b01773

Biarnés, X., Ardèvol, A., Iglesias-Fernández, J., Planas, A., and Rovira, C. (2011). Catalytic itinerary in 1,3-1,4-beta-glucanase unraveled by QM/MM metadynamics. Charge is not yet fully developed at the oxocarbenium ion-like transition state. *J. Am. Chem. Soc.* 133, 20301–20309. doi: 10.1021/ja207113e

Biarnés, X., Ardèvol, A., Planas, A., Rovira, C., Laio, A., and Parrinello, M. (2007). The conformational free energy landscape of beta-D-glucopyranose. Implications for substrate preactivation in beta-glucoside hydrolases. *J. Am. Chem. Soc.* 129, 10686–10693. doi: 10.1021/ja068411o

Biarnés, X., Nieto, J., Planas, A., and Rovira, C. (2006). Substrate distortion in the Michaelis complex of Bacillus 1,3-1,4-beta-glucanase. Insight from first principles molecular dynamics simulations. *J. Biol. Chem.* 281, 1432–1441. doi: 10.1074/jbc.M507643200

Bilyard, M. K., Bailey, H. J., Raich, L., Gafitescu, M. A., Machida, T., Iglesias-Fernandez, J., et al. (2018). Palladium-mediated enzyme activation suggests multiphase initiation of glycogenesis. *Nature* 563, 235–240. doi: 10.1038/s41586-018-0644-7

Caner, S., Zhang, X., Jiang, J., Chen, H. M., Nguyen, N. T., Overkleeft, H., et al. (2016). Glucosyl epi-cyclophellitol allows mechanism-based inactivation and structural analysis of human pancreatic alpha-amylase. *FEBS Lett.* 590, 1143–1151. doi: 10.1002/1873-3468.12143

Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* 37, D233–238. doi: 10.1093/nar/gkn663

Car, R., and Parrinello, M. (1985). Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.* 55, 2471–2474. doi: 10.1103/PhysRevLett.55.2471

Case, D. A., Darden, T. A., Cheatham, T. E., Simmerling, C., Wang, J., Duke, R., et al. (2010). *AMBER 11.* San Francisco, CA: University of California.

Cremer, D., and Pople, J. A. (1975). General definition of ring puckering coordinates. *J. Am. Chem. Soc.* 97, 1354–1358. doi: 10.1021/ja00839a011

Davies, G. J., Planas, A., and Rovira, C. (2012). Conformational analyses of the reaction coordinate of glycosidases. *Acc. Chem. Res.* 45, 308–316. doi: 10.1021/ar2001765

Ensing, B., Laio, A., Parrinello, M., and Klein, M. L. (2005). A recipe for the computation of the free energy barrier and the lowest free energy path of concerted reactions. *J. Phys. Chem. B* 109, 6676–6687. doi: 10.1021/jp045571i

Fujimoto, Z., Takase, K., Doui, N., Momma, M., Matsumoto, T., and Mizuno, H. (1998). Crystal structure of a catalytic-site mutant alpha-amylase from *Bacillus subtilis* complexed with maltopentaose. *J. Mol. Biol.* 277, 393–407. doi: 10.1006/jmbi.1997.1599

Fukui, K. (1981). The path of chemical reactions - the IRC approach. *Acc. Chem. Res.* 14, 363–368. doi: 10.1021/ar00072a001

Gamblin, D. P., Scanlan, E. M., and Davis, B. G. (2009). Glycoprotein synthesis: an update. *Chem. Rev.* 109, 131–163. doi: 10.1021/cr078291i

Guthrie, R. D., and Jencks, W. P. (1989). IUPAC recommendations for the representation of reaction mechanisms. *Acc. Chem. Res.* 22, 343–349. doi: 10.1021/ar00166a001

Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* 65, 712–725. doi: 10.1002/prot.21123

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual molecular dynamics. *J. Mol. Graph.* 14, 33–38. doi: 10.1016/0263-7855(96)00018-5

Iglesias-Fernandez, J., Raich, L., Ardèvol, A., and Rovira, C. (2015). The complete conformational free-energy landscape of β-xylose reveals a two-fold catalytic itinerary for β-xylanases. *Chem. Sci.* 6, 1167–1177. doi: 10.1039/C4SC02240H

Jensen, M. H., Mirza, O., Albenne, C., Remaud-Simeon, M., Monsan, P., Gajhede, M., et al. (2004). Crystal structure of the covalent intermediate

of amylosucrase from *Neisseria polysaccharea*. *Biochemistry* 43, 3104–3110. doi: 10.1021/bi0357762

Jin, Y., Petricevic, M., John, A., Raich, L., Jenkins, H., Portela De Souza, L., et al. (2016). A beta-mannanase with a lysozyme-like fold and a novel molecular catalytic mechanism. *ACS Cent. Sci.* 2, 896–903. doi: 10.1021/acscentsci.6b00232

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935. doi: 10.1063/1.445869

Kirschner, K. N., Yongye, A. B., Tschampel, S. M., Gonzalez-Outeirino, J., Daniels, C. R., Foley, B. L., et al. (2008). GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.* 29, 622–655. doi: 10.1002/jcc.20820

Koshland, D. E. (1953). Stereochemistry and the mechanism of enzymatic reactions. *Biol. Rev. Camb. Philos. Soc.* 28, 416–436. doi: 10.1111/j.1469-185X.1953.tb01386.x

Laio, A., and Parrinello, M. (2002). Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* 99, 12562–12566. doi: 10.1073/pnas.202427399

Laio, A., Vandevondele, J., and Rothlisberger, U. (2002). A hamiltonian electrostatic coupling scheme for hybrid Car-Parrinello molecular dynamics simulations. *J. Chem. Phys.* 116, 6941–6947. doi: 10.1063/1.1462041

Li, Y. K., Chir, J., and Chen, F. Y. (2001). Catalytic mechanism of a family 3 beta-glucosidase and mutagenesis study on residue Asp-247. *Biochem. J.* 355, 835–840. doi: 10.1042/bj3550835

Limb, M.,a,.L., Suardiaz, R., Grant, I. M., and Mulholland, A.J. (2019). Quantum mechanics/molecular mechanics simulations show saccharide distortion is required for reaction in hen egg-white lysozyme. *Chem. Eur. J.* 25, 764–768. doi: 10.1002/chem.201805250

Marianski, M., Supady, A., Ingram, T., Schneider, M., and Baldauf, C. (2016). Assessing the accuracy of across-the-scale methods for predicting carbohydrate conformational energies for the examples of glucose and alpha-maltose. *J. Chem. Theory Comput.* 12, 6157–6168. doi: 10.1021/acs.jctc.6b00876

Mirza, O., Skov, L. K., Remaud-Simeon, M., Potocki De Montalk, G., Albenne, C., Monsan, P., et al. (2001). Crystal structures of amylosucrase from *Neisseria polysaccharea* in complex with D-glucose and the active site mutant Glu328Gln in complex with the natural substrate sucrose. *Biochemistry* 40, 9032–9039. doi: 10.1021/bi010706l

Mirza, O., Skov, L. K., Sprogoe, D., Van Den Broek, L. A., Beldman, G., Kastrup, J. S., et al. (2006). Structural rearrangements of sucrose phosphorylase from Bifidobacterium adolescentis during sucrose conversion. *J. Biol. Chem.* 281, 35576–35584. doi: 10.1074/jbc.M605611200

Pauly, M., and Keegstra, K. (2008). Cell-wall carbohydrates and their modification as a resource for biofuels. *Plant J.* 54, 559–568. doi: 10.1111/j.1365-313X.2008.03463.x

Perdew, J. P., Burke, K., and Ernzerhof, M. (1996). Generalized gradient approximation made simple. *Phys. Rev. Lett.* 77, 3865–3868. doi: 10.1103/PhysRevLett.77.3865

Pinto, G. P., Bras, N. F., Perez, M. A., Fernandes, P. A., Russo, N., Ramos, M. J., et al. (2015). Establishing the catalytic mechanism of human pancreatic alpha-amylase with QM/MM methods. *J. Chem. Theory Comput.* 11, 2508–2516. doi: 10.1021/acs.jctc.5b00222

Potocki de Montalk, G., Remaud-Simeon, M., Willemot, R. M., Sarcabal, P., Planchot, V., and Monsan, P. (2000). Amylosucrase from *Neisseria polysaccharea*: novel catalytic properties. *FEBS Lett.* 471, 219–223. doi: 10.1016/S0014-5793(00)01406-X

Raich, L., Nin-Hill, A., Ardevol, A., and Rovira, C. (2016). Enzymatic cleavage of glycosidic bonds: strategies on how to set up and control a QM/MM metadynamics simulation. *Methods Enzymol.* 577, 159–183. doi: 10.1016/bs.mie.2016.05.015

Rye, C. S., and Withers, S. G. (2000). Glycosidase mechanisms. *Curr. Opin. Chem. Biol.* 4, 573–580. doi: 10.1016/S1367-5931(00)00135-6

Schramm, V. L., and Shi, W. (2001). Atomic motion in enzymatic reaction coordinates. *Curr. Opin. Struct. Biol.* 11, 657–665. doi: 10.1016/S0959-440X(01)00269-X

Skov, L. K., Mirza, O., Sprogoe, D., Dar, I., Remaud-Simeon, M., Albenne, C., et al. (2002). Oligosaccharide and sucrose complexes of amylosucrase. Structural

implications for the polymerase activity. *J. Biol. Chem.* 277, 47741–47747. doi: 10.1074/jbc.M207860200

Speciale, G., Thompson, A. J., Davies, G. J., and Williams, S. J. (2014). Dissecting conformational contributions to glycosidase catalysis and inhibition. *Curr. Opin. Struct. Biol.* 28C, 1–13. doi: 10.1016/j.sbi.2014.06.003

Tiwary, P., and Parrinello, M. (2015). A time-independent free energy estimator for metadynamics. *J. Phys. Chem. B* 119, 736–742. doi: 10.1021/jp504920s

Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C., and Bussi, G. (2014). PLUMED 2: new feathers for an old bird. *Comput. Phys. Commun.* 185, 604–613. doi: 10.1016/j.cpc.2013.09.018

Troullier, N., and Martins, J. L. (1991). Efficient pseudopotentials for plane-wave calculations. *Phys. Rev. B Condens. Matter.* 43, 1993–2006. doi: 10.1103/PhysRevB.43.1993

Zechel, D. L., and Withers, S. G. (2000). Glycosidase mechanisms: anatomy of a finely tuned catalyst. *Acc. Chem. Res.* 33, 11–18. doi: 10.1021/ar970172

Zhang, R., Li, C., Williams, L. K., Rempel, B. P., Brayer, G. D., and Withers, S. G. (2009). Directed "*in situ*" inhibitor elongation as a strategy to structurally characterize the covalent glycosyl-enzyme intermediate of human pancreatic alpha-amylase. *Biochemistry* 48, 10752–10764. doi: 10.1021/bi901400p

Benchmark of Density Functionals for the Calculation of the Redox Potential of Fe^{3+}/Fe^{2+} Within Protein Coordination Shells

Risnita Vicky Listyarini[†], Diana Sofia Gesto[†], Pedro Paiva, Maria João Ramos and Pedro Alexandrino Fernandes*

UCIBIO-REQUIMTE, Departamento de Química e Bioquímica, Faculdade de Ciências, Universidade do Porto, Porto, Portugal

Iron is a very important transition metal often found in proteins. In enzymes specifically, it is often found at the core of reaction mechanisms, participating in the reaction cycle, more often than not in oxidation/reduction reactions, where it cycles between its most common Fe(III)/Fe(II) oxidation states. QM and QM/MM computational methods that study these catalytic reaction mechanisms mostly use density functional theory (DFT) to describe the chemical transformations. Unfortunately, density functional is known to be plagued by system-specific and property-specific inaccuracies that cast a shadow of uncertainty over the results. Here we have modeled 12 iron coordination complexes, using ligands that represent amino acid sidechains, and calculated the accuracy with which the most common density functionals reproduce the redox properties of the iron complexes (specifically the electronic component of the redox potential at 0 K, $\Delta E_{elec}^{Fe^{3+}/Fe^{2+}}$), using the same property calculated with CCSD(T)/CBS as reference for the evaluation. A number of hybrid and hybrid-meta density functionals, generally with a large % of HF exchange (such as BB1K, mPWB1K, and mPW1B95) provided systematically accurate values for $\Delta E_{elec}^{Fe^{3+}/Fe^{2+}}$, with MUEs of ∼2 kcal/mol. The very popular B3LYP density functional was found to be quite precise as well, with a MUE of 2.51 kcal/mol. Overall, the study provides guidelines to estimate the inaccuracies coming from the density functionals in the study of enzyme reaction mechanisms that involve an iron cofactor, and to choose appropriate density functionals for the study of the same reactions.

Keywords: redox potencial, DFT, benchmaking, iron, quantum-chemical calculations

INTRODUCTION

Iron plays an important role in enzyme catalysis. It is present in a wide range of proteins as a cofactor, and intervenes in many biochemical processes, particularly in oxygen transport, and electron transfer. Iron can be often found incorporated in heme prosthetic groups, or it can also occur individually, as an ion cofactor in some metalloproteins, or, more unfrequently, together with sulfide, forming iron-sulfur clusters (Harding et al., 2010; Abbaspour et al., 2014).

One of the most important characteristics of iron is the fact that it can be found in two stable states of oxidation, Fe^{2+} and Fe^{3+}. It is this characteristic that allows iron to participate in a wide range of redox reactions (Broderick, 2001). In proteins, it is usually found in the active site,

coordinated with the side chains of amino acid residues and/or water molecules. Interestingly, the redox potential of the iron ion changes depending on the ligands it is complexed with. In enzymes, through evolution, the ligands to which the iron is coordinated were selected so that its redox potential became the most appropriate for the reaction catalyzed by it (Tamames and Ramos, 2010; Liu et al., 2014).

Iron is of extreme importance to almost all organisms. In adult humans, iron makes up about 0.005% of the total body weight (circa 4 g), most of which is complexed with the heme group in hemoglobin. Iron is also present in cytochromes, heme-containing proteins that are involved in many electron transfer reactions that occur in our body, from oxidative phosphorylation to the synthesis of hormones or the degradation of drugs (Abbaspour et al., 2014).

The reduction potential (or redox potential) of an atom/molecule is, by definition, a measurement of its ability to gain an electron, and therefore be reduced (Equation 1).

$$M^n + e^- \rightarrow M^{n-1} \qquad (1)$$
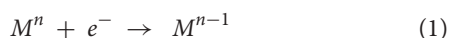
The reduction potential is not a fixed value for a given chemical entity, and can change depending on different factors, such as temperature, pH, structure, conformation, and electrostatic environment (Su et al., 2011; Ho et al., 2016). For this reason, the redox potentials of metal cofactors need to be calculated for each enzyme individually. In proteins, iron acts as both an electron donor and acceptor, depending on its oxidation state. When it is in the form of Fe(II), iron will mostly give up one of its electrons, thus being oxidized to Fe(III). On the other hand, when in the form of Fe(III), it can act as a oxidizing agent, gaining one electron, and changing its oxidation state to Fe(II). In some rarer cases, it is possible to find Fe^{4+} as a cofactor of some enzymes as well, but these cases will not be addressed in this paper.

Since iron is such an important metal in biological processes, with its redox potential changing greatly depending on the ligands coordinated to it, the study of these processes, and the accuracy of the final results, is also determined by how well we understand the redox chemistry of the iron ion. The measurement of reduction potentials in biological systems using experimental methodologies is a very complex and difficult task. The alternative is the use of high-level computational methods, which employ elaborate algorithms, equations, and approximations to simulate the system in question and achieve the most accurate result possible (Riley and Merz, 2007). However, it should be noted that in order to obtain these results one must employ methods that are extremely time consuming and CPU demanding, which is impractical for biological systems such as proteins, where the number of atoms is exceedingly high. It is possible to use less accurate methods when calculating redox potentials in such systems, but the problem in these cases is that, due to their limited accuracy, we might not know which one is the best for each chemical scenario. To solve this problem, we can benchmark the different methods, and compare these results to a reference value, ranking each method as more or less accurate.

The main focus of this work is, therefore, to perform a benchmark study of different computational methods and

evaluate their ability to simulate the iron ion in different arrangements, often found in proteins. We selected only density functional theory (DFT) (Thomas, 1927; Hohenberg and Kohn, 1964) methods because of their very good accuracy and applicability on large systems (they have a more favorable scaling with the system size than post-Hartree Fock methods). DFT methods are a good choice when it comes to simulating proteins (Uudsemaa and Tamm, 2003; Riley et al., 2007; Li et al., 2009). The problem comes when selecting which density functional to use for a specific case, since their performance can differ depending on the reaction, model, chemical system, etc. For this reason, benchmarking studies are gaining increasingly more attention, as they give us the ability to make a more informed decision on which DFT methods might be more accurate for the system we want to study.

For this benchmark study, in order to mimic the many complexes that are possible for an iron ion to have in the protein, we constructed 12 models, that span over different coordination numbers and ligands. The models try to mirror the various types of amino acid side chains (and water molecules) that can be most frequently found coordinating iron in proteins. Only one side chain was used in each model, to evaluate the individual effect of each side chain in the redox potential, albeit in proteins the iron coordination shell is made of many (up to six) side chains simultaneously. The coordination shells were completed with water molecules. Water was treated as a coordinating ligand as well.

For each of these models, we calculated the electronic component of the redox potential at 0 K ($\Delta E_{elec}^{Fe^{3+}/Fe^{2+}}$) using the different density functionals. We decided to measure only this component, instead of calculating the redox potential in its entirety for several reasons: this electronic component is the one that contributes the most for the overall potential; it is the only component that can be calculated with CCSD(T), which is the method to be used as a reference; it is the component in which the different density functionals show a larger variation in the final results. Furthermore, calculating only $\Delta E_{elec}^{Fe^{3+}/Fe^{2+}}$ is much easier and faster computationally, and the introduction of the other components in the calculation would render this study much more time consuming with not much improvement in the final results (Su et al., 2011; Ho et al., 2016).

Reference values for $\Delta E_{elec}^{Fe^{3+}/Fe^{2+}}$, calculated in each system to benchmark the accuracy of the density functionals, were calculated using two post-Hartree-Fock methods: the second order Moller-Plesset perturbation theory (MP2) method, and the single and double coupled cluster theory with perturbative triple correction [CCSD(T)]. The extrapolation energy at the complete basis set (CBS) limit was calculated using two different, widely used extrapolation schemes.

COMPUTATIONAL METHODS

Model Systems and Structure Optimization

The iron complexes found in proteins are quite diversified. Iron can make complexes with a different number of ligands, and, depending on these, the geometry of the complex will

also change. Furthermore, the side chains of the amino acid residues that interact and complex with iron have different functional groups.

In order to take all of these conditions into account in the benchmark, we decided to build 12 model systems, each mimicking a different coordination sphere. These conditions address most of the various coordination states, geometries and ligands that iron complexes can adopt. Our models comprise structures with one, two, four, or six coordinated ligands, which assume geometries of the type linear (for coordination numbers one and two), tetrahedral (for coordination number four), and octahedral (for coordination number six). As for the ligands, we first started with a set of structures in which the iron ion was coordinated exclusively to water molecules, having modeled one structure for each coordination number, with a total of four models. As for the remaining 8, we substituted one of the coordinated waters by one of the following ligands: methoxide (CH_3O^-), formate ($HCOO^-$), methanethiolate (CH_3S^-), and methylamine (NH_2CH_3). This choice of ligands reflects our objective of benchmarking the DFT methods in the context of proteins. Each of these entities represents the side chains of one amino acid that frequently participates in the formation of iron complexes in proteins: CH_3O^- mimics the side chain of serine, $CHOO^-$ that of glutamate and aspartate, CH_3S^- the side chain of cysteine and NH_2CH_3 that of lysine. The structures for all the models studied are represented on **Figure 1**.

After building each of the models, we optimized their geometries using the MP2/6-311+G(d,p) level of theory. One might argue that, for the sake of consistency, it would have been better to calculate $\Delta E_{elec}^{Fe^{3+}/Fe^{2+}}$ for each model using the minimum-energy geometries obtained with each functional we purpose to study. In reality, such study would demand a lot more computational time, for a minimal difference in the final results, as shown in previous studies (Ribeiro et al., 2010; Brás et al., 2011).

Furthermore, MP2 geometries are known to give good results without introducing biases toward any functional. All calculations were performed using the Gaussian 09 package (Frisch et al., 2009).

We optimized the geometries of each complex with both Fe(II) and Fe(III) oxidation states. It is known that, for each oxidation state, the iron ion can have two different spin states. For Fe(II), the high spin multiplicity is five and the low spin multiplicity is one. In the case of Fe(III), the high spin multiplicity is six, and the low one is two. We calculated which of these spin states gave the lowest energy for each oxidation state, and used that spin multiplicity for the remainder of our calculations. In both cases, the high spin multiplicity [five for Fe(II) and six for Fe(III)] was the one that resulted in the lowest energies (Riley and Merz, 2007) (data not shown). As a result, these were the spin states used throughout the following calculations. The $<S^2>$ values in each MP2 calculation were checked and compared to the expected values of 6.0 for Fe(II) complexes and 8.75 for Fe(III) complexes. The Gaussian 09 package includes an annihilation step to decrease the amount of spin contamination. For the Fe(II) complexes, we observed an average $<S>^2$ of 6.0079 and 6.0000 before and after annihilation, respectively. Regarding

the Fe(III) complexes, the average $<S^2>$ before and after the annihilation step were 8.7625 and 8.7501, respectively. Thus, the spin contamination is virtually null.

All calculations were made in gas-phase. As the purpose is to benchmark only and specifically the DFs, we avoided to introduce solvation, so that small differences in solvation would not introduce an unwanted contribution for differences in accuracy of the DFs that are being compared.

Calculation of the Reference $\Delta E_{elec}^{Fe^{3+}/Fe^{2+}}$ Values

In order to correctly benchmark and compare the diverse density functionals, we needed to obtain accurate values for $\Delta E_{elec}^{Fe^{3+}/Fe^{2+}}$ in each of our models.

For this study, we decided to calculate $\Delta E_{elec}^{Fe^{3+}/Fe^{2+}}$ using the very accurate CCSD(T) method (Bartlett and Purvis, 1978; Purvis and Bartlett, 1982; Pople et al., 1987), extrapolated to the CBS limit, and use these values as a reference. To this end, we carried out single point calculations using the MP2/aug-cc-pVXZ (X = 2, 3, 4) and CCSD(T)/aug-cc-pVDZ levels, in vacuum, for each optimized structure, and then we employed two different and well-known extrapolation methods to obtain the final values: one developed by Truhlar (1998) (scheme I) and another by Helgaker et al. (1997), Halkier et al. (1998) (scheme II). Both schemes begin with the extrapolation of the correlation energy to the CBS limit at the MP2 level. The difference between the correlation energies calculated using MP2 and CCSD(T) at the largest basis set possible (for this study, we used aug-cc-pVDZ) is then determined and added to the MP2/CBS value. The extrapolation of the Hartree-Fock (HF) energy is done separately from that of the correlation energy in both schemes as well.

To determine the MP2/CBS energy using scheme I, we employ Equation 2, using the results obtained from the calculations using MP2/aug-cc-pVDZ and MP2/aug-cc-pVTZ. α and β are constants taken from the literature ($\alpha = 4.93$ and $\beta = 2.13$).

$$E_{\frac{MP2}{CBS}} = \frac{3^\alpha}{3^\alpha - 2^\alpha}E_{\frac{HF}{TZ}} - \frac{2^\alpha}{3^\alpha - 2^\alpha}E_{\frac{HF}{DZ}} + \frac{3^\beta}{3^\beta - 2^\beta}E_{\frac{corr}{TZ}} - \frac{2^\beta}{3^\beta - 2^\beta}E_{\frac{corr}{DZ}}$$

(2)

As for scheme II, $E_{\frac{HF}{CBS}}$ is obtained by exponentially fitting the values obtained for the aug-cc-pVDZ, aug-cc-pVTZ, and aug-cc-pVQZ basis sets. $E_{\frac{corr}{CBS}}$ is calculated with Equation 3.

$$E_{\frac{corr}{CBS}} = \frac{4^3 E_{\frac{corr}{QZ}} - 3^3 E_{\frac{corr}{TZ}}}{4^3 - 3^3}$$

(3)

The HF energy is taken from Equation 2, whichever the scheme is used to extrapolate the correlation energy.

Benchmarking of the Density Functionals

We selected a set of 44 density functionals and tested their performance against the values for $\Delta E_{elec}^{Fe^{3+}/Fe^{2+}}$ obtained at the CCSD(T)/CBS level. This selection includes functionals belonging to different classes: 2 local density approximation (LDA) functionals, 7 generalized gradient approximation (GGA),
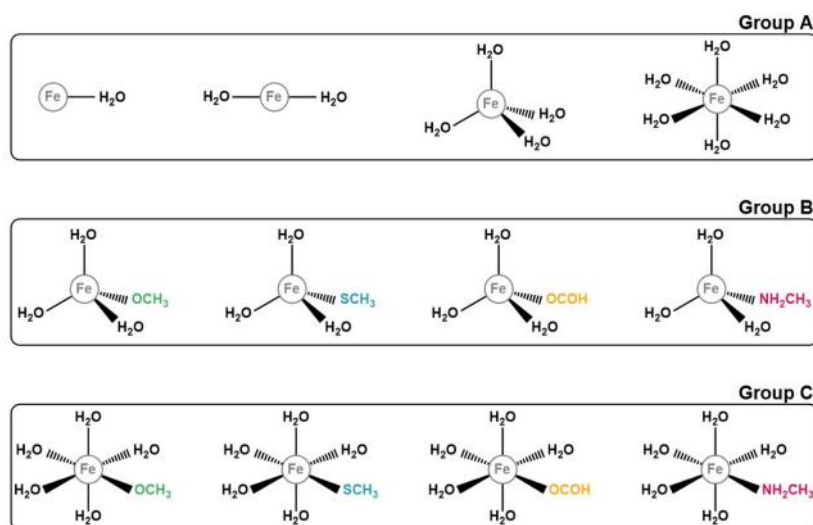
FIGURE 1 | Representation of the 12 iron complexes studied. They were divided in 3 groups according to their ligands for simplicity and for a better discussion. Group A comprises those structures which only have water molecules coordinated to the iron ion. Group B and C include the models with in which one of the ligands mimics an amino acid side chain. The difference between these two groups is their geometry: tetrahedral for group B and octahedral for group C.

5 meta-GGA (m-GGA), 12 hybrid-GGA (h-GGA), 10 hybrid-meta-GGA (hm-GGA), 4 double hybrid-GGA (hh-GGA), 1 non-separable gradient approximation (NGA), 1 hybrid-meta-NGA (hm-NGA), 1 meta-NGA (m-NGA), and 1 generalized gradient exchange (GGE) (**Table 1**). All of the benchmarking calculations were carried out using 6-311+G(2df,2p) basis set, which was selected due to the fact that, for large biological systems, it is the most CBS that can be generally used. Larger basis sets would render the calculations too difficult, making them extremely computationally demanding and time consuming. It is also worth noting that this basis set is frequently close to the DFT CBS limit, and any differences in relative energy (activation energy, reaction energy, redox potentials, etc.) from larger basis sets are usually in the tenths of a kilocalorie per mole. Another reason for the choice of the 6-311+G(2df,2p) basis set was because it is sufficiently large that it minimizes basis set truncation error within DFT, and so we can more accurately evaluate the performance of the functional without the interference from the basis set.

In order to calculate $\Delta E_{elec}^{Fe^{3+}/Fe^{2+}}$ with each of the DFT methods, we used the structures optimized at the MP2/6-311+G(d,p) level of theory, for both Fe(II) and Fe(III) and determined their energies through single point calculations. $\Delta E_{elec}^{Fe^{3+}/Fe^{2+}}$ was then calculated from the difference between each Fe(II)/Fe(III) pair for the corresponding complex. All DFT calculations were carried out using the Gaussian 09 suite (Frisch et al., 2009).

RESULTS AND DISCUSSION

Calculation of the Electronic Component of the Reduction Potential at the MP2/CBS Level

In order to obtain an accurate measure of $\Delta E_{elec}^{Fe^{3+}/Fe^{2+}}$ for each iron complex, after the initial geometry optimization of

the complexes we used high-level *ab initio* post-HF methods, with CBS extrapolation to eliminate any basis set truncation error. This was done by calculating these values first with the MP2 method, with aug-cc-pVXZ (X = 2, 3, and 4) with subsequent extrapolation to the MP2/CBS level. Extrapolation to this level was done using two different schemes, as described in the methods section of this paper. These values were used in conjunction with the CCSD(T) results (at the corresponding basis set level) in order to extrapolate the energies to the CCSD(T)/CBS level. In this manner, we ensure that we have reliable values for $\Delta E_{elec}^{Fe^{3+}/Fe^{2+}}$ for each of our 12 models, and that these values can be used as a reference to ascertain the precision of various density functionals in calculating the electronic energy contribution of Fe(II) and Fe(III) in each system. **Table 2** summarizes the results obtained for $\Delta E_{elec}^{Fe^{3+}/Fe^{2+}}$ at the MP2 level and the aug-cc-pVDZ, aug-cc-pVTZ, and aug-cc-pVQZ basis sets.

$\Delta E_{elec}^{Fe^{3+}/Fe^{2+}}$ changed by 2–3 kcal/mol when moving from aug-cc-pVDZ to the aug-cc-pVTZ basis set, and by a further ~2–3 kcal/mol when moving from aug-cc-pVTZ to the aug-cc-pVQZ basis set, emphasizing the need for the extrapolation for the CBS limit. As stated previously, we used two different schemes to determine the MP2/CBS energies. The first extrapolation scheme we used was scheme I, developed by D. Truhlar. **Table 3** shows the values obtained at the CBS limit using this scheme, as well as the differences between these energies and the corresponding ones at the aug-cc-pVXZ (X = 2, 3, and 4) basis sets.

Our results show that MP2/CBS energies obtained with scheme I are closer to those calculated with quadruple-zeta basis set than any of the other basis sets, as expected, and that the difference between the CBS and quadruple-zeta basis set is small, as can be confirmed by comparing the values for the mean signed error (MSE) and mean unsigned error (MUE) for each of them. We can also observe that the convergence of the reduction

TABLE 1 | List of all the density functionals tested in this wok.

LDA	GGA	h-GGA	hh-GGA
SPW91 (Hohenberg and Kohn, 1964; Kohn and Sham, 1965; Slater and Phillips, 1974; Perdew, 1986a; Perdew et al., 1991, 1992, 1996a)	BP86 (Perdew, 1986b; Becke, 1988)	B1LYP (Becke, 1988; Lee et al., 1988; Adamo and Barone, 1997)	B2GPPLYP (Becke, 1988; Lee et al., 1988; Grimme, 2006; Karton et al., 2008)
SVWN (Hohenberg and Kohn, 1964; Kohn and Sham, 1965; Slater and Phillips, 1974; Vosko et al., 1980)	BPBE (Becke, 1988; Perdew et al., 1996b)	B3LYP (Becke, 1988, 1996; Lee et al., 1988)	B2PLYP (Becke, 1988; Lee et al., 1988; Grimme, 2006)
hm-GGA	BPW91 (Becke, 1988; Perdew et al., 1991, 1992, 1993, 1996a; Burke et al., 1998)	B3P86 (Perdew, 1986b; Becke, 1988, 1996)	DSD-BLYP (Becke, 1988; Lee et al., 1988; Grimme, 2006; Kozuch et al., 2010)
BB1K (Becke, 1988, 1996; Zhao et al., 2004)	G96LYP (Lee et al., 1988; Gill, 1996; Adamo and Barone, 1998a)	B3PW91 (Becke, 1988, 1996; Perdew et al., 1991, 1992, 1993, 1996a; Burke et al., 1998)	mPW2PLYP (Lee et al., 1988; Perdew et al., 1991, 1992, 1993, 1996a; Adamo and Barone, 1998b; Burke et al., 1998; Schwabe and Grimme, 2006)
BMK (Boese and Martin, 2004)	HCTH407 (Hamprecht et al., 1998)	B97-1 (Hamprecht et al., 1998)	**NGA**
M05 (Zhao et al., 2005)	OLYP (Lee et al., 1988; Handy and Cohen, 2001; Hoe et al., 2001)	B97-2 (Wilson et al., 2001)	N12 (Peverati and Truhlar, 2012a)
M05-2X (Zhao et al., 2006; Zhao and Truhlar, 2008)	OPL (Perdew and Zunger, 1981; Handy and Cohen, 2001; Hoe et al., 2001)	B97-D3 (Grimme et al., 2011)	**m-NGA**
M06 (Zhao and Truhlar, 2007)	**m-GGA**	BhandH (Becke, 1988, 1993; Lee et al., 1988)	MN12-L (Peverati and Truhlar, 2012b)
M06-2X (Zhao and Truhlar, 2007)	M06-L (Zhao et al., 2006)	mPW1K (Perdew et al., 1991, 1992, 1993, 1996a; Adamo and Barone, 1998b; Burke et al., 1998; Lynch et al., 2000)	**hm-NGA**
M11 (Peverati and Truhlar, 2011a)	M11-L (Peverati and Truhlar, 2011b)	mPW1N (Perdew et al., 1991, 1992, 1993, 1996a; Adamo and Barone, 1998b; Burke et al., 1998; Kormos and Cramer, 2002)	MN12-SX (Peverati and Truhlar, 2012b,c)
mPW1B95 (Perdew et al., 1991, 1992, 1993, 1996a; Becke, 1996; Adamo and Barone, 1998b; Zhao and Truhlar, 2004)	mPWB95 (Perdew et al., 1991, 1992, 1993, 1996a; Becke, 1996; Adamo and Barone, 1998b)	PBE1PBE (Perdew et al., 1996b; Adamo and Barone, 1997)	**GGE**
mPW1K (Perdew et al., 1991, 1992, 1993, 1996a; Becke, 1996; Adamo and Barone, 1998b; Boese et al., 2003; Zhao and Truhlar, 2004)	OTPSS (Handy and Cohen, 2001; Hoe et al., 2001; Tao et al., 2003)	wB97X-D (Chai and Head-Gordon, 2008)	OVWN5 (Vosko et al., 1980; Handy and Cohen, 2001; Hoe et al., 2001)
TPSSh (Tao et al., 2003)	VSXC (Van Voorhis and Scuseria, 1998)		

potential with the basis set is slow, since at quadruple-zeta the values are still 1.40 kcal/mol (in average) away from the CBS limit.

Using scheme II, developed by T. Helgaker and co-workers, to extrapolate $\Delta E_{elec}^{Fe^{3+}/Fe^{2+}}$ to the MP2/CBS level, we obtained the results shown in **Table 4**. This scheme uses the energies calculated with the aug-cc-pVTZ and aug-cc-pVQZ basis sets for the extrapolation. Since these basis sets are more complete that the ones used in scheme I, it might be expected that the reduction potentials extrapolated using scheme II are more accurate.

Besides the $\Delta E_{elec}^{Fe^{3+}/Fe^{2+}}$ extrapolated to the MP2/CBS level using scheme II, **Table 4** also shows the difference between these values and those extrapolated using scheme I and calculated at the MP2/aug-cc-pVQZ level. From these results, we can observe that there is a slight difference between the scheme

II MP2/CBS and MP2/aug-cc-pVQZ energies, with an average of 0.88 kcal/mol. This indicates that, for these systems, the values calculated with quadruple-zeta basis sets are not accurate enough, and extrapolation to the CBS limit is indeed necessary. Furthermore, it is interesting to note that the difference between both extrapolation schemes is small (MUE = 0.52 kcal/mol), but not insignificant, especially considering that both of these methods are expected to yield very accurate results. Although in some cases this difference is small enough that it can be disregarded [0.024 kcal/mol for $Fe(H_2O)$ and 0.0830 kcal/mol for $Fe(H_2O)_3(CH_3S^-)$], in cases such as that of $Fe(H_2O)_3(CH_3O^-)$ and $Fe(H_2O)_5(CH_3O^-)$ it is actually somewhat relevant ($\sim$ 1.3 kcal/mol). Since scheme II uses larger basis sets to perform the extrapolation, we chose those energies for the remainder of the work, particularly in the extrapolation to the CCSD(T)/CBS

TABLE 2 | Results for $\Delta E_{elec}^{Fe^{3+}/Fe^{2+}}$ calculated at the MP2 level and the aug-cc-pVDZ, aug-cc-pVTZ, and aug-cc-pVQZ basis sets, for each model.

Model	$\Delta E_{elec}^{Fe^{3+}/Fe^{2+}}$ (kcal/mol)		
	MP2/aug-cc-pVDZ	MP2/aug-cc-pVTZ	MP2/aug-cc-pVQZ
Fe(H$_2$O)	−596.6545	−599.1266	−601.7583
Fe(H$_2$O)$_2$	−532.2910	−534.6568	−537.1668
Fe(H$_2$O)$_4$	−436.5451	−438.9066	−441.2415
Fe(H$_2$O)$_6$	−381.3428	−383.5432	−385.8931
Fe(H$_2$O)$_3$(CH$_3$O$^-$)	−271.8354	−274.9105	−276.9331
Fe(H$_2$O)$_3$(CH$_3$S$^-$)	−258.7914	−261.2471	−262.7309
Fe(H$_2$O)$_3$(NH$_2$CH$_3$)	−418.7231	−421.6494	−423.9952
Fe(H$_2$O)$_3$(HCOO$^-$)	−300.7855	−303.4670	−305.5606
Fe(H$_2$O)$_5$(CH$_3$O$^-$)	−246.0797	−248.7022	−250.6648
Fe(H$_2$O)$_5$(CH$_3$S$^-$)	−236.5792	−238.9616	−240.4693
Fe(H$_2$O)$_5$(NH$_2$CH$_3$)	−370.4211	−373.1151	−375.4922
Fe(H$_2$O)$_5$(HCOO$^-$)	−269.0788	−271.3712	−273.4479

TABLE 3 | $\Delta E_{elec}^{Fe^{3+}/Fe^{2+}}$ obtained at the MP2/CBS level, extrapolated using scheme I, for each model, and the differences between these new energies and the ones obtained in **Table 2**.

Model	$\Delta E_{elec}^{Fe^{3+}/Fe^{2+}}$ (kcal/mol)			
	Scheme I MP2/CBS	MP2/CBS-MP2/aug-cc-pVDZ	MP2/CBS-MP2/aug-cc-pVTZ	MP2/CBS-MP2/aug-cc-pVQZ
Fe(H$_2$O)	−602.8804	−6.23	−3.75	−1.12
Fe(H$_2$O)$_2$	−538.3929	−6.10	−3.74	−1.23
Fe(H$_2$O)$_4$	−442.5902	−6.05	−3.68	−1.35
Fe(H$_2$O)$_6$	−386.9860	−5.64	−3.44	−1.09
Fe(H$_2$O)$_3$(CH$_3$O$^-$)	−279.0221	−7.19	−4.11	−2.09
Fe(H$_2$O)$_3$(CH$_3$S$^-$)	−263.8079	−5.02	−2.56	−1.08
Fe(H$_2$O)$_3$(NH$_2$CH$_3$)	−425.3816	−6.66	−3.73	−1.39
Fe(H$_2$O)$_3$(HCOO$^-$)	−307.3227	−6.54	−3.86	−1.76
Fe(H$_2$O)$_5$(CH$_3$O$^-$)	−252.6365	−6.56	−3.93	−1.97
Fe(H$_2$O)$_5$(CH$_3$S$^-$)	−241.5428	−4.96	−2.58	−1.07
Fe(H$_2$O)$_5$(NH$_2$CH$_3$)	−376.6431	−6.22	−3.53	−1.15
Fe(H$_2$O)$_5$(HCOO$^-$)	−274.9742	−5.90	−3.60	−1.53
MSE	−	−6.09	−3.54	−1.40
MUE	−	6.09	3.54	1.40

TABLE 4 | $\Delta E_{elec}^{Fe^{3+}/Fe^{2+}}$ obtained at the MP2/CBS level, extrapolated using scheme II for each model, and the differences between these energies and those obtained using scheme I and at the MP2/aug-cc-pVQZ level.

Model	$\Delta E_{elec}^{Fe^{3+}/Fe^{2+}}$ (kcal/mol)		
	Scheme II MP2/CBS	Scheme I MP2/CBS – Scheme II MP2/CBS	Scheme II MP2/CBS - MP2/aug-cc-pVQZ
Fe(H$_2$O)	−602.8540	−0.0264	−1.0956
Fe(H$_2$O)$_2$	−538.2482	−0.1447	−1.0813
Fe(H$_2$O)$_4$	−442.1297	−0.4605	−0.8882
Fe(H$_2$O)$_6$	−386.5973	−0.3887	−0.7042
Fe(H$_2$O)$_3$(CH$_3$O$^-$)	−277.7082	−1.3139	−0.7751
Fe(H$_2$O)$_3$(CH$_3$S$^-$)	−263.7248	−0.0830	−0.9939
Fe(H$_2$O)$_3$(NH$_2$CH$_3$)	−424.9332	−0.4484	−0.9379
Fe(H$_2$O)$_3$(HCOO$^-$)	−306.5217	−0.8010	−0.9611
Fe(H$_2$O)$_5$(CH$_3$O$^-$)	−251.2897	−1.3468	−0.6250
Fe(H$_2$O)$_5$(CH$_3$S$^-$)	−241.3777	−0.1651	−0.9084
Fe(H$_2$O)$_5$(NH$_2$CH$_3$)	−376.2997	−0.3434	−0.8075
Fe(H$_2$O)$_5$(HCOO$^-$)	−274.2278	−0.7464	−0.7799
MSE	−	−0.52	−0.88
MUE	−	0.52	0.88

Calculation of the Reduction Potential at the CCSD(T)/CBS Level

The last step in the calculation of the reference values for $\Delta E_{elec}^{Fe^{3+}/Fe^{2+}}$ is the extrapolation to the CCSD(T)/CBS level. After obtaining $\Delta E_{elec}^{Fe^{3+}/Fe^{2+}}$ at MP2/CBS, this last step can be easily done by adding to this value the difference between the reduction potentials calculated with MP2 and CCSD(T) ($E_{CCSD(T)-MP2}$) at the same basis set, aug-cc-pVDZ in our case. This widely used approximation is based on the fact that the difference in correlation energy between CCSD(T) and MP2 does not depend strongly on the basis set, in particular for medium/large basis sets. Due to the high computational cost of CCSD(T) calculations, we obtained $\Delta E_{elec}^{Fe^{3+}/Fe^{2+}}$ with this method with the aug-cc-pVDZ basis set. Previous studies have shown that using a larger basis set in this correction does not significantly alter the extrapolated value, and that the largest difference between using double or triple-zeta when obtaining $E_{CCSD(T)-MP2}$ was < 0.1 kcal/mol (Jurečka and Hobza, 2002). Jurečka et al. (2006) have also previously stated that double-zeta basis sets are complete enough to obtain a good extrapolation to the CBS limit, and using larger basis sets requires more computational time without much benefit.

Table 5 presents the results $\Delta E_{elec}^{Fe^{3+}/Fe^{2+}}$ at the CCSD(T)/CBS level (as stated previously, CCSD(T)/CBS was extrapolated using the results obtained with extrapolation scheme II). We can observe that the difference between both CCSD(T) and MP2 at the CBS limit is significant, which is a result of the relevant role that correlation energy has in the case of our chemical transformation, which is expectable as an electron is being taken out from the system.

level. It is, however, worth mentioning that, even though the use of large basis sets to get more accurate results is a good practice in systems with a small number of atoms, such methods are impractical for large biological systems, in which case the system would have to be reduced significantly, introducing much larger errors.

Fortunately, DFT methods converge much faster than Post-HF methods in relation to the basis set completeness, and as such they will not be affected by these large basis set truncation errors, as seen here.

TABLE 5 | $\Delta E_{elec}^{Fe^{3+}/Fe^{2+}}$ obtained at the CCSD(T)/CBS level and at the MP2/CBS level (extrapolated using scheme II), as well as the difference between these two values for each model.

Model	$\Delta E_{elec}^{Fe^{3+}/Fe^{2+}}$ (kcal/mol)		
	CCSD(T)/CBS	MP2/CBS	CCSD(T)/CBS - MP2/CBS
$Fe(H_2O)$	−593.8053	−602.8540	9.05
$Fe(H_2O)_2$	−532.2488	−538.2482	6.00
$Fe(H_2O)_4$	−437.8827	−442.1297	4.25
$Fe(H_2O)_6$	−383.3053	−386.5973	3.29
$Fe(H_2O)_3(CH_3O^-)$	−268.8103	−277.7082	8.90
$Fe(H_2O)_3(CH_3S^-)$	−263.1074	−263.7248	0.62
$Fe(H_2O)_3(NH_2CH_3)$	−419.8000	−424.9332	5.13
$Fe(H_2O)_3(HCOO^-)$	−302.7982	−306.5217	3.72
$Fe(H_2O)_5(CH_3O^-)$	−242.1318	−251.2897	9.16
$Fe(H_2O)_5(CH_3S^-)$	−239.3788	−241.3777	2.00
$Fe(H_2O)_5(NH_2CH_3)$	−372.7251	−376.2997	3.57
$Fe(H_2O)_5(HCOO^-)$	−268.0456	−274.2278	6.18

The energies extrapolated to the MP2/CBS level are much closer to the true value than those calculated using a truncated basis set, but these are still, nonetheless, approximations with associated errors that are introduced at different points throughout our calculations. Furthermore, the conversion of the MP2 level into the CCSD(T) level at the CBS limit is also the cause of small error. Overall, it must be taken into account that these results were obtained through several approximations, each of which adding a small error to our final values. In total, we expect our reference values ($\Delta E_{elec}^{Fe^{3+}/Fe^{2+}}$ calculated at the CCSD(T)/CBS level) to have an associated uncertainty of 1 kcal/mol or less, in most of the cases. As such, a qualitative distinction between density functionals having a difference smaller than 1 kcal/mol between them should not be made.

Benchmarking of the DFT Functionals

After obtaining the reference values for $\Delta E_{elec}^{Fe^{3+}/Fe^{2+}}$ of iron in each of the complexes, we can now use these results to evaluate the performance of the 44 density functionals we have selected. It should be noted that this benchmark study does not represent the overall quality of each of the DFT, but merely indicates, for these systems in specific (which we expect to be representative of aminoacid side-chain coordination shells), which of them are better at estimating the chemical property we wish to study (iron reduction potential, in our case).

From the optimized geometries for all systems, we performed single point calculation with each of the 44 density functionals, and calculated $\Delta E_{elec}^{Fe^{3+}/Fe^{2+}}$ using the basis set 6-311+G(2df,2p). We selected a large basis set in order to minimize truncation errors, which means that the fluctuations on the results arise essentially from the density functional. The values obtained for $\Delta E_{elec}^{Fe^{3+}/Fe^{2+}}$ will be compared to the reference values calculated at the CCSD(T)/CBS level.

In order to make discussing the results easier, and because our study involves several different systems, we decided to group our models in 3 distinct groups (**Figure 1**). Group A comprises the iron complexes with only water as their ligand, with coordination numbers 1, 2, 4, and 6, whereas groups B and C include complexes which contain one ligand that represents the side chain of an amino acid. What differentiates these 2 groups is the coordination number of the iron ion: 4 for group B and 6 for group C. The iron is complexed with 3 and 5 water molecules, respectively, plus the ligand simulating the amino acid side chain.

For the next part of this paper, we will only analyze the ten functionals that gave the best performance for each case. We decided to present our results in this fashion in order to make a more focused and easy-to-read discussion. A full table displaying the results for all the DFs tested is available in the **Supporting Information** section of this paper.

We begin our discussion with the systems belonging to group A.

Table 6 represents the results obtained for the ten density functionals that showed the best performance for this group. Performance is measured as the difference between $\Delta E_{elec}^{Fe^{3+}/Fe^{2+}}$ calculated with each DFT and the reference value (MUE). We also show the maximum error (MaxE), which is the model were the difference between the value calculated with the reference value and the energy calculated with the density functional was higher.

Taking these results into account, we can further divide the DFs into 3 separate groups. Group I includes all the density functionals with MUE between 0 and 2.31 kcal/mol (which corresponds to an error below 0.1 V per electron transfer). Group II comprises functionals with MUE between 2.31 and 4.62 kcal/mol (between 0.1 and 0.2 V per electron transfer) and group III encompasses those whose MUE is higher than 4.62 kcal/mol. The density functionals whose performance puts them in group I are BB1K, mPWB1K, and mPW1N. It is interesting to note that these DFs all belong to either the hybrid-meta-GGA or hybrid-GGA class, and that in the whole table only two functionals do not belong to these classes. The first four functionals all have a rather high HF exchange percentage, higher than 40%. The very popular B3LYP functional appears at 9th place in our table as well, and its MUE puts it in group II. It is also worth noting that the $Fe(H_2O)_6$ system is the one where the largest error is more frequently found.

We could also measure the performance of the tested functionals in terms of MaxE. However, and taking the results in **Table 6** into account, we can observe that both MUE and MaxE show a similar trend, and that those functionals with a smaller MUE also tend to have a smaller MaxE, with a few exceptions. For this reason, we decided to continue to rank the DFs in terms of MUE.

For group B, we performed the same analysis as group A. As stated previously, this group consists of systems with three water molecules and one ligand coordinated to the iron ion in a tetrahedral geometry. Each ligand represents a different amino acid side chain (CH_3O^- for serine, $HCOO^-$ for glutamate and aspartate, CH_3S^- for cysteine and NH_2CH_3 for lysine), which correspond to the most common residues that participate in iron complexes of proteins.

The results for group B (**Table 7**), show that the DFs that display a better accuracy for this group are mostly similar to those that performed better for system A as well. The density functionals that belong to group I in this case are BMK and B3LYP, which are, yet again, either hybrid-meta-GGAs or hybrid-GGAs. The HF correlation percentage for most of the top ten functionals is also higher than 40%, with MN-12L as a remarkable exception. The system that contributes the most to the maximum error in this group is Fe(CH$_3$S$^-$)(H$_2$O)$_3$.

And finally, we reach our last group, group C, composed of iron complexes with coordination number 6, in which one of the ligand is a chemical group representing an amino acid side chain, and the other five are water molecules. The results obtained for the benchmarking performed for group C are present in **Table 8**. Six functionals have MUE values < 2.31 kcal/mol, placing them in group I. These are mPW1B95, PBE1PBE, B3PW91, wB97X-D, BB1K, and mPWB1K. The percentage of HF exchange is not as high for this group, but still some of the functionals placed in the top ten have more the 40% HF exchange. Again for group C, all of the functionals in group I, and indeed in the whole table, are either hybrid-meta-GGAs or hybrid-GGAs. In accordance with the results for group B, for group C it is also the complex with CH$_3$S$^-$ that contributed the most times to the

MaxE. B3LYP shows a good performance in this group as well, ranking in the 7th position. We also assessed the contribution of the dispersion effect on the iron complexes of group C, which are those that have a higher number of ligands and, therefore, where the dispersion interactions should be more noticeable. In this sense, we calculated the Grimme's dispersion correction (Grimme et al., 2010; Goerigk and Grimme, 2011) with the top ten functionals (**Table 8**) and concluded that its inclusion marginally enhances the results, as the average MUE was reduced by ∼0.15 kcal/mol. However, considering the precision of the reference CCSD(T)/CBS method (about 0.5–1 kcal/mol), we can assume that the inclusion of the dispersion correction does not significantly influence the results and, consequently, the gathered illations.

Having analyzed the performance of the 44 density functionals in estimating the reduction potentials of iron complexes in the three separate groups, we will now make an overall appreciation of their performance for all the systems studies as a whole.

Table 9 presents the ten functionals that showed the best performance in calculating $\Delta E_{elec}^{Fe^{3+}/Fe^{2+}}$ for the studied iron complexes. Three functionals MUE values that put them in group I: BB1K, mPWB1K, and mPW1B95. All of them belong to the hybrid-meta-GGA class of functionals. The very popular B3LYP

TABLE 6 | Top ten best performing functionals for group A complexes.

	Functional	Type	%HF$_{exchange}$	MUE	MaxE	Fe(H$_2$O)	Fe(H$_2$O)$_2$	Fe(H$_2$O)$_4$	Fe(H$_2$O)$_6$
1	BB1K	hm-GGA	42.00	0.67	1.76	−0.47	0.21	−0.26	1.76
2	mPWB1K	hm-GGA	44.00	0.99	1.93	0.44	1.28	0.29	1.93
3	mPW1N	h-GGA	40.60	2.23	3.49	1.21	2.55	1.68	3.49
4	BMK	hm-GGA	42.00	2.41	3.62	3.39	1.52	−1.10	3.62
5	mPW1B95	hm-GGA	31.00	2.49	4.12	−2.39	− 4.12	−2.56	0.89
6	M06-2X	hm-GGA	54.00	2.53	4.50	4.50	3.79	0.89	0.93
7	mPW1K	h-GGA	42.80	2.62	3.51	1.70	3.30	1.98	3.51
8	MN12-SX	hm-NGA	25.00/–	2.98	7.43	0.81	−1.43	2.25	7.43
9	B3LYP	h-GGA	20.00	3.03	5.41	−0.67	− 5.41	−2.61	3.42
10	mPW2PLYP	hh-GGA	55.00	3.08	4.69	−3.55	−1.89	− 4.69	−2.19

For each functional, we highlighted (red) the model with the largest difference between reference and calculated value.

TABLE 7 | Top ten best performing functionals for group B complexes.

	Functional	Type	%HF$_{exchange}$	MUE	MaxE	Fe(H$_2$O)$_3$ (CH$_3$O$^-$)	Fe(H$_2$O)$_3$ (CH$_3$S$^-$)	Fe(H$_2$O)$_3$ (NH$_2$CH$_3$)	Fe(H$_2$O)$_3$ (HCOO$^-$)
1	BMK	hm-GGA	42.00	2.02	3.71	3.71	−1.24	1.84	−1.30
2	B3LYP	h-GGA	20.00	2.10	6.22	0.02	−0.63	−1.55	− 6.22
3	wB97X-D	h-GGA	22.20/100.00	2.33	4.21	− 4.21	−0.82	0.38	−3.90
4	BB1K	hm-GGA	42.00	2.44	4.33	1.59	− 4.33	0.93	−2.92
5	mPWB1K	hm-GGA	44.00	2.53	3.97	2.61	− 3.97	1.61	−1.93
6	mPW1N	h-GGA	40.60	2.65	4.04	3.68	− 4.04	2.59	−0.27
7	mPW1B95	hm-GGA	31.00	2.72	6.02	−0.69	−2.45	−1.71	− 6.02
8	mPW1K	hm-GGA	44.00	2.96	4.30	4.22	− 4.30	3.01	0.30
9	MN12-L	m-NGA	–	2.98	4.50	4.03	4.50	0.56	−2.84
10	M06-2X	hm-GGA	54.00	3.07	6.93	6.93	−2.52	2.21	0.63

For each functional, we highlighted (red) the model with the largest difference between reference and calculated value.

TABLE 8 | Top ten best performing functionals for group C complexes.

	Functional	Type	%HF$_{exchange}$	MUE	MaxE	Fe(H$_2$O)$_5$ (CH$_3$O⁻)	Fe(H$_2$O)$_5$ (CH$_3$S⁻)	Fe(H$_2$O)$_5$ (NH$_2$CH$_3$)	Fe(H$_2$O)$_5$ (HCOO⁻)
1	mPW1B95	hm-GGA	31.00	1.65	3.26	−1.73	− 3.26	0.86	0.74
2	PBE1PBE	h-GGA	25.00	1.71	3.48	−2.51	− 3.48	−0.61	−0.25
3	B3PW91	h-GGA	20.00	1.73	2.60	−2.06	− 2.60	1.23	1.02
4	wB97X-D	h-GGA	22.20/100.00	1.99	3.43	0.38	0.77	3.43	3.39
5	BB1K	hm-GGA	42.00	2.03	4.47	0.23	− 4.47	1.68	1.76
6	mPWB1K	hm-GGA	44.00	2.29	3.98	1.05	− 3.98	1.92	2.20
7	B3LYP	h-GGA	20.00	2.38	3.22	−0.41	−3.08	3.22	2.82
8	B1LYP	h-GGA	25.00	2.81	7.23	−2.53	− 7.23	0.90	0.57
9	B97-1	h-GGA	21.00	2.82	4.10	−3.38	− 4.10	−0.35	−3.46
10	M06-2X	hm-GGA	54.00	2.91	4.77	4.77	−2.43	1.39	3.05

For each functional, we highlighted (red) the model with the largest difference between reference and calculated value.

TABLE 9 | Top ten best performing functionals for all the complexes studied.

	Functional	Type	%HF$_{exchange}$	MUE	MaxE	MaxE Complex
1	BB1K	hm-GGA	42.00	1.72	4.47	Fe(H$_2$O)$_5$(CH$_3$S⁻)
2	mPWB1K	hm-GGA	44.00	1.93	3.98	Fe(H$_2$O)$_5$(CH$_3$S⁻)
3	mPW1B95	hm-GGA	31.00	2.28	6.02	Fe(H$_2$O)$_3$(HCOO⁻)
4	BMK	hm-GGA	42.00	2.49	3.93	Fe(H$_2$O)$_5$(HCOO⁻)
5	B3LYP	h-GGA	20.00	2.51	6.22	Fe(H$_2$O)$_3$(HCOO⁻)
6	B3PW91	h-GGA	20.00	2.72	7.23	Fe(H$_2$O)$_3$(HCOO⁻)
7	mPW1N	h-GGA	40.60	2.74	4.04	Fe(H$_2$O)$_3$(CH$_3$S⁻)
8	M06-2X	hm-GGA	54.00	2.84	6.93	Fe(H$_2$O)$_3$(CH$_3$O⁻)
9	mPW1K	h-GGA	42.80	3.03	4.30	Fe(H$_2$O)$_3$(CH$_3$S⁻)
10	PBE1PBE	h-GGA	25.00	3.08	7.21	Fe(H$_2$O)$_3$(HCOO⁻)

The complex to which the MaxE corresponds to is also shown.

functional ranks 5th in this overall analysis. All functionals present in the top ten are either hybrid-meta-GGAs or hybrid-GGAs, and the percentage of HF exchange is higher than 40% in most of the cases. We can therefore conclude that both of these classes are the best at estimating the $\Delta E_{elec}^{Fe^{3+}/Fe^{2+}}$ of the iron complexes studied, and should be used when calculating the reduction potential of iron complexes.

CONCLUSION

In our study, we analyzed the performance of 44 density functionals in estimating the electronic component of the reduction potential at 0 K of 16 different iron complexes, with a special interest in iron complexes with ligands similar to some amino acid side-chains. We were able to conclude which of these DFs gave results closer to reference values calculated using very high level computational methods, with posterior extrapolation to the very accurate CCSD(T)/CBS level of theory, which was done with two different extrapolation schemes. An initial geometry optimization for each of the 16 systems was carried out with MP2/6-311+G(d,p) level of theory, followed by single-point calculations with each of the 44 selected functional with

the 6-311+G(2df,2p) basis set. From the difference between the energies obtained with Fe²⁺ and Fe³⁺, we calculated $\Delta E_{elec}^{Fe^{3+}/Fe^{2+}}$ for all complexes, and then compared the results obtained with the reference values, ranking the DFs according to this difference.

Our results show that the difference between the extrapolated values for the level MP2/CBS calculated with both extrapolation schemes is very small (average difference is 0.52 kcal/mol). For the final extrapolation to the CCSD(T)/CBS level of the reduction potentials we selected the values calculated with scheme II, since it uses results obtained with a larger basis set for the extrapolation.

The benchmarking study showed us that the best functionals to calculate the reduction potential of iron complexes, especially those associated to biological systems, belong to the hybrid or hybrid-meta-GGAs classes, and have a high percentage of HF exchange. Overall, BB1K, mPWB1K, and mPW1B95 were the functionals that performed better for all iron complexes studied (average MUE 1.72, 1.93 and 2.28 kcal/mol, respectively). The very popular functional B3LYP gave a rather good performance, ranking 5th in our list of best functionals.

As a final remark it is important to note that there is an estimate of a 1 kcal/mol error in our reference values, which means that it is very difficult to assess the overall performance of functionals that differ by <1 kcal/mol between them. Nevertheless, our results are significant to show which DFs are more likely to perform well with iron complexes that incorporate biological systems, even if we are not able to show which, among all of them, is the single best density functional.

DATA AVAILABILITY

All datasets generated for this study are included in the manuscript and/or the **Supplementary Files**.

AUTHOR CONTRIBUTIONS

RL performed most of the calculations, analyzed the results, and contributed to the writing of the manuscript.

DG and PP performed a part of the calculations, analyzed the results, and contributed to the writing of the manuscript. MR and PF planned the research, analyzed the results, and contributed to the writing of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2019.00391/full#supplementary-material

REFERENCES

Abbaspour, N., Hurrell, R., and Kelishadi, R. (2014). Review on iron and its importance for human health. *J. Res. Med sci.* 19, 164–174.

Adamo, C., and Barone, V. (1997). Toward reliable adiabatic connection models free from adjustable parameters. *Chem. Phys. Lett.* 274, 242–250. doi: 10.1016/S0009-2614(97)00651-9

Adamo, C., and Barone, V. (1998a). Implementation and validation of the Lacks-Gordon exchange functional in conventional density functional and adiabatic connection methods. *J. Comput. Chem.* 19, 418–429. doi: 10.1002/(SICI)1096-987X(199803)19:4%3C418::AID-JCC4%3E3.0.CO;2-V

Adamo, C., and Barone, V. (1998b). Exchange functionals with improved long-range behavior and adiabatic connection methods without adjustable parameters: the mPW and mPW1PW models. *J. Chem. Phys.* 108, 664–675. doi: 10.1063/1.475428

Bartlett, R. J., and Purvis, G. D. (1978). Many-body perturbation theory, coupled-pair many-electron theory, and the importance of quadruple excitations for the correlation problem. *Int. J. Quant. Chem.* 14, 561–581. doi: 10.1002/qua.560140504

Becke, A. D. (1988). Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* 38, 3098–3100. doi: 10.1103/PhysRevA.38.3098

Becke, A. D. (1993). A new mixing of hartree–fock and local density-functional theories. *J. Chem. Phys.* 98, 1372–1377. doi: 10.1063/1.464304

Becke, A. D. (1996). Density-functional thermochemistry. IV. a new dynamical correlation functional and implications for exact-exchange mixing. *J. Chem. Phys.* 104, 1040–1046. doi: 10.1063/1.470829

Boese, A. D., Martin, J. M., and Handy, N. C. (2003). The role of the basis set: Assessing density functional theory. *J. Chem. Phys.* 119, 3005–3014. doi: 10.1063/1.1589004

Boese, A. D., and Martin, J. M. L. (2004). Development of density functionals for thermochemical kinetics. *J. Chem. Phys.* 121, 3405–3416. doi: 10.1063/1.1774975

Brás, N. F., Perez, M. A. S., Fernandes, P. A., Silva, P. J., and Ramos, M. J. (2011). Accuracy of density functionals in the prediction of electronic proton affinities of amino acid side chains. *J. Chem. Theory Comput.* 7, 3898–3908. doi: 10.1021/ct200309v

Broderick, J. B. (2001). *Coenzymes and Cofactors. Encyclopedia of Life Sciences*. Chichester, UK: John Wiley and Sons, Ltd.

Burke, K., Perdew, J., Wang, Y., Dobson, J., Vignale, G., and Das, M. (1998). *Electronic Density Functional Theory: Recent Progress and New Directions*. Springer. 81–111.

Chai, J.-D., and Head-Gordon, M. (2008). Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Phys. Chem. Chem. Phys.* 10:6615. doi: 10.1039/b810189b

Frisch, M., Trucks, G., Schlegel, H., Scuseria, G., Robb, M., Cheeseman, J., et al. (2009). *Gaussian09 Revision D. 01, Gaussian Inc. Wallingford CT*. Available online at: http://www.gaussian.com

Gill, P. (1996). A new gradient-corrected exchange functional. *Mol. Phys.* 89, 433–445. doi: 10.1080/002689796173813

Goerigk, L., and Grimme, S. (2011). A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions. *Phys. Chem. Chem. Phys.* 13, 6670–6688. doi: 10.1039/c0cp02984j

Grimme, S. (2006). Semiempirical hybrid density functional with perturbative second-order correlation. *J. Chem. Phys.* 124:034108. doi: 10.1063/1.2148954

Grimme, S., Antony, J., Ehrlich, S., and Krieg, H. (2010). 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. doi: 10.1063/1.3382344

Grimme, S., Ehrlich, S., and Goerigk, L. (2011). Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* 32, 1456–1465. doi: 10.1002/jcc.21759

Halkier, A., Helgaker, T., Jørgensen, P., Klopper, W., Koch, H., Olsen, J., et al. (1998). Basis-set convergence in correlated calculations on Ne, N2, and H2O. *Chem. Phys. Lett.* 286, 243–252. doi: 10.1016/S0009-2614(98)00111-0

Hamprecht, F. A., Cohen, A. J., Tozer, D. J., and Handy, N. C. (1998). Development and assessment of new exchange-correlation functionals. *J. Chem. Phys.* 109, 6264–6271. doi: 10.1063/1.477267

Handy, N. C., and Cohen, A. J. (2001). Left-right correlation energy. *Mol. Phys.* 99, 403–412. doi: 10.1080/00268970010018431

Harding, M. M., Nowicki, M. W., and Walkinshaw, M. D. (2010). Metals in protein structures: a review of their principal features. *Crystallogr. Rev.* 16, 247–302. doi: 10.1080/0889311X.2010.485616

Helgaker, T., Klopper, W., Koch, H., and Noga, J. (1997). Basis-set convergence of correlated calculations on water. *J. Chem. Phys.* 106, 9639–9646. doi: 10.1063/1.473863

Ho, J., Coote, M. L., Cramer, C. J., and Truhlar, D. G. (2016). "Theoretical calculation of reduction potentials," in *Organic Electrochem, 5th Edn*, eds O. Hammerich and B. Speiser (CRC Press), 229–259.

Hoe, W.-M., Cohen, A. J., and Handy, N. C. (2001). Assessment of a new local exchange functional OPTX. *Chem. Phys. Lett.* 341, 319–328. doi: 10.1016/S0009-2614(01)00581-4

Hohenberg, P., and Kohn, W. (1964). Inhomogeneous electron gas. *Phys Rev.* 136:B864. doi: 10.1103/PhysRev.136.B864

Jurečka, P., and Hobza, P. (2002). On the convergence of the (ΔECCSD(T)–ΔEMP2) term for complexes with multiple H-bonds. *Chem. Phys. Lett.* 365, 89–94. doi: 10.1016/S0009-2614(02)01423-9

Jurečka, P., and Šponer, J., Cern,ý, J., Hobza, P. (2006). Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys. Chem. Chem. Phys.* 8, 1985–1993. doi: 10.1039/B600027D

Karton, A., Tarnopolsky, A., Lamère, J.-F., o., Schatz, G. C., and Martin, J. M. L. (2008). Highly accurate first-principles benchmark data sets for the parametrization and validation of density functional and other approximate methods. derivation of a robust, generally applicable, double-hybrid functional for thermochemistry and thermochemical kinetics†. *J. Phys. Chem. A* 112, 12868–12886. doi: 10.1021/jp801805p

Kohn, W., and Sham, L. J. (1965). Self-consistent equations including exchange and correlation effects. *Phys. Rev.* 140, A1133–A1138. doi: 10.1103/PhysRev.140.A1133

Kormos, B. L., and Cramer, C. J. (2002). Adiabatic connection method for X$^-$+RX nucleophilic substitution reactions (X=F, Cl). *J. Phys. Organic Chem.* 15, 712–720. doi: 10.1002/poc.548

Kozuch, S., Gruzman, D., and Martin, J. M. L. (2010). DSD-BLYP: A general purpose double hybrid density functional including spin component scaling and dispersion correction. *J. Phys. Chem. C* 114, 20801–20808. doi: 10.1021/jp1070852

Lee, C., Yang, W., and Parr, R. G. (1988). Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* 37, 785–789. doi: 10.1103/PhysRevB.37.785

Li, S., Hennigan, J. M., Dixon, D. A., and Peterson, K. A. (2009). Accurate thermochemistry for transition metal oxide clusters. *J. Phys. Chem. A* 113, 7861–7877. doi: 10.1021/jp810182a

Liu, J., Chakraborty, S., Hosseinzadeh, P., Yu, Y., Tian, S., Petrik, I., et al. (2014). Metalloproteins containing cytochrome, iron–sulfur, or copper redox centers. *Chem. Rev.* 114, 4366–4469. doi: 10.1021/cr400479b

Lynch, B. J., Fast, P. L., Harris, M., and Truhlar, D. G. (2000). Adiabatic connection for kinetics. *J. Phys. Chem. A* 104, 4811–4815. doi: 10.1021/jp000497z

Perdew, J. P. (1986a). Erratum: Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B* 34, 7406–7406. doi: 10.1103/PhysRevB.34.7406

Perdew, J. P. (1986b). Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B.* 33, 8822–8824. doi: 10.1103/PhysRevB.33.8822

Perdew, J. P., Burke, K., and Ernzerhof, M. (1996b). Generalized gradient approximation made simple. *Phys. Rev. Lett.* 77, 3865–3868. doi: 10.1103/PhysRevLett.77.3865

Perdew, J. P., Burke, K., and Wang, Y. (1996a). Generalized gradient approximation for the exchange-correlation hole of a many-electron system. *Phys. Rev. B* 54, 16533–16539. doi: 10.1103/PhysRevB.54.16533

Perdew, J. P., Chevary, J. A., Vosko, S. H., Jackson, K. A., Pederson, M. R., Singh, D. J., et al. (1992). Atoms, molecules, solids, and surfaces: applications of the generalized gradient approximation for exchange and correlation. *Phys. Rev. B* 46:6671. doi: 10.1103/PhysRevB.46.6671

Perdew, J. P., Chevary, J. A., Vosko, S. H., Jackson, K. A., Pederson, M. R., Singh, D. J., et al. (1993). Erratum: Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Phys. Rev. B* 48, 4978–4978. doi: 10.1103/PhysRevB.48.4978.2

Perdew, J. P., Ziesche, P., and Eschrig, H. (1991). *Electronic Structure of Solids' 91.* Berlin: Akademie Verlag.

Perdew, J. P., and Zunger, A. (1981). Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B* 23, 5048–5079. doi: 10.1103/PhysRevB.23.5048

Peverati, R., and Truhlar, D. G. (2011a). Improving the accuracy of hybrid meta-GGA density functionals by range separation. *J. Phys. Chem. Lett.* 2, 2810–2817. doi: 10.1021/jz201170d

Peverati, R., and Truhlar, D. G. (2011b). M11-L: a local density functional that provides improved accuracy for electronic structure calculations in chemistry and physics. *J. Phys. Chem. Lett.* 3, 117–124. doi: 10.1021/jz201525m

Peverati, R., and Truhlar, D. G. (2012a). Exchange–correlation functional with good accuracy for both structural and energetic properties while depending only on the density and its gradient. *J. Chem. Theory Comput.* 8, 2310–2319. doi: 10.1021/ct3002656

Peverati, R., and Truhlar, D. G. (2012b). An improved and broadly accurate local approximation to the exchange–correlation density functional: The MN12-L functional for electronic structure calculations in chemistry and physics. *Phys. Chem. Chem. Phys.* 14:13171. doi: 10.1039/c2cp42025b

Peverati, R., and Truhlar, D. G. (2012c). Screened-exchange density functionals with broad accuracy for chemistry and solid-state physics. *Phys. Chem. Chem. Phys.* 14:16187. doi: 10.1039/c2cp42576a

Pople, J. A., Head-Gordon, M., and Raghavachari, K. (1987). Quadratic configuration interaction. a general technique for determining electron correlation energies. *J. Chem. Phys.* 87, 5968–5975. doi: 10.1063/1.453520

Purvis, I. I. I. G. D., and Bartlett, R. J. (1982). A full coupled-cluster singles and doubles model: the inclusion of disconnected triples. *J. Chem. Phys.* 76, 1910–1918. doi: 10.1063/1.443164

Ribeiro, A. J., Ramos, M. J., and Fernandes, P. A. (2010). Benchmarking of DFT functionals for the hydrolysis of phosphodiester bonds. *J. Chem. Theory comput.* 6, 2281–2292. doi: 10.1021/ct900649e

Riley, K. E., and Merz, K. M. (2007). Assessment of density functional theory methods for the computation of heats of formation and ionization potentials of systems containing third row transition metals. *J. Phys. Chem. A* 111, 6044–6053. doi: 10.1021/jp0705931

Riley, K. E., Op't Holt, B. T., and Merz, K. M. (2007). Critical assessment of the performance of density functional methods for several atomic and molecular properties. *J. Chem. Theory Comput.* 3, 407–433. doi: 10.1021/ct600185a

Schwabe, T., and Grimme, S. (2006). Towards chemical accuracy for the thermodynamics of large molecules: new hybrid density functionals including non-local correlation effects. *Phys. Chem. Chem. Phys.* 8:4398. doi: 10.1039/b608478h

Slater, J. C., and Phillips, J. C. (1974). Quantum theory of molecules and solids Vol. 4: the self-consistent field for molecules and solids. *Phys Today.* 27, 49–50. doi: 10.1063/1.3129035

Su, N. Q., Zhang, I. Y., Wu, J., and Xu, X. (2011). Calculations of ionization energies and electron affinities for atoms and molecules: a comparative study with different methods. *Front Chem China.* 6, 269–279. doi: 10.1007/s11458-011-0256-3

Tamames, J. A. C., and Ramos, M. J. (2010). Metals in proteins: cluster analysis studies. *J Molecul Model.* 17, 429–442. doi: 10.1007/s00894-010-0733-5

Tao, J., Perdew, J. P., Staroverov, V. N., and Scuseria, G. E. (2003). Climbing the density functional ladder: nonempirical meta–generalized gradient approximation designed for molecules and solids. *Phys Rev Lett.* 91:146401. doi: 10.1103/PhysRevLett.91.146401

Thomas, L. H. (1927). "The calculation of atomic fields," in *Mathematical Proceedings of the Cambridge Philosophical Society,* Cambridge University Press. 542–548. doi: 10.1017/S0305004100011683

Truhlar, D. G. (1998). Basis-set extrapolation. *Chem Phys Lett.* 294, 45–48. doi: 10.1016/S0009-2614(98)00866-5

Uudsemaa, M., and Tamm, T. (2003). Density-functional theory calculations of aqueous redox potentials of fourth-period transition metals. *J Phys Chem A.* 107, 9997–10003. doi: 10.1021/jp0362741

Van Voorhis, T., and Scuseria, G. E. (1998). A novel form for the exchange-correlation energy functional. *J Chem Phys.* 109, 400–410. doi: 10.1063/1.476577

Vosko, S. H., Wilk, L., and Nusair, M. (1980). Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.* 58, 1200–1211. doi: 10.1139/p80-159

Wilson, P. J., Bradley, T. J., and Tozer, D. J. (2001). Hybrid exchange-correlation functional determined from thermochemical data and ab initio potentials. *J. Chem. Phys.* 115, 9233–9242. doi: 10.1063/1.1412605

Zhao, Y., Lynch, B. J., and Truhlar, D. G. (2004). Development and assessment of a new hybrid density functional model for thermochemical kinetics. *J. Phys. Chem. A* 108, 2715–2719. doi: 10.1021/jp049908s

Zhao, Y., Schultz, N. E., and Truhlar, D. G. (2005). Exchange-correlation functional with broad accuracy for metallic and nonmetallic compounds, kinetics, and noncovalent interactions. *J. Chem. Phys.* 123:161103. doi: 10.1063/1.2126975

Zhao, Y., Schultz, N. E., and Truhlar, D. G. (2006). Design of density functionals by combining the method of constraint satisfaction with parametrization for thermochemistry, thermochemical kinetics, and noncovalent interactions. *J. Chem. Theory Computat.* 2, 364–382. doi: 10.1021/ct0502763

Zhao, Y., and Truhlar, D. G. (2004). Hybrid meta density functional theory methods for thermochemistry, thermochemical kinetics, and noncovalent interactions: the MPW1B95 and MPWB1K models and

comparative assessments for hydrogen bonding and van der waals interactions. *J Phys Chem A.* 108, 6908–6918. doi: 10.1021/jp04 8147q

Zhao, Y., and Truhlar, D. G. (2007). The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Accoun.* 120, 215–241. doi: 10.1007/s00214-007-0310-x

Zhao, Y., and Truhlar, D. G. (2008). Density functionals with broad applicability in chemistry. *Accounts Chem. Res.* 41, 157–167. doi: 10.1021/ar70 0111a

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read
for greatest visibility
and readership

FAST PUBLICATION
Around 90 days
from submission
to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative,
and constructive
peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers
acknowledged by name
on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: info@frontiersin.org | +41 21 510 17 00

REPRODUCIBILITY OF RESEARCH
Support open data
and methods to enhance
research reproducibility

DIGITAL PUBLISHING
Articles designed
for optimal readership
across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics
track visibility across
digital media

EXTENSIVE PROMOTION
Marketing
and promotion
of impactful research

LOOP RESEARCH NETWORK
Our network
increases your
article's readership