Influence of the Greater Protein Environment on the Electrostatic Potential in Metalloenzyme Active Sites: the Case of Formate Dehydrogenase

Azadeh Nazemi[1], Adam H. Steeves[1], and Heather J. Kulik[1,*]

[1]*Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139*

ABSTRACT: The Mo/W containing metalloenzyme formate dehydrogenase (FDH) is an efficient and selective natural catalyst which reversibly converts CO_2 to formate under ambient conditions. A greater understanding of the role of the protein environment in determining the local properties of the FDH active site would enable rational bioinspired catalyst design. In this study, we investigate the impact of the greater protein environment on the electrostatic potential (ESP) of the active site. To model the enzyme environment, we used a combination of long-timescale classical molecular dynamics (MD) and multiscale quantum-mechanical/molecular-mechanical (QM/MM) simulations. We leverage the charge shift analysis method to systematically construct QM regions and analyze the electronic environment of the active site by evaluating the degree of charge transfer between the core active site and the protein environment. The contribution of the terminal chalcogen ligand to the ESP of the metal center is substantial and dependent on the chalcogen identity, with ESPs less negative and similar for Se and S terminal chalcogens than for O regardless of whether the Mo^{6+} or W^{6+} metal center is present. Our evaluation reveals that the orientation of the sidechains and ligand conformations will alter the relative trends in the ESP observed for a given metal center or terminal chalcogen, highlighting the importance of sampling dynamic fluctuations in the protein. Overall, our observations suggest that the terminal chalcogen ligand identity plays an important role in the enzymatic activity of FDH.

Keywords: CO_2 catalysis, formate dehydrogenase, metalloenzymes, QM/MM, electrostatic potential

1. Introduction

Metalloenzymes are proteins that catalyze key chemical transformations under mild conditions.[1,2] Understanding the structure and function of these enzymes can enable design strategies for molecular catalysts with high activity and selectivity necessary for industrial use.[3-6] CO_2 reduction remains a challenging chemical reaction that is not industrially viable[7-10] but is carried out efficiently and selectively by formate dehydrogenase (FDH) enzymes.[11-14] Mo/W-containing FDHs convert CO_2 reversibly to formate ($HCOO^-$), a chemical fuel and H_2 storage compound.[15-18] Although Mo/W ions frequently appear in nature for the interconversion of CO_2 and $HCOO^-$, only a few examples of Mo/W-based homogeneous catalysts[19-23] have been proposed for reduction of CO_2 in comparison to costly noble metal catalysts[24-27] or earth abundant non-toxic metal catalysts.[28-30] Therefore, understanding the active site of Mo/W dependent FDH to better isolate the geometric and electronic environment effects is valuable for designing novel bioinspired Mo/W-based molecular catalysts.

The highly conserved active site of Mo/W containing FDH[31-33] consists of the Mo- or W-center coordinated by one or two dithiolene cofactors, known as molybdopterin guanine dinucleotides (MGDs) (Figure 1). In addition to these MGDs, the metal is coordinated by a selenocysteine (Sec), cysteine (Cys), or serine (Ser) residue and a terminal sulfido (M=S) or oxo (M=O) ligand in a distorted hexacoordinated trigonal prismatic geometry.[34,35] Several mononuclear Mo/W complexes have been synthesized and characterized in an attempt to replicate this FDH active site.[36-40] Generally, W-dependent FDH enzymes carry out the conversion of the $CO_2/HCOO^-$ with a higher turnover frequency than the Mo-dependent counterparts.[41] Similarly, enzymes containing selenocysteine are faster than their cysteine mutant.[41] The relation between type of the chalcogen ligand and reactivity is not well

understood, although presence of a terminal chalcogen ligand seems essential for the formate oxidation.[35,38,42,43] Computational modeling has been beneficial in further ruling in or out mechanisms.[31] Further accurate quantum mechanical modeling of the influence of the greater environment on the active site electronic structure is necessary to understand subtle differences in enzyme action with the metal center and terminal chalcogen identities.

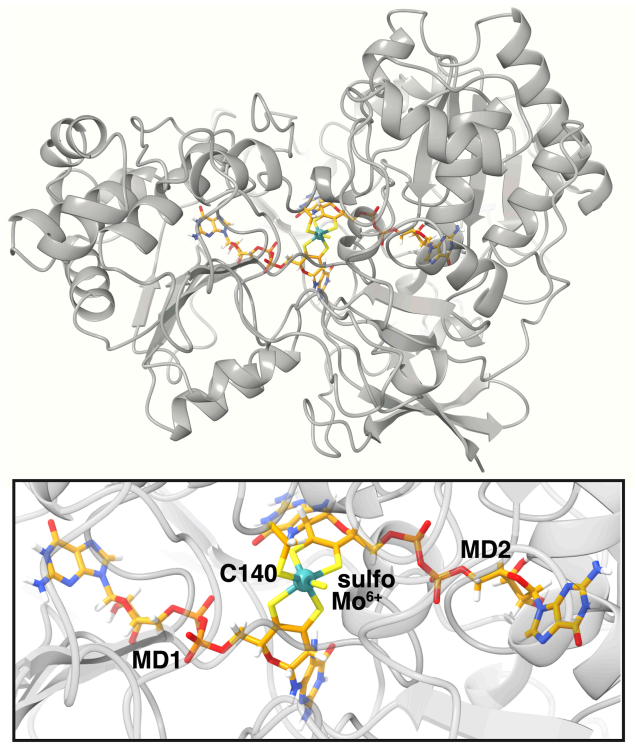

Figure 1. complete view of the FDH enzyme (PDB: 1FDO) is shown on top. The core active site is shown in orange stick representation (except Mo which is shown in cyan sphere) and annotated (bottom).

Using molecular dynamics (MD) and multilevel approaches like quantum mechanics/molecular mechanics (QM/MM) simulations help to balance the cost and accuracy in studying enzyme electronic structure along with suitable sampling.[44] Fluctuations, for example in electrostatic potential, are increasingly recognized as playing an important role in enzyme action that mandate efficient sampling to identify.[45-49] Moreover, it is essential to treat the active site and surrounding key residues with QM to quantify the bond rearrangement, polarization, and charge transfer.[50-54] Although small QM regions may produce a reasonable prediction of the

reaction mechanism, selecting a large number of residues to include in the QM region may be necessary to describe charge transfer between MM and QM regions. Similarly, an increasingly large QM region size ensures the inclusion of all residues that are important for enzyme action.[55-61] Regarding the challenges for QM region selection, systematic approaches can provide an unbiased insight toward electronic environment of the enzyme and detection of crucial residues along with improving the accuracy of the QM/MM.[62-70]

Previous studies on bioinspired catalyst design of FDH mimics found that the enzyme environment primarily affects the geometric properties of the metal center, and terminal chalcogen moieties primarily influence local electronic properties.[71] Here, we apply long-time classical MD and large-scale QM/MM simulation to further analyze the influence of the greater protein environment and metal coordinating residues on electronic properties of the FDH active site. Specifically, we focus on the charge reorganization of the protein residues in the presence and absence of the metallocofactor. Moreover, we evaluate the relationship between residue sidechain orientation and electrostatic potential at the metal center (Mo or W) in cofactors with different chalcogen (O, S, Se) ligands. Our analysis reveals the large contribution of the terminal chalcogen ligand to the ESP of the metal center, while ESPs are less negative and similar for Se and S terminal chalcogens than for O and more negative for Mo^{6+} than W^{6+}.

2. Computational Details.

Protein Structure and Preparation. An X-ray crystal structure of an oxidized form of the formate dehydrogenase (FDH) enzyme was obtained from the protein data bank (PDB ID: 1FDO[35]). The active site of 1FDO contains selenocysteine, Mo, two molybdopterin guanine dinucleotide (MGD) cofactors, in addition to a Fe_4S_4 cluster. To prepare the protein, the Fe_4S_4 cluster was removed from the crystal structure and a sulfido ligand was added manually as the

sixth coordinating ligand. The coordinating selenocysteine was converted to Cys140 (Figure 1). The protonation states of the apoenzyme residues were assigned using the H++ webserver[72-75] assuming a pH of 7.0 with all other defaults applied. Force-field parameters for the core active site were obtained from a combination of the MCPB.py[76] utility in AMBER18[77] and the parameters available in literature.[78] Standard protein residues were simulated with AMBER ff14SB[79]. The protein was solvated with 10 Å of TIP3P[80] water buffer in a periodic rectangular prism box and neutralized with Na^+ counterions. The initial AMBER topology and coordinate files were generated for the 77,778-atom system, which are provided in the Supporting Information.

MM Equilibration and Dynamics. The MD equilibration and production used the GPU-accelerated PMEMD code in AMBER 18.[77] The equilibration steps were as follows: i) restrained (1000 steps) and unrestrained (2000 steps) minimizations, ii) 10 ps NVT heating to 300 K with a Langevin thermostat with collision frequency of 1.0 ps^{-1} and a random seed, iii) 1 ns NPT equilibration using the Berendsen barostat with a pressure relaxation time of 2 ps, and iv) 100 ns of production dynamics. All MD applied SHAKE[81] with a 2 fs time-step and particle mesh Ewald with a real space 10-Å electrostatic cutoff.

QM/MM, QM cluster, and QM-only simulations. QM/MM single point energy calculations were performed on snapshots obtained from MD to obtain geometric and electronic properties of FDH enzyme active site. MD snapshots were post-processed using the center of mass utility in PyMOL[82] to generate the largest possible spherical droplet centered around each protein with a radius of 44 Å, leading to total 37,548 atoms in the QM/MM droplet structure. For all QM/MM calculations, a developer version of TeraChem v1.9[83] was used for the QM region, and AMBER18[77] was used for the MM region as well as to drive the QM/MM calculation. To

passivate covalent bonds that bridge QM and MM portions, electrostatic embedding and hydrogen link atoms were employed. The 376-atom (i.e., including 26 link atoms) QM region with a net charge of −4 was selected systematically with charge shift analysis method[63,64]. All calculations were carried out in a closed shell formalism (i.e., singlet multiplicity). The QM region was modeled with density functional theory (DFT) using the range-separated exchange-correlation functional ωPBEh[84] (ω = 0.2 $bohr^{-1}$) along with the composite LACVP* basis set (i.e., LANL2DZ effective core potential[85] on Mo and W heavy atoms along with 6-31G*[86] for the remaining atoms). The level of theory was chosen based on prior studies that show degraded performance of global hybrids (e.g., B3LYP) with increasing QM region size[58,63]. For QM cluster optimization calculations, all backbone atoms (i.e., Cα, N, C, O) in protein residues were fixed using positional constraints. The QM cluster was modeled with the same level of theory that was used for QM region of QM/MM calculations. The protein environment was mimicked with the conductor-like implicit solvent model (C-PCM)[87,88], as implemented in TeraChem[89,90] with ε = 4. To evaluate the electrostatic potential (ESP) at the metal center in QM/MM or QM cluster calculations, we tested three partial charges including Mulliken, Voronoi, and Hirshfeld charges (Supporting Information Figure S1). After comparison, the most physically reasonable partial charges were determined to be the Hirshfeld form. Using these charges, we computed the ESP at the metal center with an in-house Python script.

3. Results and Discussion.

3a. Charge Shift Analysis for QM Region Detection.

We employed charge shift analysis (CSA)[63,64] to systematically determine the size of the QM region. CSA works by identifying residues that undergo charge redistribution when the active site substrates are removed, which provides valuable insights for identifying residues to

include in the QM region for large QM or hybrid QM/MM calculations. We performed CSA on the final snapshot of a 100 ns MD simulation by selecting a large cluster of amino acids proximal to the core active site. This initially chosen set of residues consisted of 1,280 atoms (i.e., 75 amino acid residues) including 150 cofactor atoms and 26 capping link atoms (Supporting Information Figure S2). This large initial QM region contains 46 polar or charged amino acids and 29 nonpolar amino acids. Note that, in this work, we designate cysteine and proline as nonpolar amino acids. Of the overall set, 8 are positively charged and 6 negatively charged. The combination of the strongly negatively charged cofactor (net charge: -5) and the charged amino acids yields a total charge of -3. To generate the apoenzyme, the core cofactor (i.e., Mo, molybdopterins, sulfido) and active site residues (i.e., the side chain of Cys140) were extracted. In the case of the covalently bonded Cys140, we mutated the residue to an alanine to avoid bond polarization effects caused by removal of the cysteine side chain (Supporting Information Figure S2).

To determine the charge reorganization upon removal of the active site using CSA, we performed single-point QM/MM calculations on the holo and apoenzymes. We then computed partial charge sums for each amino acid in the QM region followed by the charge difference between the holo and apoenzymes for each residue:

$$\Delta q_{RES} = q_{RES}^{holo} - q_{RES}^{apo} \quad (1)$$

with a previously motivated cutoff[63,64] of Δq_{RES} = |0.05 e| applied (Supporting Information Table S1). All changes in residue charges above this cutoff are identified as those corresponding to substantive charge redistribution. After carrying out this initial filtering step (i.e., to exclude residues with Δq_{RES} < |0.05 e|), a large, 509-atom QM region is selected that consists of the

cofactor, 38 link atoms, and 21 amino acid residues (Figure 2). The 21 selected amino acids consist of four positively charged and three negatively charged amino acids, while the initial QM region included eight positively charged and six negatively charged amino acids. The total charge of the new holoenzyme is -4.

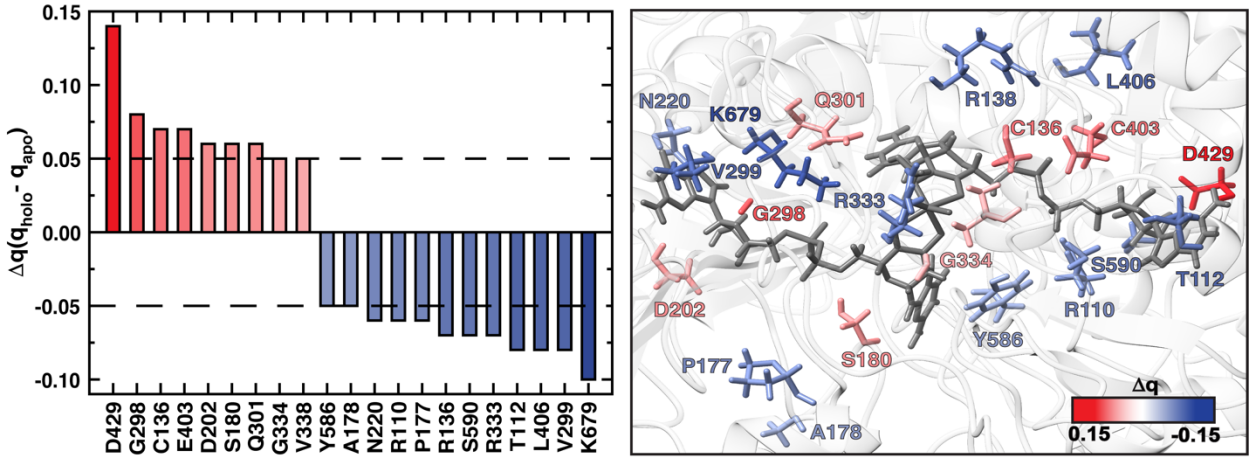

Figure 2. Results from CSA analysis halo-apo charge difference (in e) upon removal of the cofactor ($Mo^{6+,}$ molybdopterins, sulfido and side chain of C140). The 21 amino acids with a charge difference above the cutoff range of |-0.05| are shown with red or blue bars to indicate the sign of the shift (left). The cutoff is indicated by two dashed lines. In the right panel, selected residues are shown as sticks. The residues shown and labeled in blue lose partial charge upon removal of cofactor, whereas residues shown and labeled in red gain partial charge upon cofactor removal (as shown in inset color bar).

Interestingly, we observe that 7 out of 21 CSA-selected amino acids are nonpolar (i.e., (Cys136, Pro177, Ala178, Gly298, Val299, Val338, Leu406), which are not typically expected to have a large impact on the electronic environment of the QM region. Some of the amino acids detected by CSA are ca. 6–9 Å from the closest atoms in the molybdopterin cofactors (i.e., Pro177, Ala178, and Leu406) and are thus not in close proximity to the active site (Supporting Information Figure S3). For the remaining residues, the atoms closest to the core site are ca. 2-4 Å from the molybdopterins (Supporting Information Table S2).

To confirm the importance of the selected residues, we employed two density-based validation metrics[63] to evaluate whether residues such as the nonpolar residues and those distant

from the active site should be included in the QM region. These CSA validation methods depend on the direct effects of residue-substrate interactions and indirect effects from residue-residue interactions. To quantify these interactions, we generate a QM region model in QM/MM with the 21 residues detected by the initial CSA in the QM region. We then return individual residues from the QM region to the MM region to see how the core active site residues partial charges will shift (ASRS):

$$ASRS(RES) = \sqrt{\sum_{j}^{RES}(q_j^{RES \in QM} - q_j^{RES \in MM})^2} \quad (2)$$

and how the charges of the other key residues will alter background (BGD) residues [63]:

$$BGD(RES) = \sqrt{\frac{1}{N}\sum_{k}^{RES \notin ASR}(q_k^{RES \in QM} - q_k^{RES \in MM})^2} \quad (3)$$

Comparison of the charge metrics show that the direct ASRS effect dominates in most cases, but the indirect BGD outweighs ASRS for several nonpolar residues (i.e., Pro177, Ala178, Gly298, and Val299, Figure 3, top). From this analysis, we can see that Gly298 and Pro177 do have a strong indirect contribution to the electronic environment of the QM region since their BGD factors are above the |0.05 e| threshold while the ASRS factors are below this cutoff.

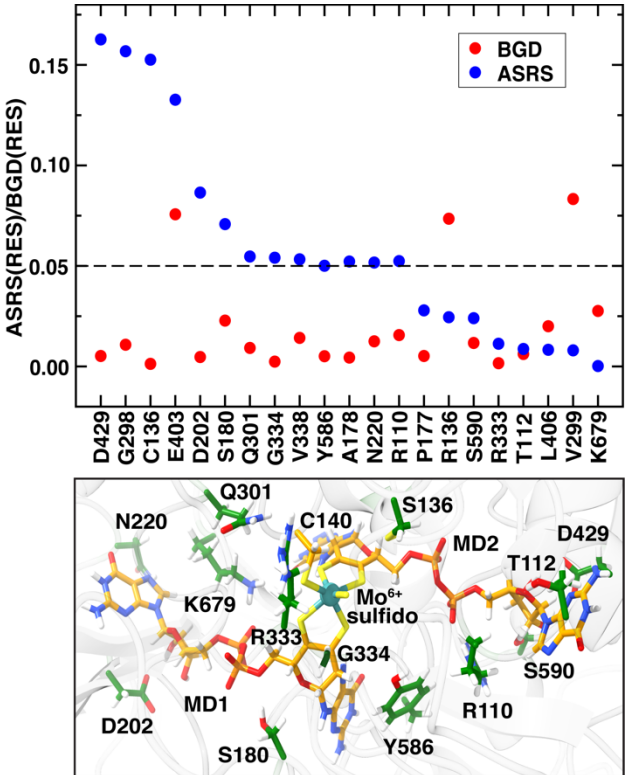

Figure 3. (top) ASRS(RES) (blue circles) and BGD(RES) (red circles) partial charge validation metrics in |e| for 21 amino acids detected by CSA approach. An |0.05 e| cutoff is indicated by a black dashed line. (bottom) Final QM region selected using the CSA method. The cofactor is shown in orange stick representation (except Mo which is shown as a cyan sphere) and annotated. The confirmed residues are shown as green sticks including 11 polar and 2 nonpolar residues, which corresponds to 376 atoms.

Among the three residues which are relatively distant from the active site (Pro177, Ala178, and Leu406), this analysis indicates Ala178 and Leu406 are false positives from the initial CSA and do not contribute significantly to the electronic environment of the QM region. While Pro177 does contribute indirectly, its relatively high distance from the active site (ca. 6 Å) motivates its removal from the final QM region to avoid introducing challenges of boundary treatment for a highly fragmented QM region. The only remaining nonpolar amino acids after our elimination of false positives are Cys136 and Gly334. Our analysis indicates they are contributing directly to the electronic environment, and they are relatively close (Cys136: 3.6 Å and Gly334: 3.8 Å) to the active site (Figure 3, top and Supporting Information Figure S4).

Analyzing the charged residues that can be expected to be more influential on the electronic environment, we observe that Arg333 is an important residue that contributes both directly to interactions with the cofactors and indirectly through interaction with Gly334. Conversely, Arg138 and Glu403 do not contribute directly or indirectly. Since they are also located relatively far (Arg133: 7.6 Å and Glu403: 4.6 Å) from the core active site, they were omitted from the final QM region (Figure 3, top and Supporting Information Figure S5). In total from the CSA evaluation and elimination of false positives, we select 13 key residues to include in the QM region. This final, validated QM region contains 376 atoms including 26 link hydrogen atoms and 150 cofactor atoms, and it has a net charge of -4 (Figure 3, bottom). Out of 8 positively charged and 6 negatively charged amino acids in the initial QM region only three positively charged and two negatively charged amino acids thus remain in the final QM region. Compared to the initial QM region, a comparable fraction of polar (i.e., 11 out of 46) residues remain but fewer (i.e., 2 out of 29) nonpolar residues are included in the final QM region.

3b. Electrostatic Potential Analysis

To gain a better understanding of the electronic environment at the metal center, we studied the electrostatic potential (ESP) using Hirshfeld partial charges (see Sec. 2 and Supporting Information Figure S2). We carried out this ESP analysis on single point QM/MM calculations for both the Mo^{6+} and W^{6+} centers using 100 snapshots from a 100 ns MD simulation (Supporting Information Figure S6). Comparing these two metal centers indicates the ESP is generally lower for the Mo center than the W center. To confirm the accuracy of the trend we observed for the ESP of the Mo^{6+} and W^{6+} centers from the single point calculations, we performed geometry optimizations on 10 snapshots for the enzymes with Mo^{6+} and W^{6+} (Supporting Information Table S3). The trends are unchanged after full optimization in the

presence of the two metal centers. This analysis shows that ESP obtained for the Mo^{6+} center on average is lower by 87.4 ± 0.9 $kJ.mol^{-1}e^{-1}$ for these 10 optimized snapshots than the ESP at W^{6+} center, which is consistent with the larger positive partial charge on the Mo^{6+} center. Thus, more favorable electrostatic potential does not necessarily correspond to increased activity observed for W-dependent FDHs.

Focusing on the Mo^{6+} center, we evaluated the contributions of both QM and MM regions to the ESP over the 100 snapshots from MD. To make this comparison, we first computed the ESP from 100 QM/MM single point snapshots on the MD trajectory into QM and MM parts (Figure 4). The ESP at the Mo^{6+} center obtained from the QM/MM simulation is a large negative value that corresponds to a stabilizing influence on the positively charged Mo^{6+} center (Figure 4). We then decomposed the ESP obtained from QM/MM into QM and MM regions (Figure 4). This comparison indicates the stabilizing contribution of the systematiclaly-constructed QM region (-1299.1 ± 16.8 $kJ.mol^{-1}e^{-1}$) is almost four times that of the MM region (306.5 ± 40.5 $kJ.mol^{-1}e^{-1}$). This analysis indicates the systematic selection of the QM region was sufficiently large such that the QM atoms play a much larger role in determining the electrostatic environment than the potentially limited fixed charge force field atoms do.

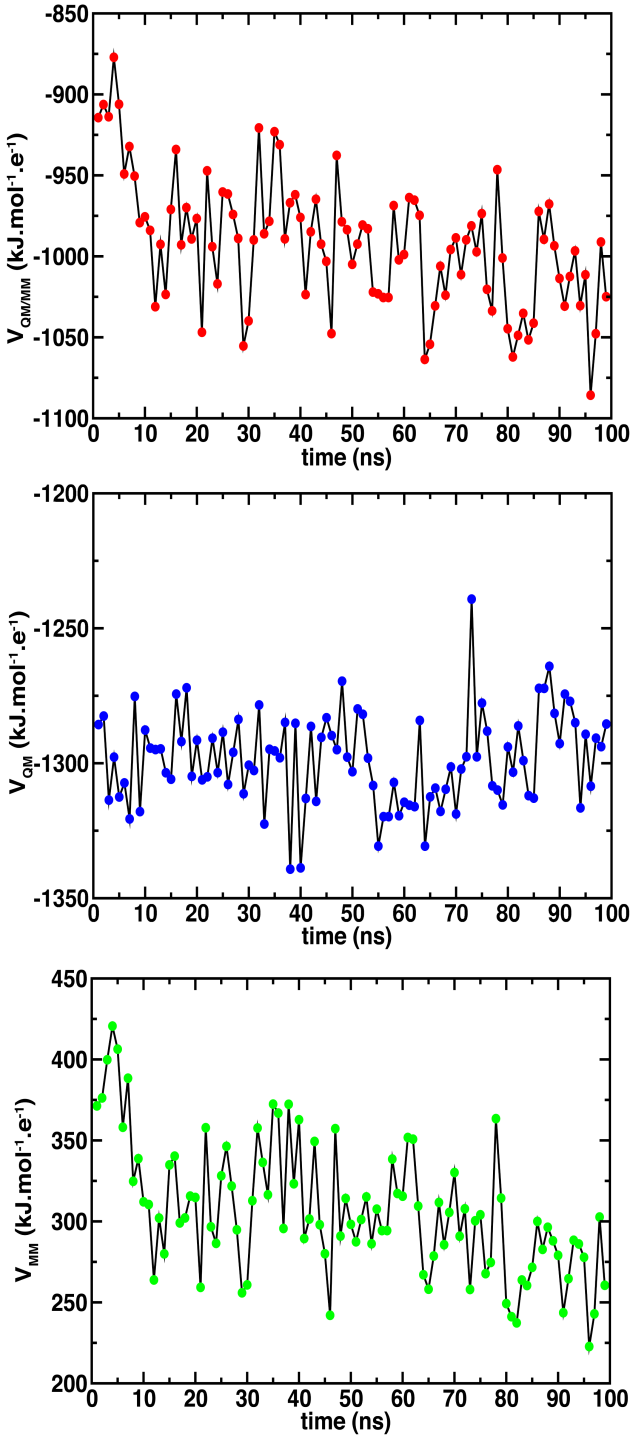

Figure 4. ESP values at Mo^{6+} site obtained from single point QM/MM calculations for 100 snapshots spaced by 1ns from MD simulations. (top) ESP obtained from combination of partial charges in QM region and point charges in MM region. (middle) ESP from only QM region. (bottom) ESP from only MM region.

To evaluate which components of the QM region have the greatest contributions to the ESP of Mo^{6+}, we calculated the ESP separately for the coordinating residues and more distant

but still surrounding QM amino acids and sidechains. The contribution to the ESP from coordinating residues (i.e., molybdopterins, sulfido, and Cys140) dominates (ca. 90%) over more distant residues (Figure 5, left). As could be expected, the contribution of the neutral side chains is very small, while the contribution of the sidechains of positively charged amino acids (i.e., Arg110, Arg333 and Lys679) are as large as the contribution of the backbone atoms of all residues and larger (ca. 2x) than the contribution of the sidechains of negatively charged amino acids (i.e., Asp202 and Asp429). This difference between positively and negatively charged amino acids can be rationalized by the greater proximity of Arg333 and Lys679 to the metal center in comparison to the negative side chains which are instead more proximal to the other end of the molybdopterin ligands.

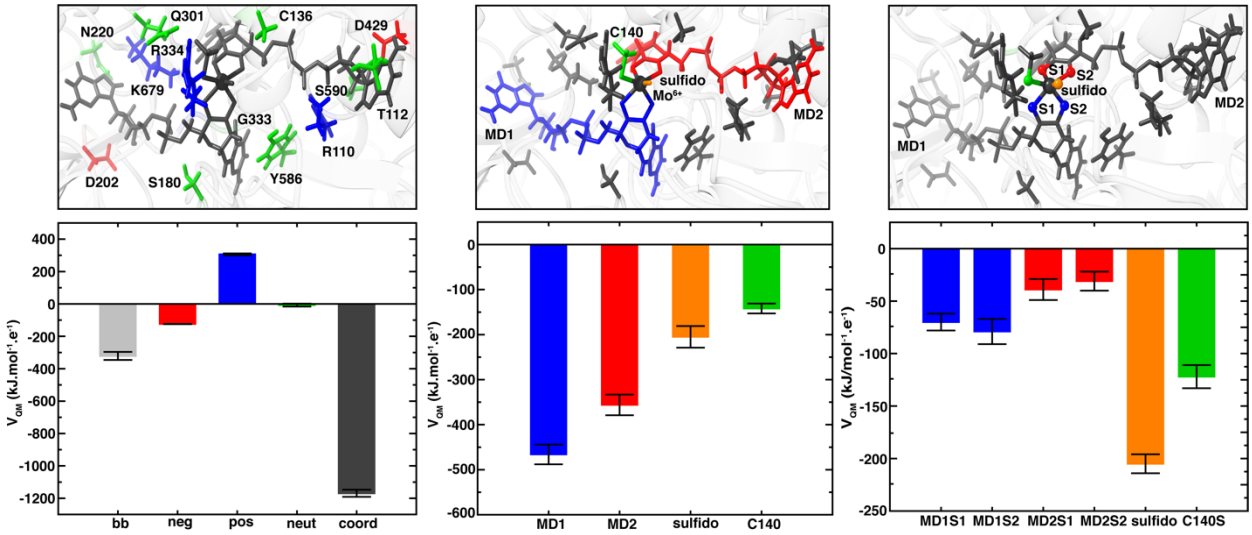

Figure 5. (left) ESP obtained from each component of QM region, coordinated residues to the metal are shown in gray, neutral, negatively charged, and positively charged amino acids are shown in purple, red, and blue, respectively. (middle) ESP obtained from coordinated residues including molybdobterins (MD1 in blue and MD2 in red), sulfido in yellow and C140 in orange. (right) The ESP obtained from six coordinated sulfur atoms, MD1 sulfurs, MD2 sulfurs, sulfido, and C140 sulfur are shown in blue, red, orange, and green, respectively.

Since the largest contributions of the QM region come from the coordinating residues and ligands, we further analyzed the contribution of individual coordinating species to the ESP at the Mo^{6+} center (Figure 5). Approximately 70% of the coordinated residues contribution to the ESP

comes from the molybdopterins, potentially due to their large size and significant number of charged functional groups. The contributions of the two molybdobpterins is asymmetric (MD1: -466 ± 22 kJ.mol^{-1}·e^{-1} is larger than MD2: -356 ± 23 kJ.mol^{-1}·e^{-1}). This difference could be due to differences in the surrounding residues in the enzyme environment. For the charged residues that are in the QM region, Arg110 and Asp429 form hydrogen bonds with the MD2 phosphate group and hydroxy groups of tetrahydrofuran, respectively, for 100% of the frames over which the ESP is evaluated (Supporting Information Figure S7). For 19% of the MD frames, MD2 also forms hydrogen bonds between the Lys679 sidechain and the carbonyl group of the molybdopterin pyrimidine ring. In contrast for MD1, hydrogen bonding interactions are sampled less frequently and with different residues. For MD1, the phosphate group and hydroxy groups of tetrohydrofuran again form hydrogen bonds, in this case with Lys679 and Asp202, respectively. Nevertheless, these hydrogen bonds are present in only around half (ca. 39-48%) of the 100 MD snapshots used for ESP evaluation.

To further investigate the impact of the charged residues on the ESP, we investigated the differences in hydrogen bonding for lower and higher relative ESP contributions from MD1 and MD2. Specifically, the difference in ESP (i.e., ESP_{MD1} - ESP_{MD2}) for the snapshot at 1 ns is largest at ~-173.5 kJ·mol^{-1}·e^{-1} and for the snapshot at 79 ns is smallest at ~-20.1 kJ·mol^{-1}·e^{-1}. Analyzing the first a snapshot with a high relative difference in ESPs, we observe an additional hydrogen bond for MD2 with Arg110, and Lys679 forms hydrogen bonds with both MD1 and MD2 through the carbonyl group of the molybdopterin pyrimidine ring and the phosphate group of MD2, respectively. On the other hand, for the case where the ESP contributions of the two molybdopterins are close, Lys679 only forms hydrogen bonds with MD1, is too far to form a productive hydrogen bond to MD2, and the number of positively charged residues hydrogen

bonding to the two molybdopterins is more balanced (Supporting Information Table S4).

Considering the number of atoms in each coordinating species, the single atom of the sulfido is much smaller than Cys140 (10 atoms) and the molybdopterins (69 atoms each), but its contribution remains substantial (Figure 2, middle). We thus specifically compared the contributions to the ESP of the coordinated sulfur atoms (i.e., four sulfur atoms from the molybdopterins, the one from Cys140 and the one from sulfido). This allowed us to determine if any of the other sulfur atoms also have an outsized impact on the ESP (Figure 5, right). As might be expected, the largest contribution among coordinated sulfurs to the ESP is from the sulfido ligand (38%, Figure 5). Following this doubly bonded species, the Cys140 sulfur has a larger contribution (22%) than the individual contributions from either molybdopterin (i.e., MD1: 13% and 14% each and MD2: 6% and 7% each). Overall, the contribution of the terminal sulfido ligand to the ESP of the Mo^{6+} center is significant, which means this terminal ligand can play an important role in the activity of the enzyme cofactor, which we will further evaluate next.

3c. Element-specific Trends in Electrostatic Potentials.

Since we observed a significant contribution of the terminal sulfido to the ESP at the metal center, we next investigated how sensitive this observation was to replacement of the terminal chalcogen (i.e., to O or Se) as well as the identity of the metal (i.e., Mo or W). We extracted ten representative snapshots from MD simulations for this comparison (Supporting Information Table S5 and Figure S8). These snapshots were selected based on their having the lowest or highest ESP evaluated with QM/MM in each 20 ns window of the 100 ns MD trajectory (Supporting Information Table S8). To take into account how the structure might change when the terminal ligand changes, we optimized the geometries of large QM clusters of all 10 snapshots for all six combinations of both Mo and W centers with the three different

terminal ligands. Despite possible differences in size or electronegativity, the ESP values obtained for Mo and W centers for each structure follow the same trend (Supporting Information Figure S9). We computed the difference between the ESP of Mo and W from the average values of over 10 structures and generally observe the Mo ESP to be more negative regardless of chalcogen identity (i.e., $\Delta ESP(Mo\text{-}W)_{oxo}$ = -82.1 ± 1.5 $kJ{\cdot}mol^{-1}{\cdot}e^{-1}$ vs $\Delta ESP(Mo\text{-}W)_{selenido}$ = -87.1 ± 1.3 $kJ.mol^{-1}{\cdot}e^{-1}$, Supporting Information Table S6). This more negative ESP at the Mo center regardless of terminal chalcogen is consistent with the more positive partial charge at the Mo center relative to W center (Supporting Information Table S7). Since the difference between Mo and W is consistent across structures, we focus on Mo for subsequent analysis.

Over the 10 optimized structures, differences in chalcogen atoms are largely consistent with the differences in electronegativity of the chalcogen atoms. That is, Mo=Se and Mo=S are more similar to each other than Mo=O (i.e., $ESP_{Mo=O}$ /$ESP_{Mo=S}$ = 1.10 ± 0.02 vs $ESP_{Mo=S}$ /$ESP_{Mo=Se}$ = 0.96 ± 0.03, Figure 6 and Supporting Information Table S8). Thus, ESP values obtained in the presence of the oxo terminal ligand are more negative than those obtained from sulfido, which are in turn slightly more negative than the case of selenido ligands (Figure 6). When we obtain relative ESP values with respect to the first structure, we notice that the ESPs occasionally do not follow this general trend (Supporting Information Figure S10). For example, structures 3 and 4 have very similar ESP values in the presence of sulfido and selenido ligands, which differ significantly from the oxo case.

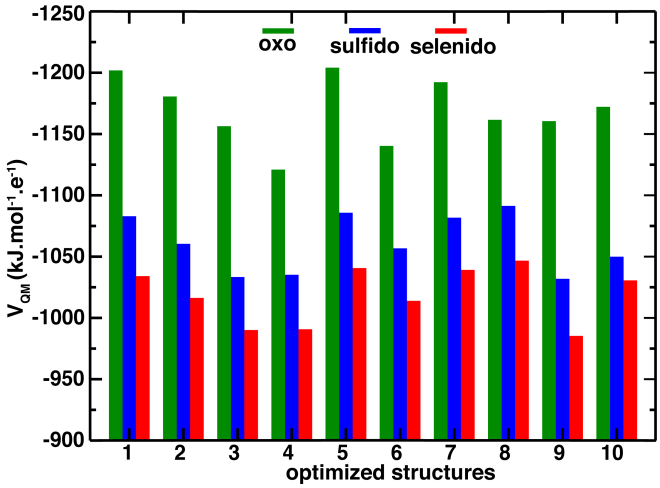

Figure 6. ESP at the Mo center for oxo (green), sulfido (blue) and selenido (red) series over 10 structures optimized from MD simulations.

To understand whether structural differences explain this discrepancy between structures, we first aligned the optimized Mo=O, Mo=S, and Mo=Se structures for the representative structure 4. The aligned Mo=S and Mo=Se structures are very similar, whereas the structure of the Mo=O intermediate differs from either Mo=S or Mo=Se. These differences are due to distinct orientations of residues or the molybdopterins that result in distinct non-covalent interactions (i.e., hydrogen bonds). Hydrogen bond strength differences could also play a role even when the same number of hydrogen bonds is present in the different structures. In the case of structure 4, the first difference observed is for the positively charged Lys679 which forms two hydrogen bonds for Mo=S or Mo=Se with both MD1 and MD2 molybdopterins (Figure 7). The hydrogen bond with MD1 is to the phosphate group and the hydrogen bond with MD2 is to the carbonyl group of the molybdopterin pyrimidine ring (Figure 7). In the case of Mo=O, Lys679 does not form a hydrogen bond with MD2 due to the orientation change in MD2, instead only forming a hydrogen bond with MD1 (Figure 7). Thus, in total, there are a higher number of hydrogen bonds for Mo=S (or Mo=Se) relative to Mo=O for Lys679. The second difference observed in

structure 4 arises from Thr112. For Mo=O, Thr112 forms one hydrogen bond through its backbone nitrogen with one of the nitrogen atoms of the imidazole ring in MD2 while for Mo=S (or Mo=Se), the same Thr112 backbone nitrogen instead forms a hydrogen bond with the carbonyl of the MD2 purine group's pyrimidine ring (Supporting Information Figure S11). Thus, the Mo=O N–H···N hydrogen bond can be expected to be weaker than the N–H···O hydrogen bond formed for Mo=S (or Mo=Se).

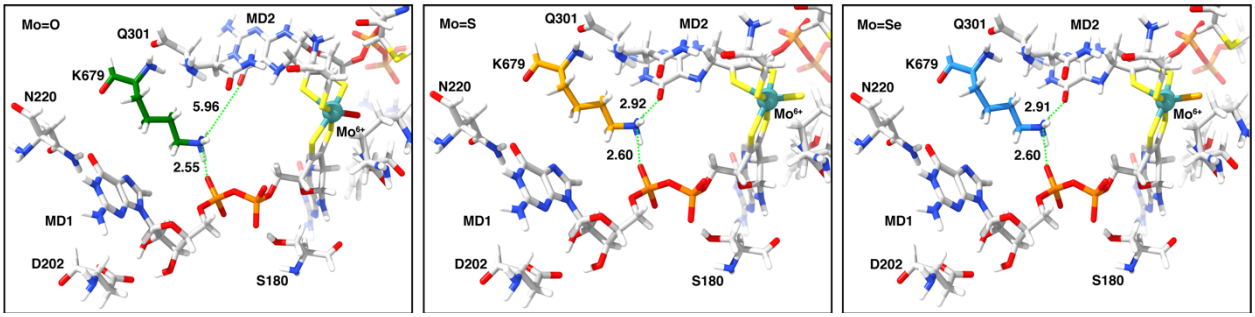

Figure 7. The optimized Mo=O (green), Mo=S (orange), and Mo=Se (blue) structures for structure 4 with only key amino acid, here Lys679, colored with the rest of the following colors: carbon in white, oxygen in red, nitrogen in blue, phosphor in orange, sulfurs in yellow, seleno in light orange and Mo in cyan. Relevant hydrogen bonding distances are shown with a light green dashed line and annotated distances in Å colored the same as the relevant structures. (left) Mo=O, (middle) Mo=S, (right) Mo=Se.

We also observe a key difference for both structures 6 and 8, in which the ESP for the Mo=O deviates from that observed for Mo=S and Mo=Se (Supporting Information Figure S10). To determine if these differences could be attributed to differences in sidechain orientations, we compared the optimized Mo=O, Mo=S, and Mo=Se structures. Alignment of the three structure 6 structures indicates the only difference arises from a conformational change in the MD2 ligand for Mo=O that increases the distance from the pterin rings to the Lys679, eliminating a hydrogen bonding interaction (Supporting Information Figure S12). Conversely, the differences in structure 8 can instead be attributed to Arg333 alone, which forms a charge-assisted hydrogen bond with the phosphate group of MD1 for the Mo=O and not Mo=S or Mo=Se (Supporting Information Figure S13). On the other hand, for Mo=O structure 6, Lys679 lacks a hydrogen

bonding interaction with a carbonyl group and no charge-assisted hydrogen bond forms, leading to a more stabilizing ESP at the metal center for Mo=O than observed for structure 8.

Interestingly, in one case (i.e., structure 10), it is instead Mo=Se that deviates from Mo=O and Mo=S structures by having a relatively higher ESP than the other two cases (i.e., Se ΔESP(1-10) = -4 $kJ{\cdot}mol^{-1}{\cdot}e^{-1}$ vs O ΔESP(1-10) = -30 $kJ{\cdot}mol^{-1}{\cdot}e^{-1}$). A key difference can be attributed to the fact that the Mo=Se structure has lost the hydrogen bond between the sidechain of Arg333 and MD1 phosphate group. While for the majority of the other structures Gln301 had formed a hydrogen bond with the backbone of Mo-coordinating Cys140, this interaction is not present in structure 10 (Figure 8). For Mo=O and Mo=S this interaction is replaced with a hydrogen bond between the Gln301 sidechain and one of the MD1 phosphate groups, but no such interaction is present for Mo=Se (Figure 8).

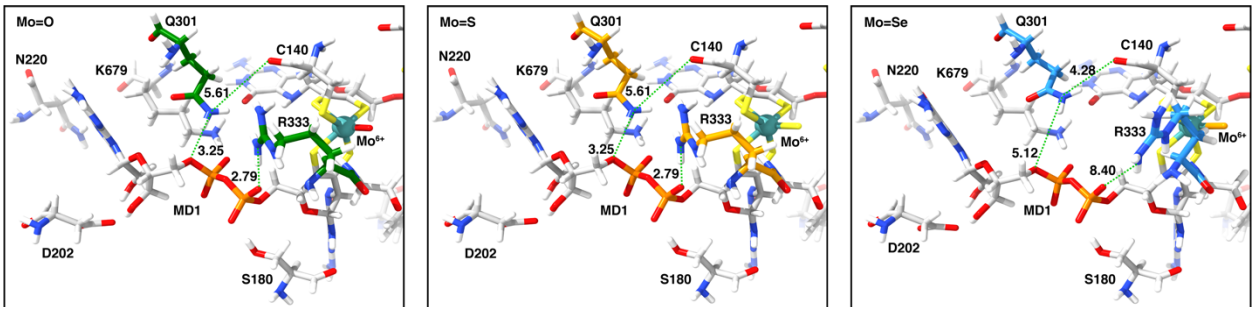

Figure 8. The optimized Mo=O (green), Mo=S (orange), and Mo=Se (blue) structures for structure 10 with key amino acids, here Gln301, colored with the rest of the following colors: carbon in white, oxygen in red, nitrogen in blue, phosphor in orange, sulfurs in yellow, seleno in light orange and Mo in cyan. Relevant hydrogen bonding distances are shown with a light green dashed line and annotated distances in Å colored the same as the relevant structures. (left) Mo=O, (middle) Mo=S, (right) Mo=Se.

Overall, orientation of residues sidechains or ligand conformations can impact the ESP at the metal center, altering the relative trends in ESP by metal center or terminal chalcogen identity from the average, expected trends. The most important residues identified in this analysis for FDH are Lys679 and Arg333, which can impact the ESP at the metal center by their hydrogen bond interactions with molybdopterins. Lys679 resides between the two

molybdopterins (i.e., MD1 and MD2) and forms a hydrogen bond with each of them. The interaction of Lys679's sidechain with MD1 is a charge-assisted hydrogen bond and it remains in all structures. Conversely, the hydrogen bonding interaction of Lys679 with MD2 is not a charge-assisted HB. In structures where reorientation of MD2 eliminates this interaction, the relative ESP is less stabilizing. Arg333 resides near the core active site, and, in some structures, its sidechain orients toward the phosphate groups of the MD1 and makes a charge-assisted hydrogen bond. The stronger charge-assisted hydrogen bond controls the position of the negative charge of the phosphate group, making the ESP less favorable at the metal center relative to those that lack this interaction. Thus, dynamic fluctuations that can be expected to occur in the enzyme play an important role in determining the electrostatic potential at the FDH active site, potentially limiting the biomimicry that can be obtained with molecular catalysts designed to resemble the FDH cofactor.

4. Conclusions.

We investigated the impact of the greater protein environment on the electronic properties of the active site of Mo/W containing FDH, which reversibly catalyzes the conversion of CO_2 to formate using a combination of long-timescale classical MD and large-scale QM/MM simulations. To systematically account for residues that influence the electronic properties of the cofactor, we employed the CSA method to probe the reorganization of charge density following extraction of the cofactor from the holoenzyme. This method identified 13 key amino acids that underwent meaningful charge reorganization following removal of the cofactor including three positively charged and two negatively charged amino acids. All residues included in the final QM region were polar except the special case of Cys136, which we designate as nonpolar in this work, and Gly334.

To better understand the electronic environment of the active site, we also studied the ESP at the metal center for Mo and W. For all structures, the ESP is lower for a Mo center than a W center consistent with the larger positive partial charge on Mo compared to W. Comparison of the ESP contributions from the QM and MM regions reveals that stabilization of the ESP by the QM region is almost four times that of the MM region, reinforcing the benefit of using the CSA method to select the QM region. To better understand the relative contributions of the QM region to the ESP, we decomposed the QM region into its components: the surrounding uncoordinated residues (i.e., negatively, positively, and neutral residues) and the metal binding residues (i.e., molybdopterins, terminal chalcogen ligand, and Cys140). From this analysis, we observed that the coordinating species (i.e., molybdopterins, sulfido) and coordinating residues have a greater influence over non-covalent residues in their contributions to the metal-center ESP. Moreover, among the sulfur atoms coordinating the metal center, the terminal sulfido ligand contributes most to the ESP of the metal. In addition to the catalytically essential sulfido, the Cys140 sulfur contributes more to the ESP than either molybdopterin. Overall, the large contribution of the terminal sulfido ligand to the ESP of the Mo^{6+} center suggests that it may play an important role in the enzymatic activity of the cofactor.

Given the significant contribution of the terminal sulfido to the ESP at the metal center, we sought to determine the sensitivity of this observation to replacement of the terminal chalcogen (i.e., O or Se) as well as the identity of the metal (i.e., Mo or W). Further investigation demonstrated that despite differences in size and electronegativity, the ESP values obtained for the Mo and W centers follow the same trend observed for structures with different terminal chalcogen ligands and ESP is more negative for Mo than W. As could be expected from differences in electronegativity, ESPs are less negative and similar for Se and S terminal

chalcogens than for O. We also observed that there was a relationship between the ESP and the number of hydrogen bonds between surrounding residues (e.g., Lys679 and Arg333) and coordinated ligands. These observations suggest opportunities (i.e., terminal chalcogen identities) and limitations (i.e., lack of enzyme environment) in mimicking the enzyme action of FDH to design novel molecular catalysts for CO_2 conversion.

AUTHOR INFORMATION

Corresponding Author

*email: hjkulik@mit.edu phone: 617-253-4584

Notes

The authors declare no competing financial interest.

ASSOCIATED CONTENT

Supporting Information. Comparison of partial charge schemes, a list of the amino acids selected for performing the charge shift analysis, charge shift analysis scheme, distances between Pro177, Ala178, and Leu406 residues, and the cofactor, distances between residues and cofactor, distances between Cys136 and Gly334 residues, and the cofactor, distances between Arg138 and Glu403 residues, and the cofactor, ESP at Mo^{6+} and W^{6+} (single point calculations), ESP at Mo^{6+} and W^{6+} (optimizations), HB interactions between charged residues and molybdopterins distances and angles between charged residues and MD1/MD2, distribution of ΔE values, representative snapshots from MD, distribution of ESP at Mo and W centers, values for ESP at Mo and W centers, Hirshfeld partial charges at Mo and W, distribution of ESP at Mo and W

centers, aligned distributions for three series (O, S, Se), Asp202 orientation for snapshots 4, MD2 conformations for snapshot 6, Arg333 orientation for snapshot 8 (PDF)

Initial AMBER topology and coordinate files for the FDH enzyme (ZIP)

ACKNOWLEDGMENT

The authors acknowledge funding from ExxonMobil Research and Engineering. H.J.K. holds a Career Award at the Scientific Interface from the Burroughs Wellcome Fund and an AAAS Marion Milligan Mason Award, which supported this work.

REFERENCES

(1) Koeller, K. M.; Wong, C.-H. Enzymes for Chemical Synthesis. *Nature* **2001,** *409*, 239.
(2) HE Schoemaker; D Mink; Wubbolts, M. Dispelling the Myths—Biocatalysis in Industrial Synthesis. *Science* **2003,** *299*, 1694-1697.
(3) Gamba, I. Biomimetic Approach to Co2 Reduction. *Bioinorg. Chem. Appl.* **2018,** *2018*, 1-14.
(4) Marchetti, L.; Levine, M. Biomimetic Catalysis. *ACS Catal.* **2011,** *1*, 1090-1118.
(5) Cook, S. A.; Hill, E. A.; Borovik, A. S. Lessons from Nature: A Bio-Inspired Approach to Molecular Design. *Biochemistry* **2015,** *54*, 4167-4180.
(6) Laureanti, J. A.; Ginovska, B.; Buchko, G. W.; Schenter, G. K.; Hebert, M.; Zadvornyy, O. A.; Peters, J. W.; Shaw, W. J. A Positive Charge in the Outer Coordination Sphere of an Artificial Enzyme Increases Co2 Hydrogenation. *Organometallics* **2020,** *39*, 1532-1544.
(7) Cotton, C. A. R.; Edlich-Muth, C.; Bar-Even, A. Reinforcing Carbon Fixation: Co2 Reduction Replacing and Supporting Carboxylation. *Curr. Opin. Biotechnol.* **2018,** *49*, 49-56.
(8) Marpani, F.; Pinelo, M.; Meyer, A. S. Enzymatic Conversion of Co 2 to Ch 3 Oh Via Reverse Dehydrogenase Cascade Biocatalysis: Quantitative Comparison of Efficiencies of Immobilized Enzyme Systems. *Biochem. Eng. J.* **2017,** *127*, 217-228.
(9) Sultana, S.; Chandra Sahoo, P.; Martha, S.; Parida, K. A Review of Harvesting Clean Fuels from Enzymatic Co2 Reduction. *RSC Adv.* **2016,** *6*, 44170-44194.
(10) Wang, W.-H.; Himeda, Y.; Muckerman, J. T.; Manbeck, G. F.; Fujita, E. Co2 Hydrogenation to Formate and Methanol as an Alternative to Photo- and Electrochemical Co2 Reduction. *Chem. Rev.* **2015,** *115*, 12936-12973.
(11) Crawford, A. M.; Cotelesage, J. J. H.; Prince, R. C.; George, G. N. In *Metallocofactors That Activate Small Molecules: With Focus on Bioinorganic Chemistry*; Ribbe, Markus W., Ed.; Springer International Publishing: Cham, 2019, DOI:10.1007/430_2018_30 10.1007/430_2018_30,63-100

(12) Cordas, C. M.; Moura, J. J. G. Molybdenum and Tungsten Enzymes Redox Properties – a Brief Overview. *Coord. Chem. Rev.* **2019,** *394*, 53-64.
(13) Maia, L. B.; Moura, I.; Moura, J. J. G. Molybdenum and Tungsten-Containing Formate Dehydrogenases: Aiming to Inspire a Catalyst for Carbon Dioxide Utilization. *Inorg. Chim. Acta* **2017,** *455*, 350-363.
(14) Maia, L. B.; Fonseca, L.; Moura, I.; Moura, J. J. G. Reduction of Carbon Dioxide by a Molybdenum-Containing Formate Dehydrogenase: A Kinetic and Mechanistic Study. *J. Am. Chem. Soc.* **2016,** *138*, 8834-8846.
(15) Goeppert, A.; Czaun, M.; Jones, J.-P.; Surya Prakash, G. K.; Olah, G. A. Recycling of Carbon Dioxide to Methanol and Derived Products – Closing the Loop. *Chem. Soc. Rev.* **2014,** *43*, 7995-8048.
(16) Appel, A. M.; Bercaw, J. E.; Bocarsly, A. B.; Dobbek, H.; DuBois, D. L.; Dupuis, M.; Ferry, J. G.; Fujita, E.; Hille, R.; Kenis, P. J. A.et al. Frontiers, Opportunities, and Challenges in Biochemical and Chemical Catalysis of Co2 Fixation. *Chem. Rev.* **2013,** *113*, 6621-6658.
(17) Hille, R.; Hall, J.; Basu, P. The Mononuclear Molybdenum Enzymes. *Chem. Rev.* **2014,** *114*, 3963-4038.
(18) Yishai, O.; Lindner, S. N.; Cruz, J. G. d. l.; Tenenboim, H.; Bar-Even, A. The Formate Bio-Economy. *Curr. Opin. Chem. Biol.* **2016,** *35*, 1-9.
(19) Gomez-Mingot, M.; Porcher, J.-P.; Todorova, T. K.; Fogeron, T.; Mellot-Draznieks, C.; Li, Y.; Fontecave, M. Bioinspired Tungsten Dithiolene Catalysts for Hydrogen Evolution: A Combined Electrochemical, Photochemical, and Computational Study. *J. Phys. Chem. B* **2015,** *119*, 13524-13533.
(20) Shiekh, B. A.; Kaur, D.; Kumar, S. Bio-Mimetic Self-Assembled Computationally Designed Catalysts of Mo and W for Hydrogenation of Co2/Dehydrogenation of Hcooh Inspired by the Active Site of Formate Dehydrogenase. *Phys. Chem. Chem. Phys.* **2019,** *21*, 21370-21380.
(21) Chakraborty, S.; Blacque, O.; Berke, H. Ligand Assisted Carbon Dioxide Activation and Hydrogenation Using Molybdenum and Tungsten Amides. *Dalton Trans.* **2015,** *44*, 6560-6570.
(22) Zhang, Y.; Williard, P. G.; Bernskoetter, W. H. Synthesis and Characterization of Pincer-Molybdenum Precatalysts for Co2 Hydrogenation. *Organometallics* **2016,** *35*, 860-865.
(23) Rebelein, J. G.; Hu, Y.; Ribbe, M. W. Differential Reduction of Co2 by Molybdenum and Vanadium Nitrogenases. *Angew. Chem., Int. Ed.* **2014,** *53*, 11543-11546.
(24) Siek, S.; Burks, D. B.; Gerlach, D. L.; Liang, G.; Tesh, J. M.; Thompson, C. R.; Qu, F.; Shankwitz, J. E.; Vasquez, R. M.; Chambers, N.et al. Iridium and Ruthenium Complexes of N-Heterocyclic Carbene- and Pyridinol-Derived Chelates as Catalysts for Aqueous Carbon Dioxide Hydrogenation and Formic Acid Dehydrogenation: The Role of the Alkali Metal. *Organometallics* **2017,** *36*, 1091-1106.
(25) Mazzotta, M. G.; Xiong, M.; Abu-Omar, M. M. Carbon Dioxide Reduction to Silyl-Protected Methanol Catalyzed by an Oxorhenium Pincer Pnn Complex. *Organometallics* **2017,** *36*, 1688-1691.
(26) Filonenko, G. A.; Conley, M. P.; Copéret, C.; Lutz, M.; Hensen, E. J. M.; Pidko, E. A. The Impact of Metal–Ligand Cooperation in Hydrogenation of Carbon Dioxide Catalyzed by Ruthenium Pnp Pincer. *ACS Catal.* **2013,** *3*, 2522-2526.

(27) Barrett, S. M.; Slattery, S. A.; Miller, A. J. M. Photochemical Formic Acid Dehydrogenation by Iridium Complexes: Understanding Mechanism and Overcoming Deactivation. *ACS Catal.* **2015,** *5*, 6320-6327.
(28) Dubey, A.; Nencini, L.; Fayzullin, R. R.; Nervi, C.; Khusnutdinova, J. R. Bio-Inspired Mn(I) Complexes for the Hydrogenation of Co2 to Formate and Formamide. *ACS Catal.* **2017,** *7*, 3864-3868.
(29) Bielinski, E. A.; Lagaditis, P. O.; Zhang, Y.; Mercado, B. Q.; Würtele, C.; Bernskoetter, W. H.; Hazari, N.; Schneider, S. Lewis Acid-Assisted Formic Acid Dehydrogenation Using a Pincer-Supported Iron Catalyst. *J. Am. Chem. Soc.* **2014,** *136*, 10234-10237.
(30) Tondreau, A. M.; Boncella, J. M. 1,2-Addition of Formic or Oxalic Acid to –N{Ch2ch2(Pipr2)}2-Supported Mn(I) Dicarbonyl Complexes and the Manganese-Mediated Decomposition of Formic Acid. *Organometallics* **2016,** *35*, 2049-2052.
(31) Dong, G.; Ryde, U. Reaction Mechanism of Formate Dehydrogenase Studied by Computational Methods. *JBIC, J. Biol. Inorg. Chem.* **2018,** *23*, 1243-1254.
(32) Maia, L. B.; Moura, J. J. G.; Moura, I. Molybdenum and Tungsten-Dependent Formate Dehydrogenases. *JBIC, J. Biol. Inorg. Chem.* **2014,** *20*, 287-309.
(33) Hartmann, T.; Schwanhold, N.; Leimkühler, S. Assembly and Catalysis of Molybdenum or Tungsten-Containing Formate Dehydrogenases from Bacteria. *Biochim. Biophys. Acta, Proteins Proteomics* **2015,** *1854*, 1090-1100.
(34) Schrapers, P.; Hartmann, T.; Kositzki, R.; Dau, H.; Reschke, S.; Schulzke, C.; Leimkühler, S.; Haumann, M. Sulfido and Cysteine Ligation Changes at the Molybdenum Cofactor During Substrate Conversion by Formate Dehydrogenase (Fdh) from Rhodobacter Capsulatus. *Inorg. Chem.* **2015,** *54*, 3260-3271.
(35) Boyington, J. C.; Gladyshev, V. N.; Khangulov, S. V.; Stadtman, T. C.; Sun, P. D. Crystal Structure of Formate Dehydrogenase H: Catalysis Involving Mo, Molybdopterin, Selenocysteine, and an Fe,S, Cluster. *Science* **1997,** *275*, 1305-1308.
(36) Groysman, S.; Holm, R. H. Synthesis and Structures of Bis(Dithiolene)Tungsten(Iv,Vi) Thiolate and Selenolate Complexes Approaches to the Active Sites of Molybdenum and Tungsten Formate Dehydrogenases. *Inorg. Chem.* **2007,** *46*, 4090.
(37) Donahue, J. P.; Goldsmith, C. R.; Nadiminti, U.; Holm, R. H. Synthesis, Structures, and Reactivity of Bis(Dithiolene)Molybdenum (Iv,Vi) Complexes Related to the Active Sites of Molybdoenzymes. *J. Am. Chem. Soc.* **1998,** *120*, 12689-12881.
(38) Musgrave, K. B.; Lim, B. S.; Sung, K. M.; Holm, R. H.; Hedman, B.; Hodgson, K. O. X-Ray Spectroscopy of Enzyme Active Site Analogues and Related Molecules: Bis(Dithiolene)Molybdenum(Iv) and -Tungsten(Iv,Vi) Complexes with Variant Terminal Ligands. *Inorg. Chem.* **2000,** *39*, 5238-5247.
(39) Jiang, J. F.; Holm, R. H. An Expanded Set of Functional Groups in Bis(Dithiolene)Tungsten(Iv,Vi) Complexes Related to the Active Sites of Tungstoenzymes, Including W-Iv-Sr and W-Vi-O(Sr). *Inorg. Chem.* **2004,** *43*, 1302.
(40) Basu, P.; Nemykin, V. N.; Sengar, R. S. Syntheses, Spectroscopy, and Redox Chemistry of Encapsulated Oxo-Mo(V) Centers: Implications for Pyranopterin-Containing Molybdoenzymes. *Inorg. Chem.* **2003,** *42*, 7489.
(41) Niks, D.; Hille, R. Molybdenum ‐ and Tungsten ‐ Containing Formate Dehydrogenases and Formylmethanofuran Dehydrogenases: Structure, Mechanism, and Cofactor Insertion. *Protein Sci.* **2018,** *28*, 111-122.

(42) Raaijmakers, H. C. A.; Romão, M. J. Formate-Reduced E. Coli Formate Dehydrogenase H: The Reinterpretation of the Crystal Structure Suggests a New Reaction Mechanism. *JBIC, J. Biol. Inorg. Chem.* **2006,** *11*, 849-854.
(43) Kirk, M. L.; Kc, K. In *Metal Ions in Life Sciences: Transition Metals and Sulfur - a Strong Relationship for Life*; Sosa Torres, Martha E. ;Kroneck, Peter M. H. , Eds.; De Gruyter, 2020; Vol. 20,313-342
(44) Cui, Q.; Pal, T.; Xie, L. Biomolecular Qm/Mm Simulations: What Are Some of the "Burning Issues"? *J. Phys. Chem. B* **2021,** *125*, 689-702.
(45) Welborn, V. V.; Head-Gordon, T. Fluctuations of Electric Fields in the Active Site of the Enzyme Ketosteroid Isomerase. *J. Am. Chem. Soc.* **2019,** *141*, 12487-12492.
(46) Ryde, U. How Many Conformations Need to Be Sampled to Obtain Converged Qm/Mm Energies? The Curse of Exponential Averaging. *J. Chem. Theory Comput.* **2017,** *13*, 5745-5752.
(47) Hu, L.; Söderhjelm, P. r.; Ryde, U. On the Convergence of Qm/Mm Energies. *J. Chem. Theory Comput.* **2011,** *7*, 761-777.
(48) Wang, L.; Fried, S. D.; Markland, T. E. Proton Network Flexibility Enables Robustness and Large Electric Fields in the Ketosteroid Isomerase Active Site. *J. Phys. Chem. B* **2017,** *121*, 9807-9815.
(49) Bím, D.; Alexandrova, A. N. Local Electric Fields as a Natural Switch of Heme-Iron Protein Reactivity. *ACS Catal.* **2021,** *11*, 6534-6546.
(50) Field, M. J.; Bash, P. A.; Karplus, M. A Combined Quantum-Mechanical and Molecular Mechanical Potential for Molecular-Dynamics Simulations. *J. Comput. Chem.* **1990,** *11*, 700–733.
(51) Bakowies, D.; Thiel, W. Hybrid Models for Combined Quantum Mechanical and Molecular Mechanical Approaches. *J. Phys. Chem.* **1996,** *100*, 10580–10594.
(52) Warshel, A.; Levitt, M. Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.* **1976,** *103*, 227–249.
(53) Senn, H. M.; Thiel, W. Qm/Mm Methods for Biomolecular Systems. *Angew. Chem., Int. Ed.* **2009,** *48*, 1198–1229.
(54) Gao, J.; Truhlar, D. G. Quantum Mechanical Methods for Enzyme Kinetics. *Annu. Rev. Phys. Chem.* **2002,** *53*, 467-505.
(55) Himo, F. Recent Trends in Quantum Chemical Modeling of Enzymatic Reactions. *J. Am. Chem. Soc.* **2017,** *139*, 6780–6786.
(56) Ufimtsev, I. S.; Luehr, N.; Martínez, T. J. Charge Transfer and Polarization in Solvated Proteins from Ab Initio Molecular Dynamics. *J. Phys. Chem. Lett.* **2011,** *2*, 1789–1793.
(57) Kulik, H. J. Large-Scale, Qm/Mm Free Energy Simulations of Enzyme Catalysis Reveal the Influence of Charge Transfer. *Phys. Chem. Chem. Phys.* **2018,** *20*, 20650–20660.
(58) Mehmood, R.; Kulik, H. J. Both Configuration and Qm Region Size Matter: Zinc Stability in Qm/Mm Models of DNA Methyltransferase. *J. Chem. Theory Comput.* **2020,** *16*, 3121-3134.
(59) Yang, Z.; Mehmood, R.; Wang, M.; Qi, H. W.; Steeves, A. H.; Kulik, H. J. Revealing Quantum Mechanical Effects in Enzyme Catalysis with Large-Scale Electronic Structure Simulation. *React. Chem. Eng.* **2019,** *4*, 298-315.

(60) Hu, L.; Soderhjelm, P.; Ryde, U. Accurate Reaction Energies in Proteins Obtained by Combining Qm/Mm and Large Qm Calculations. *J. Chem. Theory Comput.* **2012,** *9*, 640-649.
(61) Morgenstern, A.; Jaszai, M.; Eberhart, M. E.; Alexandrova, A. N. Quantified Electrostatic Preorganization in Enzymes Using the Geometry of the Electron Charge Density. *Chemical Science* **2017,** *8*, 5010-5018.
(62) Vennelakanti, V.; Nazemi, A.; Mehmood, R.; Steeves, A. H.; Kulik, H. J. Harder, Better, Faster, Stronger: Large-Scale Qm and Qm/Mm for Predictive Modeling in Enzymes and Proteins. *Curr. Opin. Struct. Biol.* **2022,** *72*, 9-17.
(63) Karelina, M.; Kulik, H. J. Systematic Quantum Mechanical Region Determination in Qm/Mm Simulation. *J. Chem. Theory Comput.* **2017,** *13*, 563-576.
(64) Kulik, H. J.; Zhang, J.; Klinman, J. P.; Martínez, T. J. How Large Should the Qm Region Be in Qm/Mm Calculations? The Case of Catechol O-Methyltransferase. *J. Phys. Chem. B* **2016,** *120*, 11381-11394.
(65) Zheng, M.; Waller, M. P. Yoink: An Interaction‐Based Partitioning Api. *J. Comput. Chem.* **2018,** *39*, 799-806.
(66) Waller, M. P.; Kumbhar, S.; Yang, J. A Density‐Based Adaptive Quantum Mechanical/Molecular Mechanical Method. *ChemPhysChem* **2014,** *15*, 3218-3225.
(67) Zheng, M.; Waller, M. P. Adaptive Quantum Mechanics/Molecular Mechanics Methods. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2016,** *6*, 369-385.
(68) Hix, M. A.; Leddin, E. M.; Cisneros, G. A. Combining Evolutionary Conservation and Quantum Topological Analyses to Determine Quantum Mechanics Subsystems for Biomolecular Quantum Mechanics/Molecular Mechanics Simulations. *J. Chem. Theory Comput.* **2021,** *17*, 4524-4537.
(69) Summers, T. J.; Daniel, B. P.; Cheng, Q.; DeYonker, N. J. Quantifying Inter-Residue Contacts through Interaction Energies. *J. Chem. Inf. Model.* **2019,** *59*, 5034-5044.
(70) Brandt, F.; Jacob, C. Systematic Qm Region Construction in Qm/Mm Calculations Based on Uncertainty Quantification. *ChemRxiv* **2021**.
(71) Liu, M.; Nazemi, A.; Taylor, M.; Nandy, A.; Duan, C.; Steeves, A.; Kulik, H. Large-Scale Screening Reveals That the Geometric Structure Matters More Than the Electronic Structure in the Bioinspired Catalyst Design of Formate Dehydrogenase Mimics. *ACS Catal.* **2022,** *12*, 383-396.
(72) Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V. H++ 3.0: Automating Pk Prediction and the Preparation of Biomolecular Structures for Atomistic Molecular Modeling and Simulations. *Nucleic Acids Res.* **2012,** *40*, W537-W541.
(73) Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A. H++: A Server for Estimating Pkas and Adding Missing Hydrogens to Macromolecules. *Nucleic Acids Res.* **2005,** *33*, W368-W371.
(74) Myers, J.; Grothaus, G.; Narayanan, S.; Onufriev, A. A Simple Clustering Algorithm Can Be Accurate Enough for Use in Calculations of Pks in Macromolecules. *Proteins: Struct., Funct., Bioinf.* **2006,** *63*, 928-938.
(75) Laban, J.; Granzin, J.; Schlukebie, G.; Robinson, D. P.; Jack, W. E.; Schildkraut, I.; Saenger, W. Three-Dimensional Structure of the Adenine-Specific DNA Methyltransferase Mtaq I in Complex with the Cofactor S-Adenosylmethionine. *Proc. Natl. Acad. Sci. U. S. A.* **1994,** *91*, 10957-10961.

(76) Li, P.; Merz, K. M. Mcpb.Py: A Python Based Metal Center Parameter Builder. *J. Chem. Inf. Model.* **2016,** *56*, 599-604.
(77) Case, D. A. B.-S., I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, I., T.E.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Greene, D.; Harris, R.; Homeyer, N.; Huang, Y.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D. J.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Smith, J.; Salomon-Ferrer, R.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. Amber 2018, University of California, San Francisco. 2018. http://www.ambermd.org. (Accessed December 7, 2021).
(78) Ferreira, P.; Cerqueira, N. M. F. S. A.; Brás, N. F.; Fernandes, P. A.; Ramos, M. J. Parametrization of Molybdenum Cofactors for the Amber Force Field. *J. Chem. Theory Comput.* **2018,** *14*, 2538-2548.
(79) James A. Maier, C. M., Koushik Kasavajhala,; Lauren Wickstrom, K. H., and Carlos Simmerling. Ff14sb: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99sb. *J. Chem. Theory Comput.* **2015,** *14* 2538-2548.
(80) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983,** *79*, 926-935.
(81) Ryckaert J-P, C. G., Berendsen HJC. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. *J. Comput. Phys.* **1977,** *23*, 327-341.
(82) Schrodinger LLC. Pymol Molecular Graphics System. 2021. http://www.pymol.org/. (Accessed December 7, 2021).
(83) Ufimtsev IS, M. T. Quantum Chemistry on Graphical Processing Units. 3. Analytical Energy Gradients, Geometry Optimization, and First Principles Molecular Dynamics. *J. Chem. Theory Comput.* **2009,** *5*, 2619-2628.
(84) Rohrdanz, M. A.; Martins, K. M.; Herbert, J. M. A Long-Range-Corrected Density Functional That Performs Well for Both Ground-State Properties and Time-Dependent Density Functional Theory Excitation Energies, Including Charge-Transfer Excited States. *J. Chem. Phys.* **2009,** *130*, 054112.
(85) Hay PJ, W. W. Ab Initio Effective Core Potetentials for Molecular Calculations. Potentials for the Transition Metal Atoms Sc to Hg. *J. Chem. Phys.* **1985,** *82*, 270−283.
(86) Hariharan PC, P. J. The Influence of Polarization Functions on Molecular-Orbital Hydrogenation Energie. *Theor. Chim. Acta* **1973,** *13*, 563-576.
(87) Lange, A. W.; Herbert, J. M. A Smooth, Nonsingular, and Faithful Discretization Scheme for Polarizable Continuum Models: The Switching/Gaussian Approach. *J. Chem. Phys.* **2010,** *133*, 244111.
(88) York, D. M.; Karplus, M. A Smooth Solvation Potential Based on the Conductor-Like Screening Model. *J. Phys. Chem. A* **1999,** *103*, 11060-11079.
(89) Liu, F.; Luehr, N.; Kulik, H. J.; Martinez, T. J. Quantum Chemistry for Solvated Molecules on Graphical Processing Units Using Polarizable Continuum Models. *J. Chem. Theory Comput.* **2015,** *11*, 3131-3144.

(90) Liu, F.; Sanchez, D. M.; Kulik, H. J.; Martinez, T. J. Exploiting Graphical Processing Units to Enable Quantum Chemistry Calculation of Large Solvated Molecules with Conductor-Like Polarizable Continuum Models. *Int. J. Quantum Chem.* **2019,** *119*, e25760.

For Table of Contents Use Only