Both Configuration and QM Region Size Matter: Zinc Stability in QM/MM Models of DNA Methyltransferase

Rimsha Mehmood[1,2] and Heather J. Kulik[1,*]

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

[2]*Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139*

ABSTRACT: Quantum-mechanical/molecular-mechanical (QM/MM) methods are essential to the study of metalloproteins, but the relative importance of sampling and degree of QM treatment in achieving quantitative predictions is poorly understood. We study the relative magnitude of configurational and QM-region sensitivity of energetic and electronic properties in a representative Zn^{2+} metal binding site of a DNA methyltransferase. To quantify property variations, we analyze snapshots extracted from 250 ns of molecular dynamics simulation. To understand the degree of QM-region sensitivity, we perform analysis using QM regions ranging from a minimal 49-atom region consisting only of the Zn^{2+} metal and its four coordinating Cys residues up to a 628-atom QM region that includes residues within 12 Å of the metal center. Over the configurations sampled, we observe that illustrative properties (e.g., rigid Zn^{2+} removal energy) exhibit large fluctuations that are well captured with even minimal QM regions. Nevertheless, for both energetic and electronic properties, we observe a slow approach to asymptotic limits with similarly large changes in absolute values that converge only with larger (ca. 300-atom) QM region sizes. For the smaller QM regions, the electronic description of Zn^{2+} binding is incomplete: the metal binds too tightly, is too stabilized by the strong electrostatic potential of MM point charges, and the Zn-S bond covalency is overestimated. Overall, this work suggests that efficient sampling with QM/MM in small QM regions is an effective method to explore the influence of enzyme structure on target properties. At the same time, accurate descriptions of electronic and energetic properties require a larger QM region than the minimal metal-coordinating residues in order to converge treatment of both metal-local bonding and the overall electrostatic environment.

1. Introduction

Multi-scale quantum mechanics/molecular mechanics (QM/MM) modeling[1-10] has emerged as the predominant method for simulating enzymes[11]. With a divide-and-conquer approach, computationally demanding QM simulation is applied to an essential region of interest, whereas the remainder of the protein is affordably described with MM, enabling a balance of computational cost and accuracy while still describing bond rearrangements, polarization, and charge transfer at a central active site.

Nevertheless, limitations remain in applying QM/MM for quantitatively predictive mechanistic study, especially on new enzymes where limited experimental guidance or chemical insight is available. The optimal QM region must be selected prior to study, and the higher formal scaling of typical QM methods (i.e., density functional theory or DFT) motivates restricting the size of the QM region. Thus, minimal QM regions of tens of atoms (i.e. ligands and a few direct residues)[9, 12-13] have most commonly been employed. The QM/MM boundary introduces artifacts into the simulation, although significant development[7, 14-23] has focused on reducing such effects and evaluating[24] the potential improvements that can be achieved by employing polarizable,[17, 24-29] instead of conventional electrostatic embedding. Even with these advances, small QM regions will restrict charge transfer between MM residues and the QM active site.[30-31]

When small QM regions are used in QM/MM simulation, they are sometimes adequate to predict physically reasonable mechanisms[32], but in other cases are missing critical residues needed to describe the essential enzyme action[33-34], and distinguishing these two scenarios is challenging. Despite the many successes of force fields essential for describing globular protein structure and dynamics, qualitatively incorrect descriptions in other cases (e.g., disordered

structures[35-36] and dynamic non-covalent interactions[37]) can limit the ability of predominantly MM-based models in small-QM region QM/MM simulation to reveal enzyme mechanism. Recent advances[30, 35, 38-44] in algorithmic efficiency and hardware have enabled quantum chemical simulation of polypeptides[35, 45] and large-scale QM (ca. 100-1000 atom) QM/MM simulation. Researchers[34, 46-62] have leveraged these advances to identify how sensitive mechanistic predictions are to QM region size in QM/MM calculations. The majority of resulting studies have revealed an exceptionally slow approach to asymptotic limits (ca. 500 atoms) for QM region property convergence including for: NMR shieldings,[51, 56] proton transfer,[47, 63] solvation effects,[48] barrier heights,[34, 54-55] forces[46], excitation energies,[52, 64-65] partial charges,[57] bond critical points[66], and electrostatic[67] or redox potentials[58-59]. In a number of cases, smaller QM regions or clusters[59, 68] have been justified (e.g., in DNA models[69] and the cytochrome P450cam metalloenzyme[70]), but more study is needed to understand when small QM regions are sufficient.

A related concern is the extent to which protein dynamical configurations affect QM/MM predictions. Naturally, as QM regions are enlarged, sampling becomes increasingly cost prohibitive. As a result, free energy simulations with enlarged QM regions have nearly exclusively been carried out with semi-empirical methods[46, 71-75] and only recently with first principles DFT[33]. Sampling configurations from classical MD and pairing them with QM/MM property evaluation has been motivated[60-61] as an efficient alternative, although this approach has been criticized when the MM-derived structures are poor approximations of relaxed QM structures[69]. Predictive QM/MM simulation should thus address not just the dependence on QM treatment but the need for sampling.

Metalloproteins present unique challenges for modeling[37, 53, 59, 68] due to the unique trade-

off between high levels of theory needed to adequately describe the metal-organic bond and the need to establish the dependence of properties on adequate sampling of variations in the greater protein environment. For this study, we select DNA methyltransferase as a prototypical metalloprotein in which Zn^{2+} plays an essential structural role[76-78]. DNMTs methylate[79] the 5' position of cytosine bases in CpG-rich islands in an essential epigenetic process[80]. Although Zn plays only a structural role, disruption of Zn binding alters DNA methylation and has been implicated in cancer[82-83]. As a result, Zn^{2+} binding has often been the focus of QM[84] and QM/MM[85] studies on DNA methyltransferases.[84-88] We choose this structural site for the focus of our own QM/MM study here to simplify comparisons of fluctuations in a single energetic quantity (i.e., Zn stability) and the extent of QM treatment without the complication of needing to compare multiple intermediates in a catalytic cycle. Nevertheless, we expect our observations will be relevant to other enzymes in which Zn plays a catalytic role or more broadly to other tetrahedrally-coordinated metals. We focus on the structurally-characterized[81] DNMT1, which has four Zn metal binding sites, three (Zn1, Zn2, and Zn4) adjacent to the DNA binding domain and one (Zn3) more distant (Figure 1). Zn is tetrahedrally coordinated, either by four Cys residues (Zn2 and Zn4) or in a three Cys/one His configuration (Zn1 and Zn3, Figure 1).

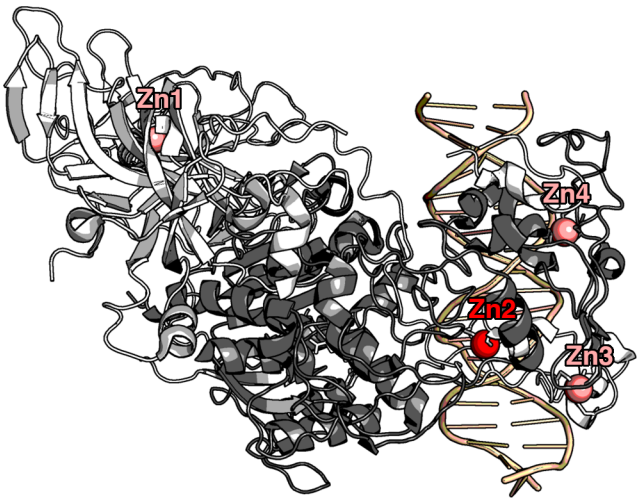

Figure 1. The structure of DNMT1 shown in cartoon form. The methyltransferase domain is colored as dark grey, the DNA strand in tan, and the rest of the protein is colored in light grey. The four Zn centers are shown as colored spheres, with the Zn2 studied in this work shown in bright red, whereas the other Zn centers (Zn1, Zn3, and Zn4) are shown in salmon.

To identify the relative importance of configuration and QM region in QM/MM descriptions of Zn^{2+} binding in DNMT1, we carry out a combination of classical molecular dynamics and large-scale QM/MM simulation. We evaluate the relative effect of configuration and QM region size on energetic and electronic descriptors of Zn^{2+} binding. Although recently developed systematic approaches[34, 54, 70, 89-91] have provided valuable insight into the most essential residues for QM treatment, they exhibit strong sensitivity to protein configuration[33, 92]. Thus, to evaluate the relative importance of QM region and configuration in a balanced manner, we instead start with a minimal QM region consisting of Zn^{2+} and its coordinating residues and then gradually expand the QM region radially outward.

The rest of this article is outlined as follows. In Sec. 2, we describe the Zn^{2+} binding site and residues incorporated in eight QM regions. In Sec. 3, we outline the computational details of this work. In Sec. 4, we evaluate configuration and QM region sensitivity of Zn^{2+} rigid binding energetics to DNMT1, Zn^{2+} charges, Zn-S bond order, key residue sidechain charges, the

electrostatic potential, as well as correlations among these properties. In Sec. 5, we provide our conclusions.

2. Protein Structure and QM Region Selection

We selected the Cys-coordinated Zn2 site of DNMT1 as the focus of our study on the effect of QM region on electronic properties (Figure 1). The Zn2 site is most buried of the four Zn^{2+}-binding sites, allowing us to increase the size of the QM region more significantly before reaching the protein boundary or other Zn sites than the alternative Zn1, Zn3, or Zn4 sites (see Figure 1). Starting from the DNMT1 X-ray crystal structure (PDB ID: 3PTA[81]), we constructed QM regions by radially increasing cutoffs around the Zn2 Zn^{2+} center in the DNMT1 crystal structure in integer values incremented by 1 Å (Figure 2). If any part of a residue was inside the radial cutoff, we included the entire residue sidechain and backbone atoms (Figure 2). Once sequent-adjacent residues are included in a QM region, this approach naturally incorporates the amide bond between them, asymptotically approaching a full QM treatment of the protein.

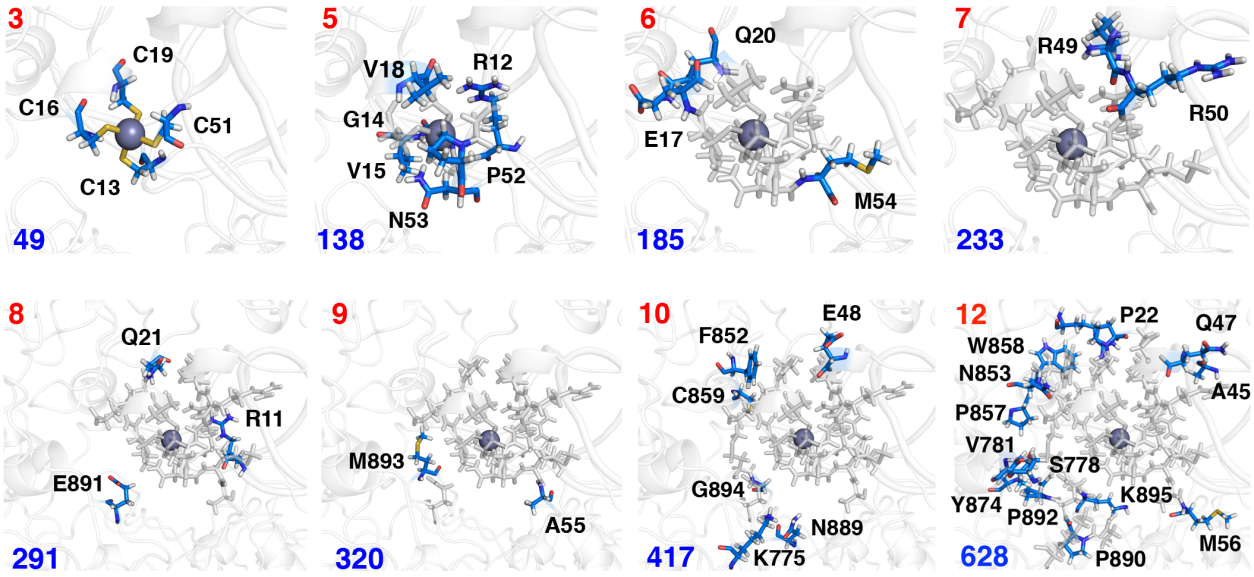

Figure 2. Structures of the DNMT1 Zn2 binding site indicating which residues are added in each increasingly large QM region from left to right and top to bottom. The QM region Zn^{2+} radial cutoff (in Å) is indicated in red at the top left of each pane, and the number of atoms including link atoms is indicated at the bottom left of each pane in blue. The residues added in each QM region are indicated with single letter residue codes and numbers and shown as sticks with

carbon atoms in light blue, oxygen in red, hydrogen in white, sulfur in yellow, and nitrogen in blue. The QM atoms from smaller regions, where applicable, are shown as white sticks, Zn^{2+} is shown as a sphere, and remaining MM atoms are shown in translucent white cartoon representation.

The smallest, minimal QM region considered in this work consists of the 49 atoms from Zn^{2+} and its negatively charged coordinating Cys residues, which corresponds to a 3-Å radial cutoff and a net charge of -2 (Figure 2). The 138-atom, 5-Å QM region adds sequence-adjacent residues to the Cys that are primarily hydrophobic (G14, V15, V18, and P52) and positively charged (R12), reducing the net charge on the QM region to -1 (Figure 2). A 4-Å cutoff would yield the same QM region as this 5-Å cutoff, thus we study only a single set of residues and refer to it as the 5-Å region. We selected six additional QM region radial cutoffs (6-10 Å and 12 Å) for a total of eight QM regions that span an order of magnitude (49 to 628 atoms) in size (Figure 3 and Supporting Information Table S1). For the largest, 12-Å QM region, in addition to distance considerations, we also aimed to preserve the charge neutrality achieved in the intermediate sized (i.e., 7-10 Å) QM regions (Figure 3). We thus excluded R10 and R784 that partly reside inside the radial cutoff as well as C46, which is negatively charged and coordinates the Zn3 site Zn^{2+} (Figure 1). For the remaining QM regions, no modifications are made, as increasing the radial cutoff generally adds a balance of hydrophobic, polar, and charged residues without strongly shifting the net charge of the QM region (Figures 2 and 3).

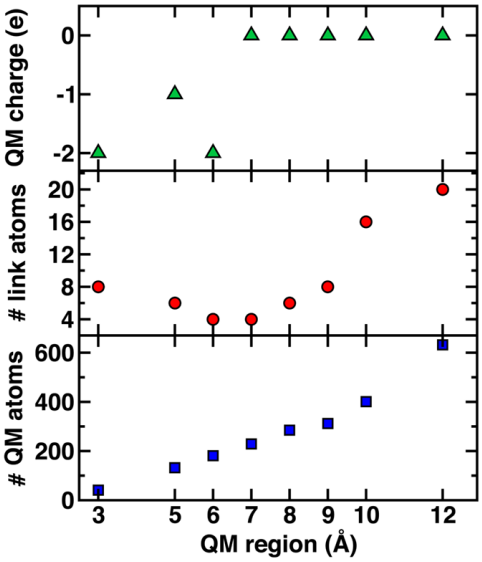

Figure 3. Characteristics of QM regions with increasing radial cutoff (in Å): the overall net QM region charge (in e, top), the number of link (i.e., passivating hydrogen) atoms (middle), and the number of atoms in the QM region (excluding link atoms, bottom).

Notably given their relatively low abundance in proteins[93], methionine residues are added in several QM regions (6 Å: M54, 9 Å: M893, 12 Å: M56), although not necessarily with sidechains oriented in the crystal structure toward Zn^{2+} (Figure 2). Given the charge neutrality achieved in most of these QM regions as well as the even distributions of residues, we expect these QM regions to be suitable for quantifying trends in electronic and energetic properties at the Zn site.

3. Computational Details

The preparation, MD simulation, and QM/MM calculation of Zn^{2+} binding energies in DNMT follows a previously introduced protocol[92], which we reiterate and expand upon for clarity. To isolate configurational effects, we employ MD for extensive sampling, and we obtain properties from QM/MM snapshots evaluated on snapshots from the MD trajectory. This approach neglects changes in the trajectory that could occur with *ab initio* MD sampling but ensures we capture changes in protein structure that could occur only at longer times.

Protein structure and preparation. Starting structures from Ref. 92 were obtained as

follows. A crystal structure of DNMT1 (PDB ID: 3PTA[81]) with a 19 base-pair DNA strand and S-adenosyl homocysteine (SAH) inhibitor was modified to replace SAH with S-adenosyl methionine (SAM) using Avogadro v1.2.0[94]. Fifty-four missing residues were added and refined with Modeller[95] loop refinement (Supporting Information Table S2). The protonation states of apoenzyme residues were assigned using the H++ webserver[96-99] assuming a pH of 7.0 with all other defaults applied, and protonation states of Zn^{2+}-coordinating-residues were manually assigned for a final holoenzyme net charge of -22 (Supporting Information Table S3). Standard protein residues were simulated with AMBER ff14SB[100], the zinc AMBER force field (ZAFF)[101] was employed for Zn^{2+} and coordinating residues, and the DNA base pairs were parameterized with the parmBSC1[102] force field (Supporting Information Table S4). For SAM, the generalized AMBER force field (GAFF)[103] was employed in conjunction with restrained electrostatic potential (RESP) charges[104] obtained using Hartree-Fock/6-31G*[105] evaluated with GAMESS-US[106], as implemented by the R.E.D. server[107-109]. The protein was solvated with no less than 15 Å of TIP3P[110] water buffer in a periodic rectangular prism box and neutralized with Na^+ counterions. The initial AMBER topology and coordinate files for the 151,473-atom system are provided in the Supporting Information.

MD Equilibration and Dynamics. The MD equilibration and production employed the GPU-accelerated form of the AMBER16 code[111]. This procedure included the following steps: 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) 250 ns of production MD. All MD employed SHAKE[112] with a 2-fs timestep and a 10-Å electrostatic cutoff.

QM/MM Calculations. For QM/MM calculations, the QM energetics were evaluated with a developer version of TeraChem[40, 113] and the MM energetics were evaluated with OpenMM[114] v.7.1.1, with TeraChem acting as the QM/MM driver. We used electrostatic embedding and hydrogen link atoms to passivate covalent bonds that span the QM-to-MM boundary.

The QM-level calculations employed DFT using the range-separated exchange-correlation functional ωPBEh[115] (ω=0.2 $bohr^{-1}$) with an LANL2DZ effective core potential[116] on Zn^{2+} and 6-31G*[105] for the remaining atoms. These choices were motivated by prior studies[34, 70] and through validation tests of the sensitivity of properties and their trends on basis set and DFT exchange-correlation choice (Supporting Information Tables S5-S6).

As described in Sec. 2, nine QM regions were selected based on increasing radial cutoffs, leading to QM region net charges that varied from -2 to 0 (see Figure 3 and Supporting Information Table S1). From production MD, 250 snapshots spaced 1 ns apart were selected for QM/MM single point energies of the holoenzyme and then repeated with the Zn^{2+} rigidly removed for a total of 500 single point energies per QM region considered. Snapshots were post-processed using the center of mass utility in PyMOL[117] to generate the largest possible spherical droplet circumscribed by the original periodic box for simulation in spherical boundary conditions in QM/MM without electrostatic cutoff. This spherical droplet was centered on the center of mass of the protein with a radius of 42 Å and contained 91,233 atoms. Representative snapshots are provided in the Supporting Information.

Analysis. Atom-wise and by-residue sums of Mulliken charges were evaluated on all QM/MM single point calculations, as obtained from TeraChem[40, 113]. For all QM regions, the wavefunction for each of the structures was post-processed with Multiwfn[118] to evaluate the Mayer bond orders[119-120] for Zn^{2+}-Cys bonds and the total bond valence, defined as sum of

Mayer bond orders with all coordinating atoms[120] around the Zn^{2+} center. For snapshots both with and without Zn^{2+}, the electrostatic potentials for each QM region were evaluated with Coulomb's law using Mulliken partial charges for atoms in the QM region and MM point charges assigned by the force field for the remaining atoms. The partial charge for Zn^{2+} was not included in any electrostatic potential calculations in order to evaluate the electrostatic potential at the Zn^{2+} site.

4. Results and Discussion

To probe changes in the electronic structure around the Zn^{2+} binding site in DNMT1 with increasing QM region size, we first carried out 250 ns of molecular dynamics simulation (see Sec. 3). From this trajectory, we extracted 250 snapshots spaced 1-ns apart to ensure they were not correlated. We then evaluated a series of electronic and energetic properties over which we examine the ability of smaller QM regions to capture the property values as well as their fluctuation with changes in configuration.

4a. Effects on Zn^{2+} Binding

We first studied the relationship between protein structure and the relative stability of Zn^{2+} in the Zn2 binding site of DNMT1. We computed Zn^{2+} rigid binding energies, ΔE, for all QM regions as:

$$\Delta E = E(\text{DNMT1-Zn}^{2+}) - E(\text{DNMT1}\cdots) - E(\text{Zn}^{2+}) \quad (1)$$

where the first two terms are the QM/MM single point energies of DNMT1 with Zn^{2+} bound and DNMT1 without Zn^{2+}, respectively, and the last term is the energy of the isolated Zn^{2+} ion. This quantity is a significant overestimate of the energy required to exchange Zn^{2+} in more stable forms (e.g., solvated Zn^{2+}) because Zn is removed without the protein allowed to relax. Nevertheless, we employ the ΔE quantity as an upper bound on the energy required to remove

Zn^{2+} and as a probe of both the QM-region and configurational sensitivity of the relative stability of Zn^{2+} in DNMT1.

As might be expected, the minimal, 3-Å QM region has a high mean ΔE value of 336 kcal/mol (Figure 4). This rigid removal of Zn^{2+} in the smallest QM region requires significant charge localization (i.e., the QM region charge changes from -2 to -4) onto only the Cys residues. Despite the expected high energetic cost of bare Zn^{2+} removal, a wide range of ΔE values is sampled from a minimum of -158 kcal/mol (at 192 ns, indicating the Zn-bound state is unstable due to close proximity of an Arg) to a maximum of 693 kcal/mol (165 ns) with a standard deviation (std. dev.) of 85 kcal/mol (Figure 4). Fluctuations are significant over the snapshots, and more significant outliers (i.e., > 1 std. dev.) occur below the mean than above, suggesting greater deviations in configurations where Zn^{2+} is unexpectedly weakly bound in this QM region (Figure 4). Since ΔE values are evaluated from differences in the energies of Zn^{2+}-bound and Zn^{2+}-absent structures, we determined whether outlier energies could be attributed to one of the two structures (Figure 4 and Supporting Information Figures S1-S2). Overall, both energies exhibit individual fluctuations that do not cancel and contribute to both weakly- and strongly-binding Zn^{2+} outlier configurations (Figure 4).

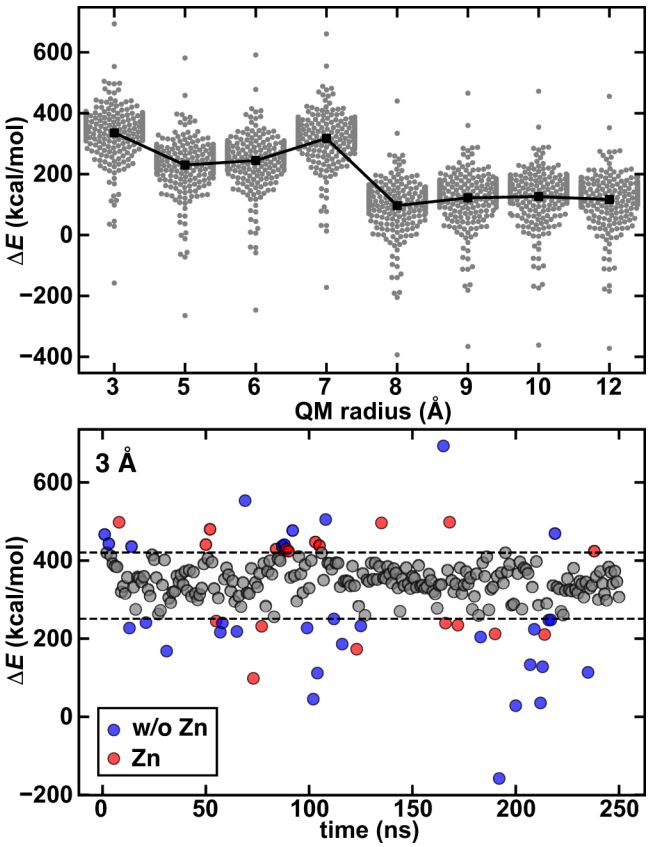

Figure 4. (top) The distribution of Δ*E* values (in kcal/mol) shown in swarm plots of 250 snapshots for all eight QM regions (3, 5, 6, 7, 8, 9, 10, and 12 Å) with mean values overlaid (black squares and lines). (bottom) Δ*E* values (in kcal/mol) for the minimal, 3-Å QM region over 250 MD snapshots. The dashed lines represent ± 1 standard deviation around the mean, and the data points within this range are shown as grey circles. Outlier points outside the dashed lines are colored blue if the deviation from the mean is larger for the energy coming from the snapshot without Zn^{2+} and red if the deviation from the mean is larger for the snapshot containing Zn^{2+}.

We repeated the evaluation of Δ*E* values with the other seven radially increasing QM regions to investigate the impact on the mean Zn-binding stability and its fluctuations (Figures 2 and 4). A near-monotonic decrease in mean Δ*E* is observed with increasing QM region, leveling off at a consistent value for the largest (i.e., 9-12 Å) QM regions (ca. 120 kcal/mol, Figure 4). The convergence of this property at ca. 300-400 atoms is consistent with prior radial convergence tests[54-55, 63, 92, 121] of energetic quantities. A decrease in Δ*E* with increasing QM treatment of the protein environment is intuitive, as the negative charge on the Cys residues can be distributed over a larger number of atoms. Increasing beyond the minimal QM region size

initially incorporates more positively charges residues (i.e., R12 in the 5-Å region), reducing the net charge and also adding adjacent, nonpolar residues (e.g., G14, V15, V18) that stabilize the Cys backbone. In cases where a negatively charged E17 is added (e.g., to the 6-Å region), this effect is compensated by addition of QM backbone atoms between Cys residues, leading to smaller changes in the mean ΔE value (Figure 4). The 7-Å region is an exception to the observation of a monotonic decrease in mean ΔE with increasing QM region, and we will shortly revisit the source of this outlier behavior. For the remaining larger QM regions, a balanced number of positively and negatively charged residues are added (e.g., 8-Å cutoff adds R11, Ql21 and E891), recovering a continued decrease in mean ΔE values consistent with trends from the smaller QM regions (Figure 4). Both the larger size (i.e., greater distance to Zn^{2+} of added residues) and good balance between oppositely charged residues added in the largest QM regions (i.e., 9, 10, and 12 Å) cause the mean ΔE values to approach a constant limit (Figure 4).

In contrast with overall observations, the mean ΔE value for the 7-Å cutoff is an outlier, increasing to a value similar to that for the minimal QM region (Figure 4). This runs counter to our initial arguments about charge and backbone stabilization because this region adds only 48 QM atoms from two positively charged arginine residues (R49 and R50) to the 6-Å cutoff (Figure 4). To quantify the role of these positively charged residues in the increased mean ΔE value, we re-computed ΔE for QM regions intermediate between the 6-Å and 7-Å cutoffs obtained by either moving R49 or R50 from the QM to MM regions (Supporting Information Figure S3). Removing either residue decreases the ΔE value, with R50 having the larger effect of the two residues, and the resulting QM region (6-Å+R49) has a lower ΔE value than the largest QM regions (Supporting Information Figure S3). Of the two arginine residues, the guanidinium nitrogen atoms of R49 are closer (6-8 Å) than those of R50 (14-15 Å), while the backbone of

latter is more proximal to the metal (Supporting Information Figure S4). The larger effect of R50 is likely attributable to the fact that it forms an amide bond with C51, but it is not immediately clear why additional QM-treated protein environment adjacent to C51 makes the rigid removal of Zn less favorable. This observation highlights the limits of heuristic rules for constructing QM regions.

We compare fluctuations within a QM region to the dependence of mean ΔE values on QM region size by recalling that mean ΔE values decrease from 336 kcal/mol in the 3-Å cutoff QM region to around 116-126 kcal/mol for the largest 9-12-Å cutoffs. These differences exceed the std. dev. of the ΔE values within a fixed QM region, which are around 85 kcal/mol for all QM regions. However, the change in mean ΔE values is less than the 800 kcal/mol range of ΔE values sampled for any fixed QM region. Thus, the absolute ΔE value prediction is sensitive to QM region size, whereas the fluctuations across configurations, as quantified through both the range and std. dev. appear to be comparable for all QM regions.

Although mean ΔE values are strongly QM-region dependent, fluctuations of ΔE values over configurations using the minimal QM region appear to be qualitatively preserved in even the largest QM regions (Figure 4 and Supporting Information Figure S5). To quantify these relationships, we calculated the Spearman's rank correlation coefficient, which is a nonparametric rank statistic that measures the strength of monotonic relationship between two variables and is less sensitive to outliers[122] than Pearson's correlation coefficient. We computed the Spearman's rank correlation coefficient between two variables, r_s, as:

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \tag{2}$$

where d_i denotes the difference between ranks of each observation and here n = 250 for the

number of snapshots compared. We observe very high r_s values (> 0.99) between all QM regions, with the highest correlations amongst the largest QM regions, a trend that is also preserved in the standard Pearson linear correlation coefficients (Supporting Information Table S7). Thus, absolute energetic properties, such as the stability of Zn in the metal binding site, can be highly sensitive to QM region size, whereas fluctuations in properties due to configurational sampling can be captured with computationally efficient, small QM regions.

4b. Electronic Descriptors of Zn^{2+} Binding

Since some differences are evident between small- and large-QM region descriptions of Zn^{2+} binding stability, we next evaluated changes in descriptions of the electronic environment of the metal. Specifically, we computed the Zn^{2+} Mulliken charge, q(Zn), and the Zn^{2+} Mayer total bond valence (MTBV) to quantify changes both with increasing QM region size and over configurations from MD (see Sec. 3). The Zn-S bond can be expected to have partial covalent character, which is confirmed through a q(Zn) value far below the +2 formal charge, even accounting for limitations of Mulliken charges[123] (Figure 5). Although this observation holds regardless of QM region, increasing the QM radial cutoff increases the average q(Zn) from the 3-Å QM region to the 6-Å region (3 Å: < 0.55 e vs. 6 Å: 0.58 e) before it decreases again slightly and levels off for the largest 8-12 Å QM regions (Figure 5). Unlike for the rigid binding energetics, no distinct behavior is observed for the 7-Å radial cutoff. The region sizes for which asymptotic limits are reached (ca. 8-12 Å) are largely consistent with what we observed for ΔE value region dependence. Significant fluctuations (range: 0.3 e, std. dev. 0.04-0.05 e) for q(Zn) are also apparent across snapshots (Figure 5 and Supporting Information Figures S6-S7). Like ΔE values, the smallest QM region largely captures essential fluctuations in q(Zn) (3-Å cutoff std. dev.: 0.044 e vs. 5-Å cutoff std. dev.: 0.046 e), but the range is smallest for this QM region

(3-Å cutoff range: 0.026 e) and increases significantly (5-Å cutoff range: 0.031 e) before decreasing again and leveling off at the largest QM region sizes (8-12-Å cutoff ranges: ca. 0.028 e, Supporting Information Figure S7). Thus, distribution characteristics change more significantly with QM region size than they do for ΔE, with agreement only being reached at around 8-Å QM radial cutoffs (see Supporting Information).

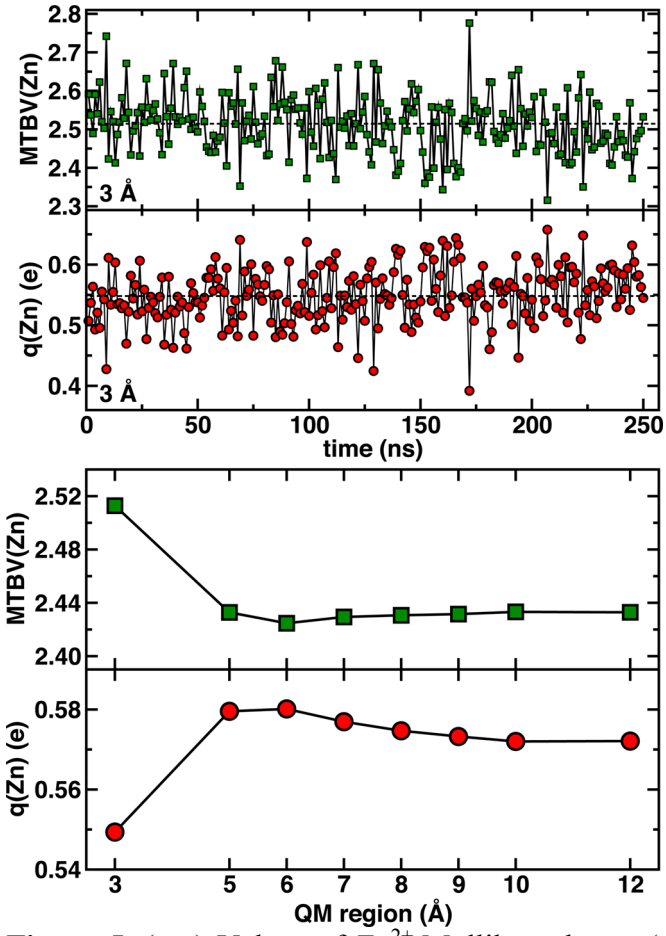

Figure 5. (top) Values of Zn^{2+} Mulliken charge (q(Zn) in e, red circles) and Mayer total bond valence (MTBV, green squares) for the 3-Å QM region over 250 1-ns-spaced MD snapshots. (bottom) Mean Zn^{2+} Mulliken charge (q(Zn) in e, red circles) and MTBV (green squares) for eight radial cutoff (i.e., 3, 5, 6, 7, 8, 9, 10, and 12 Å) QM regions.

To complement interpretation of the Zn Mulliken charges as indicative of covalent character in the Zn-S bond, we computed the Mayer total bond valence (MTBV) for Zn^{2+}, which is the effective total covalent bond order between Zn and the four Cys residues (Figure 5). Consistent with q(Zn), the mean Zn MTBV is highest in the 3-Å region (ca. 2.52), quickly

reaches asymptotic values of around 2.42 for the 7-Å and larger QM regions, and is overall indicative of partially covalent Zn-S bonding for all QM regions (Figure 5). Although the mean varies smoothly, the MTBV distribution is both wide (2.2-2.8) and changes with increasing QM region size (Figure 5 and Supporting Information Figures S8-S9). Although the std. dev. of all distributions are generally comparable (ca. 0.08), the 3-Å cutoff has the narrowest MTBV distribution (range: 0.46), which then broadens in the 5-Å cutoff (range: 0.52) and only approaches asymptotic values at 8-Å and larger radial cutoffs (range: 0.48), consistent with other properties (Figure 5 and Supporting Information Figures S8-S9).

Overall, in the minimal 3-Å QM region, the Zn-S bonds are more covalent, as evidenced through both q(Zn) and the MTBV, but even increasing the QM region slightly to incorporate more amide backbone around the Cys residues achieves relatively good agreement with the largest QM regions (Figure 5). One potential source of the fluctuations in both quantities is geometric changes in the Zn-S bond lengths over the snapshots sampled during MD. However, as indicated from limited rank correlation ($r_s < 0.1$) of these two properties to the Zn-S bond lengths, the MTBV and q(Zn) variations appear to be driven by multiple factors (Supporting Information Table S8).

Given the significant fluctuation in energetic (i.e., ΔE values) and electronic (i.e., q(Zn) and MTBV) properties, we might expect rank correlations to exist between these properties. The 3-Å QM region snapshot (172 ns) with the most neutral Zn^{2+} partial charge (0.39 e) also has the highest Zn^{2+} MTBV (2.78), providing qualitative support for such expectations (Figure 5). To obtain quantitative comparisons, we computed the r_s between pairs of the three properties for all QM regions (Supporting Information Table S9). As expected, a strong negative correlation ($r_s < -0.9$) is observed for all QM regions between q(Zn) and MTBV (Supporting Information Table

S9). However, no correlation is observed between ΔE values and either electronic property, meaning that changes in covalency across configurations do not explain Zn^{2+} stability (Supporting Information Table S9). Nevertheless, QM regions that have a consistent description of the Zn-S bond also are in broad agreement about Zn removal energies.

4c. Charge Transfer Surrounding the Zn^{2+} Binding Site

Charge transfer between the Zn^{2+} ion and surrounding residues can be expected to play an important role in describing its binding to the DNMT1 protein. We have noted substantial differences in Zn^{2+} rigid binding energies and Zn-S bonds for the minimal 3-Å QM region in comparison to larger radial cutoffs, likely due to the significant net charge on the four Cys after Zn^{2+} removal in this minimal QM region. To probe differences in charge distributions with variation of the QM region and configuration, we evaluated sums of Mulliken charges over the sidechains of the Zn^{2+}-coordinating residues (i.e., C13, C16, C19 and C51) as well as three charged arginine residues (i.e., R12, R49 and R50) that are in close proximity to the Zn^{2+} site. We excluded main chain (i.e., Cα and backbone) atoms from the sums to isolate sidechain-specific charge transfer trends (Supporting Information Figure S10).

All cysteine sidechains carry less than the expected formal -1 charge on average (ca. -0.65 to -0.55 e), and the most significant change in mean sidechain charges is observed over the three smallest QM regions (Figure 6 and Supporting Information Figure S11). Average cysteine sidechain charges approach asymptotic values at around the 8-Å QM region, but trends differ for each residue as does the sensitivity of the distribution to QM region size (Supporting Information Figure S11). For C13, we observe a gradual monotonic decrease in the magnitude of negative charge, whereas for C19 only the 3-Å region sidechain partial charge is significantly more negative than for the other seven regions (Figure 6). These differences cannot be attributed to the

addition of adjacent residues because R12 and G14 are added in the 5-Å region to extend the backbone of C13, whereas C19-adjacent Q20 is not added until the 6-Å QM region (Figure 2). The C13 and C51 residues gradually approach the most and least negative partial charge sum limits (q(C51)-q(C13) = 0.05 e), respectively (Figure 6). However, the immediately adjacent residues in both cases are relatively similar in their identity: R12 and R50 are added on the N-terminal side to both residues at intermediate distances (i.e., 5 Å and 7 Å), and the C-terminal side consists of non-polar G14 and P52, respectively, which are incorporated in the 5-Å region (Figure 2). Thus, sidechain partial charges appear sensitive to through-space interactions.

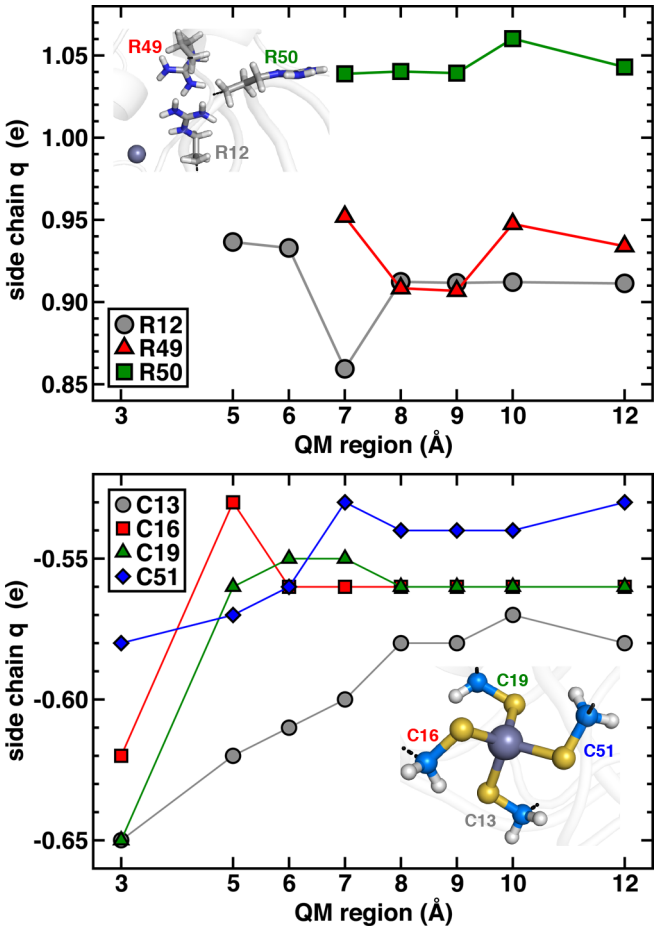

Figure 6. (top) The mean sidechain Mulliken charge sums in QM/MM for R12 (grey), R49 (red), and R50 (green) from 250 MD snapshots for all QM regions in which the residue is treated at the QM level of theory. An inset structure shows Zn^{2+} as a sphere with placement of the relevant sidechains in sticks. (bottom) The mean for sidechain Mulliken charge sums in QM/MM for Zn^{2+}-coordinating C13 (grey), C16 (red), C19 (green), and C51 (blue) from 250 MD snapshots

for all 8 QM regions. An inset structure shows Zn^{2+} as a sphere with the relevant sidechain atoms in balls and sticks.

Indeed, the C51 sidechain sum is the only one of the four cysteine residues that has an abrupt change in its average for the 7-Å QM region, likely due to its proximity to the added R49 and R50 residues (Figure 6 and Supporting Information Figures S11-S13). Focusing on the relevant arginine residues (i.e., R12, R49, and R50), we observe significant QM region dependence that is residue specific, although all three residue sidechains have charges (ca. 0.9-1.05) similar to their expected formal +1 charge (Figure 6 and Supporting Information Figure S14). The R49 residue is more positively charged than R50, and addition of E48 in the 10-Å QM region likely explains the significant increase in positive net charge on R49 over the 9-Å region (Figures 2 and 6). The significant sensitivity of R49 and R50 charge distributions to QM region could partially explain their role in shifting Zn^{2+} stability energetics and further motivates the need to balance residue addition in constructing QM regions (Supporting Information Figure S14). In comparison to the other two arginine residues, the C13-adjacent R12 is introduced earlier (at the 5-Å QM region) and exhibits monotonic decrease in net charge excluding only a discontinuity at the 7-Å QM region similar to that observed for C51 (Supporting Information Figure S14). Overall, slow, residue-specific evolution of sidechain sums for both cysteine and arginine point to distinct degrees of charge transfer as more of the protein is treated quantum mechanically, explaining why more metal-local properties are relatively slow to converge with QM region size.

4d. Contributions to the Electrostatic Potential

Extending our study of contributions to both QM region and configuration sensitivity of Zn^{2+} binding energetics and electronic properties, we evaluated the electrostatic potential (ESP) at the Zn2 site. We anticipate that the ESP could be sensitive to QM region choice due to

differences in how residues are treated at the MM vs. QM level of theory from both changes in the charge distribution on the residue and due to fewer constraints on the formal charge when the residue is modeled with QM. Since much of the character of configurational fluctuations is captured in the smallest QM region, we expect variations in binding strength over sampled configurations to be reflected in fluctuations of non-local structural contributions to the ESP that can primarily be captured from MM point charges.

We compute the ESP, V, at the Zn2 site from structures evaluated without Zn^{2+} present as an estimate of how the enzyme environment stabilizes Zn^{2+} binding (Figure 7, ESP with Zn^{2+} present, V', shown in Supporting Information Figure S15). While V is always negative, as expected to stabilize Zn^{2+} binding, the magnitude of the configurationally-averaged electrostatic potential decreases with increasing QM region size (Figure 7). This trend is at first surprising since we observed that ΔE values generally decrease with increasing QM region size, but we also noted a significant covalent contribution to the Zn-S bond (Figure 4). Thus, the ESP component is a significant but not sole contributor to Zn^{2+} binding.

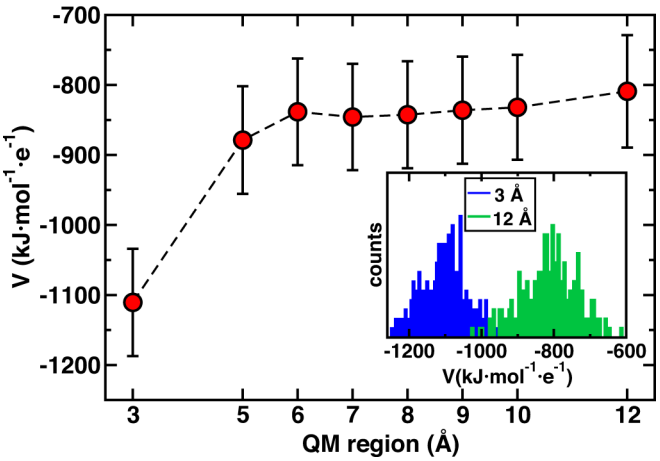

Figure 7. The average (red circle) and standard deviation (black bar) of V (in kJ/mol·e) at the Zn2 site determined from QM/MM on 250 MD snapshots evaluated without Zn^{2+} for all eight QM regions, as indicated on the x-axis. The inset shows the histogram of evaluated V values obtained for the minimal, 3-Å QM region (blue) and largest, 12-Å QM region (green).

Overall, the V for the smallest 3-Å QM region is most distinct, with almost no overlap

and a different distribution shape than the largest QM region (Figure 7). Both the mean and std. dev. across all snapshots are consistent across most larger QM regions from 6 to 12 Å in size (Figure 7). Decomposing the contributions to V from positively or negatively charged and neutral residues indicates that individual components to the ESP converge even more slowly than the total V value, suggesting distinct contributions of MM vs. QM representations to even long range electrostatics (Supporting Information Figure S16). The trend in V can be interpreted as larger QM regions distributing the negative charge more evenly over a larger QM region, leading to lower magnitudes of V at the Zn^{2+} site, which can also be observed when evaluated over structures with Zn^{2+} present (Figure 7 and Supporting Information Figure S15). Since the ESP is evaluated at the Zn2 site, we expect changes in its magnitude to have a strong contribution from the Zn-coordinating Cys residues. We computed the contribution of these residues (i.e., C13, C16, C19, and C51) to the ESP at the Zn^{2+} site for all QM regions and indeed observe a decrease in magnitude commensurate with the overall ESP decrease (Supporting Information Figure S17). The relative ESP evaluated at each Cys S site shifts from the smallest, 3-Å cutoff to moderate (e.g., 8-Å cutoff) QM region sizes (Supporting Information Figure S18). These observations support our expectation that QM interactions (e.g., charge transfer) between these Cys residues and the surrounding protein environment possible only in larger QM regions stabilizes the thiolates upon Zn^{2+} removal, decreasing both mean V and ΔE magnitudes (Figures 4 and 7).

Although both energetic and electronic properties vary across snapshots sampled from MD, any correlations between energetic and QM electronic properties examined to this point have been weak. To determine if ΔE fluctuations are primarily derived from the MM environment, we computed the MM component of QM/MM electrostatic potential for snapshots without Zn^{2+} present, V_{MM}, and compared its variation to those for the ΔE values. A weak

correlation between V_{MM} and ΔE is indeed apparent for the smallest 3-Å QM region (i.e., where the MM charges are most proximal to the Zn2 site) and remains in the largest 12-Å QM region (Supporting Information Figure S19). This observation suggests that variations in MM-atom configurations (i.e., surrounding protein environment waters, counterions, and DNA) play a significant role in property fluctuations.

4e. Correlations Among Properties

We have examined a number of metal-local and holistic electronic and energetic properties in the Zn2 binding site of DNMT1. Observations thus far point to moderate sensitivity to the QM region in QM/MM simulations, potentially due to our focus on a structural metal site rather than a catalytic active site. We observe evidence of incomplete descriptions of electrostatics and charge transfer in small DNMT1 QM regions, with property convergence motivating larger (ca. 300 atom) QM regions. At the same time, variations of each property with configuration in DNMT1 are often larger than the variations of mean values across QM region sizes (e.g., ΔE range in 3-Å cutoff: 800 kcal/mol vs. mean change from 3-Å to 12-Å ca. 220 kcal/mol), highlighting the importance of sampling in QM/MM modeling. To unify these observations, we computed linear correlations among all energetic, structural, and electronic properties computed thus far with changes in configuration (see Supporting Information). Specifically, we evaluated correlations among the ΔE, electrostatic potential without (with) Zn^{2+} present, V (V'), Mulliken charge sums of the Zn^{2+}-coordinating Cys residue sidechains, q(C13), q(C16), q(C19), and q(C51) or the nearby Arg residue sidechains, q(R12), q(R49), and q(50), the Zn^{2+} Mulliken charge, q(Zn), and Mayer total bond valence, MTBV, and the Zn-S bond distances to each of the four Zn^{2+}-coordinating Cys residues, d(C13), d(C16), d(C19), and d(C51).

To focus first on configurational effects, we computed these correlations for the 250 snapshots obtained on the 291-atom, 8-Å QM region, which we selected as the smallest representative QM region that was asymptotically converged with respect to the larger QM regions and contains key residues (e.g., R12, R49, and R50) in the QM region (Figure 2). The average values of some electronic properties correlate well (e.g., C13 and C19 sidechain charges) across the QM regions, whereas others are less correlated (e.g., C13 and R49 sidechain charges, Supporting Information Figure S20). Thus, analysis of 8-Å QM region property correlations over snapshots is likely general to other large QM regions, but observations from small QM regions could be expected to differ. For the 8-Å QM region, some of the strongest property correlations are expected, e.g. between the closely related Zn^{2+} Mulliken charge and MTBV electronic properties (Figures 5 and 8). Focusing on what governs ΔE values, the best correlations are with V' and V, but specific charges (i.e., q(R50)) and bond distances (i.e., d(C16)) also correlate somewhat (Figure 8). The MTBV and q(Zn) correlate less significantly with ΔE values than individual (e.g., Cys or Arg) sidechain charges (Figure 8). The electrostatic potential with Zn^{2+} present, V', is better correlated with both q(Zn) and the MTBV than the electrostatic potential without, V (Figure 8). However, ΔE is a quantity that measures the differences between the Zn^{2+}-bound and -absent structures, potentially explaining why ΔE does not correlate with Zn-local electronic properties alone.

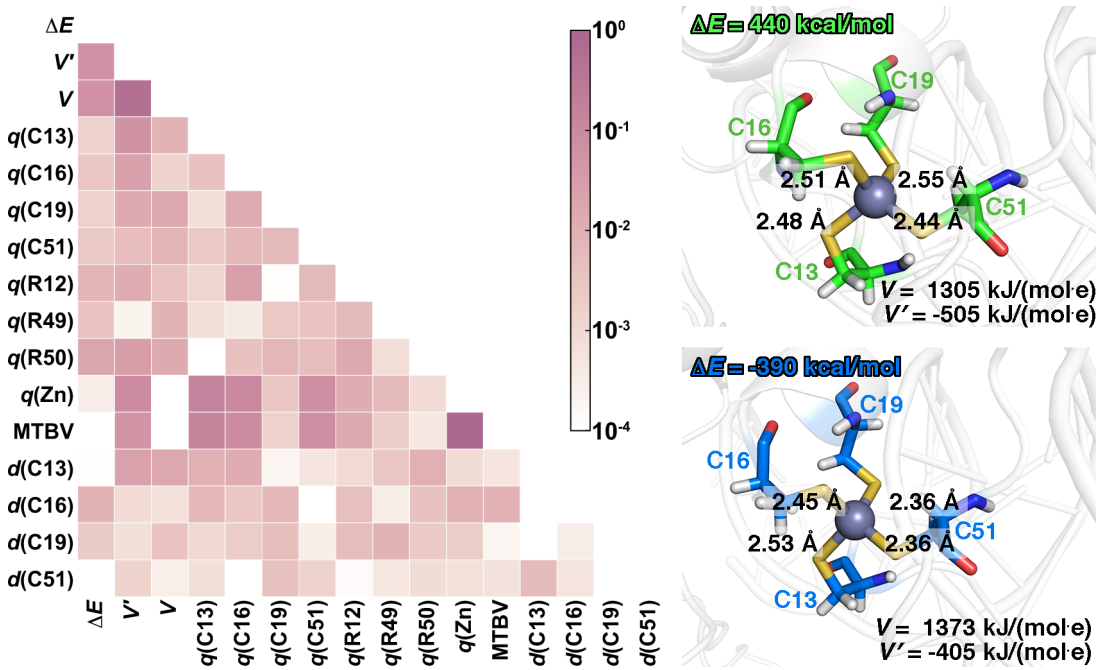

Figure 8. (left) Matrix of the log norm of squared Pearson correlation coefficients (colored as in inset colorbar) for several properties evaluated over 250 snapshots with Zn^{2+} present unless otherwise indicated. These include: ΔE; electrostatic potential with (without) Zn^{2+} present, V' (V); Mulliken charge sums of residue sidechains, q(C13), q(C16), q(C19), q(C51), q(R12), q(R49), and q(R50); q(Zn); MTBV; and the Zn-S bond distances, d(C13), d(C16), d(C19), and d(C51). (right) Representative snapshots at 165 ns (top, green) and 192 ns (bottom, blue) are shown with the 8-Å QM region ΔE value labeled on each structure. The structures are shown with key Zn-S distances and electrostatic potential values indicated in inset.

From over a dozen properties considered, no single property exhibits a singular correlation with ΔE (Figure 8). We also compared properties for snapshots that have the maximum and minimum ΔE values (165 and 192 ns, respectively, see Figure 8). Two of the four Zn-S bonds are significantly shorter in the low-ΔE configuration, and the average of the four Zn–S bond lengths decreases by around 0.07 Å (Figure 8 and Supporting Information Figure S21). Dynamic motion of the protein may be expected to destabilize Zn^{2+} bonding through compression of the Zn-S bonds, but weak ΔE correlation with individual bonds suggests this is not the primary factor in determining stability (Figure 8). Thus, it appears that a number of changes in electronic and geometric structure are responsible for the fluctuations observed across configurations. While most of such fluctuations can be captured in the smallest QM region,

descriptions of electronic and energetic properties are distinct in this smallest QM region, highlighting the importance of both suitable sampling and sufficiently large QM region size in QM/MM modeling. Thus, a strategy in which preliminary sampling using a small QM regions is coupled to large-QM region reweighting for quantifying properties, as has sometimes been advocated[124], is likely the most computationally efficient approach for modeling metal sites similar to the one studied here. To ensure that both effects are adequately accounted for in mechanistic study, automated workflows that include both sampling and systematic QM region selection[70] will be essential.

5. Conclusions

We have studied the relative magnitude of configurational and QM region sensitivity of energetic and electronic properties in a representative metal binding site of the enzyme DNMT1. From 250 ns of molecular dynamics simulation, we extracted evenly spaced snapshots and observed a significant variation in energetic and electronic properties with configuration. To simultaneously study QM region sensitivity, we expanded from a minimal, 49-atom QM region consisting only of the Zn metal and its four coordinating Cys residues and radially increased the QM region distance cutoff to include residues at least 12 Å away from the metal center in the largest, 628-atom QM region. Over the configurations sampled during MD, we observed that fluctuations in illustrative properties (e.g., rigid Zn^{2+} removal energy) were of similar magnitude in comparison to the variation of the mean value obtained with increasing QM region size. Most variations, as qualitatively observed through good rank correlations and quantified with property distribution characteristics, could be captured with even the smallest QM region.

In studying QM region dependence, we observed the mean values of energetic and electronic properties to converge only with larger (ca. 300-atom) QM region sizes. Generally,

balanced addition of positively and negatively charged residues enabled smooth, monotonic approach to asymptotic limits of key properties, and exceptions were noted when this charge balance was disrupted (e.g., Arg addition in the 7-Å region). For the smaller QM regions, the electronic description of Zn^{2+} binding is incomplete: the metal binds too tightly, is too stabilized by the strong electrostatic potential of MM point charges, and the Zn-S bond covalency is overestimated. We deduced that these incomplete descriptions are a consequence of both metal-local effects such as completing the amide bond between Cys residues and the rest of the protein backbone but also involved more non-local charge transfer effects for residues with sidechains positioned 6 Å or more away from the metal center. Correlation analysis of ΔE to other electronic and geometric properties reinforced this observation quantitatively, as no single descriptor uniquely explained variations in metal stability.

Overall, this case study on Zn^{2+} binding in DNMT1 suggests that efficient sampling with QM/MM in small QM regions is an effective method to explore the influence of enzyme structure on target properties. At the same time, descriptions of electronic and energetic properties require a larger QM region than the minimal metal-coordinating residues in order to converge treatment of both metal-local bonding and the overall electrostatic environment. We expect these conclusions to hold for similar metal-binding sites, be they catalytic or structural in nature, and new methods and software will be necessary to enable the systematic and automated evaluation of the magnitude and interplay of sampling and QM region dependence in property prediction.

ASSOCIATED CONTENT

Supporting Information.

Details of residues and numbers of atoms in QM region; list of residues added with Modeller to crystal structure; charge assignment to DNMT1 holoenzyme; ZAFF charges for DNMT1 holoenzyme; exchange-correlation functional and basis set tests; fluctuations in single point energies and ΔE values for various QM regions; effect of R49 and R50 removal on ΔE values; distance of R49 and R50 sidechain nitrogen atoms to Zn^{2+}; Spearman rank correlation of DE values between QM regions; fluctuations of Zn^{2+} Mulliken charges with QM region; fluctuations of Zn Mayer total bond valence with QM regions; Spearman's rank correlation between properties including Zn-S distances, q(Zn), ΔE, and MTBV; fluctuation in sidechain charges of relevant Cys and Arg residues; electrostatic potential distributions with QM region size evaluated from full QM/MM simulation, only from core residues, as well as the effect of MM-only point charges on smaller or larger QM regions; extended correlation matrix of properties between QM regions; distances of all Zn-S bonds over the 250 snapshots (PDF)

Detailed listing of residue protonation states in DNMT1, listing of MTBV and q(Zn) for all QM regions at each snapshot; quantitative values of correlations between properties either evaluate for fixed 8-Å QM region or between QM regions; properties of distributions with QM region size; and total energies (XLS)

Starting topology and coordinate files for AMBER Code of DNMT1; six representative snapshots of extracted spheres for QM/MM simulation of DNMT1; all 250 snapshots used from MD for the study (ZIP)

This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

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

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

Support for this research was provided by a core center grant P30-ES002109 from the National Institute of Environmental Health Sciences, National Institutes of Health. H.J.K. holds a Career Award at the Scientific Interface from the Burroughs Wellcome Fund and an AAAS Marion Milligan Mason Award, which also supported this work. The authors acknowledge Adam H. Steeves for providing a critical reading of the manuscript.

References

1. 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.
2. Bakowies, D.; Thiel, W., Hybrid Models for Combined Quantum Mechanical and Molecular Mechanical Approaches. *J. Phys. Chem.* **1996,** *100*, 10580-10594.
3. Mordasini, T. Z.; Thiel, W., Combined Quantum Mechanical and Molecular Mechanical Approaches. *Chimia* **1998,** *52*, 288-291.
4. Monard, G.; Merz, K. M., Combined Quantum Mechanical/Molecular Mechanical Methodologies Applied to Biomolecular Systems. *Acc. Chem. Res.* **1999,** *32*, 904-911.
5. Gao, J.; Truhlar, D. G., Quantum Mechanical Methods for Enzyme Kinetics. *Annu. Rev. Phys. Chem.* **2002,** *53*, 467-505.
6. Rosta, E.; Klahn, M.; Warshel, A., Towards Accurate Ab Initio QM/MM Calculations of Free-Energy Profiles of Enzymatic Reactions. *J. Phys. Chem. B* **2006,** *110*, 2934-2941.
7. Lin, H.; Truhlar, D., QM/MM: What Have We Learned, Where Are We, and Where Do We Go from Here? *Theor Chem Acc* **2007,** *117*, 185-199.
8. 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.
9. Senn, H. M.; Thiel, W., QM/MM Methods for Biomolecular Systems. *Angew. Chem., Int. Ed.* **2009,** *48*, 1198-1229.
10. Acevedo, O.; Jorgensen, W. L., Advances in Quantum and Molecular Mechanical (QM/MM) Simulations for Organic and Enzymatic Reactions. *Acc. Chem. Res.* **2009,** *43*, 142-151.
11. Gao, J.; Ma, S.; Major, D. T.; Nam, K.; Pu, J.; Truhlar, D. G., Mechanisms and Free Energies of Enzymatic Reactions. *Chem. Rev.* **2006,** *106*, 3188-3209.

12. Vidossich, P.; Florin, G.; Alfonso-Prieto, M.; Derat, E.; Shaik, S.; Rovira, C., On the Role of Water in Peroxidase Catalysis: A Theoretical Investigation of Hrp Compound I Formation. *J. Phys. Chem. B* **2010,** *114*, 5161-5169.
13. Carloni, P.; Rothlisberger, U.; Parrinello, M., The Role and Perspective of Ab Initio Molecular Dynamics in the Study of Biological Systems. *Acc. Chem. Res.* **2002,** *35*, 455-464.
14. Eurenius, K. P.; Chatfield, D. C.; Brooks, B. R.; Hodoscek, M., Enzyme Mechanisms with Hybrid Quantum and Molecular Mechanical Potentials. I. Theoretical Considerations. *Int. J. Quantum Chem.* **1996,** *60*, 1189-1200.
15. Senn, H. M.; Thiel, W., QM/MM Studies of Enzymes. *Curr. Opin. Chem. Biol.* **2007,** *11*, 182-187.
16. Monari, A.; Rivail, J.-L.; Assfeld, X., Advances in the Local Self-Consistent Field Method for Mixed Quantum Mechanics/Molecular Mechanics Calculations. *Acc. Chem. Res.* **2012,** *46*, 596-603.
17. Wang, Y.; Gao, J., Projected Hybrid Orbitals: A General QM/MM Method. *J. Phys. Chem. B* **2015,** *119*, 1213-1224.
18. Slavicek, P.; Martinez, T. J., Multicentered Valence Electron Effective Potentials: A Solution to the Link Atom Problem for Ground and Excited Electronic States. *J. Chem. Phys.* **2006,** *124*, 084107.
19. Murphy, R. B.; Philipp, D. M.; Friesner, R. A., A Mixed Quantum Mechanics/Molecular Mechanics (QM/MM) Method for Large Scale Modeling of Chemistry in Protein Environments. *J. Comput. Chem.* **2000,** *21*, 1442-1457.
20. Zhang, Y.; Lee, T.-S.; Yang, W., A Pseudobond Approach to Combining Quantum Mechanical and Molecular Mechanical Methods. *J. Chem. Phys.* **1999,** *110*, 46-54.
21. DiLabio, G. A.; Hurley, M. M.; Christiansen, P. A., Simple One-Electron Quantum Capping Potentials for Use in Hybrid QM/MM Studies of Biological Molecules. *J. Chem. Phys.* **2002,** *116*, 9578-9584.
22. von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D., Variational Optimization of Effective Atom-Centered Potentials for Molecular Properties. *J. Chem. Phys.* **2005,** *122*, 14113.
23. Wang, B.; Truhlar, D. G., Combined Quantum Mechanical and Molecular Mechanical Methods for Calculating Potential Energy Surfaces: Tuned and Balanced Redistributed Charge Algorithm. *J. Chem. Theory Comput.* **2010,** *6*, 359-369.
24. Thellamurege, N. M.; Hirao, H., Effect of Protein Environment within Cytochrome P450cam Evaluated Using a Polarizable-Embedding QM/MM Method. *J. Phys. Chem. B* **2014,** *118*, 2084-2092.
25. Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio Jr, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T., Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B* **2010,** *114*, 2549-2564.
26. Halgren, T. A.; Damm, W., Polarizable Force Fields. *Curr. Opin. Struct. Biol.* **2001,** *11*, 236-242.
27. Nåbo, L. J.; Olsen, J. M. H.; Martínez, T. J.; Kongsted, J., The Quality of the Embedding Potential Is Decisive for Minimal Quantum Region Size in Embedding Calculations: The Case of the Green Fluorescent Protein. *J. Chem. Theory Comput.* **2017,** *13*, 6230-6236.

28. Ganguly, A.; Boulanger, E.; Thiel, W., Importance of MM Polarization in QM/MM Studies of Enzymatic Reactions: Assessment of the QM/MM Drude Oscillator Model. *J. Chem. Theory Comput.* **2017,** *13*, 2954-2961.
29. Loco, D.; Lagardère, L.; Caprasecca, S.; Lipparini, F.; Mennucci, B.; Piquemal, J.-P., Hybrid QM/MM Molecular Dynamics with AMOEBA Polarizable Embedding. *J. Chem. Theory Comput.* **2017,** *13*, 4025-4033.
30. 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.
31. Nadig, G.; Van Zant, L. C.; Dixon, S. L.; Merz, K. M., Charge-Transfer Interactions in Macromolecular Systems: A New View of the Protein/Water Interface. *J. Am. Chem. Soc.* **1998,** *120*, 5593-5594.
32. Himo, F., Recent Trends in Quantum Chemical Modeling of Enzymatic Reactions. *J. Am. Chem. Soc.* **2017,** *139*, 6780-6786.
33. 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.
34. Kulik, H. J.; Zhang, J.; Klinman, J. P.; Martinez, 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.
35. Kulik, H. J.; Luehr, N.; Ufimtsev, I. S.; Martinez, T. J., Ab Initio Quantum Chemistry for Protein Structure. *J. Phys. Chem. B* **2012,** *116*, 12501-12509.
36. Rauscher, S.; Gapsys, V.; Gajda, M. J.; Zweckstetter, M.; de Groot, B. L.; Grubmüller, H., Structural Ensembles of Intrinsically Disordered Proteins Depend Strongly on Force Field: A Comparison to Experiment. *J. Chem. Theory Comput.* **2015,** *11*, 5513-5524.
37. Mehmood, R.; Qi, H. W.; Steeves, A. H.; Kulik, H. J., The Protein's Role in Substrate Positioning and Reactivity for Biosynthetic Enzyme Complexes: The Case of SyrB2/SyrB1. *ACS Catal.* **2019,** *9*, 4930-4943.
38. Ufimtsev, I. S.; Martínez, T. J., Quantum Chemistry on Graphical Processing Units. 1. Strategies for Two-Electron Integral Evaluation. *J. Chem. Theory Comput.* **2008,** *4*, 222-231.
39. Ufimtsev, I. S.; Martinez, T. J., Quantum Chemistry on Graphical Processing Units. 2. Direct Self-Consistent-Field Implementation. *J. Chem. Theory Comput.* **2009,** *5*, 1004-1015.
40. Ufimtsev, I. S.; Martínez, T. J., 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.
41. Isborn, C. M.; Luehr, N.; Ufimtsev, I. S.; Martinez, T. J., Excited-State Electronic Structure with Configuration Interaction Singles and Tamm-Dancoff Time-Dependent Density Functional Theory on Graphical Processing Units. *J. Chem. Theory Comput.* **2011,** *7*, 1814-1823.
42. Ochsenfeld, C.; Kussmann, J.; Lambrecht, D. S., Linear-Scaling Methods in Quantum Chemistry. *Rev. Comput. Chem.* **2007,** *23*, 1.
43. Eichkorn, K.; Weigend, F.; Treutler, O.; Ahlrichs, R., Auxiliary Basis Sets for Main Row Atoms and Transition Metals and Their Use to Approximate Coulomb Potentials. *Theor Chem Acc* **1997,** *97*, 119-124.
44. Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R., Auxiliary Basis Sets to Approximate Coulomb Potentials. *Chem. Phys. Lett.* **1995,** *240*, 283-290.
45. Liu, F.; Luehr, N.; Kulik, H. J.; Martínez, T. J., Quantum Chemistry for Solvated Molecules on Graphical Processing Units Using Polarizable Continuum Models. *J. Chem. Theory Comput.* **2015,** *11*, 3131-3144.

46. Solt, I.; Kulhanek, P.; Simon, I.; Winfield, S.; Payne, M. C.; Csanyi, G.; Fuxreiter, M., Evaluating Boundary Dependent Errors in QM/MM Simulations. *J. Phys. Chem. B* **2009,** *113*, 5728-5735.
47. Sumowski, C. V.; Ochsenfeld, C., A Convergence Study of QM/MM Isomerization Energies with the Selected Size of the QM Region for Peptidic Systems. *J. Phys. Chem. A* **2009,** *113*, 11734-11741.
48. Fox, S. J.; Pittock, C.; Fox, T.; Tautermann, C. S.; Malcolm, N.; Skylaris, C. K., Electrostatic Embedding in Large-Scale First Principles Quantum Mechanical Calculations on Biomolecules. *J. Chem. Phys.* **2011,** *135*, 224107.
49. Hu, L.; Söderhjelm, P. r.; Ryde, U., On the Convergence of QM/MM Energies. *J. Chem. Theory Comput.* **2011,** *7*, 761-777.
50. 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.
51. Flaig, D.; Beer, M.; Ochsenfeld, C., Convergence of Electronic Structure with the Size of the QM Region: Example of QM/MM NMR Shieldings. *J. Chem. Theory Comput.* **2012,** *8*, 2260-2271.
52. Isborn, C. M.; Goetz, A. W.; Clark, M. A.; Walker, R. C.; Martinez, T. J., 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.* **2012,** *8*, 5092-5106.
53. Liao, R.-Z.; Thiel, W., Comparison of QM-only and QM/MM Models for the Mechanism of Tungsten-Dependent Acetylene Hydratase. *J. Chem. Theory Comput.* **2012,** *8*, 3793-3803.
54. Liao, R. Z.; Thiel, W., Convergence in the QM-Only and QM/MM Modeling of Enzymatic Reactions: A Case Study for Acetylene Hydratase. *J. Comput. Chem.* **2013,** *34*, 2389-2397.
55. Sadeghian, K.; Flaig, D.; Blank, I. D.; Schneider, S.; Strasser, R.; Stathis, D.; Winnacker, M.; Carell, T.; Ochsenfeld, C., Ribose-Protonated DNA Base Excision Repair: A Combined Theoretical and Experimental Study. *Angew. Chem., Int. Ed.* **2014,** *53*, 10044-10048.
56. Hartman, J. D.; Neubauer, T. J.; Caulkins, B. G.; Mueller, L. J.; Beran, G. J., Converging Nuclear Magnetic Shielding Calculations with Respect to Basis and System Size in Protein Systems. *J. Biomol. NMR* **2015,** *62*, 327-340.
57. Vanpoucke, D. E.; Oláh, J.; De Proft, F.; Van Speybroeck, V.; Roos, G., Convergence of Atomic Charges with the Size of the Enzymatic Environment. *J. Chem. Inf. Model.* **2015,** *55*, 564-571.
58. Harris, T. V.; Szilagyi, R. K., Protein Environmental Effects on Iron-Sulfur Clusters: A Set of Rules for Constructing Computational Models for Inner and Outer Coordination Spheres. *J. Comput. Chem.* **2016,** *37*, 1681-1696.
59. Benediktsson, B.; Bjornsson, R., QM/MM Study of the Nitrogenase Mofe Protein Resting State: Broken-Symmetry States, Protonation States, and QM Region Convergence in the Femoco Active Site. *Inorg. Chem.* **2017,** *56*, 13417-13429.
60. 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.
61. Pokorna, P.; Kruse, H.; Krepl, M.; Sponer, J., QM/MM Calculations on Protein–Rna Complexes: Understanding Limitations of Classical MD Simulations and Search for Reliable Cost-Effective QM Methods. *J. Chem. Theory Comput.* **2018,** *14*, 5419-5433.

62. Prejanò, M.; Marino, T.; Russo, N., QM Cluster or QM/MM in Computational Enzymology: The Test Case of Ligw-Decarboxylase. *Front. Chem.* **2018,** *6*, 249.
63. Roßbach, S.; Ochsenfeld, C., 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.* **2017,** *13*, 1102-1107.
64. Provorse, M. R.; Peev, T.; Xiong, C.; Isborn, C. M., Convergence of Excitation Energies in Mixed Quantum and Classical Solvent: Comparison of Continuum and Point Charge Models. *J. Phys. Chem. B* **2016,** *120*, 12148-12159.
65. Milanese, J. M.; Provorse, M. R.; Alameda, E.; Isborn, C. M., Convergence of Computed Aqueous Absorption Spectra with Explicit Quantum Mechanical Solvent. *J. Chem. Theory Comput.* **2017,** *13*, 2159-2171.
66. Morgenstern, A.; Jaszai, M.; Eberhart, M. E.; Alexandrova, A. N., Quantified Electrostatic Preorganization in Enzymes Using the Geometry of the Electron Charge Density. *Chem. Sci.* **2017,** *8*, 5010-5018.
67. Yang, Z.; Liu, F.; Steeves, A. H.; Kulik, H. J., Quantum Mechanical Description of Electrostatics Provides a Unified Picture of Catalytic Action across Methyltransferases. *J. Phys. Chem. Lett.* **2019,** *10*, 3779-3787.
68. Siegbahn, P. E., 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.* **2016,** *138*, 10485-10495.
69. Das, S.; Nam, K.; Major, D. T., 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.* **2018,** *14*, 1695-1705.
70. Karelina, M.; Kulik, H. J., Systematic Quantum Mechanical Region Determination in QM/MM Simulation. *J. Chem. Theory Comput.* **2017,** *13*, 563-576.
71. Meier, K.; Thiel, W.; van Gunsteren, W. F., On the Effect of a Variation of the Force Field, Spatial Boundary Condition and Size of the QM Region in QM/MM MD Simulations. *J. Comput. Chem.* **2012,** *33*, 363-378.
72. Patra, N.; Ioannidis, E. I.; Kulik, H. J., Computational Investigation of the Interplay of Substrate Positioning and Reactivity in Catechol O-Methyltransferase. *PLoS ONE* **2016,** *11*, e0161868.
73. Lameira, J.; Bora, R. P.; Chu, Z. T.; Warshel, A., Methyltransferases Do Not Work by Compression, Cratic, or Desolvation Effects, but by Electrostatic Preorganization. *Proteins: Struct., Funct., Bioinf.* **2015,** *83*, 318-330.
74. Jindal, G.; Warshel, A., Exploring the Dependence of QM/MM Calculations of Enzyme Catalysis on the Size of the QM Region. *J. Phys. Chem. B* **2016,** *120*, 9913-9921.
75. Olsson, M. A.; Ryde, U., Comparison of QM/MM Methods to Obtain Ligand-Binding Free Energies. *J. Chem. Theory Comput.* **2017,** *13*, 2245-2253.
76. Karlin, K. D., Metalloenzymes, Structural Motifs, and Inorganic Models. *Science* **1993,** *261*, 701-708.
77. Vallee, B. L.; Auld, D. S., Zinc Coordination, Function, and Structure of Zinc Enzymes and Other Proteins. *Biochemistry* **1990,** *29*, 5647-5659.
78. Andreini, C.; Bertini, I.; Cavallaro, G.; Holliday, G. L.; Thornton, J. M., Metal Ions in Biological Catalysis: From Enzyme Databases to General Principles. *JBIC, J. Biol. Inorg. Chem.* **2008,** *13*, 1205-1218.

79. Miller, O. J.; Schnedl, W.; Allen, J.; Erlanger, B. F., 5-Methylcytosine Localised in Mammalian Constitutive Heterochromatin. *Nature* **1974,** *251*, 636.
80. Bird, A., DNA Methylation Patterns and Epigenetic Memory. *Genes Dev.* **2002,** *16*, 6-21.
81. Song, J.; Rechkoblit, O.; Bestor, T. H.; Patel, D. J., Structure of Dnmt1-DNA Complex Reveals a Role for Autoinhibition in Maintenance DNA Methylation. *Science* **2011,** *331*, 1036-1040.
82. Reichard, J. F.; Puga, A., Effects of Arsenic Exposure on DNA Methylation and Epigenetic Gene Regulation. *Epigenomics* **2010,** *2*, 87-104.
83. Bhattacharyya, D.; Boulden, A.; Foote, R.; Mitra, S., Effect of Polyvalent Metal Ions on the Reactivity of Human O 6-Methylguanine-DNA Methyltransferase. *Carcinog. - Compr. Surv.* **1988,** *9*, 683-685.
84. Topol, I. A.; Casas-Finet, J. R.; Gussio, R.; Burt, S. K.; Erickson, J. W., A Quantum-Mechanical Study of Metal Binding Sites in Zinc Finger Structures. *J. Mol. Struct.: THEOCHEM* **1998,** *423*, 13-28.
85. Drici, N.; Krallafa, M. A., Effect of Mutation on the Stabilization Energy of Hiv-1 Zinc Fingers: A Hybrid Local Self-Consistent Field/Molecular Mechanics Investigation. *JBIC, J. Biol. Inorg. Chem.* **2017,** *22*, 109-119.
86. Yang, J.; Lior-Hoffmann, L.; Wang, S.; Zhang, Y.; Broyde, S., DNA Cytosine Methylation: Structural and Thermodynamic Characterization of the Epigenetic Marking Mechanism. *Biochemistry* **2013,** *52*, 2828-2838.
87. Caulfield, T.; Medina-Franco, J. L., Molecular Dynamics Simulations of Human DNA Methyltransferase 3b with Selective Inhibitor Nanaomycin A. *J. Struct. Biol.* **2011,** *176*, 185-191.
88. Liu, L.; Shi, T.; Houk, K. N.; Zhao, Y.-L., Understanding the R882h Mutation Effects of DNA Methyltransferase Dnmt3a: A Combination of Molecular Dynamics Simulations and QM/MM Calculations. *RSC Adv.* **2019,** *9*, 31425-31434.
89. Qi, H. W.; Karelina, M.; Kulik, H. J., Quantifying Electronic Effects in QM and QM/MM Biomolecular Modeling with the Fukui Function. *Acta Phys.-Chim. Sin.* **2018,** *34*, 81-91.
90. Sumner, S.; Söderhjelm, P.; Ryde, U., Effect of Geometry Optimizations on QM-Cluster and QM/MM Studies of Reaction Energies in Proteins. *J. Chem. Theory Comput.* **2013,** *9*, 4205-4214.
91. Waller, M. P.; Kumbhar, S.; Yang, J., A Density-Based Adaptive Quantum Mechanical/Molecular Mechanical Method. *ChemPhysChem* **2014,** *15*, 3218-3225.
92. 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.
93. Qi, H. W.; Kulik, H. J., Evaluating Unexpectedly Short Non-Covalent Distances in X-Ray Crystal Structures of Proteins with Electronic Structure Analysis. *J. Chem. Inf. Model.* **2019,** *59*, 2199-2211.
94. Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R., Avogadro: An Advanced Semantic Chemical Editor, Visualization, and Analysis Platform. *J. Cheminf.* **2012,** *4*, 17.
95. Eswar, N.; Webb, B.; Marti-Renom, M. A.; Madhusudhan, M.; Eramian, D.; Shen, M. y.; Pieper, U.; Sali, A., Comparative Protein Structure Modeling Using Modeller. *Curr. Protoc. Bioinf.* **2006,** *15*, 5.6. 1-5.6. 30.

96. Labahn, J.; Granzin, J.; Schluckebier, G.; Robinson, D. P.; Jack, W. E.; Schildkraut, I.; Saenger, W., Three-Dimensional Structure of the Adenine-Specific DNA Methyltransferase M.Taq I in Complex with the Cofactor S-Adenosylmethionine. *Proc. Natl. Acad. Sci. U. S. A.* **1994,** *91*, 10957-10961.
97. 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.
98. 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.
99. 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.
100. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015,** *11*, 3696-3713.
101. Peters, M. B.; Yang, Y.; Wang, B.; Füsti-Molnár, L. s.; Weaver, M. N.; Merz Jr, K. M., Structural Survey of Zinc-Containing Proteins and Development of the Zinc Amber Force Field (Zaff). *J. Chem. Theory Comput.* **2010,** *6*, 2935-2947.
102. Ivani, I.; Dans, P. D.; Noy, A.; Pérez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goñi, R.; Balaceanu, A.; Portella, G.; Battistini, F.; Gelpí, J. L.; González, C.; Vendruscolo, M.; Laughton, C. A.; Harris, S. A.; Case, D. A.; Orozco, M., Parmbsc1: A Refined Force Field for DNA Simulations. *Nat. Methods* **2016,** *13*, 55.
103. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004,** *25*, 1157-1174.
104. Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A., A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993,** *97*, 10269-10280.
105. Harihara, P. C.; Pople, J. A., Influence of Polarization Functions on Molecular-Orbital Hydrogenation Energies. *Theor Chim Acta* **1973,** *28*, 213-222.
106. Gordon, M. S.; Schmidt, M. W., Advances in Electronic Structure Theory: GAMESS a Decade Later. *Theory Appl. Comput. Chem.: First Forty Years* **2005**, 1167-1189.
107. F. Wang, J.-P. B., P. Cieplak and F.-Y. Dupradeau R.E.D. Python: Object Oriented Programming for Amber Force Fields, Université De Picardie - Jules Verne, Sanford|Burnham Medical Research Institute, Nov. 2013. http://q4md-forcefieldtools.org/REDServer-Development/ (accessed December 31, 2019).
108. Vanquelef, E.; Simon, S.; Marquant, G.; Garcia, E.; Klimerak, G.; Delepine, J. C.; Cieplak, P.; Dupradeau, F.-Y., R.E.D. Server: A Web Service for Deriving RESP and ESP Charges and Building Force Field Libraries for New Molecules and Molecular Fragments. *Nucleic Acids Res.* **2011,** *39*, W511-W517.
109. Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P., The R.E.D. Tools: Advances in RESP and ESP Charge Derivation and Force Field Library Building. *Phys. Chem. Chem. Phys.* **2010,** *12*, 7821-7839.
110. 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.

111. D.A. Case, J. T. B., R.M. Betz, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, T. Luchko, R. Luo, B. Madej, K.M. Merz, G. Monard, P. Needham, H. Nguyen, H.T. Nguyen, I. Omelyan, A. Onufriev, D.R. Roe, A. Roitberg, R. Salomon-Ferrer, C.L. Simmerling, W. Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, D.M. York and P.A. Kollman Amber 2015, University of California, San Francisco. http://www.ambermd.org (accessed December 31, 2019).
112. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C., 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.
113. Petachem. http://www.petachem.com. (accessed December 31, 2019).
114. Eastman, P.; Friedrichs, M. S.; Chodera, J. D.; Radmer, R. J.; Bruns, C. M.; Ku, J. P.; Beauchamp, K. A.; Lane, T. J.; Wang, L.-P.; Shukla, D.; Tye, T.; Houston, M.; Stich, T.; Klein, C.; Shirts, M. R.; Pande, V. S., Openmm 4: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation. *J. Chem. Theory Comput.* **2013,** *9*, 461-469.
115. 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.
116. Hay, P. J.; Wadt, W. R., Ab Initio Effective Core Potentials for Molecular Calculations. Potentials for the Transition Metal Atoms Sc to Hg. *J. Chem. Phys.* **1985,** *82*, 270-283.
117. Schrodinger, L. L. C. The PyMOL Molecular Graphics System, Version 1.7.4.3. http://www.pymol.org (accessed December 31, 2019).
118. Lu, T.; Chen, F., Multiwfn: A Multifunctional Wavefunction Analyzer. *J. Comput. Chem.* **2012,** *33*, 580-592.
119. Mayer, I., Charge, Bond Order and Valence in the Ab Initio SCF Theory. *Chem. Phys. Lett.* **1983,** *97*, 270-274.
120. Mayer, I., Bond Order and Valence: Relations to Mulliken's Population Analysis. *Int. J. Quantum Chem.* **1984,** *26*, 151-154.
121. Kulik, H. J.; Seelam, N.; Mar, B.; Martinez, T. J., Adapting DFT+U for the Chemically-Motivated Correction of Minimal Basis Set Incompleteness. *J. Phys. Chem. A* **2016,** *120*, 5939-5949.
122. Gauthier, T. D., Detecting Trends Using Spearman's Rank Correlation Coefficient. *Environ. Forensics* **2001,** *2*, 359-362.
123. Fonseca Guerra, C.; Handgraaf, J. W.; Baerends, E. J.; Bickelhaupt, F. M., Voronoi Deformation Density (VDD) Charges: Assessment of the Mulliken, Bader, Hirshfeld, Weinhold, and VDD Methods for Charge Analysis. *J. Comput. Chem.* **2004,** *25*, 189-210.
124. Hu, L.; Söderhjelm, P.; Ryde, U., Accurate Reaction Energies in Proteins Obtained by Combining QM/MM and Large QM Calculations. *J. Chem. Theory Comput.* **2013,** *9*, 640-649.

TOC Graphic