Combining Evolutionary Conservation and Quantum Topological Analyses to Determine QM Subsystems for Biomolecular QM/MM Simulations

Mark A. Hix, Emmett M. Leddin, G. Andrés Cisneros^{*†}

Department of Chemistry, University of North Texas, Denton, TX 76201

E-mail: andres@unt.edu

Abstract

Selection of residues and other molecular fragments for inclusion in the QM region for QM/MM simulations is an important step for these calculations. Here, we present an approach that combines protein sequence/structure evolution and electron localization function (ELF) analyses. The combination of these two analysis allows the determination of whether a residue needs to be included in the QM subsystem, or can be represented by the MM environment. We have applied this approach on two systems previously investigated by QM/MM simulations, 4-oxalocrotonate tautomerase (4OT), and ten-eleven translocation-2 (TET2), that provide examples where fragments may or may not need to be included in the QM subsystem. Subsequently, we present the use of this approach to determine the appropriate QM subsystem to calculate the minimum energy path (MEP) for the reaction catalyzed by human DNA polymerase λ (Pol λ) with a third cation in the active site. Our results suggest that the combination of protein evolutionary and ELF analyses provides insights on

residue/molecular fragment selection for QM/MM simulations.

Introduction

The atomistic investigation of enzymatic reaction mechanisms can provide useful insights for various applications.¹ A popular approach to investigate reaction mechanisms in enzymes is the hybrid Quantum Mechanics/Molecular Mechanics (QM/MM) method. This approach combines two levels of theory, QM and MM, to allow the investigation of large systems by using a QM approach to represent a small region, at a minimum the atoms/molecules directly involved in the reaction, while the rest of the system is treated with a molecular mechanics (MM) method.²⁻⁵

The choice of molecular fragments to be included in the QM subsystem and approaches used to determine them is a topic of interest in the literature.⁶⁻¹³ Several factors play a role in this selection, including (but not limited to) the overall computational time (wall-time) of the calculation, the QM level of theory (semi-empirical or *ab initio*)/basis set, the MM force field, the type of simulation to be performed (e.g. ground state reactivity, excited state reactivity, electronic excitation calculations, QM/MM MD), etc. The use of polarizable force fields has been shown to provide a better description of the MM environment, which can aid in the reduction of the number of atoms that need to be treated at the QM level.¹⁴⁻²⁰

One of the ultimate goals of selecting the fragments for the QM subsystem is to achieve a balance between simulation accuracy and computational efficiency.¹³ All QM/MM calculations necessarily include the Minimum Active Region (MAR) in the QM subsystem, which we define here as the atoms or molecules *directly* involved in the chemical bond breaking/forming processes for the reaction under investigation

Inter-molecular interactions between molecular fragments surrounding the MAR and the QM subsystem may or may not be approximated appropriately by the classical environment. In these cases, it is necessary to include additional fragments that may participate indirectly

in the bond making and/or breaking processes. To this end, several methods have been proposed to determine the molecular fragments that need to be included in the QM region.

The QM fragment selection can be performed based on geometric considerations with respect to the MAR, such as including all atoms within a sphere or ellipsoid centered around the MAR,^{7,10,11} or including fragments that comprise the 1st and 2nd shell of residues around the MAR.²¹⁻³¹ Heuristic approaches that consider specific interactions of molecular fragments with the MAR have also been proposed. One such method uses covalent interactions, where residues which are accessible via a "covalent walk" from the MAR and lie within a specific radius are included.³² Another approach relies on including fragments that form hydrogen bonds with molecules in the MAR.³³

Other approaches for the determination of molecular fragments to be included in the QM subsystem rely on the calculation of electronic structure descriptors such as the charge deletion analysis⁸ or the charge shift analysis.^{34,35} The latter approach relies on the use of localized Fukui functions to determine the impact of the inclusion of additional residues beyond the MAR in enzyme catalysis. In this approach, single protein residues outside the MAR are included in the QM subsystem and changes to the localized Fukui function are evaluated. Subsequently, the residues that result in changes greater than a given threshold are included in the QM region.^{34,35} A similar approach relies on the use of QTAIM analysis to predict changes to enzymatic reaction barriers based on changing electric fields.³⁶

In many cases, the inclusion of additional fragments in the QM subsystem results in increased accuracy in the description of the system, however, this inclusion is commensurate with increased computational cost.^{24,26,28,37,38}

Although there have been a variety of approaches developed for the selection of QM fragments, we are not aware of any procedure that considers protein evolutionary conservation for QM subsystem determination. Evolutionary conservation is a critical feature of enzyme families, and provides clues to catalytic activity and proficiency.^{39,40} Therefore, it seems natural that the selection of QM fragments in simulations of enzymatic catalysis would

benefit from the use of evolutionary conservation analysis. In this contribution we describe a method to select relevant fragments for the QM subsystem that relies on the use of protein sequence/structure evolutionary conservation to identify catalytically relevant residues; combined with a quantum interpretative technique to investigate the effects of specific fragments on the QM wavefunction.

Most proteins are classified into families based on structural and functional similarity, with variations on enzymatic activity, cellular localization, expression levels, and other characteristics.⁴¹ These protein families often have structurally-conserved amino acid sequences across all members, with these residues frequently shown to be relevant to protein structure and/or function.⁴²⁻⁴⁴ Pairs of unrelated proteins have also been shown to have conserved amino acids at critical positions in their sequences.⁴⁵ Some conserved residues may be of interest in investigations of enzymatic reactions as well, as previous studies have shown certain amino acids may be important for catalysis by interacting with reactant or transition states along the reaction path, without being directly involved in the breaking or forming of chemical bonds.^{22,46,47}

As indicated above, the use of chemical descriptors based on QM can be used to investigate the effects of the environment on the QM wavefunction. The electron localization function (ELF) provides a way to investigate how regions of strong electron pairing are influenced in biological systems.⁴⁸ The ELF was initially proposed to calculate the range of electron pairing.^{49,50} Since the ELF results in a continuous and differentiable scalar field, a topological analysis can be carried out (similar to QTAIM). The resulting scalar and derivative fields can be divided into regions (basins). These basins have chemically intuitive interpretations including cores, lone pairs, bonds, etc. ELF has been shown to provide useful information for organic and enzymatic reactions.^{51,52} Additionally, since the basins do not overlap and the electron population within each basin can be calculated, a multipolar expansion can be performed, to gain even further insights on these chemically relevant molecular regions.⁵³

Here we present the use of an approach that combines protein sequence/structure evolutionary conservation and ELF analyses to determine the effects of various conserved residues and other molecular fragments on three enzyme systems: 4-oxalocrotonate (4OT), ten-eleven translocation-2 (TET2), and human DNA polymerase λ (Pol λ). The remainder of the paper is as follows: the next section describes the computational methods for the evolutionary and ELF analysis, followed by a description of the computational approaches used to study the three enzymatic systems. Subsequently, the results of the conservation and ELF analysis on the two test systems, 4OT and TET2, are presented to show the applicability of ELF to the determination of appropriate fragments in the QM subsystem. Pol λ provides a test system to determine the applicability of the method, followed by concluding remarks.

Computational Methods

Peptide sequence and structure alignments were performed using the T-Coffee Server in both Espresso mode and M-Coffee mode, and with Uniprot with CLUSTALO.^{54–61} Conserved residues on the target systems from these multiple alignments were assessed by visual inspection to determine their proximity to the active site.

The three systems considered herein, 4OT, TET2, and Pol λ , have been studied by QM/MM previously.^{22–25,52,62,63} The details for the preparation of the simulation systems and computation of the individual paths of these three systems have been described previously. Briefly, all systems correspond to the wild type structures. In every case, the structures have been checked with PROPKA and MolProbity to assign ionizable residue protonation states and check side-chain rotamers. All systems are solvated in TIP3P water, neutralized with the corresponding number of required counterions and subjected to molecular dynamics. For 4OT, the structure is based on pdbid: 1BJP,⁶⁴ the QM subsystem includes the N-terminal Pro (P1), 2-oxo-4-hexenedioate and (in some cases) R39'' or R61' (see below).^{22,62}

For TET2, the reference system, based on pdbid 4NM6,⁶⁵ is the same as the one used in

the original calculations.²⁴ Here the QM subsystem includes the ferryl intermediate and all 1st-shell ligands including coordinating residues (H1382, D1384, H1881), succinate, water, 5-hydroxymethyl-Cytosine (substrate), 2nd-shell water, T1372 and Y1902. The two systems for the investigation of the role of different molecular fragments in the active site do not include either the 2nd-shell water or T1372/Y1902 in the QM subsystem (see below). All TET2 systems were calculated in the intermediate spin quintet state with Fe in the +3 oxidation state ferromagnetically coupled to the oxygen atom.²⁴

In the case of Pol λ the system is based on pdbid: 2PFO,^{66,67} the QM subsystem comprises the two known (catalytic and nucleotide-binding) Mg²⁺, 1st-coordination shell residues (D427, D429, D490), templating base, incoming nucleotide, cation-coordinating waters, a proposed 3rd Mg(II) cation with coordinating waters, and (in some cases) R386, R420, and K472 (see below).^{25,52,63} QM/MM calculations (single point or optimization) were done at the B3LYP/6-31+G(d), ω B97X-D/6-311G(d,p), and B3LYP/6-31G(d) for 4OT, TET2 and Pol λ , respectively with AMBER parm99 for the MM environment.⁶⁸ All calculations were performed with LICHEM interfacing Gaussian16 and TINKER.⁶⁹⁻⁷²

ELF analysis was performed with the TopMOD package to determine basin populations at selected basins involved with the specific reactions.⁷³ For several calculations, the Mod_wfn utility within the TopMOD package was also used to generate minimal wavefunctions only on the reactive atoms. Each system was subdivided into 200 grid points along each axis for a total of 8,120,601 segments and calculated using very high accuracy mode. Multiwfn was used for wavefunction analysis of the reactive atoms.⁷⁴ ELF heat maps were generated on a 200x200 grid using a plane defined by three reactive atoms. Close-ups of the heat map for Pol λ were generated using a 500x500 grid. VMD and UCSF Chimera were used for visualizing the iso-surface values and creating images.^{75,76}

Results

System Selection

The ultimate goal of the combined protein sequence/structure evolution and ELF analyses procedure is to determine whether this procedure can provide insights on whether a particular residue or molecular fragment should be included in the QM subsystem, or if it may be approximated by the MM potential. To this end, we have chosen systems that have previously been investigated via QM/MM and where specific molecular fragments have been shown to be appropriately represented by the QM or MM subsystems.

Figure 1: Condensed sequence alignments (left) and image of the extended active site of the corresponding enzyme (4OT, TET2 & Polλ) (right) of five members of the 4OT superfamily from pseudomonas putida (top), human TET (middle), and human X family DNA polymerases (bottom). Residues are highlighted on the alignment to match the color on the structure on the right. Active site residues highlighted/colored in blue. Sequence alignments of complete sequences for the corresponding enzyme for different species is provided in Figure S1.

4-oxalocrotonate tautomerase (4OT) catalyzes the tautomerization of 2-oxo-4-hexenedioate to 2-oxo-3-hexenedioate. The structure of 4OT is a so-called trimer of dimers comprised of 6 active sites (Figure 2, and has been shown to exhibit "half-of-the-sites" reactivity.⁷⁷

Each active site includes residues from three different monomers, thus, the residues are designated as unprimed, single-primed, or double-primed to denote that the residues belong to different monomers (see Figure 2).⁷⁸ The reaction is carried out by an N-terminal P1 acting as a general base by abstracting the proton from the substrate, and proceeds in two steps (Figure 2).^{22,23,62,79,80} Several charged residues are located in the active site including R11', R39'', and R61' (Figure 2). These residues are highly conserved (as well as most of the rest of the monomer sequence) across homologs of 4OT, including the α/β heterohexamer 4OT (Figure 1).⁸¹

Figure 2: 4OT, clockwise from left: 4OT structure (pdb id: 5TIG) with each monomer in the hexamer uniquely colored,⁸² scheme for the reaction mechanism catalyzed by 4OT, and close up of active site with conserved and active site residues.

The original QM/MM simulations showed that no general acid is required in the reaction,^{22,62} subsequently confirmed experimentally by Metanis *et al.*;⁸³ instead, R39'' stabilizes the TS1, intermediate, and TS2 structures by electrostatic interactions. QM/MM calculations have also shown that it is necessary to consider the full hexameric structure in order to provide the correct electrostatic environment for the active site under consideration.²³

TET2 is part of the TET family, itself part of the Fe/ α -ketoglutarate (α kg) super-family

that catalyzes the sequential oxidation of 5-methylcytosine (5mC) $\rightarrow$ 5-hydroxymethylcytosine (5hmC) $\rightarrow$ 5-formylcytosine (5fC) $\rightarrow$ 5-carboxycytosine (5caC) (Figure 3).³⁸ A combined experimental and theoretical study showed that an active site scaffold formed by two highly conserved residues (Figure 1), T1372 and Y1902, helps to position the substrate in the active site to allow the oxidation by a ferryl (Fe(IV)=O) intermediate to proceed.³⁸ Recent QM/MM simulations have investigated the effect of other mutations around the active site for the 5mC to 5hmC oxidation step.⁸⁴ Separate QM/MM simulations showed that the rate-limiting step for the oxidation of 5hmC to 5fC proceeds through a ferryl intermediate (characteristic of Fe/ α k family enzymes) via a hydrogen atom abstraction from the hydroxyl of 5hmC, while a T1372A variant produces a change in orientation of 5hmC in the active site, resulting in a higher barrier.²⁴

Figure 3: TET2, clockwise from left: TET2 structure (pdb id: 4NM6), scheme for the reaction mechanism catalyzed by TET2, and close up of modeled active site with 5mC, conserved/active site residues and ordered water (W2).

Additionally, the results from these QM/MM simulations suggest that a water molecule in the "2nd-shell" of the ferryl polarizes the oxyl that performs the H atom abstraction (Figure 3).²⁴ Ordered water molecules in enzyme active sites are known to play important roles in catalysis.^{85,86} The second-shell ligand effect on ferryl-based catalysis of C-H bond breaking (activation) has been reported previously in other Fe/ α k family enzymes like

AlkB,⁸⁷ inorganic complexes,⁸⁸ and Fe-based metal-organic frameworks.⁸⁹

Pol λ belongs to the X-family of polymerases and is involved in fixing double-stranded breaks in the non-homologous end-joining (NHEJ) pathway by catalyzing the nucleotide addition that fills small 1–2 residue gaps.⁹⁰ The reaction mechanism of Pol λ proceeds via the deprotonation of the O3' on the primer base and subsequent nucleophilic attack on the α -phosphate of the incoming nucleotide with concomitant formation of pyrophosphate (Figure 4).²⁵ There are five residues conserved across human DNA polymerases in a second-shell around the active site (Figure 1).^{25,63,91} Of these five, two are conserved across the X-family — for Pol λ , these correspond to K472 and R488.

Figure 4: Pol λ , counterclockwise from left: Pol λ structure (pdbid: 2PFO), scheme for the reaction mechanism catalyzed by Pol λ , and close up of active site with conserved and active site residues.

The presence of two metals in the active site, termed catalytic and nucleotide-binding, has been extensively supported in DNA polymerases.^{25,92–95} Recently, a third divalent cation has been observed in the active sites of Pol η and Pol β (another X-family polymerase).^{96–98} Combined QM/MM simulations and structural/biochemical studies on the role of this third

cation in the reaction mechanism of Pol β ,^{99,100} and QM/MM simulations on the reaction mechanism of Pol η with a third cation,¹⁰¹ indicate that this cation serves to stabilize the product structure and avoid the reverse (pyrophosphorolysis) reaction. Prior QM/MM simulations on reactant and product structures of Pol λ including a third cation in the active site found that the third metal resulted in a negative reaction energy,^{63,69} compared with a positive reaction energy with only two cations.²⁵

We have chosen these systems to test and validate our combined procedure given the role of different conserved residues and/or conserved waters in these systems. 4OT has been selected because it has been previously shown that a minimal representation of the active site is sufficient for an accurate QM/MM evaluation of the reaction path. Conversely, for TET2 it is important to include a water molecule in the QM subsystem that is located in the 2nd-coordination shell (2nd-shell), and not directly involved in the making/breaking of covalent bonds or directly coordinating the metal center.

Based on this, we have performed a series of single point calculations on previously optimized structures of 4OT and TET2 with different residues/molecular fragments in the QM subsystem (see below) to obtain the required wavefunctions for the ELF calculations to investigate the effects of QM or MM representations of critical residues/molecules around the active site. Subsequently, we have applied the combined procedure on the reactant structure of Pol λ , including the putative third cation, to determine which residues to include in the QM subsystem followed by a full reaction path optimization with the third cation in the active site.

QM Subsystem Investigation for 4OT and TET2

The catalytic proficiency of enzymes is given in large part by electrostatic stabilization provided by the protein environment. Therefore, if a residue or molecular fragment is reasonably approximated by the MM environment, the electronic charge distribution for the QM subsystem should be similar whether that residue or fragment is included in the QM subsystem

or located in the MM environment. This should hold for both the reactant, and transition state (TS) structures.

We have performed ELF analysis on previously optimized reactant and TS structures for 4OT and TET2 to test this hypothesis. These two systems have been chosen to test this approach because they provide examples of systems where nearby molecular fragments have been shown to be suitably approximated by the MM environment on the one hand (4OT), and to be required to be included in the QM subsystem on the other (TET2). Additionally, various residues located in the 2nd- and 3rd- shell are conserved and are important for catalysis (Figure 1).

TopMod also provides the ability to generate modified wavefunction files for a subset of the atoms. We have performed ELF analyses for the reactant and TS structures of 4OT with a minimal active region and including R61' for the full wavefunction (see below, and Figure S2), and a modified wavefunction including only the three atoms involved in the bond breaking/forming process. In all cases the ELF analyses result in nearly identical results. Therefore, for all other systems only reduced wavefunctions have been considered for computational efficiency.

Reactant Structure Analysis

As noted above, the structure of 4OT is composed of six monomers arranged in a trimer-of-dimers configuration that includes six active sites (Figure 2). 4OT exhibits high sequence and structure monomeric conservation (Figure 1). Each of the six active sites includes residues from three different monomers including the N-terminal catalytic P1, R11', R39'' and R61'. The last two residues, R39'' and R61', form hydrogen bond interactions (H-bonds) with the substrate. Additionally, L8 forms a backbone H-bond with the substrate that is important for catalysis.¹⁰² All of these residues including L8, R11', R39'' and R61' are conserved (Figure 1). Previous calculations showed that the minimum energy path (MEP) and associated free energy path (FEP) can be calculated with only P1 and the substrate in

the QM subsystem.^{22,23,62}

Figure 5: ELF analysis for 4OT reactant structures. **a)** ELF of reactive atoms comparing minimal QM (yellow surface) with R39'' (red wireframe). **b)** ELF of reactive atoms comparing minimal QM (yellow surface) with R61' (cyan wireframe). **c)** Multipolar decomposition analyses for the valence basin of the transferring proton. **d)** 2D ELF Heatmap for Minimal QM. **e)** 2D ELF Heatmap for QM with R39''. **f)** 2D ELF Heatmap for QM with R61'.

ELF calculations including multipolar decomposition (up to 2nd moments) based on wavefunctions from single point energies were performed for three different 4OT systems, including a minimal QM subsystem (P1 and substrate), one including P1, substrate and R39'', and a third with P1, substrate and R61' (Figures 5 and S1). In all cases, the total population for the basins associated with the atoms involved in the reaction is consistent between systems, with only minor changes (± 0.02 e) for a small number of core and valence basins on heavy atoms. One exception occurs for the valence basin of O14 in the full wavefunction ELF analysis, which exhibits a difference of -0.1 for the R61' compared with the minimal systems.

The calculated first and second moments for the three 4OT systems are also similar, with maximum differences of ± 0.02 and ± 0.03 for individual components respectively. The basin corresponding to the bond that will be broken during the first step of the reaction (C2-H4) is largely unchanged between systems, with population remaining the same, dipole moment changing by -0.003, and quadrupole moment changing by -0.006 (Figure 5c). These data indicate that there is no appreciable difference on the electronic charge distribution for the reactant structure independent of whether the arginine sidechains are represented by the MM potential, or explicitly included in the QM subsystem.

Three systems have been considered for TET2. The first involves a QM region comprised of the substrate and full first coordination sphere of the Fe: Fe(III), oxo (ferromagnetically-coupled, intermediate spin quintet), H1382, D1384, H1881, succinate, one 1st-shell H₂O, 5hmC, the 2nd-shell H₂O, T1372 and Y1902. The second system (no 2nd-shell H₂O) excludes the second shell water (Figures 3, S2 and S3), and the third system that was considered, excludes the conserved T1372 and Y1902 (no T1372/Y1902).

Figure 6 presents the ELF analysis for the three TET2 systems (also see Figures S3 and S4). Comparison of the system that includes the conserved T1372/Y1902 (large QM) with the system where these residues are represented by the MM environment suggests that these two conserved residues can be approximated by the MM environment without large changes on the electronic charge distribution on the reactive atoms, consistent with the results reported by Waheed *et al.*⁸⁴ Comparing the TET2 active site with and without T1372 and Y1902 included in the QM region, we find that the population of the valence basin at the metal-coordinating oxyl changes by 0.04, the dipole moment is negligibly changed by 0.021, and the quadrupole moment is changed by 0.004 (see Figure 6c).

Figure 6: ELF analysis of TET2 reactant structures. **a)** ELF of oxyl and O-H atoms on 5hmC (reactive atoms) comparing large QM (yellow surface) with QM excluding T1372/Y1902 (cyan wireframe). **b)** ELF of oxyl and O-H atoms on 5hmC (reactive atoms) comparing large QM (yellow surface) with QM excluding second shell water (red wireframe). **c)** Multipolar decomposition analyses for the valence basin of the oxyl atom. **d)** 2D ELF Heatmap for large QM. **e)** 2D ELF Heatmap for QM excluding T1372/Y1902. **f)** 2D ELF Heatmap for QM excluding second shell water.

Conversely, the comparison of the ELF analysis for the TET2 system with all fragments in the QM region, with the TET2 system without the 2nd-shell water included in the QM region shows that the population of the basin at the metal-coordinating oxyl changes by -1.16, the dipole moment changes by a magnitude of -1.128, and the quadrupole moment changes by a magnitude of -1.367 (see Figure 6c). Interestingly, this change in the ELF is not apparent from visual inspection of the 3D or 2D surfaces, unlike those of the TS (see below). Thus, the multipolar decomposition provides a more detailed insight on the

electronic charge distributions on the ELF basins. These results indicate that the second-shell water has a significant effect on the electronic charge distribution. To investigate this further, we performed the same analysis for the TS structures (see below).

Transition State ELF Analysis

The ELF analysis for the first TS of the proton abstraction from 2o4hex catalyzed by 4OT is shown in Figure 7 (Figure S5). The comparison of the ELF surfaces and multipolar decomposition show no differences in the electronic charge distribution between the minimal QM region compared with the systems that include either R39'' or R61', similar to what is observed in the reactant. These results indicate that the classical potential provides a good approximation around the QM subsystem in this case. Moreover, these results are consistent with the experimental confirmation of the absence of an explicit general acid in the reaction, and instead, R39'' only providing electrostatic stabilization for the 4OT catalyzed reaction.⁸³

Figure 7: **a)** ELF of reactive atoms comparing minimal QM (yellow surface) with R39'' (red wireframe). **b)** ELF of reactive atoms comparing minimal QM (yellow surface) with R61' (cyan wireframe). **c)** Multipole Analysis for reactive proton. **d)** ELF Heatmap for Minimal QM. **e)** ELF Heatmap for QM with R39''. **f)** ELF Heatmap for QM with R61'.

For the TET2 ELF analysis, similar effects to the reactant are observed. The TS structures for the 5hmC oxidation to 5fC catalyzed by wild type TET2 correspond to the hydrogen atom transferring from the hydroxyl O on the substrate to the oxyl atom on the ferryl.²⁴ The scaffold formed by the highly conserved T1372 and Y1902 residues (Figure 1) has been shown to be an integral part of the active site by orienting the substrate in the active site.³⁸ Comparing the ELF analyses of the TS structures when these two residues are included in the QM subsystem with the system where these residues are in the MM environment shows no changes on the basin associated with the transferring H atom as shown in Figure 8 (and Figure S6).

On the other hand, significant differences are observed when looking at the system with the 2nd-shell water represented by the classical force field. In this case, a significant reduction

in the population is observed with a concomitant change in the first and second moments of the basin. Here, the changes are such that the ELF basins for the lone pairs on the hydroxyl O are significantly different (Figures 8b and S5). Moreover, the representation of the water by the classical potential results in a significant overpolarization of the Fe atom inducing a subvalence splitting (Figure S6). This subvalence splitting effect on transition metals has been observed in other enzymatic systems.¹⁰³

Figure 8: **a)** ELF of truncated wavefunction around reactive atoms in TET2 transition state with and without T1372 and Y1902 in the QM region. **b)** ELF of truncated wavefunction around reactive atoms in TET2 transition state with and without second shell water in the QM region. **c)** comparison of electron localization and multipolar moments of the iron-oxyl electronic basin for full QM region, QM without T1372 and Y1902, and QM without second shell water. **d)** ELF heatmap through region of reactive atoms Fe, O, H, and O with the full QM region. **e)** ELF heatmap through region of reactive atoms with T1372 and Y1902 represented as point charges in the MM region. **f)** ELF heatmap through region of reactive atoms with second shell water represented as point charges in the MM region.

QM Subsystem Investigation for Pol λ

To test the combined evolution/ELF approach, we have applied it to the investigation of the reaction mechanism of DNA synthesis by Pol λ with three Mg²⁺ cations in the active site. Previous calculations that investigated the mechanism of Pol λ involved only two metal cations.^{25,103} These simulations indicated that several 2nd-shell residues are catalytically important, subsequently confirmed experimentally,¹⁰⁴ some of which are conserved across human polymerases (Figure 1).^{25,63,91}

Additionally, the calculated reaction energy for the DNA synthesis step with two Mg²⁺ or two Mn²⁺ cations was reported to be endoergic.²⁵ As mentioned above, a third metal has been reported to provide stabilization for the product in two other polymerases, Pol η and Pol β .^{99–101,105} This stabilization for the product has also been observed for Pol λ ,^{63,69} although the role of the third cation on the reaction mechanism of Pol λ has not been investigated.

Therefore, we investigated the effect of three different conserved residues in the 2nd-shell of Pol λ , R386, R420, and K472, (Figures 1, and S6) on the electronic distribution of the active site. Five systems have been tested including the original system (minimal QM region), three systems including only one of these 2nd-shell residues, and one more with all three residues included in the QM region.

The ELF analysis for the original system, when compared with systems where a single one of these additional residues show no difference on the basins associated with the breaking of the P α –O bond (Figure S7). The most robust test of the evolutionarily conserved second-shell residues compared the original QM region against a QM region with R386, R420, and K472 included (Figure 9). The ELF analysis reveals minimal differences between the original system and the system including all three residues. The population changes by 0.01 in the large system for the reactive phosphate bond. The slight differences in the reacting bond’s electronic environment engendered by the three residues either alone or collectively suggest that it is not necessary to include R386, R420, or K472 in the QM region.

Figure 9: a) ELF for Pol λ without R386, R420, and K472 in the QM region (blue). The reaction centers are outlined in purple. b) ELF for Pol λ with R386, R420, and K472 in the QM region (pink). The reaction centers are outlined in purple. c) Comparison of electron localization with (pink) and without these three residues (blue). d) ELF heatmap of Pol λ without additional amino acids on the plane passing through the reactive phosphate bond. e) Closeup of the reactive phosphate bond without additional amino acids in QM region. f) ELF heatmap of Pol λ with R386, R420, and K472 on the plane passing through the reactive phosphate bond. g) Closeup of the reactive phosphate bond with three additional amino acids in QM region.

A path optimization of the DNA synthesis reaction with the third Mg^{2+} was performed using the quadratic string method (QSM) combined with a restrained MM procedure as implemented in LICHEM.⁷⁰ The QSM path was approximated with 9 total images (7 structures between the optimized reactant and product) and optimized with an RMSD tolerance of 5×10^{-4} and 0.05 Å for the QM and MM subsystems respectively. The QM subsystem consists of 117 atoms including the original 2-cation system,²⁵ plus the new third cation, four water molecules, and 4 pseudobonds.^{63,69}

The calculated minimum energy path including the third cation, without any of the 2nd-shell residues, is similar to the originally reported 2-cation mechanism. The MEP for the 3-cation system suggests a two-step mechanism with TS1 associated with the deprotonation

of the O3' on the primer base, followed by the nucleophilic attack of the deprotonated O3' on the P α of the incoming nucleotide. The calculated energy barriers for the approximate TS structures are 13.8 and 17.5 kcal/mol for the deprotonation and nucleophilic attack respectively (Figure 10). The rate-limiting barrier for the present mechanism is similar to the previously reported barrier of 17.6 kcal/mol with only 2 cations. These barriers are consistent with the experimental barrier approximated by transition state theory of 16.6 kcal/mol.²⁵ In addition, the calculated product (approximated from the optimized reactant) shows a difference of less than 0.2 Å RMSD compared with the experimental structure (Figure S8), suggesting that the minimal QM subsystem is sufficient for this simulation.

Figure 10: Critical structures along the MEP, a) reactant, b) TS1, c) TS2, d) product, and QSM optimized path e) for the DNA synthesis reaction catalyzed by Pol λ with a third Mg²⁺ in the active site.

Two differences are observed between the previous and current reaction mechanism. The first is the fact that the calculated reaction energy is endoergic, $\Delta E = -0.5$ kcal/mol, for the 3-cation mechanism compared with $\Delta E = 2.3$ kcal/mol for the 2-cation Mg²⁺ calculated path. Additionally, the present MEP indicates that the proton transfer is lower in energy than the rate-limiting step with an intermediate energy plateau in between the two barriers, whereas in the two-cation mechanism both steps were almost iso-energetic with a high energy plateau. Thus, the current results suggest that the third cation stabilizes the product, but

does not affect the rate-limiting step in the DNA synthesis reaction catalyzed by Pol λ .

Conclusion

A new procedure that combines protein sequence/structure evolution and ELF analyses has been developed to determine the size of the QM subsystem for QM/MM simulations. The use of evolutionary analysis provides a tool to determine potentially catalytically important protein residues. The use of ELF analysis provides a tool to investigate the effects of different environments on the electronic charge distribution of the QM subsystem, and how this distribution is polarized depending on whether a particular fragment is represented by the classical potential, or included in the QM region. Two systems, 4OT and TET2, were shown to provide positive and negative controls. In 4OT, a charged residue is reasonably approximated by the MM environment, even though it directly interacts with the substrate. For TET2, two highly conserved residues can be represented by the classical potential; however, a 2nd-shell water needs to be included in the QM region. The procedure was further tested by calculating the MEP associated with the DNA synthesis step catalyzed by Pol λ with a third cation in the active site. The results indicate that three conserved residues around the active site can be represented by the MM environment, and the resulting MEP with the third cation is consistent with previous experimental results for Pol λ and results for two other polymerases including a third cation.

Supporting Information Available

Analysis for full wavefunction files for 4OT and TET2, reduced wavefunction ELF analysis for 4OT and TET2, and reduced wavefunction analysis for Pol λ with single 2nd-shell residues. Movie of the 3-cation Pol λ reaction mechanism based on the optimized structures along the MEP. Sequence/structure analysis files for all systems; ELF population and multipolar decomposition outputs; and cartesian coordinates of optimized structures along the MEP

for DNA Pol λ are available free of charge at <https://pubs.acs.org>.

Acknowledgement

This work was funded by NIH R01GM108583 and NSF CHE-1856162. Computing time from CASCaM with partial support from NSF CHE-1531468 is gratefully acknowledged.

References

- (1) Amaro, R. E.; Mulholland, A. J. Multiscale methods in drug design bridge chemical and biological complexity in the search for cures. *Nat. Rev. Chem.* **2018**, *2*.
- (2) Senn, H. M.; Thiel, W. QM/MM methods for biomolecular systems. *Angewandte Chemie - International Edition* **2009**, *48*, 1198–1229.
- (3) Groenhof, G. Introduction to QM/MM simulations. *Methods in Molecular Biology* **2013**, *924*, 43–66.
- (4) 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.
- (5) Dziedzic, J.; Head-Gordon, T.; Head-Gordon, M.; Skylaris, C. K. Mutually polarizable QM/MM model with in situ optimized localized basis functions. *J. Chem. Phys.* **2019**, *150*.
- (6) Solt, I.; Kulhánek, P.; Simon, I.; Winfield, S.; Payne, M. C.; Csányi, G.; Fuxreiter, M. Evaluating Boundary Dependent Errors in QM/MM Simulations. *J. Phys. Chem. B* **2009**, *113*, 5728–5735.
- (7) 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. Theo. Comp.* **2012**, *8*, 2260–2271.
- (8) 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.
- (9) Hartman, J.; Neubauer, T.; Caulkins, B.; Mueller, L.; G.J.O., B. Converging nuclear magnetic shielding calculations with respect to basis and system size in protein systems. *J. Biomol. NMR* **2015**, *62*, 327–340.

- (10) 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. *The Journal of Physical Chemistry B* **2016**, *120*, 11381–11394.
- (11) Jindal, G.; Warshel, A. Exploring the Dependence of QM/MM Calculations of Enzyme Catalysis on the Size of the QM Region. *Journal of Physical Chemistry B* **2016**, *120*, 9913–9921.
- (12) 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. Theo. Comp.* **2018**, *14*, 1695–1705.
- (13) Cui, Q.; Pal, T.; Xie, L. Biomolecular QM/MM Simulations: What Are Some of the "Burning Issues"? *J. Phys. Chem. B* **2021**, *0*, null, PMID: 33401903.
- (14) Sumner, S.; Söderhjelm, P.; Ryde, U. Effect of Geometry Optimizations on QM-Cluster and QM/MM Studies of Reaction Energies in Proteins. *J. Chem. Theo Comput.* **2013**, *9*, 4205–4214.
- (15) Olsen, J. M. H.; List, N. H.; Kristensen, K.; Kongsted, J. Accuracy of Protein Embedding Potentials: An Analysis in Terms of Electrostatic Potentials. *J. Chem. Theo. Comput.* **2015**, *11*, 1832–1842.
- (16) Kratz, E. G.; Walker, A. R.; Lagardère, L.; Lipparini, F.; Piquemal, J.-P.; Cisneros, G. A. LICHEM: A QM/MM program for simulations with multipolar and polarizable force fields. *J. Comput. Chem.* **2016**, *37*, 1019–1029.
- (17) Loco, D.; Lagardère, L.; Caprasecca, S.; Lipparini, F.; Mennucci, B.; Piquemal, J.-P. Hybrid QM/MM Molecular Dynamics with AMOEBA Polarizable Embedding. *J. Chem. Theo. Comput.* **2017**, *13*, 4025–4033.

- (18) Nabo, 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. Theo. Comput.* **2017**, *13*, 6230–6236.
- (19) Loco, D.; Lagardère, L.; Cisneros, G. A.; Scalmani, G.; Frisch, M.; Lipparini, F.; Menucci, B.; Piquemal, J.-P. Towards large scale hybrid QM/MM dynamics of complex systems with advanced point dipole polarizable embeddings. *Chem. Sci.* **2019**, *10*, 7200–7211.
- (20) Nochebuena, J.; Naseem-Khan, S.; Cisneros, G. A. Development and Application of QM/MM Methods with Advanced Polarizable Potentials. 2020.
- (21) Cisneros, G. A.; Wang, M.; Silinski, P.; Fitzgerald, M. C.; Yang, W. Theoretical and experimental determination on two substrates turned over by 4-oxalocrotonate tautomerase. *Journal of Physical Chemistry A* **2006**, *110*, 700–708.
- (22) Cisneros, G. A.; Liu, H.; Zhang, Y.; Yang, W. Ab initio QM/MM study shows there is no general acid in the reaction catalyzed by 4-oxalocrotonate tautomerase. *Journal of the American Chemical Society* **2003**, *125*, 10384–10393.
- (23) Wu, P.; Cisneros, G. A.; Hu, H.; Chaudret, R.; Hu, X.; Yang, W. Catalytic mechanism of 4-oxalocrotonate tautomerase: Significances of protein - Protein interactions on proton transfer pathways. *Journal of Physical Chemistry B* **2012**, *116*, 6889–6897.
- (24) Torabifard, H.; Cisneros, G. A. Insight into wild-type and T1372E TET2-mediated 5hmC oxidation using ab initio QM/MM calculations. *Chemical Science* **2018**, *9*, 8433–8445.
- (25) Cisneros, G. A.; Perera, L.; García-Díaz, M.; Bebenek, K.; Kunkel, T. A.; Pedersen, L. G. Catalytic mechanism of human DNA polymerase λ with Mg^{2+} and Mn^{2+}

- from ab initio quantum mechanical/molecular mechanical studies. *DNA Repair* **2008**, *7*, 1824–1834.
- (26) Chaudret, R.; Piquemal, J. P.; Andrés Cisneros, G. Correlation between electron localization and metal ion mutagenicity in DNA synthesis from QM/MM calculations. *Physical Chemistry Chemical Physics* **2011**, *13*, 11239–11247.
- (27) Casalino, L.; Nierzwicki, L.; Jinek, M.; Palermo, G. Catalytic Mechanism of Non-Target DNA Cleavage in CRISPR-Cas9 Revealed by Ab Initio Molecular Dynamics. *ACS Catalysis* **2020**, 13596–13605.
- (28) Fang, D.; Lord, R. L.; Cisneros, G. A. Ab initio QM/MM calculations show an intersystem crossing in the hydrogen abstraction step in dealkylation catalyzed by AlkB. *Journal of Physical Chemistry B* **2013**, *117*, 6410–6420.
- (29) Pratt, A. C.; Dewage, S. W.; Pang, A. H.; Biswas, T.; Barnard-Britson, S.; Cisneros, G. A.; Tsodikov, O. V. Structural and computational dissection of the catalytic mechanism of the inorganic pyrophosphatase from *Mycobacterium tuberculosis*. *Journal of Structural Biology* **2015**, *192*, 76–87.
- (30) Fang, D.; Cisneros, G. A. Alternative Pathway for the Reaction Catalyzed by DNA Dealkylase AlkB from Ab Initio QM/MM Calculations. *Journal of Chemical Theory and Computation* **2014**, *10*, 5136–5148.
- (31) Cisneros, G. A.; Perera, L.; Sehaaper, R. M.; Pedersen, L. C.; London, R. E.; Pedersen, L. G.; Darden, T. A. Reaction mechanism of the ϵ subunit of *E. coli* DNA polymerase III: Insights into active site metal coordination and catalytically significant residues. *Journal of the American Chemical Society* **2009**, *131*, 1550–1556.
- (32) Solt, I.; Kulhánek, P.; Simon, I.; Winfield, S.; Payne, M. C.; Csányi, G.; Fuxreiter, M. Evaluating boundary dependent errors in QM/MM simulations. *Journal of Physical Chemistry B* **2009**, *113*, 5728–5735.

- (33) DeYonker, N. J.; Webster, C. E. A Theoretical Study of Phosphoryl Transfers of Tyrosyl-DNA Phosphodiesterase i (Tdp1) and the Possibility of a "dead-End" Phosphohistidine Intermediate. *Biochemistry* **2015**, *54*, 4236–4247.
- (34) Qi, H. W.; Karelina, M.; Kulik, H. J. Quantifying electronic effects in QM and QM/MM biomolecular modeling with the Fukui function. *Wuli Huaxue Xuebao/ Acta Physico - Chimica Sinica* **2017**, *34*, 81–91.
- (35) Karelina, M.; Kulik, H. J. Systematic Quantum Mechanical Region Determination in QM/MM Simulation. *Journal of Chemical Theory and Computation* **2017**, *13*, 563–576.
- (36) Fuller, J.; Wilson, T. R.; Eberhart, M. E.; Alexandrova, A. N. Charge Density in Enzyme Active Site as a Descriptor of Electrostatic Preorganization. *Journal of Chemical Information and Modeling* **2019**, *59*, 2367–2373.
- (37) De La Lande, A.; Alvarez-Ibarra, A.; Hasnaoui, K.; Cailliez, F.; Wu, X.; Mineva, T.; Cuny, J.; Calaminici, P.; López-Sosa, L.; Geudtner, G.; Navizet, I.; Iriepa, C. G.; Salahub, D. R.; Köster, A. M. Molecular simulations with in-deMon2k QM/MM, a tutorial-review. *Molecules* **2019**, *24*.
- (38) Liu, M. Y.; Torabifard, H.; Crawford, D. J.; DeNizio, J. E.; Cao, X. J.; Garcia, B. A.; Cisneros, G. A.; Kohli, R. M. Mutations along a TET2 active site scaffold stall oxidation at 5-hydroxymethylcytosine. *Nature Chemical Biology* **2017**, *13*, 181–187.
- (39) Ribeiro, A.; Tyzack, J.; Borkakoti, N.; Holliday, G.; J.M., T. A global analysis of function and conservation of catalytic residues in enzymes. *J. Biol. Chem.* **2020**, *295*, 314–324.
- (40) Arnold, F. H. Directed Evolution: Bringing New Chemistry to Life. *Ang. Chem. Intl. Ed.* **2018**, *57*, 4143–4148.

- (41) Choi, I. G.; Kim, S. H. Evolution of protein structural classes and protein sequence families. *Proceedings of the National Academy of Sciences of the United States of America* **2006**, *103*, 14056–14061.
- (42) Panchenko, A. R.; Madej, T. Structural similarity of loops in protein families: Toward the understanding of protein evolution. *BMC Evolutionary Biology* **2005**, *5*, 1–8.
- (43) Spellerberg, B.; Martin, S.; Weber-Heynemann, J.; Schnitzler, N.; Lütticken, R.; Rozdzinski, E.; Podbielski, A. Lmb, a protein with similarities to the LraI adhesin family, mediates attachment of *Streptococcus agalactiae* to human laminin. *Infection and Immunity* **1999**, *67*, 871–878.
- (44) Salter, J. D.; Bennett, R. P.; Smith, H. C. The APOBEC Protein Family: United by Structure, Divergent in Function. *Trends in Biochemical Sciences* **2016**, *41*, 578–594.
- (45) Friedberg, I.; Margalit, H. Persistently conserved positions in structurally similar, sequence dissimilar proteins: Roles in preserving protein fold and function. *Protein Science* **2009**, *11*, 350–360.
- (46) Whitman, C. P. The 4-oxalocrotonate tautomerase family of enzymes: How nature makes new enzymes using a β - α - β structural motif. *Archives of Biochemistry and Biophysics* **2002**, *402*, 1–13.
- (47) Navaratnam, N.; Sarwar, R. An overview of cytidine deaminases. *International Journal of Hematology* **2006**, *83*, 195–200.
- (48) Piquemal, J.-P.; Pilmé, J.; Parisel, O.; Gérard, H.; Fourré, I.; Bergès, J.; Gourlaouen, C.; De La Lande, A.; Van Severen, M.-C.; Silvi, B. What can be learnt on biologically relevant systems from the topological analysis of the electron localization function? *Intl. J. Quant. Chem.* *108*, 1951–1969.

- (49) Becke, A. D.; Edgecombe, K. E. A simple measure of electron localization in atomic and molecular systems. **1990**, *92*, 5397–5403.
- (50) Silvi, B.; Savin, A. Classification of chemical bonds based on topological analysis of electron localization functions. *Nature* **1994**, *371*, 683–686.
- (51) Gillet, N.; Chaudret, R.; Contreras-García, J.; Yang, W.; Silvi, B.; Piquemal, J.-P. Coupling Quantum Interpretative Techniques: Another Look at Chemical Mechanisms in Organic Reactions. *J. Chem. Theo. Comp.* **2012**, *8*, 3993–3997.
- (52) Fang, D.; Chaudret, R.; Piquemal, J.-P.; Cisneros, G. A. Toward a Deeper Understanding of Enzyme Reactions Using the Coupled ELF/NCI Analysis: Application to DNA Repair Enzymes. *J. Chem. Theo. Comp.* **2013**, *9*, 2156–2160.
- (53) Pilme, J.; Piquemal, J.-P. Advancing beyond charge analysis using the electronic localization function: Chemically intuitive distribution of electrostatic moments. *J. Comput. Chem.* **2008**, *29*, 1440–1449.
- (54) Notredame, C.; Higgins, D. G.; Heringa, J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **2000**, *302*, 205–217.
- (55) O’Sullivan, O.; Suhre, K.; Abergel, C.; Higgins, D. G.; Notredame, C. 3DCoffee: combining protein sequences and structures within multiple sequence alignments. *J Mol Biol* **2004**, *340*, 385–395.
- (56) Poirot, O.; Suhre, K.; Abergel, C.; O’Toole, E.; Notredame, C. 3DCoffee@igs: a web server for combining sequences and structures into a multiple sequence alignment. *Nucleic Acids Res* **2004**, *32*, 37–40.
- (57) Armougom, F.; Moretti, S.; Poirot, O.; Audic, S.; Dumas, P.; Schaeli, B.; Keduas, V.; Notredame, C. Espresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic Acids Res* **2006**, *34*, W604–608.

- (58) Di Tommaso, P.; Moretti, S.; Xenarios, I.; Orobittg, M.; Montanyola, A.; Chang, J. M.; Taly, J. F.; Notredame, C. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res* **2011**, *39*, W13–17.
- (59) Moretti, S.; Armougom, F.; Wallace, I. M.; Higgins, D. G.; Jongeneel, C. V.; Notredame, C. The M-Coffee web server: a meta-method for computing multiple sequence alignments by combining alternative alignment methods. *Nucleic Acids Res* **2007**, *35*, W645–648.
- (60) Wallace, I. M.; O’Sullivan, O.; Higgins, D. G.; Notredame, C. M-Coffee: combining multiple sequence alignment methods with T-Coffee. *Nucleic Acids Res* **2006**, *34*, 1692–1699.
- (61) Bateman, A. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **2019**, *47*, D506–D515.
- (62) Cisneros, G. A.; Wang, M.; Silinski, P.; Fitzgerald, M. C.; Yang, W. The protein backbone makes important contributions to 4-oxalocrotonate tautomerase enzyme catalysis: Understanding from theory and experiment. *Biochemistry* **2004**, *43*, 6885–6892.
- (63) Walker, A. R.; Cisneros, G. A. Computational Simulations of DNA Polymerases: Detailed Insights on Structure/Function/Mechanism from Native Proteins to Cancer Variants. *Chemical Research in Toxicology* **2017**, *30*, 1922–1935.
- (64) Taylor, A. B.; Czerwinski, R. M.; Jr., W. H. J.; Whitman, C. P.; Hackert, M. L. Crystal Structure of 4-Oxalocrotonate Tautomerase Inactivated by 2-Oxo-3-pentynoate at 2.4 Å Resolution: Analysis and Implications for the Mechanism of Inactivation and Catalysis. *Biochemistry* **1998**, *37*, 14692.
- (65) Hu, L.; Li, Z.; Cheng, J.; Rao, Q.; Gong, W.; Liu, M.; Shi, Y. G.; Zhu, J.; Wang, P.;

- Xu, Y. Crystal Structure of TET2-DNA Complex: Insight into TET-Mediated 5mC Oxidation. *Cell* **2013**, *155*, 1545–1555.
- (66) García-Díaz, M.; Bebenek, K.; Sabariego, R.; Domínguez, O.; Rodríguez, J.; Kirchhoff, T.; García-Palomero, E.; Picher, A. J.; Juárez, R.; Ruiz, J. F.; Kunkel, T. A.; Blanco, L. DNA Polymerase λ , a Novel DNA Repair Enzyme in Human Cells. *J. Biol. Chem.* **2002**, *277*, 13184–13191.
- (67) Garcia-Diaz, M.; Bebenek, K.; Krahn, J. M.; Blanco, L.; Kunkel, T. A.; Pedersen, L. C. A Structural Solution for the DNA Polymerase λ -Dependent Repair of DNA Gaps with Minimal Homology. *Molecular Cell* **2004**, *13*, 561–572.
- (68) Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguet, L.; Huang, H.; Miguels, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; Simmerling, C. ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theo. Comput.* **2020**, *16*, 528–552.
- (69) Kratz, E. G.; Walker, A. R.; Lagardère, L.; Lipparini, F.; Piquemal, J. P.; Cisneros, G. A. LICHEM: A QM/MM program for simulations with multipolar and polarizable force fields. *J. Comput. Chem.* **2016**, *37*, 1019–1029.
- (70) Gokcan, H.; Vázquez-Mongelongo, E. A.; Cisneros, G. A. LICHEM 1.1: Recent Improvements and New Capabilities. *J. Chem. Theo. Comp.* **2019**, *15*, 3056–3065.
- (71) Frisch, M. J. et al. Gaussian 16 Revision B.01. 2016; Gaussian Inc. Wallingford CT.
- (72) Ponder, J. *TINKER, Software Tools for Molecular Design, Version 5.0: the most updated version for the TINKER program can be obtained from J.W. Ponder's WWW site at <http://dasher.wustl.edu/tinker>*; Washington University: St. Louis, 2004.
- (73) Noury, S.; Krokidis, X.; Fuster, F.; Silvi, B. Computational tools for the electron localization function topological analysis. *Comput. Chem.* **1999**, *23*, 597–604.

- (74) Lu, T.; Chen, F. Multiwfn: A multifunctional wavefunction analyzer. *Journal of Computational Chemistry* **2012**, *33*, 580–592.
- (75) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **1996**, *14*, 33 – 38.
- (76) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **2004**, *25*, 1605–1612.
- (77) Azurmendi, H. F.; Miller, S. G.; Whitman, C. P.; Mildvan, A. S. Half-of-the-sites binding of reactive intermediates and their analogues to 4-Oxalocrotonate Tautomerase and induced structural assymetry of the enzyme. *Biochemistry* **2005**, *44*, 7725–7737.
- (78) Harris, T. K.; Czerwinski, R. M.; Johnson, W. H.; Legler, P. M.; Abeygunawardana, C.; Massiah, M. A.; Stivers, J. T.; Whitman, C. P.; Mildvan, A. S. Kinetic, stereochemical, and structural effects of mutations of the active site arginine residues in 4-oxalocrotonate tautomerase. *Biochemistry* **1999**, *38*, 12343–12357.
- (79) Tuttle, T.; Keinan, E.; Thiel, W. Understanding the enzymatic activity of 4-oxalocrotonate tautomerase and its mutant analogues: a computational study. *J. Phys. Chem. B* **2006**, *110*, 19685–19695.
- (80) Tuttle, T.; Thiel, W. Substrate orientation in 4-oxalocrotonate tautomerase and its effect on QM/MM energy profiles. *J. Phys. Chem. B* **2007**, *111*, 7665–7674.
- (81) Burks, E. A.; Fleming, C. D.; Mesecar, A. D.; Whitman, C. P.; Pegan, S. D. Kinetic and structural characterization of a heterohexamer 4-oxalocrotonate tautomerase from *chloroflexus aurantiacus* J-10-fl: Implications for functional and structural diversity in the tautomerase superfamily. *Biochemistry* **2010**, *49*, 5016–5027.

- (82) Stack, T. M. M.; Li, W.; Johnson, W. H.; Zhang, Y. J.; Whitman, C. P. Inactivation of 4-Oxalocrotonate Tautomerase by 5-Halo-2-hydroxy-2,4-pentadienoates. *Biochem.* **2018**, *57*, 1012–1021.
- (83) Metanis, N.; Brik, A.; Dawson, P. E.; Keinan, E. Electrostatic interactions dominate the catalytic contribution of Arg39 in 4-oxalocrotonate tautomerase. *J. Am. Chem. Soc.* **2004**, *126*, 12726–12727.
- (84) Waheed, S. O.; Chaturvedi, S. S.; Karabencheva-Christova, T. G.; Christov, C. Z. Catalytic Mechanism of Human Ten-Eleven Translocation-2 (TET2) Enzyme: Effects of Conformational Changes, Electric Field, and Mutations. *ACS Catalysis* **2021**, *11*, 3877–3890.
- (85) Knight, J. D. R.; Hamelberg, D.; McCammon, J. A.; Kothary, R. The role of conserved water molecules in the catalytic domain of protein kinases. *Proteins: Structure, Function, and Bioinformatics* **2009**, *76*, 527–535.
- (86) Decaneto, E.; Vasilevskaya, T.; Kutin, Y.; Ogata, H.; Grossman, M.; Sagi, I.; Havenith, M.; Lubitz, W.; Thiel, W.; Cox, N. Solvent water interactions within the active site of the membrane type I matrix metalloproteinase. *Phys. Chem. Chem. Phys.* **2017**, *19*, 30316–30331.
- (87) Fang, D.; Lord, R. L.; Cisneros, G. A. Ab Initio QM/MM Calculations Show an Intersystem Crossing in the Hydrogen Abstraction Step in Dealkylation Catalyzed by AlkB. *J. Phys. Chem. B* **2013**, *117*, 6410–6420.
- (88) Anderson, M. E.; Brařda, B.; Hiberty, P. C.; Cundari, T. R. Revealing a Decisive Role for Secondary Coordination Sphere Nucleophiles on Methane Activation. *J. Am. Chem. Soc.* **2020**, *142*, 3125–3131.
- (89) Vitillo, J. G.; Lu, C. C.; Cramer, C. J.; Bhan, A.; Gagliardi, L. Influence of First

- and Second Coordination Environment on Structural Fe(II) Sites in MIL-101 for C–H Bond Activation in Methane. *ACS Cat.* **2021**, *11*, 579–589.
- (90) Yamtich, J.; Sweasy, J. B. DNA polymerase Family X: Function, structure, and cellular roles. *Biochimica et Biophysica Acta* **2010**, *1804*, 1136–50.
- (91) Frouin, I.; Toueille, M.; Ferrari, E.; Shevelev, I.; Hübscher, U. Phosphorylation of human DNA polymerase λ by the cyclin-dependent kinase Cdk2/cyclin A complex is modulated by its association with proliferating cell nuclear antigen. *Nucleic Acids Research* **2005**, *33*, 5354–5361.
- (92) Steitz, T. A. DNA Polymerases: Structural Diversity and Common Mechanisms. *Journal of Biological Chemistry* **1999**, *274*, 17395–17398.
- (93) Cowan, J. A. Structural and catalytic chemistry of magnesium-dependent enzymes. *Biometals* **2002**, *15*, 225–235.
- (94) Garcia-Diaz, M.; Bebenek, K.; Krahn, J. M.; Blanco, L.; Kunkel, T. A.; Pedersen, L. C. A structural solution for the DNA polymerase lambda-dependent repair of DNA gaps with minimal homology. *Mol Cell* **2004**, *13*, 561–572.
- (95) Garcia-Diaz, M.; Bebenek, K.; Gao, G.; Pedersen, L. C.; London, R. E.; Kunkel, T. A. Structure–function studies of DNA polymerase lambda. *DNA Repair* **2005**, *4*, 1358 – 1367, The Dale W. Mosbaugh Commemorative DNA Repair Issue.
- (96) Nakamura, T.; Zhao, Y.; Yamagata, Y.; Hua, Y. J.; Yang, W. Watching DNA polymerase η make a phosphodiester bond. *Nature* **2012**, *487*, 196–201.
- (97) Freudenthal, B. D.; Beard, W. A.; Shock, D. D.; H., W. S. Observing a DNA polymerase choose right from wrong. *Cell* **2013**, *154*, 157–168.
- (98) Gao, Y.; Yang, W. Capture of a third Mg^{2+} is essential for catalyzing DNA synthesis. *Science* **2016**, *352*, 1334–1337.

- (99) Perera, L.; Freudenthal, B. D.; Beard, W. A.; Shock, D. D.; Pedersen, L. G.; Wilson, S. H. Requirement for transient metal ions revealed through computational analysis for DNA polymerase going in reverse. *Proc. Natl. Acad. Sci.* **2015**, *112*, E5228–E5236.
- (100) Perera, L.; Greudenthal, B. D.; Beard, W. A.; Pedersen, L. G.; Wilson, S. H. Revealing the role of the product metal in DNA polymerase β catalysis. *Nucl. Ac. Res.* **2017**, *45*, 2736–2745.
- (101) Stevens, D. R.; Hammes-Schiffer, S. Exploring the Role of the Third Active Site Metal Ion in DNA Polymerase η with QM/MM Free Energy Simulations. *J. Am. Chem. Soc.* **2018**, *140*, 8965–8969.
- (102) Cisneros, G. A.; Wang, M.; Silinski, P.; Fitzgerald, M.; Yang, W. The protein backbone makes important contributions to 4-Oxalocrotonate Tautomerase enzyme catalysis: Understanding from theory and experiment. *Biochemistry* **2004**, *43*, 6885–6892.
- (103) Chaudret, R.; Piquemal, J.-P.; Cisneros, G. A. Correlation between electron localization and metal ion mutagenicity in DNA synthesis from QM/MM calculations. *Phys. Chem. Chem. Phys.* **2011**, *13*, 11239–11247.
- (104) Bebenek, K.; Garcia-Diaz, M.; Zhou, R.-Z.; Povirk, L. F.; Kunkel, T. A. Loop 1 modulates the fidelity of DNA polymerase λ . *Nuc. Ac. Res.* **2010**, *38*, 5419–5431.
- (105) Perera, L.; Essmann, U.; Berkowitz, M. L. Effect of the treatment of long-range forces on the dynamics of ions in aqueous solutions. **1995**, *102*, 450 – 456.

TOC Graphic

```

1 MDPGILKAFPKRKKIHADASSK
1 MAPKKPPA-----RAAAA-K
1 MAAKGRN-----RSPS--
1 MDPGIVKAFPKRKKSHADLSSK
1 MDPGILKAFPKRKKIHADASSK
1 MEPRGILKAFPKRKKMRADSDRG
1 MELOGIVKAFPKRKKTRDDSPKV
1 MEPRGIVKAFPKRKKVRDDSGKS
1 * : :
  
```

Sequence/
Structure
Alignment