

Computational Investigation of a Series of Small Molecules as Lead Compounds for Lysyl hydroxylase-2 (LH2) Inhibition

Yazdan Maghsoud,^{1,2} Erik Antonio Vázquez-Montelongo,^{1,3} Xudong Yang,⁴ Chengwen Liu,⁴ Zhifeng Jing,⁴ Juhoon Lee,⁵ Matthew Harger,⁴ Ally K. Smith,¹ Miguel Espinoza,¹ Hou-Fu Guo,⁶ Jonathan M. Kurie,⁷ Kevin N. Dalby,^{*,5} Pengyu Ren,^{*,4} and G. Andrés Cisneros^{*1,2,8}

¹*Department of Chemistry, University of North Texas, Denton, Texas 76201, USA*

²*Present address: Department of Chemistry and Biochemistry, The University of Texas at Dallas, Richardson, Texas 75080, USA*

³*Present address: Department of Physical Medicine and Rehabilitation, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA*

⁴*Department of Biomedical Engineering, The University of Texas at Austin, Austin, Texas 78712, USA*

*E-mail: pren@utexas.edu

⁵*Division of Chemical Biology and Medicinal Chemistry, College of Pharmacy, The University of Texas at Austin, Austin, Texas 78712, USA*

*E-mail: dalby@austin.utexas.edu

⁶*Molecular and Cellular Biochemistry, College of Medicine, The University of Kentucky, Lexington, Kentucky 40536, USA*

⁷*Department of Thoracic/Head and Neck Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas 77005, USA.*

⁸*Present address: Department of Physics, The University of Texas at Dallas, Richardson, Texas 75080, USA*

*E-mail: andres@utdallas.edu

ABSTRACT

The catalytic function of Lysyl hydroxylase-2 (LH2), a member of the Fe(II)/ α KG-dependent oxygenase superfamily, is to catalyze the hydroxylation of lysine to hydroxylysine in collagen, resulting in stable hydroxylysine aldehyde-derived collagen cross-links (HLCCs). Reports show that high amounts of LH2 lead to the accumulation of HLCCs, causing fibrosis and specific types of cancer metastasis. Some members of the Fe(II)/ α KG-dependent family have also been reported to have intramolecular O₂ tunnels, which aid in transporting one of the required co-substrates into the active site. While LH2 can be a promising target to combat these diseases, efficacious inhibitors are still lacking. We have used computational simulations to investigate a series of forty-four small molecules as lead compounds for LH2 inhibition. Tunneling analyses indicate the existence of several intra-molecular tunnels. The lengths of the calculated O₂-transporting tunnels in holoenzymes are relatively longer than the apoenzyme suggesting that the ligands may affect the enzyme's structure and possibly block (at least partially) the tunnels. The sequence alignment analysis between LH enzymes from different organisms shows that all the amino acid residues with the highest occurrence rate in the oxygen tunnels are conserved. Our results suggest that the enolate form of diketone compounds establishes stronger interactions with the Fe(II) in the active site. Branching the enolate compounds with functional groups such as phenyl and pyridinyl enhances the interaction with various residues around the active site. Our results provide information about possible leads for further LH2 inhibition design and development.

INTRODUCTION

Lysyl hydroxylases (LHs) belong to the iron(II)/ α -ketoglutarate-dependent (Fe(II)/ α KG-dependent) enzyme superfamily, with more than sixty oxygenases, using Fe(II) and 2-oxoglutarate (α KG) as cofactor and co-substrate (Figure 1A).¹ The enzymatic oxidation reaction catalyzed in the active site of these enzymes occurs by activating a non-heme Fe with concomitant oxidation of α KG into succinate and carbon dioxide; followed by an H abstraction and hydroxyl rebound to hydroxylate the substrate (Figure 1B).¹⁻³ The LH subfamily consists of LH1, LH2, and LH3, encoded by the procollagen-lysine,2-oxoglutarate 5-dioxygenase genes *PLOD1*, *PLOD2*, and *PLOD3*, respectively.⁴⁻⁶ All three enzymes catalyze the hydroxylation of lysine to hydroxylysine on collagen (Figure 1C), with LH2 being the only modifier for telopeptidyl lysine residues.⁷

The role of LH2 in the human body is the formation of highly stable hydroxylysine aldehyde-derived collagen cross-links (HLCCs) mediated by the telopeptidyl lysine residues.^{8, 9} These crosslinks are more durable than lysine-derived collagen cross-links (LCCs) that form in the absence of LH2.¹⁰ The large stabilization and resistance of HLCCs to collagenase cleavage are necessary to stabilize the extracellular matrix,⁸ as the deficiency of HLCCs is seen in patients with Bruck syndrome suffering from deformed and fragile bones.¹¹ On the other hand, excessive accumulation of HLCCs by high levels of LH2 expression can cause fibrosis.^{12, 13} Moreover, pathological studies on sarcoma¹⁴ and metastasis in lung and breast cancers¹⁵⁻¹⁷ show that the high levels of LH2 directly contribute to these conditions. Thus, LH2 is a promising target for potential inhibitors to

cope with these diseases. Various assays have been developed to determine LH2 activity, some of which are currently employed to develop possible LH2 inhibitors.¹⁸⁻²²

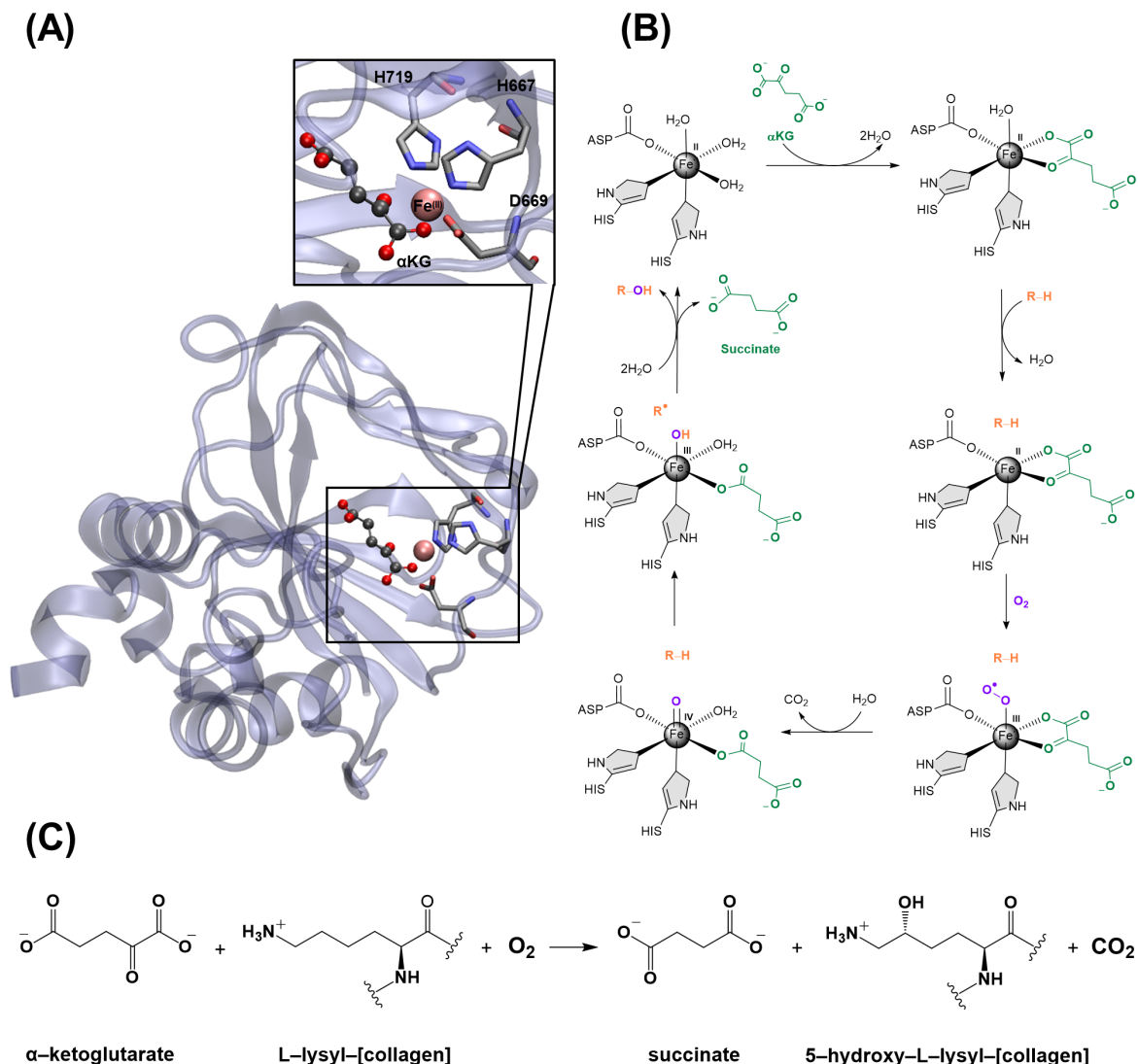


Figure 1. **(A)** Crystal Structure of full-length Human Lysyl Hydroxylase LH3 and its active site complexed with α KG (PDB ID: 6FXK).²³ **(B)** The general catalytic mechanism of the α KG-dependent hydroxylase superfamily. **(C)** Hydroxylation of lysine to hydroxylysine by the LH2 in the presence of molecular oxygen.

The hydroxylation reaction performed by LH2 is carried out via a reaction mechanism consistent with other enzymes in the Fe/ α KG-dependent superfamily. This mechanism involves molecular oxygen to activate the non-heme Fe to carry out the required oxidations. Experimental and computational studies have shown that various enzymes have intra-molecular tunnels formed by flexible hydrophobic residues to transport molecular oxygen into the active site.²⁴⁻³⁰ This feature has also been observed in some Fe/ α KG-dependent superfamily enzymes such as AlkB.³¹⁻³⁵ Yu *et al.* proposed the possibility of oxygen diffusion by intra-molecular tunnels based on the original AlkB crystal structure.³⁴ Subsequently, a computational investigation showed the likelihood of the

existence of two O₂-transporting tunnels in this enzyme by various computational approaches.³⁵

Classical atomistic molecular dynamics (MD) simulations have become a useful tool to aid in lead development and inhibitor design.³⁶⁻³⁹ Many applications use classical non-polarizable force fields implemented in various software packages such as OPLS,⁴⁰ CHARMM,^{41, 42} AMBER,^{43, 44} and GROMOS.⁴⁵ Polarizable force fields such as the Atomic Multipole Optimized Energetics for Biomolecular Applications (AMOEBA) can also be employed.⁴⁶⁻⁴⁸ In cases where transition metals are involved, an accurate description of the electronic structure around the metal may be required. One approach that can be used in these cases is quantum mechanics/molecular mechanics (QM/MM). This method combines two levels of theory and can be used to study the interaction energies, reaction energies, and reaction mechanisms in chemical and biochemical systems.⁴⁹⁻⁵⁵

A wide variety of combinations can be used to treat the QM subsystem and the MM region, such as empirical valence bond, semiempirical, or *ab initio* Hamiltonians for the QM, and non-polarizable or polarizable force fields for the MM region. It has been shown that, similar to MD, considering polarization in QM/MM simulations is important and can improve the description of the inter- and intra-molecular interactions.⁵⁶⁻⁶² Thus, combining high-level QM and polarizable/anisotropic force fields can help achieve more reliable results.⁶³⁻⁶⁶

This contribution presents a combined polarizable MD and QM/MM investigation of a series of diketone-, enol, and enolate-based ligands as lead compounds to inhibit LH2. Tunnel analyses are also presented for LH2 to investigate the possibility of the existence of molecular oxygen transport tunnels and whether the various ligands may impact them. The remainder of the manuscript is organized as follows; the Methods section describes the various computational approaches used. Subsequently, results for MD, QM/MM, and tunneling simulations are presented and discussed, followed by concluding remarks.

METHODOLOGY

Molecular Docking and MD Simulations

The homology model for LH2 was constructed by using SWISS-MODEL server.⁶⁷ The PLOD2 sequence (residue 598 to 737) was submitted. LH3 PDB structure 6fxm.1.A with 64.29% sequence identity was used as the template. Next, the protonation states of the output structure at pH 7 were predicted by using propka⁶⁸ via PDB2PQR server⁶⁹ with H atoms added/removed on the ionizable residues accordingly. The initial structures of LH2–ligand complexes were prepared using GOLD molecular docking software (version 5.8.1).⁷⁰ Initially, 16 ligands were considered in this work, followed by 28 modified compounds based on the initial results. A visual check of each docked structure was performed to confirm the ligands were in the proper position of the LH2 binding pocket. Then these complex structures were solvated and neutralized in the AMOEBA water box. Additional NaCl ions were added to reach the physiological concentration of 0.15 M. The resulted simulation boxes are about 70 * 90 * 70 cubic Angstrom with total atoms of ~44000. After 1 ns equilibration, the production MD simulations were carried out under NPT ensemble (298 K and 1 bar) for 5 ns for each protein-ligand complex. The RESPA

integrator^{71, 72}, Monte Carlo barostat⁷³ and BUSSI thermostat⁷⁴ were employed in the simulations. To accelerate the simulation, the integration time step was chosen to be 2 fs, and the induced dipole moment was converged to 0.0001 Debye. Following the conventional settings used by AMOEBA simulations, the cutoff distances for van der Waals interactions and real space electrostatics are 12 Å and 7 Å, respectively. The long-range interactions were treated with Ewald summation method as implemented in the Tinker software package. MD trajectories were saved every 50 ps for a total simulation time of 5 ns for each system. The generated trajectories were subjected to a clustering analysis and further investigation using QM/MM calculations described below. All MD simulations were performed using the Tinker software implemented via the OpenMM plugin⁷⁵ on Nvidia GPU cards (GTX 1070 or RTX 2070).

Clustering and QM/MM Calculations

The *k*-means clustering analysis⁷⁶ was performed on the MD trajectories of the apoenzyme and selected lead compounds based on the distances between the Fe²⁺ and the donor atoms of the coordinated residues (H666, D668, and H718), ligand (N/A for apoenzyme), and water molecules (whenever applicable). The LICHEM^{77, 78} (Layered Interacting CHEmical Models) program was used in combination with Gaussian16⁷⁹ and TINKER^{80, 81} for all energy and optimization calculations. The ω B97X-D/6–31G(d,p)^{82, 83} level of theory and AMOEBA¹⁸⁸⁴ force field were employed for the QM region and the MM environment, respectively. The QM/MM long-range electrostatic correction (QM/MM-LREC) method⁸⁵ was applied with a 25 Å cutoff for the QM subsystem coupled with the Particle Mesh Ewald (PME) method⁸⁶ for the MM calculations. The QM subsystem includes the Fe²⁺, ligand molecule, H666, D668, H718, and one or two water molecules depending on the ligand to complete the octahedral coordination sphere around the divalent cation. In contrast, the remaining residues and solvent molecules were assigned to the MM subsystem. The MM environment consists of two subregions; the active region, in which all the protein atoms and the solvent are flexible inside a sphere of 30 Å radius, and the frozen subregion, where the remaining MM atoms are kept fixed. Boundary conditions across the QM and MM subsystems where covalent bonds were cut were described using the pseudobond approach⁸⁷. Interaction energies ($\Delta E^{Interact.}$) were calculated using the following approach:

$$\Delta E_{Lig./Env.}^{Interact.} = [QM_{Active\ site+ligand} - (QM_{Active\ site} + QM_{Ligand})] + [MM_{System+Ligand} - MM_{System}] \quad 1$$

$$\Delta E_{Lig./Wat.}^{Interact.} = QM_{Ligand, PCM\ Water} - QM_{Ligand, Gas\ Phase} \quad 2$$

$$\Delta \Delta E^{Interact.} = \Delta E_{Lig./Env.}^{Interact.} - \Delta E_{Lig./Wat.}^{Interact.} \quad 3$$

Eqns. 1 and 2 give the interaction energy between the ligand, the solvated protein, and water, respectively. Eq. 3 reports the difference between the two calculated interaction energies. The terms $QM_{Active\ site+ligand}$ and $QM_{Active\ site}$ correspond to the gas-phase energies of the respective systems including basis set superposition error (BSSE) correction using the counterpoise approach.^{88, 89}

Non-covalent interaction and thermal fluctuation indexes analyses

Non-covalent interaction (NCI), average non-covalent interaction (aNCI), and thermal fluctuation index (TFI) analyses were used as implemented in the Multiwfn3.7⁹⁰ software, employing promolecular densities⁹¹. This analysis has a default RGB color code for plotting isosurfaces. Blue surfaces represent strong-attractive interactions; red surfaces refer to strong-repulsive interactions, and green ones show weak attractive/repulsive interactions. The thermal fluctuation index (TFI) visually indicates the variations between a set of NCI surfaces and is color-coded like the NCI. Blue (red) surfaces in this analysis mean that the TFI is smaller (larger), and the non-covalent interactions do not fluctuate that much (fluctuate a lot). NCI analysis was studied on QM/MM optimized structures, while aNCI and TFI analyses were used for the last 5ns of MD trajectories. A cubic grid of 200 au and an isovalue of 0.4 au with the color scale of $-0.05 \text{ au} < \text{sign}(\lambda_2)\rho < 0.05 \text{ au}$ were used for all surfaces. VMD⁹² software package was used to render images and visualize MD trajectories and surfaces.

Sequence Alignment

Forty-one lysyl hydroxylase (LH) enzymes coded by PLOD1, PLOD2, and PLOD3 genes from fourteen different organisms were aligned with the Human LH2 (UniProtKB: O00469) via the Clustal Omega program accessed via UniProtKB.⁹³ A specific portion of the studied enzymes (V538 to P737 in human LH2), formed by a sequence of 199 residues corresponding to the catalytically active region for the hydroxylase reaction²³ were selected for the further multiple alignments by the Espresso mode of the T-Coffee server.⁹⁴⁻⁹⁷ Lastly, conserved residues on the target systems from these multiple alignments were assessed by visual inspection to determine their proximity and positions toward the active site.

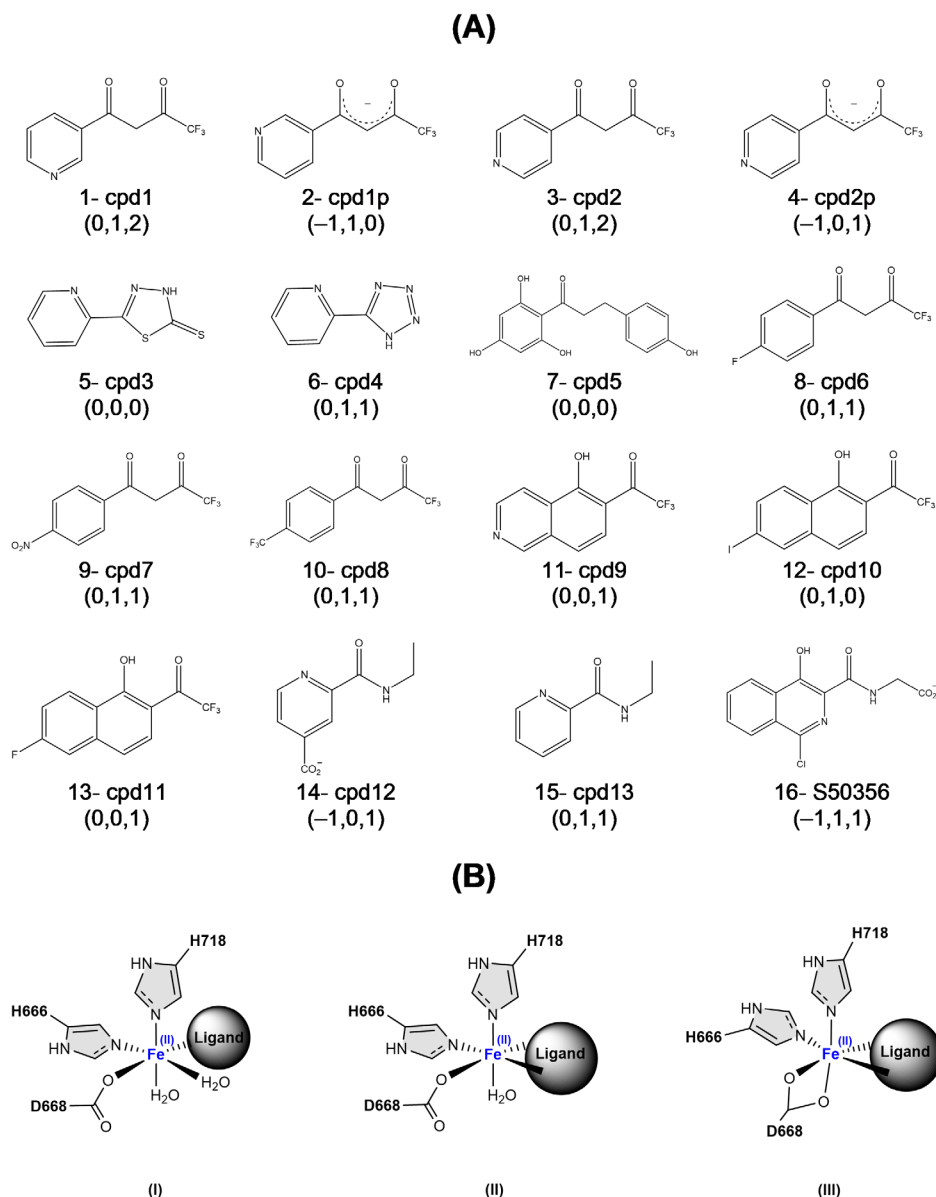
O₂-transporting tunnels

Intramolecular tunnel analysis was done with the CAVER 3.01 algorithm⁹⁸ implemented in Caver Analyst 2.0.⁹⁹ The *k*-means clustering analysis was used on the MD trajectories of selected ligand systems to identify representative structures to perform the tunneling analysis. The Fe²⁺ was considered as the starting point of the tunnels in all cases. The minimal radius of the computed tunnels (Min. probe radius) was set to 0.9 Å, the maximum depth of the surface region (shell depth) was set to 4.0 Å, and the radius of the shell probe (shell radius), defining which parts of the Voronoi diagram represent the bulk solvent was set to 3.0 Å.¹⁰⁰

RESULTS AND DISCUSSION

The natural co-substrate of the LH2 enzyme, αKG, which is the deprotonated form of 2-oxoglutaric acid, coordinates to the iron cation of the active site via its carbonyl oxygens in a bidentate form (see Figure 1A). Thus, we started with a series of structures having diketone, enol, or enolate ligand skeletons branched with aromatic rings and various functional groups (see Scheme 1A). We also considered aromatic compounds with two N: donor atoms (cpd3 and cpd4 in Scheme 1A) and picolinamide-based compounds with

C=O: and N: donors (cpd12 and cpd13 in Scheme 1A) to study the interaction tendencies of different combinations of donor atoms and negative charges. We analyzed the QM/MM interaction energies between each lead compound (ligand) and its environment ($\Delta E_{Lig./Env.}^{Interact.}$) in three configurations: First, the docking configuration with the best score. Second, the last frame of a 5 ns of MD simulation (MD1) with AMOEBA_{Bio18} using NPT ensemble at 298 K and 1 bar, with 1 kcal mol⁻¹ Å⁻² of positional restraints on all protein atoms (excluding solvent, ions, and ligand). Third, the last frame of a 5 ns MD simulation without restraints (MD2) with the same conditions as the second configuration.



Scheme 1. **(A)** The first set of test compounds. The numbers in parentheses show the ligand's charge and number of coordinated waters to the Fe cation in the active site in MD1, and MD2, respectively. **(B)** Possible coordination geometries of the ligands to Fe(II) in the active site.

Figure 2A compares the calculated interaction energies for the first set of ligands. The interaction energy differences ($\Delta\Delta E^{Interact.}$) indicate that the neutral compounds have the weakest interaction energies (detailed results in Table S1). Overall, $\Delta E_{Lig./Wat.}^{Interact.}$ represents ~20–50% of $\Delta E_{Lig./Env.}^{Interact.}$. The negatively charged ligands, cpd1p, cpd2p, cpd12, and S50356 (IDs: 2, 4, 14, and 16) have the most negative interaction energies, and their corresponding ($\Delta E_{Lig./Wat.}^{Interact.}/\Delta E_{Lig./Env.}^{Interact.}$) is 0.42, 0.50, 0.48, and 0.44, respectively. Two compounds, cpd3 and cpd13, show positive $\Delta\Delta E^{Interact.}$ as shown in Figure 2A. This is due to the fact that these neutral compounds are the only two that coordinate the metal cation in a mono-dentate fashion, and the interaction with the rest of the protein environment is not sufficient to stabilize the repulsive energy between the ligand and the active site, resulting in an overall unfavorable interaction.

Clustering analysis was performed on the ensemble generated from the MD trajectory (without restraints) for the ligands with the largest interaction energy differences, compounds cpd1p, cpd2p, cpd12, and S50356. The calculated averaged QM/MM interaction energies ($\Delta\Delta E_{Avg.}^{Interact.}$) on the selected representatives of the clusters for ligands cpd1p, cpd2p, and cpd12 shown in Figure 2B are similar (−103.1, −109.5, and −94.6 kcal mol^{−1}, respectively), while this value for S50356 is slightly larger (−112.2 kcal mol^{−1}). Calculated standard deviations for cpd12 and S50356 show the largest and smallest values of ±58.91 and ±6.26 kcal mol^{−1}, respectively. Although compounds cpd1p and cpd2p are similar, the standard deviation of cpd1p is smaller than cpd2p (±16.27 and ±31.98 kcal mol^{−1}, respectively).

To understand the nature of these differences in the calculated QM/MM interaction energies, we employed averaged non-covalent interaction index (aNCI) and thermal fluctuation index (TFI) analyses. The non-covalent interaction index (NCI) is a qualitative analysis that visually identifies the non-covalent interactions between the molecule of interest and the surrounding residues. The thermal fluctuation index is another qualitative analysis showing fluctuation and instability in the no-covalent interactions between the compound and the surrounding residues. aNCI in Figure 2C shows continuous and blue-green surfaces among cpd12 and the surrounding residues, which corresponds to van der Waals interactions, and red surfaces with well-defined blue spots for the strong repulsive-attractive interactions (cpd12-NO₂ group and LH2-Arg residue). This visual analysis shows that overall, cpd12 has many and consistent interactions with the neighboring residues. Based on the TFI surfaces for cpd12 in Figure 2D, the high standard deviation is due to the high fluctuations (red surfaces) around the carboxylate and ethyl groups. In the case of ligand S50356 (ID: 16), despite the carboxylate group, more stability is seen due to a polycyclic aromatic hydrocarbon (see Figure S3). For cpd1p and cpd2p (ID: 2 and 4), the nitrogen atom's position in the pyridine group affects the enolate's binding to the Fe⁺² ion (see Figures S1 and S2).

Based on these results, it became evident that the negative charge significantly increases the interaction tendency, but the values of $\Delta\Delta E_{Avg.}^{Interact.}$ among the ligands with the most robust interactions were relatively similar. Thus, the specific effect of the negative charge, was still unclear. In other words, it is unclear whether the enolate-based compounds with the negative charge on the donor oxygen or the anionic compounds with the negative charge far from the donor atoms are better ligands. To better understand

this issue, we broadened the combinations of the second series to have diketone-, enolate-, and picolinamide-based structures branched with more aromatic rings and various functional groups (see scheme 2).

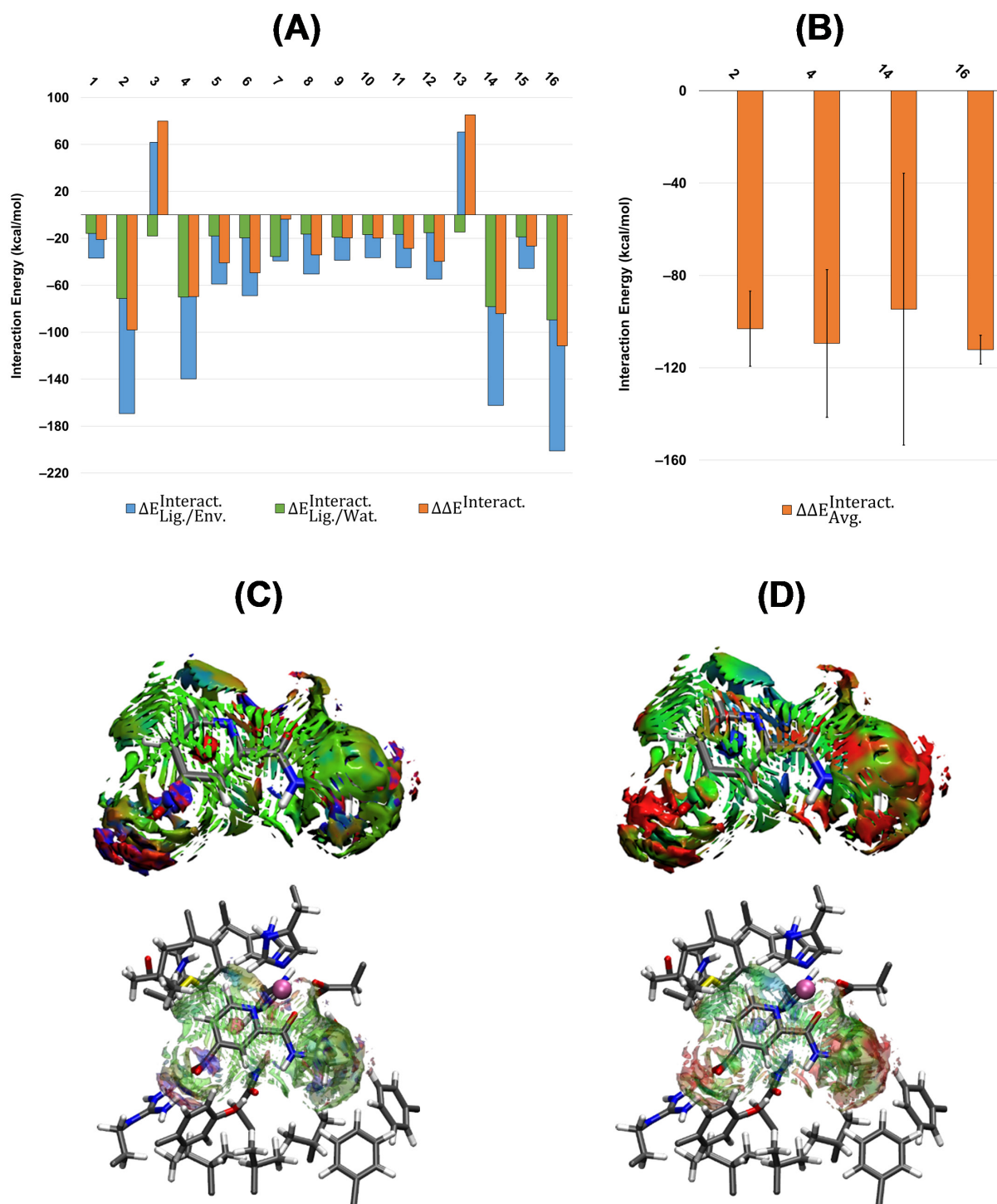


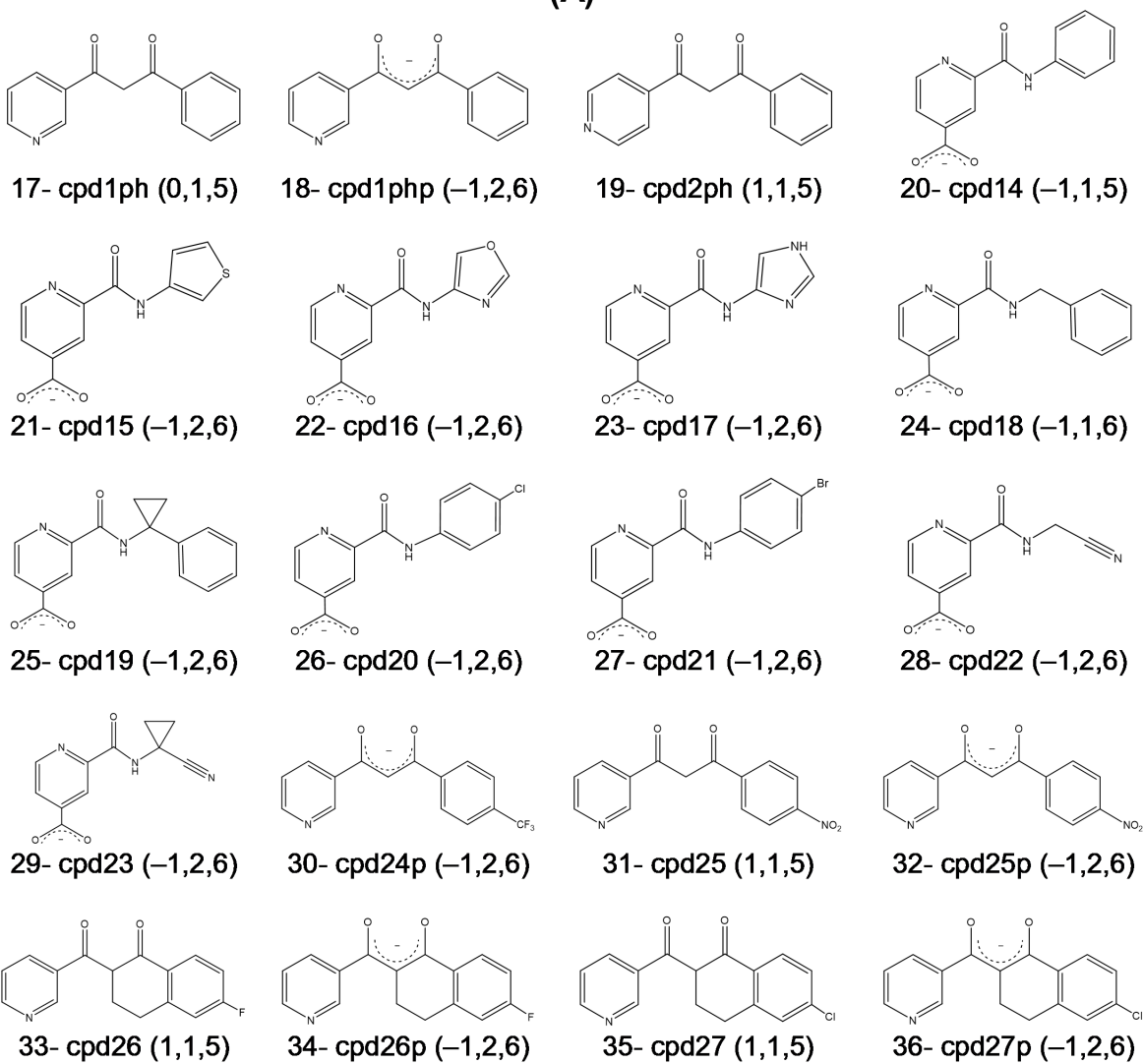
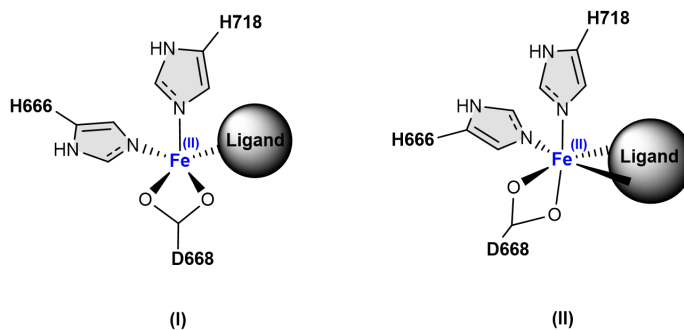
Figure 2. (A) Calculated QM/MM interaction energies (kcal mol⁻¹) for the first set of test compounds (ID: 1–16). (B) Average QM/MM interaction energies ($\Delta \Delta E_{\text{Avg.}}^{\text{Interact.}}$) and the standard deviation for the compounds with the most negative interaction energies. All the calculations are at ω B97X-D/6–31G(d,p)

level of theory and AMOEBAbio18 force field. **(C)** Averaged non-covalent index (aNCI) and **(D)** thermal fluctuation index (TFI) analyses for cpd12 (ID: 14). The isosurface cutoff for NCI and TFI is 0.4 au, and the data is plotted in the color range $-0.05 < \text{sign}(\lambda_2)\rho < 0.05$ au.

In the second set of the lead compounds listed in Scheme 2A, an equilibrated LH2 system with one of the selected ligands from the first set (cpd1p, cpd2p, cpd12, or S50356) was used, and the ligand was replaced with a new one, followed by QM/MM optimization. The results of the interaction energies are shown in Figure 3A. Our QM/MM optimizations showed distorted square pyramidal or octahedral geometries based on the ligand's coordination type (see Scheme 2B). All the neutral diketone compounds (IDs: 17, 19, 31, 33, and 35) are coordinated in a monodentate mode, while all the picolinamide- and enolate-based compounds are bidentate-coordinated (except IDs: 20 and 24).

Based on the results of the interaction energies, calculated values of $\Delta E_{\text{Lig./Wat.}}^{\text{Interact.}}$ are ~30–50% of $\Delta E_{\text{Lig./Env.}}^{\text{Interact.}}$, similar to the first set of ligands. Most of the ligands have considerable negative values of $\Delta\Delta E^{\text{Interact.}}$, except for cpd1ph, cpd2ph, cpd25, cpd26, and cpd27 with the corresponding values of -43.3 , -33.6 , -21.6 , -34.9 , and -45.2 kcal mol⁻¹, respectively. All the mentioned $\Delta\Delta E^{\text{Interact.}}$ correspond to the neutral, monodentate ligands. Moreover, once the enolate form of cpd1ph, cpd25, cpd26, and cpd27, which are cpd1php, cpd25p, cpd26p, and cpd27p coordinate to the active site, the $\Delta\Delta E^{\text{Interact.}}$ significantly improve to -115.0 , -124.5 , -143.9 , and -144.7 kcal mol⁻¹, respectively. These results are consistent with our previous observation for the first set, which showed that the enolate ligands have stronger interactions than the neutral ligands. We also realized that the negative charge effect on the interaction energy is significantly higher when closer to the active site. In other words, the enolate-based compounds having a negative charge on their donor oxygen have stronger interaction with the active site compared to the picolinamide-based ligands in which a $-\text{COO}^-$ group is far from the active site.

Non-covalent interactions between the second set's ligands and their surrounding residues were another important factor that needed to be considered. Interaction energies showed that enolate-based compounds are more energetically favorable, while the surrounding residues should stabilize a good candidate structurally. The main reason to branch the second series of the compounds with more aromatic rings, substituted with electron-donating and electron-withdrawing groups, was to see the effect of expanding the ligand's size on the non-covalent interactions. The residues that show the most interactions with the ligands are F651, V653, Y655, L663, N675, C690, I730, and V732 (Table S2 and Figure S4). As mentioned before, based on our $\Delta\Delta E^{\text{Interact.}}$ values in Figure 3A, compound cpd27p, and cpd26p exhibit the strongest interactions, while cpd2ph had one of the weakest interactions. However, the NCI plots for these compounds in Figure 3B–D show negligible differences in the numbers of interacting amino acid residues among these compounds. This observation may come from the fact that the second series of designed compounds are made from an equilibrated LH2 system with one of the selected ligands of the first set, subjected to QM/MM optimization without MD simulations.

(A)**(B)**

Scheme 2. (A) The second set of the studied ligands. Numbers in parentheses are ligand's charge, ligand's coordination mode (monodentate or bidentate), and coordination number of the complex, respectively. **(B)** Different geometries of the active site in coordination with the lead compounds.

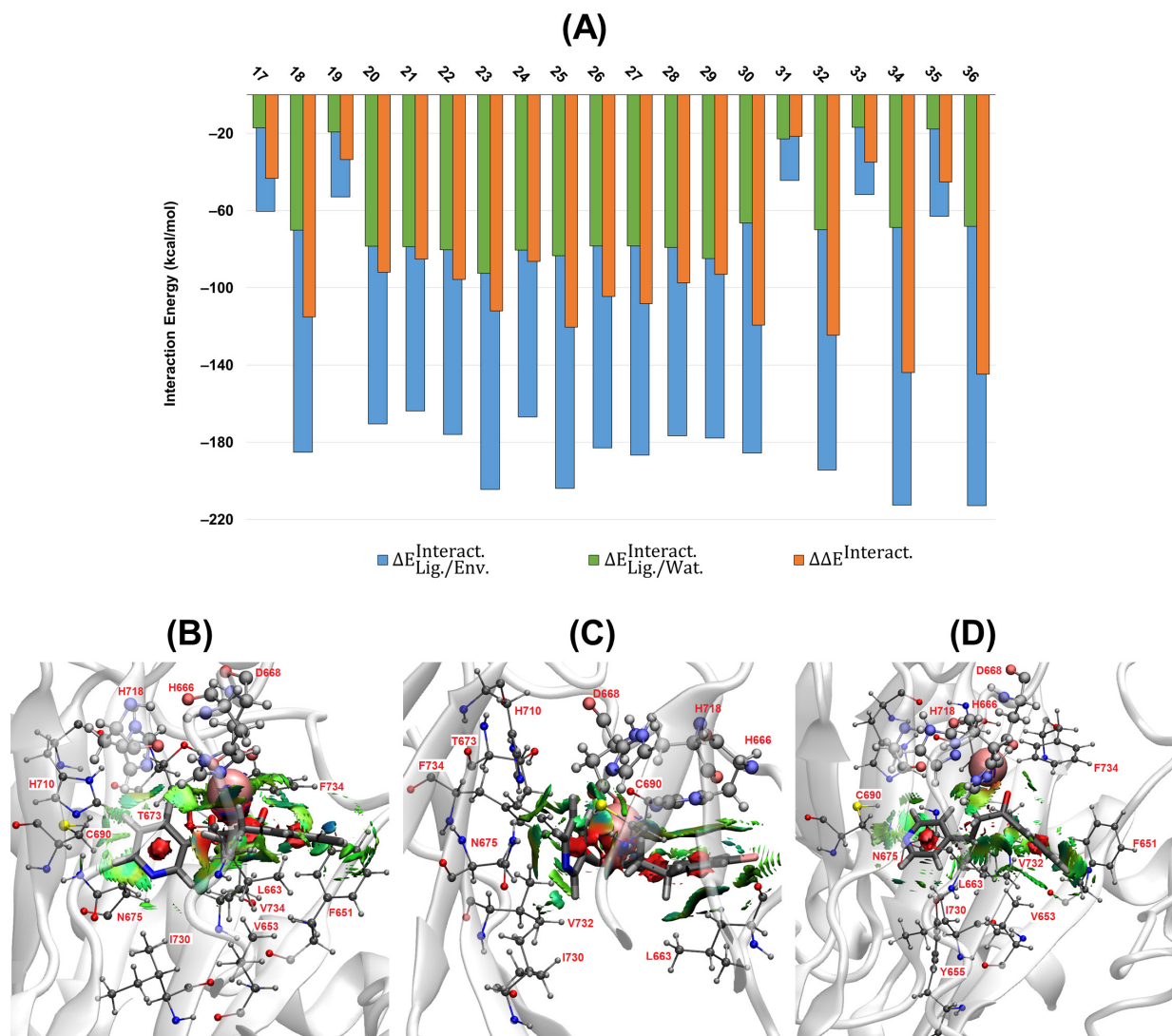


Figure 3. (A) Calculated QM/MM interaction energies (kcal mol⁻¹) for the second set of the studied ligands (ID: 17–36) at ω B97X-D/6–31G(d,p) level of theory and AMOEBAbio18 force field. The plot of the non-covalent interactions (NCI) between the ligand and surrounding amino acid residues for **(B)** cpd27p, **(C)** cpd26p, and **(D)** cpd2ph. The isosurface cutoff for the NCI is 0.4 au, and the data is plotted in the color range $-0.05 < \text{sign}(\lambda_2)\rho < 0.05$ au. Residues of the active site (H666, D668, and H718) are shown in ball-and-sticks, ligand atoms are in thick licorice, and Fe²⁺ is in a pink vdW sphere, and the surrounding residues with the non-covalent interactions are in thin licorice.

Our observations from the first and the second series of the compounds indicate that the enolate-based ligands have stronger interaction with the active site. Additionally, several residues around the active site have attractive non-covalent interactions with compounds of the second set, which means all the designed ligands of this series are stabilized by the environment. Previous studies have reported O₂-transporting tunnels in other α KG-dependent superfamily members³¹⁻³⁵, based on this, we hypothesize that O₂-transporting tunnels might also exist in LH2. Therefore, before designing the third set of enolate-based compounds, we studied the existence of O₂-transporting tunnels in the apoenzyme. We performed a 5 ns MD simulation on the apo structure and applied *k*-means clustering on the trajectories. The geometry of the active site was considered

Figure 4. (A) O₂-transporting tunnels with largest calculated percentages along the trajectory observed in the apoenzyme. Calculated tunnels are colored in blue, green, and red, respectively, based on the tunnel's length (blue: shortest, red: longest). The intersection of the tunnels inside the protein is the active site, which is not shown for clarity. The molecular surface of the protein is shown in magnification to give a better view of the enzyme's cavities. **(B)** Close-up of the approximate positions of the tunnel-lining residues with more than ~80% presence in all the calculated tunnels/representatives. **(C)** Condensed sequence alignment between human LH2 (UniProtKB: O00469) and selected members of the LH family. Key: "*" indicates residues conserved in all the sequences, while ":" and "." indicate highly and weakly conserved residues.

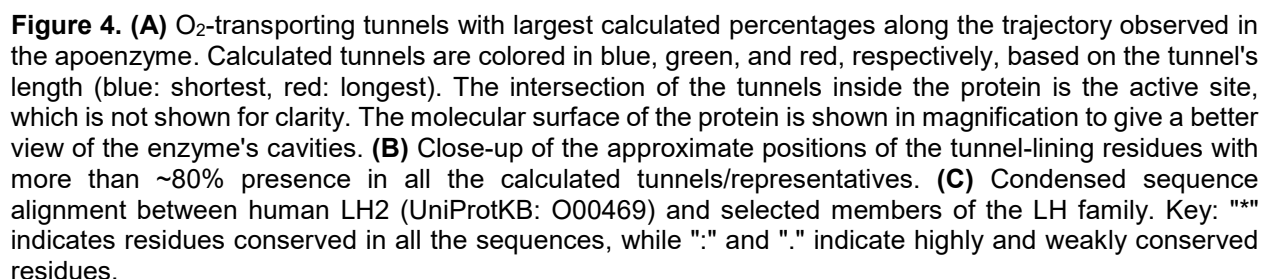


Figure 4A shows the three major tunnels that are accessible for the apoenzyme with the priority/availability of 0.92/63% (blue), 0.83/51% (green), and 0.52/32% (red). There were some other accessible tunnels, but their availability and priority were much less than the three mentioned tunnels (detailed results in Table S11). It should be noted that, in addition to the tunnel's availability, two other factors are essential when considering a calculated tunnel as a major tunnel; the cost and priority of the tunnel.^{100, 101} Here, the cost of each tunnel, which is defined as the balance between the width and length of the tunnel (Å), was considered the priority (e^{-cost}) of the tunnels. The tunnels with a cost less than ~0.7 (wide and short) and a higher priority than ~0.5 were considered the threshold for selecting the tunnels.

Several common residues are present in most of the representatives' tunnels. Table S3 shows that the tunnel-defining residues present in more than ~80% of the tunnels are V653, Y655, L663, H666, D668, T673, N675, C690, H710, H718, E719, G720, I730, V732, and F734. These residues are shown in Figure 4B. A sequence alignment between all the three LH enzymes of Homo sapiens (LH1, LH2, and LH3) and fourteen other organisms shows that all the tunnel-defining residues are conserved (see Figure 4C and Tables S5 and S6).

Based on our observations from the last two series of compounds and the presence of the oxygen-transporting tunnels in the apoenzyme with conserved amino acids, we designed a third set of lead compounds consisting of eight enolate-based structures with more aromatic rings substituted with electron-donating and electron-withdrawing groups (see Figure 5). One of the goals for the third set was to see if the new ligands could affect the oxygen transport by blocking the putative tunnels, as well as establishing further interactions with conserved residues around the active site. Furthermore, we calculated the interaction energies between each compound and the active site and studied their non-covalent interactions with the surrounding amino acids. All the ligands were parametrized, and holo-structures were subjected to MD simulations. The stability of the whole structure during the MD simulations has been illustrated by RMSD of the core part of LH2 bound with the compounds of this set (see Figure S6). At the same time, the main interactions between Fe ion and the surrounding active site (H666, D668 and H718) in those complexes have been kept stable. The specific data for those distances have been recorded in Table S7. The 5 ns of each structure's MD trajectories were taken for further clustering analysis to select the best representative structures. The k-means clustering results and the related QM/MM optimization energies for each representative are given in Table S8. The geometries of the active site in all the representatives for each compound are also shown in Figures S7–S14.

Like the apo structure, we used each compound's six representatives to calculate the O₂-transporting tunnels. The close-ups of the major calculated tunnels for each compound are given in Figure 5. The length/width of the three major tunnels in the apo structure—selected by the criteria as mentioned earlier—were 2.2/1.9 (blue), 5.8/1.7 (green), and 10.2/0.9 Å (red) (Table S4). Additionally, it is observed that the holo-structures have relatively longer tunnels. This may be due to the structural effects of the incoming ligands after binding to the active site and elongating the tunnel. A comparison of the tunnel availability between apo and holo-structures indicates a decrease in the tunnel availability by all the compounds. This reduction is 27%–51% and 31%–53% for

the blue and green tunnels, respectively, with the highest availability in the apo structure. Analysis of the conserved residues lining the tunnels for the holo systems is consistent with the apo enzyme tunnel analysis (Tables S12–S19).

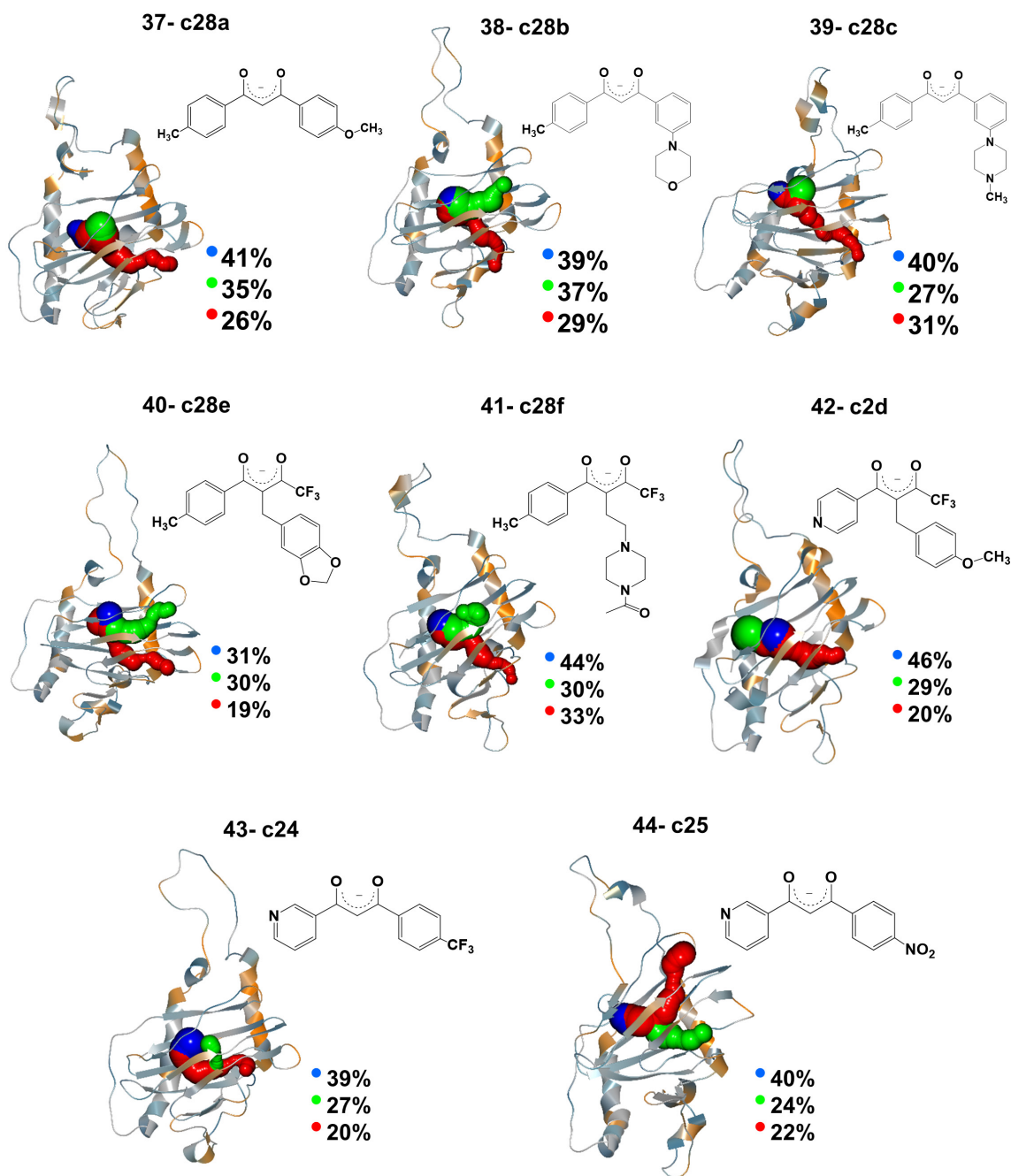


Figure 5. O₂-transporting tunnels with largest calculated percentages along the trajectory observed in the third set of lead compounds. The 2D structure of each ligand is also given next to its tunnel graph. Calculated tunnels are colored in blue, green, and red, respectively, based on the increase in the tunnel's length. The intersection of the tunnels inside the protein is the active site of the inhibition reaction, but the active site's residues are not shown for more clarity.

The interaction energies between the most stable representative of each ligand (the representative with the zero value of relative optimization energy in Table S8) and the active site were calculated for each ligand in the third set. Calculated interaction energies in Figure 6A show that 33–63% of the $\Delta E_{Lig./Env.}^{Interact.}$ is due to $\Delta E_{Lig./Wat.}^{Interact.}$. Based on the calculated values of $\Delta\Delta E^{Interact.}$, ligands c2d and c24 have the weakest interaction with the active site (–37.9 and –48.2 kcal mol^{–1}, respectively), while c28b and c28a have the strongest interactions, –199.3 and –147.4 kcal mol^{–1}, respectively (Table S1).

The aNCI analysis for the third set of ligands in Table S10 shows that the amino acid residues interacting with most of the ligands from this third set are F651, Y655, C690, R728, and F734. The aNCI graphs in Figure 6B qualitatively show that c28b and c28a, which have the strongest interaction with the active site, also have non-covalent interactions with more surrounding residues than c2d and c24, with the weakest interaction with the active site (supplementary aNCI graphs in Figure S15). The QM/MM interaction energies and the aNCI show that c28a and c28b are energetically and structurally more favorable than the other ligands. These two compounds also had a significant blocking effect on the O₂-transporting tunnel.

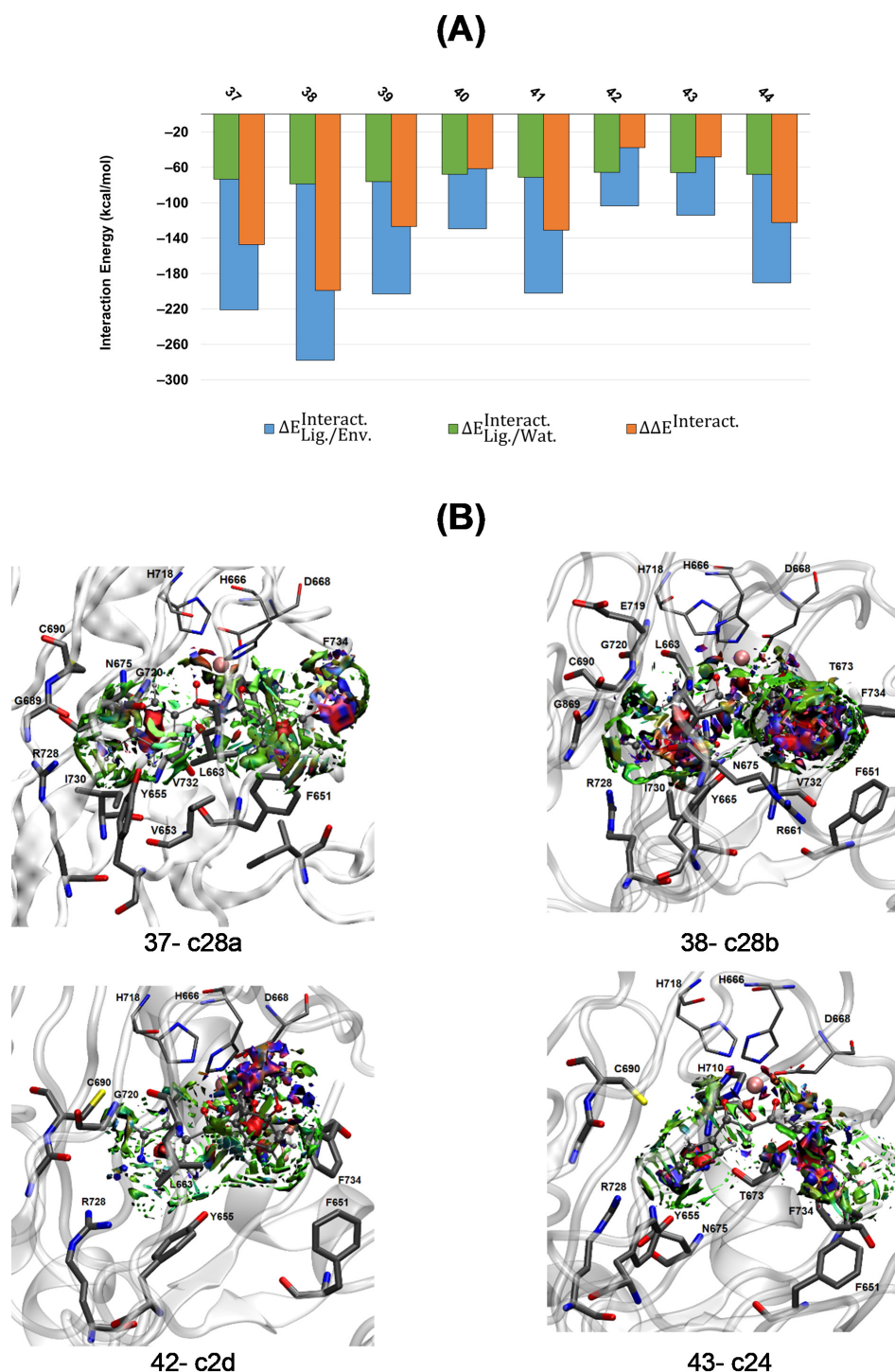


Figure 6. (A) Calculated QM/MM interaction energies (kcal mol^{-1}) for the most stable representative of the studied ligands of the third set (ID: 37–44) at $\omega\text{B97X-D/6-31G(d,p)}$ level of theory and AMOEBA_{bio}18 force field. **(B)** Averaged non-covalent interactions (aNCl) between the ligand and its surrounding residues for c28a, c28b, c2d, and c24. The isosurface cutoff for the NCI is 0.4 a.u., and the data is plotted in the color range $-0.05 < \text{sign}(\lambda_2)\rho < 0.05$ au. Residues of the active site (H666, D668, and H718) are shown in thin licorice, ligand atoms are in ball-and-sticks, Fe^{2+} is in a pink vdW sphere, and the surrounding residues with the non-covalent interactions are in thick licorice. Hydrogens are not shown for more clarity.

CONCLUSION

Forty-four small molecules were investigated in three sequential sets as lead compounds for LH2 inhibition. The first set of compounds was designed to have one or two oxygen, nitrogen, or a combination of both to coordinate with the iron cation of the active site. Results showed that diketone-based compounds do not interact strongly with the active site, while the enolate form has considerably large negative interaction energies. We designed compounds with various functional groups for the second set but with the main skeleton of diketone, enol, enolate (to see the effect of the negative charge and the structural change), and picolinamide-based compounds having oxygen and nitrogen donor atoms. Results showed that enolate-based compounds have the most potent interactions with the active site, similar to the first set. Intramolecular O₂-transport tunnel analysis was carried out on apo-LH2 to determine the feasibility of the existence of this feature in LH2. We found three major tunnels with the availability of 63%, 51%, and 32%, in which a series of tunnel-defining residues had more than 80% of repetition. Further investigation on the evolutionary conservation of amino acid residues via the multiple sequence alignment on LH enzymes (LH1, LH2, and LH3) in forty-one organisms showed that all the residues that show more than 80% presence in the tunnels are conserved. For the third set of ligands, new enolate-based compounds branched with more rings and functional groups were designed, to attempt to block the O₂-transport tunnels and enhance the interactions with the active site and the surrounding residues. Results showed that the availability of the tunnels is relatively higher in the apo structure than in the holo-structures with the third set and some of the compounds in this third set may block the tunnels. The interaction energies and the non-covalent interactions showed that compounds 37- C28a and 38- C28b have the strongest interactions with the active site and non-covalent interactions with most of the surrounding amino acid residues.

ACKNOWLEDGMENTS

This work was partially supported by National Institutes of Health (R01GM108583, R01GM106137, R01GM114237, R01CA105155 and K99CA225633) and National Science Foundation (CHE-2217856, and CHE-1856173), the Cancer Prevention and Research Institute of Texas (award numbers RP160657, RP210088, and RP160652.), and the Welch Foundation (award number F-1390). Computational time provided by the University of Texas at Dallas CyberInfrastructure Facilities and the University of North Texas CASCaM CRUNTCh3 high-performance cluster partially supported by NSF grant nos. CHE-1531468 and OAC-2117247. Additional computing time from XSEDE Project TG-CHE160044 is gratefully acknowledged. E.A.V.-M. thanks CONACyT for funding.

CONFLICTS

PR is co-founder of Qubit Pharmaceuticals.

REFERENCES

1. Loenarz, C.; Schofield, C. J., Expanding chemical biology of 2-oxoglutarate oxygenases. *Nature chemical biology* **2008**, *4* (3), 152-156.
2. Hausinger, R. P., Fe (II)/ α -ketoglutarate-dependent hydroxylases and related enzymes. *Critical reviews in biochemistry and molecular biology* **2004**, *39* (1), 21-68.
3. Flashman, E.; Schofield, C. J., The most versatile of all reactive intermediates? *Nature chemical biology* **2007**, *3* (2), 86-87.
4. Hautala, T.; Byers, M. G.; Eddy, R. L.; Shows, T. B.; Kivirikko, K. I.; Myllyla, R., Cloning of human lysyl hydroxylase: Complete cDNA-derived amino acid sequence and assignment of the gene (PLOD) to chromosome 1p36. 3→ p36. 2. *Genomics* **1992**, *13* (1), 62-69.
5. Xu, Y.; Baldassare, M.; Fisher, P.; Rathbun, G.; Oltz, E. M.; Yancopoulos, G. D.; Jessell, T. M.; Alt, F. W., LH-2: a LIM/homeodomain gene expressed in developing lymphocytes and neural cells. *Proceedings of the National Academy of Sciences* **1993**, *90* (1), 227-231.
6. Passoja, K.; Rautavuoma, K.; Ala-Kokko, L.; Kosonen, T.; Kivirikko, K. I., Cloning and characterization of a third human lysyl hydroxylase isoform. *Proceedings of the National Academy of Sciences* **1998**, *95* (18), 10482-10486.
7. Takaluoma, K.; Lantto, J.; Myllyharju, J., Lysyl hydroxylase 2 is a specific telopeptide hydroxylase, while all three isoenzymes hydroxylate collagenous sequences. *Matrix biology* **2007**, *26* (5), 396-403.
8. Scott, I.; Yamauchi, M.; Sricholpech, M., Lysine post-translational modifications of collagen. *Essays in biochemistry* **2012**, *52*, 113-133.
9. Pornprasertsuk, S.; Duarte, W. R.; Mochida, Y.; Yamauchi, M., Lysyl hydroxylase - 2b directs collagen cross - linking pathways in MC3T3 - E1 cells. *Journal of Bone and Mineral Research* **2004**, *19* (8), 1349-1355.
10. Bailey, A.; Peach, C. M.; Fowler, L., Chemistry of the collagen cross-links. Isolation and characterization of two intermediate intermolecular cross-links in collagen. *Biochemical Journal* **1970**, *117* (5), 819-831.
11. Ha - Vinh, R.; Alanay, Y.; Bank, R. A.; Campos - Xavier, A. B.; Zankl, A.; Superti - Furga, A.; Bonafé, L., Phenotypic and molecular characterization of Bruck syndrome (osteogenesis imperfecta with contractures of the large joints) caused by a recessive mutation in PLOD2. *American journal of medical genetics Part A* **2004**, *131* (2), 115-120.
12. Van Der Slot, A. J.; Zuurmond, A.-M.; Van Den Bogaardt, A. J.; Ulrich, M. M.; Middelkoop, E.; Boers, W.; Ronday, H. K.; DeGroot, J.; Huizinga, T. W.; Bank, R. A., Increased formation of pyridinoline cross-links due to higher telopeptide lysyl hydroxylase levels is a general fibrotic phenomenon. *Matrix biology* **2004**, *23* (4), 251-257.
13. van der Slot, A. J.; Zuurmond, A.-M.; Bardoel, A. F.; Wijmenga, C.; Pruijs, H. E.; Sillence, D. O.; Brinckmann, J. r.; Abraham, D. J.; Black, C. M.; Verzijl, N., Identification of PLOD2 as telopeptide lysyl hydroxylase, an important enzyme in fibrosis. *Journal of Biological Chemistry* **2003**, *278* (42), 40967-40972.
14. Eisinger-Mathason, T. K.; Zhang, M.; Qiu, Q.; Skuli, N.; Nakazawa, M. S.; Karakasheva, T.; Mucaj, V.; Shay, J. E.; Stangenberg, L.; Sadri, N., Hypoxia-dependent modification of collagen networks promotes sarcoma metastasis. *Cancer discovery* **2013**, *3* (10), 1190-1205.
15. Gilkes, D. M.; Bajpai, S.; Chaturvedi, P.; Wirtz, D.; Semenza, G. L., Hypoxia-inducible factor 1 (HIF-1) promotes extracellular matrix remodeling under hypoxic conditions by inducing P4HA1, P4HA2, and PLOD2 expression in fibroblasts. *Journal of Biological Chemistry* **2013**, *288* (15), 10819-10829.
16. Gilkes, D. M.; Bajpai, S.; Wong, C. C.; Chaturvedi, P.; Hubbi, M. E.; Wirtz, D.; Semenza, G. L., Procollagen lysyl hydroxylase 2 is essential for hypoxia-induced breast cancer metastasis. *Molecular cancer research* **2013**, *11* (5), 456-466.
17. Chen, Y.; Terajima, M.; Yang, Y.; Sun, L.; Ahn, Y.-H.; Pankova, D.; Puperi, D. S.; Watanabe, T.; Kim, M. P.; Blackmon, S. H., Lysyl hydroxylase 2 induces a collagen cross-link switch in tumor stroma. *The Journal of clinical investigation* **2015**, *125* (3), 1147-1162.
18. Luo, L.; Pappalardi, M. B.; Tummino, P. J.; Copeland, R. A.; Fraser, M. E.; Grzyska, P. K.; Hausinger, R. P., An assay for Fe (II)/2-oxoglutarate-dependent dioxygenases by enzyme-coupled detection of succinate formation. *Analytical biochemistry* **2006**, *353* (1), 69-74.

19. McNeill, L.; Bethge, L.; Hewitson, K.; Schofield, C., A fluorescence-based assay for 2-oxoglutarate-dependent oxygenases. *Analytical biochemistry* **2005**, 336 (1), 125-131.
20. Guo, H.-F.; Cho, E. J.; Devkota, A. K.; Chen, Y.; Russell, W.; Phillips Jr, G. N.; Yamauchi, M.; Dalby, K. N.; Kurie, J. M., A scalable lysyl hydroxylase 2 expression system and luciferase-based enzymatic activity assay. *Archives of biochemistry and biophysics* **2017**, 618, 45-51.
21. Devkota, A. K.; Veloria, J. R.; Guo, H.-F.; Kurie, J. M.; Cho, E. J.; Dalby, K. N., Development of a high-throughput lysyl hydroxylase (LH) assay and identification of small-molecule inhibitors against LH2. *SLAS DISCOVERY: Advancing Life Sciences R&D* **2019**, 24 (4), 484-491.
22. Rhoads, R. E.; Udenfriend, S., Decarboxylation of alpha-ketoglutarate coupled to collagen proline hydroxylase. *Proceedings of the National Academy of Sciences of the United States of America* **1968**, 60 (4), 1473.
23. Scietti, L.; Chiapparino, A.; De Giorgi, F.; Fumagalli, M.; Khorauli, L.; Nergadze, S.; Basu, S.; Olieric, V.; Cucca, L.; Banushi, B., Molecular architecture of the multifunctional collagen lysyl hydroxylase and glycosyltransferase LH3. *Nature communications* **2018**, 9 (1), 1-11.
24. Ruscio, J. Z.; Kumar, D.; Shukla, M.; Prisant, M. G.; Murali, T.; Onufriev, A. V., Atomic level computational identification of ligand migration pathways between solvent and binding site in myoglobin. *Proceedings of the National Academy of Sciences* **2008**, 105 (27), 9204-9209.
25. Tomita, A.; Kreutzer, U.; Adachi, S.-i.; Koshihara, S.-y.; Jue, T., 'It's hollow': the function of pores within myoglobin. *Journal of Experimental Biology* **2010**, 213 (16), 2748-2754.
26. Colloc'h, N.; Gabison, L.; Monard, G.; Altarsha, M.; Chiadmi, M.; Marassio, G.; Sopkova-de Oliveira Santos, J.; El Hajji, M.; Castro, B.; Abirini, J. H., Oxygen pressurized X-ray crystallography: probing the dioxygen binding site in cofactorless urate oxidase and implications for its catalytic mechanism. *Biophysical journal* **2008**, 95 (5), 2415-2422.
27. Pesce, A.; Nardini, M.; Dewilde, S.; Capece, L.; Martí, M. A.; Congia, S.; Salter, M. D.; Blouin, G. C.; Estrin, D. A.; Ascenzi, P., Ligand migration in the apolar tunnel of Cerebratulus lacteus mini-hemoglobin. *Journal of Biological Chemistry* **2011**, 286 (7), 5347-5358.
28. Elber, R., Ligand diffusion in globins: simulations versus experiment. *Current opinion in structural biology* **2010**, 20 (2), 162-167.
29. Di Russo, N. V.; Conduro, H. L.; Li, K.; Bruner, S. D.; Roitberg, A. E., Oxygen diffusion pathways in a cofactor-independent dioxygenase. *Chemical science* **2015**, 6 (11), 6341-6348.
30. Shadrina, M. S.; English, A. M.; Peshherbe, G. H., Effective simulations of gas diffusion through kinetically accessible tunnels in multisubunit proteins: O₂ pathways and escape routes in T-state deoxyhemoglobin. *Journal of the American Chemical Society* **2012**, 134 (27), 11177-11184.
31. Li, C.; Junaid, M.; Almuqri, E. A.; Hao, S.; Zhang, H., Structural analysis of a phosphonate hydroxylase with an access tunnel at the back of the active site. *Acta Crystallographica Section F: Structural Biology Communications* **2016**, 72 (5), 362-368.
32. Chaplin, V. D.; Valliere, M. A.; Hangasky, J. A.; Knapp, M. J., Investigations on the role of a solvent tunnel in the α -ketoglutarate dependent oxygenase factor inhibiting HIF (FIH). *Journal of inorganic biochemistry* **2018**, 178, 63-69.
33. Yu, B.; Hunt, J. F., Enzymological and structural studies of the mechanism of promiscuous substrate recognition by the oxidative DNA repair enzyme AlkB. *Proceedings of the National Academy of Sciences* **2009**, 106 (34), 14315-14320.
34. Yu, B.; Edstrom, W. C.; Benach, J.; Hamuro, Y.; Weber, P. C.; Gibney, B. R.; Hunt, J. F., Crystal structures of catalytic complexes of the oxidative DNA/RNA repair enzyme AlkB. *Nature* **2006**, 439 (7078), 879-884.
35. Torabifard, H.; Cisneros, G. A., Computational investigation of O₂ diffusion through an intra-molecular tunnel in AlkB; influence of polarization on O₂ transport. *Chemical science* **2017**, 8 (9), 6230-6238.
36. Buch, I.; Giorgino, T.; De Fabritiis, G., Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proceedings of the National Academy of Sciences* **2011**, 108 (25), 10184-10189.
37. Costa, R. A.; Cruz, J. N.; Nascimento, F. C.; Silva, S. G.; Silva, S. O.; Martelli, M. C.; Carvalho, S. M.; Santos, C. B.; Neto, A. M.; Brasil, D. S., Studies of NMR, molecular docking, and molecular dynamics simulation of new promising inhibitors of cruzaine from the parasite Trypanosoma cruzi. *Medicinal Chemistry Research* **2019**, 28 (3), 246-259.

38. Śledź, P.; Caflisch, A., Protein structure-based drug design: from docking to molecular dynamics. *Current opinion in structural biology* **2018**, *48*, 93-102.
39. Zhao, H.; Caflisch, A., Molecular dynamics in drug design. *European journal of medicinal chemistry* **2015**, *91*, 4-14.
40. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J., Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society* **1996**, *118* (45), 11225-11236.
41. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I., CHARMM general force field: A force field for drug - like molecules compatible with the CHARMM all - atom additive biological force fields. *Journal of computational chemistry* **2010**, *31* (4), 671-690.
42. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S. a.; Karplus, M., CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *Journal of computational chemistry* **1983**, *4* (2), 187-217.
43. Lindorff - Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E., Improved side - chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function, and Bioinformatics* **2010**, *78* (8), 1950-1958.
44. Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham III, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P., AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications* **1995**, *91* (1-3), 1-41.
45. Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F., A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force - field parameter sets 53A5 and 53A6. *Journal of computational chemistry* **2004**, *25* (13), 1656-1676.
46. Wu, J. C.; Chattree, G.; Ren, P., Automation of AMOEBA polarizable force field parameterization for small molecules. *Theoretical chemistry accounts* **2012**, *131* (3), 1-11.
47. 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., Current status of the AMOEBA polarizable force field. *The journal of physical chemistry B* **2010**, *114* (8), 2549-2564.
48. Ren, P.; Ponder, J. W., Consistent treatment of inter - and intramolecular polarization in molecular mechanics calculations. *Journal of computational chemistry* **2002**, *23* (16), 1497-1506.
49. Magalhães, R. P.; Fernandes, H. S.; Sousa, S. F., Modelling enzymatic mechanisms with QM/MM approaches: current status and future challenges. *Israel Journal of Chemistry* **2020**, *60* (7), 655-666.
50. Ahmadi, S.; Barrios Herrera, L.; Chehelamirani, M.; Hostaš, J.; Jalife, S.; Salahub, D. R., Multiscale modeling of enzymes: QM - cluster, QM/MM, and QM/MM/MD: a tutorial review. *International Journal of Quantum Chemistry* **2018**, *118* (9), e25558.
51. Duarte, F.; Amrein, B. A.; Blaha-Nelson, D.; Kamerlin, S. C., Recent advances in QM/MM free energy calculations using reference potentials. *Biochimica et Biophysica Acta (BBA)-General Subjects* **2015**, *1850* (5), 954-965.
52. Brunk, E.; Rothlisberger, U., Mixed quantum mechanical/molecular mechanical molecular dynamics simulations of biological systems in ground and electronically excited states. *Chemical reviews* **2015**, *115* (12), 6217-6263.
53. van der Kamp, M. W.; Mulholland, A. J., Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology. *Biochemistry* **2013**, *52* (16), 2708-2728.
54. Senn, H. M.; Thiel, W., QM/MM methods for biomolecular systems. *Angewandte Chemie International Edition* **2009**, *48* (7), 1198-1229.
55. Lin, H.; Truhlar, D. G., QM/MM: what have we learned, where are we, and where do we go from here? *Theoretical Chemistry Accounts* **2007**, *117* (2), 185-199.

56. Nakano, H.; Yamamoto, T., Variational calculation of quantum mechanical/molecular mechanical free energy with electronic polarization of solvent. *The Journal of Chemical Physics* **2012**, *136* (13), 134107.
57. Boulanger, E.; Thiel, W., Solvent boundary potentials for hybrid QM/MM computations using classical drude oscillators: a fully polarizable model. *Journal of chemical theory and computation* **2012**, *8* (11), 4527-4538.
58. Illingworth, C.; Parkes, K.; Snell, C.; Marti, S.; Moliner, V.; Reynolds, C., The effect of MM polarization on the QM/MM transition state stabilization: application to chorismate mutase. *Molecular Physics* **2008**, *106* (12-13), 1511-1515.
59. Giese, T. J.; York, D. M., Charge-dependent model for many-body polarization, exchange, and dispersion interactions in hybrid quantum mechanical/molecular mechanical calculations. *The Journal of chemical physics* **2007**, *127* (19), 194101.
60. Gao, J.; Alhambra, C., A hybrid semiempirical quantum mechanical and lattice-sum method for electrostatic interactions in fluid simulations. *The Journal of chemical physics* **1997**, *107* (4), 1212-1217.
61. Slipchenko, L. V., Solvation of the excited states of chromophores in polarizable environment: Orbital relaxation versus polarization. *The Journal of Physical Chemistry A* **2010**, *114* (33), 8824-8830.
62. Gao, J., Energy components of aqueous solution: Insight from hybrid QM/MM simulations using a polarizable solvent model. *Journal of computational chemistry* **1997**, *18* (8), 1061-1071.
63. Giovannini, T.; Riso, R. R.; Ambrosetti, M.; Puglisi, A.; Cappelli, C., Electronic transitions for a fully polarizable qm/mm approach based on fluctuating charges and fluctuating dipoles: linear and corrected linear response regimes. *The Journal of Chemical Physics* **2019**, *151* (17), 174104.
64. Giovannini, T.; Puglisi, A.; Ambrosetti, M.; Cappelli, C., Polarizable QM/MM approach with fluctuating charges and fluctuating dipoles: the QM/FQFμ model. *Journal of Chemical Theory and Computation* **2019**, *15* (4), 2233-2245.
65. Vázquez-Montelongo, E. A.; Vázquez-Cervantes, J. E.; Cisneros, G. A., Polarizable ab initio QM/MM Study of the Reaction Mechanism of N-tert-Butyloxycarbonylation of Aniline in [EMIm][BF₄]. *Molecules* **2018**, *23* (11), 2830.
66. Hagras, M. A.; Glover, W. J., Polarizable embedding for excited-state reactions: dynamically weighted polarizable QM/MM. *Journal of Chemical Theory and Computation* **2018**, *14* (4), 2137-2144.
67. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L., SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research* **2018**, *46* (W1), W296-W303.
68. Olsson, M. H.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H., PROPKA3: consistent treatment of internal and surface residues in empirical p K a predictions. *Journal of chemical theory and computation* **2011**, *7* (2), 525-537.
69. Jurrus, E.; Engel, D.; Star, K.; Monson, K.; Brandi, J.; Felberg, L. E.; Brookes, D. H.; Wilson, L.; Chen, J.; Liles, K., Improvements to the APBS biomolecular solvation software suite. *Protein Science* **2018**, *27* (1), 112-128.
70. Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R., Development and validation of a genetic algorithm for flexible docking. Edited by F. E. Cohen. *Journal of Molecular Biology* **1997**, *267* (3), 727-748.
71. Humphreys, D. D.; Friesner, R. A.; Berne, B. J., A Multiple-Time-Step Molecular Dynamics Algorithm for Macromolecules. *The Journal of Physical Chemistry* **1994**, *98* (27), 6885-6892.
72. Qian, X.; Schlick, T., Efficient multiple-time-step integrators with distance-based force splitting for particle-mesh-Ewald molecular dynamics simulations. *The Journal of Chemical Physics* **2002**, *116* (14), 5971-5983.
73. Frenkel, D.; Smit, B.; Ratner, M. A., *Understanding molecular simulation: from algorithms to applications*. Academic press San Diego: 1996; Vol. 2.
74. Bussi, G.; Zykova-Timan, T.; Parrinello, M., Isothermal-isobaric molecular dynamics using stochastic velocity rescaling. *The Journal of Chemical Physics* **2009**, *130* (7), 074101.
75. Harger, M.; Li, D.; Wang, Z.; Dalby, K.; Lagardère, L.; Piquemal, J.-P.; Ponder, J.; Ren, P., Tinker-OpenMM: Absolute and relative alchemical free energies using AMOEBA on GPUs. *Journal of Computational Chemistry* **2017**, *38* (23), 2047-2055.
76. Likas, A.; Vlassis, N.; Verbeek, J. J., The global k-means clustering algorithm. *Pattern recognition* **2003**, *36* (2), 451-461.
77. Gökcen, H.; Vázquez-Montelongo, E. A.; Cisneros, G. A., LICHEM 1.1: recent improvements and new capabilities. *Journal of chemical theory and computation* **2019**, *15* (5), 3056-3065.

78. Kratz, E. G.; Walker, A. R.; Lagardère, L.; Lipparini, F.; Piquemal, J. P.; Andrés Cisneros, G., LICHEM: A QM/MM program for simulations with multipolar and polarizable force fields. *Journal of computational chemistry* **2016**, *37* (11), 1019-1029.
79. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16 Rev. C.01*, Wallingford, CT, 2016.
80. Ponder, J. W., TINKER: Software tools for molecular design. *Washington University School of Medicine, Saint Louis, MO* **2004**, 3.
81. Rackers, J. A.; Wang, Z.; Lu, C.; Laury, M. L.; Lagardère, L.; Schnieders, M. J.; Piquemal, J.-P.; Ren, P.; Ponder, J. W., Tinker 8: software tools for molecular design. *Journal of chemical theory and computation* **2018**, *14* (10), 5273-5289.
82. Chai, J.-D.; Head-Gordon, M., Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Physical Chemistry Chemical Physics* **2008**, *10* (44), 6615-6620.
83. Chai, J.-D.; Head-Gordon, M., Systematic optimization of long-range corrected hybrid density functionals. *The Journal of chemical physics* **2008**, *128* (8), 084106.
84. Zhang, C.; Lu, C.; Jing, Z.; Wu, C.; Piquemal, J.-P.; Ponder, J. W.; Ren, P., AMOEBA polarizable atomic multipole force field for nucleic acids. *Journal of chemical theory and computation* **2018**, *14* (4), 2084-2108.
85. Kratz, E. G.; Duke, R. E.; Cisneros, G. A., Long-range electrostatic corrections in multipolar/polarizable QM/MM simulations. *Theoretical chemistry accounts* **2016**, *135* (7), 1-9.
86. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A smooth particle mesh Ewald method. *The Journal of chemical physics* **1995**, *103* (19), 8577-8593.
87. Fang, D.; Chaudret, R.; Piquemal, J.-P.; Cisneros, G. A. s., Toward a deeper understanding of enzyme reactions using the coupled ELF/NCI analysis: application to DNA repair enzymes. *Journal of Chemical Theory and Computation* **2013**, *9* (5), 2156-2160.
88. Simon, S.; Duran, M.; Dannenberg, J., How does basis set superposition error change the potential surfaces for hydrogen - bonded dimers? *The Journal of chemical physics* **1996**, *105* (24), 11024-11031.
89. Galano, A.; Alvarez-Idaboy, J. R., A new approach to counterpoise correction to BSSE. *Journal of computational chemistry* **2006**, *27* (11), 1203-1210.
90. Lu, T.; Chen, F., Multiwfn: a multifunctional wavefunction analyzer. *Journal of computational chemistry* **2012**, *33* (5), 580-592.
91. Johnson, E. R.; Keinan, S.; Mori-Sánchez, P.; Contreras-García, J.; Cohen, A. J.; Yang, W., Revealing noncovalent interactions. *Journal of the American Chemical Society* **2010**, *132* (18), 6498-6506.
92. Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *Journal of molecular graphics* **1996**, *14* (1), 33-38.
93. UniProt: the universal protein knowledgebase in 2021. *Nucleic acids research* **2021**, *49* (D1), D480-D489.
94. Armougom, F.; Moretti, S.; Poirot, O.; Audic, S.; Dumas, P.; Schaeli, B.; Keduas, V.; Notredame, C., Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic acids research* **2006**, *34* (suppl_2), W604-W608.
95. Di Tommaso, P.; Moretti, S.; Xenarios, I.; Orobity, 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 research* **2011**, *39* (suppl_2), W13-W17.
96. Notredame, C., higgins, DJ, heringa, J. T-Coffee: A novel method for multiple sequence alignments. *J. Mol. Biol* **2000**, *302*, 205-217.

97. Zhulidov, P. A.; Bogdanova, E. A.; Shcheglov, A. S.; Vagner, L. L.; Khaspekov, G. L.; Kozhemyako, V. B.; Matz, M. V.; Meleshkevitch, E.; Moroz, L. L.; Lukyanov, S. A., Simple cDNA normalization using kamchatka crab duplex - specific nuclease. *Nucleic acids research* **2004**, *32* (3), e37-e37.
98. Petřek, M.; Otyepka, M.; Banáš, P.; Košinová, P.; Koča, J.; Damborský, J., CAVER: a new tool to explore routes from protein clefts, pockets and cavities. *BMC bioinformatics* **2006**, *7* (1), 1-9.
99. Jurcik, A.; Bednar, D.; Byska, J.; Marques, S. M.; Furmanova, K.; Daniel, L.; Kokkonen, P.; Brezovsky, J.; Strnad, O.; Stourac, J., CAVER Analyst 2.0: analysis and visualization of channels and tunnels in protein structures and molecular dynamics trajectories. *Bioinformatics* **2018**, *34* (20), 3586-3588.
100. Chovancova, E.; Pavelka, A.; Benes, P.; Strnad, O.; Brezovsky, J.; Kozlikova, B.; Gora, A.; Sustr, V.; Klvana, M.; Medek, P., CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. **2012**.
101. Pavelka, A.; Sebestova, E.; Kozlikova, B.; Brezovsky, J.; Sochor, J.; Damborsky, J., CAVER: algorithms for analyzing dynamics of tunnels in macromolecules. *IEEE/ACM transactions on computational biology and bioinformatics* **2015**, *13* (3), 505-517.