Dynamic Charge Distribution as a Key Driver of Catalytic Reactivity in an Artificial Metalloenzyme

David W. Kastner[1,2], Clorice R. Reinhardt[1], Husain Adamji[1], Melissa T. Manetsch,[1] Yuriy Román-Leshkov[1,3], and Heather J. Kulik[1,3]*

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

[2]*Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

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

ABSTRACT: Miniature artificial enzymes such as mimochromes provide a simplified platform to extract design principles for engineering rate enhancements beyond that of natural enzymes, although design optimizations have largely focused on geometric properties, leaving the impact of the electronic environment unexplored. To investigate how the electronic environment influences reactivity, we carry out classical and *ab initio* molecular dynamics (MD) simulations, supervised machine learning (ML), and statistical analysis of a series of mimochromes, MC6, MC6*, and MC6*a. Our classical MD simulations reveal a correlation between increased protein–heme contact and improved reactivity, confirming the importance of geometry, while *ab initio* MD simulations provide insight into the electronic environment, showing the electrostatic potential (ESP) at the metal center also correlates with reactivity. Quantum mechanical calculations of sulfoxidation and hydroxylation reactions demonstrate that the negative ESP at the metal center and active site electric field stabilize the highest-energy intermediate. Furthermore, using ML classifiers, we identify critical residues such as Lys12 and Asp18 in MC6*a that demonstrate charge-coupling patterns that explain differences in reactivity. This suggests that the reactivity series in mimochromes is primarily driven by key aspects of partial charge distribution dynamics, which should guide the engineering of next-generation metalloenzymes.

1

1. Introduction.

Enzymes are versatile catalysts that enable the biosynthesis of complex chemical compounds with high chemo-, regio-, and stereospecificity under mild conditions.[1] The growing need for green catalysts has sparked interest in applying computational approaches and directed evolution to engineer non-native enzymes with enhanced reactivity and the capability to perform novel chemical transformations.[2-7] Heme proteins, which contain a protoporphyrin IX cofactor typically with a coordinated iron cation, are of particular interest as they catalyze a remarkable array of challenging reactions, including hydroxylation,[8] halogenation,[9] peroxidation,[10] sulfoxidation,[11] and C–C bond cleavage[12], and can be tuned by modifying the protein scaffold and the coordinating environment of the reactive high-valent metal center.[13] The tunability of the protein and metal environments has served as inspiration for an array of *de novo* designed enzymes that display equivalent or superior reactivities to their natural counterparts and can carry out reactions not found in nature.[14-17] However, despite successes, the complexity of natural enzymes has made it difficult to extract design principles from structure–function relationships and quantify the core requirements for catalytic activity.[18] To facilitate the investigation of the requirements for the protein environment in enzymatic heme catalysis, several miniaturized heme-based enzyme systems have been developed.[19-24]

Exemplary of an miniaturized metalloenzyme system, the family of *de novo* mini-enzymes known as mimochromes consist of a five-coordinate heme iron consisting of a deuteroporphyrin covalently bound to two nonapeptides with the proximal peptide coordinating iron through an axial histidine (Figure 1).[25] Due to the simplicity of the protein environment, the reactivity of the mimochrome family has been successfully tuned and optimized over several generations through principles of rational design.[26] There are three main catalytically active mimochromes exhibiting

2

increasing reactivity, MC6[27] < MC6*[28] < MC6*a,[29] with the catalytic efficiency of the most reactive mimochrome, MC6*a, exceeding that of natural horseradish peroxidase.[29] In the first modification, MC6 was converted to MC6* through the change of an acidic residue (Glu3) on the lower helix,[28] termed the TD chain, to a hydrophobic residue, which eliminated a salt bridge (Glu3···Arg27) (Figure 1). Eliminating the salt bridge was hypothesized to allow Arg27 to interact with the active site.[30,31] In the second iteration, two polar residues, Glu20 and Ser23, were both changed to a non-canonical amino acid, aminoisobutyric acid (Aib), which was hypothesized to enhance the stability of the upper helix, also termed the D chain (Figure 1).[29] A comparison of nuclear magnetic resonance (NMR) chemical shifts for MC6, MC6*, and MC6*a revealed below-average values for the Aib residues, suggesting increased protein–heme contact between the porphyrin ring and the Aib residues.[29] While experimental evidence has provided valuable insights into each iterative mimochrome improvement in activity or structural shifts, the precise mechanisms by which individual substitutions dictate reactivity differences between MC6, MC6*, and MC6*a are still poorly understood.
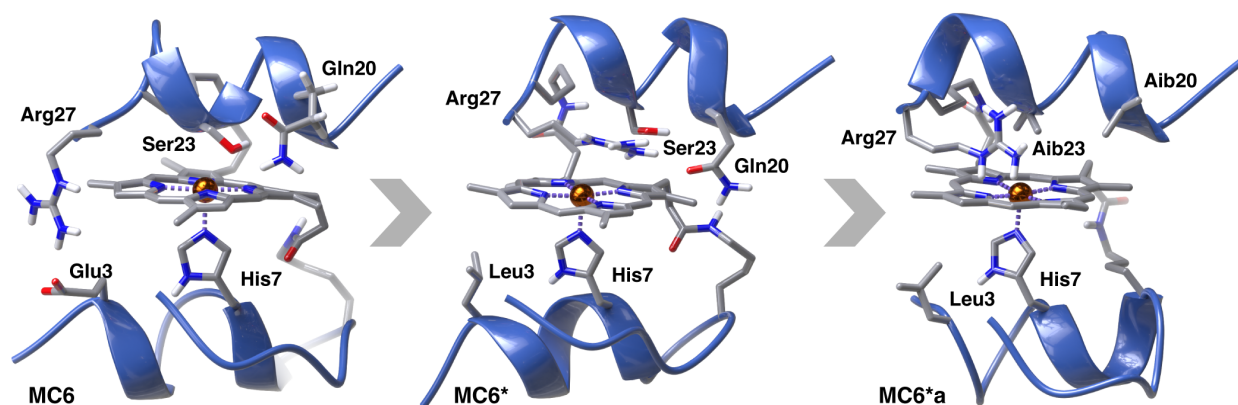

Figure 1. Representative structures of (left) MC6, (middle) MC6*, and (right) MC6*a obtained from the centroids of our MD simulations. Substitution sites and important residues are labeled on the structures. Dark blue cartoon representations are used for the secondary structures and atoms are colored as follows: carbon in gray, nitrogen in blue, oxygen in red, hydrogen in white, and iron in brown. Coordinating bonds are shown as purple dashed lines.

As described above, the current explanations for differences in mimochrome reactivity focus on variations of the protein secondary structure but have not explored potential electronic effects. While geometric and electronic effects are interconnected, even small differences in the electronic environment have shown to have large effects on heme catalysis through changes in the electrostatic potential,[32] electric fields,[32-35] and local dielectric constants.[36] Of particular interest are the effects of dynamic partial charge coupling, which has been shown computationally to be important in small proteins including to distinguish wild type and mutant species[37-39] and provides an electrostatic corollary to geometric allosteric coupling.[40] Unlike structural differences, interrogating differences in the electronic environment is typically difficult to probe experimentally for enzyme systems and often must rely extensively on computational studies.[41-45] While computational studies have indeed been performed on the MC6 variant of these three mimochromes (i.e., the least reactive one),[46] no computational analysis of the changes in reactivity across the mimochrome series has been carried out.

In this work, we perform extensive computational analysis across three iterations of the mimochrome family (i.e., MC6, MC6*, and MC6*a) and evaluate the specific electronic properties and their influence on and correlation with experimentally observed reactivities. To investigate the role of the electronic environment, we carry out extensive classical and *ab initio* molecular dynamics (MD) simulations of the three mimochromes and observe stark differences in their charge dynamics. Based on the *ab initio* molecular dynamics simulations, we evaluate the electrostatic potential (ESP) at the metal center, local electric fields, and patterns of charge coupling that reveal striking differences between mimochromes that correlate with the experimentally observed reactivity trends, while more conventionally applied backbone geometric correlation analyses show few differences. These results are further corroborated by supervised

machine learning (ML) models, which classify the three mimochromes based only on charge-derived features and provide feature importance scores that identify residues that distinguish the behavior of the variants, highlighting opportunities for engineering the electronic environment when designing *de novo* artificial metalloenzymes.

2. Computational Details.

2a. Protein Structure and Preparation.

We used the QM/MM-optimized structure of MC6 from Perrella et al.[46] as a starting point for the our workflow, as no crystal structure has been solved and reported for MC6, MC6*, or MC6*a (Supporting Information Figure S1). To create MC6*, we introduced the substitution Glu3Leu, and we introduced the substitutions Gln20Aib and Ser23Aib on MC6* to create MC6*a using the ChimeraX substitution tool.[47,48] All aspartates and glutamates were assigned a -1 charge and arginines and lysines were given +1 charges (Supporting Information Table S1). The optimized structures had atom counts and net charges of 487 and -1 for MC6, 491 and 0 for MC6*, and 489 and 0 for MC6*a, respectively. The resulting structures were then geometry-optimized with density functional theory (DFT). We employed the long-range corrected, range-separated hybrid ωPBEh functional (ω = 0.2 bohr^{-1})[49] using TeraChem developer version 1.9.[50,51] All calculations employed a basis set consisting of the LANL2DZ effective core potential[52] for Fe and 6-31G* for other atoms.[50,51] We employed an implicit conductor-like polarizable continuum model (C-PCM) with epsilon set to 80[53,54], as implemented in TeraChem.[55,56]

Utilizing the tleap utility in the AMBER software suite, we generated force field topology and coordinate files for each mimochrome. The topology and inpcrd files for all MD simulations are provided in the Zenodo repository.[57] To accurately model the non-standard Aib residues in MC6*a, we used the implicitly polarized Q (IPolQ) scheme force field ff15ipq[58] developed for

5

protein mimetics and the recommended SPC/E$_b$ water model for all classical MD (see Sec. 2b).[59] The core active site parameters for the porphyrin ring, covalently bound lysines, and the metal center both with and without an oxo present were obtained using AMBER's Metal Center Parameter Builder (MCPB).[60] The MCPB.py v3.0 script was employed to compute charges with the ChgModB method at the B3LYP/LACVP*[51,61] level of theory assuming a charge state of 0 and a singlet spin state without the oxo and a charge state of +1 and a doublet spin state with the oxo. The Seminario method[62] was used to derive additional force field parameters (Supporting Information Table S2 and S3). The default MCPB atom naming convention was used (Supporting Information Figures S2 and S3). Each protein system was solvated with a 15 Å SPC/E$_b$ water box with periodic boundary conditions and Na$^+$ counterions.[63] The resulting final atom counts for the three mimochromes were: 8,078 for MC6, 8,054 for MC6*, and 8,325 for MC6*a.

2b. Classical Molecular Dynamics Simulations and Analysis.

We performed MD simulations for MC6, MC6*, and MC6*a using AMBER18 and the GPU-accelerated particle mesh Ewald (PME)[64] molecular dynamics (PMEMD) code.[65,66] The equilibration protocol was as follows: i) 1,000 steps of hydrogen atom minimization, 2,000 steps of sidechain minimization with a fixed backbone, and 2,000 steps of unrestrained protein minimization; ii) controlled NVT heating from 0 to 300 K over 10 ps using the Langevin thermostat and a collision frequency of 5.0 ps^{-1}; and iii) a 1 ns NPT simulation with the Berendsen barostat and a 2 ps relaxation time. Following equilibration, we collected 250 ns of production dynamics for each mimochrome, employing 2 fs time steps, and the SHAKE algorithm.[67] Electrostatics were treated with the PME method with a real-space cutoff of 10 Å.[64] Seven 250 ns independent simulations were generated for each mimochrome system starting from the QM/MM optimized structure from Perrella et al.[46]. Additionally, to obtain representative equilibrated

structures of MC6, MC6*, and MC6*a in the oxo-bound state, we ran seven independent 250 ns molecular dynamics simulations for each mimochrome in the oxo-bound state for use in electric field analysis. A full flowchart of all calculations and analysis carried out in this work is provided in Supporting Information Figure S1. We evaluated the secondary structure of the three mimochromes using the definitions of secondary structure of proteins (DSSP). For the analysis the non-covalent interactions between the porphyrin ring and the protein component, we employed MMPBSA.py as implemented in AMBER18,[68,69] using the generalized Born (GB)[70] approximation and OBC1 model. We extracted snapshots from each MD simulation that were equally spaced by 50 ps and taken from the largest cluster, which was identified with DBSCAN. We calculated the residue-wise contributions to protein–porphyrin interactions and decomposed the interactions into electrostatics and can der Waals contributions.[71,72] The root mean squared fluctuations (RMSFs) were calculated by aligning the seven independent classical MD simulations to the starting structure and then calculating the average per-residue RMSF with MDAnalysis.[73]

2c. Ab Initio Molecular Dynamics Simulations.

Representative frames were selected using the DBSCAN clustering algorithm in CPPTRAJ to obtain clusters and their most representative frames (i.e., cluster centroids).[74] The centroid of the largest cluster from each independent simulation was then used as input to for semi-empirical quantum mechanical (SQM) MD simulations (Supporting Information Tables S4–S6 and Figures S2–S4).[74,75] The SQM MD simulations used GFN2-xTB[76], which incorporates D3 dispersion[77], as implemented in the TeraChem v1.9 developer version.[78,79] We employed the conductor-like polarizable continuum model (C-PCM)[54,80] using a dielectric constant of 80 for water and a cavity derived from 1.2 times Bondi's van der Waals radii[81] for all atoms, as implemented in TeraChem.[55,56] We modeled MC6, MC6*, and MC6*a in the closed-shell singlet (i.e., singlet

7

multiplicity, $2S+1 = 1$) for computational efficiency in sampling protein dynamics, assuming the metal center is Fe(II). For subsequent catalytic cycle modeling, we used revised spin state assignments using Fe(III) as a doublet, as both Fe(II) and Fe(III) are observed experimentally (see Sec. 2d)[82,83]. The simulations were executed within the NVT ensemble, with an initial temperature of 300 K and a timestep of 0.5 fs. We employed the Langevin thermostat with a damping frequency of 1 ps^{-1}. We recorded both charges and structures at every timestep over 20 ps (i.e., 40,000 steps) for each of the 8 independent simulations for each of the three mimochromes. This comparatively brief timescale for the SQM MD versus the classical MD was in line with earlier studies that demonstrated convergence of charge distributions within tens of picoseconds.[38,84] Although explicit solvation would be preferrable, we used an implicit solvent model to avoid complications arising from equilibrating solvent degrees of freedom.

2d. Quantum Mechanical Calculations.

A subset of 400 evenly-spaced structures (i.e., every 0.2 ps) were extracted from the SQM MD runs for evaluation of properties with range-separated hybrid DFT. For each snapshot, DFT partial charges were collected at the ωPBEh/LACVP* level of theory using implicit C-PCM solvation with the Mulliken, Voronoi, Hirshfeld, and ADCH partial charge schemes. To mitigate the limitations of partial charge schemes,[85] we summed the charges over each residue.[38,86,87] We then estimated the ESP at the metal center with each of the partial charge schemes (Supporting Information Figures S13–S15).[45,88] The choice of partial charge scheme did not influence the overall trends, which were in agreement among most methods (Supporting Information Figures S16–S19).[45] For consistency with previous work, we used the Mulliken partial charge scheme for analyses using by-residue summed partial charges,[38,39] and we used the Hirshfeld partial charge scheme for calculations involving the ESP at the metal center (Supporting Information Figure

S1).[45] Next, the charge covariance and mutual information (MI) were computed using sklearn and numpy with the nearest neighbors set to 10 for MI using an in-house script following previous work.[38] We computed the ESP at the metal center following Coulomb's law across various components including all the atoms of the mimochrome, the lower helix, the upper helix, the heme group, and the coordinating histidine. The ESP at the metal center was calculated with an in-house python script, which is provided in the Zenodo repository.[57]

For modeling the reaction mechanisms, the starting structures were selected for MC6, MC6*, and MC6*a from a DFT-computed snapshot of the *ab-initio* MD for each of the mimochromes. This snapshot was chosen from 3200 DFT snapshots (see Sec. 2d) as the single structure from DFT analysis that had the most stabilizing ESP at the metal center. These structures are provided in the Zenodo repository.[57] Due to the size of the systems, full geometry optimizations for the sulfoxidation reaction were first performed on the mimochrome reactant state with HOOH bound (R) with no atoms fixed. For subsequent steps, intermediate structures (IM1, IM2, IM3, and P, see Sec. 3e) were constructed from the optimized HOOH-bound (R) structures, and geometry optimizations were performed with the lower helix and backbone atoms fixed. For the hydroxylation reaction, the same procedure was followed with full geometry optimizations first performed on the mimochrome reactant state with HOOH bound (R') with no atoms fixed. For subsequent steps, intermediate structures (IM1', IM2', IM3', IM4'$_\sigma$, IM4'$_\pi$, and P', see Sec. 3e) were constructed from the optimized HOOH bound (R') structures and geometry optimizations were performed with the lower helix and backbone atoms fixed. The lower helix was frozen as it does not interact with the active site.

Energies for all optimized intermediates were calculated at the ωPBEh/LACVP* level of theory (Supporting Information Tables S9 and S10). The electronic configuration of all

9

intermediates were modeled in a doublet $(2S + 1 = 2)$ spin state, as the ground state was the doublet for all intermediates except for the product (Supporting Information Table S11 and Figure S54). For IM4' of the hydroxylation reaction, both the σ and π pathways were explored (i.e., the intermediate was modeled as a singlet or a triplet), although they were roughly the same in energy (Supporting Information Table S12 and Figure S55). A cluster of four waters was used as the proton acceptor across all steps of the sulfoxidation and hydroxylation reactions, and its reference geometry was identical in all mimochromes (Supporting Information Text S3 and Figure S53). For the calculation of electric field projections, we employed an iterative Hirshfeld approach to compute the partial charge, dipole, and multipole moment of each atom in the extended protein environment, using the software package Multiwfn[89] (Supporting Information Text S2). From the seven trajectories for each mimochrome in the oxo-bound state (see Sec. 2b) used in electric field analysis (see Sec. 3e), we clustered the trajectories using the DBSCAN[74] algorithm to obtain four representative centroids from each trajectory, yielding 28 snapshots of the Fe-oxo state for each mimochrome.

2e. Interpretable Machine Learning Models.

We formulated two distinct temporal feature sets: the first containing residue-summed Mulliken charges (23 features), and the second comprising all pairwise distances (351 features). The features corresponding to substituted amino acids (i.e., those that varied between mimics) were omitted from both sets as including the charges for the mutated residues would lead to perfect assignments. We next employed random forest (RF)[90] and multi-layer perceptron (MLP)[91] multi-class classifier models with a 60-20-20 partitioning strategy for training, testing, and validation sets, respectively. We implemented a temporal dataset split, allocating the initial 60% of each independent simulation for training, the following 20% for validation, and the final 20% for

testing. This approach was chosen over a random split to avoid data leakage, as molecular dynamics simulation points are temporally correlated. A random split would risk providing the model with information from adjacent time points across training, validation, and test sets, thus compromising the independence of the test set and potentially inflating the model's performance metrics. The RF models were built and trained using scikit-learn v1.3.1.[92] Hyperparameter searching was performed using a grid search and 3-fold cross-validation (Supporting Information Table S7). The multi-class MLP classifier model was trained with PyTorch version 2.0,[93] with the ReLU activation function[94] applied between layers, and utilizing the Adam optimizer.[95] Hyperparameter searching for the MLP was determined with Optuna v3.3.0 across 200 trials (Supporting Information Table S8). The evaluation of model performance was conducted on the test set for each data split by examining the area under the receiver operating characteristic (ROC) curve (AUC) and analyzing confusion matrices. To determine the most critical features for the MLP models, we performed SHapley Additive exPlanations (SHAP) analysis.[96,97] For the SHAP interpretation MLP models, we designated 100 data points as the background set and 156 data points for the test set. To determine the most critical features for the RF models, we used Gini feature importance[90] analysis as implemented in scikit-learn.

3. Results and Discussion.

3a. Differences in Classical Dynamics Across Closely Related Mimochromes.

To explore conformational and dynamic differences among MC6, MC6*, and MC6*a, we performed extensive classical molecular dynamics (MD) simulations (see Sec. 2b). The MD simulations revealed structural variation among the mimochromes due to their corresponding amino acid substitutions. DSSP indicated comparable α-helical content among MC6, MC6*, and MC6*a, aligning with previous experimental results (Supporting Information Figure S7).

11

Although some studies suggest that α-helical content in MC6*a increases when solvated with a 50% aqueous solution of trifluoroethanol (TFE), this phenomenon has not been observed experimentally in water.[28,29]

Next, we explored the experimentally suggested "hangman effect" for Arg27 in MC6* and MC6*a.[26] We performed residue-wise root-mean-squared fluctuation (RMSF) calculations across all the MD simulations for each mimochrome and found that the RMSF was higher for the residue Arg27 in MC6* (4.3 Å) and MC6*a (3.9 Å) than in MC6 (3.5 Å) (Supporting Information Figures S8 and S9). The enhanced flexibility of Arg27, as revealed by RMSF calculations, supports its proposed function in facilitating the hangman effect. However, even in MC6, we observed instances of the hangman effect due to frequent loss of the Glu3⋯Arg27 salt bridge during the MD simulations. This observation is corroborated by a previous computational study that identified the Glu3⋯Arg27 interaction as the weakest ion pair in MC6 during those classical MD simulations.[46] In our simulations, the C-terminus exhibited the highest degree of flexibility for all mimochromes, usually appearing disordered (Supporting Information Figures S7–S9).

Next, we investigated the role of non-native amino acids Aib20 and Aib23 in the increased reactivity of MC6*a. Specifically, we quantified whether the D chain (i.e., the upper helix that only interacts non-covalently with the porphyrin) was positioned more centrally above the porphyrin ring in MC6*a as expected.[26,29] The distances between residues 20 and 23 of the D chain and the iron center were measured over all MD simulations for each of the mimochromes (Figure 2). Our distance measurements indicate that, on average, the D chain is situated closer to the iron center for MC6*a (5.5 Å) than MC6* (5.9 Å) and is furthest in MC6 (6.4 Å). The ordering is consistent with experimentally observed reactivity: MC6 < MC6* < MC6*a. We also investigated how the Aib residues and the proximity of the D chain affects the formation of a hydrophobic

binding pocket. We calculated the number of water molecules within 6 Å of iron for all three mimochromes and found that water was better excluded from the active site of MC6*a than for MC6 and MC6* (Supporting Information Figure S10). The exclusion of water from the substrate binding site suggests a secondary function of Aib20 and Aib23 in the formation of a more hydrophobic binding pocket, enhancing the binding affinity for indole substrates commonly used with mimochromes.[98]
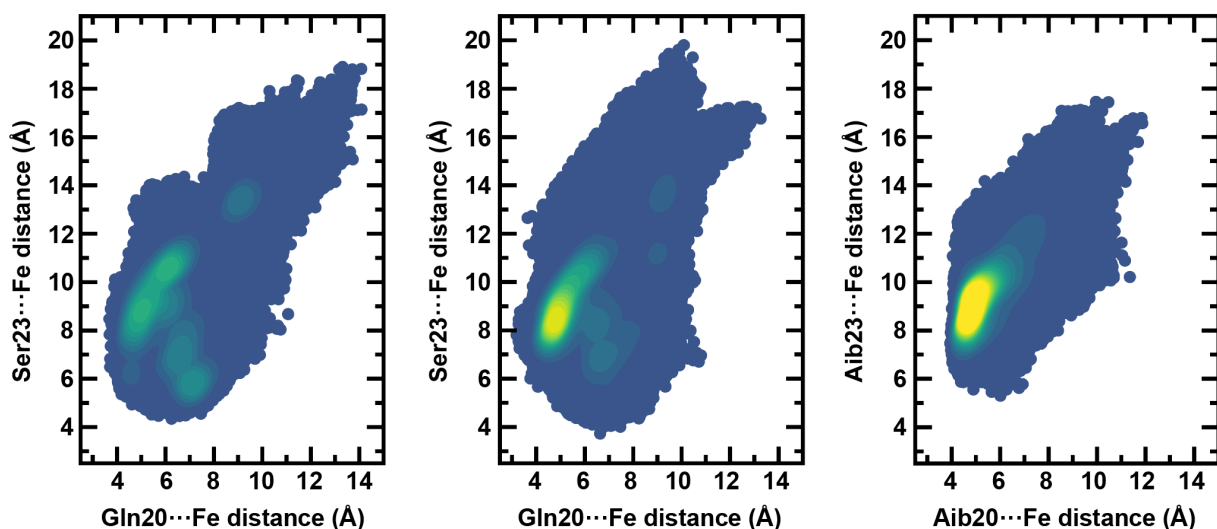

Figure 2. Kernel density estimation (KDE) of the positioning of the D chain for each mimochrome using the distance between the α-carbon of amino acid 20 and Fe (x-axis) and the distance between the α-carbon of amino acid 23 and Fe (y-axis) as reference: (left) MC6, (middle) MC6*, and (right) MC6*a. Dark blue regions represent individual MD frames while a transition towards yellow signifies higher density areas.

To explore the non-covalent interactions between the D chain and the porphyrin ring further, we conducted an energy decomposition analysis using the generalized Born model (GBSA).[68,70] We found significant differences between the observed noncovalent interactions for MC6*a compared to MC6 and MC6*, which were similar to each other. Moreover, in MC6*a, Aib23 demonstrated the most substantial interaction with the porphyrin among the D chain residues (-4.5 kcal/mol) and exhibited more favorable interactions with the heme group relative to Aib20 (-0.6 kcal/mol, Supporting Information Figure S11). Thus, the contributions of Aib20 in

13

positioning the D chain are likely indirect as Aib20–porphyrin contact was infrequent. In MC6 and MC6*, the strongest interaction was between Arg27 and the porphyrin (-9.7 and -7.8 kcal/mol, respectively). While energetically favorable, the interaction between Arg27 and the porphyrin pushes the D chain further from the porphyrin ring as Arg27 was often positioned between the D chain and the porphyrin ring during dynamics (Supporting Information Figures S12 and S13). However, in MC6*a, the hydrophobic interaction of Aib23 prevented the insertion of Arg27 between the D chain and the porphyrin ring, which led to a significantly lower Arg27–porphyrin interaction energy for MC6*a (-1.8 kcal/mol) compared to MC6 and MC6*, as measured with GBSA.

3b. Differences in Metal Electrostatic Environments Across Mimochromes.

After identifying specific variations in the classical dynamics of MC6, MC6*, and MC6*a, we investigated how the identified geometric changes correspond to differences in the electrostatic environment at the active site. Given the variation in D chain–porphyrin distances, we focused on the ESP at the metal center (see Sec. 2d). To determine the components of the mimochrome that had the largest influence on the ESP, we calculated the contribution to the ESP at the metal center caused by each structural component: the TD chain, the D chain, the TD chain without the coordinating histidine, the porphyrin ring, and the coordinating histidine alone. To capture geometric variations, the ESP at the metal center was calculated for 3200 DFT-computed snapshots derived from the 8 SQM MD simulations for each mimochrome (see calculation details in Supporting Information Figure S1). From our analysis, we observed that the experimentally determined reactivity, MC6 < MC6* < MC6*a,[26] aligns with increasingly negative values for the contribution from the D chain and increasingly positive values for the TD chain. That is, the most reactive mimochrome, MC6*a, had the most negative ESP for the contribution from the D chain

14

and the most positive ESP for the TD chain, followed by MC6* and then MC6 (Figure 3). Furthermore, the average ESP contribution from the D chain decreases from MC6 (78 kJ mol^{-1} e^{-1}) to MC6* (49 kJ mol^{-1} e^{-1}) to MC6*a where it becomes negative (-86 kJ mol^{-1} e^{-1}). This negative ESP for the most reactive mimochrome, MC6*a, suggests an electrostatic stabilization for the positively charged metal center that could be expected to significantly alter the reactivity.
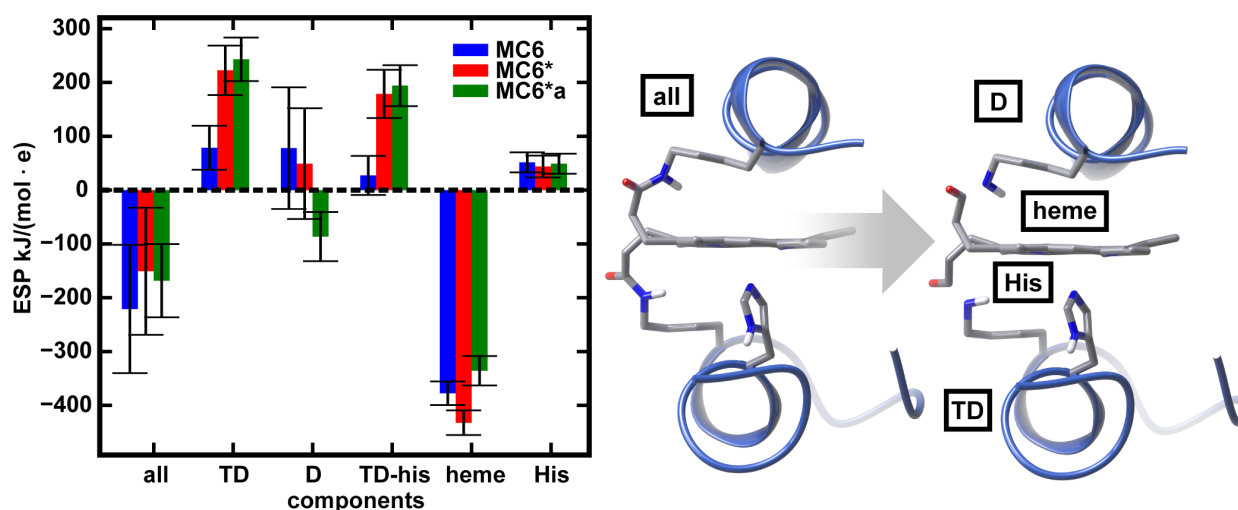

Figure 3. (left) The value of the ESP at the metal center estimated using the Hirshfeld charges for various structural components of MC6 (blue), MC6* (red), and MC6*a (green). The ESP was computed for the entire mimochrome (all), the TD chain (TD), and D chain (D), the TD chain without histidine (TD-his), the porphyrin ring (heme), and histidine (his) alone. The standard deviations over snapshots are shown on the plots with error bars. (right) The structures of the main components: all, TD, D, heme, and his. TD-his is the TD chain without histidine. Dark blue cartoon representations are used for the secondary structures and atoms are colored as follows: carbon in gray, nitrogen in blue, oxygen in red, and hydrogen in white.

We next investigated if specific D-chain residues influence the calculated trends in the ESP at the metal center by examining the correlation between the distances from the center of mass of amino acids to the iron atom and the variations in ESP. Two charged amino acids played a significant role in ESP variations: Asp18 and Arg27. The correlation between the Asp18···Fe distance and the ESP of the D chain is pronounced, especially in MC6 and MC6*, with shorter Asp18···Fe distances resulting in lower ESP values (Figure 4 and Supporting Information Figure

15

S21). The average Asp18···Fe distance was shorter for MC6*a (12.8 Å) and MC6* (12.6 Å) than for MC6 (13.7 Å). Conversely, the relationship between the average Arg27···Fe distance and the D chain ESP shows increasingly longer distances for MC6 (6.7 Å), MC6* (7.6 Å), and MC6*a (8.3 Å) in line with an increasingly negative ESP (Supporting Information Figure S22). In addition to its proposed role in the hangman effect, the ability of Arg27 to alter the ESP at the metal center suggests an additional role in tuning the electrostatic environment. The introduction of the residue Aib23 in MC6*a likely contributes to the longer Arg27···Fe distances due to favorable hydrophobic interactions (-4.6 kcal/mol from GBSA) with the porphyrin. The increased hydrophobic interactions prevent the charged Arg27 residue from coming closer to the active site by decreasing the distance between the D chain and the iron center. We then checked if the distance between the D-chain itself, measured at its centroid, and the iron center also correlated with a more negative ESP. We found that the distance between the D-chain centroid and the metal center increased consistently from MC6*a (6.6 Å), to MC6* (7.4 Å) to MC6 (7.8 Å), further demonstrating the relationship between the positioning of the D chain and the ESP at the metal center (Supporting Information Figure S23).
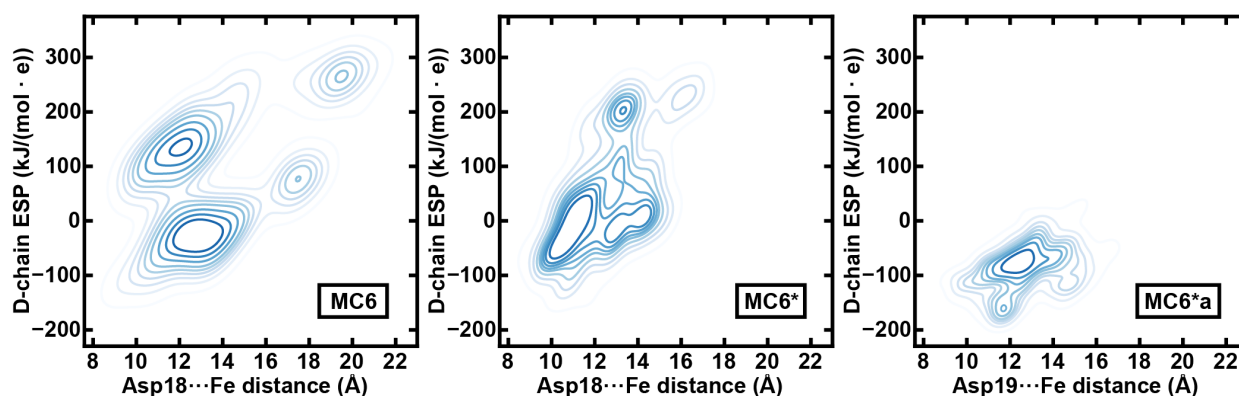

Figure 4. Kernel density estimation (KDE) plots showing the correlation between the distance between the centroid of Asp18 and the iron center and the ESP at the metal center of the D chain for the three mimochromes: (left) MC6, (middle) MC6*, and (right) MC6*a. Dark blue lines indicate regions densely populated with points while a transition towards white signifies regions more sparsely populated with points.

16

3c. Charge-derived features distinguish mimochromes.

Following our identification of the ESP at the metal center of the D and TD chains as distinguishing features across mimochromes, we sought to identify patterns in charge distribution at the individual residue level. While we expect differences in charges for the three substituted residues that distinguish MC6, MC6*, and MC6*a, we investigated if any of the 27 shared residues exhibit differences in charge distribution. To aid in the identification of qualitative distinguishing differences in charge distribution between the mimochromes, we trained supervised machine learning (ML) classifier models (i.e., RF and MLP, see Sec. 2) to distinguish patterns in charge differences among the mimochromes from their *ab initio* MD trajectories.[99] The models were trained as multi-class classifiers to distinguish *ab initio* MD frames as belonging to MC6, MC6*, or MC6*a based on by-residue summed Mulliken partial charges of the shared residues. These models showed good performance with AUC scores of 0.92–0.96 for the model RF and 0.88–0.92 for the MLP model (Figure 5 and Supporting Information Figures S24 and S25). The model performance could be improved if we also included features related to the ESP of each mimochrome because the ESP strongly distinguishes the three mimochromes (Supporting Information Text S1 and Figures S26–S28). From feature importance analysis on the RF model, we found Asp18 to have the highest feature importance as measured by Gini impurity (Figure 6). We also found Arg27, Lys12, and Glu19 to have high feature importance scores. The ranking of Arg27 as the feature with the second-highest feature importance is of interest given its involvement in the hangman effect and its suggested importance in the catalytic improvement between MC6 and MC6*. Lys12 is one of the charged amino acids on the TD chain and is involved in several non-covalent interactions. Finally, Glu19 is believed to be key in maintaining the overall structure

17

of the mimochromes through the Arg11···Glu19 interchain salt bridge. Conversely, residues Ser8 and Ile13 showed the lowest feature importance scores and their charges likely do not contribute to the increased reactivity observed across the MC6, MC6*, MC6*a series, despite the proximity in sequence space of Ile13 to the key Lys12 residue. Residues with low feature importance scores such as Ser8 and Ile13 could be potential targets for future mutagenesis experiments to further optimize the electrostatic environment for reactivity.
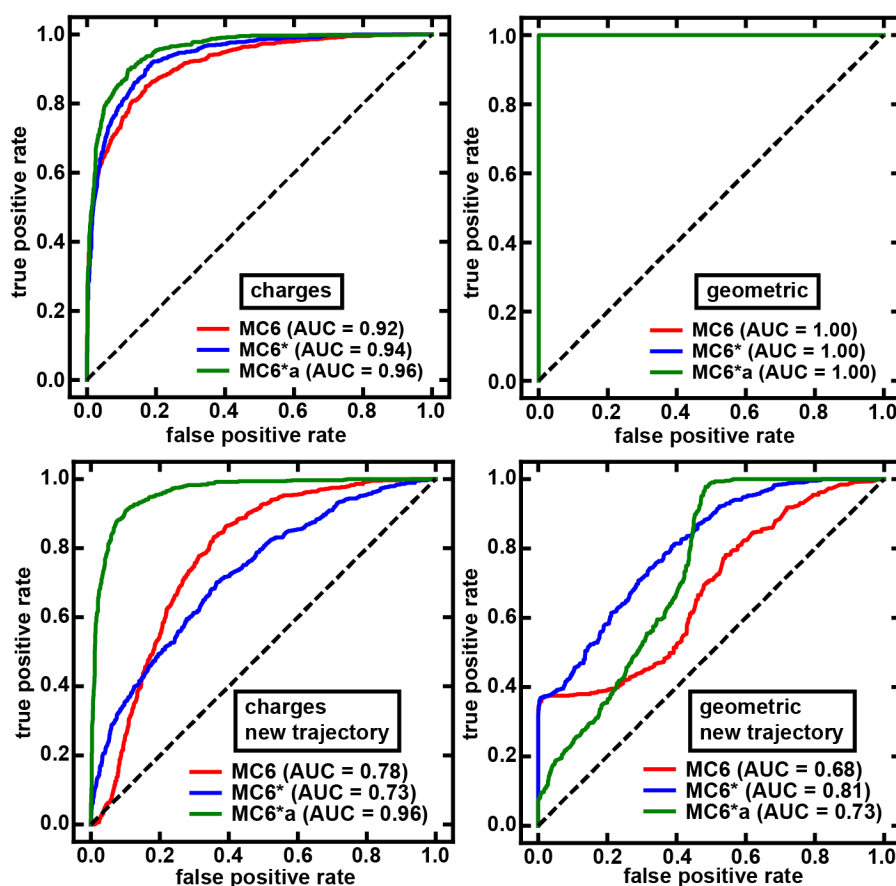

Figure 5. Receiver operating characteristic (ROC) curve for RF models for classification of individual snapshots as belonging to trajectories of one of three mimochromes. The models were trained on (top, left) by-residue summed charges of residues shared by the three mini-proteins, (top, right) geometric pair-wise distance features, (bottom, left) by-residue summed charges on an unseen test trajectory, and (bottom, right) pairwise distances for all shared residues on an unseen test trajectory. ROC-AUC curves are shown for MC6 (red), MC6* (blue), and MC6*a (green). The area under the curve (AUC) values are indicated in the legend. A diagonal dotted line represents the line of no-discrimination.
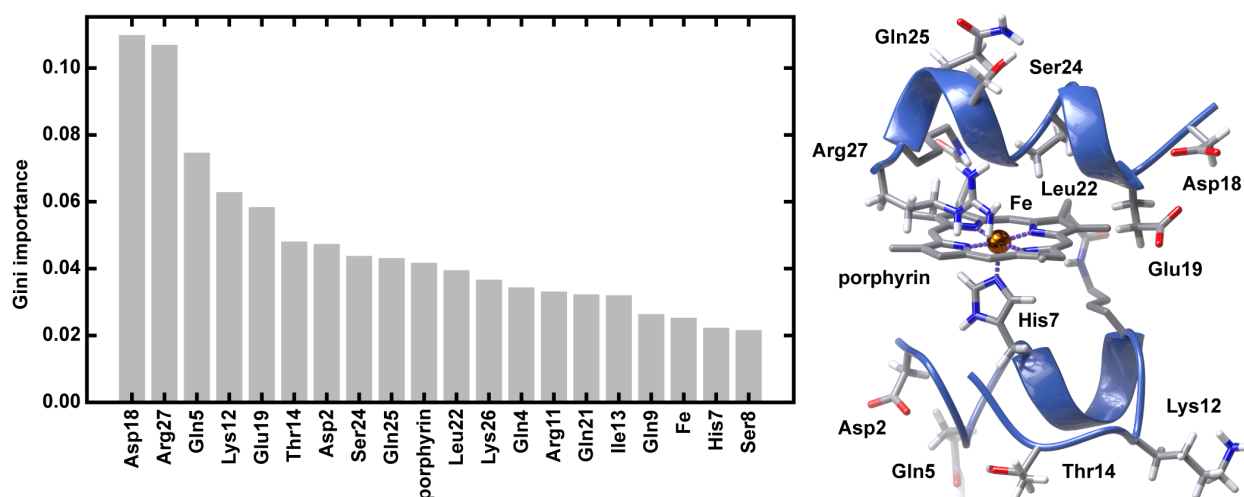
18

Figure 6. Ranked Gini feature importance for the RF model trained on charge-derived features. (left) The RF model was trained on by-residue summed charges, indicated with their three-letter code and residue index. (right) Representative structure of MC6*a with the residues with high feature importance scores shown with stick representation. Dark blue cartoon representations are used for the secondary structures and atoms are colored as follows: carbon in gray, nitrogen in blue, oxygen in red, hydrogen in white, and iron in brown. Metal coordinating bonds are shown as purple dashed lines.

Given the high computational cost of obtaining key charge-derived features, we repeated our investigation to ascertain if similar patterns are observable in the geometric characteristics of the mimochromes that could have been obtained from a purely classical MD trajectory. To test whether geometric features could distinguish MC6, MC6*, and MC6*a, we trained RF and MLP models on centroid-to-centroid pairwise distances of all residues shared between the mimochromes from the same DFT-computed snapshots of the *ab initio* MD trajectories (351 features). We found that the models trained on geometric features were also able to almost perfectly classify the mimochromes (i.e., test set ROC-AUC scores of 1.0 for all mimochromes, Figure 5 and Supporting Information Figures S29 and S30). The pairwise distances with the highest feature importance based on the Gini score of the RF model were Lys26–Arg27 and Gln4–Arg27 (Supporting Information Figure S31). However, to test if the models could generalize to unseen trajectories, we retrained the models with a new train-validation-test split where we used six trajectories for training, one trajectory for validation, and one trajectory for testing. Repeating this same approach

19

for the charge-derived features, we found that the models trained on the geometric features showed a notable decrease in AUC (0.68–0.81, Figure 5 and Supporting Information Figure S32 and S33). Conversely, charge-derived features were able to more accurately classify the mimochromes even for an unseen trajectory (AUC: 0.73–0.96), corresponding to a much smaller drop in performance, and this performance was even higher (AUC: 0.98–1.00) when ESP features were included (Figure 5 and Supporting Information Figures S34–S37). The fact that the models trained on by-residue summed partial charges were able to generalize to not just randomly held out data but also to held out, unseen trajectories with only 23 features highlights the importance of charge distribution in understanding mimochrome systems.

3d. Charge-Coupling Interactions Distinguish Mimochromes.

In our preceding investigation of the ESP at the metal center of MC6, MC6*, and MC6*a, we identified specific amino acid residues that were important in modulating the electrostatic environment. Next, we turned our focus to charge-coupling interactions between residue pairs that we could expect to collectively alter the electronic environment around the heme cofactor. To quantify charge-coupling interactions, we utilized linear cross-correlation[100] and mutual information[40,101,102] analyses from information theory of the by-residue-summed Mulliken partial charges on each residue, as introduced previously for charge-coupling in proteins.[38,39] Cross correlation captures the linear dependence between the charges in residue pairs, while mutual information captures both linear and nonlinear relationships through the statistical dependence of charges between residues. We computed cross correlation and mutual information scores for all DFT-computed snapshots on SQM trajectories (see Sec. 2). The cross-correlation and mutual information heatmaps reveal distinct patterns of charge-coupling for the MC6, MC6*, and MC6*a mini-proteins (Figure 7 and Supporting Information Figure S38). While we observe a significant

20

number of charge-coupling interactions, including both short- and long-range interactions, there is a notable increase in charge coupling for MC6*a near the Aib substitution sites on the D chain (Figure 7). Specifically, we observed that the Aib mutations lead to a change in the charge coupling of Glu19 such that its charge coupling pattern resembles that of Asp18, which is not seen in MC6 and MC6* (Figure 7).
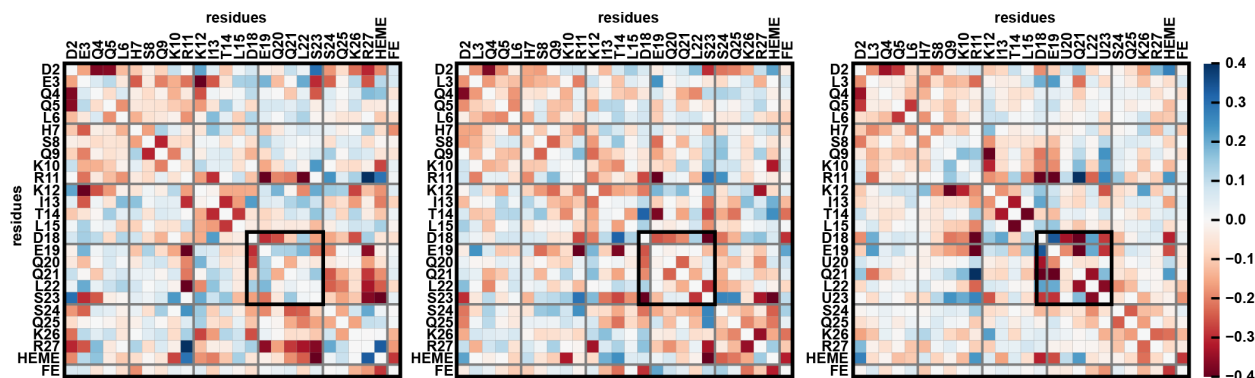

Figure 7. Charge cross-correlation heatmaps of charge dynamics from SQM simulations. Cross-correlation plots for (left) MC6, (right) MC6*, and (right) MC6*a. Each amino acid pair is indicated by its one-letter code and is colored according to the color bar on right. Aib residues are indicated with the letter U. Amino acids are ordered based on residue number. The charge coupling hotspot residues are marked with a black box. The trivial case of same-residue correlations on the diagonal have been omitted. The N-terminus acetyl and the C-terminus amide caps are not shown as no notable charge coupling interactions were observed for these caps.

We find strong charge coupling between residue 20 and Asp18, where residue 20 is an Aib residue in MC6*a, and a Gln in MC6 and MC6* (Figure 8). The Asp18···Aib20 charge-coupling interaction is increasingly evident as the substitution to the nonpolar hydrophobic Aib residue leads to increased charge coupling compared to the polar Gln residue. The trend in charge coupling strength between Asp18 and residue 20 aligns with the experimental reactivity trend: MC6*a > MC6* > MC6. In MC6*a, we observe a clear inverse correlation in the charge fluctuations of Asp18 and Aib20 as they have dynamic rearrangements in their charge densities (Figures 8 and 9). In addition to the correlation with Aib20, Asp18 also shows increased charge coupling with Gln21 in MC6*a that correlates with reactivity (Supporting Information Figure S39). For this

21

residue pair, the cross-correlation scores are -0.27 for MC6, -0.37 for MC6*, and -0.55 for MC6*a, respectively. This close relationship of the density distributions on these two residues in MC6*a is also evident when comparing the individual trajectory in charge density for the two residues (Supporting Information Figure S40).
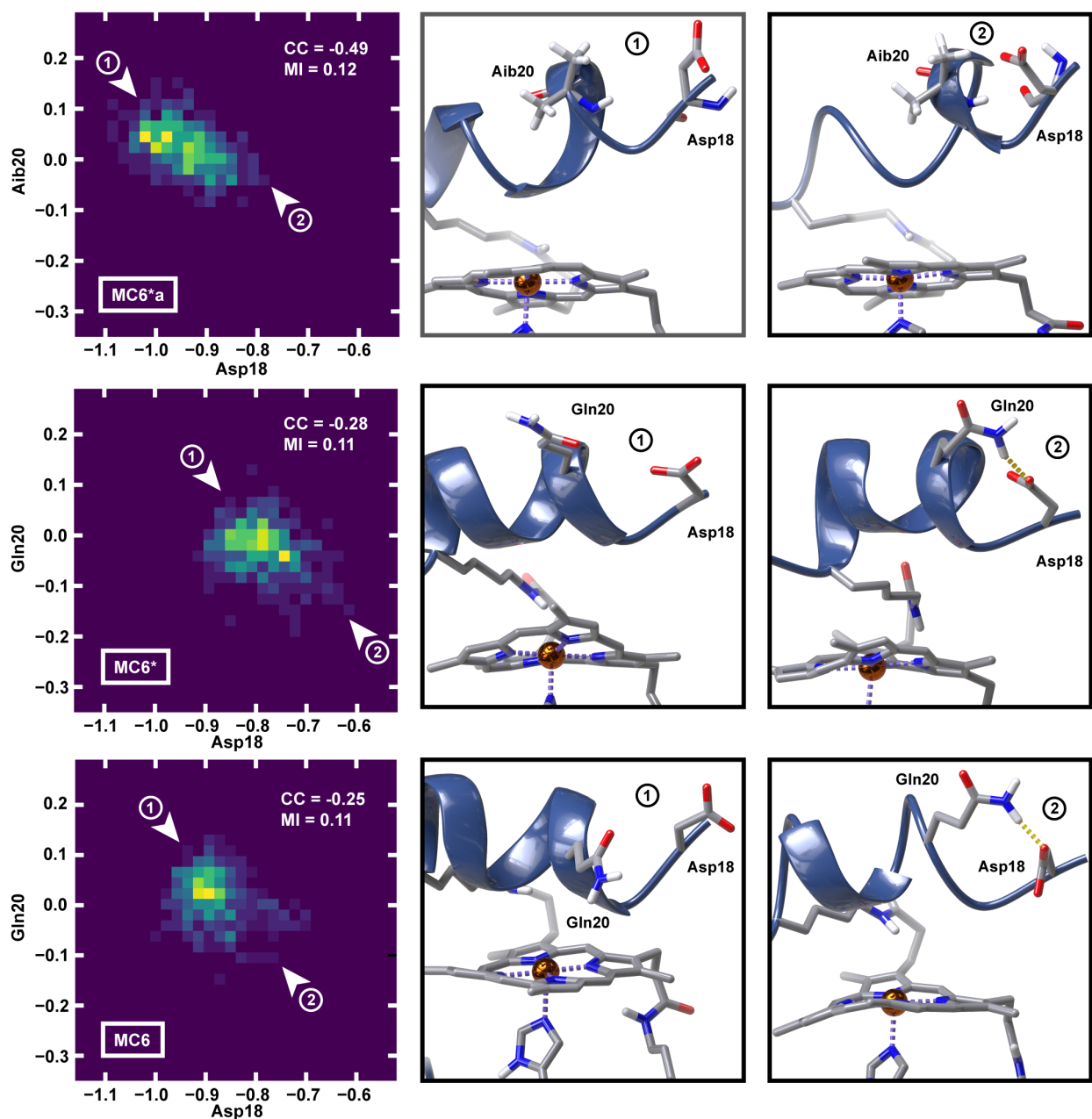

Figure 8. Two-dimensional distributions of charges showing the charge-coupling interaction between Asp18 and residue 20 with representative structures from the extremes of the distribution. Residue 20 is an Aib in MC6*a and a Gln in MC6 and MC6*. (left) Charge coupling distribution

22

for Asp18 and residue 20. (middle) Representative structure for Asp18 with a relatively negative charge. (right) Representative structure for Asp18 with a relatively positive charge. Charge coupling plots and structures are shown for (top) MC6*a, (middle) MC6*, and (bottom) MC6. The cross-correlation (CC) and mutual information (MI) scores are shown in inset. Regions with high populations are indicated with yellow and regions with low populations are indicated with purple in the distributions. Dark blue cartoon representations are used for the secondary structures with purple dashed lines for the coordinating bonds. The atoms are colored as follows: carbon in gray, nitrogen in blue, oxygen in red, hydrogen in white, and iron in brown.
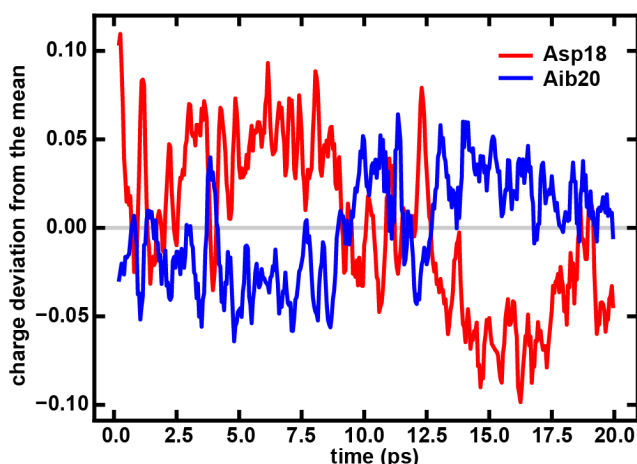

Figure 9. Time-dependent charge deviation from the mean for Asp18 (red) and Aib20 (blue) over time for MC6*a. The time-dependent charge deviation was calculated for 400 DFT-computed snapshots spaced 0.2 ps apart for MC6*a. The depicted snapshots were taken from SQM MD trajectory 3. The y-axis represents the deviation of the charge from its average value, while the x-axis indicates the elapsed time in picoseconds (ps).

In addition to the charge-coupling interactions near the substitution sites on the D chain, we also observed strong coupling interactions on the TD chain. One particularly strong coupling interaction that also distinguishes the mimochromes in alignment with the experimental reactivity trend is between Gln9 and Lys12. The Gln9···Lys12 cross correlation description of the coupling interaction is strongest for MC6*a (-0.67), weaker for MC6* (-0.50), and weakest for MC6 (-0.11, Supporting Information Figure S41). Analysis of Gln9···Lys12 coupling across the simulations shows a consistent transfer of charge from Lys12 to Gln9 (Supporting Information Figure S42). The increase in charge coupling for MC6*a and MC6* is likely due to the Glu3Leu mutation, which was introduced to MC6 to increase inter-chain flexibility by breaking the Glu3···Arg27 salt

bridge. While most charge-coupling interactions were localized to the same chain, we also identified Arg11···Glu19 as an instance of an intra-chain charge-coupling spanning the D and TD chains. The Arg11···Glu19 coupling interaction is not unique to MC6*a, and robust coupling is seen in all mimochromes and across the trajectories, as indicated by their cross-correlation scores (MC6: -0.56, MC6*: -0.60, and MC6*a: -0.56, Supporting Information Figures S43 and S44). The observation of a strong salt bridge interaction such as Arg11···Glu19 is not surprising and could have likely been inferred from geometric analysis. Conversely, the other correlations of Glu19 are substantially changed in MC6*a and mirror those of Asp18.

Next, we examine charge-coupling interactions closer to metal center. More coupling is evident between amino acids and the porphyrin ring than between the amino acids and iron. In MC6 and MC6*, we observe strong charge coupling between Ser23 and the porphyrin ring indicated by high cross-correlation scores of -0.46 and -0.58, respectively. However, there is no evidence of charge coupling between Aib23 and the porphyrin in MC6*a, with a much lower cross-correlation score of -0.14 (Supporting Information Figure S45). In the classical and *ab initio* MD of MC6 and MC6*, Ser23 is most often positioned directly above iron and is often found occupying the open coordination site. This interaction leads to very clear time-dependent coupling where Ser loses charge (i.e., becomes significantly more positively charged), and the porphyrin accumulates charge (i.e., becomes significantly more negatively charged) upon coordination (Supporting Information Figure S46). The Ser23Aib substitution that results in the increased reactivity of MC6*a likely improves reactivity by maintaining an open coordination site. We also observe a charge-coupling interaction between Ser23 and Arg27 for MC6 and MC6* with cross-correlation scores of -0.43 and -0.39, respectively. However, we do not observe charge coupling between Aib23 and Arg27 with a cross-correlation score of -0.17 in MC6*a (Supporting

24

Information Figures S47 and S48). The lack of charge coupling between Arg27 and active site residues in MC6*a is in line with our results indicating a more negative ESP at the metal center for MC6*a.

The electrostatic analysis and charge-coupling depict significant contributions to electronic structure differences in the mimochrome variants. We were interested in whether geometric backbone motions (i.e., of alpha carbon, C_α, motions) typically used in protein analysis were equally revealing. The resulting C_α correlation heatmaps for MC6, MC6*, and MC6*a show a strikingly different picture from the electronic analysis, depicting highly similar patterns for all three mimochromes and failing to capture the same hotspots as charge coupling relevant to Aib substitutions (Supporting Information Text S2 and Figure S49).

3e. Reactivity and Mechanistic Insights.

Finally, we investigated reaction energetics to quantify reactivity differences between MC6, MC6*, and MC6*a for two closely related reaction mechanisms. We explored the experimentally studied mimochrome-catalyzed sulfoxidation reaction,[26] which converts thioanisole to methyl phenyl sulfoxide, and the hydroxylation of propane, which has not been observed, but which the enzymes may be capable of on the basis of heme chemistry (Supporting Information Figures S50–S52). Based on our analysis of the ESP at the metal center, we expected that a more negative ESP for the D chain correlated with experimentally observed reactivity trends across the mimochrome series. Therefore, we selected the structure with the most negative metal-center ESP for the upper D chain from the DFT-computed snapshots of the *ab-initio* MD for each of the mimochromes, and generated QM cluster models for each intermediate (Supporting Information Figures S53–S55).

25

In the proposed sulfoxidation mechanism,[103] the proximal oxygen of the hydrogen peroxide bound, Fe(III)–HOOH, species (R) is deprotonated and stabilized by a water cluster, forming what is referred to in the heme community as the Cpd 0 intermediate, Fe(III)–OOH (IM1). We assume the deprotonation is performed by a cluster of water molecules as mimochromes, especially MC6*a, lack a dedicated proton acceptor residue (see Sec. 2). The proton is then shuttled to the distal oxygen of IM1 to form the Fe(III)–OOH$_2^+$ intermediate (IM3).[103,104] O–O cleavage then leads to the generation of what is referred to in the heme community as Cpd 1, Fe(IV)=O (IM3),[98] and the release of water. IM3 then oxidizes thioanisole releasing methyl phenyl sulfoxide and returning iron to the Fe(III) state (P) (Supporting Information Figures S50 and S51).[105] The mechanism for our proposed hydroxylation reaction follows the same pathway for the generation of the reactive Fe(IV)=O intermediate, proceeding through Fe(III)–HOOH (R'), Fe(III)–OOH (IM1'), Fe(III)–OOH$_2^+$ (IM2'), to generate Fe(IV)=O (IM3'), which abstracts a hydrogen from the substrate followed by a radical rebound, yielding the hydroxylated product (Supporting Information Figure S52).[8]

The computed relative energies for the intermediates of the sulfoxidation and hydroxylation reactions reveal intermediate IM2 to be the highest in energy for MC6 (16.6 kcal/mol), MC6* (16.0 kcal/mol), and MC6*a (12.4 kcal/mol), and the decreasing energy across the series correlates with the experimentally observed reactivity trend, with MC6*a being the lowest in energy (Supporting Information Tables S9 and S10). This stabilization may be a result of the more negative ESP at the metal center for the upper D chain for MC6*a, which stabilizes the partial positive charge on the IM2 intermediate. We observe the most stabilizing effect for MC6*a, which was the only mimochrome with a negative ESP at the metal center for the upper D chain (Figure 3). In both reactions, we found the overall reaction favorability to also align with

known reactivity trends. For the sulfoxidation reaction, we found the reaction thermodynamics to be increasingly favorable from MC6 to MC6* to MC6*a by around 3.6 kcal/mol (i.e., MC6: -17.2 kcal/mol, MC6*: -19.2 kcal/mol, MC6*a: -20.8 kcal/mol, Figure 10). For the hydroxylation reaction that shares many of the same intermediates, we observe a similar trend (MC6: -40.3 kcal/mol, MC6*: -42.3 kcal/mol, MC6*a: -43.9 kcal/mol, Supporting Information Figure S59). The optimized geometries of all intermediates for the hydroxylation and sulfoxidation reactions are provided in theZenodo repository.[57]
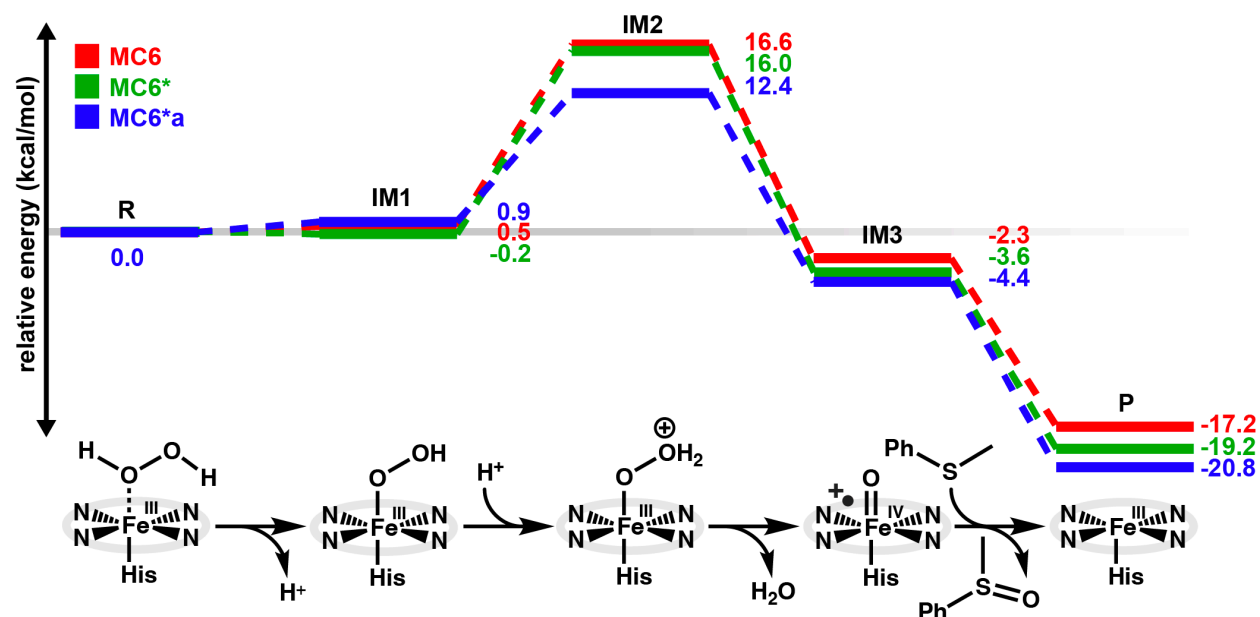

Figure 10. Relative energy profiles in kcal/mol for the sulfoxidation of thioanisole by MC6 (red), MC6* (green), and MC6*a (blue). The skeleton structures of the intermediates are depicted below the reaction profile.

Building on our analysis of the ESP at the metal center, where we observed a negative ESP at the metal center for the D chain and a positive ESP for the TD chain, we further explored the impact of their contributions to the electric field in the active site. Specifically, we focused our investigation on the electric field projected across the Fe-oxo bond, as its strength has been correlated with peroxidase activity in heme enzymes and has been shown to similarly influence

IM2 and IM3.[32,35] The projected electric field across the Fe-oxo bond was calculated for each structure from the sum of electric field contributions from each computed atomic partial charges, atomic dipole, and atomic multipole moment in the extended protein environment (Supporting Information Text S2). The results reveal a moderate increase in the average strength of the projected electric field along the Fe-oxo bond that correlates with reactivity: MC6 (-64.2 MV/cm) < MC6* (-75.0 MV/cm) < MC6*a (-79.8 MV/cm) (Figure 11). We observed the biggest difference in the projected electric field between the least reactive mimochrome MC6 and the two more reactive mimochromes, MC6* and MC6*a, whereas the difference between MC6* and MC6*a was more modest. Upon inspecting the structures with the strongest and weakest projected electric fields, a stronger electric field was observed when the D chain formed the expected α-helical fold and a weaker electric field when it was more disordered (Figure 11 and Supporting Information Figure S60). When the D chain was more structured (i.e., had greater helical content), two negatively charged residues, Asp18 and Glu19 (i.e., charges with strong coupling with a number of residues identified previously), were positioned closer to the active site, leading to a more negative ESP at the metal center on the upper D chain (see Figures 3 and 4). Asp18 and Glu19 also interact with Arg11 on the lower TD chain and position it closer to the active site, further influencing the electric field (Figure 11). The increased electric field strength in MC6* and MC6*a may contribute to their increased reactivity, suggesting a potential avenue to further improve mimochrome reactivity in future generations.
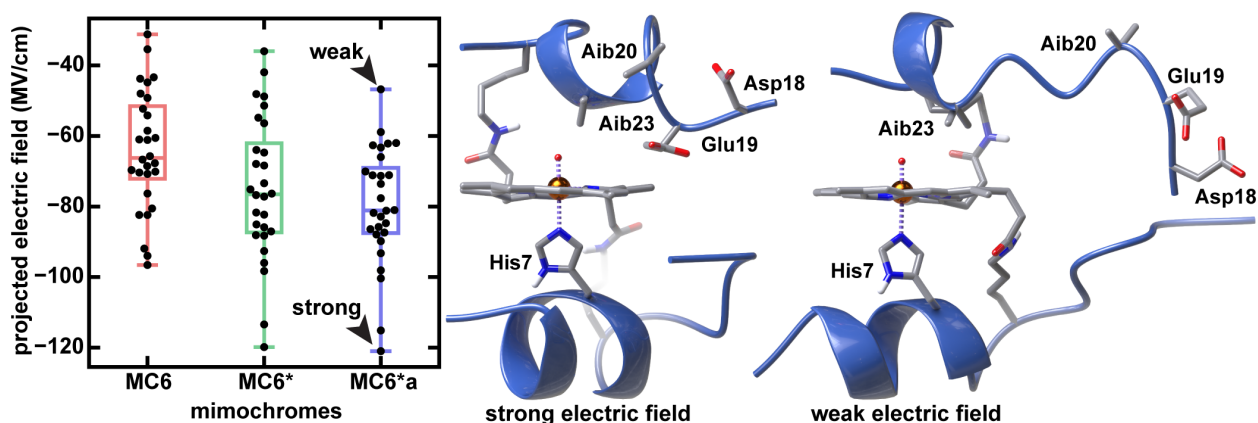
Figure 11. Calculated projected electric fields in MV/cm for MC6 (red), MC6* (green), and MC6*a (blue). The MC6*a structure with the strongest electric field (middle) and weakest electric fields (right) are depicted. Asp18, Glu19, and all Aib residues are labeled on the structures. Dark blue cartoon representations are used for the secondary structures and atoms are colored as follows: carbon in gray, nitrogen in blue, oxygen in red, hydrogen in white, and iron in brown. Coordinating bonds are shown as purple dashed lines.

4. Conclusions.

We performed extensive computational analysis of three iterations of the mimochrome family, MC6, MC6*, and MC6*a, to investigate the specific electronic properties that influence and correlate with the experimentally observed reactivities. Through classical and *ab initio* MD simulations, supervised ML models, and data-driven methods, we uncovered stark differences in the dynamic fluctuation of the electron density across the mimochromes. Our investigation revealed that the ESP at the metal center, local electric fields, and patterns of partial charge coupling exhibited differences among the mimochromes, correlating with known reactivity trends.

We observed significant shifts in the electrostatic environment from MC6 to MC6* to MC6*a: the ESP contributions at the metal center from the upper D chain were increasingly negative (i.e., stabilizing), whereas the ESP contribution from the lower TD chain were increasingly positive (i.e., less stabilizing). Using cross-correlation and mutual information analysis to analyze our *ab initio* MD simulations, we identified residues involved in key charge-coupling interactions, such as Asp18 and Arg27. These observations were corroborated by ML

29

model training and feature importance analysis. While we observed significant differences in charge coupling between the mimochromes, we did not observe notable differences in C_α motions between mimochromes, highlighting the importance of charge-derived features in understanding mimochrome systems and confirming observations from analysis of the ML model. QM calculations further revealed the intermediate between Cpd 0 and Cpd 1 to be the highest in energy, which is stabilized by the more negative ESP of the D chain in MC6*a. In line with reactivity trends, we observed increasingly strong electric fields projected across the Fe(IV)=O bond from MC6 to MC6* to MC6*a, which favor the oxo formation steps and the sulfoxidation reaction overall.

Our comprehensive analysis of dynamic electronic properties in mimochromes underscores the importance of considering the electronic environment when designing *de novo* artificial metalloenzymes. Our analysis also suggests opportunities for further improvement of mimochromes. Future generations of mimochromes could be improved by further tuning the ESP at the metal center and the local electric fields. While key charge-coupling interactions should not be disrupted, such as those involving Lys12, Asp18, and Glu19, residues with less relevance in discriminating mimochromes, such as Ser8 and Ile13, could be targeted through mutagenesis to further enhance the stabilizing ESP, electric fields, and charge coupling with an emphasis on stabilization of the formation of the intermediate between Cpd 0 and Cpd 1. The insights gained from this study advance our understanding of the structure–function relationships in mimochromes and reveal key design principles that can be used for engineering artificial enzymes with rationally-tuned charge dynamics and enhanced catalytic efficiencies.

AUTHOR INFORMATION

Corresponding Author

*email:hjkulik@mit.edu

Notes

The authors declare no competing financial interest.

Supporting Information.

Additional details for the MC6, MC6*, and MC6*a mimochrome systems, including residue numbering convention, protonation states, and parameters; clustering statistics for MD trajectories and centroids; DSSP, RMSF, water shell, GBSA analyses; electrostatic potential residue correlation plots; random forest and multi-layer perceptron model hyperparameters; model performance and feature importance scores; mutual information heatmaps; charge coupling heatmaps and time-dependent charge deviation plots; sulfoxidation and hydroxylation reaction details and energies; electric field calculation details and structural comparisons (PDF).

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

REFERENCES

(1)   Winkler, C. K.; Schrittwieser, J. H.; Kroutil, W. Power of Biocatalysis for Organic
        Synthesis. *ACS Cent. Sci.* **2021,** *7*, 55-71.
(2)   Kiss, G.; Çelebi-Ölçüm, N.; Moretti, R.; Baker, D.; Houk, K. N. Computational Enzyme
        Design. *Angew. Chem. Int. Ed.* **2013,** *52*, 5700-5725.
(3)   Huang, P.-S.; Boyken, S. E.; Baker, D. The Coming of Age of De Novo Protein Design.
        Nature **2016,** *537*, 320-327.
(4)   Miller, D. C.; Athavale, S. V.; Arnold, F. H. Combining Chemistry and Protein
        Engineering for New-to-Nature Biocatalysis. *Nat. Synth.* **2022,** *1*, 18-23.
(5)   Hassan, I. S.; Fuller, J. T.; Dippon, Vanessa N.; Ta, A. N.; Danneman, M. W.;
        McNaughton, B. R.; Alexandrova, A. N.; Rovis, T. Tuning through-Space Interactions
        Via the Secondary Coordination Sphere of an Artificial Metalloenzyme Leads to
        Enhanced Rh(Iii)-Catalysis. *Chem. Sci.* **2022,** *13*, 9220-9224.
(6)   Chaturvedi, S. S.; Bím, D.; Christov, C. Z.; Alexandrova, A. N. From Random to
        Rational: Improving Enzyme Design through Electric Fields, Second Coordination
        Sphere Interactions, and Conformational Dynamics. *Chem. Sci.* **2023,** *14*, 10997-11011.
(7)   Kastner, D. W.; Nandy, A.; Mehmood, R.; Kulik, H. J. Mechanistic Insights into
        Substrate Positioning That Distinguish Non-Heme Fe(Ii)/A-Ketoglutarate-Dependent
        Halogenases and Hydroxylases. *ACS Catal.* **2023,** *13*, 2489-2501.
(8)   Ortiz de Montellano, P. R. Hydrocarbon Hydroxylation by Cytochrome P450 Enzymes.
        Chem. Rev. **2010,** *110*, 932-948.
(9)   Vaillancourt, F. H.; Yeh, E.; Vosburg, D. A.; Garneau-Tsodikova, S.; Walsh, C. T.
        Nature's Inventory of Halogenation Catalysts: Oxidative Strategies Predominate. *Chem.
        Rev.* **2006,** *106*, 3364-3378.
(10)   Shaik, S.; de Visser, S. P.; Ogliaro, F.; Schwarz, H.; Schröder, D. Two-State Reactivity
        Mechanisms of Hydroxylation and Epoxidation by Cytochrome P-450 Revealed by
        Theory. *Curr. Opin. Chem. Biol.* **2002,** *6*, 556-567.
(11)   Meunier, B.; de Visser, S. P.; Shaik, S. Mechanism of Oxidation Reactions Catalyzed by
        Cytochrome P450 Enzymes. *Chem. Rev.* **2004,** *104*, 3947-3980.
(12)   Guengerich, F. P.; Yoshimoto, F. K. Formation and Cleavage of C–C Bonds by
        Enzymatic Oxidation–Reduction Reactions. *Chem. Rev.* **2018,** *118*, 6573-6655.
(13)   Arnold, F. H. Directed Evolution: Bringing New Chemistry to Life. *Angew. Chem. Int.
        Ed.* **2018,** *57*, 4143-4148.
(14)   Privett, H. K.; Kiss, G.; Lee, T. M.; Blomberg, R.; Chica, R. A.; Thomas, L. M.; Hilvert,
        D.; Houk, K. N.; Mayo, S. L. Iterative Approach to Computational Enzyme Design. *Proc.
        Nat. Acad. Soc.* **2012,** *109*, 3790-3795.
(15)   Jeschek, M.; Reuter, R.; Heinisch, T.; Trindler, C.; Klehr, J.; Panke, S.; Ward, T. R.
        Directed Evolution of Artificial Metalloenzymes for in Vivo Metathesis. *Nature* **2016,**
        537, 661-665.
(16)   Kan, S. B. J.; Lewis, R. D.; Chen, K.; Arnold, F. H. Directed Evolution of Cytochrome C
        for Carbon–Silicon Bond Formation: Bringing Silicon to Life. *Science* **2016,** *354*, 1048-
        1051.

(17) Ma, N.; Fang, W.; Liu, C.; Qin, X.; Wang, X.; Jin, L.; Wang, B.; Cong, Z. Switching an Artificial P450 Peroxygenase into Peroxidase Via Mechanism-Guided Protein Engineering. *ACS Catal.* **2021,** *11*, 8449-8455.

(18) Chalkley, M. J.; Mann, S. I.; DeGrado, W. F. De Novo Metalloprotein Design. *Nat. Rev. Chem.* **2022,** *6*, 31-50.

(19) Lu, Y.; Berry, S. M.; Pfister, T. D. Engineering Novel Metalloproteins: Design of Metal-Binding Sites into Native Protein Scaffolds. *Chem. Rev.* **2001,** *101*, 3047-3080.

(20) Rosenblatt, M. M.; Wang, J.; Suslick, K. S. De Novo Designed Cyclic-Peptide Heme Complexes. *Proc. Nat. Acad. Soc.* **2003,** *100*, 13140-13145.

(21) Polizzi, N. F.; Wu, Y.; Lemmin, T.; Maxwell, A. M.; Zhang, S.-Q.; Rawson, J.; Beratan, D. N.; Therien, M. J.; DeGrado, W. F. De Novo Design of a Hyperstable Non-Natural Protein–Ligand Complex with Sub-Å Accuracy. *Nat. Chem.* **2017,** *9*, 1157-1164.

(22) Kalvet, I.; Ortmayer, M.; Zhao, J.; Crawshaw, R.; Ennist, N. M.; Levy, C.; Roy, A.; Green, A. P.; Baker, D. Design of Heme Enzymes with a Tunable Substrate Binding Pocket Adjacent to an Open Metal Coordination Site. *J. Amer. Chem. Soc.* **2023,** *145*, 14307-14315.

(23) Polanco, E. A.; Opdam, L. V.; Passerini, L.; Huber, M.; Bonnet, S.; Pandit, A. An Artificial Metalloenzyme That Can Oxidize Water Photocatalytically: Design, Synthesis, and Characterization. *Chem. Sci.* **2024,** *15*, 3596-3609.

(24) Nastri, F.; D'Alonzo, D.; Leone, L.; Zambrano, G.; Pavone, V.; Lombardi, A. Engineering Metalloprotein Functions in Designed and Native Scaffolds. *Trends Biochem. Sci.* **2019,** *44*, 1022-1040.

(25) Chino, M.; Leone, L.; Zambrano, G.; Pirro, F.; D'Alonzo, D.; Firpo, V.; Aref, D.; Lista, L.; Maglio, O.; Nastri, F.et al. Oxidation Catalysis by Iron and Manganese Porphyrins within Enzyme-Like Cages. *Biopolymers* **2018,** *109*, e23107.

(26) Leone, L.; Chino, M.; Nastri, F.; Maglio, O.; Pavone, V.; Lombardi, A. Mimochrome, a Metalloporphyrin-Based Catalytic Swiss Knife†. *Biotechnol. Appl. Biochem.* **2020,** *67*, 495-515.

(27) Nastri, F.; Lista, L.; Ringhieri, P.; Vitale, R.; Faiella, M.; Andreozzi, C.; Travascio, P.; Maglio, O.; Lombardi, A.; Pavone, V. A Heme–Peptide Metalloenzyme Mimetic with Natural Peroxidase-Like Activity. *Chem. Eur. J.* **2011,** *17*, 4444-4453.

(28) Vitale, R.; Lista, L.; Cerrone, C.; Caserta, G.; Chino, M.; Maglio, O.; Nastri, F.; Pavone, V.; Lombardi, A. An Artificial Heme-Enzyme with Enhanced Catalytic Activity: Evolution, Functional Screening and Structural Characterization. *Org. Biomol. Chem.* **2015,** *13*, 4859-4868.

(29) Caserta, G.; Chino, M.; Firpo, V.; Zambrano, G.; Leone, L.; D'Alonzo, D.; Nastri, F.; Maglio, O.; Pavone, V.; Lombardi, A. Enhancement of Peroxidase Activity in Artificial Mimochrome Vi Catalysts through Rational Design. *ChemBioChem* **2018,** *19*, 1823-1826.

(30) Lee, C. H.; Dogutan, D. K.; Nocera, D. G. Hydrogen Generation by Hangman Metalloporphyrins. *J. Amer. Chem. Soc.* **2011,** *133*, 8775-8777.

(31) Bediako, D. K.; Solis, B. H.; Dogutan, D. K.; Roubelakis, M. M.; Maher, A. G.; Lee, C. H.; Chambers, M. B.; Hammes-Schiffer, S.; Nocera, D. G. Role of Pendant Proton Relays and Proton-Coupled Electron Transfer on the Hydrogen Evolution Reaction by Nickel Hangman Porphyrins. *Proc. Nat. Acad. Soc.* **2014,** *111*, 15001-15006.

(32)   Bím, D.; Alexandrova, A. N. Local Electric Fields as a Natural Switch of Heme-Iron Protein Reactivity. *ACS Catal.* **2021,** *11*, 6534-6546.

(33)   Shaik, S.; de Visser, S. P.; Kumar, D. External Electric Field Will Control the Selectivity of Enzymatic-Like Bond Activations. *J. Amer. Chem. Soc.* **2004,** *126*, 11746-11749.

(34)   Shaik, S.; Mandal, D.; Ramanan, R. Oriented Electric Fields as Future Smart Reagents in Chemistry. *Nat. Chem.* **2016,** *8*, 1091-1098.

(35)   Peng, W.; Yan, S.; Zhang, X.; Liao, L.; Zhang, J.; Shaik, S.; Wang, B. How Do Preorganized Electric Fields Function in Catalytic Cycles? The Case of the Enzyme Tyrosine Hydroxylase. *J. Amer. Chem. Soc.* **2022,** *144*, 20484-20494.

(36)   Song, Y.; Mao, J.; Gunner, M. R. Electrostatic Environment of Hemes in Proteins: Pkas of Hydroxyl Ligands. *Biochemistry* **2006,** *45*, 7949-7958.

(37)   Kulik, H. J. Large-Scale Qm/Mm Free Energy Simulations of Enzyme Catalysis Reveal the Influence of Charge Transfer. *Phys. Chem. Chem. Phys.* **2018,** *20*, 20650-20660.

(38)   Yang, Z.; Hajlasz, N.; Steeves, A. H.; Kulik, H. J. Quantifying the Long-Range Coupling of Electronic Properties in Proteins with Ab Initio Molecular Dynamics. *Chem. Methods* **2021,** *1*, 362-373.

(39)   Steeves, A. H.; Kulik, H. J. Insights into the Stability of Engineered Mini-Proteins from Their Dynamic Electronic Properties. *Electron. Struct.* **2022,** *4*, 034005.

(40)   Cortina, G. A.; Kasson, P. M. Excess Positional Mutual Information Predicts Both Local and Allosteric Mutations Affecting Beta Lactamase Drug Resistance. *Bioinformatics* **2016,** *32*, 3420-3427.

(41)   Nandy, A.; Adamji, H.; Kastner, D. W.; Vennelakanti, V.; Nazemi, A.; Liu, M.; Kulik, H. J. Using Computational Chemistry to Reveal Nature's Blueprints for Single-Site Catalysis of C–H Activation. *ACS Catal.* **2022,** *12*, 9281-9306.

(42)   Jurich, C.; Yang, Z. J. High-Throughput Computational Investigation of Protein Electrostatics and Cavity for Sam-Dependent Methyltransferases. *Protein Sci.* **2023,** *32*, e4690.

(43)   Eberhart, M. E.; Wilson, T. R.; Johnston, N. W.; Alexandrova, A. N. Geometry of Charge Density as a Reporter on the Role of the Protein Scaffold in Enzymatic Catalysis: Electrostatic Preorganization and Beyond. *J. Chem. Theory Comput.* **2023,** *19*, 694-704.

(44)   Yan, S.; Ji, X.; Peng, W.; Wang, B. Evaluating the Transition State Stabilization/Destabilization Effects of the Electric Fields from Scaffold Residues by a Qm/Mm Approach. *J. Phys. Chem. B* **2023,** *127*, 4245-4253.

(45)   Nazemi, A.; Steeves, A. H.; Kastner, D. W.; Kulik, H. J. Influence of the Greater Protein Environment on the Electrostatic Potential in Metalloenzyme Active Sites: The Case of Formate Dehydrogenase. *J. Phys. Chem. B* **2022,** *126*, 4069-4079.

(46)   Perrella, F.; Raucci, U.; Chiariello, M. G.; Chino, M.; Maglio, O.; Lombardi, A.; Rega, N. Unveiling the Structure of a Novel Artificial Heme-Enzyme with Peroxidase-Like Activity: A Theoretical Investigation. *Biopolymers* **2018,** *109*, e23225.

(47)   Goddard, T. D.; Huang, C. C.; Meng, E. C.; Pettersen, E. F.; Couch, G. S.; Morris, J. H.; Ferrin, T. E. Ucsf Chimerax: Meeting Modern Challenges in Visualization and Analysis. *Protein Sci.* **2018,** *27*, 14-25.

(48)   Pettersen, E. F.; Goddard, T. D.; Huang, C. R. C.; Meng, E. E. C.; Couch, G. S.; Croll, T. I.; Morris, J. H.; Ferrin, T. E. Ucsf Chimerax: Structure Visualization for Researchers, Educators, and Developers. *Protein Sci.* **2021,** *30*, 70-82.

(49)   Rohrdanz, M. A.; Martins, K. M.; Herbert, J. M. A Long-Range-Corrected Density Functional That Performs Well for Both Ground-State Properties and Time-Dependent Density Functional Theory Excitation Energies, Including Charge-Transfer Excited States. *J. Chem. Phys.* **2009,** *130*, 054112.

(50)   Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988,** *37*, 785-789.

(51)   Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994,** *98*, 11623-11627.

(52)   Hay, P. J.; Wadt, W. R. Ab Initio Effective Core Potentials for Molecular Calculations. Potentials for the Transition Metal Atoms Sc to Hg. *J. Chem. Phys.* **1985,** *82*, 270-283.

(53)   Lange, A. W.; Herbert, J. M. A Smooth, Nonsingular, and Faithful Discretization Scheme for Polarizable Continuum Models: The Switching/Gaussian Approach. *J. Chem. Phys.* **2010,** *133*, 244111.

(54)   York, D. M.; Karplus, M. A Smooth Solvation Potential Based on the Conductor-Like Screening Model. *J. Phys. Chem. A* **1999,** *103*, 11060-11079.

(55)   Liu, F.; Luehr, N.; Kulik, H. J.; Martínez, T. J. Quantum Chemistry for Solvated Molecules on Graphical Processing Units Using Polarizable Continuum Models. *J. Chem. Theory Comput.* **2015,** *11*, 3131-3144.

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

(57)   Kastner, D. W.; Reinhardt, C. R.; Adamji, H.; Manetsch, M. T.; Roman-Leshkov, Y.; Kulik, H. J. Dynamic Charge Distribution as a Key Driver of Catalytic Reactivity in an Artificial Metalloenzyme. *Zenodo* **2024**, DOI:10.5281/zenodo.11330706 10.5281/zenodo.11330706.

(58)   Debiec, K. T.; Cerutti, D. S.; Baker, L. R.; Gronenborn, A. M.; Case, D. A.; Chong, L. T. Further Along the Road Less Traveled: Amber Ff15ipq, an Original Protein Force Field Built on a Self-Consistent Physical Model. *J. Chem. Theory Comput.* **2016,** *12*, 3926-3947.

(59)   Takemura, K.; Kitao, A. Water Model Tuning for Improved Reproduction of Rotational Diffusion and Nmr Spectral Density. *J. Phys. Chem. B* **2012,** *116*, 6279-6287.

(60)   Li, P. F.; Merz, K. M. Mcpb.Py: A Python Based Metal Center Parameter Builder. *J. Chem. Inf. Model.* **2016,** *56*, 599-604.

(61)   Lee, C. T.; Yang, W. T.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron-Density. *Phys. Rev. B* **1988,** *37*, 785-789.

(62)   Seminario, J. M. Calculation of Intramolecular Force Fields from Second-Derivative Tensors. *Int. J. Quantum Chem.* **1996,** *60*, 1271-1277.

(63)   Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983,** *79*, 926-935.

(64)   Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993,** *98*, 10089-10092.

(65)   Gotz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with Amber on Gpus. 1. Generalized Born. *J. Chem. Theory Comput.* **2012,** *8*, 1542-1555.

(66)   Salomon-Ferrer, R.; Gotz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with Amber on Gpus. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013,** *9*, 3878-3888.

(67)   Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. *J. Comput. Phys.* **1977,** *23*, 327-341.

(68)   Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. Mmpbsa.Py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **2012,** *8*, 3314-3321.

(69)   Massova, I.; Kollman, P. A. Combined Molecular Mechanical and Continuum Solvent Approach (Mm-Pbsa/Gbsa) to Predict Ligand Binding. *Perspect. Drug Discovery Des.* **2000,** *18*, 113-135.

(70)   Tsui, V.; Case, D. A. Theory and Applications of the Generalized Born Solvation Model in Macromolecular Simulations. *Biopolymers* **2001,** *56*, 275-291.

(71)   Onufriev, A.; Bashford, D.; Case, D. A. Exploring Protein Native States and Large-Scale Conformational Changes with a Modified Generalized Born Model. *Proteins* **2004,** *55*, 383-394.

(72)   Hou, T. J.; Wang, J. M.; Li, Y. Y.; Wang, W. Assessing the Performance of the Mm/Pbsa and Mm/Gbsa Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2011,** *51*, 69-82.

(73)   Gowers, R. J.; Linke, M.; Barnoud, J.; Reddy, T. J. E.; Melo, M. N.; Seyler, S. L.; Domanski, J.; Dotson, D. L.; Buchoux, S.; Kenney, I. M. "Mdanalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations," Los Alamos National lab.(LANL), Los Alamos, NM (United States), 2019.

(74)   Roe, D. R.; Cheatham, T. E. Ptraj and Cpptraj: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013,** *9*, 3084-3095.

(75)   Mehmood, R.; Qi, H. W.; Steeves, A. H.; Kulik, H. J. The Protein's Role in Substrate Positioning and Reactivity for Biosynthetic Enzyme Complexes: The Case of Syrb2/Syrb1. *ACS Catal.* **2019,** *9*, 4930-4943.

(76)   Bannwarth, C.; Ehlert, S.; Grimme, S. Gfn2-Xtb—an Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **2019,** *15*, 1652-1671.

(77)   Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (Dft-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010,** *132*, 154104.

(78)   Ufimtsev, I. S.; Martinez, T. J. Quantum Chemistry on Graphical Processing Units. 3. Analytical Energy Gradients, Geometry Optimization, and First Principles Molecular Dynamics. *J. Chem. Theory Comput.* **2009,** *5*, 2619-2628.

(79)   Seritan, S.; Bannwarth, C.; Fales, B. S.; Hohenstein, E. G.; Isborn, C. M.; Kokkila-Schumacher, S. I. L.; Li, X.; Liu, F.; Luehr, N.; Snyder, J. W.et al. Terachem: A Graphical Processing Unit-Acceleratedlectronic Structure Package Forlarge-Scaleab Initio Molecular Dynamics. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2021,** *11*, e1494.

36

(80) Lange, A. W.; Herbert, J. M. A Smooth, Nonsingular, and Faithful Discretization Scheme for Polarizable Continuum Models: The Switching/Gaussian Approach. *J. Chem. Phys.* **2010,** *133*.

(81) Bondi, A. Van Der Waals Volumes and Radii. *J. Phys. Chem.* **1964,** *68*, 441-451.

(82) Ranieri, A.; Monari, S.; Sola, M.; Borsari, M.; Battistuzzi, G.; Ringhieri, P.; Nastri, F.; Pavone, V.; Lombardi, A. Redox and Electrocatalytic Properties of Mimochrome Vi, a Synthetic Heme Peptide Adsorbed on Gold. *Langmuir* **2010,** *26*, 17831-17835.

(83) Vitale, R.; Lista, L.; Lau-Truong, S.; Tucker, R. T.; Brett, M. J.; Limoges, B.; Pavone, V.; Lombardi, A.; Balland, V. Spectroelectrochemistry of Feiii- and Coiii-Mimochrome Vi Artificial Enzymes Immobilized on Mesoporous Ito Electrodes. *Chem. Commun.* **2014,** *50*, 1894-1896.

(84) Yang, Z. Y.; Mehmood, R.; Wang, M. Y.; Qi, H. W.; Steeves, A. H.; Kulik, H. J. Revealing Quantum Mechanical Effects in Enzyme Catalysis with Large-Scale Electronic Structure Simulation. *React. Chem. Eng.* **2019,** *4*, 298-315.

(85) Fonseca Guerra, C.; Handgraaf, J.-W.; Baerends, E. J.; Bickelhaupt, F. M. Voronoi Deformation Density (Vdd) Charges: Assessment of the Mulliken, Bader, Hirshfeld, Weinhold, and Vdd Methods for Charge Analysis. *J. Comput. Chem.* **2004,** *25*, 189-210.

(86) Karelina, M.; Kulik, H. J. Systematic Quantum Mechanical Region Determination in Qm/Mm Simulation. *J. Chem. Theory Comput.* **2017,** *13*, 563-576.

(87) Kulik, H. J.; Zhang, J.; Klinman, J. P.; Martínez, T. J. How Large Should the Qm Region Be in Qm/Mm Calculations? The Case of Catechol O-Methyltransferase. *J. Phys. Chem. B* **2016,** *120*, 11381-11394.

(88) Hirshfeld, F. L. Bonded-Atom Fragments for Describing Molecular Charge Densities. *Theor. Chim. Acta* **1977,** *44*, 129-138.

(89) Lu, T.; Chen, F. Multiwfn: A Multifunctional Wavefunction Analyzer. *J. Comput. Chem.* **2012,** *33*, 580-592.

(90) Tin Kam, H. Proceedings of 3rd International Conference on Document Analysis and Recognition, 1995; p 278-282 vol.1.

(91) Rumelhart, D. E.; Hinton, G. E.; Williams, R. J. Learning Representations by Back-Propagating Errors. *Nature* **1986,** *323*, 533-536.

(92) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011,** *12*, 2825–2830.

(93) Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.et al. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*; Curran Associates Inc., 2019,Article 721

(94) Krizhevsky, A.; Sutskever, I.; Hinton, G. E. Imagenet Classification with Deep Convolutional Neural Networks. *Communications of the ACM* **2017,** *60*, 84–90.

(95) Kingma, D. P.; Ba, J. ICLR, 2015.

(96) Lundberg, S. M.; Lee, S.-I. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*; Curran Associates Inc.: Long Beach, California, USA, 2017.

(97) Lundberg, S. M. Shap. https://github.com/slundberg/shap. (Accessed June 12, 2023).

(98) Leone, L.; Muñoz-García, A. B.; D'Alonzo, D.; Pavone, V.; Nastri, F.; Lombardi, A. Peptide-Based Metalloporphyrin Catalysts: Unveiling the Role of the Metal Ion in Indole Oxidation. *J. Inorg. Biochem.* **2023,** *246*, 112298.

(99)   Fleetwood, O.; Kasimova, M. A.; Westerlund, A. M.; Delemotte, L. Molecular Insights from Conformational Ensembles Via Machine Learning. *Biophys. J.* **2020,** *118*, 765-780.

(100)   Ichiye, T.; Karplus, M. Collective Motions in Proteins: A Covariance Analysis of Atomic Fluctuations in Molecular Dynamics and Normal Mode Simulations. *Proteins* **1991,** *11*, 205-217.

(101)   McClendon, C. L.; Friedland, G.; Mobley, D. L.; Amirkhani, H.; Jacobson, M. P. Quantifying Correlations between Allosteric Sites in Thermodynamic Ensembles. *J. Chem. Theory Comput.* **2009,** *5*, 2486-2502.

(102)   Guo, J.; Zhou, H.-X. Protein Allostery and Conformational Dynamics. *Chem. Rev.* **2016,** *116*, 6503-6515.

(103)   Schlichting, I.; Berendzen, J.; Chu, K.; Stock, A. M.; Maves, S. A.; Benson, D. E.; Sweet, R. M.; Ringe, D.; Petsko, G. A.; Sligar, S. G. The Catalytic Pathway of Cytochrome P450cam at Atomic Resolution. *Science* **2000,** *287*, 1615-1622.

(104)   Harris, D. L.; Loew, G. H. Identification of Putative Peroxide Intermediates of Peroxidases by Electronic Structure and Spectra Calculations. *J. Amer. Chem. Soc.* **1996,** *118*, 10588-10594.

(105)   Wang, B.; Li, C.; Dubey, K. D.; Shaik, S. Quantum Mechanical/Molecular Mechanical Calculated Reactivity Networks Reveal How Cytochrome P450cam and Its T252a Mutant Select Their Oxidation Pathways. *J. Amer. Chem. Soc.* **2015,** *137*, 7379-7390.

For Table of Contents Use Only