

Mechanistic Insights into Enzyme Catalysis from Explaining Machine-Learned Quantum Mechanical and Molecular Mechanical Minimum Energy Pathways

Zilin Song, Francesco Trozzi, Hao Tian, Chao Yin, and Peng Tao*

Cite This: *ACS Phys. Chem Au* 2022, 2, 316–330

Read Online

ACCESS |



Metrics & More



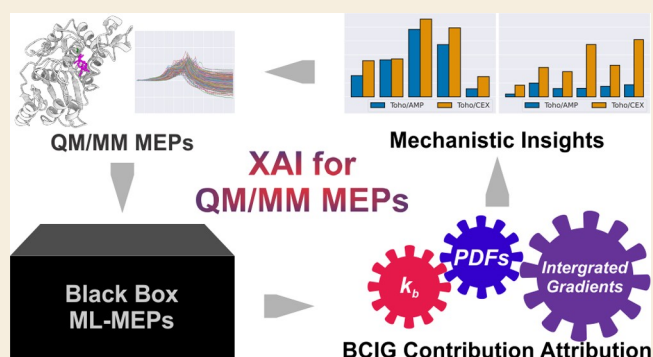
Article Recommendations



Supporting Information

ABSTRACT: With the increasing popularity of machine learning (ML) applications, the demand for explainable artificial intelligence techniques to explain ML models developed for computational chemistry has also emerged. In this study, we present the development of the Boltzmann-weighted cumulative integrated gradients (BCIG) approach for effective explanation of mechanistic insights into ML models trained on high-level quantum mechanical and molecular mechanical (QM/MM) minimum energy pathways. Using the acylation reactions of the Toho-1 β -lactamase and two antibiotics (ampicillin and cefalexin) as the model systems, we show that the BCIG approach could quantitatively attribute the energetic contribution in one system and the relative reactivity of individual steps across different systems to specific chemical processes such as the bond making/breaking and proton transfers. The proposed BCIG contribution attribution method quantifies chemistry-interpretable insights in terms of contributions from each elementary chemical process, which is in agreement with the validating QM/MM calculations and our intuitive mechanistic understandings of the model reactions.

KEYWORDS: QM/MM, minimum energy pathways, deep learning, explainable artificial intelligence, class A β -lactamases, Toho-1, β -lactam antibiotic resistance, ampicillin, cefalexin



INTRODUCTION

Machine learning (ML) has emerged with a great promise to approximate the target function of any form regardless of the a priori knowledge about the underlying correlations among variables. The application of various ML techniques has also advanced theoretical chemistry in various subjects,^{1–4} which have been suffering from either the extensive computational demands of high levels of theory^{5–8} or the high dimensionality of the chemical and/or conformational spaces.^{9–14} Although ML could be introduced to many topics that require accurate and efficient approximations, its performance and effectiveness have been limited by the feature representation and interpretability of the model.¹⁵ In addition to the routinely applied feature representations, unsupervised models and rational statistical procedures have been developed and applied to extract robust feature vectors from the chemical feature space.^{16–18} In particular, considerable pioneering efforts have focused on the development of suitable descriptors and accurate ML models for approximating hybrid quantum mechanical/molecular mechanical (QM/MM) potentials.^{19–23}

With the booming popularity of ML, interest to interpret deep learning (DL) neural networks have synergistically risen as a subfield of great importance, namely the explainable artificial intelligence (XAI).²⁴ The DL models being

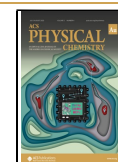
interpretable not only elevate our understanding of the learning algorithms but also constitute responsible ML/DL-assisted decision making.²⁵ Practically, XAI techniques attribute the predicted outcome of DL models to individual feature contributions, therefore, rationalizing the driving forces behind the decision flow in the models that are black boxes. While explicit indicators for feature contributions are straightforward in linear models²⁶ and are incorporated by design in specific ensemble-based models,²⁷ explaining neural networks is in general hindered by the high nonlinearity accumulated through the activations of the hidden layers. Based on the assumption that the predicted nonlinear surface could be approximated as linear at local regions, effective importance attribution methods have been proposed based on model gradients.²⁸ The state-of-the-art XAI techniques, such as the integrated gradients (IG)²⁹ and the layer-wise relevance

Received: January 19, 2022

Revised: March 16, 2022

Accepted: March 16, 2022

Published: May 18, 2022



propagation,³⁰ have demonstrated great promise in various explaining tasks such as medical diagnosis³¹ and cheminformatics applications.³² In addition, interpretable-by-design DL architectures effectively making the use of the attention mechanism have been developed for challenging explaining tasks such as protein–ligand binding affinities.^{33,34} Whereas DL models have proven to accurately predict various chemical properties, we are particularly interested in that if the DL models could also be *post hoc* explained to gain mechanistic insights into enzyme catalysis.

In this study, the acylation between Toho-1, a class A serine-based β -lactamase (AS β L), and two β -lactam antibiotics, ampicillin (AMP) and cefalexin (CEX), were used as the model reactions (Figure 1a,b). Theoretical efforts have revealed that the acylation of AS β Ls could undergo two distinct pathways, which are distinguished by the reactant protonation states of Lys73 and Glu166 (Figure 1c).^{35–40} Hermann et al.^{35,36} first reported that the acylation is initialized by a deprotonated Glu166 acting as the general base, leading to the Glu166 as the sole base pathway which produces the acyl-enzyme product (R1-AE). Meroueh et al.³⁷ further proposed that the deprotonated Lys73 and protonated Glu166 could concertedly mediate the acylation (R2-AE). Whereas the acylation mechanisms of β -lactamases are considered similar, in general, the reactivity of β -lactamases varies with different β -lactam families.³⁸ The understanding of the determinant factors that impact the acylation activity of different ligand scaffolds is challenging but critical for future optimizations of β -lactam drugs.³⁹ Our recent study on the acylation of Toho-1 by AMP and CEX showed that the R1-AE pathways are viable for AMP but prohibitive for CEX, while the R2-AE mechanism is generally accessible for both AMP and CEX.⁴⁰ Herein, we aim to unravel the conformational and energetic factors that prime the different pathways observed for the acylation of Toho-1 by AMP and CEX. We first present 800 minimum energy pathways (MEPs) that were obtained from a robust computational workflow combining constrained molecular dynamics (MD) and QM/MM chain-of-state calculations. Feed-forward neural networks were designed to accurately predict the MEP profiles using conformational features selected from an ad hoc univariate variable analysis. In light of the pioneering works on XAI, we developed a thermodynamics-aware contribution attribution method to interpret the ML MEP models. We show that quantitative mechanistic insights into various aspects of enzyme catalysis can be derived from interpreting purpose-oriented ML-MEP models in terms of energy contributions from chemistry-interpretable factors.

METHODS

For clarity, we first present the theoretical framework for contribution attribution of ML/DL models that were trained on an ensemble of QM/MM MEPs. For an ML-MEP model, which is designated as F and is trained on a dataset of P MEPs, the contribution of a chemical process c on the p -th MEP can be represented as the “pathway-wise” contribution attributing function A_{path}

$$\text{contribution}_c^p = A_{\text{path}}(F, c, p) \quad (1)$$

The overall contribution of chemical process c in the ensemble of P MEPs is the sum of the contribution $_c^p$ weighted by the accessibility (the Boltzmann factor) of p -th MEP, which is

$$\text{contribution}_c^P = C_N \sum_{p=1}^P \left(\exp\left(-\frac{\Delta E^p}{RT}\right) \times A_{\text{path}}(F, c, p) \right) \quad (2)$$

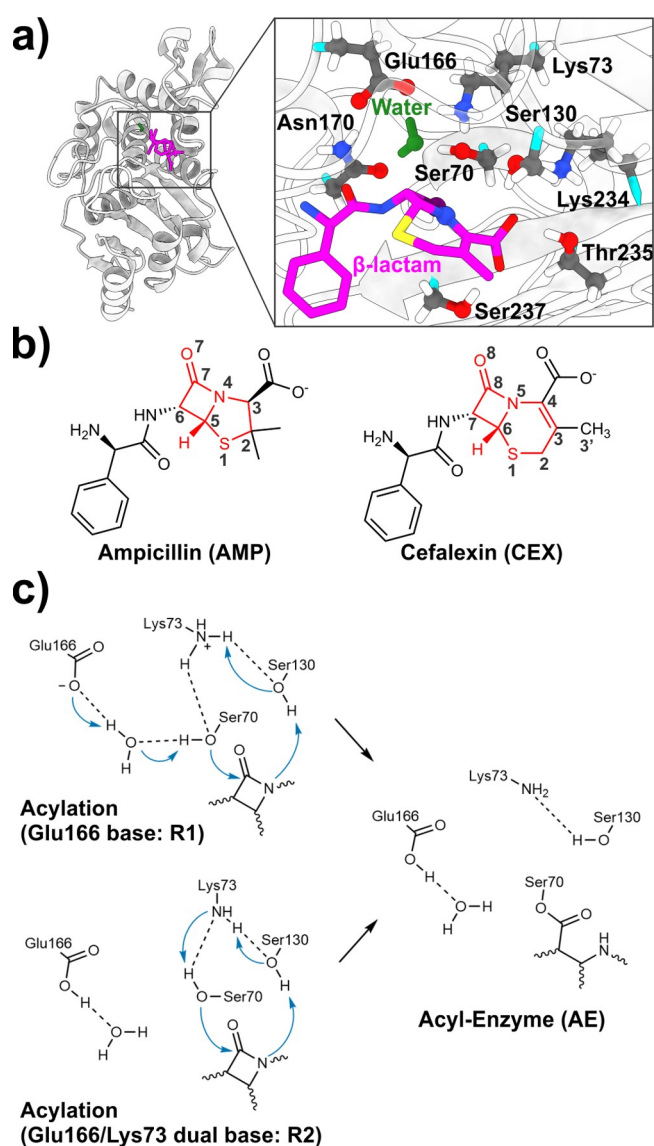


Figure 1. Toho-1 β -lactamase, the β -lactam antibiotics, and the acylation pathways of Toho-1. (a) Toho-1/ β -lactam complex and the selection of QM atoms. The carbon atoms of the amino acid residues are colored in dark gray. The $\text{C}\alpha$ – $\text{C}\beta$ bonds (used as the QM/MM boundary for the amino acid side chains) are colored in cyan. The carbon atoms of the β -lactam ligand are colored in magenta, except the carbonyl carbon in β -lactams, which is colored in dark purple. The catalytic water molecule is uniformly colored in green. All nitrogen, oxygen, sulfur, and hydrogen atoms are colored in blue, red, yellow, and white, respectively. All hydrogen atoms in the ligand are omitted for clarity. (b) Chemical structures of AMP and CEX. The penam and cephem scaffolds are colored in red; (c) Acylation pathways using Glu166 as the general base (R1-AE) and using Glu166/Lys73 concertedly as the general base (R2-AE).

where ΔE^p is the energy barrier of the p -th MEP, R is the ideal gas constant, T is the temperature, and C_N is a normalization factor.

The exponential averaging implicitly assumes sample completeness in the MEP datasets, which is mostly impractical for actual MEP calculations. In practice, the direct application of Boltzmann weights would lead to numerical instability with a limited number of sampled MEPs. Alternatively, a probability density function (PDF), $\text{PDF}(\Delta E^p)$, could be introduced to smoothen the density of MEP barriers. In its simplest form, the PDF could be a single Gaussian function. Furthermore, in cases where the sampled barrier distribution failed to approximate a Gaussian distribution, alternative density estimators

such as Gaussian mixture models (GMMs) or Kernel density estimations could be employed for better approximation. Nonetheless, introducing PDF to eq 2 yields

$$\text{contribution}_c^p = C'_N \sum_{p=1}^P \left(\exp\left(-\frac{\Delta E^p}{RT}\right) \times \text{PDF}(\Delta E^p) \times A_{\text{path}}(F, c, p) \right) \quad (3)$$

Note that the normalization factor C'_N is

$$C'_N = \frac{P}{\sum_{p=1}^P \exp\left(-\frac{\Delta E^p}{RT}\right) \times \text{PDF}(\Delta E^p)} \quad (4)$$

In our implementation, 2-component GMMs, which approximate the distribution of MEP barriers using a weighted sum of two independent Gaussian, were used as the PDF for MEP barriers.

If the chain-of-states replica path method (RPM, see [Computational Details](#))^{41–43} is used for MEP calculations, each transition path is represented by a series of discrete replicated structures (replicas) that connect the reactant and product. The “pathway-wise” contribution A_{path} could be calculated from the “replica-wise” attribution function A_{replica} of c to the energy of the r -th replica on the p -th MEP

$$A_{\text{path}}(F, c, p) = \sum_{r=1}^M A_{\text{replica}}(F, c, p, r) \quad (5)$$

where M is the total number of replicas in each MEP.

As proposed by Sundararajan et al.,²⁹ for a DL model F , the contribution of the i -th feature x_i of the feature vector \mathbf{x} corresponding to a specific sample point can be calculated as the IG along a path $\gamma(\alpha)$ that connects the sample point with feature vector \mathbf{x} (where $\alpha = 1$) and a baseline with feature vector \mathbf{x}' (where $\alpha = 0$)

$$\text{integratedGrad}_i^{\gamma}(x) = \int_{\alpha=0}^1 \frac{\partial F(\gamma(\alpha))}{\partial \gamma(\alpha)} \frac{\partial \gamma(\alpha)}{\partial \alpha} d\alpha \quad (6)$$

In our case, the reactant states on the MEPs were selected as the baselines, and the contribution of c at the r -th replica was the integrated partial derivatives (with regard to c) through the intermediate replicas preceding the r -th replica along the MEP. Accordingly, eq 6 is adapted for the discrete reaction pathway as

$$A_{\text{replica}}(F, c, p, r) = \text{IntegratedGrad}_c^p(x) = \sum_{i=1}^r \frac{\partial F(\mathbf{x}(p, i))}{\partial \mathbf{x}(p, i)} \frac{\partial \mathbf{x}(p, i)}{\partial i} \quad (7)$$

where i is the index of the pathway replicas, and $\mathbf{x}(p, i)$ is the feature vector of the i -th replica on the p -th MEP.

The representation of c must be determined to expand the first partial derivative in eq 7. As noted, c represents a “chemical process” that includes (but is not limited to) bond making/breaking and proton transfers. The progress of the chemical process is commonly represented by the linear combination of multiple order parameters such as atomic distances, often referred to as the reaction coordinates or collective variables. However, we note that the correlation between the atomic distances is highly nonlinear in the evolution along the optimal reaction path obtained from the chain-of-state calculations. Therefore, instead of feeding the reduced representation of linear-combined atomic distances, we used a set of atomic distances that accounts for all chemical processes during the acylation for the training of the DL model F , that is

$$E(p, i) \cong F(\mathbf{x}(p, i)) \quad (8)$$

where $E(p, i)$ is the energy of the i -th replica on the p -th MEP. Furthermore, owing to the nonlinearity of the correlation between the feature dimensions, the gradients of F with regard to c , which usually

correspond to multiple feature dimensions in \mathbf{x} , cannot be calculated analytically. Therefore, the first partial derivative in eq 7 is computed numerically

$$\frac{\partial F(\mathbf{x}(p, i))}{\partial \delta(\mathbf{x}(p, i))} = [F(\mathbf{x}(p, i) + \delta\epsilon(\mathbf{x}(p, i+1) - \mathbf{x}(p, i))) - F(\mathbf{x}(p, i) - \delta\epsilon(\mathbf{x}(p, i) - \mathbf{x}(p, i-1)))] / [Dist(\mathbf{x}(p, i) + \delta\epsilon(\mathbf{x}(p, i+1) - \mathbf{x}(p, i)), \mathbf{x}(p, i)) - Dist(\mathbf{x}(p, i) - \delta\epsilon(\mathbf{x}(p, i) - \mathbf{x}(p, i-1)), \mathbf{x}(p, i))] \quad (9)$$

where ϵ denotes a small perturbative factor (0.01 as in the current study); δ acts as the selector for the feature dimensions included in the chemical process (c). In practice, δ is a multihot-encoded mask to ensure that the perturbation ϵ is applied only to the features \mathbf{x} that represent the chemical process of interest; $Dist(A, B)$ stands for the distance metric that accounts for the pathway curvature. The gradient on the pathway curvature reads

$$\frac{\partial \mathbf{x}(p, i)}{\partial i} = [Dist(\mathbf{x}(p, i) + \delta\epsilon(\mathbf{x}(p, i+1) - \mathbf{x}(p, i)), \mathbf{x}(p, i)) - Dist(\mathbf{x}(p, i) - \delta\epsilon(\mathbf{x}(p, i) - \mathbf{x}(p, i-1)), \mathbf{x}(p, i))] / [E((i+1) - (i-1))] \quad (10)$$

Combining eqs 7, 9, and 10, the contribution of the chemical process c to the r -th replica on the p -th pathway is calculated as the integrated partial gradients of the ML-MEP model F as

$$A_{\text{replica}}(F, c, p, r) = \sum_{i=1}^r [F(\mathbf{x}(p, i) + \delta\epsilon(\mathbf{x}(p, i+1) - \mathbf{x}(p, i))) - F(\mathbf{x}(p, i) - \delta\epsilon(\mathbf{x}(p, i) - \mathbf{x}(p, i-1)))] / 2\epsilon \quad (11)$$

Accordingly, the contribution of c along one MEP $A_{\text{path}}(F, c, p)$ can be calculated by cumulatively summing the integrated partial gradients (eq 5). The sign of $A_{\text{path}}(F, c, p)$ gives the interaction between the chemical processes, whereas its absolute values give the perturbative response of the ML-MEP model regarding different c . Therefore, the absolute values of $A_{\text{path}}(F, c, p)$ were used to calculate the weighted contributions in eq 3.

In this study, eq 3 is termed the Boltzmann-weighted cumulative IG (BCIG) contributions. We proceed with the computational details that implement the BCIG approach for interpreting the acylation reactions of Toho/AMP and Toho/CEX.

Computational Details

System Setup. The crystal structure of the Toho-1/benzylpenicillin complex with high resolution (PDB entry: SKMW)⁴⁴ was used as the template to generate the target systems. Toho/AMP and Toho/CEX systems were constructed from the crystal structure of the Toho-1/benzylpenicillin complex, respectively. The parameters for AMP and CEX were obtained from the CHARMM general force field (CGenFF).^{45,46} The three engineered mutations (Ala166/Asn274/Asn276) in the crystal structure were modified to Glu166/Arg274/Arg276 as in the wild type enzyme. All titratable residues were protonated under neutral pH based on PropKa calculations and experimental neutron diffraction results.^{47,48} Both systems were solvated in 80 Å cubic water boxes. Sodium and chloride ions were added to neutralize the total charges in the systems. The CHARMM36 force field (C36),⁴⁹ CGenFF,⁵⁰ and TIP3P parameters were used to treat the protein, ligands, and solvent molecules, respectively. The solvent molecules were held rigid with the SHAKE constraint.⁵¹ The nonbonding interactions excluding long-range electrostatic interactions between the atoms using molecular mechanics (MM) were explicitly treated within 10 Å and were smoothed out to zero at 12 Å. The periodic boundary conditions were employed in all three dimensions, and the particle mesh Ewald method⁵² was used for the summation of classical long-range electrostatic interactions.

The solvated systems were then subjected to 500 steps of steepest descent minimizations followed by 5000 steps of adopted basis Newton–Raphson (ABNR) minimizations using the MM potentials. In order to obtain reliable orientations of the active site, we further optimized the system for 5000 ABNR steps using the semiempirical third-order density-functional tight-binding (DFTB3) potential with the 3OB parameter set for the QM region and CHARMM36 force field for the MM region (DFTB3/3OB/C36).^{53,54} The DFTB3/3OB/C36 potential has been used extensively for simulating biomolecules, including β -lactamases,⁵⁵ for its computational efficiency. While the size of the QM region converges slowly with respect to physicochemical properties,^{56,57} we note that we did not perform extensive benchmarks on the dependency of energetic profiles to the QM selections. Instead, the QM atoms are picked as the balance between interactions with the reacting atoms and computational cost. The QM region includes Ser70, Lys73, Ser130, Glu166, Asn170, Lys234, Thr235, Ser237, ligands, and catalytic water (Figure 1a). Our selections yield 125 and 123 atoms in the QM subsystems of Toho/AMP and Toho/CEX, respectively. The QM part of the amino acid residues was capped by hydrogen link atoms through their α -C β bonds. In all QM/MM simulations, the classical charges on the MM host atoms ($C\alpha$) were removed to prevent the unphysical overpolarizations to the QM density.

After energy minimization, both systems were subjected to heating and equilibration simulations using classical force fields. All QM atoms were kept frozen during the heating and equilibration stage in order to preserve the QM-optimized geometries of the active site. Accordingly, the QM link atoms (used in the DFTB3 minimizations) were temporarily removed, and the masses of the MM hosting carbon atoms were rescaled to their unit atomic mass (12.0110). Both systems were heated from 110 to 310 K with a temperature increment of 10 K per 1.5 ps followed by canonical (NVT) equilibrations at 310 K for 150 ps. The temperatures during the heating and the NVT equilibration dynamics were regulated *via* explicit velocity rescaling. Both Toho/AMP and Toho/CEX systems were then subjected to 150 ps isothermal–isobaric (NPT) equilibration dynamics. During this stage, the temperature of the systems was maintained *via* the Hoover thermostat⁵⁸ at 310 K, while the pressure was coupled to 1 atm with the Langevin piston method.⁵⁹

Based on reaction mechanisms of the target reaction, we created the reactant state suitable for the mechanism with Glu166 as the sole base (R1), the reactant state suitable for the mechanism with Lys73 and Glu166 as dual bases (R2), and the acyl-enzyme (AE) state for Toho/AMP and Toho/CEX systems using the DFTB3/3OB/C36 potential with necessary distance-based restraints. A total of six states (R1, R2, and AE of Toho/AMP and Toho/CEX) were subjected to extensive conformational sampling with constrained NVT simulations for 150 ns. During the constrained dynamics, the reacting function groups (the hydroxyls of the Ser70/Ser130, the Lys73 amino group, the Glu166 carboxyl group, the carbonyl–nitrogen bond of the ligands, and the catalytic water) were fixed in place to retain their QM optimized orientations, and the rest of the QM atoms were allowed to move freely. The snapshots used for the MEP calculations were taken from the last 120 ns of the constrained MD trajectories with a time interval of 1.2 ns for each system. Briefly, a total number of 600 snapshots (100 snapshots from three states of two systems) were selected as the starting conformations for the MEP calculations.

QM/MM MEPs. The 600 starting conformations were all optimized using the DFTB3 potential. During the optimizations, the surrounding MM residues within 4 Å of the QM region were allowed to move while the remaining of the system was fixed. The corresponding product/reactant states were generated from the starting conformations. Specifically, AE states were generated from each R1 or R2 conformation from the simulations as the product state for each R1 \rightarrow AE or R2 \rightarrow AE pathway. For each AE state sampled from the simulations, both R1 and R2 reactant states are generated, leading to two independent AE \rightarrow R1 and AE \rightarrow R2 pathways. In total, 800 pairs of either R1/AE or R2/AE states were generated for the Toho/ligand complexes and were used as the reactant/product

pairs for the MEP optimizations using the RPM with holonomic constraints.

The accuracy of the QM potential is crucial for the quality of the MEP profiles. The DFTB3 method has been previously validated to produce accurate geometries at stable states. However, it was also reported that this approach would largely overestimate the bond dissociation energies between the heavy atoms.⁶⁰ Alternatively, a plausible approach for obtaining high-quality pathway profiles is to first optimize the MEP geometry under a fast, parameterized semiempirical QM method and refine the single-point energies using higher *ab initio*/DFT QM levels of theory along the reaction pathway represented as a chain of replicas.⁶¹ This approach requires that the MEPs optimized at the semiempirical QM level of theory are close to the MEPs optimized at higher *ab initio*/DFT QM levels of theory to achieve the goals of avoiding large errors of the energetics using semiempirical QM models such as DFTB3 and high computational cost of pathway optimization using the higher QM levels of theory. We compared the MEPs optimized using DFTB3/3OB/C36 and B3LYP-D3/6-31G**/C36 levels of theory, respectively (Figure S1). Additionally, the 3OB-f parameter set,⁵³ with optimal stretching vibrational frequencies for the C=C, C=N, and C=O species, was also employed as comparison (referred to as DFTB3/3OB-f/C36). Overall, the DFTB3/3OB/C36 level of theory overestimates the bond lengths of the tetrahedral intermediates compared to that of DFT calculations. On the other hand, DFTB3/3OB-f/C36 MEPs closely resemble the B3LYP-D3/6-31G**/C36 MEPs in terms of key bond distances during the acylation processes. Therefore, DFTB3/3OB-f/C36 was applied to carry out all the pathway optimization calculations in this study. The B3LYP hybrid functional^{62,63} with the 6-31+G** basis set⁶⁴ plus the D3 dispersion correction⁶⁵ (B3LYP-D3/6-31+G**/C36) were employed as the high-level QM counterpart to refine the single-point energies along the optimized MEPs. We note that the combination of the levels of theory (DFTB3 and B3LYP-D3) has been proposed previously for enzyme catalysis simulations.^{66,67} It has also been demonstrated that both the DFTB3 and B3LYP methods are applicable for studying concerted reaction steps in QM/MM simulations.^{68,69}

The initial guesses of the MEPs were created by linearly intercepting the Cartesian space between the reactants and the acyl-enzyme products using 50 replicated structures. The constrained RPM method,^{41–43} which enforces equal mass-weighted root-mean-square deviation distances between adjacent replicas, was applied to optimize the initial linear chain-of-replicas. The atomic weights of the migrating protons were scaled by a factor of 50 in order to capture continuous proton transfers along the pathway trajectory. A kinetic energy potential force constant of 0.05 kcal mol⁻¹ Å⁻² was applied for all MEP optimizations to facilitate the smoothness and convergence of the transition path. All 800 Toho/AMP and Toho/CEX acylation reaction MEPs were subjected to single-point calculations at the B3LYP-D3/6-31+G**/C36 level of theory. These 800 MEPs (40,000 replica conformations and energies) were grouped into four datasets: the Toho/AMP: R1-AE pathways, the Toho/AMP: R2-AE pathways, the Toho/CEX: R1-AE pathways, and the Toho/CEX: R2-AE pathways (Figure S2).

Machine Learning. A proper selection of feature representation is critical to the performance and interpretability of the ML models. Various feature vectors built from conformational and/or physical descriptors have been proposed to accurately predict molecular properties or potentials.^{15,14} In this study, we note that the goal of the ML regressions is not only to predict the single-point total energies of the systems but also to bridge the conformational change with the energy evolution along the optimized MEPs, which mainly attributes to the displacement of the reacting atoms in the QM regions. Therefore, the initial selection of features covered (1) the atomic distances between the chemical-bonded and hydrogen-bonded heavy atoms in the QM region and (2) the hydrogen-donor/acceptor distances between the reacting hydrogens and surrounding QM heavy atoms. This initial selection of features was then subjected to a univariate analysis against the energy profiles using mutual information as the metric. In each pathway dataset, the atomic

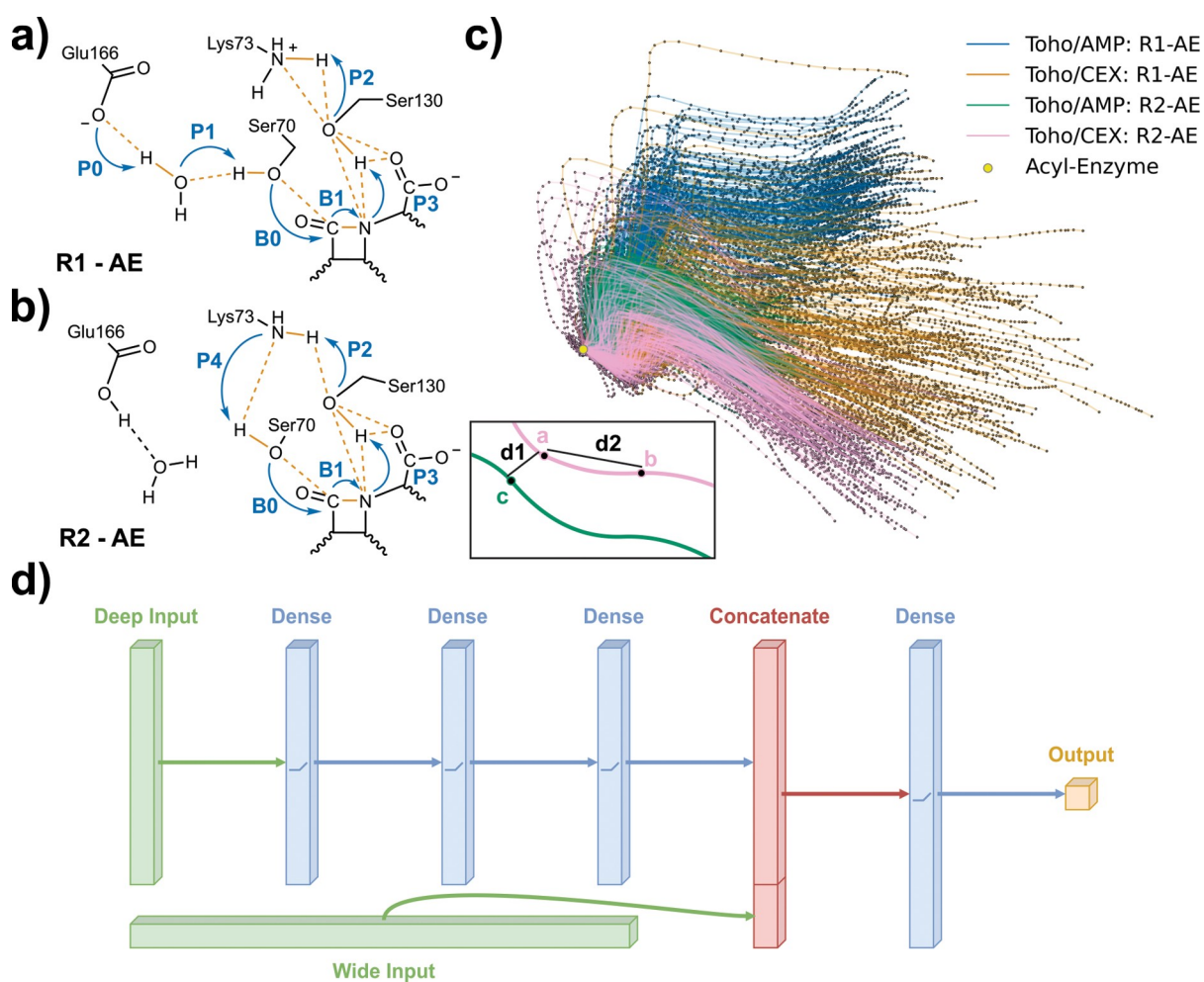


Figure 2. Selected features, the 2D representation of the pathway conformations, and the architecture of the ML-MEP models. The selected features and chemical processes of (a) Toho/AMP: R1-AE and Toho/CEX: R1-AE datasets and (b) Toho/AMP: R2-AE and Toho/CEX: R2-AE datasets. The atomic distances that are included in feature vectors are noted in orange lines and the chemical processes are noted in blue; see also Table S1. (c) 2D principal component dimensionality reduction of the pairwise inter-heavy-atom distances in the QM region, and a schematic demonstration for the loss of pathway context of the replicas; (d) architecture of the QM/MM MEP learning deep-and-wide neural network.

distances sharing higher mutual information than the intuitive reaction coordinates were included in the feature vectors. The features selected based on datasets of the same mechanism (e.g., R1-AE pathways for both Toho/AMP and Toho/CEX complexes) were merged to produce unified selections for either R1-AE or R2-AE mechanisms across different systems, respectively. The final selection of features on the R1-AE and R2-AE pathway datasets was illustrated in Figure 2a,b.

In order to reflect the conformational change along each MEP, the selected bond lengths as features were offset by subtracting their corresponding values in the acyl-enzyme conformation as the product state in the pathway. Furthermore, for each MEP, the feature vectors on the replicas are rescaled by dividing the largest values along each feature dimension

$$\mathbf{x}_{\text{new}}^p = \frac{\mathbf{x}^p}{\max(|\mathbf{x}^p|)} \quad (12)$$

where \mathbf{x}^p represents the feature vectors of all replicas on the p -th MEP. Similarly, for each MEP, the replica energy profiles were also offset using the energies of the acyl-enzyme product states as the reference (0 kcal mol⁻¹).

Notably, our preprocessing of input features enables the independent representation of conformational and energetic evolutions along each MEP: each pathway is treated individually without reference to other MEPs in the dataset. However, the largely reduced

feature representation also blurred the context of each replica, leading to the loss of pathway identity. As shown in Figure 2c, replicas that belong to different pathways could distribute closer than their neighboring replica on the same pathway ($d_1 < d_2$). Using the reduced feature representation, it is important to distinguish the pathway identity for replicas, especially those that are close to each other in the reduced feature space. In this regard, the distinguishability between different pathways should be retrieved by further hardcoding the pathway affinity of the replicas in a “one-hot” manner.

With these considerations, a feed-forward neural network with the “deep-and-wide” learning architecture (DaWNN) proposed by Cheng et al.⁷⁰ was implemented for the learning of the QM/MM MEPs (Figure 2d). The conformational features were fed to the deep side of the input layers and were embedded through four hidden layers. The encoded pathway identities were introduced from the wide entrance and were concatenated to the deep embedding flow before the last hidden layer. The single-point replica energies were obtained from a linear output layer. The rectifier linear unit was used to activate all hidden neuron nodes. The dropout strategy⁷¹ was applied for all hidden layers to promote the generalizability of the neural networks and prevent overfitting. The dropout rate (0.1) and the number of neuron units (256) on the hidden layers were tuned *via* a grid search strategy on a 10% pathway stratified validation set. Practically, we constructed the validation set by randomly picking five replicas from each of the pathways carrying 50 replicas. The standard mean squared error (MSE) was used as the objective loss function to train the ML-

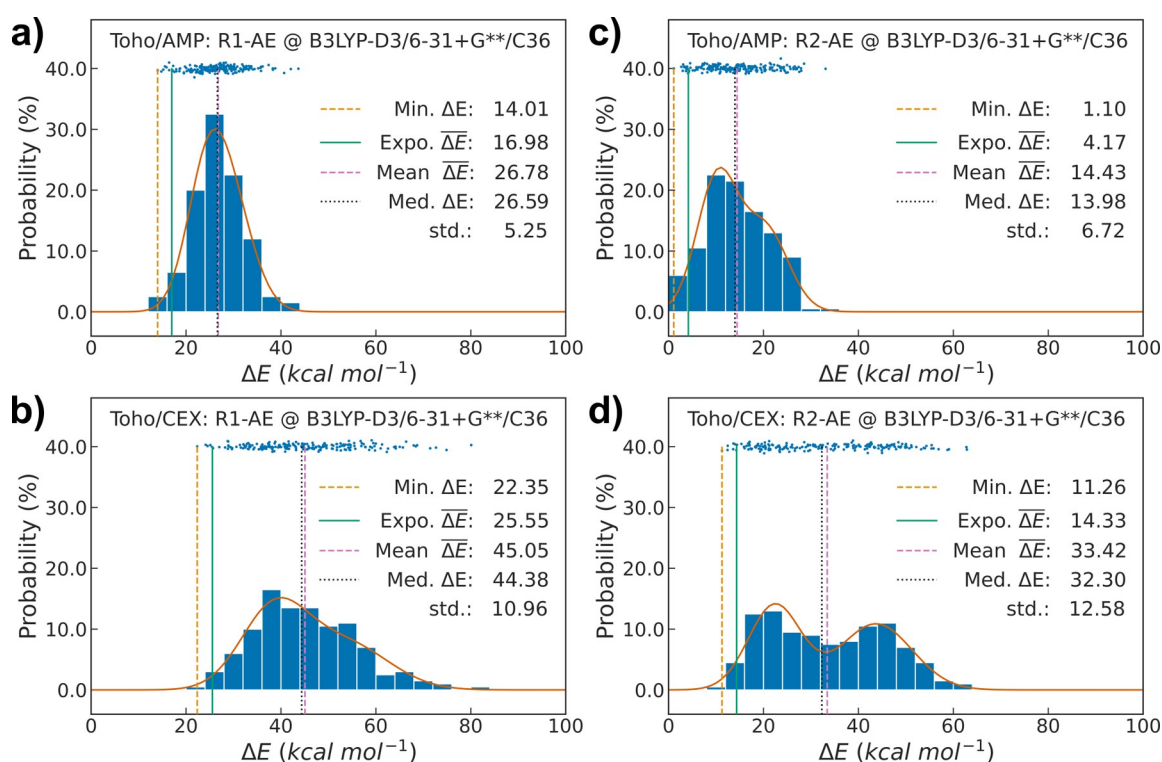


Figure 3. The distribution of the acylation barriers (ΔE) at B3LYP-D3/6-31+G**/C36 level of theory. (a) Toho/AMP: R1-AE acylation pathways; (b) Toho/CEX: R1-AE acylation pathways; (c) Toho/AMP: R2-AE acylation pathways; and (d) Toho/CEX: R2-AE acylation pathways. The scatters show the locations of the energy barriers. The width of the histograms is 4 kcal mol⁻¹. The red curves note the density estimation from the GMMs. The labels “min. ΔE ”, “expo. $\overline{\Delta E}$ ”, “mean $\overline{\Delta E}$ ”, “med. ΔE ”, and “std” refer to the minimum, exponential average, arithmetic average, median, and standard deviation of the energy barriers from the QM/MM MEP profiles, respectively.

MEP models. All models were trained with the AdaM optimizer⁷² for 300 epochs with a sample batch size of 25.

Implementation Details. All MD simulations were performed using CHARMM and OpenMM.^{73,74} All QM/MM calculations were performed with the built-in DFTB module of CHARMM or its interface to Q-Chem.^{75–77} The postprocessing of the pathway conformations, featurization, and the density estimating GMMs took advantage of the MDAnalysis and the scikit-learn package.^{78,79} The MEP-learning neural networks were implemented with TensorFlow.⁸⁰ All molecular graphics were prepared with UCSF ChimeraX.⁸¹

RESULTS

QM/MM MEPs

Pioneering mechanistic studies have suggested that the acylation half of the AS β L catalysis is the rate-limiting step of the β -lactam hydrolysis,^{35,36} allowing the comparison between the theoretical acylation barriers and the experimental k_{cat} values using the Eyring equation. According to the Toho-1 catalysis kinetic studies by Shimizu-Ibuka et al.,⁸² the acylation barrier of Toho/AMP is ~ 15.5 kcal mol⁻¹. Moreover, the enzyme kinetics reported by Nitani et al.⁸³ on Toho-1 with the Arg274Asn/Arg276Asn-engineered mutants showed that the acylation barrier of AMP is lower compared to that of CEX (determined from the $k_{\text{cat}}/k_{\text{M}}$ ratio) by ~ 1.7 kcal mol⁻¹. While both mutations were known to not participate in the acylation process, the acylation barrier of wild type Toho/CEX can be estimated as ~ 17.2 kcal mol⁻¹.

The barrier distributions of the calculated QM/MM MEPs are plotted in Figure 3. While the free-energy barriers of a certain reaction path could be obtained from the exponential

average of the barriers from a set of MEPs, the number of minimum energy barriers needed to achieve an accurate estimation is prohibitively high. According to the numerical simulations of Ryde,⁸⁴ if the barriers were Gaussian distributed, the number of pathway samples required for an accurate exponential averaged free-energy barrier is >5000 for all four pathways investigated in this study. Fortunately, the relative rankings of free-energy barriers for different mechanisms can be robustly estimated from the arithmetic averaged barrier. In particular, in cases where the difference between mean averages of the modeled systems is larger than ~ 4.8 kcal mol⁻¹, reliable comparisons of reactivity can be assessed from ~ 70 barrier samples with regard to the largest standard deviation of the pathway barriers in the current study (12.58 kcal mol⁻¹ of the Toho/CEX: R2-AE pathways, Figure 3d).⁸⁴ With 200 barrier samples for each mechanism, we accordingly rank the relative activity of the four mechanisms as Toho/AMP: R2-AE > Toho/AMP: R1-AE > Toho/CEX: R2-AE > Toho/CEX: R1-AE.

Both R1-AE and R2-AE acylation pathways of Toho/AMP are accessible as they show lower mean averaged barriers than the Toho/CEX ones. The Toho/AMP pathways also resemble the energetic landscapes of AS β L/penam acylation reported in previous studies.³⁷ The exponential-averaged acylation barrier is 16.98 kcal mol⁻¹ for the Toho/AMP: R1-AE pathway (Figure 3a), which is 12.81 kcal mol⁻¹ higher than the Toho/AMP: R2-AE pathway (4.17 kcal mol⁻¹, Figure 3c). The lowest Toho/AMP: R1-AE barrier is shown to be 14.01 kcal mol⁻¹, which is lower than the experimental acylation barriers of ~ 15.5 kcal mol⁻¹. As for Toho/CEX acylation, the pathway *via* the R2-AE mechanism confers an exponential-averaged barrier

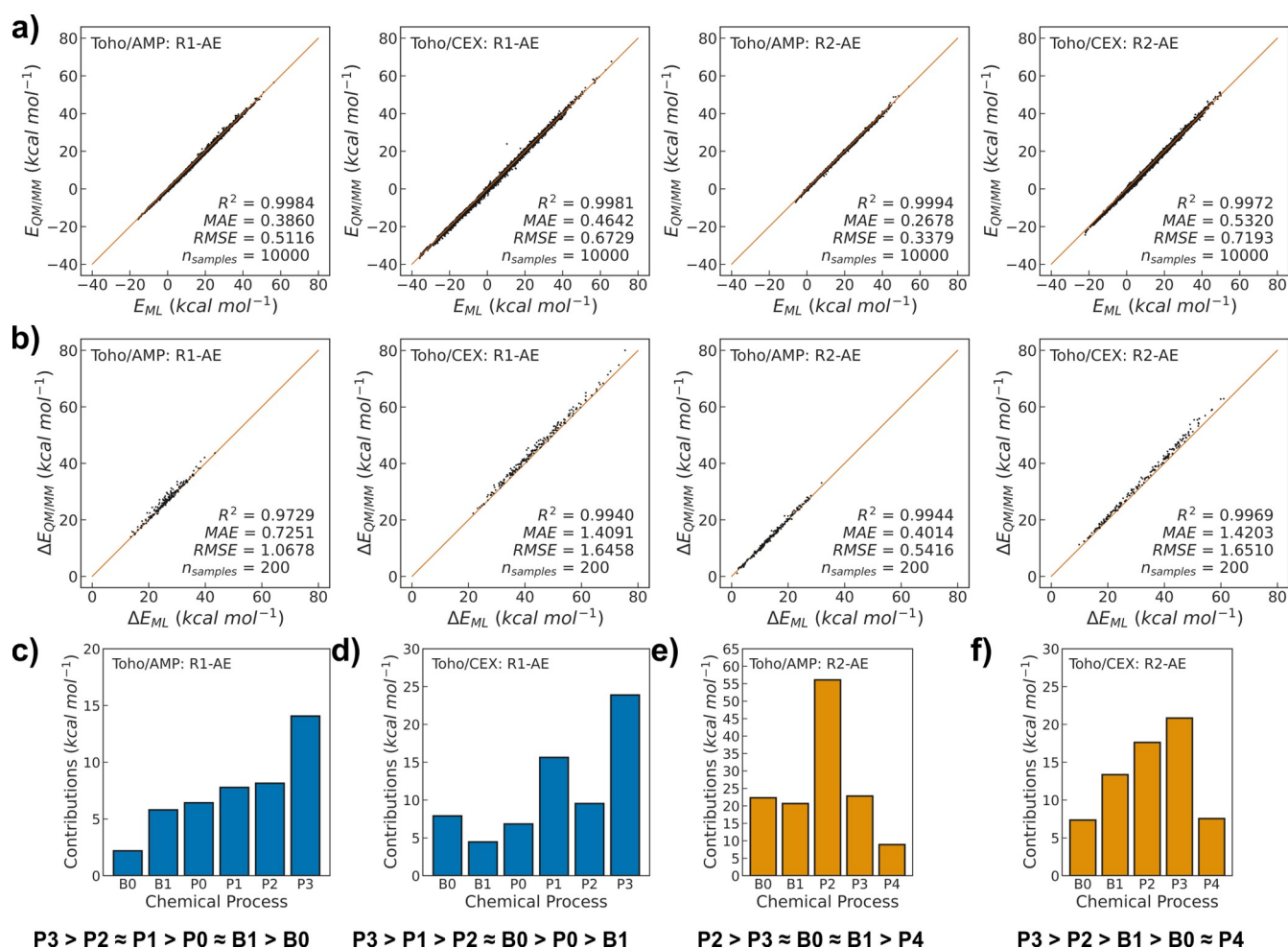


Figure 4. Predictive performance and the BCIG contributions of the ML-MEP models. The predictive performance of (a) replica energies and (b) pathway barriers of (left to right) the Toho/AMP: R1-AE, Toho/CEX: R1-AE, Toho/AMP: R2-AE, and Toho/CEX: R2-AE models. The BCIG contributions of the models: (c) Toho/AMP: R1-AE; (d) Toho/CEX: R1-AE; (e) Toho/AMP: R2-AE; and (f) Toho/CEX: R2-AE. The rankings (highest to lowest) of the BCIG contributions are noted below.

of 14.33 kcal mol⁻¹ (Figure 3d), which is 11.22 kcal mol⁻¹ lower than its R1-AE alternative (25.55 kcal mol⁻¹, Figure 3b). Furthermore, as the lowest energy barrier found on Toho/CEX: R1-AE pathways (22.35 kcal mol⁻¹) is much higher than the estimated experimental barrier (~17.2 kcal mol⁻¹), these pathways are considered generally inaccessible. The viable acylation path for Toho/CEX is therefore verified to be the R2-AE mechanism.⁴⁰

Energetics Interpreted from BCIG Contributions

In order to delineate the BCIG contributions of the chemical processes during the acylation, each MEP dataset was used to train an ML-MEP model to learn the correlation between the evolution of energetics and conformations along the MEPs.

The prerequisite for model expandability is that the ML/DL model should correctly resemble the underlying variable correlations, which also facilitate accurate predictions. The importance of accounting for the pathway identity is reflected in the predictive performances of our DaWNN architecture and conventional deep neural networks (DNN) with identical hyperparameters and training configurations. As shown in Figure 4a, the DaWNN models could accurately predict the replica-wise energy with R^2 scores >0.995. Meanwhile, the predictive performance of the DNN models dramatically decreases ($R^2 < 0.940$) in the absence of pathway distinguish-

ability (Figure S3). On the other hand, the DaWNN models could correctly predict the barrier heights with RMSE < 2 kcal mol⁻¹ (Figure 4b). While the improvement of including the pathway identity of each replica is evident, intuitively, the identity encoding extended the input feature space to higher dimensions and embedded the replicas into their respective pathways, allowing each pathway replica to be independently correlated to its own conformational/energetic context.

To interpret the ML-MEP models in terms of energetic contributions, we first grouped the input features (as atomic distances) by the chemical processes that they are involved in (Figure 2a,b, Table S1). The contributions of each chemical process are quantified using the proposed BCIG metrics. At this stage, the BCIG measurements are not ready to be quantitatively compared between the ML-MEP models that were trained from different pathway datasets (see below).

For the Toho/AMP: R1-AE pathways (Figure 4c), the highest BCIG contributions come from the concerted proton transfers from Lys73 to the thiazolidine nitrogen (P2 and P3). The proton abstraction of Glu166 carboxyl (P0 and P1) was assessed to be moderately rate-limiting as they pose higher contributions than the nucleophilic attacks of the Ser70 hydroxyl to the β -lactam carbonyl (B0 and B1). The contributions for the Toho/CEX: R1-AE pathways also

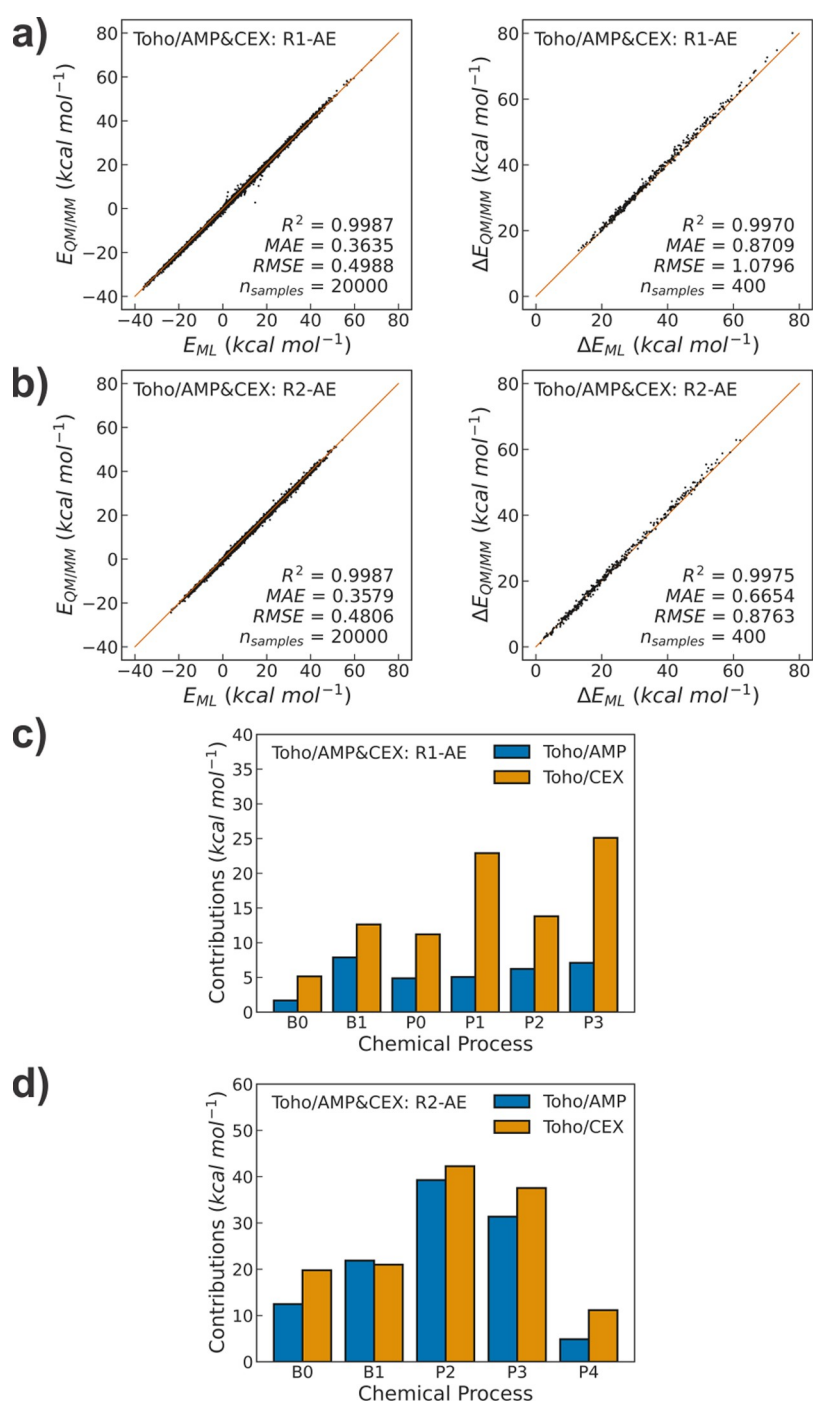


Figure 5. Predictive performance and the BCIG contributions of the unified ML-MEP models. The predictive performance of (left to right) the replica energies and the pathway barriers of (a) Toho/AMP&CEX: R1-AE and (b) Toho/AMP&CEX: R2-AE models. The BCIG contributions of (c) Toho/AMP&CEX: R1-AE and (d) Toho/AMP&CEX: R2-AE models.

suggest that the protonation of the cephem nitrogen (P2 and P3) is the determinant factor with the highest contribution (Figure 4d). However, the deprotonation of Ser70 O γ (P1) and its nucleophilic attach to the cephem carbonyl carbon (B0) were shown to considerably contribute to the reaction profiles of Toho/CEX: R1-AE. On the R2-AE acylation pathways, the concerted proton transfers from Lys73 to the β -lactam nitrogen, bridged by Ser130 hydroxyl (P2 and P3), remains as the reaction step of the highest BCIG contributions in both Toho/AMP and Toho/CEX systems (Figure 4e,f). Interestingly, in the Toho/AMP: R2-AE pathways, the highest

individual contribution comes from P2 (the proton transfer between Lys73 and Ser130), while in Toho/CEX: R2-AE, it was determined as P3 (the protonation of the cephem nitrogen). In summary, on all reactive pathways (Figure 4c–f), the protonation of the β -lactam nitrogen (P2 and P3) is deemed to be the highest energy contribution, in agreement with our previous investigations on TEM-1/benzylpenicillin acylation.⁸⁵

Noteworthy, even if the contributions between two systems cannot be directly compared, the ordering of the contributions from various chemical processes could qualitatively hint at the

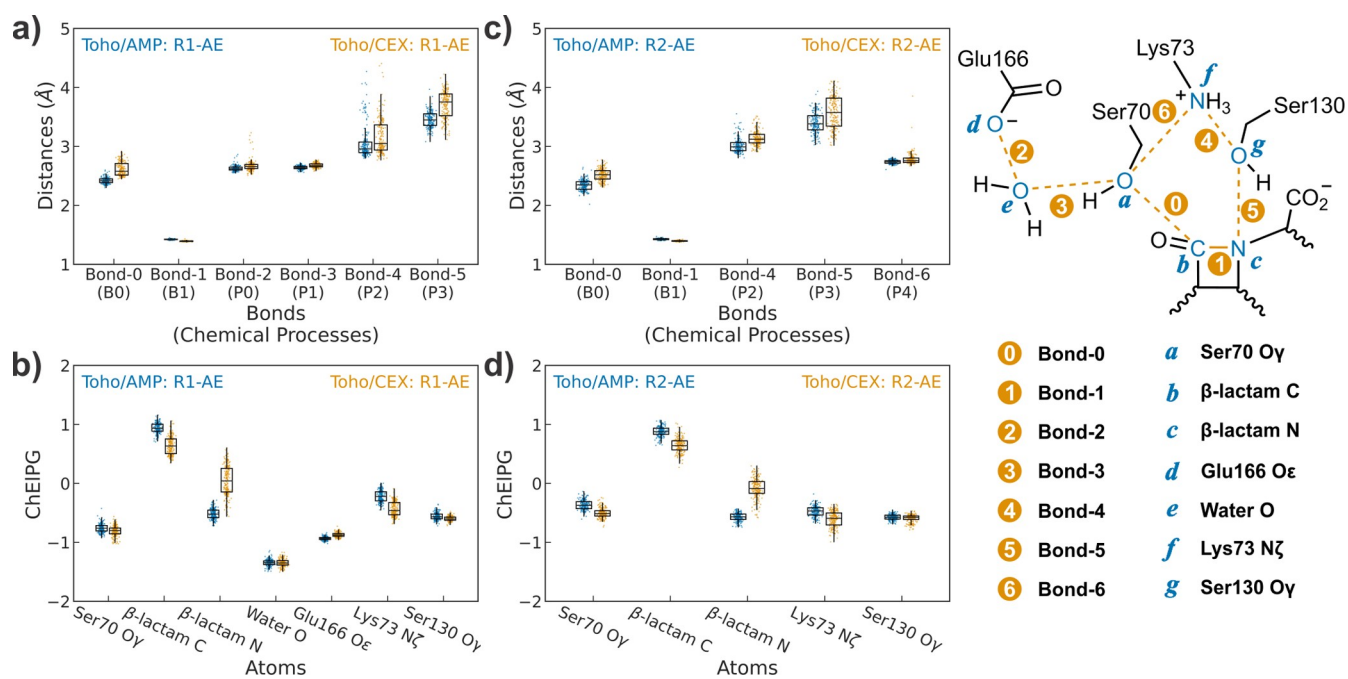


Figure 6. Atomic distances between the ChElPG charges on critical heavy atoms in all reactant states at the B3LYP-D3/6-31+G**/C36 level. The (a) atomic distances and (b) ChElPG charge profiles of the Toho/AMP: R1-AE and Toho/CEX: R1-AE pathways. The (c) atomic distances and (d) ChElPG charge profiles of the Toho/AMP: R2-AE and Toho/CEX: R2-AE pathways. The meanings of the labels on x -axes are defined in the symbolic legend (right panel).

major factors that differentiate the acylation kinetics of different systems. Specifically, the ranking of contributions from the nucleophilic attack of Ser70 O γ (B0 and P1) are higher for the Toho/CEX: R1-AE pathways (Figure 4d) than that of the Toho/AMP: R1-AE pathways (Figure 4c). This suggests that the formation of the tetrahedral intermediates for the CEX acylation is one of the main causes for its decreased acylation activity. On the other hand, the ranked order of BCIG contributions on the R2-AE pathways suggests that the tetrahedral collapsing (B1 and P3) is the major source of the kinetic differences, as P3 and B1 are respectively ranked the second- and the fourth-highest BCIG contributions in the Toho/AMP contributions (Figure 4e), and increase to the first and the third as in the Toho/CEX contributions (Figure 4f), respectively. These implications are related to the BCIG-explained reactivities.

Reactivity Interpreted from BCIG Contributions

As aforementioned, the BCIG contributions cannot be compared among different models trained from different MEP datasets due to that the distributions of input features from different pathway datasets are not guaranteed to be identical. In this case, the BCIG contributions were incompatible with each other as the models from which the contributions were derived do not share the same input space (Figure 2c). A solution is to derive BCIG contributions from a new model that is trained on both datasets, which allows the ML model to extract knowledge from the union input space of different systems. Therefore, the pathway datasets of the same mechanism were merged to produce the Toho/AMP&CEX: R1-AE and Toho/AMP&CEX: R2-AE datasets. Additional ML-MEP models with the DaWNN architecture were trained on the combined datasets with extended pathway identity encodings. We note that the mixed ML-MEP models could deep-embed all the input MEP samples through their hidden

layers from both systems, adapting their trainable weights to the mixing of training samples from different sources. While this mixed embedding does not impact the prediction accuracy of the model, from the perspective of sample affiliations (Toho/AMP or Toho/CEX), the model's knowledge on the learned PES landscape of one system is likely biased. In other words, when the DaWNNs are trained from a dataset with MEPs from different systems, the BCIG-explained reactivity contribution of the same feature from one system (e.g., the Toho/AMP: R1-AE pathways) can be compared to its contribution to the MEPs of the other system (the Toho/CEX: R1-AE pathways), as the knowledge landscape of the ML-MEP models is aware of the coexistence and the merged sample distribution of both systems in the training data. Meanwhile, the reactivity contributions of different features from the same system cannot be compared because the contributions of all feature dimensions in one system are biased by the other.

With the above awareness, we first examined the predictive performance of the ML-MEP models trained on the unified datasets. As shown in Figure 5a,b, the model prediction on the replica energies or the pathway barriers remain at the same predictive accuracies (RMSEs < 2 kcal mol⁻¹ and R² > 0.995) compared with the models trained on individual datasets (Figure 4a,b). The scalability of the DaWNN architecture for learning MEPs is demonstrated.

The BCIG contributions are computed for the reactivity-explaining models: Toho/AMP&CEX: R1-AE and Toho/AMP&CEX: R2-AE. For the R1-AE pathways, the BCIG contributions to all chemical processes are much higher in the Toho/CEX acylation MEPs than the Toho/AMP ones (Figure 5c). As expected, the correlated P2 and P3 contributions largely increase for the Toho/CEX R1-AE acylation pathways, reflecting the less active protonation of the cephem nitrogen. In addition, the nucleophilic serine attack (B0 and P1) in

Toho/CEX is more than three times higher than that in Toho/AMP, which aligns with the increment of rankings of B0 and P1 for the energetic contribution in Toho/CEX (Figure 4c,d). While the acylation pathways are initialized by the nucleophilic serine addition, the high BCIG contributions attributed to this process indicate that the R1-AE acylation pathways for Toho/CEX are unfavored compared to Toho/AMP. Combining the enzyme kinetics discussed above, the interpretation of BCIG contributions shows that the acylation pathway using solely Glu166 as the general base is turned off for Toho/CEX due to its incapability to activate the serine attack on the cephem carbonyl.

On the other hand, the contributions in the Toho/AMP&CEX: R2-AE models demonstrated the same trend (Figure 5d): the BCIG values for most chemical processes in Toho/CEX pathways are higher than those of the Toho/AMP pathways. The differences of the BCIG contributions in the two systems mainly come from the residue-involved processes: B0, P4 (the serine nucleophilic attack to β -lactam carbonyl), and P2, P3 (the concerted proton transfers to protonate the β -lactam nitrogen). Interestingly, the BCIG contribution from C–N bond breaking (B1) for Toho/CEX concerted base acylation is shown to be slightly lower than that for Toho/AMP. In brief, the BCIG contributions from the Toho/AMP&CEX: R2-AE MEP learning model show that the energy contributions of concerned chemical processes on the Toho/CEX pathways are moderately higher than the Toho/AMP pathways, suggesting lower acylation activity for the Toho/CEX: R2-AE pathways.

Validation of the BCIG Explanations

The mechanistic insights into the Toho/ β -lactam energetic and reactivity contributions derived from the ML-MEP models using the BCIG metric are further validated by additional QM/MM protocols. Specifically, we measured the atomic distances between critical heavy atoms and performed the ChEIPG population analysis⁸⁶ on the 800 reactant conformations in all MEPs.

We first validate the BCIG-explained energy contribution for each system. For the Glu166 mediated acylation pathways (R1-AE), the concerted proton transfers from Lys73 to the β -lactam nitrogen (P2 and P3) are determined as the reaction steps of the highest energy contributions in both Toho/AMP and CEX pathways (Figure 4c,d). The proton transfer processes in enzyme catalysis are known to highly correlate with the strength of the hydrogen-bonding interactions between the proton donor and the acceptor. Here, the atomic distances among Lys73 N ζ , Ser70 O γ , and the β -lactam nitrogen (P2 and P3) are shown to be the largest compared to that of other proton transfers (P0 and P1, Figure 6a), suggesting the weaker the hydrogen actions, the harder the proton transfers. A similar conclusion could also be drawn on the R2-AE acylation pathways (Figure 4e,f), where the P2/P3 atomic distances are larger than that of P4 (Figure 6c).

The BCIG-explained reactivities generally state that most of the chemical processes in Toho/AMP acylation pathways are more reactive than the Toho/CEX ones, regardless of the acylation mechanism. For the R1-AE acylation pathways that are prohibitive for Toho/CEX, it is shown that the atomic distances between the Lys73 amino and the Ser130 hydroxyl (P2) and the Ser130 hydroxyl to the cephem nitrogen (P3) are generally longer than those in Toho/AMP (Figure 6a). Moreover, the ChEIPG charges on the β -lactam nitrogen

atoms in CEX are much higher than that of the penam nitrogen of AMP (Figure 6b). These suggest the poor proton affinity of the CEX nitrogen atoms, which lead to higher BCIG contributions of P2 and P3 observed in Figure 5c. The contributions from the Ser70 addition processes (B0, P1) are another source of the deactivation of R1-AE acylation pathways for CEX. The key reactant distances in the two systems also suggest that the serine attack in CEX is less favorable than that of AMP due to the longer B0 distance in Toho/CEX (Figure 6a). Meanwhile, the ChEIPG analysis shows no significant change in the density population of Ser70 O γ upon binding of different ligands, but the atomic charges on the CEX carbonyl carbon are found to decrease (Figure 6b). Such electrostatic features are expected for the cephem scaffolds as the extended π -conjugations introduced by the C3=C4 double bond could delocalize the densities on the β -lactam carbonyl and nitrogen, leading to stronger resistance to the nucleophilic serine addition for CEX acylation. In the BCIG-explained ML-MEPs, this is reflected as the increased contributions from B0 and P1 observed in Toho/CEX: R1-AE pathways (Figure 5c).

The general mechanistic insights made for the different reactivity in R1-AE acylation also hold for the R2-AE acylation pathways: the cephem scaffold delocalizes the density on the β -lactam, promoting its resistance to the protonation of the dihydrothiazine nitrogen (P2, P3, Figure 6c) and the nucleophilic serine addition (B0, P4, Figure 6d). As for the C–N bond breaking (B1) in different systems, the BCIG contribution from the pairwise distance between two bonded atoms refers *not* to the bond strength between the atoms but the impact to the energy profiles when perturbations are applied to the progress of that distance (eq 3). Specifically, the BCIG contribution of B1 in the R2-AE hybrid models being close (Figure 5d) does not indicate that the “bond strengths” of the scissile C–N bonds are similar. Instead, it suggests that the “sensitivity” of the energetics of the QM/MM MEPs to the perturbation of the atomic distance between the carbonyl carbon and the nitrogen are similar. This is further supported by the reactant conformations and the atomic charges. As shown in Figure 6c, the β -lactam C–N bond distances are generally similar in different systems; it is expected that the BCIG contributions of B1 are close in Toho/CEX: R2-AE and Toho/AMP: R2-AE. On the other hand, the distributions of ChEIPG charges of the bonded atoms (the β -lactam C, N) largely deviate between the two systems (Figure 6d), suggesting very different densities and thus the “bond strength” of the C–N scissile bonds in Toho/AMP and Toho/CEX systems. Alternatively, the relative “bond strength” of the β -lactam carbonyl nitrogen is reflected from the contribution of the serine attack (B0) and the nitrogen protonation (P3), which are concerted processes to the ring-opening (B1).

DISCUSSIONS

In this work, we propose the contribution attribution method, BCIG, as an XAI technique to interpret ML-MEP models for mechanistic insights into enzyme catalysis. Using the BCIG contributions, we demonstrate how to extract interpretable mechanistic information that is difficult to quantify due to the concerted nature of the enzymatic reaction steps.

The BCIG method is helpful to discern the thermodynamics that governs all chemical transitions. However, since the ML-MEP models are trained based on the limited samples of MEPs, the direct application of the Boltzmann factor to the

cumulative IG could be limited by the size of the training set. Therefore, a PDF was introduced to smoothen the density of barriers distributions. Pioneering efforts have extensively applied equivalent density smoothening assumptions, such as Gaussian functions, for investigating or promoting the numerical convergence of free-energy methods.^{84,87,88} Moreover, von der Esch et al.⁸⁹ reported that the adiabatic mapped desuccinylation barriers of Sirtuin five could reasonably resemble a Gaussian distribution with ~ 150 samples. In this regard, we further examined the applicability of Gaussian distributions of our MEP barriers by a goodness-of-fit test. It was shown that Toho/CEX: R2-AE barriers failed to approximate a reasonable Gaussian distribution (Table S2). However, we note that the type of the PDFs does not seem to qualitatively impact the BCIG results (Table S3, Figure S4). Furthermore, numerical features of BCIG are characterized with a bootstrapping-based convergence test (Supporting Information Methods). We note that the BCIG contributions in all four pathway datasets converge when ~ 170 MEPs were used (Figure S5).

Moreover, our BCIG approach inherits the advantage of “model implementation invariance” from the IG formalism of Sundararajan et al.,²⁹ with the contributions attributed to the feature inputs being independent of the ML/DL algorithms used. Based on the model implementation invariance, if two models predict the same output on all input examples, the gradients of the predicted hypersurface should also be equal everywhere, and thus the invariance of the IG and the BCIG contributions. On the other hand, it is notable that the “path implementation invariance” of the IG approach is not preserved in the BCIG formalism for obvious reasons: it is vital to account for the contribution of the chemical process along a specific transition path (MEPs as in the current study). Thus, the gradients calculated for each replica along the path are cumulatively summed to achieve path specificity for the feature contributions.

It is also worth mentioning that our explanations are purpose-oriented. It is preferred to build individual ML-MEP models using datasets for distinct chemical processes, such as Toho/AMP acylation pathways or Toho/CEX acylation pathways in this study. Therefore, we intentionally trained multiple ML-MEP models to learn different aspects of Toho-1-catalyzed acylation. Practically, the energetic contributions in each acylation pathway dataset are interpreted from models that learn only on a particular pathway dataset, and the reactivity of the chemical processes between different systems is interpreted from models trained on per mechanism merged datasets.

We further note the extensibility of the current workflow based on the BCIG approach. In terms of data preparation, any MEP sampling method that effectively explores the transition region could be applied to construct the MEP datasets. Thanks to the “model implementation invariance,” the ML/DL algorithm used to learn the chemical MEPs are not limited to the DaWNN as we used. However, it is important to monitor the transformations applied to the model input and rationally design the learning strategies. Herein, we sequentially applied feature selections and pathway-wise feature rescaling, resulting in the largely dimensionality-reduced feature space and the loss of the pathway context for each replica. Accordingly, this is solved by hardcoding the pathway identity, which leads to the DaWNN architecture. Lastly, the *c* term to be explained is not limited to “chemical processes” as

in the current study. The only requirement for the representation of *c* is that it should describe a certain interaction of interest, which is generally open to any physical descriptors that are known to correlate with chemical transitions.

CONCLUSIONS

In this study, we first presented a QM/MM computational workflow that achieves fast sampling of QM/MM MEPs for enzyme catalysis. Based on the observation that the 3OB-f DFTB3 parameter set could correctly resemble the bond dissociation distances during the β -lactam acylation reaction, we optimized 800 MEP conformations and refined the single-point energies using B3LYP-D3/6-31+G**/C36 calculations. The energetics from this computational workflow are in good agreement with previous calculations based on DFT/MM calculations.⁴⁰

ML-MEP models with high performance and scalability using the DaWNN architecture were developed for the ML of the QM/MM MEPs of enzyme catalysis. Compared to conventional DNN models, the DaWNN architecture achieves much higher accuracy in learning the energetic profiles from the conformational evolutions along the QM/MM MEPs. Furthermore, the DaWNN model is shown to be highly scalable to the training size or the source of the training data without significant loss in performance.

Inspired by the IG approach for explaining ML/DL models, we further developed the BCIG approach to interpreting the ML-MEP models for mechanistic insights into enzyme catalysis. Using Toho/AMP and Toho/CEX as the model systems, the energetic and the reactivity contributions of the processes with different substrates are quantified by the BCIG attributions. The conformational factors that differentiate the Toho-1 acylation activities of AMP and CEX were identified. The BCIG contributions quantified that the cephem scaffold was less susceptible to the nucleophilic serine addition and the protonation of the β -lactam nitrogen than the penam. Moreover, we presented a purpose-oriented training-explaining strategy to focus on mode interpretability. Whereas the different ML-MEP models are trained and interpreted for specific mechanistic aspects, we have shown that the interpretations of different models give consistent mechanistic insights that agree with our intuitive mechanistic understandings and the validating QM/MM calculations on the modeled systems.

With the developments of robust ML/DL-based QM/MM potential functions and the implementations of thermodynamics-aware XAI approaches for protein-related systems, one could foresee the emergence of the “full-stack” (simulation to analysis) ML/DL-assisted QM/MM studies on enzyme catalytic mechanisms in the near future.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsphyschemau.2c00005>.

Technical details of the convergence test on BCIG; tables of notations of chemical processes, statistical normality tests, and BCIG results using different PDFs; figures of the benchmark QM/MM MEP calculations, all QM/MM MEPs, predictive performance of DNN

models, BCIG results using Gaussian function as the PDF, and the results of the convergence test (PDF)

AUTHOR INFORMATION

Corresponding Author

Peng Tao – Department of Chemistry, Center for Research Computing, Center for Drug Discovery, Design, and Delivery (CD4), Southern Methodist University, Dallas, Texas 75205, United States; orcid.org/0000-0002-2488-0239; Email: ptao@smu.edu

Authors

Zilin Song – Department of Chemistry, Center for Research Computing, Center for Drug Discovery, Design, and Delivery (CD4), Southern Methodist University, Dallas, Texas 75205, United States; orcid.org/0000-0002-3127-3971

Francesco Trozzi – Department of Chemistry, Center for Research Computing, Center for Drug Discovery, Design, and Delivery (CD4), Southern Methodist University, Dallas, Texas 75205, United States; orcid.org/0000-0001-5538-9189

Hao Tian – Department of Chemistry, Center for Research Computing, Center for Drug Discovery, Design, and Delivery (CD4), Southern Methodist University, Dallas, Texas 75205, United States; orcid.org/0000-0002-0186-9811

Chao Yin – Department of Chemistry, Center for Research Computing, Center for Drug Discovery, Design, and Delivery (CD4), Southern Methodist University, Dallas, Texas 75205, United States

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acspchemau.2c00005>

Author Contributions

Z.S.: conceptualization, methodology, software, investigation, formal analysis, validation, visualization, data curation, writing—original draft, writing—review and editing; F.T.: formal analysis, validation, writing—review and editing; H.T.: visualization; C.Y.: writing—review and editing; P.T.: supervision, conceptualization, writing—review and editing, project administration, and funding acquisition.

Notes

The authors declare no competing financial interest.

All (intermediate) data in the current study, including coordinates, energies, ChElPG charges, and featurized arrays, are available at: doi.org/10.5281/zenodo.5701371. The python codes developed for calculating the BCIG contributions are available at: github.com/smutaogroup/bcig_toho.

ACKNOWLEDGMENTS

The authors are grateful to Prof. Timothy Palzkill of BCM for discussions on the catalytic mechanisms of AS β LS. The authors are also grateful to Dr. Xiaoliang Pan and Prof. Yihan Shao of OU for fruitful discussions. This material is based upon work supported by the National Science Foundation under a CAREER grant no. 1753167. Computational time was provided by the Southern Methodist University's Center for Research Computing.

REFERENCES

- (1) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine Learning for Molecular and Materials Science. *Nature* **2018**, *559*, 547–555.
- (2) Vamathevan, J.; Clark, D.; Czodrowski, P.; Dunham, I.; Ferran, E.; Lee, G.; Li, B.; Madabhushi, A.; Shah, P.; Spitzer, M.; Zhao, S. Applications of Machine Learning in Drug Discovery and Development. *Nat. Rev. Drug Discovery* **2019**, *18*, 463–477.
- (3) Behler, J. Four Generations of High-Dimensional Neural Network Potentials. *Chem. Rev.* **2021**, *121*, 10037–10072.
- (4) Keith, J. A.; Vassilev-Galindo, V.; Cheng, B.; Chmiela, S.; Gastegger, M.; Müller, K.-R.; Tkatchenko, A. Combining Machine Learning and Computational Chemistry for Predictive Insights into Chemical Systems. *Chem. Rev.* **2021**, *121*, 9816–9872.
- (5) Li, Z.; Kermode, J. R.; De Vita, A. Molecular Dynamics with On-the-Fly Machine Learning of Quantum-Mechanical Forces. *Phys. Rev. Lett.* **2015**, *114*, 096405.
- (6) Schütt, K. T.; Arbabzadah, F.; Chmiela, S.; Müller, K. R.; Tkatchenko, A. Quantum-Chemical Insights from Deep Tensor Neural Networks. *Nat. Commun.* **2017**, *8*, 13890.
- (7) Bleiziffer, P.; Schaller, K.; Riniker, S. Machine Learning of Partial Charges Derived from High-Quality Quantum-Mechanical Calculations. *J. Chem. Inf. Model.* **2018**, *58*, 579–590.
- (8) Smith, J. S.; Nebgen, B. T.; Zubatyuk, R.; Lubbers, N.; Devereux, C.; Barros, K.; Tretiak, S.; Isayev, O.; Roitberg, A. E. Approaching Coupled Cluster Accuracy with a General-Purpose Neural Network Potential through Transfer Learning. *Nat. Commun.* **2019**, *10*, 2903.
- (9) Wang, J.; Olsson, S.; Wehmeyer, C.; Pérez, A.; Charron, N. E.; de Fabritiis, G.; Noé, F.; Clementi, C. Machine Learning of Coarse-Grained Molecular Dynamics Force Fields. *ACS Cent. Sci.* **2019**, *5*, 755–767.
- (10) Bonk, B. M.; Weis, J. W.; Tidor, B. Machine Learning Identifies Chemical Characteristics That Promote Enzyme Catalysis. *J. Am. Chem. Soc.* **2019**, *141*, 4108–4118.
- (11) Noé, F.; Olsson, S.; Köhler, J.; Wu, H. Boltzmann Generators: Sampling Equilibrium States of Many-Body Systems with Deep Learning. *Science* **2019**, *365*, No. eaaw1147.
- (12) Dral, P. O.; Owens, A.; Dral, A.; Csányi, G. Hierarchical Machine Learning of Potential Energy Surfaces. *J. Chem. Phys.* **2020**, *152*, 204110.
- (13) Han, J.; Zhang, L.; Car, R. Deep Potential: A General Representation of a Many-Body Potential Energy Surface. *Comput. Phys. Commun.* **2018**, *23*, 629.
- (14) Hansen, K.; Biegler, F.; Ramakrishnan, R.; Pronobis, W.; von Lilienfeld, O. A.; Müller, K.-R.; Tkatchenko, A. Machine Learning Predictions of Molecular Properties: Accurate Many-Body Potentials and Nonlocality in Chemical Space. *J. Phys. Chem. Lett.* **2015**, *6*, 2326–2331.
- (15) Hansen, K.; Montavon, G.; Biegler, F.; Fazli, S.; Rupp, M.; Scheffler, M.; von Lilienfeld, O. A.; Tkatchenko, A.; Müller, K.-R. Assessment and Validation of Machine Learning Methods for Predicting Molecular Atomization Energies. *J. Chem. Theory Comput.* **2013**, *9*, 3404–3419.
- (16) Glielmo, A.; Husic, B. E.; Rodriguez, A.; Clementi, C.; Noé, F.; Laio, A. Unsupervised Learning Methods for Molecular Simulation Data. *Chem. Rev.* **2021**, *121*, 9722–9758.
- (17) Bertoni, M.; Duran-Frigola, M.; Badia-i-Mompel, P.; Pauls, E.; Orozco-Ruiz, M.; Guitart-Pla, O.; Alcalde, V.; Diaz, V. M.; Berenguer-Llgero, A.; Brun-Heath, I.; Villegas, N.; de Herrerros, A. G.; Aloy, P. Bioactivity Descriptors for Uncharacterized Chemical Compounds. *Nat. Commun.* **2021**, *12*, 3932.
- (18) Trozzi, F.; Wang, X.; Tao, P. UMAP as a Dimensionality Reduction Tool for Molecular Dynamics Simulations of Biomacromolecules: A Comparison Study. *J. Phys. Chem. B* **2021**, *125*, 5022–5034.
- (19) Zhang, Y.-J.; Khorshidi, A.; Kastlunger, G.; Peterson, A. A. The Potential for Machine Learning in Hybrid QM/MM Calculations. *J. Chem. Phys.* **2018**, *148*, 241740.

- (20) Shen, L.; Yang, W. Molecular Dynamics Simulations with Quantum Mechanics/Molecular Mechanics and Adaptive Neural Networks. *J. Chem. Theory Comput.* **2018**, *14*, 1442–1455.
- (21) Zhang, P.; Shen, L.; Yang, W. Solvation Free Energy Calculations with Quantum Mechanics/Molecular Mechanics and Machine Learning Models. *J. Phys. Chem. B* **2019**, *123*, 901–908.
- (22) Bösel, L.; Thürlmann, M.; Riniker, S. Machine Learning in QM/MM Molecular Dynamics Simulations of Condensed-Phase Systems. *J. Chem. Theory Comput.* **2021**, *17*, 2641–2658.
- (23) Pan, X.; Yang, J.; Van, R.; Epifanovsky, E.; Ho, J.; Huang, J.; Pu, J.; Mei, Y.; Nam, K.; Shao, Y. Machine-Learning-Assisted Free Energy Simulation of Solution-Phase and Enzyme Reactions. *J. Chem. Theory Comput.* **2021**, *17*, 5745–5758.
- (24) Gunning, D.; Stefik, M.; Choi, J.; Miller, T.; Stumpf, S.; Yang, G. Z. XAI—Explainable Artificial Intelligence. *Sci. Rob.* **2019**, *4*, No. eaay7120.
- (25) Murdoch, W. J.; Singh, C.; Kumbier, K.; Abbasi-Asl, R.; Yu, B. Definitions, Methods, and Applications in Interpretable Machine Learning. *Sci. Rob.* **2019**, *116*, 22071–22080.
- (26) Grömping, U. Variable Importance in Regression Models. *Wiley Interdiscip. Rev.: Comput. Stat.* **2015**, *7*, 137–152.
- (27) Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32.
- (28) Samek, W.; Montavon, G.; Lapuschkin, S.; Anders, C. J.; Müller, K.-R. Explaining Deep Neural Networks and Beyond: A Review of Methods and Applications. *Proc. IEEE* **2021**, *109*, 247–278.
- (29) Sundararajan, M.; Taly, A.; Yan, Q. Axiomatic Attribution for Deep Networks. In *Proceedings of the 34th International Conference on Machine Learning*; Precup, D., Teh, Y. W., Eds.; Proceedings of Machine Learning Research; PMLR, 2017; Vol. 70, pp 3319–3328.
- (30) Binder, A.; Bach, S.; Montavon, G.; Müller, K.-R.; Samek, W. Layer-Wise Relevance Propagation for Deep Neural Network Architectures. In *Information Science and Applications (ICISA) 2016*; Kim, K. J., Joukov, N., Eds.; Lecture Notes in Electrical Engineering; Springer Singapore: Singapore, 2016; Vol. 376, pp 913–922.
- (31) Yang, Y.; Tresp, V.; Wunderle, M.; Fasching, P. A. Explaining Therapy Predictions with Layer-Wise Relevance Propagation in Neural Networks. *2018 IEEE International Conference on Healthcare Informatics (ICHI)*; IEEE: New York, NY, 2018; pp 152–162.
- (32) Cho, H.; Lee, E. K.; Choi, I. S. Layer-Wise Relevance Propagation of InteractionNet Explains Protein–Ligand Interactions at the Atom Level. *Sci. Rep.* **2020**, *10*, 21155.
- (33) Gao, K. Y.; Fokoue, A.; Luo, H.; Iyengar, A.; Dey, S.; Zhang, P. Interpretable Drug Target Prediction Using Deep Neural Representation. *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*; International Joint Conferences on Artificial Intelligence Organization: Stockholm, Sweden, 2018; pp 3371–3377.
- (34) Karimi, M.; Wu, D.; Wang, Z.; Shen, Y. Explainable Deep Relational Networks for Predicting Compound–Protein Affinities and Contacts. *J. Chem. Inf. Model.* **2021**, *61*, 46–66.
- (35) Hermann, J. C.; Hensen, C.; Ridder, L.; Mulholland, A. J.; Höltje, H.-D. Mechanisms of Antibiotic Resistance: QM/MM Modeling of the Acylation Reaction of a Class A β -Lactamase with Benzylpenicillin. *J. Am. Chem. Soc.* **2005**, *127*, 4454–4465.
- (36) Hermann, J. C.; Ridder, L.; Höltje, H.-D.; Mulholland, A. J. Molecular Mechanisms of Antibiotic Resistance: QM/MM Modelling of Deacylation in a Class A β -Lactamase. *Org. Biomol. Chem.* **2006**, *4*, 206–210.
- (37) Meroueh, S. O.; Fisher, J. F.; Schlegel, H. B.; Mobashery, S. Ab Initio QM/MM Study of Class A β -Lactamase Acylation: Dual Participation of Glu166 and Lys73 in a Concerted Base Promotion of Ser70. *J. Am. Chem. Soc.* **2005**, *127*, 15397–15407.
- (38) Palzkill, T. Structural and Mechanistic Basis for Extended-Spectrum Drug-Resistance Mutations in Altering the Specificity of TEM, CTX-M, and KPC β -Lactamases. *Front. Mol. Biosci.* **2018**, *5*, 16.
- (39) Fisher, J. F.; Meroueh, S. O.; Mobashery, S. Bacterial Resistance to β -Lactam Antibiotics: Compelling Opportunism, Compelling Opportunity. *Chem. Rev.* **2005**, *105*, 395–424.
- (40) Song, Z.; Trozzi, F.; Palzkill, T.; Tao, P. QM/MM Modeling of Class A β -Lactamases Reveals Distinct Acylation Pathways for Ampicillin and Cefalexin. *Org. Biomol. Chem.* **2021**, *19*, 9182.
- (41) Brokaw, J. B.; Haas, K. R.; Chu, J.-W. Reaction Path Optimization with Holonomic Constraints and Kinetic Energy Potentials. *J. Chem. Theory Comput.* **2009**, *5*, 2050–2061.
- (42) Lee Woodcock, H.; Hodošček, M.; Sherwood, P.; Lee, Y. S.; Schaefer, H. F., III; Brooks, B. R. Exploring the Quantum Mechanical/Molecular Replica Path Method: A Pathway Optimization of the Chorismate to Prephenate Claisen Rearrangement Catalyzed by Chorismate Mutase. *Theor. Chem. Acc.* **2003**, *109*, 140–148.
- (43) Tao, P.; Hodošček, M.; Larkin, J. D.; Shao, Y.; Brooks, B. R. Comparison of Three Chain-of-States Methods: Nudged Elastic Band and Replica Path with Restraints or Constraints. *J. Chem. Theory Comput.* **2012**, *8*, 5035–5051.
- (44) Langan, P. S.; Vandavasi, V. G.; Weiss, K. L.; Cooper, J. B.; Ginell, S. L.; Coates, L. The Structure of Toho1 β -Lactamase in Complex with Penicillin Reveals the Role of Tyr105 in Substrate Recognition. *FEBS Open Bio* **2016**, *6*, 1170–1177.
- (45) Vanommeslaeghe, K.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. *J. Chem. Inf. Model.* **2012**, *52*, 3144–3154.
- (46) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *J. Chem. Inf. Model.* **2012**, *52*, 3155–3168.
- (47) Li, H.; Robertson, A. D.; Jensen, J. H. Very Fast Empirical Prediction and Rationalization of Protein PKa Values. *Proteins: Struct., Funct., Bioinf.* **2005**, *61*, 704–721.
- (48) Tomanicek, S. J.; Standaert, R. F.; Weiss, K. L.; Ostermann, A.; Schrader, T. E.; Ng, J. D.; Coates, L. Neutron and X-Ray Crystal Structures of a Perdeuterated Enzyme Inhibitor Complex Reveal the Catalytic Proton Network of the Toho-1 β -Lactamase for the Acylation Reaction. *J. Biol. Chem.* **2013**, *288*, 4715–4722.
- (49) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ , ψ and Side-Chain X1 and X2 Dihedral Angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257–3273.
- (50) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., Jr. CHARMM General Force Field: A Force Field for Drug-like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2010**, *31*, 671–690.
- (51) 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.
- (52) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An $N \cdot \log(N)$ Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (53) Gaus, M.; Cui, Q.; Elstner, M. DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB). *J. Chem. Theory Comput.* **2011**, *7*, 931–948.
- (54) Gaus, M.; Goez, A.; Elstner, M. Parametrization and Benchmark of DFTB3 for Organic Molecules. *J. Chem. Theory Comput.* **2013**, *9*, 338–354.
- (55) Fritz, R. A.; Alzate-Morales, J. H.; Spencer, J.; Mulholland, A. J.; van der Kamp, M. W. Multiscale Simulations of Clavulanate Inhibition Identify the Reactive Complex in Class A β -Lactamases and Predict the Efficiency of Inhibition. *Biochemistry* **2018**, *57*, 3560–3563.
- (56) Kulik, H. J.; Zhang, J.; Klinman, J. P.; Martinez, T. J. How Large Should the QM Region Be in QM/MM Calculations? The Case of Catechol O-Methyltransferase. *J. Phys. Chem. B* **2016**, *120*, 11381–11394.

- (57) Roßbach, S.; Ochsenfeld, C. Influence of Coupling and Embedding Schemes on QM Size Convergence in QM/MM Approaches for the Example of a Proton Transfer in DNA. *J. Chem. Theory Comput.* **2017**, *13*, 1102–1107.
- (58) Evans, D. J.; Holian, B. L. The Nose–Hoover Thermostat. *J. Chem. Phys.* **1985**, *83*, 4069–4074.
- (59) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant Pressure Molecular Dynamics Simulation: The Langevin Piston Method. *J. Chem. Phys.* **1995**, *103*, 4613–4621.
- (60) Zhou, Y.; Ojeda-May, P.; Nagaraju, M.; Kim, B.; Pu, J. Mapping Free Energy Pathways for ATP Hydrolysis in the E. Coli ABC Transporter HlyB by the String Method. *Molecules* **2018**, *23*, 2652.
- (61) Plotnikov, N. V.; Kamerlin, S. C. L.; Warshel, A. Paradynamics: An Effective and Reliable Model for Ab Initio QM/MM Free-Energy Calculations and Related Tasks. *J. Phys. Chem. B* **2011**, *115*, 7950–7962.
- (62) 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.
- (63) Becke, A. D. A New Mixing of Hartree–Fock and Local Density-functional Theories. *J. Chem. Phys.* **1993**, *98*, 1372–1377.
- (64) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian—Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1972**, *56*, 2257–2261.
- (65) Grimme, S. Semiempirical GGA-Type Density Functional Constructed with a Long-Range Dispersion Correction. *J. Comput. Chem.* **2006**, *27*, 1787–1799.
- (66) Yang, Z.; Twidale, R. M.; Gervasoni, S.; Suardiaz, R.; Colenso, C. K.; Lang, E. J. M.; Spencer, J.; Mulholland, A. J. Multiscale Workflow for Modeling Ligand Complexes of Zinc Metalloproteins. *J. Chem. Inf. Model.* **2021**, *61*, 5658.
- (67) Yagi, K.; Ito, S.; Sugita, Y. Exploring the Minimum-Energy Pathways and Free-Energy Profiles of Enzymatic Reactions with QM/MM Calculations. *J. Phys. Chem. B* **2021**, *125*, 4701–4713.
- (68) Kaiyawet, N.; Lonsdale, R.; Rungrotmongkol, T.; Mulholland, A. J.; Hannongbua, S. High-Level QM/MM Calculations Support the Concerted Mechanism for Michael Addition and Covalent Complex Formation in Thymidylate Synthase. *J. Chem. Theory Comput.* **2015**, *11*, 713–722.
- (69) Goyal, P.; Yang, S.; Cui, Q. Microscopic Basis for Kinetic Gating in Cytochrome c Oxidase: Insights from QM/MM Analysis. *Chem. Sci.* **2015**, *6*, 826–841.
- (70) Cheng, H.-T.; Koc, L.; Harmsen, J.; Shaked, T.; Chandra, T.; Aradhya, H.; Anderson, G.; Corrado, G.; Chai, W.; Ispir, M.; Anil, R.; Haque, Z.; Hong, L.; Jain, V.; Liu, X.; Shah, H. Wide & Deep Learning for Recommender Systems. *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems*; ACM: Boston, MA, USA, 2016; pp 7–10.
- (71) Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
- (72) Kingma, D. P.; Ba, J. Adam: A Method for Stochastic Optimization. 2017, arXiv:1412.6980. arXiv.org e-Print archive. <https://arxiv.org/abs/1412.6980> (accessed 3 March 2022).
- (73) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caffisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.
- (74) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLoS Comput. Biol.* **2017**, *13*, No. e1005659.
- (75) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. A QM/MM Implementation of the Self-Consistent Charge Density Functional Tight Binding (SCC-DFTB) Method. *J. Phys. Chem. B* **2001**, *105*, 569–585.
- (76) Woodcock, H. L.; Hodošček, M.; Gilbert, A. T. B.; Gill, P. M. W.; Schaefer, H. F., III; Brooks, B. R. Interfacing Q-Chem and CHARMM to Perform QM/MM Reaction Path Calculations. *J. Comput. Chem.* **2007**, *28*, 1485–1502.
- (77) Epifanovsky, E.; Gilbert, A. T. B.; Feng, X.; Lee, J.; Mao, Y.; Mardirossian, N.; Pokhilko, P.; White, A. F.; Coons, M. P.; Dempwolff, A. L.; Gan, Z.; Hait, D.; Horn, P. R.; Jacobson, L. D.; Kaliman, I.; Kussmann, J.; Lange, A. W.; Lao, K. U.; Levine, D. S.; Liu, J.; McKenzie, S. C.; Morrison, A. F.; Nanda, K. D.; Plasser, F.; Rehn, D. R.; Vidal, M. L.; You, Z.-Q.; Zhu, Y.; Alam, B.; Albrecht, B. J.; Aldossary, A.; Alguire, E.; Andersen, J. H.; Athavale, V.; Barton, D.; Begun, K.; Behn, A.; Bellonzi, N.; Bernard, Y. A.; Berquist, E. J.; Burton, H. G. A.; Carreras, A.; Carter-Fenk, K.; Chakraborty, R.; Chien, A. D.; Closser, K. D.; Cofer-Shabica, V.; Dasgupta, S.; de Wergifosse, M.; Deng, J.; Diedenhofen, M.; Do, H.; Ehlert, S.; Fang, P.-T.; Fatehi, S.; Feng, Q.; Friedhoff, T.; Gayvert, J.; Ge, Q.; Gidofalvi, G.; Goldey, M.; Gomes, J.; González-Espinoza, C. E.; Gulania, S.; Gunina, A. O.; Hanson-Heine, M. W. D.; Harbach, P. H. P.; Hauser, A.; Herbst, M. F.; Hernández Vera, M.; Hodecker, M.; Holden, Z. C.; Houck, S.; Huang, X.; Hui, K.; Huynh, B. C.; Ivanov, M.; Jász, Á.; Ji, H.; Jiang, H.; Kaduk, B.; Kähler, S.; Khistyayev, K.; Kim, J.; Kis, G.; Klunzinger, P.; Koczor-Benda, Z.; Koh, J. H.; Kosenkov, D.; Koulias, L.; Kowalczyk, T.; Krauter, C. M.; Kue, K.; Kunitsa, A.; Kus, T.; Ladžánszki, I.; Landau, A.; Lawler, K. V.; Lefrançois, D.; Lehtola, S.; Li, R. R.; Li, Y.-P.; Liang, J.; Liebenthal, M.; Lin, H.-H.; Lin, Y.-S.; Liu, F.; Liu, K.-Y.; Loipersberger, M.; Luenser, A.; Manjanath, A.; Manohar, P.; Mansoor, E.; Manzer, S. F.; Mao, S.-P.; Marenich, A. V.; Markovich, T.; Mason, S.; Maurer, S. A.; McLaughlin, P. F.; Menger, M. F. S. J.; Mewes, J.-M.; Mewes, S. A.; Morgante, P.; Mullinax, J. W.; Oosterbaan, K. J.; Paran, G.; Paul, A. C.; Paul, S. K.; Pavošević, F.; Pei, Z.; Prager, S.; Proynov, E. I.; Rák, Á.; Ramos-Cordoba, E.; Rana, B.; Rask, A. E.; Rettig, A.; Richard, R. M.; Rob, F.; Rossomme, E.; Scheele, T.; Scheurer, M.; Schneider, M.; Sergueev, N.; Sharada, S. M.; Skomorowski, W.; Small, D. W.; Stein, C. J.; Su, Y.-C.; Sundstrom, E. J.; Tao, Z.; Thirman, J.; Tornai, G. J.; Tsuchimochi, T.; Tubman, N. M.; Veccham, S. P.; Vydrov, O.; Wenzel, J.; Witte, J.; Yamada, A.; Yao, K.; Yeganeh, S.; Yost, S. R.; Zech, A.; Zhang, I. Y.; Zhang, X.; Zhang, Y.; Zuev, D.; Aspuru-Guzik, A.; Bell, A. T.; Besley, N. A.; Bravaya, K. B.; Brooks, B. R.; Casanova, D.; Chai, J.-D.; Coriani, S.; Cramer, C. J.; Cserey, G.; DePrince, A. E.; DiStasio, R. A.; Dreuw, A.; Dunitz, B. D.; Furlani, T. R.; Goddard, W. A.; Hammes-Schiffer, S.; Head-Gordon, T.; Hehre, W. J.; Hsu, C.-P.; Jagau, T.-C.; Jung, Y.; Klamt, A.; Kong, J.; Lambrecht, D. S.; Liang, W.; Mayhall, N. J.; McCurdy, C. W.; Neaton, J. B.; Ochsenfeld, C.; Parkhill, J. A.; Peverati, R.; Rassolov, V. A.; Shao, Y.; Slipchenko, L. V.; Stauch, T.; Steele, R. P.; Subotnik, J. E.; Thom, A. J. W.; Tkatchenko, A.; Truhlar, D. G.; Van Voorhis, T.; Wesolowski, T. A.; Whaley, K. B.; Woodcock, H. L.; Zimmerman, P. M.; Faraji, S.; Gill, P. M. W.; Head-Gordon, M.; Herbert, J. M.; Krylov, A. I. Software for the Frontiers of Quantum Chemistry: An Overview of Developments in the Q-Chem 5 Package. *J. Chem. Phys.* **2021**, *155*, 084801.
- (78) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J. Comput. Chem.* **2011**, *32*, 2319–2327.
- (79) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
- (80) Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G. S.; Davis, A.; Dean, J.; Devin, M. Tensorflow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. 2016, arXiv:1603.04467. arXiv.org e-Print archive. <https://arxiv.org/abs/1603.04467> (accessed 3 March 2022).

(81) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Meng, 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.

(82) Shimizu-Ibuka, A.; Oishi, M.; Yamada, S.; Ishii, Y.; Mura, K.; Sakai, H.; Matsuzawa, H. Roles of Residues Cys69, Asn104, Phe160, Gly232, Ser237, and Asp240 in Extended-Spectrum β -Lactamase Toho-1. *Antimicrob. Agents Chemother.* **2011**, *55*, 284–290.

(83) Nitanaï, Y.; Shimamura, T.; Uchiyama, T.; Ishii, Y.; Takehira, M.; Yutani, K.; Matsuzawa, H.; Miyano, M. The Catalytic Efficiency (K_{cat}/K_m) of the Class A β -Lactamase Toho-1 Correlates with the Thermal Stability of Its Catalytic Intermediate Analog. *Biochim. Biophys. Acta Protein Proteomics* **2010**, *1804*, 684–691.

(84) Ryde, U. How Many Conformations Need to Be Sampled to Obtain Converged QM/MM Energies? The Curse of Exponential Averaging. *J. Chem. Theory Comput.* **2017**, *13*, 5745–5752.

(85) Song, Z.; Zhou, H.; Tian, H.; Wang, X.; Tao, P. Unraveling the Energetic Significance of Chemical Events in Enzyme Catalysis via Machine-Learning Based Regression Approach. *Commun. Chem.* **2020**, *3*, 134.

(86) Breneman, C. M.; Wiberg, K. B. Determining Atom-Centered Monopoles from Molecular Electrostatic Potentials. The Need for High Sampling Density in Formamide Conformational Analysis. *J. Comput. Chem.* **1990**, *11*, 361–373.

(87) Boresch, S.; Woodcock, H. L. Convergence of Single-Step Free Energy Perturbation. *Mol. Phys.* **2017**, *115*, 1200–1213.

(88) Hu, W.; Li, P.; Wang, J.-N.; Xue, Y.; Mo, Y.; Zheng, J.; Pan, X.; Shao, Y.; Mei, Y. Accelerated Computation of Free Energy Profile at Ab Initio Quantum Mechanical/Molecular Mechanics Accuracy via a Semiempirical Reference Potential. 3. Gaussian Smoothing on Density-of-States. *J. Chem. Theory Comput.* **2020**, *16*, 6814–6822.

(89) von der Esch, B.; Dietschreit, J. C. B.; Peters, L. D. M.; Ochsenfeld, C. Finding Reactive Configurations: A Machine Learning Approach for Estimating Energy Barriers Applied to Sirtuin 5. *J. Chem. Theory Comput.* **2019**, *15*, 6660–6667.