Electronic Structure

PAPER

Graph-learning guided mechanistic insights into imipenem hydrolysis in GES carbapenemases

RECEIVED
2 April 2022

REVISED
6 June 2022

ACCEPTED FOR PUBLICATION
16 June 2022

PUBLISHED
30 June 2022

Zilin Song and 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 of America

* Author to whom any correspondence should be addressed.

E-mail: ptao@smu.edu

Keywords: class A β-lactamases, carbapenemases, GES-5, imipenem, QM/MM, minimum energy pathways, graph leaning

Supplementary material for this article is available online

Abstract

Pathogen resistance to carbapenem antibiotics compromises effective treatments of superbug infections. One major source of carbapenem resistance is the bacterial production of carbapenemases which effectively hydrolyze carbapenem drugs. In this computational study, the deacylation reaction of imipenem (IPM) by GES-5 carbapenemases (GES) is modeled to unravel the mechanistic factors that facilitate carbapenem resistance. Hybrid quantum mechanical/molecular mechanical (QM/MM) calculations are applied to sample the GES/IPM deacylation barriers on the minimum energy pathways (MEPs). In light of the recent emergence of graph-based deep-learning techniques, we construct graph representations of the GES/IPM active site. An edge-conditioned graph convolutional neural network (ECGCNN) is trained on the acyl-enzyme conformational graphs to learn the underlying correlations between the GES/IPM conformations and the deacylation barriers. A perturbative approach is proposed to interpret the latent representations from the graph-learning (GL) model and extract essential mechanistic understanding with atomistic detail. In general, our study combining QM/MM MEPs calculations and GL models explains mechanistic landscapes underlying the IPM resistance driven by GES carbapenemases. We also demonstrate that GL methods could effectively assist the post-analysis of QM/MM calculations whose data span high dimensionality and large sample-size.

1. Introduction

Bacterial resistance to antibiotic drugs compromises effective treatments of pathogen infections. The resistance to β-lactam drugs, especially carbapenems that are of great clinical importance, deserves particular interests due to the increasing infections of β-lactamases-producing superbugs [1, 2]. β-lactamases are bacterial-produced enzymes that are able to effectively hydrolyze and confer board resistance to the antibiotic family of β-lactams [3]. Based on the catalytic mechanism, β-lactamases are divided into four classes: A, B, C and D [4]. The class A serine-based β-lactamases (ASβLs) hydrolyze β-lactams via an acylation–deacylation process [5]. While carbapenem antibiotics could inhibit the deacylation half of the hydrolysis in many ASβLs, resistance conferred by novel ASβL-carbapenemases has also emerged [6]. In this regard, detailed mechanistic understanding of the ASβL-mediated carbapenem deacylation are vital for the design and optimization of future antibiotic drugs [7].

The GES family of ASβLs has demonstrated functional diversity that constitutes its broad resistance to β-lactams. The first clinically isolated GES-1 enzyme was reported to be an extended-spectrum β-lactamase with very low resistance to carbapenems, while its single mutant Gly170Ser variant (GES-5) was shown to effectively hydrolyze carbapenems such as imipenem (IPM, figure 1(a)) [8–14]. The general β-lactam deacylation mechanism has been extensively investigated by pioneering theoretical efforts [15–25]. Based on their hybrid quantum mechanical/molecular mechanical (QM/MM) mechanistic study, Hermann *et al* proposed

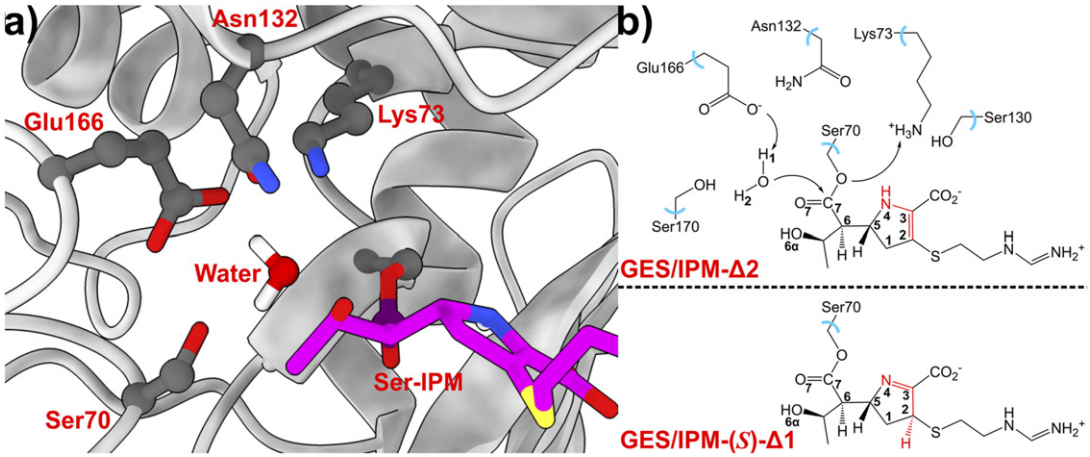

Figure 1. The GES/IPM complex and the deacylation reaction. (a) The GES/IPM-Δ2 acyl-enzyme complex. All O, N, S, and H atoms are colored in red, blue, yellow, and white, respectively. The carbon atoms on amino acid residues are colored in black. The carbon atoms on the IPM ligand molecule are colored in magenta, except for the deacylating carbonyl carbon (IPM C7) which is colored in dark purple; (b) the deacylation mechanism and the atoms included in the QM region. The QM/MM boundary bonds are marked by blue lines. The structural differences of the acylated IPM-Δ1 and IPM-Δ2 tautomers are highlighted in red.

that the ASβLs deacylation of β-lactams is a concerted one step process (figure 1(b)) [21]. Briefly, the deacylating water molecule first attacks the acylated electrophilic β-lactam carbon and synergistically donates its proton to the deprotonated GES Glu166, which acts as the general base. The scissile bond between Ser70 Oγ and the β-lactam carbonyl carbon is cleaved upon the proton transfer from the fully protonated Lys73 amino to Ser70 Oγ, which retrieves the Ser70 hydroxyl and completes the deacylation. Moreover, the deacylation of carbapenems is further complicated by possible Δ2-to-Δ1 pyrroline tautomerization on the carbapenem scaffold during the acyl-enzyme state (figure 1(b)) [8, 9]. While the interconversion between the two tautomer states is reported, it has been proposed that the Δ1-pyrroline tautomer state is more inert to ASβL deacylation [10].

Paralleling previous theoretical efforts based on QM/MM are the applications of various machine-learning and deep-learning (DL) techniques in computational enzymology [26–30]. DL has been repeatedly shown as the state-of-the-art computational tool for datasets with high complexity and/or dimensionality, making preferrable for assisting the understanding of enzymatic reactions [31]. Graph-learning (GL) is an emerging subarea that applies DL-based techniques on graph-structured data, which encodes data representation using vertices and interconnecting edges [32, 33]. Promoted by the robust graph representation of chemical structures, various GL models have achieved ground-breaking performances on molecular property predictions [34–36]. Enlightened by the robustness of graph neural nets (GNN) for learning hidden knowledge from molecular structures, we are interested in whether GL-based techniques could also extract essential mechanistic insights from graph representations of the enzyme active site.

In this computational study, we combined QM/MM minimum energy pathways (MEPs) and GL to unravel important mechanistic factors of GES-5-mediated imipenem deacylation reactivities. We first present the computational scheme for sampling 1000 GES-5/IPM deacylation QM/MM MEPs. We built graph representations of the acyl-enzymes states and trained a GL model to learn the deacylation barriers on the QM/MM MEPs. Essential structural factors that impact the barrier heights were quantified by the perturbative response of the graph-learned hidden representations. Guided by the knowledge basis extracted from the GL model, we provided useful insights that further advance our mechanistic understanding of the conformational factors that determine the IPM deacylation reactivity in GES-5.

2. Computational details

2.1. System setup

PDB preprocessing. We built the simulation system from the crystal structure of GES-5/IPM acyl-enzyme complex (PDB entry 4H8R) [8]. The protein-ligand complex and all solvent water oxygens resolved in the PDB were retained and the rest of the crystal structure was removed. The protonation states of all titratable residues were compared with PropKa results [37] and were protonated under neutral pH in their standard protonation states. Specifically, all Asp and Glu residues were deprotonated, all Arg and Lys residues were fully protonated, all His residues adopted the neural anionic π tautomer form (protonation on Nδ1). In addition, all Cys residues were fully protonated, except for Cys69 and Cys238, whose sulfhydryl groups were

patched as the disulfide bridge conserved in most ASβL-carbapenemases [9, 11]. We note that the residue IDs ('resid') in the 4H8R PDB file is inconsistent with the standard numbering of the GES-5 protein. Supporting table S1 (https://stacks.iop.org/EST/4/034001/mmedia) is provided to relate the standard residue IDs to their corresponding 'resid' presented in the PDB and coordinate files.

Solvation, neuralization, and force field potentials. The acyl-enzyme complex was solvated in an 80 Å cubic water box. The total charge of the system was neutralized to zero via additions of 14 sodium ions and 10 chloride ions. The CHARMM36 force field (C36) [38], the CHARMM general force field [39], and the CHARMM modified TIP3P water models [40] were applied for the GES-5 protein, the IPM ligand, and the solvent waters, respectively. The atom types and parameters of the ligand molecule (IPM) were assigned by the CGenFF server (https://cgenff.umaryland.edu) [41, 42]. We note that the CGenFF parameters describe the IPM molecule in its 'standalone/unbound' form. The IPM parameters are supplied in the supporting data.

In all simulations, the SHAKE constraint [43] was used to keep all solvent molecules as rigid bodies except the deacylating water. The nonbonding interactions were explicitly treated within 10 Å. The van der Waals part of the nonbonding interactions were smoothed out to zero at 12 Å. The classical coulomb interactions beyond 10 Å were calculated by particle mesh Ewald [44] under periodic boundary conditions.

System initial minimizations. The simulation solvent box was first subjected to 300 steps of steepest decent minimizations with the protein complex fixed in place. After, the system was further relaxed for 3000 adopted basis Newton–Raphson steps with the Ser70-IPM adduct fixed in place. The classical potentials were unable to simulate the covalent bonds between IPM C7 and Ser70 Oγ. We therefore introduced the semi-empirical third order density functional tight binding (DFTB3) method [45] with the 3OB parameters [46] on the active site for further system setups. In all QM/MM simulations, the amino acid residues were partitioned at the Cα–Cβ bonds and the atomic charges on the MM hosting Cα atoms were removed. The following groups were treated with the DFTB3 potential: Ser70, Lys73, Ser130, Asn132, Glu166, Ser170, the deacylating water, and the IPM ligand. We then moved the Ser70 Hγ1 onto the IPM N4 to create the IPM-Δ2 pyrroline tautomer, which was used for subsequent minimizations, equilibrations, and configurational samplings. The system was further optimized for 5000 ANBR steps using the DFTB3/3OB/C36 potential with no positional constraints.

Equilibration dynamics. The equilibration dynamics was performed with the DFTB3/3OB/C36 hybrid potential. The system was firstly heated from 110 K to 310 K at a 10 K per ps temperature increment. A 200 ps isothermal–isobaric (NPT) equilibration was performed with the temperature and pressure of the system maintained at 310 K and 1 atm using the Hoover thermostat and the Langevin piston method [47, 48].

Conformational sampling. We returned to the pure MM treatment of the whole system to efficiently sample the external MM configurations of the starting conformations used for the MEP calculations. The QM link atoms were temporarily removed from the simulation box and the MM host Cα atoms were reset to its atomic unit mass (12.0110 a.u.). The reacting groups, including the Ser70 Oγ, the Lys73 Nζ amino, the Glu166 Oε1, the deacylating water, and the IPM β-lactam/pyrroline bicyclic rings, were held in place to retain their QM-optimized orientations. The system was subjected to 200 ns MD simulations and snapshots were recorded at a 400 ps interval, producing a total number of 500 configurations. All MM and QM/MM MD trajectories in this study were integrated at timesteps of 1 fs. We note that no sampling of the GES/IPM-Δ1 states was performed at this stage and all 500 configurations are in IPM-Δ2 states.

2.2. QM/MM MEPs

Creation of the IPM-Δ1 states. The protonated N4 on the IPM pyrroline undergoes tautomerization in most ASβL/IPM acyl-enzyme intermediates, leading to two possible tautomer configurations on the IPM pyrroline ring: IPM-Δ1 and IPM-Δ2 (figure 1(b)). In their crystallography study, Smith *et al* observed that the *S* isomeric state of the Δ1 tautomer is the dominant configuration in the acyl-enzyme complex in a GES-5 Cys69Gly variant [9]. Accordingly, the *S* isomeric form was selected for all Δ1 tautomer states since that the Cys69Gly mutation has been verified to not alter the conformational architecture of the active site as with the wild-type GES-5. We note that the *S* stereoisomer has been also reported in other ASβL/IPM crystal complexes such as KPC-2 [49]. The aforenoted DFTB3/MM partitioning was rebuilt for all 500 selected GES/IPM-Δ2 configurations. We then created the corresponding GES/IPM-Δ1 tautomer by moving the IPM N4 proton to C2 on its *S* isomer side.

Chain-of-states pathway optimizations. The GES/IPM Δ1 and Δ2 configurations were optimized with the DFTB3/3OB/C36 potential to create the acyl-enzyme reactant states. During the single point optimizations, the MM residues with at least one atom that approached within 4 Å of the QM active site were defined as the melting region and were allowed to move while the rest of the system was fixed. For each of the optimized reactant configurations, the deacylated GES/IPM configurations were created by minimizations imposed with necessary distance-based restraints and were further optimized to generate the corresponding stable product states. The initial guess of the MEP was generated by linearly interpolating the Cartesian space between each pair of the reactant/product configurations with 36 replicated structures (36 replicas per MEP).

The MEPs were optimized with the chain-of-states replica path method with holonomic constrains [50]. Only the melting region was allowed to move during the QM/MM optimizations. A kinetic energy restraint potential with a force constant of 0.05 kcal mol^{-1} $Å^{-2}$ was applied to improve the convergence of the transition path optimizations. A mass scaling factor of 50 was applied on the migrating protons (the Lys73 Hζ1 and the water hydrogens) to facilitate continuous proton displacements along the MEPs. We note that the QM/MM MEP optimizations were performed at the DFTB3 level using the 3OB-f parameter set with optimized vibrational frequencies for C–O, C–N and C–C species [46].

To obtain accurate energetic profiles, the B3LYP hybrid density functional [51, 52] with Grimme's D3 dispersion correction [53] and the 6-31+G** basis set [54] (B3LYP-D3/6-31+G**/C36) was used to refine the single point energetics along the DFTB3/3OB-f/C36 optimized MEP geometries. The combination of DFTB3 and B3LYP-D3 levels of theory has been previously proposed and tested on similar systems [30, 55, 56]. In addition, we compared the MEP configurations optimized at DFTB3/3OB-f/C36 and B3LYP/6-31G*/C36 levels of theory (supporting figure S1). It is shown that the 3OB-f parameters provide better tetrahedral intermediate configurations. In summary, a total number of 1000 pathways (500 for each IPM tautomer state) and 36 000 single point B3LYP-D3/MM calculations were performed in the current study (supporting figures S2 and S3).

2.3. Featurization

The goal of our GL task is to connect the acyl-enzyme configurations to the deacylation barriers of the QM/MM MEPs. We represent the atoms and interatomic distances as vertices and edges, formulating graph representations of the acyl-enzyme configurations for GES/IPM deacylation (figure 2(a)). The atoms included in the graph were selected according to the following procedure. First, the atoms on the reaction coordinates were automatically included as initial vertices. We then expanded the selection to any neighboring atom that is covalently bonded or forming hydrogen bonds with the initial selection in at least one of the reactant configurations, leading to 19 and 21 atoms/vertices in the GES/IPM-Δ1 and GES/IPM-Δ2 graphs, respectively. Döntgen *et al* proposed that atomic partial charges are preferred descriptors for activation barrier predictions [57]. Accordingly, each vertex was represented as a vector that one-hot encodes the element type multiplied by its partial charge from natural population analysis (NPA) [58] on the B3LYP-D3 densities. On the other hand, the edges were naturally defined as the reaction coordinates, the chemical bonds, and the hydrogen-bonding interactions between the selected vertices. Additionally, the atoms in the same residue that are not directly covalent-bonded are also linked (figure 2(a)). We note that all edges defined here are bi-directional and are distinguished by their edge feature, which is the min-max scaled distance between the connected atoms.

2.4. Conformational clustering

Prior to the GL of the deacylation barriers, our visual inspection on the acyl-enzyme states identified several conformational patterns. Therefore, we performed conformational clustering analysis to elaborate the conformational modes of the non-reacting functional groups upon deacylation. We note that our attempts with unsupervised clustering algorithms (also assisted by dimensionality reductions) provides no meaningful clustering as verified from our visual inspection, which suggests the high-complexity of the conformational space on the GES/IPM active site with the reduced edge representations. Alternatively, we manually filtered the metric distances based on two criteria: the metric distance (1) should manifest high variance (>0.5 Å) to clearly distinguish conformational differences (supporting figure S4); and (2) should not lead to small clusters with fewer than 10 snapshots. Eventually, the distances of Asn132 Oδ—IPM HO6α and IPM O6α—Lys73 Hζ1 (figure 2(a)) were used to successfully divide the snapshots into four clusters, noted as clusters A, B, C and D.

2.5. Edge-conditioned graph-learning

With extensive efforts on the development and application of GL models in the field of computational chemistry, we refer the readers to recent reviews for a comprehensive picture of graph-based learning techniques [32–34]. We herein provide only a brief introduction to GL method used in this study.

The main difference between Euclidean data and graph-structured data is that graph data explicitly encode the connectivity as the preferred interaction between the vertices. Accordingly, the strategy for updating the hidden state on vertex v through a graph convolutional layer follows a message passing scheme of three steps: (1) for each vertex v in the graph, the directed edge $(w \to v)$ from the neighboring vertices w is first encoded by a *Message*() function; (2) the encoded interaction is introduced to the hidden representation of v by an *Aggregate*() function; (3) the output hidden state of the layer is then produced by an *Update*() function from the aggregated representation of v. Formally, for most graph convolution layers, the hidden state $\mathbf{x}_v^{(l)}$ of the vertex v at the lth layer is given by:

$$\mathbf{x}_v^{(l)} = \text{Update}\left(\mathbf{x}_v^{(l-1)}, \text{Aggregate}\left(\mathbf{x}_v^{(l-1)}, \text{Message}\left(\mathbf{x}_v^{(l-1)}, \mathbf{x}_w^{(l-1)}, \mathbf{e}_{w\to v}^{(l-1)}\right)\right)\right) \quad (1)$$

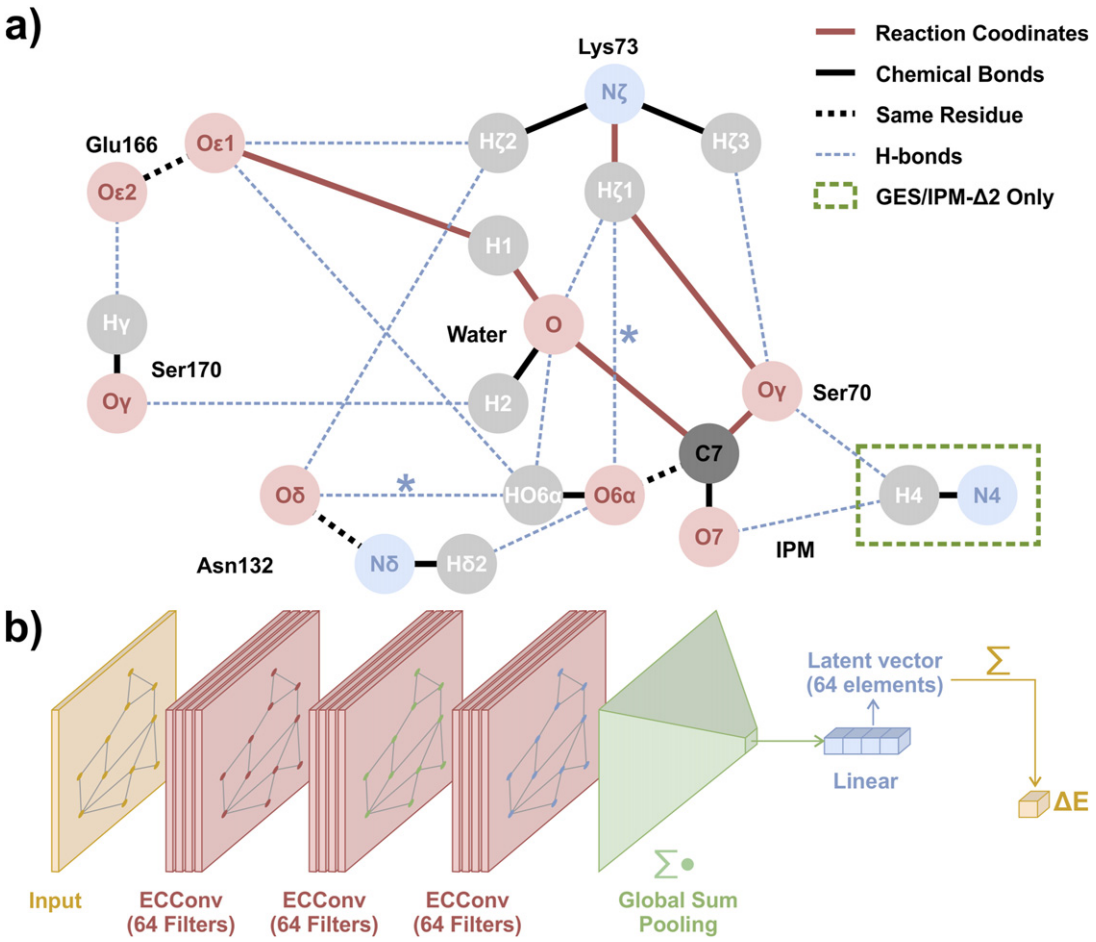

Figure 2. The graph representations and the learning model. (a) The graph representation of GES/IPM acyl-enzyme conformations. The hydrogen, carbon, nitrogen, and oxygen atoms are noted by gray, black, blue, and red spheres, respectively. The edges as reaction coordinates, chemical bonds, hydrogen bonding interactions are shown as red solid lines, black solid lines, and blue dashed lines, respectively. The black dotted lines denote the edges between two atoms of the same residue. The asterisks highlight the edges used as the metrics for conformational clustering. The green box shows the vertices and edges that are included only in GES/IPM-Δ2 graphs. (b) The ECGCNN model architecture.

where $\mathbf{x}_w$ denotes the hidden representations of the vertices connected to v; $\mathbf{e}_{w\to v}^{(l-1)}$ denotes the representation of edges from $\mathbf{x}_w$ to $\mathbf{x}_v$. While the *Aggregate*() function is limited to a handful of operators, GL schemes of different types are mostly distinguished by the implementations of *Message*() and *Update*().

In our case, while the NPA charges are useful to distinguish the atom types of each vertex in one graph, we note that they were also shown to be invariant with regard to different acyl-enzyme state configurations (supporting figure S5), leading to similar vertex representations on different graphs. In this regard, the atomic distances which distinguish different reactant configurations should be explicitly introduced in the GL as edge features. Accordingly, the edge-conditioned graph convolutional neural net (ECGCNN) model proposed by Simonovsky *et al* [59] was implemented for the learning of GES/IPM deacylation barriers. This type of graph convolution encodes edge features in the message function by learning a hidden representation of the edge features and use it as the weight matrix that multiplies the neighboring hidden states $\mathbf{x}_w^{(l-1)}$:

$$\text{Message}\left(\mathbf{x}_w^{(l-1)}, \mathbf{e}_{w\to v}^{(l-1)}\right) = \sum_w \mathbf{x}_w^{(l-1)} A\left(\mathbf{e}_{w\to v}^{(l-1)}\right) \tag{2}$$

where $A()$ is a differentiable function at the edge-conditioned convolutional layer l that maps $\mathbf{e}_{w\to v}^{(l-1)}$ to the shape of the weight matrix applied to $\mathbf{x}_w^{(l-1)}$. We implemented $A()$ as two-layer perceptron: the first layer of $A()$ was activated by a parametric rectified linear unit (PReLU) function [60] and the second was linearly activated. The message was aggregated to $\mathbf{x}_v^{(l-1)}$ by summing to $\mathbf{x}_v^{(l-1)} w_v^{(l)}$, where $w_v^{(l)}$ is the learnable weights at the ECConv layer l. Finally, the updated hidden representation $\mathbf{x}_v^{(l)}$ on the lth layer was produced by a PReLU activation.

The ECGCNN model implemented in the current study adopts three edge-conditioned learning layers, each of which learns the vertex hidden representations with 64 convolutional filters (figure 2(b)). After the third layer, the hidden representation of the graph learned by each filter was read out by a global sum pooling

operator, i.e., by summing all vertex hidden states learned per filter. The hidden representation of the graph was then transformed by a learned weight vector at the final linear layer (without bias units) to produce a 64-dimensional latent vector. The model prediction of the deacylation barrier is obtained from the reduced sum of this latent vector.

We trained the ECGCNN model using the GES/IPM acyl-enzyme reactant configurations. Specifically, 85% (850 acyl-enzyme conformations) of the input data was selected as the training set and the rest (150 conformations) as the validation set. Furthermore, in order to avoid sample unbalance in the training, both training and validation sets were selected in a stratified manner with regard to different IPM tautomer states. The learnable weights were optimized by an adaptive momentum (AdaM) optimizer [61] at a learning rate of 0.001 against the standard loss of mean-squared-error. The model was trained for 750 epochs with a sample batch size of 25. During the training process, the ECGCNN model producing the lowest validation loss was retained as the final model.

As aforenoted, the NPA charges are mostly invariant with regard to the GES/IPM acyl-enzyme configurations, we therefore focused on the edge features for the ECGCNN model explanations to gain mechanistic insights. We assessed the edge importance by the following perturbative approach. For each graph denoting an acyl-enzyme configuration, we first obtained the 64-dimensional latent vector produced from the final linear transform, which is defined as the baseline, h^{base}. We then created a perturbed graph omitting an edge e and feed it through the forward pass of the ECGCNN model, producing a perturbed latent vector, h_e^{pert}. Since the reduced sum of the latent vector gives the predicted deacylation barrier, the contribution attributed to e is defined as the perturbative response of the latent vector upon excluding e from the input graph. Eventually, we assess the importance of e with the one-norm distance between h^{base} and h_e^{pert}:

$$\text{Importance}\,(e) = \sum \left| h_e^{\text{pert}} - h^{\text{base}} \right| \tag{3}$$

We note that the displacement of latent vectors as the indicator for feature importance has been applied extensively to visualize the knowledge basis of various DL models [62]. Moreover, the above attribution method using latent displacement measures edge importance in no sense of an 'energy contribution'. Intuitively, it accounts for the 'sensitivity' of the graph-learned hidden representation to the absence of the edge e in the graph.

2.6. Implementation details

All MD simulations were performed using CHARMM and OpenMM [63, 64]. All QM/MM calculations were performed with the DFTB module of CHARMM and the CHARMM/Q-Chem interface [65–68]. The postprocessing and featurization of the molecular conformations took advantage of the MDAnalysis package [69]. All Uniform Manifold Approximation Project (UMAP) dimensionality reduction analysis were performed with the UMAP-learn library [70]. The Welch statistical tests were conducted using the SciPy infrastructure [71] and the threshold for statistical significance was defined as $p < 0.001$. The ECGCNN model was implemented with PyTorch and PyTorch Geometric [72, 73]. All molecular graphics were prepared with UCSF ChimeraX [74].

3. Results

3.1. GES/IPM deacylation barriers

It has been proposed that the deacylation half of ASβL-mediated IPM hydrolysis is the rate-limiting step of the overall hydrolysis [11]. Frase *et al* measured the kinetic rate of GES-5/IPM deacylation (k_3) as 0.45 s^{-1} under room temperature, which approximates a deacylation barrier of 17.62 kcal mol^{-1} with the Eyring equation [13]. Moreover, Kalp *et al* suggested that the deacylation efficiencies of IPM in ASβLs are also correlated with the tautomerization states of the pyrroline ring, where the IPM-Δ2 tautomer is more active for deacylation than the IPM-Δ1 species [10]. In our calculations (figure 3), the exponential-averaged deacylation barrier of the IPM-Δ1 tautomers in GES-5 (23.83 kcal mol^{-1}) is higher than that of IPM-Δ2 (21.17 kcal mol^{-1}). The IPM-Δ2 tautomer is therefore deemed to be the more active deacylating species.

3.2. Conformational analysis

The relative deacylation activities are further supported by the clustering analysis on the acyl-enzyme conformations. As noted above, all acyl-enzyme configurations are divided into four clusters using two metric distances (figure 4(a)). In configurational clusters of A, B and D, the deacylation barriers of IPM-Δ1 states are higher than Δ2 with statistical significance (Welch test $p < 0.001$, supporting figure S6). Additionally, the C cluster which is statistically inconclusive also shows averagely higher deacylation barrier in IPM-Δ1.

On the other hand, the representative configurations (cluster centroids) of each cluster demonstrate distinct hydrogen bonding patterns around the IPM 6α-hydroxyethyl group (figures 4(c) and (d)). The distance

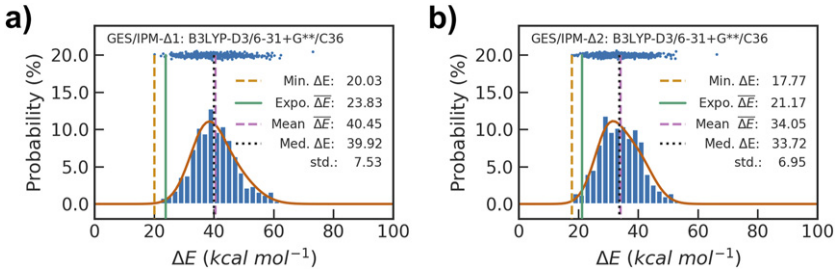

Figure 3. The deacylation barrier distributions of GES/IPM pathways. (a) GES/IPM-Δ1; (b) GES/IPM-Δ2. See also supporting figures S2 and S3.

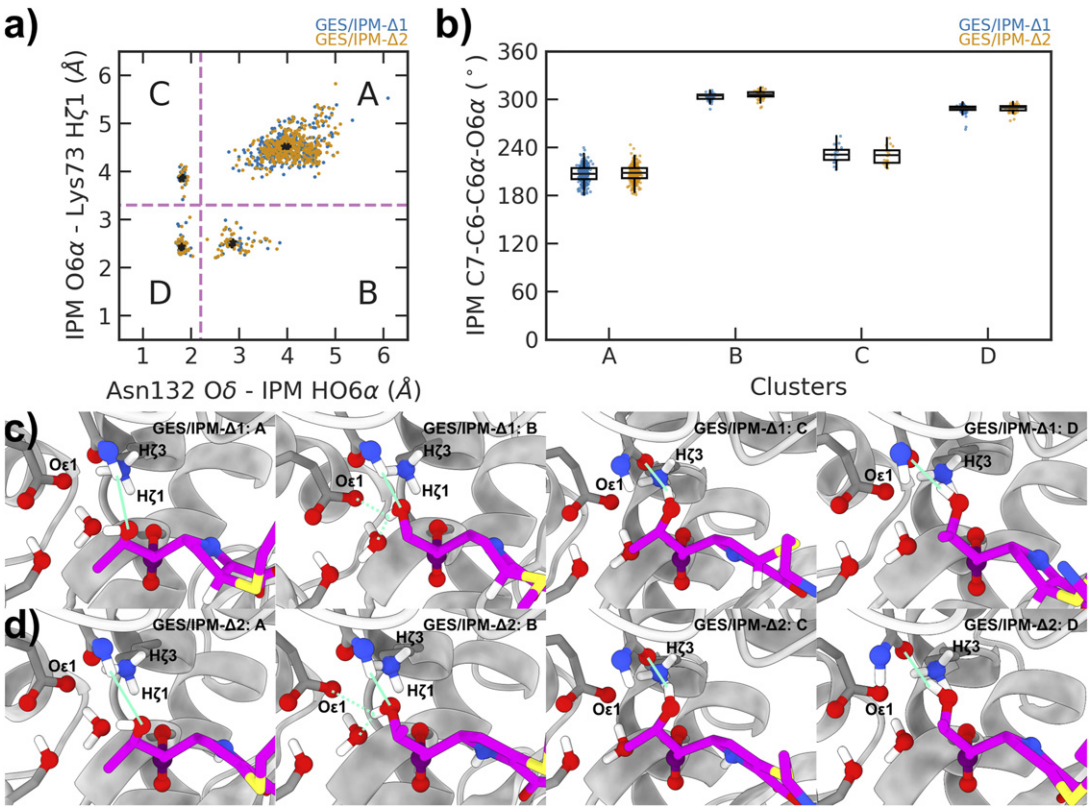

Figure 4. The conformational clustering of the acyl-enzyme states, the energy barriers of the clusters, and the representative conformations in each cluster. (a) The clustering of acyl-enzyme configurations using two distance metrics (pink dashed lines); (b) the distribution of the IPM C7-C6-C6α-O6α dihedral angle of each conformational cluster; the representative acyl-enzyme configurations in clusters of (c) GES/IPM-Δ1; and (d) GES/IPM-Δ2. The key hydrogen-bonding interactions involving the IPM 6α-hydroxyl group and Asn132 are highlighted by cyan solid lines. The Glu166 Oε1, and the Lys73 Hζ1, Hζ3 are labeled. The carbon atoms of the β-lactam ligand in colored in magenta, except the deacylating β-lactam carbon which is colored in dark purple. Other hydrogen, carbon, nitrogen, oxygen, and sulfur atoms are colored in white, dim gray, blue, red, and yellow, respectively.

between Asn132 Oδ and IPM HO6α indicates the hydrogen bonding interactions between the acceptor Asn132 Oδ and the donor IPM O6α. This distance distinguishes the clusters A and B (not hydrogen-bonded) from cluster C and D (hydrogen-bonded). The conformational clusters are further identified by the distance between Lys Hζ1 and IPM HO6α, which is correlated with the distance distribution of IPM O6α—IPM C7 (supporting figure S6) and the dihedral angle of IPM C7-C6-C6α-O6α (figure 4(b)), both of which characterize the orientation of the IPM 6α-hydroxyethyl. The orientation of the 6α-hydroxyethyl in carbapenems and related compounds has been known to be flexible in acyl-enzyme structures [23–25, 75, 76]. In particular, Chudyk *et al* recently reported that the C7-C6-C6α-O6α dihedral profiles of carbapenems in multiple ASβL-carbapenemases adopted two orientations at ∼190 and ∼300 degrees [25]. Herein, we report that GES-5 also follows this structural feature as this dihedral rests at ∼210 degrees in conformational clusters A and C and ∼300 degrees in clusters B and D.

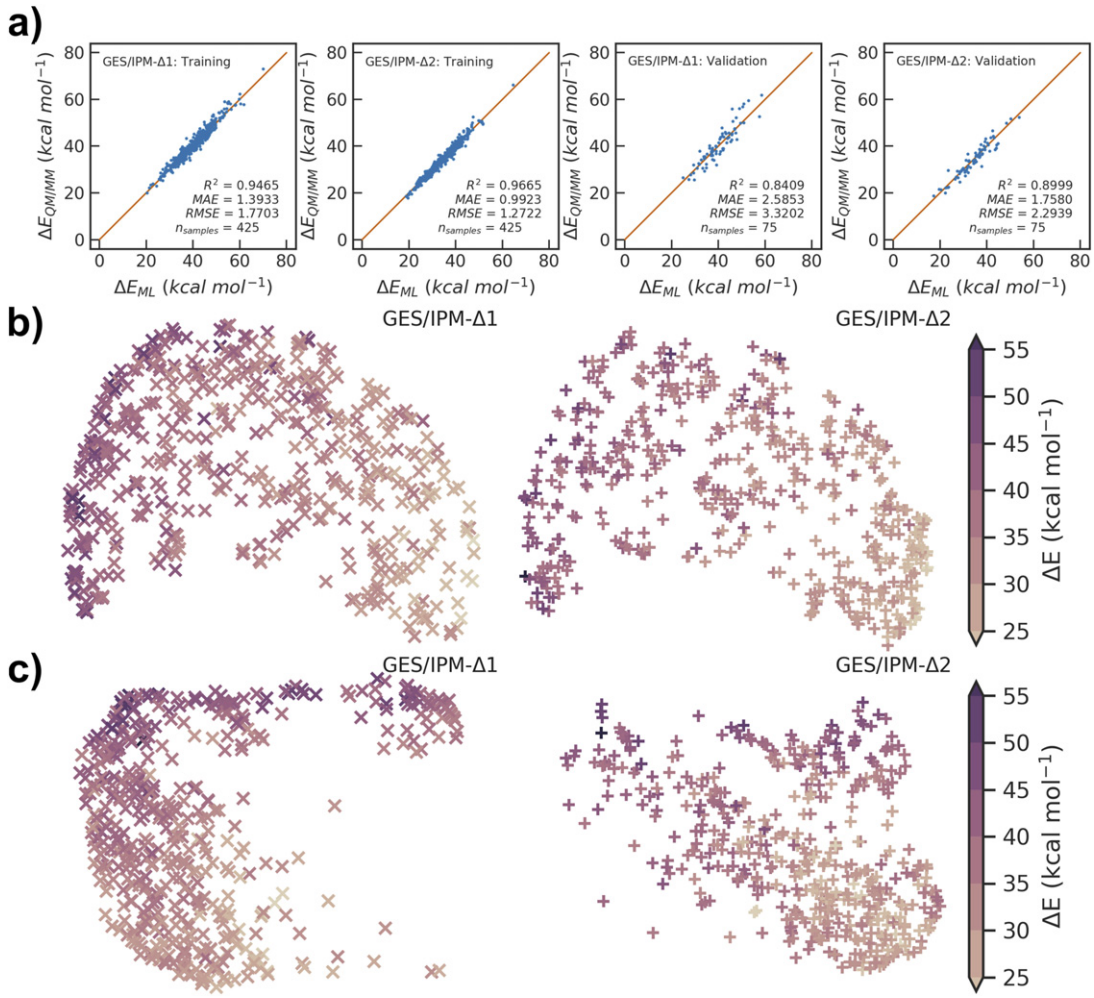

Figure 5. The regression performance of the ECGCNN model and the UMAP dimensionality reduced visualization of graph features. (a) The regression performance of the GNN model on the training and validation sets of two systems. The UMAP dimensionality reductions on the (b) distances as reaction coordinates; (c) the 64-element latent vectors from the ECGCNN model. The UMAP representations of GES/IPM-Δ1 and GES/IPM-Δ2 were plotted under the same scale.

3.3. Graph-learning

The predictive performance of the ECGCNN model on the deacylation barrier is firstly evaluated (figure 5(a)). The deacylation barriers of the training set can be predicted within 2.0 kcal mol^{-1} mean-absolute-errors (MAEs). The predictions on the validation set are with lower accuracy (MAE $<$ 3.0 kcal mol^{-1}), but the predicted barrier heights are in good correlation with the calculated values ($R^2 > 0.8$). Therefore, we conclude that our ECGCNN model could reasonably predict the deacylation barriers based on the graph representations of the acyl-enzyme configuration.

We further investigated whether the knowledge basis of the structural factors on the barrier heights are also properly encoded by the GL latent representations. We applied the UMAP method which has been demonstrated to effectively retain the conformational features in the reduced representation [77]. The low-dimension UMAP representation of the reaction coordinate distances showed clear correlation to the deacylation barrier (figure 5(b)): the configurations of high-to-low deacylation barriers are distributed from left-to-right. However, the acyl-enzyme configurations from different pyrroline tautomerization states are mixed together, suggesting that different tautomer states are conformationally indistinguishable. On the other hand, the UMAP-projected GL latent vectors are able to preserve the barrier distributions as the high-to-low deacylation barriers spans from top-left to lower-right (figure 5(c)). Meanwhile, it also shows that the IPM-Δ1 and Δ2 tautomers can be distinguished from each other, suggesting that the ECGCNN model could effectively capture and properly encode the hidden patterns underlying the conformation-barrier relationship in GES/IPM deacylation.

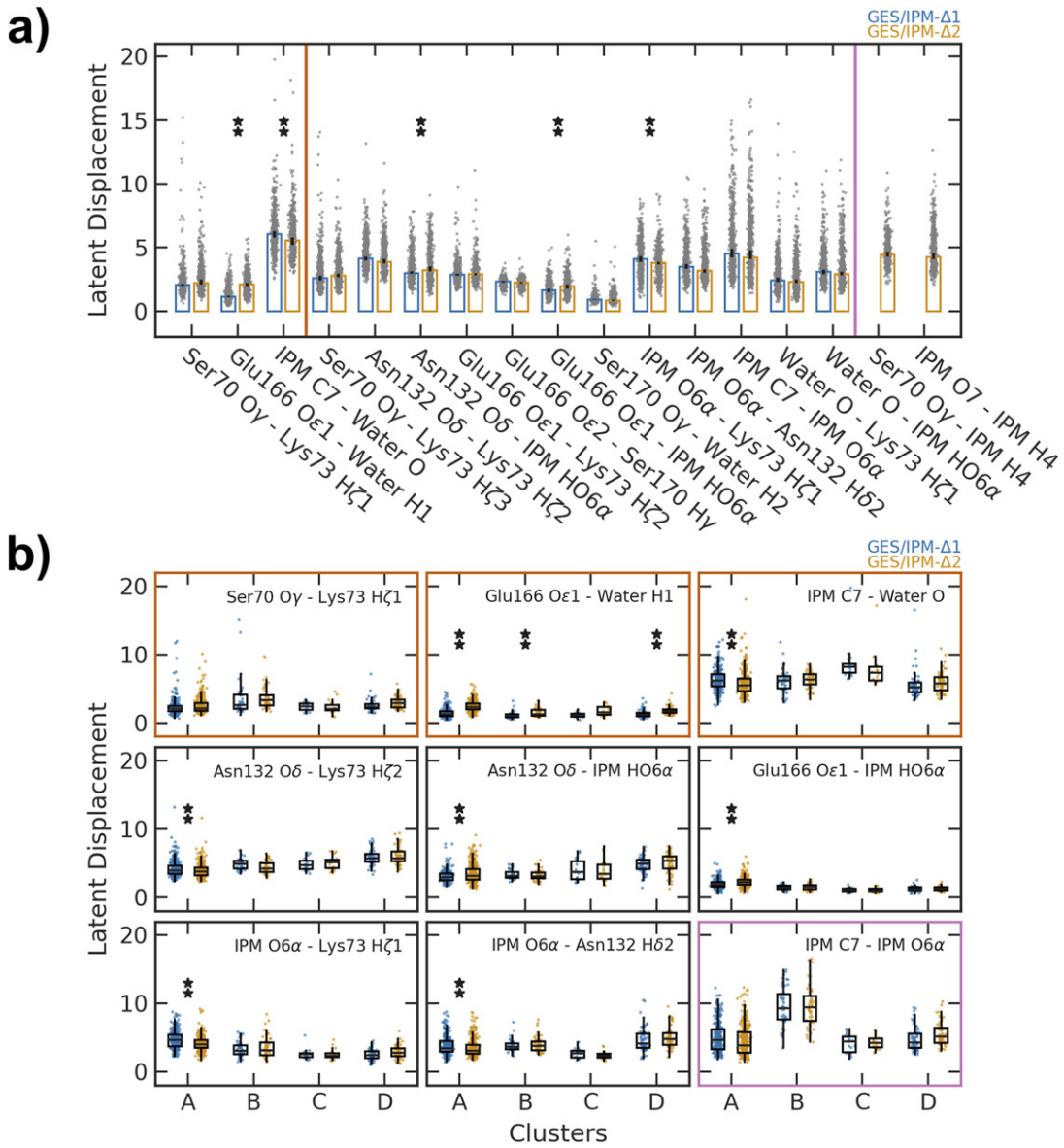

Figure 6. The displacements of latent vectors upon edge removal from the reactant graph. (a) The displacement of latent vectors arranged by edges excluded in each system. The height of bar plots denotes the expectation values of latent displacements and the black solid lines denote the error bounds estimated from 20% bootstrapping tests, see also supporting table S2 and supporting figure S7; (b) the latent vectors displacements that show very significant difference in at least one of the clusters. The sub-plots with orange and black axis are inter-residue reaction coordinates and inter-residue hydrogen bonding interactions, respectively. The symbol '$**$' denotes that the mean difference in barrier distributions between two tautomer states are statistically significant (Welch test $p < 0.001$), see also supporting figure S9.

3.4. Perturbative responses of the graph-learned representations

The ECGCNN model was unboxed using the perturbative response of the graph-learned hidden representations. The overall one-norm displacements of the 64-dimensional latent vector upon edge exclusion are shown in figure 6. The reaction coordinates of the nucleophilic attack of the deacylating water (IPM C7—Water O) and the protonation of the GLU166 (Glu166 Oε1—Water H1) showed difference with statistical significance between the IPM tautomerization states (figure 6(a)). Compared with the IPM-Δ1 tautomers, the deacylation barriers of the Δ2 states are less sensitive to the water attack on the Ser70-IPM ester carbon and are more dependent on the protonating-distance of Glu166.

On the other hand, the hydrogen bonding interactions of larger latent displacement mostly include the IPM 6α-hydroxyl group. Specifically, the interacting groups around IPM 6α-hydroxyl includes Asn132, Glu166, and the deacylating water (figure 2(a)), all of which are in close contact with Lys73. Notably, in both deacylation pathways, the deacylation barriers are shown to be most sensitive to the exclusion of the edge IPM C7—IPM O6α, which indicates the orientation of the 6α-hydroxyethyl. Lastly, while the Welch tests have shown that the latent vector displacements upon excluding three edges (Asn132 Oδ1—IPM HO6α, Glu166 Oε1—IPM

HO6α, and IPM O6α—Lys73 Hζ1) are statistically different between the two tautomer states, the magnitude of these differences is relatively small.

Due to the diverse orientations of the IPM 6α-hydroxyethyl, we further investigated the potential impact to the deacylation barriers from each conformational cluster. The deacylation barriers of IPM-Δ2 are more sensitive than those of Δ1 to the reaction coordinates of Glu166 protonation (Glu166 Oε1—water H1) in clusters A, B and D, while cluster C is inconclusive yet with a higher average perturbative response (figure 6(b)). As for the interactions concerning the IPM 6α side chain groups, all statistically significant comparisons between two tautomer states are found in the A cluster, while the actual differences are small. The key difference between the conformational clusters is the IPM O6α—IPM C7 edges. As shown in figure 6(b), while all clusters showed high latent response upon the IPM O6α—IPM C7 edge removal, cluster B is shown to have latent displacements significantly larger than the others. Such observation can be directly correlated with the local hydrogen bonding interactions to the water attack, as the cluster B is the only cluster that has its IPM 6α-hydroxyl group donating hydrogen bonds to the water molecule or the general base (Glu166, see also supporting figure S8).

4. Discussion

In this study, we investigated the mechanistic basis underlying the deacylation reactions of GES-5 and IPM. Our samplings of the QM/MM MEPs revealed that the deacylation activity of IPM is related to the tautomerization states on its pyrroline rings. The IPM-Δ1 tautomer is shown to be more inert to the deacylating water attack as it gives a higher exponential-averaged deacylation barrier with statistical significance (figure 3(a), supporting figure S6), which aligns with previous reports for ASβLs/IPM deacylation [9, 10].

To further unravel the structural impact on the deacylation barriers from the reactant acyl-enzyme configurations, we implemented a ECGCNN model to learn the deacylation barriers from abstract graph representations of the GES/IPM active site. While the performance of similar GL architectures for atomization energy prediction has reached chemical accuracy (errors within 1.0 kcal mol^{-1}) [34–36], we note that the prediction of activation energies from stable states is a much harder learning task [78, 79]. Intuitively, the activation barrier as the energy difference between the reactant and the transition states should be predicted from the conformational difference between the two states. Herein, we trained the ECGCNN model with only the conformational information of the reactant states, which is a more generalizable approach as the MEPs are normally not known *a prior* for most activation energy predicting scenarios. This strategy can be further justified by that the conformational transition along the MEP on the potential energy surfaces could generally be considered as dependent on the reactant configurations. Nonetheless, our GL model could predict the GES/IPM deacylation barrier with accuracy within 3 kcal mol^{-1} (figure 5(a)). We also investigated the model-learned hidden representation of the conformational graphs. It is shown by UMAP visualizations that using only the distance information could not distinguish the GES/IPM complex between different tautomerization states (figure 5(b)). Alternatively, the visualization on the model-learned latent vectors showed a clear separating boundary of the IPM-Δ1 and Δ2 conformational graphs (figure 5(c)), demonstrating the efficacy of the ECGCNN model to extract hidden correlations underlying the graph representations to the deacylation barrier heights.

We further propose that the mechanistic factors determining the deacylation barrier heights can be unraveled by directly probing the learned graph representations of the GL model, which is quantified by the latent displacement of the model latent vector upon edge exclusion. In particular, we investigated the correlation between the model latent displacement and the pyrroline taumerization states on IPM. For the three reaction coordinates during the deacylation, the deacylating water attack (IPM C7—water O) is shown to be the most sensitive edge connection for the GL hidden representation in both tautomerization states (figure 6(a)). The nucleophilic water attack is therefore deemed to be the rate limiting reaction step of the overall deacylation. In the IPM-Δ2 states, excluding potential hydrogen bonding interactions donated by IPM N4 (Ser70 Oγ—IPM H4 and IPM O7—IPM H4) leads to moderately high perturbative responses in the GL hidden representations (figure 6(a)). As opposed to IPM-Δ1, these hydrogen bonding interactions in Δ2 concertedly led to statistically significant lower latent displacement upon exclusion of the water attack in IPM-Δ2.

Intuitively, upon tetrahedral formation, the protonated pyrroline nitrogen (N4) of IPM-Δ2 could act as a hydrogen bond donor to the nearby ester oxygens (Ser70 Oγ and IPM O7, supporting figure S2(a)). Hydrogen bonds donated by IPM-N4 can be identified on more than half of GES/IPM-Δ2 deacylating MEPs (supporting figure S10). Hydrogen bond interactions to the Ser70-IPM ester oxygens could stabilize the tetrahedral intermediate in the way of an oxyanion hole, resulting in lower energy barriers conferred by the water attack during the deacylation and thus the lower response observed on the latent displacements. In general, our perturbative approach on the GL model suggests these hydrogen bonding interactions as one of the important mechanistic factors that underly the deacylation kinetics found for different IPM pyrroline tautomerization states.

However, the contribution from the IPM N4 protonation does not fully explain the deacylation resistance of carbapenem scaffold to most ASβLs. This is seen by the deacylation efficacy of most early-generation cephems with their equivalent β-lactam nitrogen staying deprotonated during the entire hydrolysis [80]. In this regard, the major structural feature that distinguishes carbapenems from other β-lactam families is that they carry the 6α-hydroxyethyl side chain at the C6 position. Accordingly, the exclusion of the edge between IPM O6α and IPM C7, which highly correlates with the orientation of the IPM 6α-hydroxyethyl side chain (figure 4(b), supporting figure S8), demonstrates the highest latent sensitivity (figure 6(a)) among all non-reaction coordinate distances. Guided by this observation, we identified four distinct patterns of hydrogen bonding of GES/IPM acyl-enzyme conformations from a clustering analysis, which further revealed that the IPM 6α-hydroxyethyl group can adopt two orientations. Specifically, the IPM C7-C6-C6α-O6α dihedral angle rests at ∼210° degrees in clusters A and C, and ∼300 degrees in clusters B and D. Similar observations were reported recently by Chudyk *et al* for other ASβL-carbapenemases [25].

The relative deacylation activities between the two tautomer states were also validated by conformational clustering analysis. In all four clusters, the IPM-Δ2 pyrroline tautomers are more deacylation active than the IPM-Δ1 (supporting figure S6), implying that the IPM pyrroline nitrogen (N4) protonation is mostly responsible for the lower deacylation barrier observed for Δ2. On the other hand, the most significant difference between the conformation clusters is that the IPM O6α—IPM C7 edge exclusion in cluster B showed much higher perturbative response than the other three clusters. The uniqueness of cluster B is that its IPM O6α could serve as hydrogen bond donor to the deacylation water oxygen or the Glu166 Oε1 (supporting figure S8), which decreases the nucleophilicity of the water molecule or the general base Oε and thus the deacylation activity. While the orientation of the 6α side chain is not explicitly encoded in the graph representations, we show that our ECGCNN model could capture essential information that best encodes the conformation—barrier relationship in the QM/MM MEPs.

While it is known that the deacylation reaction involves only the three conserved active site residues (Ser70, Lys73, and Glu166), the role of Ser170 in GES-5, which enables carbapenem deacylation as opposed to Gly170 in GES-1, is therefore of particular value to understand the functional evolution of GES towards carbapenemases [13, 14]. Ser170 coordinates the deacylating water molecule and poses it towards the Ser70-IPM ester carbon, preparing it for the deacylating attack [12]. Residue at the same position has demonstrated similar functions in other ASβLs [81]. The coordination effect of Ser170 is demonstrated by the relatively invariant inter-residue distances around Ser170 (supporting figure S7(a)). In our GL model, it is expected that edges concerning the Ser170 hydroxyl are not determinant to the predicted deacylation barrier heights since no MEP data provided the deacylation barrier of GES-5 in the absence of Ser170. Another structural feature introduced by Ser170 is that its side chain could bury and partially block the deacylating water and the Glu166 general base from approaching the solvent molecules. Here, we showed that additional hydrogen bonding donated by the IPM 6α-hydroxyl (cluster B) could lead to higher deacylation barriers. Similar effects could be expected from potential hydrogen bonding of the deacylating water and Glu166 with the solvent waters. Relatedly, Hirvonen *et al* reported that the hydration environment of the general base was another determinant factor that regulates carbapenem activity in class D β-lactamases [22]. In the case of GES-5, removing the side chain of Ser170 would also partially expose the Glu166 carboxyl to the external solvent molecules, which reduces its nucleophilicity as the general base.

5. Conclusion

In summary, we calculated 1000 QM/MM MEPs for the deacylation between GES-5 and IPM with two tautomer states. Based on the MEP dataset, we developed an ECGCNN model that reasonably predicts the deacylation barrier from the of graph representations of GES/IPM acyl-enzyme conformations. The mechanism underlying the deacylation reactivity of GES-5 has been revealed with atomistic details using the displacement response of the ECGCNN-learned representations upon edge exclusion. Guided by this perturbative approach, we delineated the mechanisms of two major factors that impact the deacylation reactivity in GES/IPM hydrolysis. The protonation on IPM-Δ2 N4 could facilitate an internal oxyanion hole as the hydrogen bonding donated to the Ser70-IPM ester oxygens, which potentially stabilizes the tetrahedral intermediate and is reflected as the small latent vector displacement to the exclusion of IPM C7—water O edges in the ECGCNN model. Alternatively, the IPM 6α-hydroxyethyl group could adopt two orientations and interacts with the reacting groups and thus impacts the deacylating barrier. In the ECGCNN model, this impact is demonstrated as significant latent displacements upon the removal of edges relevant to the IPM 6α side chain. Most importantly, while no explicit representation of the hydroxyethyl orientation was encoded in the feature representation, our ECGCNN model could capture this hidden information and manifest high sensitivity to the IPM O6α—IPM C7 distance which is highly correlated with the IPM 6α-hydroxyethyl orientation.

Finally, our study demonstrates the potential for DL/GL methods in assisting the mechanistic understandings of enzyme catalysis.

Acknowledgments

We are grateful to the anonymous referees for their very constructive and thorough suggestions to our work. We are grateful to Prof. Jing Huang (Westlake University) for helpful suggestions. 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.

Conflict of interest

The authors declare no competing interest.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: 10.5281/zenodo.6473976.

ORCID iDs

Zilin Song https://orcid.org/0000-0002-3127-3971
Peng Tao https://orcid.org/0000-0002-2488-0239

Reference

[1] Queenan A M and Bush K 2007 Carbapenemases: the versatile β-lactamases *Clin. Microbiol. Rev.* **20** 440–58
[2] Nordmann P, Dortet L and Poirel L 2012 Carbapenem resistance in enterobacteriaceae: here is the storm! *Trends Mol. Med.* **18** 263–72
[3] Paterson D L and Bonomo R A 2005 Extended-spectrum β-lactamases: a clinical update *Clin. Microbiol. Rev.* **18** 657–86
[4] Bush K and Jacoby G A 2010 Updated functional classification of β-lactamases *Antimicrob. Agents Chemother.* **54** 969–76
[5] Palzkill T 2018 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.* **5** 16
[6] Walther-Rasmussen J and Høiby N 2007 Class A carbapenemases *J. Antimicrob. Chemother.* **60** 470–82
[7] Llarrull L I, Testero S A, Fisher J F and Mobashery S 2010 The future of the β-lactams *Curr. Opin. Microbiol.* **13** 551–7
[8] Smith C A, Frase H, Toth M, Kumarasiri M, Wiafe K, Munoz J, Mobashery S and Vakulenko S B 2012 Structural basis for progression toward the carbapenemase activity in the GES family of β-lactamases *J. Am. Chem. Soc.* **134** 19512–5
[9] Smith C A, Nossoni Z, Toth M, Stewart N K, Frase H and Vakulenko S B 2016 Role of the conserved disulfide bridge in class A carbapenemases *J. Biol. Chem.* **291** 22196–206
[10] Kalp M and Carey P R 2008 Carbapenems and SHV-1 β-lactamase form different acyl-enzyme populations in crystals and solution *Biochemistry* **47** 11830–7
[11] Swarén P *et al* 1998 X-ray analysis of the NMC-A β-lactamase at 1.64 Å resolution, a class A carbapenemase with broad substrate specificity *J. Biol. Chem.* **273** 26714–21
[12] Stewart N K, Smith C A, Frase H, Black D J and Vakulenko S B 2015 Kinetic and structural requirements for carbapenemase activity in GES-type β-lactamases *Biochemistry* **54** 588–97
[13] Frase H, Shi Q, Testero S A, Mobashery S and Vakulenko S B 2009 Mechanistic basis for the emergence of catalytic competence against carbapenem antibiotics by the GES family of β-lactamases *J. Biol. Chem.* **284** 29509–13
[14] Frase H, Toth M, Champion M M, Antunes N T and Vakulenko S B 2011 Importance of position 170 in the inhibition of GES-type β-lactamases by clavulanic acid *Antimicrob. Agents Chemother.* **55** 1556–62
[15] Fonseca F, Chudyk E I, van der Kamp M W, Correia A, Mulholland A J and Spencer J 2012 The basis for carbapenem hydrolysis by class A β-lactamases: a combined investigation using crystallography and simulations *J. Am. Chem. Soc.* **134** 18275–85
[16] Chudyk E I, Limb M A L, Jones C, Spencer J, van der Kamp M W and Mulholland A J 2014 QM/MM simulations as an assay for carbapenemase activity in class A β-lactamases *Chem. Commun.* **50** 14736–9
[17] Song Z, Zhou H, Tian H, Wang X and Tao P 2020 Unraveling the energetic significance of chemical events in enzyme catalysis via machine-learning based regression approach *Commun. Chem.* **3** 134
[18] Kotsakis S D, Miriagou V, Tzelepi E and Tzouvelekis L S 2010 Comparative biochemical and computational study of the role of naturally occurring mutations at ambler positions 104 and 170 in GES β-lactamases *Antimicrob. Agents Chemother.* **54** 4864–71
[19] Meroueh S O, Fisher J F, Schlegel H B and Mobashery S 2005 *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.* **127** 15397–407
[20] Hermann J C, Hensen C, Ridder L, Mulholland A J and Höltje H-D 2005 Mechanisms of antibiotic resistance: QM/MM modeling of the acylation reaction of a class A β-lactamase with benzylpenicillin *J. Am. Chem. Soc.* **127** 4454–65
[21] Hermann J C, Ridder L, Höltje H-D and Mulholland A J 2006 Molecular mechanisms of antibiotic resistance: QM/MM modelling of deacylation in a class A β-lactamase *Org. Biomol. Chem.* **4** 206–10
[22] Hirvonen V H A, Mulholland A J, Spencer J and van der Kamp M W 2020 Small changes in hydration determine cephalosporinase activity of OXA-48 β-lactamases *ACS Catal.* **10** 6188–96

[23] Hirvonen V H A, Weizmann T M, Mulholland A J, Spencer J and van der Kamp M W 2022 Multiscale simulations identify origins of differential carbapenem hydrolysis by the OXA-48 β-lactamase *ACS Catal.* **12** 4534–44
[24] Stojanoski V, Hu L, Sankaran B, Wang F, Tao P, Prasad B V V and Palzkill T 2021 Mechanistic basis of OXA-48-like β-lactamases hydrolysis of carbapenems *ACS Infect. Dis.* **7** 445–60
[25] Chudyk E, Beer M, Limb A L M, Jones A C, Spencer J, van der Kamp W M and Mulholland A 2022 QM/MM simulations reveal the determinants of carbapenemase activity in class A β-lactamases *ChemRxiv* (accessed 21 Jun 2022) 10.26434/chemrxiv-2022-4jdc5
[26] Vamathevan J *et al* 2019 Applications of machine learning in drug discovery and development *Nat. Rev. Drug Discovery* **18** 463–77
[27] Gokcan H and Isayev O 2022 Learning molecular potentials with neural networks *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **12** e1564
[28] Latallo M J, Cortina G A, Faham S, Nakamoto R K and Kasson P M 2017 Predicting allosteric mutants that increase activity of a major antibiotic resistance enzyme *Chem. Sci.* **8** 6484–92
[29] Cortina G A, Hays J M and Kasson P M 2018 Conformational intermediate that controls KPC-2 catalysis and beta-lactam drug resistance *ACS Catal.* **8** 2741–7
[30] Song Z, Trozzi F, Tian H, Yin C and Tao P 2022 Mechanistic insights into enzyme catalysis from explaining machine-learned quantum mechanical and molecular mechanical minimum energy pathways *ACS Phys. Chem. Au* https://doi.org/10.1021/acsphyschemau.2c00005
[31] Kulik H *et al* 2022 Roadmap on machine learning in electronic structure *Electron. Struct.* https://doi.org/10.1088/2516-1075/ac572f
[32] Wieder O, Kohlbacher S, Kuenemann M, Garon A, Ducrot P, Seidel T and Langer T 2020 A compact review of molecular property prediction with graph neural networks *Drug Discovery Today: Technologies* **37** 1–12
[33] Zhou J, Cui G, Hu S, Zhang Z, Yang C, Liu Z, Wang L, Li C and Sun M 2020 Graph neural networks: a review of methods and applications *AI Open* **1** 57–81
[34] Gilmer J, Schoenholz S S, Riley P F, Vinyals O and Dahl G E 2017 Neural message passing for quantum chemistry *Proc. 34th Int. Conf. Machine Learning (PMLR)* pp 1263–72
[35] Wang X, Li Z, Jiang M, Wang S, Zhang S and Wei Z 2019 Molecule property prediction based on spatial graph embedding *J. Chem. Inf. Model.* **59** 3817–28
[36] Hao Z, Lu C, Huang Z, Wang H, Hu Z, Liu Q, Chen E and Lee C 2020 ASGN: an active semi-supervised graph neural network for molecular property prediction *Proc. 26th ACM SIGKDD Int. Conf. Knowledge Discovery & Data Mining (ACM)* pp 731–52
[37] Li H, Robertson A D and Jensen J H 2005 Very fast empirical prediction and rationalization of protein pKa values *Proteins* **61** 704–21
[38] Best R B, Zhu X, Shim J, Lopes P E M, Mittal J, Feig M and MacKerell A D 2012 Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ_1 and χ_2 dihedral angles *J. Chem. Theory Comput.* **8** 3257–73
[39] Vanommeslaeghe K *et al* 2010 CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields *J. Comput. Chem.* **31** 671–90
[40] Jorgensen W L, Chandrasekhar J, Madura J D, Impey R W and Klein M L 1983 Comparison of simple potential functions for simulating liquid water *J. Chem. Phys.* **79** 926–35
[41] Vanommeslaeghe K and MacKerell A D 2012 Automation of the CHARMM general force field (CGenFF) I: bond perception and atom typing *J. Chem. Inf. Model.* **52** 3144–54
[42] Vanommeslaeghe K, Raman E P and MacKerell A D 2012 Automation of the CHARMM general force field (CGenFF) II: assignment of bonded parameters and partial atomic charges *J. Chem. Inf. Model.* **52** 3155–68
[43] Ryckaert J-P, Ciccotti G and Berendsen H J C 1977 Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of *n*-Alkanes *J. Comput. Phys.* **23** 327–41
[44] Darden T, York D and Pedersen L 1993 Particle mesh Ewald: AnN·log(N) method for Ewald sums in large systems *J. Chem. Phys.* **98** 10089–92
[45] Gaus M, Cui Q and Elstner M 2011 DFTB3: extension of the self-consistent-charge density-functional tight-binding method (SCC-DFTB) *J. Chem. Theory Comput.* **7** 931–48
[46] Gaus M, Goez A and Elstner M 2013 Parametrization and benchmark of DFTB3 for organic molecules *J. Chem. Theory Comput.* **9** 338–54
[47] Evans D J and Holian B L 1985 The Nose–Hoover thermostat *J. Chem. Phys.* **83** 4069–74
[48] Feller S E, Zhang Y, Pastor R W and Brooks B R 1995 Constant pressure molecular dynamics simulation: the Langevin piston method *J. Chem. Phys.* **103** 4613–21
[49] Furey I M, Mehta S C, Sankaran B, Hu L, Prasad B V V and Palzkill T 2021 Local interactions with the Glu166 base and the conformation of an active site loop play key roles in carbapenem hydrolysis by the KPC-2 β-lactamase *J. Biol. Chem.* **296** 100799
[50] Brokaw J B, Haas K R and Chu J-W 2009 Reaction path optimization with holonomic constraints and kinetic energy potentials *J. Chem. Theory Comput.* **5** 2050–61
[51] Lee C, Yang W and Parr R G 1988 Development of the Colle–Salvetti correlation-energy formula into a functional of the electron density *Phys. Rev.* B **37** 785–9
[52] Becke A D 1993 A new mixing of Hartree–Fock and local density-functional theories *J. Chem. Phys.* **98** 1372–7
[53] Grimme S 2006 Semiempirical GGA-type density functional constructed with a long-range dispersion correction *J. Comput. Chem.* **27** 1787–99
[54] Ditchfield R, Hehre W J and Pople J A 1971 Self-consistent molecular-orbital methods: IX. An extended Gaussian-type basis for molecular-orbital studies of organic molecules *J. Chem. Phys.* **54** 724–8
[55] Yagi K, Ito S and Sugita Y 2021 Exploring the minimum-energy pathways and free-energy profiles of enzymatic reactions with QM/MM calculations *J. Phys. Chem.* B **125** 4701–13
[56] Yang Z, Twidale R M, Gervasoni S, Suardíaz R, Colenso C K, Lang E J M, Spencer J and Mulholland A J 2021 Multiscale workflow for modeling ligand complexes of zinc metalloproteins *J. Chem. Inf. Model.* **61** 5658–72
[57] Döntgen M, Fenard Y and Heufer K A 2020 Atomic partial charges as descriptors for barrier heights *J. Chem. Inf. Model.* **60** 5928–31
[58] Reed A E, Weinstock R B and Weinhold F 1985 Natural population analysis *J. Chem. Phys.* **83** 735–46
[59] Simonovsky M and Komodakis N 2022 Dynamic edge-conditioned filters in convolutional neural networks on graphs *Proc. IEEE Conf. Computer Vision and Pattern Recognition* pp 3693–702
[60] He K, Zhang X, Ren S and Sun J 2015 Delving deep into rectifiers: surpassing human-level performance on imagenet classification *Proc. IEEE Int. Conf. Computer Vision* pp 1026–34

[61] Kingma D P and Ba J 2017 Adam: a method for stochastic optimization arXiv:1412.6980 (accessed 21 Jun 2022)
[62] Krizhevsky A, Sutskever I and Hinton G E 2017 ImageNet classification with deep convolutional neural networks *Commun. ACM* **60** 84–90
[63] Brooks B R *et al* 2009 CHARMM: the biomolecular simulation program *J. Comput. Chem.* **30** 1545–614
[64] Eastman P *et al* 2017 OpenMM 7: rapid development of high performance algorithms for molecular dynamics *PLoS Comput. Biol.* **13** e1005659
[65] Cui Q, Elstner M, Kaxiras E, Frauenheim T and Karplus M 2001 A QM/MM implementation of the self-consistent charge density functional tight binding (SCC-DFTB) method *J. Phys. Chem.* B **105** 569–85
[66] Woodcock H L, Hodošček M and Brooks B R 2007 Exploring SCC-DFTB paths for mapping QM/MM reaction mechanisms *J. Phys. Chem.* A **111** 5720–8
[67] Epifanovsky E *et al* 2021 Software for the Frontiers of quantum chemistry: an overview of developments in the Q-Chem 5 package *J. Chem. Phys.* **155** 084801
[68] Woodcock H L, Hodošček M, Gilbert A T B, Gill P M W, Schaefer H F III and Brooks B R 2007 Interfacing Q-Chem and CHARMM to perform QM/MM reaction path calculations *J. Comput. Chem.* **28** 1485–502
[69] Michaud-Agrawal N, Denning E J, Woolf T B and Beckstein O 2011 MDAnalysis: a Toolkit for the analysis of molecular dynamics simulations *J. Comput. Chem.* **32** 2319–27
[70] McInnes L, Healy J, Saul N and Großberger L 2018 UMAP: uniform manifold approximation and projection *J. Open Source Softw.* **3** 861
[71] Virtanen P *et al* 2020 SciPy 1.0: fundamental algorithms for scientific computing in Python *Nat. Methods* **17** 261–72
[72] Paszke A *et al* 2019 PyTorch: an imperative style, high-performance deep learning library arXiv:1912.01703 (accessed 21 Jun 2022)
[73] Fey M and Lenssen J E 2019 Fast graph representation learning with PyTorch geometric arXiv:1903.02428 (accessed 21 Jun 2022)
[74] Pettersen E F, Goddard T D, Huang C C, Meng E C, Couch G S, Croll T I, Morris J H and Ferrin T E 2021 UCSF ChimeraX: structure visualization for researchers, educators, and developers *Protein Sci.* **30** 70–82
[75] Pemberton O A, Zhang X and Chen Y 2017 Molecular basis of substrate recognition and product release by the *Klebsiella pneumoniae* carbapenemase (KPC-2) *J. Med. Chem.* **60** 3525–30
[76] Nukaga M, Bethel C R, Thomson J M, Hujer A M, Distler A, Anderson V E, Knox J R and Bonomo R A 2008 Inhibition of class A β-lactamases by carbapenems: crystallographic observation of two conformations of meropenem in SHV-1 *J. Am. Chem. Soc.* **130** 12656–62
[77] Trozzi F, Wang X and Tao P 2021 UMAP as a dimensionality reduction tool for molecular dynamics simulations of biomacromolecules: a comparison study *J. Phys. Chem.* B **125** 5022–34
[78] Grambow C A, Pattanaik L and Green W H 2020 Deep learning of activation energies *J. Phys. Chem. Lett.* **11** 2992–7
[79] Lewis-Atwell T, Townsend P A and Grayson M N 2022 Machine learning activation energies of chemical reactions *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* e1593
[80] Song Z, Trozzi F, Palzkill T and Tao P 2021 QM/MM modeling of class A β-lactamases reveals distinct acylation pathways for ampicillin and cefalexin *Org. Biomol. Chem.* **19** 9182–9
[81] Brown N G, Shanker S, Prasad B V V and Palzkill T 2009 Structural and biochemical evidence that a TEM-1 β-lactamase N170G active site mutant acts via substrate-assisted catalysis *J. Biol. Chem.* **284** 33703–12