

UNRAVELING MOLECULAR MECHANISMS OF ANTIBIOTIC RESISTANCE THROUGH
MULTISCALE SIMULATIONS AND EXPLAINABLE MACHINE LEARNING

Approved by:

Prof. Peng Tao
Associate Professor of Chemistry

Prof. Doran I.G. Bennett
Assistant Professor of Chemistry

Prof. John D. Buynak
Professor of Chemistry

Prof. Xinlei (Sherry) Wang
Professor of Statistical Science

UNRAVELING MOLECULAR MECHANISMS OF ANTIBIOTIC RESISTANCE THROUGH
MULTISCALE SIMULATIONS AND EXPLAINABLE MACHINE LEARNING

A Dissertation Presented to the Graduate Faculty of the

Dedman College

Southern Methodist University

in

Partial Fulfillment of the Requirements

for the degree of

Doctor of Philosophy

with a

Major in Theoretical and Computational Chemistry

by

Zilin Song

M.Eng., Chemical Engineering, Wuhan Institute of Technology, China

May 14, 2022

Copyright (2022)

Zilin Song

All Rights Reserved

ACKNOWLEDGMENTS

This work could not have been possible without the experience of my wonderful advisor, Prof. Peng Tao.

I would like to thank Prof. Doran Bennett, Prof. John Buynak, and Prof. Xinlei Wang, who have kindly accepted to be on my dissertation committee. I am grateful to Prof. Jing Huang (Westlake), Prof. Yihan Shao (OU), Prof. Timothy Palzkill (BCM), and Prof. Xiaobo Chen (UMKC) for helpful suggestions. I would also like to thank Prof. Elfi Kraka, Prof. Michael Lattman, Dr. Robert Kalesky, Amit Kumar, and Ms. Mandy Graham for their precious help on various occasions during my Ph.D.

I am sincerely grateful to all my friends in the chemistry department, especially Drs. Weiwei An, Feng Wang, and Hongyu Zhou as well as the soon-to-be-Drs. Nischal Karki, Bo Li, Francesco Trozzi, and Tong Wu (CMU), for stimulating discussions on research as well as being great friends in life. I am thankful to my group mates for being the most pleasant colleagues to work with.

Above all, my deepest gratitude goes to my father and mother, who have been constantly supportive to me during the last four years.

Unraveling Molecular Mechanisms of Antibiotic Resistance Through
Multiscale Simulations and Explainable Machine Learning

Advisor: Professor Peng Tao
Doctor of Philosophy conferred May 14, 2022
Dissertation completed April 19, 2022

Pathogen resistance to β -lactam antibiotics compromises effective treatments of superbug infections. One major source of β -lactam resistance is the bacterial production of β -lactamases, which could effectively hydrolyze β -lactam drugs. In this thesis, the hydrolysis of various β -lactam antibiotics by class A serine-based β -lactamases (AS β Ls) were investigated using hybrid Quantum Mechanical / Molecular Mechanical (QM/MM) minimum energy pathway (MEP) calculations and explainable machine learning (ML) approaches. The TEM-1/benzylpenicillin acylation reaction with QM/MM chain-of-states reaction pathways was firstly revisited. I proposed two decomposition methods for energy contribution analysis based on perturbing ML regression models. Both methods were shown to be model implementation invariant and successfully bridged the discrepancies between two pioneering mechanistic studies. The Toho-1 AS β L acylations of ampicillin and cefalexin were then investigated. I reported that the acylation pathway selection can be ligand dependent: ampicillin could undergo acylation *via* Lys73 or Glu166 acting as the general base while cefalexin acylation is limited to Lys73 as the general base. An explainable artificial intelligence (XAI) method, the Boltzmann-weighted Cumulative Integrated Gradients (BCIG), was developed to explain the different acylation pathway viability found for ampicillin and cefalexin. Lastly, conformational factors determining the GES-5/imipenem deacylation activity was investigated using edge-conditioned convolutional graph-learning (GL) methods. Critical

mechanistic insights were derived from perturbative response of the GL latent representations, which explained the different deacylation reactivity between the two imipenem pyrroline tautomer states and identified the orientation of the carbapenem 6 α -hydroxyethyl as the key factor that impacts the deacylation barrier heights. In summary, my thesis focuses on bridging QM/MM chain-of-states reaction pathway calculations and explainable ML to derive essential mechanistic insights into β -lactam resistance driven by AS β Ls.

TABLE OF CONTENTS

LIST OF FIGURES	x
LIST OF TABLES	xiv
LIST OF SYMBOLS AND ACRONYMS	xv
1. INTRODUCTION	1
2. METHODOLOGIES	6
2.1 Classical Molecular Mechanical Potentials	6
2.2 Hybrid Quantum Mechanical/Molecular Mechanical Potentials.....	8
2.3 Chain-of-states MEP Optimizations	12
2.4 Machine Learning and Model Explanations	16
3. REVISITING TEM-1/BENZYL PENICILLIN ACYLATION MECHANISM WITH MACHINE-LEARNING ENERGY CONTRIBUTION ANALYSIS	19
3.1 TEM-1 β -Lactamases.....	19
3.2 Computational Details	20
3.3 Summary of Results	22
3.4 Conclusion	29
4. DISTINCT ACYLATION PATHWAYS OF CTX-M β -LACTAMASES WITH AMPICILLIN AND CEFALOXIN IDENTIFIED FROM QM/MM.....	31
4.1 CTX-M β -Lactamases.....	31
4.2 Computational Details	32
4.3 Summary of Results	34

4.4 Conclusion	38
5. ACYLATION OF CTX-M-44 AS β L EXPLAINED FROM MACHINE-LEARNED QM/MM MINIMUM ENERGY PATHWAYS	40
5.1 Boltzmann-weighted Cumulative Integrated Gradients.....	40
5.2 Computational Details	44
5.3 Summary of Results.....	46
5.4 Conclusion	51
6. MECHANISTIC ANALYSIS OF CARBAPENEMASES RESISTANCE TO IMPENEM GUIDED BY GRAPH-LEARNING	53
6.1 GES-5 β -Lactamases.....	53
6.2 Computational Details	54
6.3 Summary of Results.....	60
6.4 Conclusion	65
7. EMPIRICAL FORCE FIELD PARAMETRIZATION OF 2-AMINOTHIAZOLE	66
7.1 Parametrization Scheme of CGenFF	66
7.2 Parameter Searching Strategy	67
7.3 Parametrization of 2-Aminothiazole: Atomic Partial Charges	68
7.4 Parametrization of 2-Aminothiazole: Equilibrium Terms	70
7.5 Parametrization of 2-Aminothiazole: Force Constants.....	70
7.6 Comparison of QM and MM optimized geometries.....	72
8. CONCLUSIONS.....	73
9. BIBLIOGRAPHY	75

10.APPENDIX.....	92
Published and Submitted Manuscripts (First Author Only)	92
Natural Internal Coordinates for 2-aminothiazole with explanations.....	93
Force Field for 2-Aminothiazole Compatible with CHARMM36/CGenFF.....	94
Force Field for Cephalothin Compatible with CHARMM36/CGenFF	95

LIST OF FIGURES

Figure 1.1 Chemical structures of representative β -lactam families and mechanism of β -lactam hydrolysis in AS β Ls. (a) The β -lactam four membered-ring, the penicillin, cephalosporin, and carbapenem scaffolds; (b) The hydrolysis mechanisms of β -lactam mediated by AS β Ls.	2
Figure 2.1 The QM/MM partitioning on biomolecular simulating systems.	9
Figure 2.2 The single link-atom approach for treating QM/MM boundary potential.	10
Figure 2.3 The optimization of the MEP on a Müller potential using chain-of-states methods with a steepest decent optimizer. (a) The RPM with holonomic constraints; (b) The string method with force projections. Black dots refer the initial guess and red dots the final optimized path, gray trajectories are the optimization steps.	13
Figure 3.1 Acylation mechanism of AS β Ls and the structure of TEM-1/benzylpenicillin Michaelis complex. (a) Acylation mechanism of TEM-1 and benzylpenicillin with Glu166 acting as a general base. The β -lactam scissile bond is noted in red; (b) Crystal structure of TEM-1 complexed with benzylpenicillin and the selection of QM atoms.	20
Figure 3.2 QM/MM chain-of-states pathway profiles. Reaction pathways calculated from DFTB3/mio/C36, B3LYP/6-31++G**/ C36, and B3LYP-D3/6-31++G**/ C36 levels of theory. The black rectangles highlight the tetrahedral intermediates region along the energy profiles.	23
Figure 3.3 Feature selections and notations of chemical processes, and regression performance of the models. (a) Selection of atomic distances as features and chemical events; (b) RMSE predicted by three regression models.	25
Figure 3.4 Intrinsic energy contributions measured on DFTB3/mio/C36 and B3LYP/6-31++G**/C36 reaction pathway profiles. The ‘S’, ‘G’, and ‘K’ labels represent results from SVR, GPR, and KRR models, respectively.	26
Figure 3.5 Dynamic energy contributions. The dynamic energy contribution measured from SVR, GPR and KRR regression models trained on energetic profiles calculated at B3LYP/6-31++G**:/CHARMM level of theory. The values provided are the average over the reaction progress.	28

Figure 4.1 Mechanisms of acylation in AS β LS and structures of the model substrates. (a) The general mechanism of β -lactam acylation mediated by AS β L; (b) Structures of ampicillin (AMP) and cefalexin (CEX). 32

Figure 4.2 Conformations of R1 reactant states. The conformations of (a) Toho/AMP:R1; (b) Toho/CEX:R1; (c) Toho/CEX:R1a. The hydrogen bonding interactions are noted as blue dashed lines. 34

Figure 4.3 Energy profiles and the ChElPG charges of key atoms along the acylation pathways in Toho-1 hydrolysis. (a) The acylation profiles of Toho/AMP; The ChElPG charges along (b) the Toho/AMP: R1 to AE1 pathway, and (c) the Toho/AMP: R2 to AE1 pathway; (d) The energy profile and the ChElPG charge profiles of the refined Toho/CEX: R1a to AE1 pathways, which is calculated from inserting 18 replicas between replica 24 and 31; (e) The acylation profiles of Toho/CEX; The ChElPG charges along (f) the Toho/CEX: R1 to AE1 pathway, (g) the Toho/CEX: R2 to AE1 pathway, and (h) the Toho/CEX: R1a to AE1 pathway. The vertical black solid lines in (a) and (d) indicate the location of AE1 and AE2. Numbers in parentheses and brackets denote the local minimum and maximum values of important states along the reaction path. Note that only ChElPG charge values of β -lactam carbonyl carbon (blue) and nitrogen (orange) are shown in (b), (c), (f), (g), (h). 35

Figure 5.1 Selected features, 2D representation of the pathway conformations, and architecture of the ML-MEP models. The selected features and chemical processes of (a) the Toho/AMP: R1-AE and Toho/CEX: R1-AE datasets and (b) the Toho/AMP: R2-AE and Toho/CEX: R2-AE datasets. The atomic distances that are included in feature vectors are noted in orange lines, the chemical processes are noted in blue; (c) The 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) The architecture of the QM/MM MEP learning deep-and-wide neural network. 45

Figure 5.2 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. 47

Figure 5.3 The predictive performance and the BCIG contributions of the ML-MEP models. The predictive performance of (a) the replica energies and (b) the 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. 48

Figure 5.4 The 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) the Toho/AMP&CEX: R1-AE; and (b) the Toho/AMP&CEX: R2-AE models. The BCIG

contributions of (c) the Toho/AMP&CEX: R1-AE; and (d) the Toho/AMP&CEX: R2-AE models.	50
Figure 6.1 The GES/IPM complex and the deacylation reaction. (a) The GES/IPM acyl-enzyme complex. The deacylating water and the IPM molecule are colored in green and magenta, respectively; (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- Δ 1 and IPM- Δ 2 tautomer are highlighted in red.	
Figure 6.2 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- Δ 2 graphs. (b) The ECGCNN model architecture.	56
Figure 6.3 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 deacylation barriers in each cluster. ** denotes that the mean difference in barrier distributions between two tautomer states are statistically significant (Welch test $p < 0.001$); The representative acyl-enzyme configurations in clusters of (c) GES/IPM- Δ 1; and (d) GES/IPM- Δ 2. The key hydrogen-bonding interactions involving the IPM 6 α -hydroxyl group and Asn132 are highlighted by cyan solid lines. The Glu166 O ϵ 1, and the Lys73 H ζ 1, H ζ 3 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.	
Figure 6.4 Distribution of NPA charges on each atom included in the graph. Entries on the left side of the solid orange line are vertices (atoms) related to the reaction coordinates. Entries on the right side of the solid purple line are vertices (atoms) present only in GES/IPM- Δ 2 graphs..	59
Figure 6.5 The deacylation barrier distributions of GES/IPM pathways. (a) GES/IPM- Δ 1; (b) GES/IPM- Δ 2.	60
Figure 6.6 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. Note that the UMAP representations of GES/IPM- Δ 1 and GES/IPM- Δ 2 were plotted under the same scale.	

Figure 6.7 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; (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. ** denotes that the mean difference in barrier distributions between two tautomer states are statistically significant (Welch test $p < 0.001$).	63
Figure 6.8 The distribution of the IPM C7-C6-C6 α -O6 α dihedral angle of each conformational cluster.	64
Figure 7.1 Flowchart for CGenFF parametrization and chemical structure of 2-aminothiazole..	67
Figure 7.2 Demonstration of adaptive grid search in a 2D variable space.	68
Figure 7.3 Water interactions considered for 2-Aminothiazole.....	69
Figure 7.4 CGenFF parametrization for vibrational frequencies, the frequency marked in the red dashed box produced by the scissoring of the amino group.	71
Figure 7.5 Comparison of QM (in orange) and MM (in multi-color) minimized geometry: (a) with unfitted force field parameters; (b) with force field parameters fitted for equilibrium geometry; (c) with force field parameters fitted for vibrational frequencies.	72

LIST OF TABLES

Table 3.1 Comparison of acylation energy barriers of the current and previous works.	24
Table 4.1 The mean atomic distances between key reacting heavy atoms in the DFTB3/3OB/C36 dynamics. Parenthesis denote the standard deviation (units: Å).	33
Table 4.2 The computational and experimental catalytic barriers of AMP and CEX acylation catalyzed by Toho-1.	38
Table 7.1 Deviation to QM water interaction energies of initial and optimized atomic charges.	69
Table 7.2 Initial and optimized equilibrium bond length and bond angles.	70
Table 7.3 Initial and optimized force constants, units vary with entries.	70

LIST OF SYMBOLS AND ACRONYMS

AS β L	-	Class A β -Lactamase
AMP	-	Ampicillin
BCIG	-	Boltzmann-weighted Cumulative Integrated Gradients
CEX	-	Cefalexin
CGenFF	-	CHARMM General Force Fields
DFT	-	Density Functional Theory
DL	-	Deep-learning
ECGCNN	-	Edge-Conditioned Graph Convolutional Neural Network
FDS	-	First Depth Search
GL	-	Graph-learning
GMM	-	Gaussian Mixture Model
IG	-	Integrated Gradients
IPM	-	Imipenem
MAE	-	Mean Absolute Errors
MEP	-	Minimum Energy Pathway
ML	-	Machine-learning
MSE	-	Mean Squared Error
NIC	-	Natural Internal Coordinates
NPA	-	Natural Population Analysis

PCA	-	Principal Component Analysis
PDF	-	Probability Distribution Function
QM/MM	-	Quantum Mechanical / Molecular Mechanical
RMS	-	Root Mean Square
RPM	-	Replica Path Method
XAI	-	Explainable Artificial Intelligence

To all my failed calculations & rejected manuscripts.

1. INTRODUCTION

Bacterial resistance to antibiotic drugs compromises effective clinical treatments of pathogen infections and poses severe threat to global health. While being obviously of high clinical values, antibiotics are also economically vital for industries threatened by bacterial infections, such as husbandry.¹ Antibiotic molecules are biologically-active compounds that kill bacterial strains or disrupt their binary fissions.² The β -lactam antibiotics is the major class of antibiotic drugs and is among the first-discovered antibiotics.³⁻⁵

The β -lactams drugs function by inhibiting bacterial cell wall synthesis and thus disrupt bacterial reproduction.³ All β -lactam antibiotics share the common structural feature of carrying a β -lactam four member ring as the central functional group (Figure 1.1a).⁴ Common β -lactam-based antibiotic families are distinguished by their extended scaffolds: Penam (Penicillins), Cephem (Cephalosporins), and Carbapenem (Figure 1.1a).⁴⁻⁶ Since their application, β -lactams have demonstrated clinical effectiveness against bacterial infections. However, the abuse of β -lactam drugs has also elevated many bacterial strains to β -lactamases-producing superbugs, which could effectively inactivate common β -lactam antibiotics families.⁷⁻¹⁸

β -Lactamases are bacterial-produced enzymes that are able to effectively hydrolyze and confer board resistance to β -lactams.^{16,17} Based on their mechanisms of action, β -lactamases are divided into four classes: A, B, C, and D.¹⁹ The class A serine-based β -lactamases (AS β LS) represent a severe threat due to their prevalence in infectious strains and affinity to a wide range

of β -lactams.^{7,8,17,18} AS β LS are characterized by their conserved functional residues at the active site: Ser70, Lys73, Ser130, and Glu166. A widely-accepted catalytic mechanism has been proposed that β -lactam hydrolysis in AS β LS is a serine-mediated acylation-deacylation process (Figure 1.1b). The acylation pathways have shown flexibility as this process could be mediated by either Glu166 along or concertedly with Lys73 as the general base.¹⁷ On the other hand, the deacylation pathways of AS β LS can only be mediated by Glu166 as the general base.

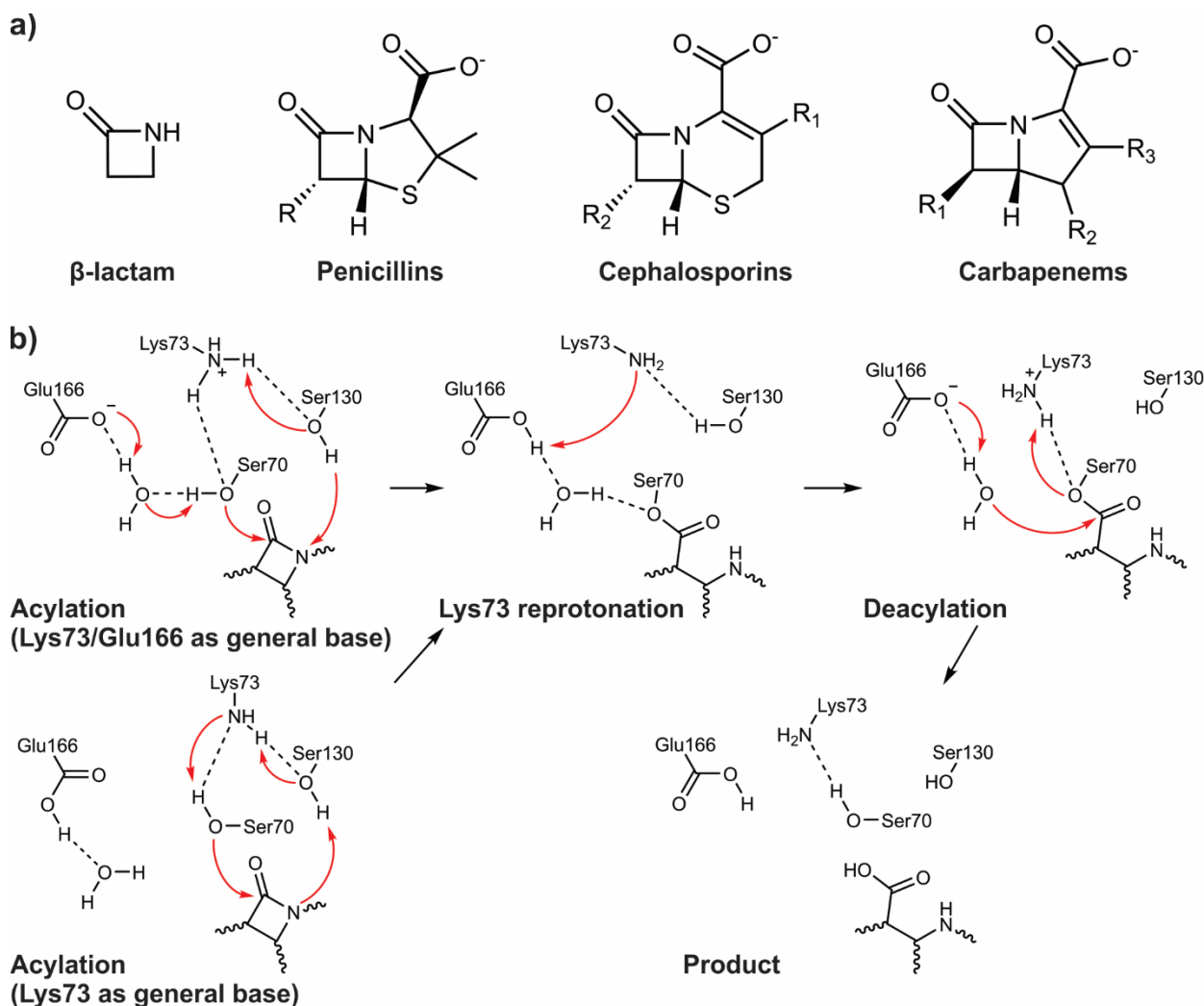


Figure 1.1 Chemical structures of representative β -lactam families and mechanism of β -lactam hydrolysis in AS β LS. (a) The β -lactam four membered-ring, the penicillin, cephalosporin, and carbapenem scaffolds; (b) The hydrolysis mechanisms of β -lactam mediated by AS β LS.

In this thesis, the hydrolysis reactions of β -lactams catalyzed by multiple representative AS β LS are investigated using multiscale Quantum Mechanical / Molecular Mechanical (QM/MM) simulations. Detailed molecular mechanisms of β -lactam resistance driven by AS β LS were revealed by explainable ML methods. The rest of the thesis is organized as follows:

In chapter 2, the computational methods used are introduced, including: the basics of molecular mechanics, the QM/MM as a computational technique for simulating catalytic reactions in biomacromolecules (enzymes), the chain-of-states optimization algorithms for determining optimal reaction pathways, Deep-Learning (DL) and Graph-Learning (GL) methods, and explainability of ML models.

In chapter 3, a classical model of β -lactam resistance, benzylpenicillin acylation in TEM-1, is revisited with QM/MM chain-of-states calculations and ML regression models. Two types of energy contribution attributing methods, coined the perturbation-based Intrinsic Energy Contribution and Dynamic Energy Contribution, were defined to linearly quantify the energy contributions from the chemical events that are concerted in nature. Both methods are shown to be quantitatively consistent and model-implementation invariant, which are tested on three ML-based regression models. I found that the two reacting phases during acylation, tetrahedral formation and tetrahedral collapsing, are partially concerted steps during the acylation. Moreover, this study also bridged and explained the discrepancy between the conclusions of two pioneering QM/MM studies on the rate limiting steps of this reaction.

In chapter 4, extensive QM/MM calculations were performed to identify the acylation pathways adopted by different β -lactam classes. Briefly, I investigated the Toho-1 AS β L acylation with ampicillin (AMP) and cefalexin (CEX) antibiotics, which belong to the β -lactam family of penams and cepheams, respectively. It was found that the Glu166 mediated acylation pathways are

viable for AMP but prohibitive for CEX. The acylation pathway selection for general penam and cephem scaffolds are further discussed.

In chapter 5, based on the initial QM/MM investigations on Toho-1/AMP and Toho-1/CEX, a computational scheme was proposed to achieve the fast sampling of high-quality QM/MM minimum energy pathways (MEPs). A DL neural network with the deep-and-wide architecture was implemented to successfully learn the QM/MM MEPs within chemical accuracy ($< 1.0 \text{ kcal mol}^{-1}$). I further developed an explainable artificial intelligence (XAI) approach to explain the kinetics and mechanistic difference observed for AMP and CEX acylation of AS β Ls. This XAI method, coined Boltzmann-weighted Cumulative Integrated Gradients (BCIG), is based on the Integrated Gradients (IG) approach. In my QM/MM validating calculations, BCIG could correctly attribute energy contributions of individual chemical processes/steps that aligns with chemistry intuitions.

In chapter 6, GL methods were applied to study the resistance conferred by GES-5 AS β L against imipenem. The GES-5 deacylation of imipenem is critical for understand the molecular mechanisms underlying carbapenem resistance. I investigated two tautomer states on the imipenem pyrroline ring, which are known to correlate with the deacylation kinetics of carbapenems. An edge-conditioned graph convolutional neural network (ECGCNN) was implemented to accurately predict the deacylation barrier from the graph representation of the GES-5/Imipenem acyl-enzyme configurations. A perturbative approach was proposed to guide the mechanistic understanding of the deacylation mechanism. The imipenem pyrroline tautomerization states and the 6 α -hydroxyethyl rotamer was revealed to impact the energy barriers of the deacylation. The potential of DL/GL methods for post analysis QM/MM calculations was demonstrated.

Lastly, in chapter 7, The force field parametrization of 2-aminothiazole molecule, which is a common fragment found in β -lactams, was presented. The parametrization protocol follows that proposed for CHARMM General Force Fields (CGenFF). Briefly, all atom types (thus the Van der Waals parameters) were taken from CGenFF. The atomic partial charges on each atom were optimized from fitting the water interaction energies. The equilibrium bond lengths and angle were fitted to the equilibrium geometries optimized from QM calculations. The classical force constants and dihedral terms were parametrized to reproduce the vibrational frequencies at high QM level. The optimized parameters were attached by the end of the thesis.

Above all, this thesis explores the possibility of ML-assisted analysis for QM/MM calculations to extract essential mechanistic insights. Several criteria for probing the ML models trained on QM/MM MEPs were proposed to effectively unravel the mechanistic basis of AS β L-mediated β -lactam hydrolysis. This information is useful to understand AS β L evolution under selective pressure posed by the application of antibiotic drugs.

2. METHODOLOGIES

2.1 Classical Molecular Mechanical Potentials

The idea of classical MD is to integrate particle displacement through time according to the Newtonian equation of motion,

$$m_i \ddot{\mathbf{r}}_i = -\nabla U(\mathbf{r}_i) \quad (2.1)$$

where m_i and $\mathbf{r}_i$ are the mass and coordinate of the particle i at time t , respectively. $\nabla U(\mathbf{r}_i)$ is the gradient of the potential function which describes the interactions between the particles. In classical molecular mechanics, the potential U normally refers to a molecular force field. In its simplest form, classical additive molecular force field is composed of a bonded part and a nonbonded part,^{20–22}

$$U^{MM} = U_{Bonded}^{MM} + U_{Nonbonded}^{MM} \quad (2.2)$$

The bonded part adopts the Hooke's law of elastic springs to account for the bond stretching, bending, and torsional potentials contributed by bonded atom groups,

$$U_{Bonded}^{MM} = \sum_{Stretching} k_b (b - b_0)^2 + \sum_{Bending} k_\theta (\theta - \theta_0)^2 + \sum_{Torsion} k_\varphi (1 + \cos(n\varphi - \delta)) \quad (2.3)$$

where the k_b , k_θ , and k_φ denote the force constants of the bond stretching, bending, and torsional terms, respectively; b_0 and θ_0 denote the bond lengths and angles at equilibrium; n and δ are the

periodicity and phase shift of the rotational potentials of dihedral angles. Note that one dihedral angle type could adopt multiple torsional terms to correctly reproduce the rotational profile.

The nonbonded part of the force field potential includes the nonbonded interactions between all atom pairs and is normally scaled or discarded for the atom pairs already included in the bonded terms. Normally, the nonbonded potential is calculated as the sum of the classical electrostatic and Van der Waals interactions,

$$U_{Nonbonded}^{MM} = \sum_{Van\ der\ Waals} \epsilon_{ij} \left[\left(\frac{R_{min,ij}}{\mathbf{r}_{ij}} \right)^{12} - \left(\frac{R_{min,ij}}{\mathbf{r}_{ij}} \right)^6 \right] + \frac{1}{4\pi\epsilon_0} \sum_{Coulomb} \frac{q_i q_j}{\mathbf{r}_{ij}} \quad (2.4)$$

In the Van der Waals part, ϵ_{ij} is the well-depth of the 6-12 Lennard-Jones potential and $R_{min, ij}$ is the Van der Waals radius. In the electrostatic part, ϵ_0 is the dielectric constant, q_i and q_j are the partial atomic charges on the interacting atoms. $\mathbf{r}_{ij}$ is the atomic distance between atoms i and j .

Additional correction terms to the minimal force field model are commonly introduced for improving the accuracy of the classical potentials. In the most popular CHARMM force fields for proteins²³, for example, Urey-Bradley terms that define pseudo-bonds between the 1-3 atom pairs in an angle are introduced to accurately reproduce the vibrational spectra,

$$U_{Urey-Bradley}^{MM} = \sum_{Urey-Bradley} k_u (u - u_0)^2 \quad (2.5)$$

where k_u is the spring force constant on the pseudo-bond and u_0 the equilibrium length between the two atoms. Improper dihedral terms are incorporated for better treatments of the out-of-plane bending motions,

$$U_{Improper}^{MM} = \sum_{Improper} k_\omega (\omega - \omega_0)^2 \quad (2.6)$$

where k_{ω} is the spring force constant on the improper dihedral and ω_0 the equilibrium angle. Furthermore, numerical grid-based potential correction such as the CMAP method has been implemented for the general improvement of protein backbone sampling.²⁴

Modern force field parameters are optimized from fitting procedures targeting experimental data and high-level QM calculations. In order to reduce the vast number of parameters needed to define a molecule, one adopts the idea of atom typing to identically treat the potentials contributed by atoms under the same physical environment, based on which the bond types, angle types, and dihedral types could be accordingly assigned. While each additive force field family adopts functional forms with different correction terms and fits against different target data, this thesis focuses only on the parametrization protocols adopted by CHARMM (Chapter 7).

Classical force field potentials are robust descriptors of biomacromolecules such as proteins. The relatively cheap computational demand of molecular force field potential and gradient computations permits classical MD simulations to be performed at longer time-scales needed to access the molecular properties of various biophysical processes.²¹

2.2 Hybrid Quantum Mechanical/Molecular Mechanical Potentials

The classical potentials based on force field functions ignore the electron degrees of freedom and are held to several intrinsic limitations. Its quadratic functional form of the bonded terms does not permit the dissociation of bonded atom pairs nor the formation of new bonds. The QM/MM approach, which treats specific regions of interest at QM level and the rest at MM level, has been proposed to enable the simulations of chemical reactions in complex biomolecular systems (Figure 2.1).^{25–29}

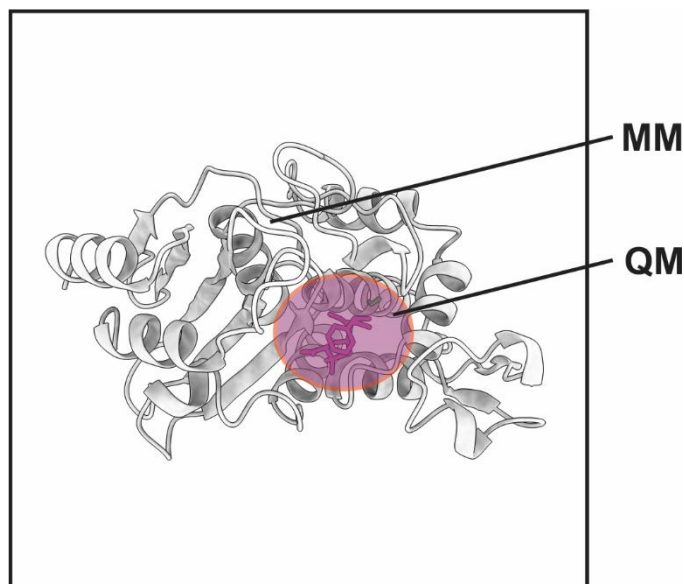


Figure 2.1 The QM/MM partitioning on biomolecular simulating systems.

In the additive QM/MM formulism, the hybrid potential is the sum of the QM and MM Hamiltonians and the QM/MM coupling,

$$\hat{H}_{total} = \hat{H}_{QM} + \hat{H}_{MM} + \hat{H}_{QM/MM} \quad (2.7)$$

The essential focus of the QM/MM approach is the coupling scheme applied to treat the interaction between QM and MM regions. In practice, the coupling Hamiltonian $\hat{H}_{QM/MM}$ is given as

$$\hat{H}_{QM/MM} = \hat{H}_{QM/MM}^{Van\ der\ Waals} + \hat{H}_{QM/MM}^{Electrostatic} + \hat{H}_{QM/MM}^{Bonded} \quad (2.8)$$

The $\hat{H}_{QM/MM}^{Bonded}$ term notes the QM/MM boundary condition which splits a molecule into a QM fragment and a MM fragment. For example, the side chains of amino acid residues which participate the chemical reactions are normally partitioned into the QM region while the backbones are treated at MM level. This term is commonly implemented in the single “link-atom” approach,

where pseudo-hydrogen atoms are introduced to partition the chemical bond into the QM and MM region (Figure 2.2).³⁰ These pseudo-hydrogens are used to complement the valence of the QM hosting atoms and is only treated quantum mechanically. Additional research efforts for developing accurate QM/MM boundary potentials include the frozen orbitals,³¹ generalized hybrid orbitals,³² or pseudobonds.³³ While conceptually more physical, the improvements from these methods over the single link-atom scheme are inconclusive.³⁰

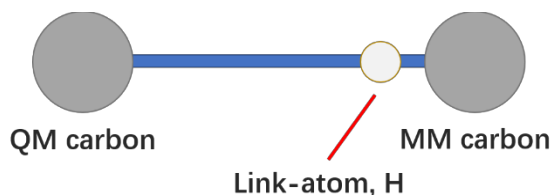


Figure 2.2 The single link-atom approach for treating QM/MM boundary potential.

Essentially, the treatments applied to the nonbonded interactions between QM and MM regions distinguish the embedding scheme of the QM region into the MM environments.²⁶ In the simplest mechanical embedding, all QM/MM coupling interactions are treated purely at MM level,

$$\hat{H}_{QM/MM}^{Van\ der\ Waals} = \sum_{A,M} \epsilon_{A,M} \left[\left(\frac{R_{min,ij}}{\mathbf{r}_{A,M}} \right)^{12} - \left(\frac{R_{min,ij}}{\mathbf{r}_{A,M}} \right)^6 \right] \quad (2.9)$$

$$\hat{H}_{QM/MM}^{Electrostatic} = \frac{1}{4\pi\epsilon_0} \sum_{A,M} \frac{q_A q_M}{\mathbf{r}_{A,M}} \quad (2.10)$$

where label A denotes the QM atoms and M the MM atoms.

The major defect of the mechanical embedding scheme is that for fixed-charge force field models, the partial charges on the QM atoms used for computing the QM/MM electrostatics are

invariant with regard to possible QM atom displacements.²⁸ This poses an unrealistic approximation when bond making or breaking occurs between the QM atoms. Furthermore, the absence of MM environments from the QM Hamiltonian also presents another limitation that the QM subsystem is actually computed in vacuum.

As a general improvement over the mechanical embedding, the most popular electrostatic embedding scheme has been proposed to incorporate the MM point charges into the QM wave functions as one-electron integrals in the Fock matrix, allowing the QM density to be polarized by the external field of point charges. The electrostatic part of QM/MM coupling is then given as

$$\hat{H}_{QM/MM}^{Electrostatic} = - \int d\mathbf{r} \rho(\mathbf{r}) \sum_M \frac{q_M}{|\mathbf{r} - \mathbf{r}_M|} + \frac{1}{4\pi\epsilon_0} \sum_{A,M} \frac{Q_A q_M}{r_{A,M}} \quad (2.11)$$

where the $\rho(\mathbf{r})$ is the density distribution of the QM electrons and Q_A is the atomic charge of the QM nuclei. Accordingly, the $\hat{H}_{QM/MM}^{Bonded}$ term in the electrostatic embedding scheme needs special care. The linking hydrogen atoms are normally placed on the QM-MM bonds and maintain realistic link atom-QM host bond lengths (~ 1.1 Å for C-H bonds). Under the electrostatic embedding scheme, the point charges on the MM hosts would appear too close to the link atom and induce unphysical over-polarizations to the QM density. As a general fix, the partial charges on the MM hosts are deleted or shifted in common practice.^{34,35}

Semiempirical parametrized QM methods such as the Tight-Binding models has been developed for accurate and efficient sampling of reaction profiles.^{36,37} Notably, in many parametrized QM models, such as the third order Density Functional Tight Binding (DFTB3) model,^{38,39} the QM/MM electrostatic interaction was simply implemented as the Coulomb potential between the QM Mulliken charges and MM point charges. Robust simplifications as such

not only lead to more accurate treatment of boundary potentials but also enable efficient QM/MM electrostatic calculation using Ewald summation⁴⁰ and related methods⁴¹.

Finally, the QM/MM electrostatic embedding scheme does not consider the MM electronic degrees of freedom. This signifies the direction for future developments of the QM/MM model to the polarization embedding scheme, where the MM polarizability is included by means of polarizable force fields.^{20,21}

2.3 Chain-of-states MEP Optimizations

Sampling rare events and locating optimal transition paths are important for detailed mechanistic understanding of the dynamics on the system potential.^{42–44} As a common practice, enhanced sampling methods based on reaction coordinates or collective variables explore the (free) energy surface *via* bias potentials that forcingly drive the sampling towards the desired high energy states (rare events), which are thermodynamically inaccessible under equilibrium time scales.⁴² Essential information such as free energies can be obtained from debiasing the sampled ensembles.^{45,46} Alternatively, enhanced sampling methods in the chain-of-states regime explore the potential/free energy landscape by representing the transition path with a series of discrete conformations (termed replicas or images) between the minimized/equilibrium states.^{47–50} This chain of replicas is subjected to optimizations with inter-replica interactions (restraints or holonomic constraints) to prevent the intermediate replicas from falling to the low energy basin (Figure 2.3), thus locating the high energy rare events along the transition path. In general, different chain-of-states methods are characterized by the type of replica interaction applied during the pathway optimization. This inter-replica interactions could be presented as restraining quadratic potentials in the generalized coordinates space (such as the Nudged Elastic Band methods^{51–53}) or root-mean-square (RMS) space (such as the Replica Path Method, RPM, with restraints)⁴⁸, implicit

holonomic constraints *via* the reparameterization trick (such as the String Methods^{50,54–57} and its simplified version without force projections⁵⁸), or explicit holonomic constraints in the 2-norm distance space (such as the RPM with constraints⁴⁹). This thesis does not aim to exhaustively review the chain-of-states and related free energy calculations. Instead, the algorithm adopted by the RPM with holonomic constraints from which all MEPs were optimized in the rest of the thesis is formally detailed. An example implementation of the RPM and related chain-of-states methods on a Muller potential is provided at: github.com/ZL-Song/MullerPot.

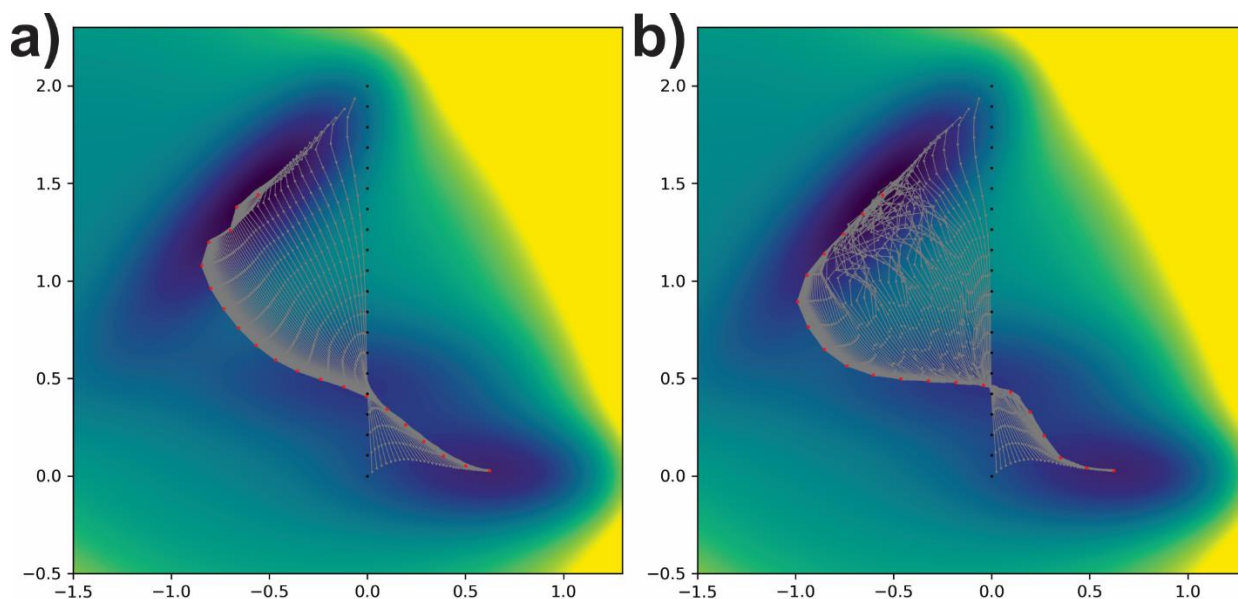


Figure 2.3 The optimization of the MEP on a Müller potential using chain-of-states methods with a steepest decent optimizer. (a) The RPM with holonomic constraints; (b) The string method with force projections. Black dots refer the initial guess and red dots the final optimized path, gray trajectories are the optimization steps.

The p -norm distance between two vectors is noted as $\|\mathbf{a} - \mathbf{b}\|_p$. Accordingly, the 2-norm distance between the i -th and $i+1$ -th replica (with coordinate vectors $\mathbf{r}_i$ and $\mathbf{r}_{i+1}$, respectively), $\Delta l_{i,i+1}$ is given as

$$\Delta l_{i,i+1} = \|\mathbf{r}_i - \mathbf{r}_{i+1}\|_2 \quad (2.12)$$

and its gradients with regard to $\mathbf{r}_i$,

$$\frac{\partial \Delta l_{i,i+1}}{\partial \mathbf{r}_i} = \frac{\mathbf{r}_i - \mathbf{r}_{i+1}}{\|\mathbf{r}_i - \mathbf{r}_{i+1}\|_2} \quad (2.13)$$

$$\frac{\partial \Delta l_{i,i+1}}{\partial \mathbf{r}_i} = \frac{\mathbf{r}_i - \mathbf{r}_{i+1}}{\Delta l_{i,i+1}} \quad (2.14)$$

It is trivial to note that $\Delta l_{i,i+1}$ could be calculated as mass-weighted best-fit distances.

Referring to the original implementation of Brokaw et al.,⁴⁹ the RPM with holonomic constraints seeks to optimize a MEP between to minimal states as a chain of i replicas with equal 2-norm distances between adjacent replicas, that is

$$\Delta l_{0,1} = \Delta l_{1,2} = \dots = \Delta l_{i-2,i-1} = \bar{\Delta l} \quad (2.15)$$

The search for the constrained replica coordinates during each pathway optimization step follows the common iterative solution of optimization problem under holonomic constraints.⁵⁹ Practically, following each optimization step which changes the replica coordinates, the replicas are heterogeneously distributed along the chain and are updated using an iterative procedure with Lagrangian multipliers to satisfy the desired constraint conditions. In the RPM with holonomic constraint method, the desired distance between all pairs of adjacent replicas, $\bar{\Delta l}$, is the average 2-norm distance from the initial heterogeneously distributed replica path. Therefore, the constrained replica coordinates $\mathbf{r}_i$ is solved by the following update functions,

$$(\mathbf{r}_i)^{(n+1)} = \left(\mathbf{r}_i + \lambda_{i-1} \frac{\partial \Delta l_{i-1,i}}{\partial \mathbf{r}_i} + \lambda_i \frac{\partial \Delta l_{i,i+1}}{\partial \mathbf{r}_i} \right)^{(n)} \quad (2.16)$$

$$(\mathbf{r}_i)^{(n+1)} = \left(\mathbf{r}_i - \lambda_{i-1} \frac{\mathbf{r}_{i-1} - \mathbf{r}_i}{\Delta l_{i-1,i}} + \lambda_i \frac{\mathbf{r}_i - \mathbf{r}_{i+1}}{\Delta l_{i,i+1}} \right)^{(n)} \quad (2.17)$$

where the superscript (n) denotes the n -th iteration, λ_i is the Lagrangian multiplier at $\Delta l_{i,i+1}$. The convergence of the MEP is met when

$$(\Delta l_{i,i+1})^{(n+1)} - \bar{\Delta l} \cong 0 \quad (2.18)$$

Since Δl is a function of λ , expand $(\Delta l_{i,i+1})^{(n+1)}$ via the Taylor series with regard to $\lambda^{(n)}$ to the first order yields

$$(\Delta l_{i,i+1})^{(n)} + \left(\frac{\partial \Delta l_{i,i+1}}{\partial \lambda} \right)^{(n)} (\lambda^{(n+1)} - \lambda^{(n)}) - \bar{\Delta l} = 0 \quad (2.19)$$

Assuming convergence gives $\lambda_i^{(n+1)} = 0$ and descending along the negative gradients:

$$(\Delta l_{i,i+1})^{(n)} + \left(\frac{\partial \Delta l_{i,i+1}}{\partial \lambda} \right)^{(n)} (-\lambda^{(n)}) - \bar{\Delta l} = 0 \quad (2.20)$$

$$\bar{\Delta l} - (\Delta l_{i,i+1})^{(n)} = \left(\frac{\partial \Delta l_{i,i+1}}{\partial \lambda} \right)^{(n)} (\lambda^{(n)}) \quad (2.21)$$

Plugging equations 2.14 and 2.17 to equation 2.21 gives

$$\bar{\Delta l} - (\Delta l_i)^{(n)} = \left(\frac{\mathbf{r}_i - \mathbf{r}_{i+1}}{\Delta l_i} \right)^{(n)} \left(\frac{\mathbf{r}_{i-1} - \mathbf{r}_i}{\Delta l_{i-1}} \lambda_{i-1} + 2 \frac{\mathbf{r}_i - \mathbf{r}_{i+1}}{\Delta l_i} \lambda_i - \frac{\mathbf{r}_{i+1} - \mathbf{r}_{i+2}}{\Delta l_{i+1}} \lambda_{i+1} \right)^{(n-1)} \quad (2.22)$$

Finally, by enumerating equation 2.22 for all replicas ($i = 0, \dots, i-1$, equation 2.15), one obtains a tridiagonal matrix of λ which can be solved in most linear algebra packages. Iteratively

solving this matrix for λ and $\mathbf{r}$ will converge all inter-replica distances to $\bar{\Delta}l$. The convergence threshold (equation 2.18) is normally 10^{-8} Å.

2.4 Machine Learning and Model Explanations

ML has emerged with great promise to approximate target function of any form regardless of the *a priori* knowledge about the underlying correlations among input variables. The applications of various ML techniques have also advanced theoretical chemistry in various subjects^{60–63}, which have been suffering from either the extensive computational demands of high levels of theories^{64–69}, or the high dimensionality of the chemical and/or conformational spaces^{70–74}. Although ML could be introduced to many topics that require accurate and efficient approximations, its performance and effectiveness have been limited by feature representations and model interpretability.⁷⁵ In addition to the routinely applied feature representations, unsupervised models and rational statistical procedures have been developed to extract robust feature vectors from the chemical feature space.^{76,77} In particular, considerable pioneering efforts have focused on the development of suitable descriptors and accurate DL neural networks for approximating hybrid Quantum Mechanical / Molecular Mechanical (QM/MM) potentials.^{78–83}

Being an emerging subarea of ML, GL applies DL-based techniques on graph-structured data. Graph structures could encode data representation using vertices and interconnecting edges.⁸⁴ Promoted by the robust graph representation of chemical structures, various GL models have achieved ground-breaking performances on molecular property predictions.^{85–89} In this thesis, I provide only a brief introduction to ECGCNN model used in Chapter 6.

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 (features). Thus,

the strategy for updating the hidden state on vertex v through a graph convolutional layer follows a message passing scheme^{84,90} of three steps: (1) For each vertex v in the graph, the directed edge $(w \rightarrow v)$ from the neighboring vertices w is first encoded by a *Message()* function; (2) The encoded message is introduced to the hidden representation of v by an *Aggregate()* function; (3) The output hidden state of the graph convolutional layer is then produced by an *Update()* function from the aggregated representation of v . In the most general form, a GL convolutional layer updates the hidden state $\mathbf{x}_v^{(l)}$ of the vertex v at the l -th layer by

$$\mathbf{x}_v^{(l)} = \text{Update} \left(\mathbf{x}_v^{(l-1)}, \text{Aggregate} \left(\mathbf{x}_v^{(l-1)}, \text{Message} \left(\mathbf{x}_v^{(l-1)}, \mathbf{x}_w^{(l-1)}, \mathbf{e}_{w \rightarrow v}^{(l-1)} \right) \right) \right) \quad (2.23)$$

where $\mathbf{x}_w$ denotes the hidden representations of the vertices connected to v ; $\mathbf{e}_{w \rightarrow v}^{(l-1)}$ 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()*. The edge-conditioned GL⁹¹ for ECGCNN incorporates the edge features into the *Message()* function by learning a hidden representation for $\mathbf{e}^{(l)}$ and uses it as the weight matrix that is multiplied to the neighboring hidden states $\mathbf{x}_w^{(l-1)}$

$$\text{Message} \left(\mathbf{x}_w^{(l-1)}, \mathbf{e}_{w \rightarrow v}^{(l-1)} \right) = \sum_w \mathbf{x}_w^{(l-1)} A \left(\mathbf{e}_{w \rightarrow v}^{(l-1)} \right) \quad (2.24)$$

where $A()$ is a differentiable function at the edge-conditioned convolutional layer, l . In practice, $A()$ is normally implemented as multilayer perceptron that maps $\mathbf{e}_{w \rightarrow v}^{(l-1)}$ to the shape of weight matrix applicable to $\mathbf{x}_w^{(l-1)}$.

With the booming popularity of ML, interests to interpret DL/GL neural networks have synergistically risen as a subfield of great importance, namely the XAI.⁹² The ML models being

interpretable not only elevates our understanding of the learning algorithms, but also constitutes responsible DL/GL-assisted decision making. Practically, XAI techniques attribute the predicted outcome of DL/GL models to individual feature contributions, therefore rationalize the driving forces behind the decision flow in the models that are black-boxes. While explicit indicators for feature contributions are straightforward in linear models^{93,94} and are incorporated by design in specific ensemble-based models⁹⁵, explaining neural networks is in general hindered by the high nonlinearity accumulated through the activations of the hidden layers.⁹⁶

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.⁹⁷ The state-of-the-art XAI techniques, such as the IG⁹⁸ and the Layer-wise Relevance Propagation⁹⁹, have demonstrated great promise in various explaining tasks such as medical diagnosis^{100,101} and cheminformatics applications^{102,103}. Alternatively, feature importance can also be assigned based on data perturbations and/or model re-learning. As one of common practice, one could drop or permute a feature and re-learn the model, the performance difference between the original and the re-learned models can be used as the indicator for the contribution from that feature. While conceptually more intuitive, cautions have to be taken for this approach that the perturbation introduced to the feature must not change the native distribution of the data on which the learning models were trained.¹⁰⁴

In this thesis, both gradient-based and perturbation-based explanation techniques are presented, developed, and applied to ML models trained on QM/MM MEPs of AS β L-driven β -lactam resistance.

3. REVISITING TEM-1/BENZYL PENICILLIN ACYLATION MECHANISM WITH MACHINE-LEARNING ENERGY CONTRIBUTION ANALYSIS

3.1 TEM-1 β -Lactamases

TEM-1 is a representative AS β L and the most common β -lactamase among Gram-negative bacterial strains.¹⁷ Numerous experimental and computational studies have been carried out to delineate the functions of the residues at the catalytic binding site.^{105–113} Based on these studies, one widely accepted mechanism was proposed that Glu166 acts as a general base during the acylation process of benzylpenicillin hydrolysis (Figure 3.1a). The hydroxyl group of Ser70 first attacks the β -lactam carbonyl carbon to form a tetrahedral intermediate, with its proton delivered to the bridging catalytic water. The catalytic water molecule in turn donates a proton to the deprotonated carboxyl group of Glu166. Upon the formation of the tetrahedral intermediate, the fully protonated Lys73 activates the nearby Ser130 to protonate the β -lactam nitrogen, which cleaves the β -lactam scissile bond and completes the acylation half of β -lactam hydrolysis. Other residues including Asn170 and Ser235 were also validated to contribute hydrogen bonding interactions that are critical for the formation of the Michaelis complex between TEM-1 and the benzylpenicillin substrate.^{112,113}

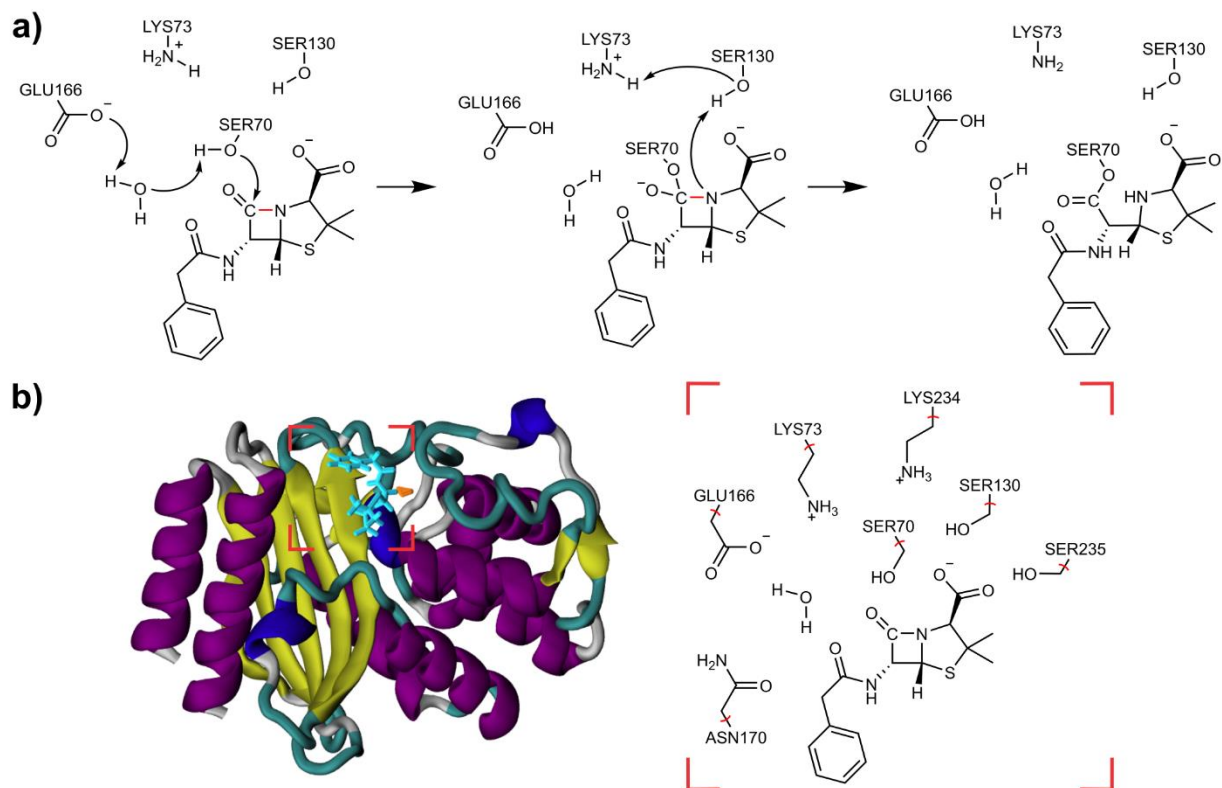


Figure 3.1 Acylation mechanism of ASβLs and the structure of TEM-1/benzylpenicillin Michaelis complex. (a) Acylation mechanism of TEM-1 and benzylpenicillin with Glu166 acting as a general base. The β-lactam scissile bond is noted in red; (b) Crystal structure of TEM-1 complexed with benzylpenicillin and the selection of QM atoms.

3.2 Computational Details

QM/MM calculations. All hybrid QM/MM multiscale calculations in the present study were conducted by interfacing CHARMM¹¹⁴ with SCC-DFTB¹¹⁵ or Q-Chem^{116,117}. All MD simulations were performed by OpenMM¹¹⁸. The acyl-enzyme product of TEM-1 with benzylpenicillin was obtained from the X-ray crystal structure (PDB id: 1FQG)¹¹⁹ and the mutant residue Asn166 was modified to Glu166 as in the wild type TEM-1. The residues were then protonated according to previous studies.^{105,106} The system was solvated and sodium and chloride ions were added to

balance the total charge of the system. In order to fully relax the system, classical mechanic minimization and equilibration were performed with the CHARMM36 force field for proteins²³, CHARMM general force field (CGenFF)^{120–122} for the penicillin molecule and TIP3P model for water¹²³. The structure of the QM/MM initial pathway calculation was taken from the trajectory of a 10 ns MD simulation at 300 K. The RPM with holonomic constraints⁴⁹ was applied for reaction pathway calculations. All the pathway calculations were carried out with 50 replicas.

Reaction pathway sampling. The initial pathway was calculated from DFTB3/mio/C36 level of theory with any residue in the outer 15 Å of QM region selected as the unfrozen MM region. Based on the initial pathway, multiple reaction pathways were sampled. Firstly, three replicas representing reactant (r), transition (t) and product (p) states were selected. 200 ns MD simulations were performed on each of the selected replicas. During the MD runs, all the atoms in the QM region were fixed and snapshots were taken every 0.1 ps. 2-dimensional Principal Component Analysis (PCA) were performed on the MD trajectories with the pairwise C α distances as input. The PCA results was grouped into six clusters, and the snapshots that are the closest to the centers of each cluster were chosen as the representative structures. A total of eighteen representative structures were then selected. In order to retain the consistency among the QM/MM pathway's energetic profiles, a common MM region was used, which is selected to be the union set of residues within the outer 10 Å of all representative QM regions. Geometry optimizations were then performed on the selected representative structures. Lastly, based on those representative structures, eighteen RPM calculations were carried out to obtain the MEPs.

Machine-learning protocols. The scikit-learn package¹²⁴ was employed for various machine-learning protocols including dimensionality reduction, clustering, and regression. The hydrogen bonding interactions are identified via the Baker-Hubbard criteria¹²⁵ as implemented in MDTraj¹²⁶.

The radial basis function was used as the kernel function for all regression models: Support Vector regression (SVR), Gaussian Process regression (GPR), and Kernel Ridge regression (KRR). For the training-validation process of models, the leave-one-group-out cross-validation regression analysis was employed in the validation step; the hyper-parameters of the models were tuned *via* a grid search strategy.

3.3 Summary of Results

TEM-1/Benzylpenicillin Acylation Profiles. As highlighted in the black rectangles in Figure 3.2, the carbonyl tetrahedral intermediate state could be obtained from all B3LYP reaction pathway optimizations. However, 16 out of 18 DFTB3 optimized pathways demonstrated that the intermediates are lower in energy than the reactant, whereas all B3LYP pathways show that the energies of tetrahedral states are well elevated from the reactant. In addition, the tetrahedral intermediates from my DFTB3 calculations are structurally different from Hermann et al. The average distance between Ser70 O γ and the carbonyl carbon is 2.1 Å, comparing to 1.45 Å reported by Hermann et al.¹⁰⁵ Also, it is noted that tetrahedral intermediates from my DFTB3 calculations are accompanied by a hydronium formed by the catalytic water and negatively charged Glu166, whereas Hermann et al.¹⁰⁵ observed a neutral catalytic water and protonated Glu166. Such disagreement could originate from the fundamental difference between the QM methodologies. Although the DFTB3/mio/C36 optimized pathways provide acylation barriers that are in good agreements with experiments, the configurational changes along the chain-of-replicas may not be reliable.

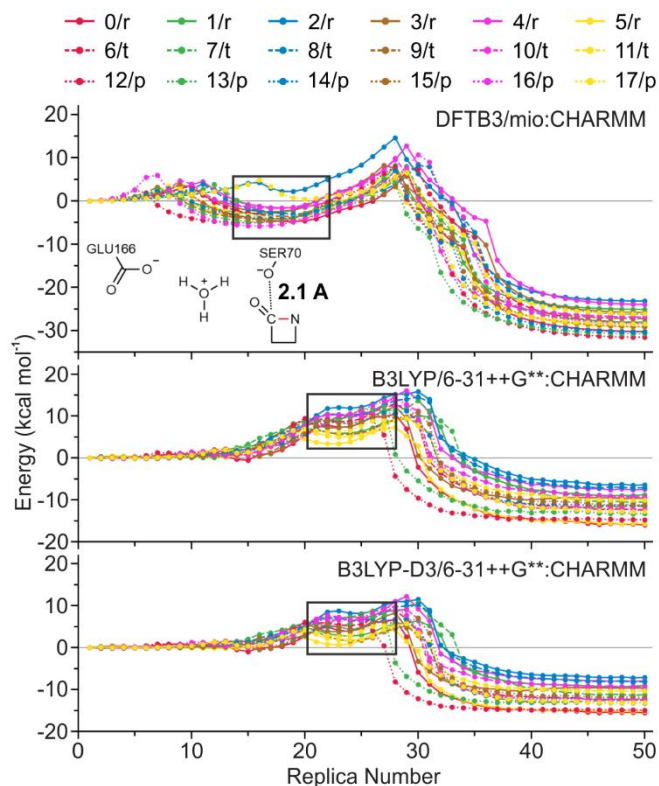


Figure 3.2 QM/MM chain-of-states pathway profiles. Reaction pathways calculated from DFTB3/mio/C36, B3LYP/6-31++G**/ C36, and B3LYP-D3/6-31++G**/ C36 levels of theory. The black rectangles highlight the tetrahedral intermediates region along the energy profiles.

On the other hand, my B3LYP optimized reaction pathways agree with the results reported by Meroueh et al.,¹⁰⁶ showing that the potential energies of the tetrahedral intermediate are elevated from the reactant. Moreover, the dispersion corrected B3LYP calculations generally led to 3-5 kcal mol⁻¹ decrease of the activation barriers during the acylation, which is consistent with previous observations^{127,128}. Detailed barrier results of the acylation are compared with previous computational and experimental studies in Table 3.1.

Table 3.1 Comparison of acylation energy barriers of the current and previous works.

Source	Energy barriers (kcal mol ⁻¹) [a]			Method [b]
	MC-TI	TI-AE	Overall	
This study	3.6(3)	11.4(1)	11.4(1)	DFTB3/mio/C36, RPM
This study	8.6(9)	3.8(7)	11.9(4)	B3LYP/6-31++G**/C36, RPM
Hermann et al.	19.6	16.4	19.6	AM1/CHARMM, PES
Hermann et al.	8.7	7.1	8.7	B3LYP/6-31+G*/C36, PES
Meroueh et al. [c]	22.0	N/D	22.0	MP2/6-31+G*/AMBER, PES
Gibson et al.	-	-	12.6(7)	293.15 K, Exp
Sirot et al.	-	-	13.0(5)	310.15 K, Exp

[a] MC-TI: Michaelis complex to tetrahedral intermediate; TI-AE: tetrahedral intermediate collapsing to acyl-enzyme product;

[b] RPM: Chain-of-states RPM calculations, averaged over 18 pathways; IRC: Intrinsic Reaction Coordinate calculation; PES: Potential Energy Surface; Exp: derived from k_{cat} via the Eyring equation;

[c] This study uses penicillanic acid instead of benzylpenicillin, the experimental acylation barrier of penicillanic acid is estimated to be 16 – 17 kcal mol⁻¹.

MEP regression models. Predictive PES models were trained to bridge the conformational descriptors of each replica to its corresponding energy. An appropriate selection of features is critical for the performance of machine-learning predictions. In my case, a total of 105 pairwise distances between bonded atoms – either through chemical bonding or hydrogen bonding – in the QM region are considered as initial features. As the size of the dataset (900 replicas) is relatively small compared to the dimension (105 features), regression models are expected to fit poorly and unstably. In order to reduce the dimension of feature vectors, a recursive feature elimination analysis using SVR model with linear kernel function was first performed on both DFTB3/mio/C36 and B3LYP/6-31++G**/C36 pathways. Based on the selected features and my prior knowledge with TEM-1/Benzylpenicillin hydrolysis, 15 interatomic distances were selected

and used to construct the feature vector (Figure 3.3a). Moreover, the overall prediction quality of regression models on B3LYP pathways are promising with the root mean squared error lower than 2.0 kcal mol⁻¹(Figure 3.3b).

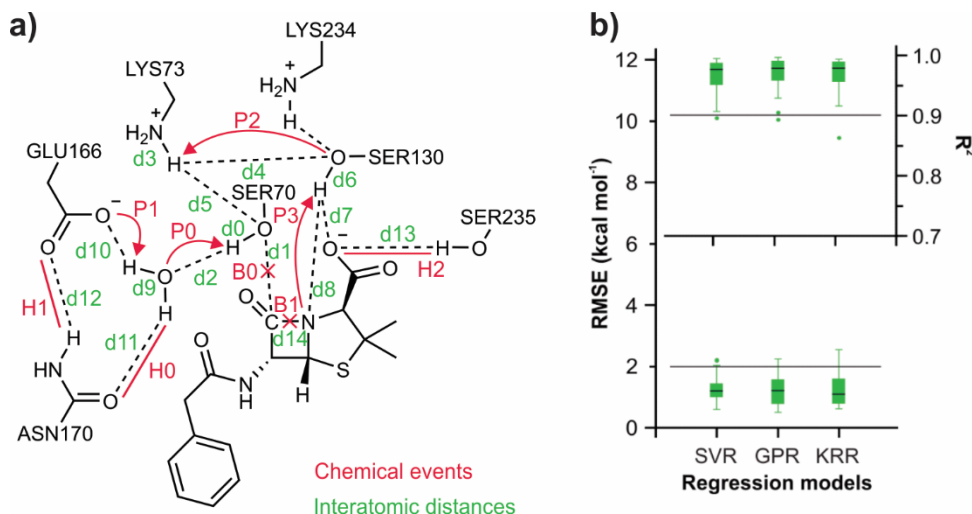


Figure 3.3 Feature selections and notations of chemical processes, and regression performance of the models. (a) Selection of atomic distances as features and chemical events; (b) RMSE predicted by three regression models.

Intrinsic energy contribution. One universal criterion to measure variable contribution is the decrease in prediction performance when a certain feature is dropped out from the model.¹⁰⁴ Practically, I measure the joint contribution of feature subset by the difference between the fitting performance of a predictive model trained from full input feature set and the same model trained with the target feature subset set to zero. In this regard, the intrinsic energy contribution is defined as the RMSE between the predicted energetic pathway profiles of the two models, that is

$$I_{a, \text{intrinsic}} = \left(\frac{1}{R} \sum_{r=1}^R \left(f(\mathbf{A}^{(r)}) - f_{a=0}(\mathbf{A}^{(r)}) \right)^2 \right)^{\frac{1}{2}} \quad (3.1)$$

where R is the total number of replicas on each pathway; f is the trained regression model; $f_{a=0}$ is the same model trained from input data with the target feature subset set to zero; $\mathbf{A}^{(r)}$ is the input feature vector at the r -th replica. For numerical comparisons between different regression models, the measurement used is the percentage of each intrinsic contribution over the sum of all feature subgroups (Figure 3.4).

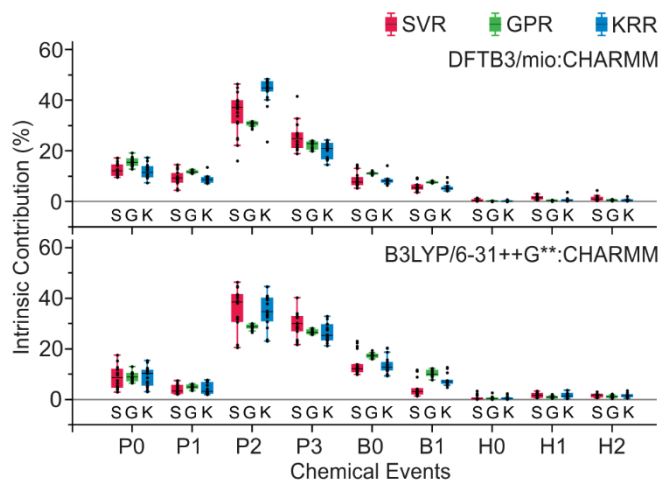


Figure 3.4 Intrinsic energy contributions measured on DFTB3/mio/C36 and B3LYP/6-31++G**/C36 reaction pathway profiles. The ‘S’, ‘G’, and ‘K’ labels represent results from SVR, GPR, and KRR models, respectively.

The intrinsic contribution provides a quantitative insight into the energy contribution of each reaction step to the overall energetic profile. Generally, all regression models give the same statistical rankings of the energy contributions from each chemical event: P2 and P3 are the decisive processes during the reaction; P0, P1, B0 and B1 pose less impact to the overall energetic; Hydrogen bonds (H0, H1 and H2) are considered to be the least critical events. The intrinsic energy contribution measured using the GPR model is the most numerically stable, whereas the SVR

model gives the largest deviation among the testing cases. As for pathway profiles decomposed at different QM levels of theory, the intrinsic contributions are compatible to each other.

Dynamic Energy Contribution. The intrinsic energy contribution reflects the overall energetic contribution of a certain chemical process to the energetic profile. Alternatively, a dynamic energy contribution along the reaction progress could be determined by the model response to small perturbations applied to each feature subset, which could be implemented from the first-order Taylor series:

$$I_{a,dynamic}^{(r)} = |f(\mathbf{A}^{(r)} + \mathbf{P}^{(r)}) - f(\mathbf{A}^{(r)})| = \left| \left(\frac{\partial f(\mathbf{A}^{(r)})}{\partial \mathbf{A}^{(r)}} \right) (\mathbf{P}^{(r)}) \right| \quad (3.2)$$

where $\mathbf{P}^{(r)}$ is the perturbation applied to the MEP regression models. The partial gradients are computed numerically

$$\frac{\partial f(\mathbf{A}^{(r)})}{\partial \mathbf{A}^{(r)}} = \frac{f(\mathbf{A}^{(r)} + \mathbf{P}^{(r)}) - f(\mathbf{A}^{(r)} - \mathbf{P}^{(r)})}{2\mathbf{P}^{(r)}} \quad (3.3)$$

formally $\mathbf{P}^{(r)}$ is given as

$$\mathbf{P}^{(r)} = \mathbf{\Gamma}^{(r)} \mathbf{E}^{(r)} \quad (3.4)$$

where $\mathbf{\Gamma}^{(r)}$ is a one-hot encoded mask for selecting feature dimensions to which the perturbation $\mathbf{E}^{(r)}$ is applied, and

$$\mathbf{E}^{(r)} = p(\mathbf{A}^{(r-1)} - \mathbf{A}^{(r+1)}) \quad (3.5)$$

where p is the amount of perturbation applied and is set as 0.01. Finally,

$$I_{a,dynamic}^{(r)} = \left| f\left(\mathbf{A}^{(r)} + p\mathbf{\Gamma}^{(r)}(\mathbf{A}^{(r-1)} - \mathbf{A}^{(r+1)})\right) - f\left(\mathbf{A}^{(r)} - p\mathbf{\Gamma}^{(r)}(\mathbf{A}^{(r-1)} - \mathbf{A}^{(r+1)})\right) \right| \quad (3.6)$$

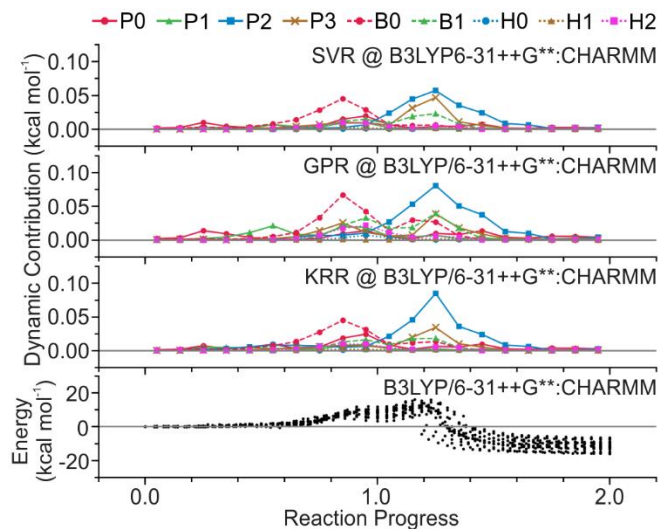


Figure 3.5 Dynamic energy contributions. The dynamic energy contribution measured from SVR, GPR and KRR regression models trained on energetic profiles calculated at B3LYP/6-31++G** level of theory. The values provided are the average over the reaction progress.

As shown in Figure 3.5, the acylation is initialized by the proton transfer between Ser70 and the catalytic water (P0). During the first transition to the tetrahedral intermediate, the bond formation between Ser70 O γ and carbonyl carbon (B0) is deemed to be the most energetic dominant event. Notably, the protonation of the thiazolidine nitrogen (P3) is concerted in this process. The rate determining events of the acylated product formation are the dual protonation of Ser130 (P2) and the β -lactam nitrogen (P3) together with the cleavage of the β -lactam scissile bond (B1). In addition, the dynamic contribution measurement is also regression model-independent, and the small differences in turn reflect the difference in the predicted PES of the regression models. During the formation of tetrahedral intermediate, the rate determining event is shown to be the bond formation between Ser70 hydroxyl oxygen and the β -lactam carbonyl carbon. As for the formation of acyl-enzyme product, the dual-proton transfer from Lys73 to β -lactam nitrogen,

bridged by the Ser130 hydroxyl group, becomes the rate determining event. The dynamic energy contributions are consistent with the intrinsic contribution measurements as they identify the same critical chemical events during the acylation. Generally, the dynamic energy contribution qualitatively reveals the time windows and spans of chemical events and quantitatively reflects their underlying correlations.

3.4 Conclusion

Qualitative agreement between the dynamic and the intrinsic energy contribution assessments is observed. The underlying correlations between the transfers of different protons are validated in the dynamic energy contribution. Notably, the protonation of Ser130 hydroxyl group and the thiazolidine nitrogen are found to be concerted with the formation of tetrahedral intermediate, indicating that the acylation reaction is a one-step 4-proton-transfer process. Isolating such proton transfers from the tetrahedral formation process has led to conflicted estimations on the overall reaction barrier or the stepwise activation energy (Table 3.1). Moreover, dynamic energy contributions reveal that the rate limiting events of the acylation are the proton transfers from Lys73 to β -lactam nitrogen *via* the bridging Ser130 hydroxyl group, opposing to previous QM/MM calculations, in which the tetrahedral formation is concluded to be the rate limiting step. It should be further emphasized that the present study serves as a further complement, not criticism, to previous high-level insightful QM/MM computational studies on the mechanisms of β -lactamases driven antibiotic resistance.

In summary, I presented novel regression models with machine-learning component to quantify the energetic contributions from, as well as the correlations among, individual chemical process during enzyme catalysis with high degrees of freedom. Such quantitative measurements serve as a useful energetic-decomposing analysis to the enzymatic reaction pathway and reflect

the detailed underlying mechanism. This study also serves as a proof of the concept for extending the application of machine-learning techniques to probe complex enzymatic reaction mechanisms in high degrees of freedom configurational space.

4. *DISTINCT ACYLATION PATHWAYS OF CTX-M β -LACTAMASES WITH AMPICILLIN AND CEFALLEXIN IDENTIFIED FROM QM/MM*

4.1 CTX-M β -Lactamases

CTX-M is a representative AS β L group and has been identified as an immediate menace to commonly prescribed β -lactam antibiotics.⁸⁻¹⁰ The CTX-M enzyme class is characterized by its enhanced catalytic efficiency ($k_{\text{cat}}/K_{\text{M}}$) against cephalosporin antibiotic families.¹⁷ The hydrolysis of most cephalosporins deviates from that of other β -lactams by bearing a leaving group at its C3' position. Expelling the C3' leaving group would trigger a series of rearrangements, allowing its dihydrothiazine nitrogen to stay as an unprotonated imine after the acylation.⁶ However, an exception is cefalexin (CEX) which adopts a C3' methyl as a poor leaving group; The protonation of the CEX cephem amine is thus inevitable (Figure 4.1a). CEX also poses enhanced resistance against CTX-M hydrolysis compared to other early generations of penicillin or cephalosporins. In particular, Nitnai et al.¹²⁹ showed that the catalytic efficiency ($k_{\text{cat}}/K_{\text{M}}$) of CEX hydrolysis mediated by Toho-1 (also known as CTX-M-44) is $0.119 \mu\text{M}^{-1} \text{s}^{-1}$, which is 17-fold lower than that of ampicillin (AMP, $2.11 \mu\text{M}^{-1} \text{s}^{-1}$). Whereas AMP and CEX structurally differ only in their signature penam/cephem bicyclic rings (Figure 4.1b), the cephem scaffold of CEX evidently showed higher hydrolysis resistance even to the CTX-M enzyme class. In this study, the acylation pathways of AMP and CEX hydrolysis in Toho-1 was investigated using QM/MM chain-of-states calculations.

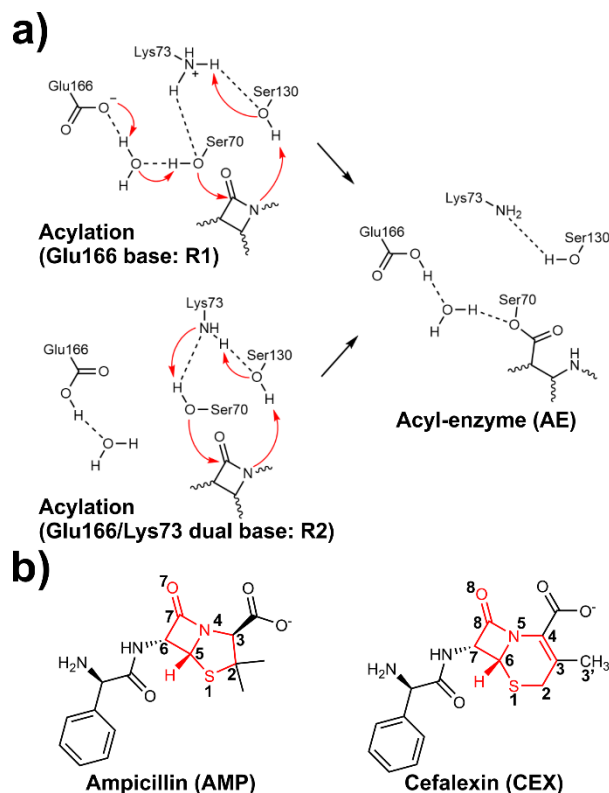


Figure 4.1 Mechanisms of acylation in ASβLs and structures of the model substrates. (a) The general mechanism of β-lactam acylation mediated by ASβL; (b) Structures of ampicillin (AMP) and cefalexin (CEX).

4.2 Computational Details

System Preparation. The high-resolution crystal structures of Toho-1/benzylpenicillin (PDB entry: 5KMW)¹³⁰ and Toho-1/cephalothin (PDB entry: 2ZQ9)¹²⁹ acyl-enzyme complexes were used as template systems to create structures for Toho-1/AMP and Toho-1/CEX complexes. The topology files of AMP and CEX were derived from CGenFF^{121,122}. The ligand topologies in the template systems were then substituted to create initial structures for Toho-1/AMP and Toho-1/CEX complexes. Systems with alternative protonation states on Lys73 and Glu166 were prepared to account for acylation pathways *via* different general base residues.¹³¹ A total of 4

enzyme-ligand models were created, protonated, optimized, and equilibrated using a semi-empirical QM/MM scheme with DFTB3 level of theory³⁸ and the 3OB parameter set³⁹ (DFTB3/3OB) as the QM potential and C36²³ as the MM counterpart. The atomic distances between the key reacting heavy atoms during a 100 ps molecular dynamic simulation using the DFTB3/3OB/C36 potential are shown in Table 4.1. The initial structures of the pathway calculations were selected as the snapshots that have the minimal inter-heavy-atom distances between the reacting functional groups of the four residues (Ser70, Lys73, Ser130, and Glu166), the catalytic water and the β -lactam.

Table 4.1 The mean atomic distances between key reacting heavy atoms in the DFTB3/3OB/C36 dynamics. Parenthesis denote the standard deviation (units: Å).

Atom pairs	Toho/AMP:R1	Toho/CEX:R1	Toho/AMP:R2	Toho/CEX:R2
Ser70 O γ – AMP C7 / CEX C8	2.43 (0.17)	2.58 (0.18)	2.44 (0.17)	2.57 (0.18)
Lys73 N ζ – Ser130 O γ	2.85 (0.15)	2.95 (0.32)	3.07 (0.25)	3.15 (0.32)
Ser130 O γ – AMP N4 / CEX N5	3.60 (0.23)	3.86 (0.26)	3.67 (0.31)	3.63 (0.31)
Ser70 O γ – Watercat O	2.65 (0.10)	2.65 (0.09)	–	–
Glu166 O ϵ 2 – Watercat O	3.06 (0.23)	2.77 (0.17)	–	–
Ser70 O γ – Lys73 N ζ	–	–	2.88 (0.13)	2.93 (0.17)

MEP calculations. A total of five structures (noted as Toho/AMP: R1, R2, and Toho/CEX: R1, R1a, R2) were chosen from the production trajectories. These five frames were then subjected to calculations at DFT level. The DFT QM region covers important active site fragments: β -lactams, the catalytic water, the surrounding residues (Ser70, Lys73, Ser130, Glu166, Asn170, Lys234, Thr235, Ser237), together with a surrounding solvent molecule for the reaction pathway calculations. The hybrid density functional B3LYP^{132,133} was used in conjunction with Pople’s 6-

31G double ζ basis set^{134–136} for the QM atoms (B3LYP/6-31G/C36). The RPM with holonomic constraints⁴⁹ was applied for all pathway optimizations and the energetic profiles on the B3LYP/6-31G/C36 optimized MEPs were further refined with the augmented 6-31++G** basis set and the D3 dispersion corrections (B3LYP-D3/6-31++G**/C36). The ChElPG scheme¹³⁷ was employed for the charge population analysis along the chain-of-states.

4.3 Summary of Results

Initial Conformations. The optimized reactant structures of Toho/AMP differ from Toho/CEX by the hydrogen bonding networks between the penam/cephem carboxylate and the residues Thr235, Ser237 (Figure 4.2). Practically, the Ser237 hydroxyl is generally outside of the H-bonding region of the AMP carboxylic group. The reactant configuration is therefore stabilized by a water molecule serving as the H-bond bridge between the Ser237 hydroxyl and the AMP carboxylate (Toho/AMP:R1, Figure 4.2a). Meanwhile, the CEX adopts a more flexible binding pattern: the hydroxyl group from Ser237 could either form direct hydrogen interacting to the substrate carboxyl group (Toho/CEX:R1, Figure 4.2b) or to a solvent water molecule (Toho/CEX:R1a, Figure 4.2c).

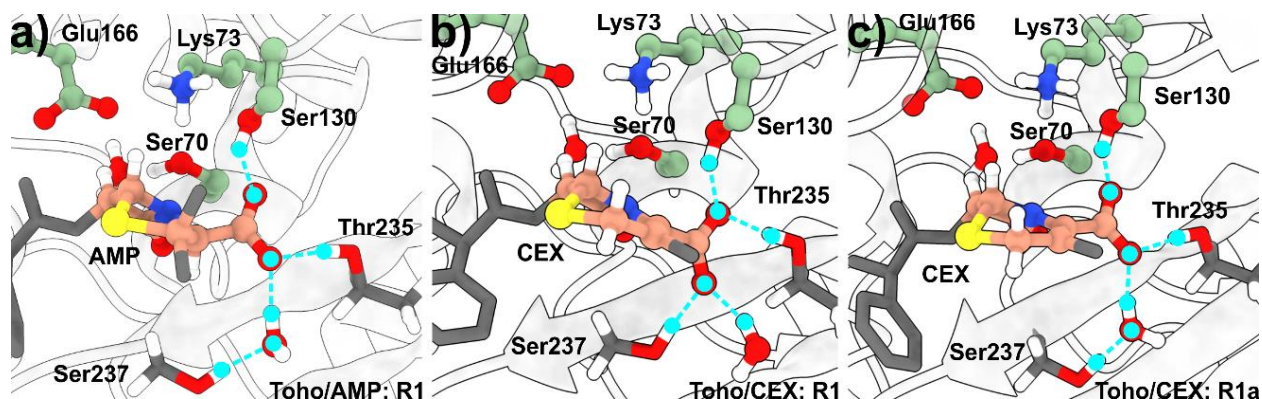


Figure 4.2 Conformations of R1 reactant states. The conformations of (a) Toho/AMP:R1; (b) Toho/CEX:R1; (c) Toho/CEX:R1a. The hydrogen bonding interactions are noted as blue dashed lines.

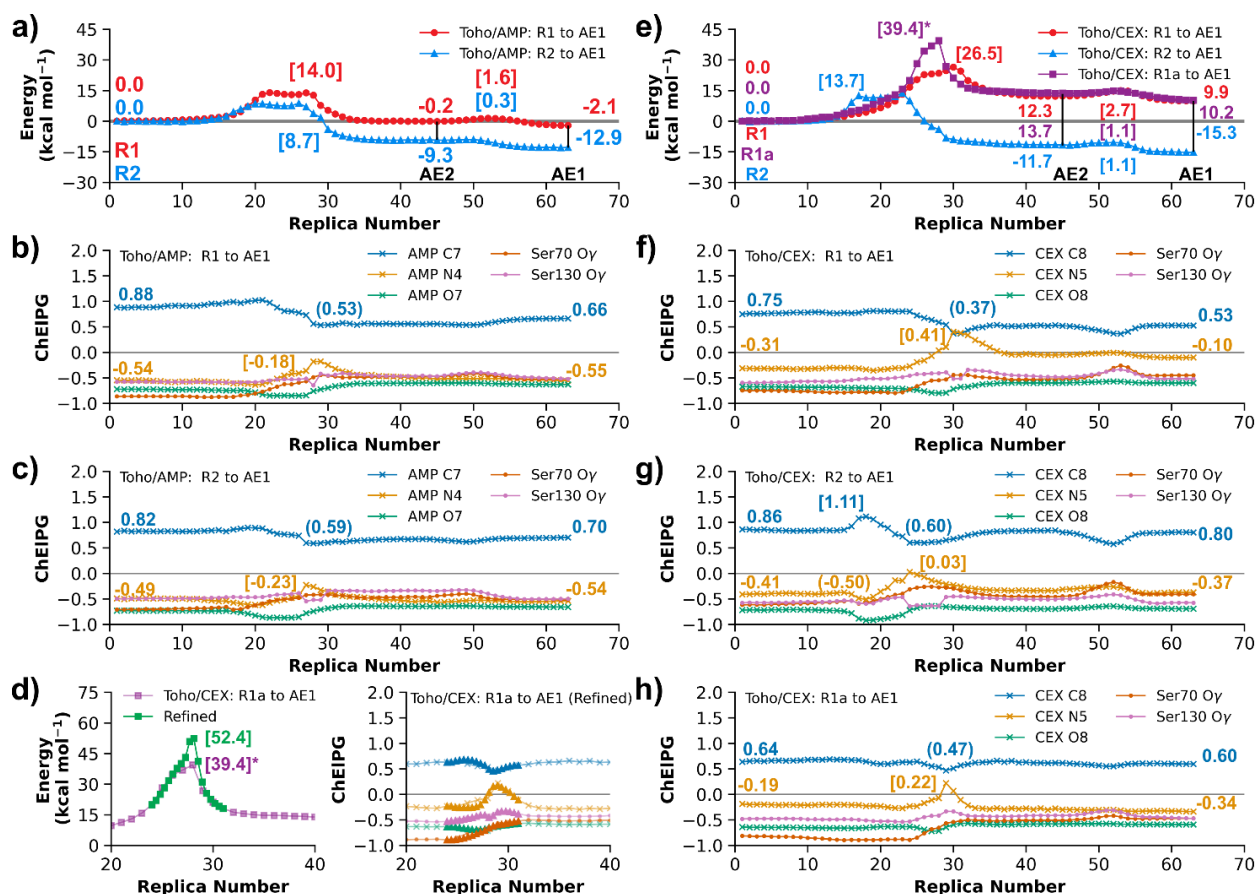


Figure 4.3 Energy profiles and the ChElPG charges of key atoms along the acylation pathways in Toho-1 hydrolysis. (a) The acylation profiles of Toho/AMP; The ChElPG charges along (b) the Toho/AMP: R1 to AE1 pathway, and (c) the Toho/AMP: R2 to AE1 pathway; (d) The energy profile and the ChElPG charge profiles of the refined Toho/CEX: R1a to AE1 pathways, which is calculated from inserting 18 replicas between replica 24 and 31; (e) The acylation profiles of Toho/CEX; The ChElPG charges along (f) the Toho/CEX: R1 to AE1 pathway, (g) the Toho/CEX: R2 to AE1 pathway, and (h) the Toho/CEX: R1a to AE1 pathway. The vertical black solid lines in (a) and (d) indicate the location of AE1 and AE2. Numbers in parentheses and brackets denote the local minimum and maximum values of important states along the reaction path. Note that only ChElPG charge values of β -lactam carbonyl carbon (blue) and nitrogen (orange) are shown in (b), (c), (f), (g), (h).

Toho/AMP Acylation Profiles. My calculated Toho/AMP acylation pathways (Figure 4.3) closely resemble the potential energy landscapes reported by Meroueh et al.,¹⁰⁶ the energy barrier for the acylation using Glu166 as general base (14.0 kcal mol⁻¹) is moderately higher than that of

Lys73/Glu166 concerted base ($8.7 \text{ kcal mol}^{-1}$). The Toho/AMP acylation pathways agree with both acylation mechanisms, indicating that either Lys73 or Glu166 could mediate the acylation process in Toho/AMP hydrolysis. The ChElPG charge profiles of the Toho/AMP pathways align with the intuitive understanding of the reaction mechanism. As shown in Figure 4.3b and 4.3c, the decreasing charge population on AMP O7 between replica 20 to 27 is synergetic to the increasing charge on Ser70 O γ , suggesting the formation of tetrahedral intermediate (with a formal charge of -1 on AMP O7) during the serine addition. Furthermore, the locations of maximal charge profiles on AMP N4 are also correlated with the replica with the highest energy along the reaction progress, showing that the protonation of AMP N4 is strongly correlated with the rate of acylation, agreeing with previous observations.

Toho/CEX Acylation Profiles. Toho/CEX acylation demonstrates a different catalytic mechanism, as shown in Figure 4.3e. The acylation barrier using Glu166 as the general base is prohibitively high ($26.5 \text{ kcal mol}^{-1}$). In particular, the corresponding barrier further increases to $52.4 \text{ kcal mol}^{-1}$ when cefalexin substrate adopts a similar binding pattern as ampicillin (Toho/CEX:R1a to AE1, Figure 4.3e). These leave Lys73 as the inevitable candidate to mediate deprotonation of the Ser70 hydroxyl during CEX acylation, which confers an energetic barrier of $13.7 \text{ kcal mol}^{-1}$ (Toho/CEX:R2 to AE1). Further mechanistic insights can be derived from the ChElPG charge profiles. On the Glu166-mediated Toho/CEX acylation pathways (Figure 4.3f, 4.3g, 4.3h), a stable tetrahedral intermediate indicated by the temporarily decreased charge on β -lactam carbonyl oxygen (as in the corresponding Toho/AMP pathways) is less synergetic to the formation of the tetrahedral intermediate. Moreover, the charge on the cephem nitrogen is largely increased to 0.41 (Figure 4.3f) and 0.22 (Figure 4.3h) upon the barrier replica, which evidently suggests its poor proton affinity to accept the proton transfer from Ser130. Alternatively, the dual-

base mediated Toho/CEX: R2 to AE1 pathways (Figure 4.3g) demonstrates a similar charge profile to the corresponding AMP acylation pathway. Interestingly, an increase of ChEIPG charge on CEX C8 is seen uniquely upon the formation of tetrahedral intermediate on this pathway (Figure 4.3g, replica 18). Intuitively, the lone pair on Ser70 O γ in the R2 configurations are oriented towards the ligand carbonyl carbon, potentially activating the conjugated π orbital on the β -lactam bicyclic. While the π -conjugation in AMP (N4-C7=O8) is localized to the β -lactam scissile C-N bond, it is extended along the cephem bicyclic (C3=C4-N5-C8=O8) in CEX. The temporary charge increment on CEX C8 can therefore be interpreted as the consequence of breaking the more delocalized π -conjugation on the cephem scissile bond during the nucleophilic attack of Ser70 O γ . Accordingly, this explanation is also supported by the observation that the tetrahedral intermediates on Toho/AMP and Toho/CEX pathways do not significantly differ from each other in terms of heavy atom conformations.

Comparison with Experiments. The computational barriers are further correlated with experimental kinetic studies (Table 4.2). Nitanaï et al.¹²⁹ reported that the catalytic barrier (calculated from k_{cat}) of Toho/AMP hydrolysis is ~ 14.9 kcal mol⁻¹, slightly lower by ~ 1.7 kcal mol⁻¹ than that of CEX (~ 16.6 kcal mol⁻¹). In my calculations, both acylation barriers for Toho/AMP are sufficiently lower than the experimentally determined catalytic barrier, suggesting that the acylation mechanism previously developed for AS β LS are applicable to Toho-1/AMP as well. In contrast, the only viable reaction pathway for CEX is the Lys73/Glu166 dual base mechanism. The pathway that uses Glu166 as the only general base greatly exceeded the experimental barrier (16.6 kcal mol⁻¹) by 9.9 kcal mol⁻¹. Briefly, my results provide good correlation to the experimental barriers.

Table 4.2 The computational and experimental catalytic barriers of AMP and CEX acylation catalyzed by Toho-1.

Source	Systems	Energy barriers (kcal mol ⁻¹)	Method [a]
Shimizu-Ibuka et al. ¹³⁸ [b]	Toho-1/AMP	15.5	303.15K, Exp
Nitanai et al. ¹²⁹	Toho-1/AMP	14.9	303.15K, Exp
This study	Toho-1/AMP	8.7 / 14.0 [c]	B3LYP-D3/C36
Nitanai et al. ¹²⁹	Toho-1/CEX	16.6	303.15K, Exp
This study	Toho-1/CEX	13.7 / 26.5 [c]	B3LYP-D3/C36

[a]The experimental (Exp) catalytic barrier of Toho/AMP were derived from k_{cat} via the Eyring equations, the acylation barrier of Toho/CEX were derived from the ratio of k_{cat}/K_M to Toho/AMP;

[b]This study used the wild-type Toho-1 as the enzyme host while others used the Arg274Asn/Arg276Asn Toho-1 mutant as the enzyme host;

[c]Values before “/” report the barrier of the Lys73/Glu166 concerted base acylation pathway. Values after “/” report the Glu166 sole base acylation pathway.

4.4 Conclusion

In this study, I demonstrate that the AMP and CEX acylation energy landscapes differ from each other during Toho-1 hydrolysis. In my calculations of both systems, the R1 pathway, which is mediated solely by Glu166 as the base, confers a higher (potential) energy barrier than the R2 pathways. Herein, the R1 acylation pathway is shown to be energetically prohibitive for CEX, leaving the Lys73/Glu166 dual base mechanism as the main viable pathway for its acylation. In the case of AMP, whereas the investigated acylation barrier *via* the Glu166 sole base mechanism is sufficiently lower than the experimentally determined kinetics, the viability of the R1 pathway is not evidently clear from the potential barrier alone. However, unlike Toho/CEX, I note that the ChElPG charge profiles in Toho/AMP acylation demonstrate a similar pattern for the R1 and R2

pathways, suggesting that the R1 acylation mechanism is at least competitive to the R2 alternatives.

My calculations with CEX acylation also shed light onto the hydrolysis of other cephalosporins. As noted above, CEX mechanistically stands out in the cephalosporin family as its β -lactam nitrogen has to be protonated upon the formation of the acyl-enzyme product. However, common cephalosporins such as cephalothin and cefotaxime show higher catalytic efficiency (k_{cat}/K_M), which suggests a much lower acylation barrier than that of CEX. Such observations suggest that the cephem nitrogen may not be protonated during the entire acylation processes of other cephalosporins. Through their crystallographic study, Olmos et al.¹³⁹ recently reported that the departure of the C3' leaving group is clearly simultaneous to the serine attack during the AS β Ls/cefotaxime acylation, supporting the above hypothesis. In this regard, the protonation of the cephem nitrogen, which was also previously validated as the rate limiting step, could be avoided, and leading to the higher acylation rates observed in other early generations of cephalosporins.

5. ACYLATION OF CTX-M-44 AS β L EXPLAINED FROM MACHINE-LEARNED QM/MM MINIMUM ENERGY PATHWAYS

5.1 Boltzmann-weighted Cumulative Integrated Gradients

In order to verify the intuitive hypothesis on the acylation of Toho/AMP and Toho/CEX, I developed the XAI method, coined Boltzmann-weighted Cumulative IG (BCIG), for ML models trained on high quality DFT/MM MEPs, which is formulated as follows. 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}

$$Contribution_c^p = A_{path}(F, c, p) \quad (5.1)$$

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

$$Contribution_c^P = C_N \sum_{p=1}^P \left(\exp\left(-\frac{\Delta E^p}{RT}\right) \times A_{path}(F, c, p) \right) \quad (5.2)$$

where ΔE^p is the energy barrier of the p -th MEP, R is the ideal gas constant, T is the temperature, 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, direct application of Boltzmann weights would lead to numerical instability with a limited number of sampled MEPs.

Alternatively, a Probability Density Function (PDF), $PDF(\Delta E^p)$, could be introduced to smooth the density of MEP barriers. In its simplest form, the PDF could be a single Gaussian function. Further, 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 equation 5.2 yields

$$Contribution_c^P = C'_N \sum_{p=1}^P \left(\exp\left(-\frac{\Delta E^p}{RT}\right) \times PDF(\Delta E^p) \times A_{path}(F, c, p) \right) \quad (5.3)$$

Note that the normalization factor C'_N is

$$C'_N = \frac{P}{\sum_{p=1}^P \exp\left(-\frac{\Delta E^p}{RT}\right) \times PDF(\Delta E^p)} \quad (5.4)$$

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

If the chain-of-states RPM 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:

$$A_{path}(F, c, p) = \sum_{r=1}^M A_{replica}(F, c, p, r) \quad (5.5)$$

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

As proposed by Sundararajan et al.⁹⁸, 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 $\alpha = 0$)

$$IntegratedGrad_i^{\gamma}(\mathbf{x}) = \int_{\alpha=0}^1 \frac{\partial F(\gamma(\alpha))}{\partial \gamma(\alpha)} \frac{\partial \gamma(\alpha)}{\partial \alpha} d\alpha \quad (5.6)$$

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

$$A_{replica}(F, \mathbf{c}, p, r) = IntegratedGrad_{\mathbf{c}}^p(\mathbf{x}) = \sum_{i=1}^r \frac{\partial F(\mathbf{x}(p, i))}{\partial \mathbf{x}(p, i)} \frac{\partial \mathbf{x}(p, i)}{\partial i} \quad (5.7)$$

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

The representation of $\mathbf{c}$ must be determined to expand the first partial derivative in equation 5.7. As noted, $\mathbf{c}$ represents a ‘chemical process’ that includes (but is not limited to) bond making/breaking and proton transfers. The progress of 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, I note that the correlation between the atomic distances is highly nonlinear in the evolution along the optimal reaction path obtained from the chain-of-states calculations. Therefore, instead of feeding the reduced representation of

linear-combined atomic distances, I used a set of atomic distances that accounts for all chemical process during the acylation for the training of the DL model F , that is,

$$E(p, i) \cong F(\mathbf{x}(p, i)) \quad (5.8)$$

where $E(p, i)$ is the energy of i -th replica on the p -th MEP. Further, due to the nonlinearity of the correlation between the feature dimensions, the gradients of F with regard to $\mathbf{c}$, which usually corresponds to multiple feature dimensions in $\mathbf{x}$, cannot be calculated analytically. Therefore, the first partial derivative in equation 5.7 is computed numerically

$$\frac{\partial F(\mathbf{x}(p, i))}{\partial \delta(\mathbf{x}(p, i))} = \frac{F(\mathbf{x}(p, i) + \delta \varepsilon(\mathbf{x}(p, i+1) - \mathbf{x}(p, i))) - F(\mathbf{x}(p, i) - \delta \varepsilon(\mathbf{x}(p, i) - \mathbf{x}(p, i-1)))}{\text{Dist}(\mathbf{x}(p, i) + \delta \varepsilon(\mathbf{x}(p, i+1) - \mathbf{x}(p, i)), \mathbf{x}(p, i) - \delta \varepsilon(\mathbf{x}(p, i) - \mathbf{x}(p, i-1)))} \quad (5.9)$$

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 ($\mathbf{c}$). In practice, δ is a multi-hot encoded mask to ensure that the perturbation ε is applied only to the features $\mathbf{x}$ that represent the chemical process of interests; $\text{Dist}(A, B)$ stands for the distance metric that accounts for the pathway curvature. The gradient on the pathway curvature reads:

$$\frac{\partial \mathbf{x}(p, i)}{\partial i} = \frac{\text{Dist}(\mathbf{x}(p, i) + \delta \varepsilon(\mathbf{x}(p, i+1) - \mathbf{x}(p, i)), \mathbf{x}(p, i) - \delta \varepsilon(\mathbf{x}(p, i) - \mathbf{x}(p, i-1)))}{\varepsilon \times ((i+1) - (i-1))} \quad (5.10)$$

Combining equations. 5.7, 5.9, 5.10, the contribution of the chemical process $\mathbf{c}$ to the r -th replica on the p -th pathway is calculated as the integrated partial gradients of the ML-MEPs model F as:

$$A_{\text{replica}}(F, \mathbf{c}, p, r) = \sum_{i=1}^r \frac{F(\mathbf{x}(p, i) + \delta \varepsilon(\mathbf{x}(p, i+1) - \mathbf{x}(p, i))) - F(\mathbf{x}(p, i) - \delta \varepsilon(\mathbf{x}(p, i) - \mathbf{x}(p, i-1)))}{2\varepsilon} \quad (5.11)$$

Accordingly, the contribution of c along one MEP $A_{path}(F, c, p)$ can be calculated by cumulatively summing the integrated partial gradients (equation 5.5). The sign of $A_{path}(F, c, p)$ 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 $A_{path}(F, c, p)$ were used to calculate the weighted contributions in equation 5.3. I proceed with the computational details that implements the BCIG approach for interpreting the acylation reactions of Toho/AMP and Toho/CEX studied in the last chapter.

5.2 Computational Details

System Preparation. Based on reaction mechanisms of the target reaction studied in the last chapter, I created the R1, R2, and AE states for wild type Toho/AMP and Toho/CEX systems using the DFTB3/3OB/C36 potential with necessary distance-based restraints. A total of six states were subjected to extensive conformational sampling with constrained NVT simulations for 150 ns. During the constrained dynamics, 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. 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.

MEP calculations. 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 system were fixed. The corresponding product/reactant states were generated from the starting conformations. 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 constraints⁴⁹.

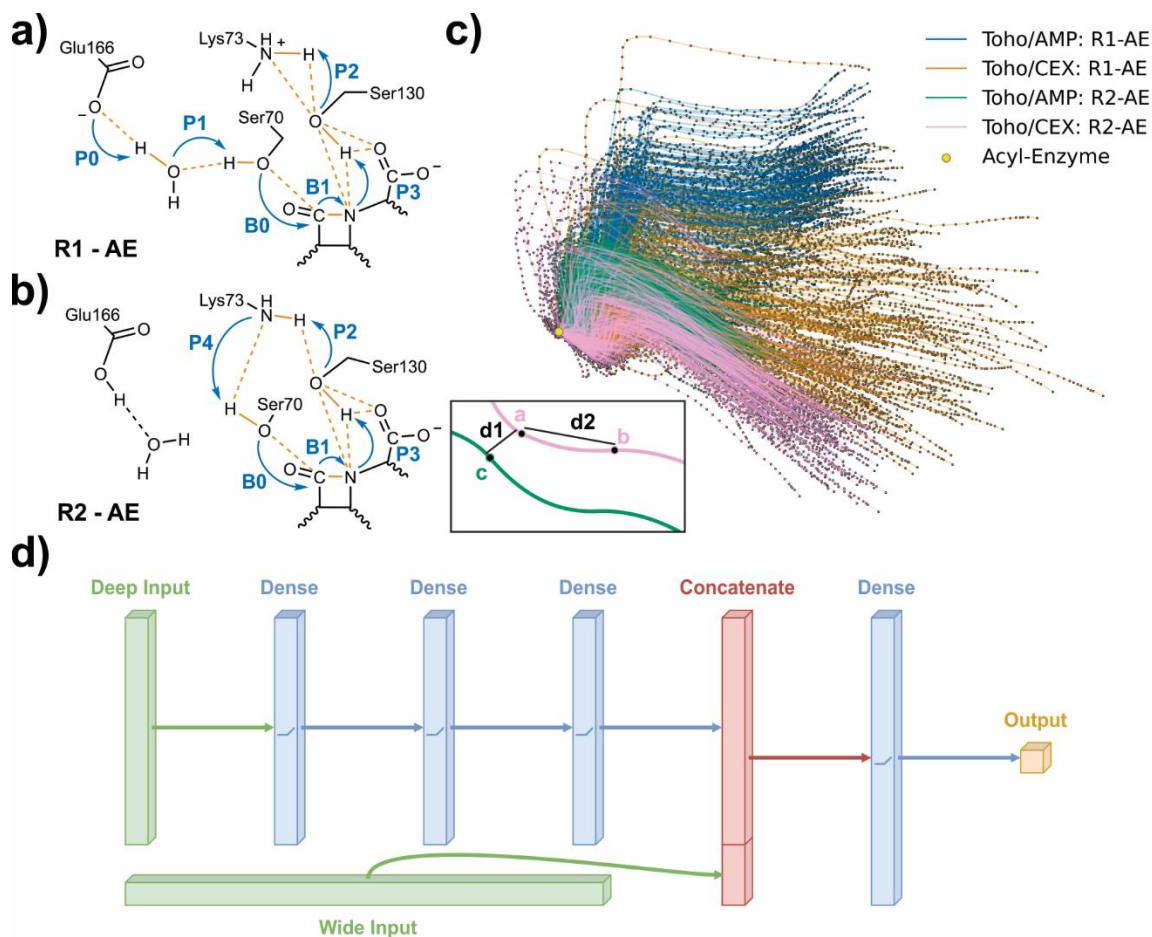


Figure 5.1 Selected features, 2D representation of the pathway conformations, and architecture of the ML-MEP models. The selected features and chemical processes of (a) the Toho/AMP: R1-AE and Toho/CEX: R1-AE datasets and (b) the Toho/AMP: R2-AE and Toho/CEX: R2-AE datasets. The atomic distances that are included in feature vectors are noted in orange lines, the chemical processes are noted in blue; (c) The 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) The architecture of the QM/MM MEP learning deep-and-wide neural network.

ML-MEP models. 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. The final selection of features on the R1-AE and R2-AE pathway datasets were illustrated in Figure 5.1a, 5.1b.

A feed-forward neural network with the ‘Deep-and-Wide’ learning architecture (DaWNN) proposed by Cheng et al.¹⁴⁰ was implemented for the learning of the QM/MM MEPs (Figure 5.1d). The dropout strategy¹⁴¹ was applied for all hidden layers to promote the generalizability of the neural networks and prevent over-fitting. 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% path-wise stratified validation set. Practically, I constructed the validation set by randomly picking 5 replicas from each of the pathway carrying 50 replicas. The standard mean squared error (MSE) was used as the objective loss function to train the ML-MEP models. All models were trained with the AdaM¹⁴² optimizer for 300 epochs with a sample batch-size of 25.

5.3 Summary of Results

QM/MM MEPs. The barrier distributions of the calculated QM/MM MEPs are plotted in Figure 5.2. 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 exponential averaged acylation barrier is 16.98 kcal mol⁻¹ for the Toho/AMP: R1-AE pathway (Figure 5.2a), which is 12.81 kcal mol⁻¹ higher than the Toho/AMP: R2-AE pathways (4.17 kcal mol⁻¹, Figure 5.2c). 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 of 14.33 kcal mol⁻¹ (Figure 5.2d), which is 11.22 kcal mol⁻¹ lower than its R1-AE alternative (25.55 kcal mol⁻¹, Figure 5.2b). Further, 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 as generally inaccessible. The viable acylation path for Toho/CEX is therefore verified to be the R2-AE mechanism, as stated in the previous chapter.

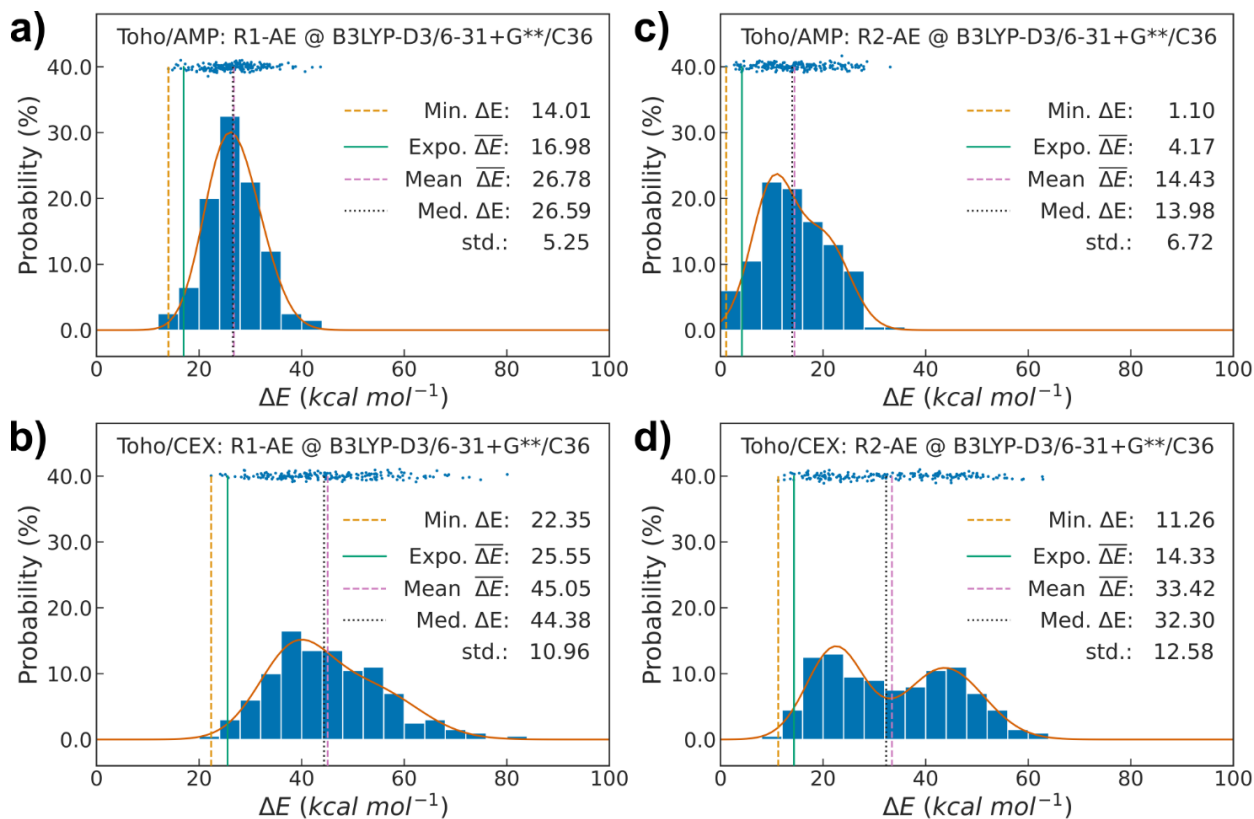


Figure 5.2 The distribution of the acylation barriers (ΔE) at B3LYP-D3/6-31+G**/C36 level of theory. (a) Toho/AMP: R1-AE acylation pathways; (b) Toho/CEX: R1-AE acylation pathways; (c) Toho/AMP: R2-AE acylation pathways, and; (d) Toho/CEX: R2-AE acylation pathways. The scatters show the locations of the energy barriers. The width of the histograms is 4 kcal mol⁻¹. The red curves note the density estimation from the GMMs.

Energetics interpreted from BCIG contributions. As shown in Figure 5.3a and 5.3b, the DaWNN models could accurately predict the replica-wise energy with R^2 scores > 0.995 and the barrier heights with $RMSE < 2$ kcal mol $^{-1}$.

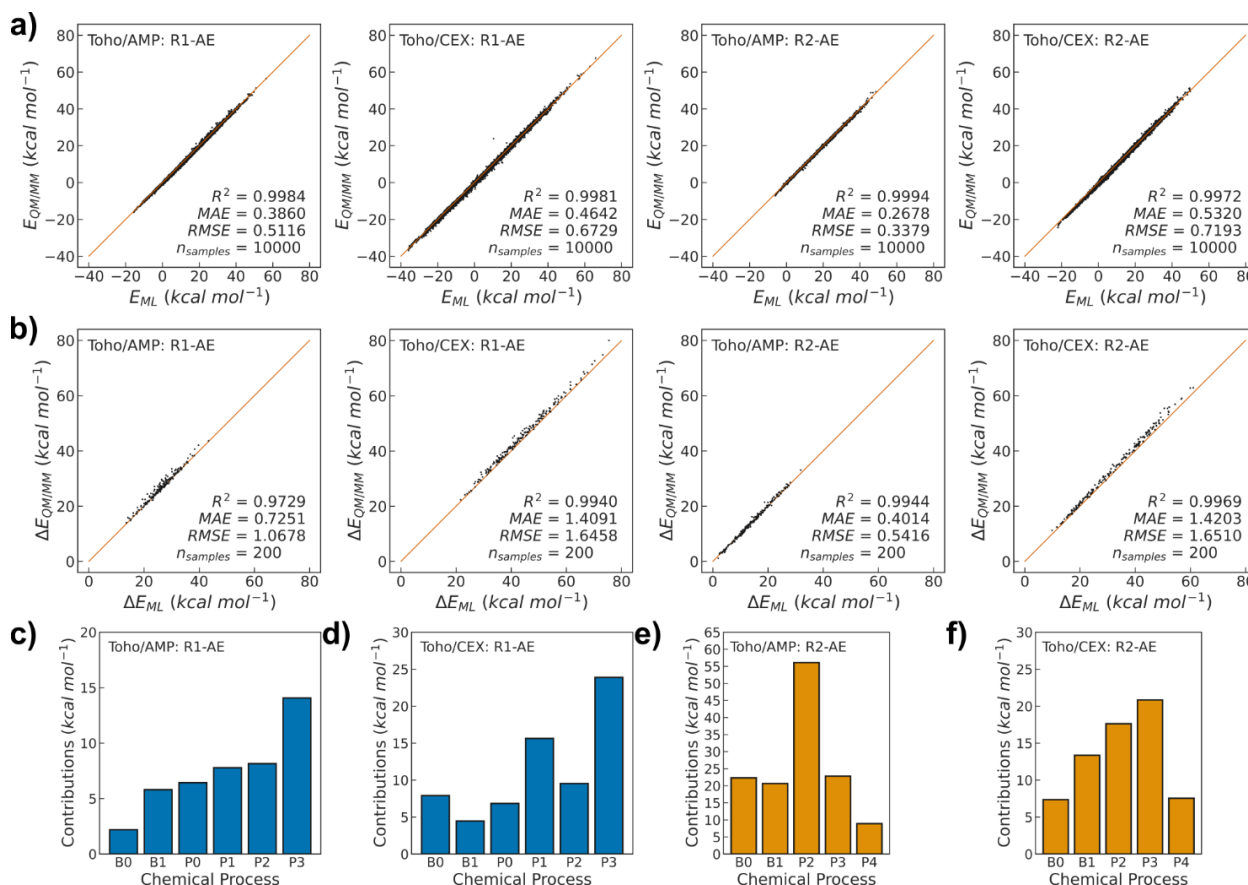


Figure 5.3 The predictive performance and the BCIG contributions of the ML-MEP models. The predictive performance of (a) the replica energies and (b) the pathway barriers of (left to right) the Toho/AMP: R1-AE, Toho/CEX: R1-AE, Toho/AMP: R2-AE, and Toho/CEX: R2-AE models. The BCIG contributions of the models: (c) Toho/AMP: R1-AE; (d) Toho/CEX: R1-AE; (e) Toho/AMP: R2-AE; and (f) Toho/CEX: R2-AE.

For the Toho/AMP: R1-AE pathways (Figure 5.3c), 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) were 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 suggest that the protonation of the cephem nitrogen (P2 and P3) is the determinant factor with the highest contribution (Figure 5.3d). However, the deprotonation of Ser70 O γ (P1) and its nucleophilic attack 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 5.3e, 5.3f). 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). I note that the BCIG derived from different systems are not to be compared with each other since they explain different models which were trained on different datasets with distinct distributions. Alternatively, one can train models on mixed datasets to allow comparison of BCIG from different systems to explain for relative reactivity. However, the BCIG values in this case for different chemical process are not to be compared with each other, since the distribution of different feature is biased by the artificial mixing of the datasets.

Reactivity interpreted from BCIG contributions. Additional DaWNN models trained on the mixed datasets (Toho/AMP&CEX: R1-AE and Toho/AMP&CEX: R2-AE) As shown in Figure 5.4a, 5.4b, the model prediction on the replica energies or the pathway barriers remain the same predictive accuracies (RMSEs < 2 kcal mol⁻¹ and R² > 0.995) compared with the models trained on individual datasets, demonstrating the scalability of the DaWNN architecture for learning MEPs.

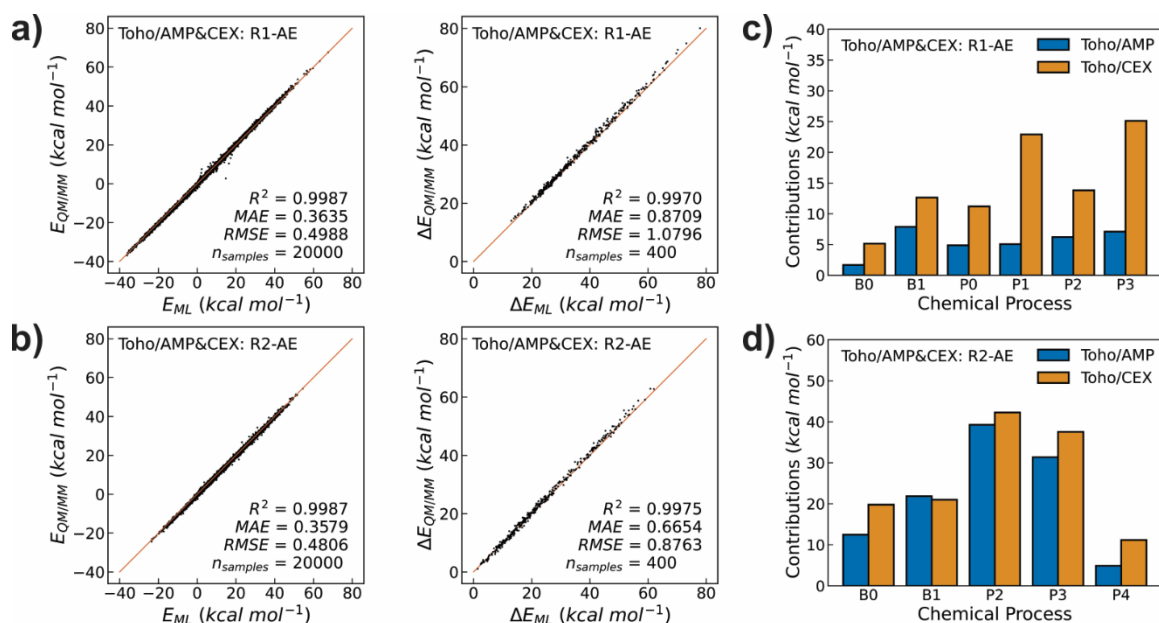


Figure 5.4 The predictive performance and the BCIG contributions of the unified ML-MEP models. The predictive performance of (left to right) the replica energies and the pathway barriers of (a) the Toho/AMP&CEX: R1-AE; and (b) the Toho/AMP&CEX: R2-AE models. The BCIG contributions of (c) the Toho/AMP&CEX: R1-AE; and (d) the Toho/AMP&CEX: R2-AE models.

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 on all chemical processes are much higher in the Toho/CEX acylation MEPs than the Toho/AMP ones (Figure 5.4c). 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. While the acylation pathways are initialized by the nucleophilic serine addition, the high BCIG contributions attributed for this process indicates that the R1-AE acylation pathways for Toho/CEX is unfavored comparing 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 5.4d): the BCIG values for most chemical processes in Toho/CEX pathways are higher than the Toho/AMP pathways. The differences of the BCIG contributions in two systems mainly come from the residue-evolved 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 CEX concerted base acylation is shown to be slightly lower than that for 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.

5.4 Conclusion

In this study, I presented a QM/MM computational workflow that achieves fast sampling of QM/MM MEPs for enzyme catalysis. Firstly, I 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 demonstrated in chapter 4.

ML-MEP models with high performance and scalability using the DaWNN architecture were developed for the machine-learning of the QM/MM MEPs of enzyme catalysis. Comparing 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. Further, 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 approach⁹⁸ for explaining ML/DL models, I further developed the BCIG approach to interpret the ML-MEP models for mechanistic insights in 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, I 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, I have shown that the interpretations of different models give consistent mechanistic insights that agrees with my intuitive mechanistic understandings on the modeled systems.

6. MECHANISTIC ANALYSIS OF CARBAPENEMASES RESISTANCE TO IMIPENEM GUIDED BY GRAPH-LEARNING

6.1 GES-5 β -Lactamases

The GES family of AS β LS has demonstrated functional diversity that constitutes its broad resistance to β -lactams.^{5,12,13,18} 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 6.1a).¹⁴³ The general β -lactam deacylation mechanism has been extensively investigated by pioneering theoretical efforts.^{144–146} Based on their hybrid QM/MM mechanistic study, Hermann et al.¹¹⁰ proposed that the AS β LS deacylation of β -lactams is a concerted one step process (Figure 6.1b). 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 O γ and the β -lactam carbonyl carbon is cleaved upon the proton transfer from the fully protonated Lys73 amino to Ser70 O γ , which retrieves the Ser70 hydroxyl and completes the deacylation. Moreover, the deacylation of carbapenems is further complicated by possible Δ 2-to- Δ 1 pyrroline tautomerization on the carbapenem scaffold during the acyl-enzyme state (Figure 6.1b).^{147,148} While the interconversion between the two tautomer states is reported, it has been proposed that the Δ 1-pyrroline tautomer state is more inert to AS β L deacylation.^{149,150}

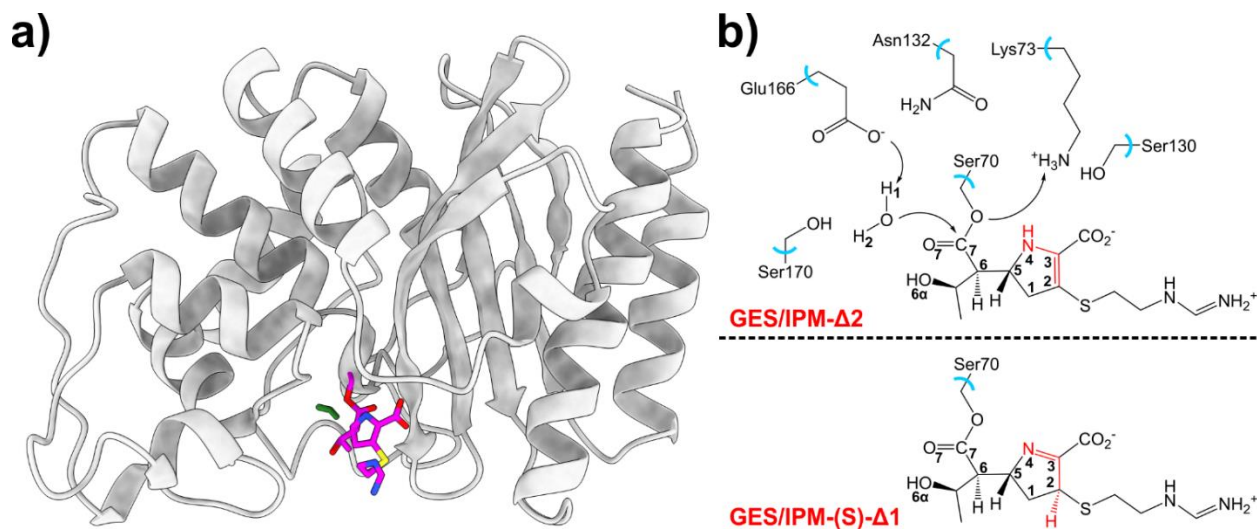


Figure 6.1 The GES/IPM complex and the deacylation reaction. (a) The GES/IPM acyl-enzyme complex. The deacylating water and the IPM molecule are colored in green and magenta, respectively; (b) The deacylation mechanism and the atoms included in the QM region. The QM/MM boundary bonds are marked by blue lines. The structural differences of the acylated IPM- $\Delta 1$ and IPM- $\Delta 2$ tautomer are highlighted in red.

6.2 Computational Details

System Preparation I built the simulation system from the crystal structure of GES-5/IPM acyl-enzyme complex (PDB: 4H8R).¹⁵¹ The sulfhydryl groups in Cys69 and Cys238 were patched as the conserved disulfide bridge in most AS β L-carbapenemases.¹⁵² The prepared systems were subjected to extensive MD sampling after proper equilibration dynamics. The reacting groups, including the Ser70 O γ , the Lys73 N ζ amino, the Glu166 O ϵ 1, the deacylating water, and the IPM β -lactam/pyrroline bicyclic scaffold, 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 representative configurations.

QM/MM MEPS. The protonated N4 on the IPM pyrroline ring undergoes tautomerization in most AS β L/IPM acyl-enzyme intermediates, leading to two possible tautomer configurations on the IPM pyrroline ring: IPM- Δ 1 and IPM- Δ 2 (Figure 6.1b). In their crystallography study, Smith et al. observed that the *S* isomeric state of the Δ 1 tautomer is the dominant configuration in the acyl-enzyme complex in a GES-5 Cys69Gly variant. Accordingly, the *S* isomeric form was selected for all Δ 1 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-5.¹⁵² For each representative configuration, both IPM tautomerization states were built and were used to calculate the deacylation MEPS. A total number of 1,000 pathways using the RPM with holonomic constraints⁴⁹ and 36,000 single point B3LYP-D3/MM energies were calculated in the current study.

Featurization. The goal of my GL task is to connect the acyl-enzyme configurations to the deacylation barriers of the QM/MM MEPS. I represent the atoms and interatomic distances as vertices and edges, formulating graph representations of the acyl-enzyme configurations for GES/IPM deacylation (Figure 6.2a). The atoms included in the graph was selected according to the following procedure. First, the atoms on the reaction coordinates were automatically included as initial vertices. I then expanded the selection to any neighboring atoms that are covalently-bonded or forming hydrogen-bonding interactions with the initially atoms in at least one of the reactant configurations, leading to 19 and 21 atoms/vertices in the GES/IPM- Δ 1 and GES/IPM- Δ 2 graphs, respectively. Each vertex was represented as a vector that one-hot encodes the element type multiplied by its partial charge from Natural Population Analysis (NPA)¹⁵³ 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.

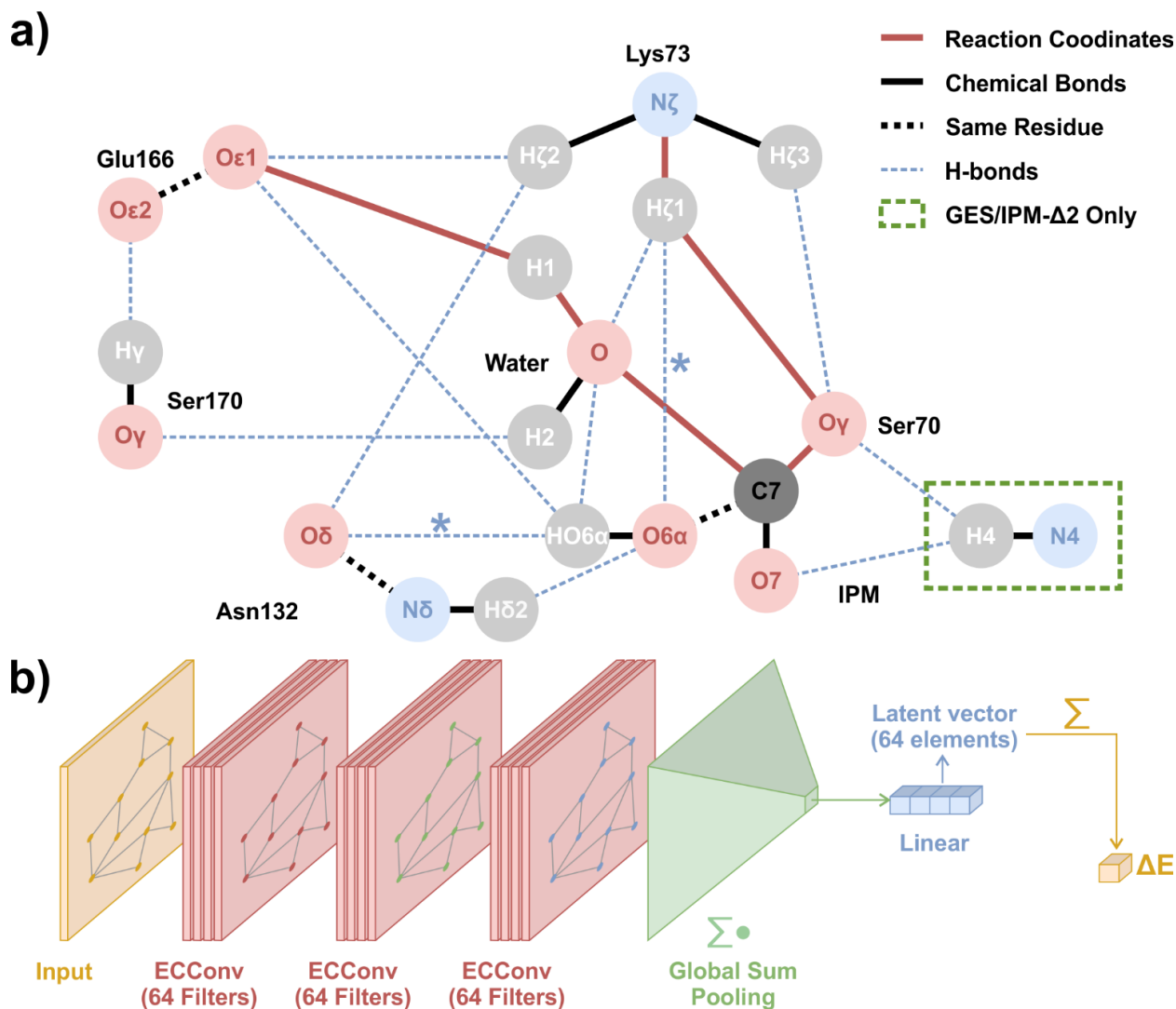


Figure 6.2 The graph representations and the learning model. (a) The graph representation of GES/IPM acyl-enzyme conformations. The hydrogen, carbon, nitrogen, and oxygen atoms are noted by gray, black, blue, and red spheres, respectively. The edges as reaction coordinates, chemical bonds, hydrogen bonding 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- $\Delta 2$ graphs. (b) The ECGCNN model architecture.

Conformational clustering. Prior to the GL of the deacylation barriers, my visual inspection on the acyl-enzyme states identified several conformational patterns. Therefore, I performed

conformational clustering analysis to elaborate the conformational modes of the non-reacting functional groups upon deacylation. I note that my attempts with unsupervised clustering algorithms (also assisted by dimensionality reduction) provides no meaningful clustering as verified from my visual inspection, which suggests the high-complexity of the conformational space on the GES/IPM active site with the reduced edge representations. Alternatively, I manually filtered the metric distances based on two criteria: the metric distance (1) should manifest high variance ($> 0.5 \text{ \AA}$) to clearly distinguish conformational differences; and (2) should not lead to small clusters with fewer than 10 snapshots. Eventually, the distances of Asn132 O δ – IPM HO6 α and IPM O6 α – Lys73 H ζ 1 were used to successfully divide the snapshots into four clusters (Figure 6.3), noted as cluster A, B, C, and D.

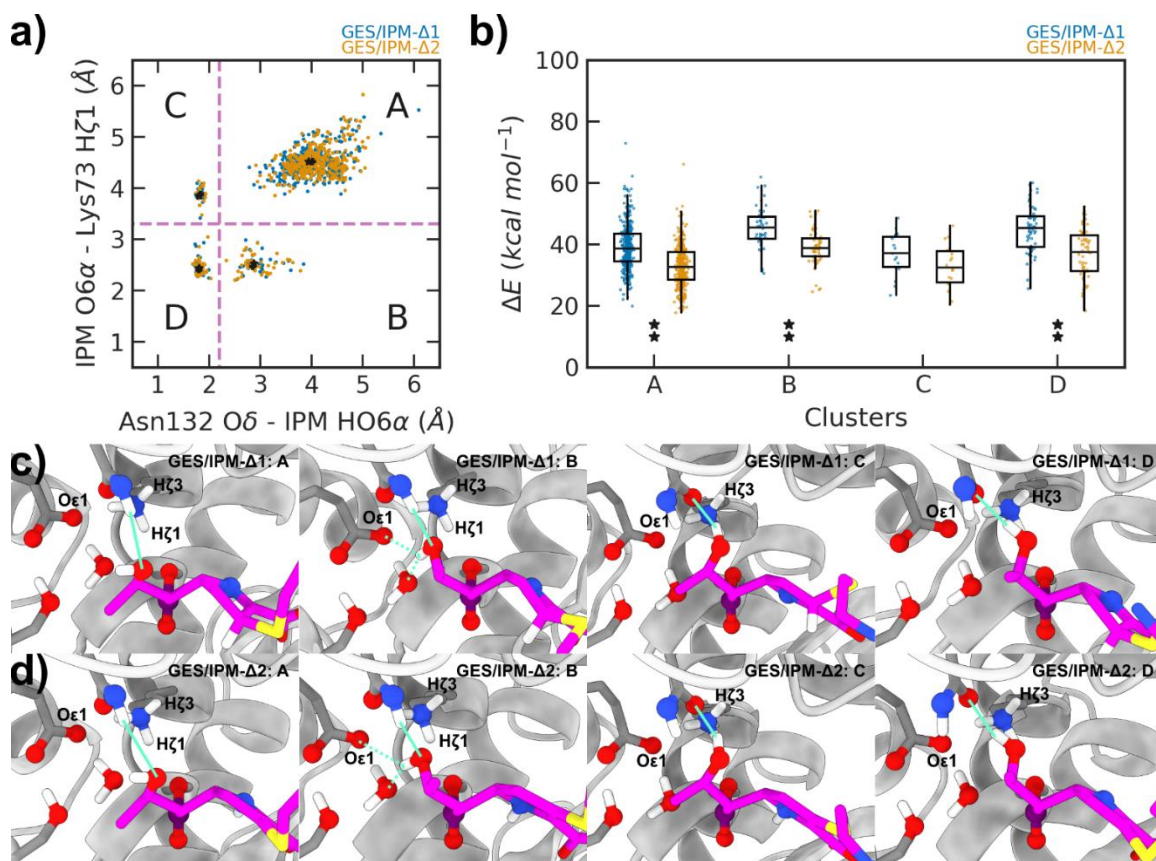


Figure 6.3 The conformational clustering of the acyl-enzyme states, the energy barriers of the clusters, and the representative conformations in each cluster. (a) The clustering of acyl-enzyme

configurations using two distance metrics (pink dashed lines); (b) The deacylation barriers in each cluster. ** denotes that the mean difference in barrier distributions between two tautomer states are statistically significant (Welch test $p < 0.001$); The representative acyl-enzyme configurations in clusters of (c) GES/IPM- $\Delta 1$; and (d) GES/IPM- $\Delta 2$. The key hydrogen-bonding interactions involving the IPM 6 α -hydroxyl group and Asn132 are highlighted by cyan solid lines. The Glu166 O ϵ 1, and the Lys73 H ζ 1, H ζ 3 are labeled. The carbon atoms of the β -lactam ligand are 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.

GL Model for Predicting Deacylation Barriers. The ECGCNN model⁹¹ implemented in the current study adopts three edge-conditioned learning layers, each of which learns the vertex hidden representations with 64 convolutional filters (Figure 6.2b). 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.

I trained the ECGCNN model using all 1,000 GES/IPM configurations. 150 configurations were randomly selected as a 15% validation set stratified with regard to the IPM pyrroline tautomer states. The learnable weights were optimized by an AdaM optimizer¹⁴² at a learning rate of 0.001 against the standard loss of mean-squared-error (MSE). 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.

Perturbative Response of the GL Model. The NPA charges are mostly invariant with regard to the GES/IPM acyl-enzyme configurations (Figure 6.4), I therefore focus on the edge features for

the ECGCNN model explanations to gain mechanistic insights. I assess the edge importance by the following perturbative approach. For each graph denoting an acyl-enzyme configuration, I first obtain the 64-dimensional latent vector produced from the final linear transform, which is defined as the baseline, h^{base} . I then feed a perturbed graph omitting an edge e that flows through the forward pass of the ECGCNN model and produces 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, I assess the importance of e with the 1-norm distance between h^{base} and h_e^{pert} :

$$Importance(e) = \sum |h_e^{pert} - h^{base}| \quad (6.1)$$

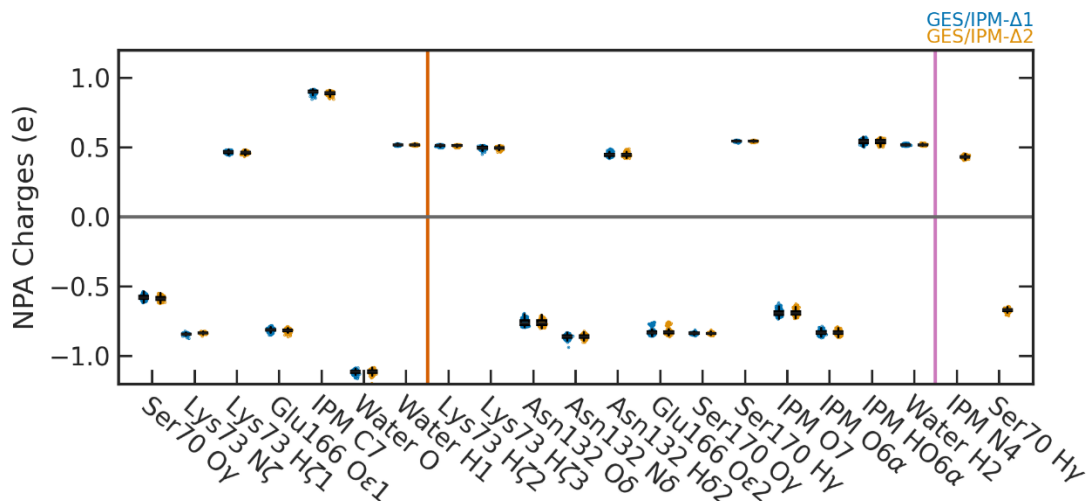


Figure 6.4 Distribution of NPA charges on each atom included in the graph. Entries on the left side of the solid orange line are vertices (atoms) related to the reaction coordinates. Entries on the right side of the solid purple line are vertices (atoms) present only in GES/IPM- Δ 2 graphs.

6.3 Summary of Results

GES/IPM Deacylation Barriers. It has been proposed that the deacylation half of AS β L-mediated IPM hydrolysis is the rate-limiting step of the overall hydrolysis. 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 $17.62\text{ kcal mol}^{-1}$ with the Eyring equation.¹⁵⁴ Moreover, Kalp et al suggested that the deacylation efficiencies of IPM in AS β Ls are also correlated with the tautomerization states of the pyrroline ring, where the IPM- $\Delta 2$ tautomer is more active for deacylation than the IPM- $\Delta 1$ species. In my calculations (Figure 6.5), the exponential-averaged deacylation barriers of the IPM- $\Delta 1$ tautomers in GES-5 ($23.83\text{ kcal mol}^{-1}$) is higher than that of IPM- $\Delta 2$ ($21.17\text{ kcal mol}^{-1}$). The IPM- $\Delta 2$ tautomer is therefore deemed to be the more active deacylating species.

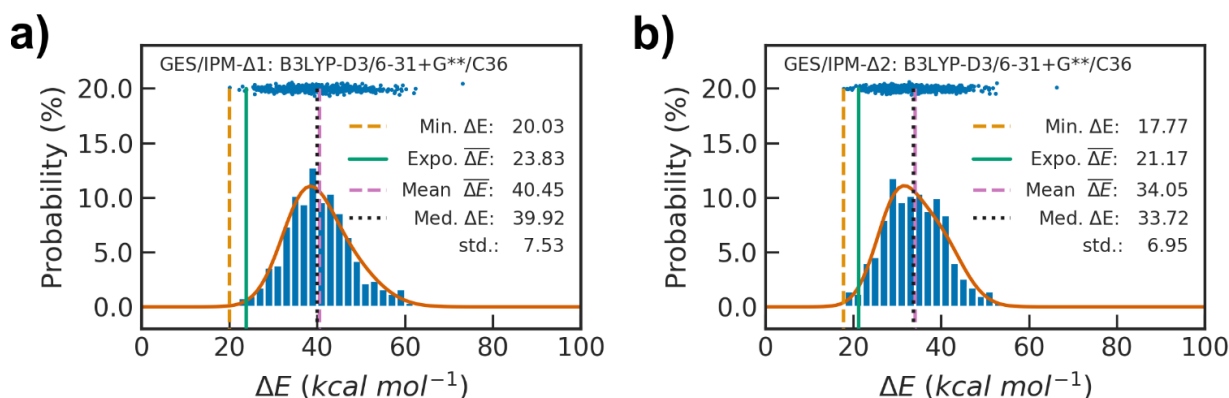


Figure 6.5 The deacylation barrier distributions of GES/IPM pathways. (a) GES/IPM- $\Delta 1$; (b) GES/IPM- $\Delta 2$.

GL Representations. The predictive performance of the ECGCNN model on the deacylation barrier is firstly evaluated (Figure 6.6a). The deacylation barriers of the training set can be predicted within 2.0 kcal mol^{-1} mean absolute errors (MAE). The predictions on the validation set

are with lower accuracy ($\text{MAE} < 3.0 \text{ kcal mol}^{-1}$), but the predicted barrier heights are in good agreement with the calculated values ($R^2 > 0.8$). Therefore, I conclude that my ECGCNN model could reasonably predict the deacylation barriers based on the graph representations of the acyl-enzyme configuration.

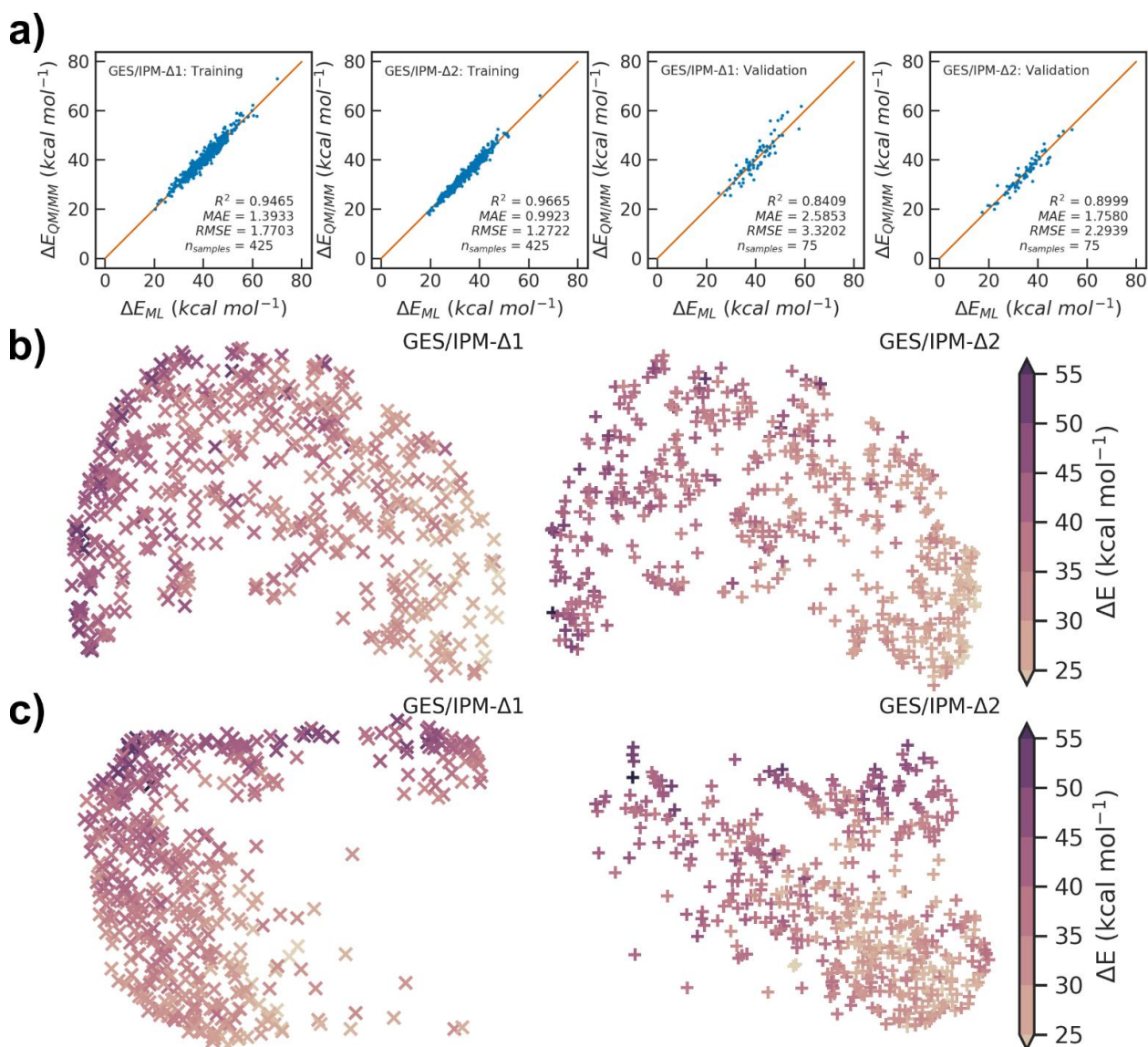


Figure 6.6 The regression performance of the ECGCNN model and the UMAP dimensionality reduced visualization of graph features. (a) The regression performance of the GNN model on the training and validation sets of two systems. The UMAP dimensionality reductions on the (b) distances as reaction coordinates; (c) the 64-element latent vectors from the ECGCNN model. Note that the UMAP representations of GES/IPM-Δ1 and GES/IPM-Δ2 were plotted under the same scale.

I further investigated if the knowledge basis of the structural factors on the barrier heights are also properly encoded by the GL latent representations. The low-dimension UMAP¹⁵⁵ representation of the reaction coordinate distances showed clear correlation to the deacylation barrier (Figure 6.6b): 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 6.6c). Meanwhile, it also shows that the IPM- $\Delta 1$ and $\Delta 2$ 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.

Perturbative Responses of the GL Representations. The ECGCNN model was unboxed using the perturbative response of the graph-learned hidden representations. The overall 1-norm displacements of the 64-dimensional latent vector upon edge exclusion are shown in Figure 6.7. The reaction coordinates of the nucleophilic attack of the deacylating water (IPM C7 – Water O) and the protonation of the GLU166 (Glu166 O ϵ 1 – Water H1) showed statistically significant difference between the IPM tautomerization states (Figure 6.7a). Compared with the IPM- $\Delta 1$ tautomers, the deacylation barriers of the $\Delta 2$ states are less sensitive to the water attack on the Ser70-IPM ester carbon and are more dependent on the protonating-distance of Glu166. Notably, in both deacylation pathways, the deacylation barriers are shown to be most sensitive to the exclusion of the edge IPM C7 – IPM O6 α , which specifies the orientation of the 6 α -hydroxyethyl. Lastly, while the Welch tests have shown that the latent vector displacements upon excluding three

edges (Asn132 O δ 1 – IPM HO6 α , Glu166 O ϵ 1 – IPM HO6 α , and IPM O6 α – Lys73 H ζ 1) are statistically different between the two tautomer states, those differences in values seem small.

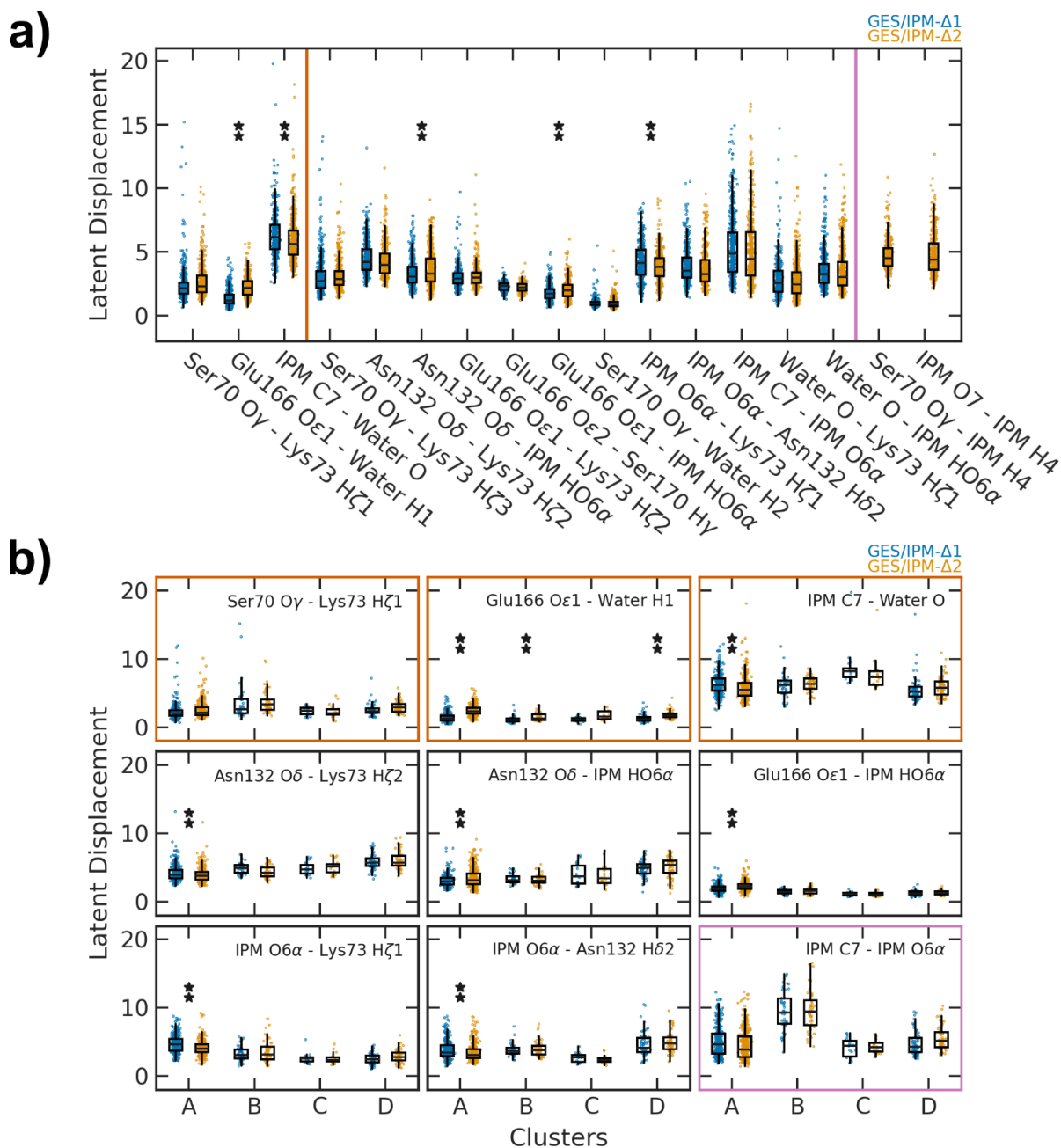


Figure 6.7 The displacements of latent vectors upon edge removal from the reactant graph. (a) The displacement of latent vectors arranged by edges excluded in each system; (b) The latent vectors displacements that show very significant difference in at least one of the clusters. The sub-plots with orange and black axis are inter-residue reaction coordinates and inter-residue hydrogen

bonding interactions, respectively. ** denotes that the mean difference in barrier distributions between two tautomer states are statistically significant (Welch test $p < 0.001$).

Due to the diverse orientations of the IPM 6 α -hydroxyethyl, I further investigated the potential impact to the deacylation barriers from each conformational cluster. The deacylation barriers of IPM- $\Delta 2$ are more sensitive than those of $\Delta 1$ to the reaction coordinates of Glu166 protonation (Glu166 O ϵ 1 – water H1) in clusters A, B, and D, while cluster C is inconclusive yet with a higher average perturbative response (Figure 6.7b). As for the interactions concerning the IPM 6 α side chain groups, all statistically significant comparisons between two tautomer states are found for the A cluster, while the actual differences are shown to be small. The key difference between the conformational clusters is the IPM O6 α – IPM C7 edges. As shown in Figure 6.7b, while all clusters showed high latent response upon the IPM O6 α – 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 6 α -hydroxyl group donating hydrogen bonds to the water molecule or the general base (Glu166, Figure 6.8).

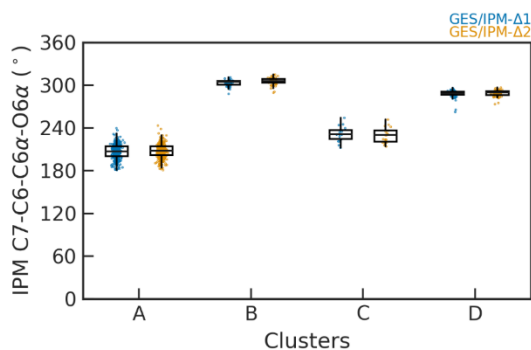


Figure 6.8 The distribution of the IPM C7-C6-C6 α -O6 α dihedral angle of each conformational cluster.

6.4 Conclusion

In summary, I calculated 1,000 QM/MM MEPs for the deacylation between GES-5 and IPM with two tautomer states. Based on the MEP dataset, I 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, I delineate the mechanisms of two major factors that impact the deacylation reactivity in GES/IPM hydrolysis. The protonation on IPM- $\Delta 2$ 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 6 α -hydroxyethyl group could adopt two orientations and interacts with the reacting groups and thus impacts the deacylating barrier. In the ECGCNN model, this is demonstrated as significant latent displacements upon the removal of edges relevant to the IPM 6 α side chain. Most importantly, while no explicit representation of the hydroxyethyl orientation was encoded in the feature representation, my ECGCNN model could capture this hidden information and manifest high sensitivity to the IPM O6 α – IPM C7 distance which is highly correlated with the IPM 6 α -hydroxyethyl orientation. Finally, my study demonstrates the potential for DL/GL methods in assisting the mechanistic understandings of enzyme catalysis.

7. EMPIRICAL FORCE FIELD PARAMETRIZATION OF 2-AMINOTHIAZOLE

7.1 Parametrization Scheme of CGenFF

Although CGenFF provides automatic atom typing and parameter assignment for small bio-molecules based on the property of parameter-transferability, the generated parameters are not as accurate and needs to be further refined.^{121,122} The original workflow for CGenFF parameter refinement is shown in Figure 7.1 and is described as follows:

- Atom types (Van der Waals parameters) are transferred from CGenFF;
- Atomic partial charges are parameterized to HF/6-31G* water interaction energies from at MP2/6-31G* optimized geometries;
- Equilibrium bond length & angles are fitted to reproduce the molecular structure from MP2/6-31G* calculations;
- Force constants are optimized to reproduce the vibrational frequencies from MP2/6-31G* calculations; The vibrational frequencies are calculated by the MOLVIB module integrated in CHARMM, which requires Pulay's Natural Internal Coordinate (NIC)^{156,157} as the input format to specify both fragment topology and coordinates;
- Multiplicity and phase shift on torsional terms are determined by the connectivity of the dihedrals and thus need not to be changed;
- Divide and conquer: when dealing with large molecules with more than 20 atoms, the molecule is divided into fragments and each fragment is parametrized independently. The

parametrized fragments are then joined together and the energy terms on the connecting part is further optimized using QM PES profile as the target data.

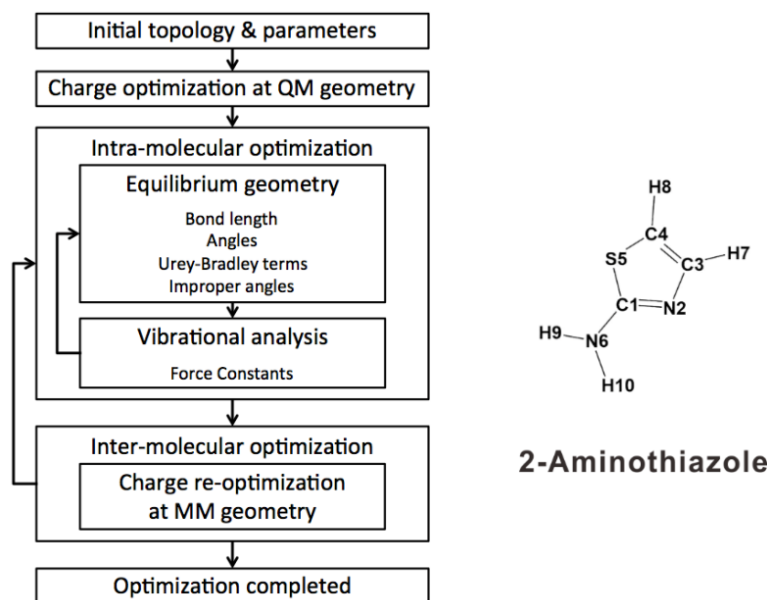


Figure 7.1 Flowchart for CGenFF parametrization and chemical structure of 2-aminothiazole.

7.2 Parameter Searching Strategy

Searching for the optimal combination of force field parameters is a heavy task due to the dimensionality of the parameters and simple exhaustive method is computationally unaffordable. Accordingly, an adaptive grid searching strategy is adopted for the automation of parameter searching. As demonstrated in Figure 7.2, instead of generating a tightly distributed point grid, a diffused grid was firstly generated with three points separated by large point span on each direction. The optimal parameter combination was chosen as the reference point and a new grid with smaller span was generated and tested. Finally, this iterative process converges when the difference between the MM results and QM data is smaller than criteria threshold. Moreover, the

grid points were generated using the First Depth Search (FDS) algorithm to avoid revisiting duplicated parameter combinations. This strategy is also important as in most cases, the variable space consists of more than eight dimensions, leaving exhaustive generation of all possible parameter set extremely time-consuming.

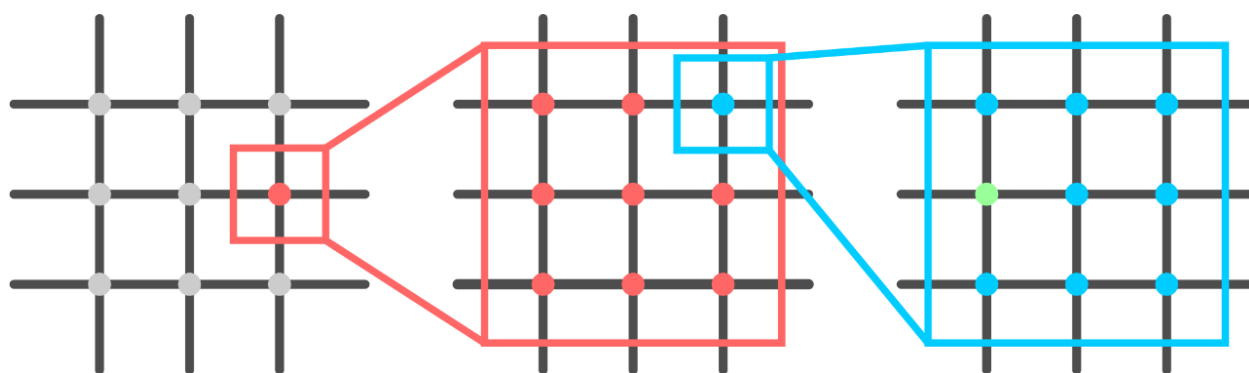


Figure 7.2 Demonstration of adaptive grid search in a 2D variable space.

7.3 Parametrization of 2-Aminothiazole: Atomic Partial Charges

As shown in Figure 7.3, seven water interactions were firstly constructed. the optimal interaction energies and distances are calculated using rigid PES scans on the intermolecular distance between H₂O and the target atom. The deviation between the water interaction energies at QM (HF/6-31G*/MP2/6-31G*) level of theory and MM level of theory are all larger than 1 kcal/mol, as presented in Table 7.1. After the optimization of atomic charges, the maximum deviation in the water interaction energies is reduced to less than 0.07 kcal/mol. Notably, water interaction on H9 is not optimized independently as the compromise for keeping H9 and H10 indistinguishable.

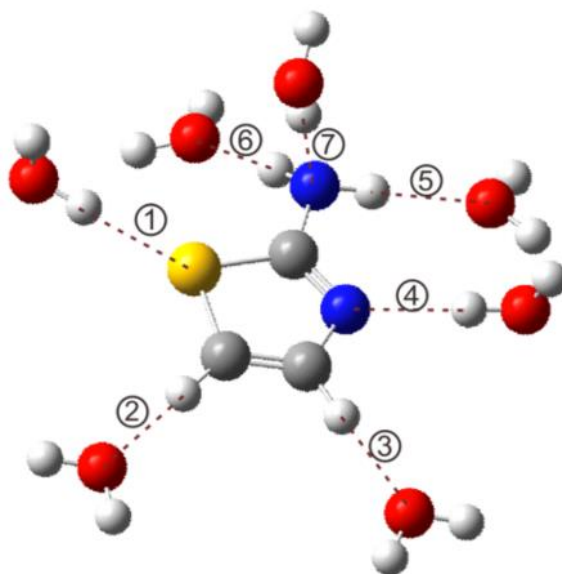


Figure 7.3 Water interactions considered for 2-Aminothiazole.

Table 7.1 Deviation to QM water interaction energies of initial and optimized atomic charges.

No. Interaction / Atom	Initial charge	ΔE (kcal mol ⁻¹)	Fitted charge	ΔE (kcal mol ⁻¹)
1 / S ₅	-0.053	-3.201	0.470	0.003
2 / C ₄	-0.185	-3.344	-0.659	-0.027
3 / C ₃	0.209	-3.319	-0.146	-0.068
4 / N ₂	-0.620	-2.568	-0.440	0.005
5 / H ₁₀	0.355	-1.354	0.164	0.010
6 / H ₉	0.355	-1.778	0.164	-0.926
7 / N ₆	-0.760	-2.593	-0.479	0.032

7.4 Parametrization of 2-Aminothiazole: Equilibrium Terms

The equilibrium terms are fitted to MP2/6-31G* minimized structures. The initial and fitted parameters are shown in Table 7.2. The comparison between the QM and MM minimized geometry are presented in the following sections.

Table 7.2 Initial and optimized equilibrium bond length and bond angles.

Equilibrium Bond Length or Angles	Initial	Fitted
C1-N6	1.4000 Å	1.3779 Å
N2-C1-N6	123.00°	127.55°
S5-C1-N6	119.80°	116.60°
C1-N6-H9	115.00°	117.55°

7.5 Parametrization of 2-Aminothiazole: Force Constants

The force constants are optimized to reproduce the vibrational frequencies at MP2/6-31G*, the optimized force constants are shown in Table 7.3.

Table 7.3 Initial and optimized force constants, units vary with entries.

Force Constants	Initial	Fitted
C1-N6	330.00	370.00
N2-C1-N6	45.80	45.00
S5-C1-N6	25.00	58.00
C1-N6-H9	45.00	68.00
C3-N2-C1-N6	3.0000	7.0000
N2-C1-N6-H9	0.3200	2.5000
S5-C1-N6-H9	0.3200	2.3000
C4-S5-C1-N6	4.0000	5.0000

The comparison between the vibrational frequencies calculated at QM level of theory and MM level of theory are presented in Figure 7.4. The underfitting of the frequency in the red dashed box is contributed (>95%) by the scissoring motion of the NH₂ group. Such a vibrational mode could be expressed by a linear combination of three bond angles, shown as α , β and γ in Figure 7.4. The γ angle is a well-parametrized bond bending term in the original CGenFF and was adopted in the current fragment due to the transferability of force field parameters. Modifying this bending term violates the parametrization protocol as this bending term is shared by a number of well-parametrized fragments in the CGenFF. Thus, optimizing this frequency mode demands introducing new atom types and is beyond necessity. Regardless, this discrepancy between QM and MM vibrational frequency is left aside with no further optimization.

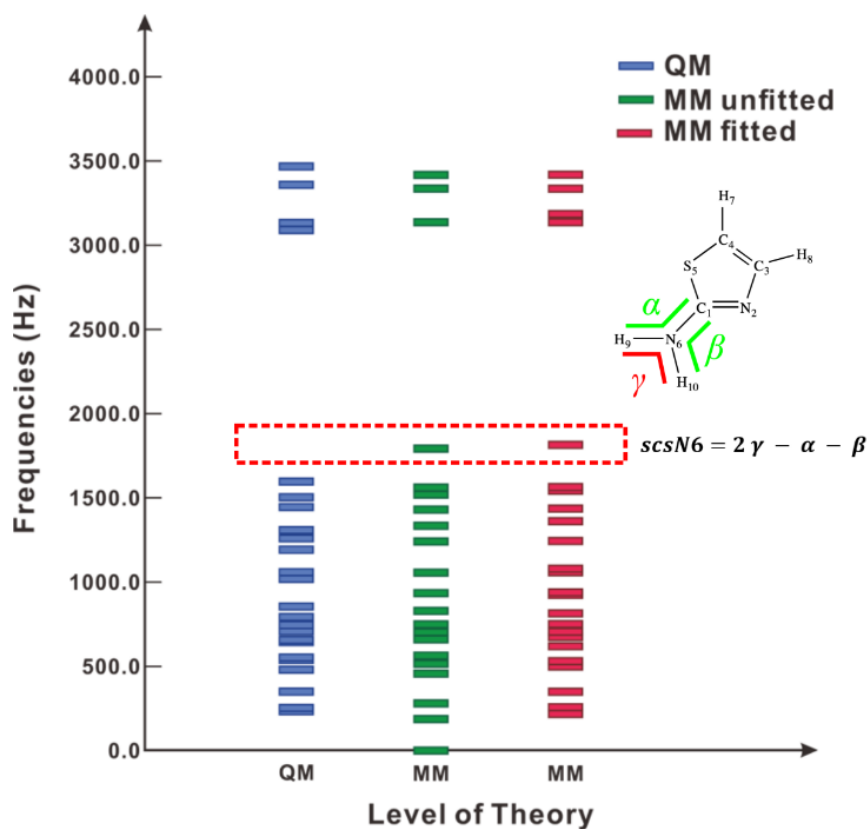


Figure 7.4 CGenFF parametrization for vibrational frequencies, the frequency marked in the red dashed box produced by the scissoring of the amino group.

7.6 Comparison of QM and MM optimized geometries

The performance of the MM force field parameters on 2-aminothiazole are benchmarked on the basis of minimized structures, as shown in Figure 7.5. It can be seen that, the unfitted CGenFF parameters produce a symmetrical conformation, whereas the MP2/6-31G* geometries give an asymmetrical conformation with the conjugation C-C double bond and the lone pairs on the amino group. The minimal conformation is improved after parametrizing for equilibrium structures, but the torsional angle is not optimal. Upon fitted for vibrational frequencies, the MM minimization could reasonably reproduce the QM minimized conformation. Although compromise was made during the parameter searching, the performance of the CGenFF parameters for 2-aminothiazole is satisfying.

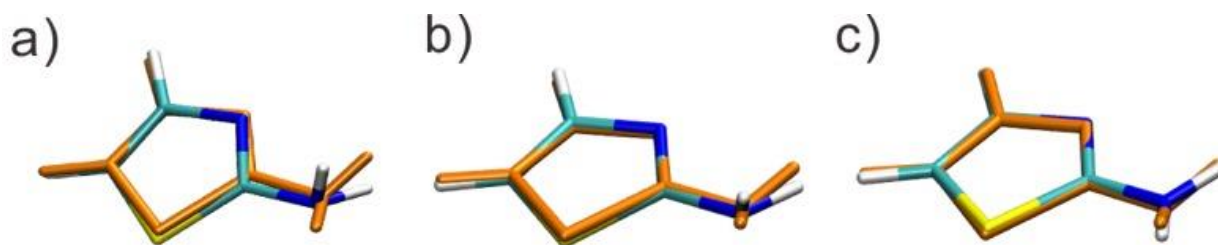


Figure 7.5 Comparison of QM (in orange) and MM (in multi-color) minimized geometry: (a) with unfitted force field parameters; (b) with force field parameters fitted for equilibrium geometry; (c) with force field parameters fitted for vibrational frequencies.

8. CONCLUSIONS

In this thesis, the molecular mechanisms of antibiotic resistance driven by AS β LS are unraveled with QM/MM MEP calculations and explainable ML models.

The detailed molecular mechanism of the acylation reaction between TEM-1, a penicillinase, and benzylpenicillin is firstly revisited. I proposed two types of perturbation-based importance attribution method for understanding the energy contributions per reaction step. Both methods provide consistent measures of the energy contributions and demonstrated that the collapsing of the TEM-1/Benzylpenicillin tetrahedral intermediates is the rate limiting reaction step of the overall acylation. Along this line, the discrepancies between two pioneering studies on the same system is also bridged.

The detailed molecular mechanism of the acylation reaction between Toho-1, a cephalosporinase, and AMP/CEX is then investigated. I proposed the BCIG approach, an XAI method, to explain the machine-learned QM/MM MEPs. The DL model of learning QM/MM MEPs adopts the DaWNN architecture and achieved high accuracy and scalability. The BCIG approach could attribute variable contributions of the overall reaction profile to individual reaction steps. The BCIG metric are validated on additional QM/MM calculations and the intuitive understanding of the reaction. Moreover, based on this study combining QM/MM and XAI, I proposed that the acylation of Toho-1/CEX does not happen using Glu166 as the only general base, rather, it has to be mediated by Lys73/Glu166 as the concerted base.

The detailed molecular mechanism of the deacylation activity of GES-5, a carbapenemase, and IPM is also studied. I built graph representations of the acyl-enzyme active site of GES-5/IPM, and applied GL model to predict the deacylation barrier heights calculated from the QM/MM MEPs. The GL model adopts a edge-conditioned learning scheme and effectively utilizes atomic distances as the edge information to reach a prediction error of $< 3 \text{ kcal mol}^{-1}$. I proposed that the conformational factors that regulate the deacylation activity could be probed from a perturbative edge-removing scheme on the GL model. Accordingly, I concluded that the tautomerization states on the IPM pyrroline ring and the orientations of the IPM 6 α -hydroxyethyl side group are two major factors that impacts and deacylation activity in IPM hydrolysis by GES-5.

Further discussions in this thesis focused on developing classical molecular force fields for β -lactam molecules. The parametrization of a common β -lactam fragment, 2-aminothiazole, is detailed. The parametrized force field files for the 2-aminothiazole and cephalothin were appended as appendix in the hope that it would be helpful to future researches. However, I note that those force field files are not extensively tested nor peer-reviewed.

In summary, my Ph.D. efforts are mostly devoted to understand molecular mechanisms underlying AS β L-driven β -lactam resistance combining QM/MM and XAI. Using three AS β L enzymes with distinct functions, I have shown that these ML/DL/GL models can be explained with chemical relevance. In other words, the ML/DL/GL models trained on reasonable chemical datasets could also encode essential hidden mechanistic information, which could be unraveled using XAI methods for mechanistic insights. To conclude, I proved that ML and model explanation methods could be effectively introduced for reliable QM/MM post-analysis, which extracts chemical insights of great importance and details for enzyme mechanism studies.

9. BIBLIOGRAPHY

- (1) Gould, I. M. A Review of the Role of Antibiotic Policies in the Control of Antibiotic Resistance. *J. Antimicrob. Chemother.* **1999**, 43 (4), 459–465. <https://doi.org/10.1093/jac/43.4.459>.
- (2) *Antimicrobial Agents: Antibacterials and Antifungals*; Bryskier, A., Ed.; ASM Press: Washington, D.C, 2005.
- (3) Fisher, J. F.; Meroueh, S. O.; Mobashery, S. Bacterial Resistance to β -Lactam Antibiotics: Compelling Opportunism, Compelling Opportunity. *Chem. Rev.* **2005**, 105 (2), 395–424. <https://doi.org/10.1021/cr030102i>.
- (4) Llarrull, L. I.; Testero, S. A.; Fisher, J. F.; Mobashery, S. The Future of the β -Lactams. *Curr. Opin. Microbiol.* **2010**, 13 (5), 551–557. <https://doi.org/10.1016/j.mib.2010.09.008>.
- (5) Papp-Wallace, K. M.; Endimiani, A.; Taracila, M. A.; Bonomo, R. A. Carbapenems: Past, Present, and Future. *Antimicrob. Agents Chemother.* **2011**, 55 (11), 4943–4960. <https://doi.org/10.1128/AAC.00296-11>.
- (6) Hoemann, M. Z. Penicillin and Cephalosporin Antibiotics. In *Bioactive Heterocyclic Compound Classes*; Dinges, J., Lamberth, C., Eds.; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2013; pp 237–253. <https://doi.org/10.1002/9783527664450.ch15>.
- (7) Chaibi, E. B.; Sirot, D.; Paul, G.; Labia, R. Inhibitor-Resistant TEM-Lactamases: Phenotypic, Genetic and Biochemical Characteristics. *J. Antimicrob. Chemother.* **1999**, 43 (4), 447–458. <https://doi.org/10.1093/jac/43.4.447>.
- (8) Cantón, R.; Coque, T. M. The CTX-M β -Lactamase Pandemic. *Curr. Opin. Microbiol.* **2006**, 9 (5), 466–475. <https://doi.org/10.1016/j.mib.2006.08.011>.
- (9) Cantón, R.; González-Alba, J. M.; Galán, J. C. CTX-M Enzymes: Origin and Diffusion. *Front Microbiol* **2012**, 3, 110–110. <https://doi.org/10.3389/fmicb.2012.00110>.

- (10) Rossolini, G. M.; D'Andrea, M. M.; Mugnaioli, C. The Spread of CTX-M-Type Extended-Spectrum β -Lactamases. *Clinical Microbiology and Infection* **2008**, *14*, 33–41. <https://doi.org/10.1111/j.1469-0691.2007.01867.x>.
- (11) D'Andrea, M. M.; Arena, F.; Pallecchi, L.; Rossolini, G. M. CTX-M-Type β -Lactamases: A Successful Story of Antibiotic Resistance. *International Journal of Medical Microbiology* **2013**, *303* (6–7), 305–317. <https://doi.org/10.1016/j.ijmm.2013.02.008>.
- (12) Nordmann, P.; Dortet, L.; Poirel, L. Carbapenem Resistance in Enterobacteriaceae: Here Is the Storm! *Trends Mol. Med.* **2012**, *18* (5), 263–272. <https://doi.org/10.1016/j.molmed.2012.03.003>.
- (13) Queenan, A. M.; Bush, K. Carbapenemases: The Versatile β -Lactamases. *Clin. Microbiol. Rev.* **2007**, *20* (3), 440–458. <https://doi.org/10.1128/CMR.00001-07>.
- (14) Paterson, D. L.; Bonomo, R. A. Extended-Spectrum-Lactamases: A Clinical Update. *Clin. Microbiol. Rev.* **2005**, *18* (4), 657–686. <https://doi.org/10.1128/CMR.18.4.657-686.2005>.
- (15) Bradford, P. A. Extended-Spectrum-Lactamases in the 21st Century: Characterization, Epidemiology, and Detection of This Important Resistance Threat. *Clin. Microbiol. Rev.* **2001**, *14* (4), 933–951. <https://doi.org/10.1128/CMR.14.4.933-951.2001>.
- (16) Hall, B. G.; Barlow, M. Evolution of the Serine β -Lactamases: Past, Present and Future. *Drug Resistance Updates* **2004**, *7* (2), 111–123. <https://doi.org/10.1016/j.drug.2004.02.003>.
- (17) 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. <https://doi.org/10.3389/fmolb.2018.00016>.
- (18) Walther-Rasmussen, J.; Høiby, N. Class A Carbapenemases. *J. Antimicrob. Chemother.* **2007**, *60* (3), 470–482. <https://doi.org/10.1093/jac/dkm226>.
- (19) Bush, K.; Jacoby, G. A. Updated Functional Classification of β -Lactamases. *Antimicrob. Agents Chemother.* **2010**, *54* (3), 969–976. <https://doi.org/10.1128/AAC.01009-09>.
- (20) Nerenberg, P. S.; Head-Gordon, T. New Developments in Force Fields for Biomolecular Simulations. *Current Opinion in Structural Biology* **2018**, *49*, 129–138. <https://doi.org/10.1016/j.sbi.2018.02.002>.
- (21) Jing, Z.; Liu, C.; Cheng, S. Y.; Qi, R.; Walker, B. D.; Piquemal, J.-P.; Ren, P. Polarizable Force Fields for Biomolecular Simulations: Recent Advances and Applications. *Annu. Rev. Biophys.* **2019**, *48* (1), 371–394. <https://doi.org/10.1146/annurev-biophys-070317-033349>.

- (22) Huang, J.; MacKerell, A. D. Force Field Development and Simulations of Intrinsically Disordered Proteins. *Current Opinion in Structural Biology* **2018**, *48*, 40–48. <https://doi.org/10.1016/j.sbi.2017.10.008>.
- (23) 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* (9), 3257–3273. <https://doi.org/10.1021/ct300400x>.
- (24) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat Methods* **2017**, *14* (1), 71–73. <https://doi.org/10.1038/nmeth.4067>.
- (25) Liu, M.; Wang, Y.; Chen, Y.; Field, M. J.; Gao, J. QM/MM through the 1990s: The First Twenty Years of Method Development and Applications. *Isr. J. Chem.* **2014**, *54* (8–9), 1250–1263. <https://doi.org/10.1002/ijch.201400036>.
- (26) Boulanger, E.; Harvey, J. N. QM/MM Methods for Free Energies and Photochemistry. *Current Opinion in Structural Biology* **2018**, *49*, 72–76. <https://doi.org/10.1016/j.sbi.2018.01.003>.
- (27) Senn, H. M.; Thiel, W. QM/MM Studies of Enzymes. *Current Opinion in Chemical Biology* **2007**, *11* (2), 182–187. <https://doi.org/10.1016/j.cbpa.2007.01.684>.
- (28) Riccardi, D.; Schaefer, P.; Yang, Yu, H.; Ghosh, N.; Prat-Resina, X.; König, P.; Li, G.; Xu, D.; Guo, H.; Elstner, M.; Cui, Q. Development of Effective Quantum Mechanical/Molecular Mechanical (QM/MM) Methods for Complex Biological Processes. *J. Phys. Chem. B* **2006**, *110* (13), 6458–6469. <https://doi.org/10.1021/jp056361o>.
- (29) Gao, J.; Xia, X. A Priori Evaluation of Aqueous Polarization Effects through Monte Carlo QM-MM Simulations. *Science* **1992**, *258* (5082), 631–635. <https://doi.org/10.1126/science.1411573>.
- (30) Reuter, N.; Dejaegere, A.; Maigret, B.; Karplus, M. Frontier Bonds in QM/MM Methods: A Comparison of Different Approaches. *J. Phys. Chem. A* **2000**, *104* (8), 1720–1735. <https://doi.org/10.1021/jp9924124>.
- (31) Murphy, R. B.; Philipp, D. M.; Friesner, R. A. Frozen Orbital QM/MM Methods for Density Functional Theory. *Chemical Physics Letters* **2000**, *321* (1–2), 113–120. [https://doi.org/10.1016/S0009-2614\(00\)00289-X](https://doi.org/10.1016/S0009-2614(00)00289-X).

- (32) Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. A Generalized Hybrid Orbital (GHO) Method for the Treatment of Boundary Atoms in Combined QM/MM Calculations. *J. Phys. Chem. A* **1998**, *102* (24), 4714–4721. <https://doi.org/10.1021/jp9809890>.
- (33) Zhang, Y. Pseudobond Ab Initio QM/MM Approach and Its Applications to Enzyme Reactions. *Theor Chem Acc* **2006**, *116* (1–3), 43–50. <https://doi.org/10.1007/s00214-005-0008-x>.
- (34) Cui, Q.; Karplus, M. Molecular Properties from Combined QM/MM Methods. I. Analytical Second Derivative and Vibrational Calculations. *The Journal of Chemical Physics* **2000**, *112* (3), 1133–1149. <https://doi.org/10.1063/1.480658>.
- (35) Cui, Q.; Karplus, M. Molecular Properties from Combined QM/MM Methods. 2. Chemical Shifts in Large Molecules. *J. Phys. Chem. B* **2000**, *104* (15), 3721–3743. <https://doi.org/10.1021/jp994154g>.
- (36) Niehaus, T. A.; Elstner, M.; Frauenheim, Th.; Suhai, S. Application of an Approximate Density-Functional Method to Sulfur Containing Compounds. *Journal of Molecular Structure: THEOCHEM* **2001**, *541* (1), 185–194. [https://doi.org/10.1016/S0166-1280\(00\)00762-4](https://doi.org/10.1016/S0166-1280(00)00762-4).
- (37) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, Th.; Suhai, S.; Seifert, G. Self-Consistent-Charge Density-Functional Tight-Binding Method for Simulations of Complex Materials Properties. *Phys. Rev. B* **1998**, *58* (11), 7260–7268. <https://doi.org/10.1103/PhysRevB.58.7260>.
- (38) 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* (4), 931–948. <https://doi.org/10.1021/ct100684s>.
- (39) Gaus, M.; Goez, A.; Elstner, M. Parametrization and Benchmark of DFTB3 for Organic Molecules. *J. Chem. Theory Comput.* **2013**, *9* (1), 338–354. <https://doi.org/10.1021/ct300849w>.
- (40) Ewald, P. P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.* **1921**, *369* (3), 253–287. <https://doi.org/10.1002/andp.19213690304>.
- (41) 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* (12), 10089–10092. <https://doi.org/10.1063/1.464397>.
- (42) Banushkina, P. V.; Krivov, S. V. Optimal Reaction Coordinates. *WIREs Comput Mol Sci* **2016**, *6* (6), 748–763. <https://doi.org/10.1002/wcms.1276>.

- (43) Lu, X.; Fang, D.; Ito, S.; Okamoto, Y.; Ovchinnikov, V.; Cui, Q. QM/MM Free Energy Simulations: Recent Progress and Challenges. *null* **2016**, *42* (13), 1056–1078. <https://doi.org/10.1080/08927022.2015.1132317>.
- (44) Dewyer, A. L.; Argüelles, A. J.; Zimmerman, P. M. Methods for Exploring Reaction Space in Molecular Systems: Exploring Reaction Space in Molecular Systems. *WIREs Comput Mol Sci* **2018**, *8* (2), e1354. <https://doi.org/10.1002/wcms.1354>.
- (45) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13* (8), 1011–1021. <https://doi.org/10.1002/jcc.540130812>.
- (46) Hub, J. S.; de Groot, B. L.; van der Spoel, D. G. wham—A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *J. Chem. Theory Comput.* **2010**, *6* (12), 3713–3720. <https://doi.org/10.1021/ct100494z>.
- (47) Jónsson, H.; Mills, G.; Jacobsen, K. W. Nudged Elastic Band Method for Finding Minimum Energy Paths of Transitions. In *Classical and Quantum Dynamics in Condensed Phase Simulations*; WORLD SCIENTIFIC: LERICI, Villa Marigola, 1998; pp 385–404. https://doi.org/10.1142/9789812839664_0016.
- (48) Woodcock, H. L.; Hodošček, M.; Sherwood, P.; Lee, Y. S.; Schaefer III, H. F.; Brooks, B. R. Exploring the Quantum Mechanical/Molecular Mechanical Replica Path Method: A Pathway Optimization of the Chorismate to Prephenate Claisen Rearrangement Catalyzed by Chorismate Mutase. *Theor. Chem. Acc.* **2003**, *109* (3), 140–148. <https://doi.org/10.1007/s00214-002-0421-3>.
- (49) 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* (8), 2050–2061. <https://doi.org/10.1021/ct9001398>.
- (50) E, W.; Ren, W.; Vanden-Eijnden, E. String Method for the Study of Rare Events. *Phys. Rev. B* **2002**, *66* (5), 052301. <https://doi.org/10.1103/PhysRevB.66.052301>.
- (51) Chu, J.-W.; Trout, B. L.; Brooks, B. R. A Super-Linear Minimization Scheme for the Nudged Elastic Band Method. *J. Chem. Phys.* **2003**, *119* (24), 12708–12717. <https://doi.org/10.1063/1.1627754>.
- (52) Henkelman, G.; Uberuaga, B. P.; Jónsson, H. A Climbing Image Nudged Elastic Band Method for Finding Saddle Points and Minimum Energy Paths. *The Journal of Chemical Physics* **2000**, *113* (22), 9901–9904. <https://doi.org/10.1063/1.1329672>.

- (53) Herbol, H. C.; Stevenson, J.; Clancy, P. Computational Implementation of Nudged Elastic Band, Rigid Rotation, and Corresponding Force Optimization. *J. Chem. Theory Comput.* **2017**, *13* (7), 3250–3259. <https://doi.org/10.1021/acs.jctc.7b00360>.
- (54) Maragliano, L.; Fischer, A.; Vanden-Eijnden, E.; Ciccotti, G. String Method in Collective Variables: Minimum Free Energy Paths and Isocommittor Surfaces. *The Journal of Chemical Physics* **2006**, *125* (2), 024106. <https://doi.org/10.1063/1.2212942>.
- (55) Vanden-Eijnden, E.; Venturoli, M. Revisiting the Finite Temperature String Method for the Calculation of Reaction Tubes and Free Energies. *The Journal of Chemical Physics* **2009**, *130* (19), 194103. <https://doi.org/10.1063/1.3130083>.
- (56) Vanden-Eijnden, E.; Venturoli, M. Exact Rate Calculations by Trajectory Parallelization and Tilting. *The Journal of Chemical Physics* **2009**, *131* (4), 044120. <https://doi.org/10.1063/1.3180821>.
- (57) E, W.; Ren, W.; Vanden-Eijnden, E. Finite Temperature String Method for the Study of Rare Events. *J. Phys. Chem. B* **2005**, *109* (14), 6688–6693. <https://doi.org/10.1021/jp0455430>.
- (58) E, W.; Ren, W.; Vanden-Eijnden, E. Simplified and Improved String Method for Computing the Minimum Energy Paths in Barrier-Crossing Events. *The Journal of Chemical Physics* **2007**, *126* (16), 164103. <https://doi.org/10.1063/1.2720838>.
- (59) 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* (3), 327–341. [https://doi.org/10.1016/0021-9991\(77\)90098-5](https://doi.org/10.1016/0021-9991(77)90098-5).
- (60) Kulik, H.; Hammerschmidt, T.; Schmidt, J.; Botti, S.; Marques, M. A. L.; Boley, M.; Scheffler, M.; Todorović, M.; Rinke, P.; Oses, C.; Smolyanyuk, A.; Curtarolo, S.; Tkatchenko, A.; Bartok, A.; Manzhos, S.; Ihara, M.; Carrington, T.; Behler, J.; Isayev, O.; Veit, M.; Grisafi, A.; Nigam, J.; Ceriotti, M.; Schütt, K. T.; Westermayr, J.; Gastegger, M.; Maurer, R.; Kalita, B.; Burke, K.; Nagai, R.; Akashi, R.; Sugino, O.; Hermann, J.; Noé, F.; Pilati, S.; Draxl, C.; Kuban, M.; Rigamonti, S.; Scheidgen, M.; Esters, M.; Hicks, D.; Toher, C.; Balachandran, P.; Tamblyn, I.; Whitelam, S.; Bellinger, C.; Ghiringhelli, L. M. Roadmap on Machine Learning in Electronic Structure. *Electron. Struct.* **2022**. <https://doi.org/10.1088/2516-1075/ac572f>.
- (61) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine Learning for Molecular and Materials Science. *Nature* **2018**, *559* (7715), 547–555. <https://doi.org/10.1038/s41586-018-0337-2>.
- (62) 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 Discov.* **2019**, *18* (6), 463–477. <https://doi.org/10.1038/s41573-019-0024-5>.

- (63) Dral, P. O. Quantum Chemistry in the Age of Machine Learning. *J. Phys. Chem. Lett.* **2020**, *11* (6), 2336–2347. <https://doi.org/10.1021/acs.jpcllett.9b03664>.
- (64) Behler, J. Four Generations of High-Dimensional Neural Network Potentials. *Chem. Rev.* **2021**, *121* (16), 10037–10072. <https://doi.org/10.1021/acs.chemrev.0c00868>.
- (65) Schütt, K. T.; Arbabzadah, F.; Chmiela, S.; Müller, K. R.; Tkatchenko, A. Quantum-Chemical Insights from Deep Tensor Neural Networks. *Nat Commun* **2017**, *8* (1), 13890. <https://doi.org/10.1038/ncomms13890>.
- (66) 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* (1), 2903. <https://doi.org/10.1038/s41467-019-10827-4>.
- (67) Faber, F. A.; Hutchison, L.; Huang, B.; Gilmer, J.; Schoenholz, S. S.; Dahl, G. E.; Vinyals, O.; Kearnes, S.; Riley, P. F.; von Lilienfeld, O. A. Prediction Errors of Molecular Machine Learning Models Lower than Hybrid DFT Error. *J. Chem. Theory Comput.* **2017**, *13* (11), 5255–5264. <https://doi.org/10.1021/acs.jctc.7b00577>.
- (68) Dral, P. O.; Owens, A.; Dral, A.; Csányi, G. Hierarchical Machine Learning of Potential Energy Surfaces. *J. Chem. Phys.* **2020**, *152* (20), 204110. <https://doi.org/10.1063/5.0006498>.
- (69) 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* (9), 096405. <https://doi.org/10.1103/PhysRevLett.114.096405>.
- (70) Noé, F.; Olsson, S.; Köhler, J.; Wu, H. Boltzmann Generators: Sampling Equilibrium States of Many-Body Systems with Deep Learning. *Science* **2019**, *365* (6457), eaaw1147. <https://doi.org/10.1126/science.aaw1147>.
- (71) Van Oort, C. M.; Ferrell, J. B.; Remington, J. M.; Wshah, S.; Li, J. AMPGAN v2: Machine Learning-Guided Design of Antimicrobial Peptides. *J. Chem. Inf. Model.* **2021**, *61* (5), 2198–2207. <https://doi.org/10.1021/acs.jcim.0c01441>.
- (72) Agarwal, S.; Dugar, D.; Sengupta, S. Ranking Chemical Structures for Drug Discovery: A New Machine Learning Approach. *J. Chem. Inf. Model.* **2010**, *50* (5), 716–731. <https://doi.org/10.1021/ci9003865>.
- (73) Hooft, F.; Pérez de Alba Ortiz, A.; Ensing, B. Discovering Collective Variables of Molecular Transitions via Genetic Algorithms and Neural Networks. *J. Chem. Theory Comput.* **2021**, *17* (4), 2294–2306. <https://doi.org/10.1021/acs.jctc.0c00981>.

- (74) Fleetwood, O.; Kasimova, M. A.; Westerlund, A. M.; Delemotte, L. Molecular Insights from Conformational Ensembles via Machine Learning. *Biophysical Journal* **2020**, *118* (3), 765–780. <https://doi.org/10.1016/j.bpj.2019.12.016>.
- (75) Murdoch, W. J.; Singh, C.; Kumbier, K.; Abbasi-Asl, R.; Yu, B. Definitions, Methods, and Applications in Interpretable Machine Learning. *Proc Natl Acad Sci USA* **2019**, *116* (44), 22071–22080. <https://doi.org/10.1073/pnas.1900654116>.
- (76) Glielmo, A.; Husic, B. E.; Rodriguez, A.; Clementi, C.; Noé, F.; Laio, A. Unsupervised Learning Methods for Molecular Simulation Data. *Chem. Rev.* **2021**, *121* (16), 9722–9758. <https://doi.org/10.1021/acs.chemrev.0c01195>.
- (77) 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 Herreros, A. G.; Aloy, P. Bioactivity Descriptors for Uncharacterized Chemical Compounds. *Nat Commun* **2021**, *12* (1), 3932. <https://doi.org/10.1038/s41467-021-24150-4>.
- (78) 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* (24), 241740. <https://doi.org/10.1063/1.5029879>.
- (79) 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* (4), 901–908. <https://doi.org/10.1021/acs.jpcc.8b11905>.
- (80) Bösel, L.; Thürlmann, M.; Riniker, S. Machine Learning in QM/MM Molecular Dynamics Simulations of Condensed-Phase Systems. *J. Chem. Theory Comput.* **2021**, *17* (5), 2641–2658. <https://doi.org/10.1021/acs.jctc.0c01112>.
- (81) 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* (9), 5745–5758. <https://doi.org/10.1021/acs.jctc.1c00565>.
- (82) Shen, L.; Wu, J.; Yang, W. Multiscale Quantum Mechanics/Molecular Mechanics Simulations with Neural Networks. *J. Chem. Theory Comput.* **2016**, *12* (10), 4934–4946. <https://doi.org/10.1021/acs.jctc.6b00663>.
- (83) Shen, L.; Yang, W. Molecular Dynamics Simulations with Quantum Mechanics/Molecular Mechanics and Adaptive Neural Networks. *J. Chem. Theory Comput.* **2018**, *14* (3), 1442–1455. <https://doi.org/10.1021/acs.jctc.7b01195>.

- (84) Wieder, O.; Kohlbacher, S.; Kuenemann, M.; Garon, A.; Ducrot, P.; Seidel, T.; Langer, T. A Compact Review of Molecular Property Prediction with Graph Neural Networks. *Drug Discov. Today Technol.* **2020**, *37*, 1–12. <https://doi.org/10.1016/j.ddtec.2020.11.009>.
- (85) Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; Dahl, G. E. Neural Message Passing for Quantum Chemistry. In *Proceedings of the 34th International Conference on Machine Learning*; PMLR, 2017; pp 1263–1272.
- (86) Hao, Z.; Lu, C.; Huang, Z.; Wang, H.; Hu, Z.; Liu, Q.; Chen, E.; Lee, C. ASGN: An Active Semi-Supervised Graph Neural Network for Molecular Property Prediction. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*; ACM, 2020; pp 731–752. <https://doi.org/10.1145/3394486.3403117>.
- (87) Chen, D.; Gao, K.; Nguyen, D. D.; Chen, X.; Jiang, Y.; Wei, G.-W.; Pan, F. Algebraic Graph-Assisted Bidirectional Transformers for Molecular Property Prediction. *Nat Commun* **2021**, *12* (1), 3521. <https://doi.org/10.1038/s41467-021-23720-w>.
- (88) Kearnes, S.; McCloskey, K.; Berndl, M.; Pande, V.; Riley, P. Molecular Graph Convolutions: Moving beyond Fingerprints. *J Comput Aided Mol Des* **2016**, *30* (8), 595–608. <https://doi.org/10.1007/s10822-016-9938-8>.
- (89) Wang, X.; Li, Z.; Jiang, M.; Wang, S.; Zhang, S.; Wei, Z. Molecule Property Prediction Based on Spatial Graph Embedding. *J. Chem. Inf. Model.* **2019**, *59* (9), 3817–3828. <https://doi.org/10.1021/acs.jcim.9b00410>.
- (90) Zhou, J.; Cui, G.; Hu, S.; Zhang, Z.; Yang, C.; Liu, Z.; Wang, L.; Li, C.; Sun, M. Graph Neural Networks: A Review of Methods and Applications. *AI Open* **2020**, *1*, 57–81. <https://doi.org/10.1016/j.aiopen.2021.01.001>.
- (91) Simonovsky, M.; Komodakis, N. Dynamic Edge-Conditioned Filters in Convolutional Neural Networks on Graphs. *arXiv:1704.02901* **2017**.
- (92) Gunning, D.; Stefik, M.; Choi, J.; Miller, T.; Stumpf, S.; Yang, G.-Z. XAI—Explainable Artificial Intelligence. *Sci. Robot.* **2019**, *4* (37), eaay7120. <https://doi.org/10.1126/scirobotics.aay7120>.
- (93) Cortes, C.; Vapnik, V. Support-Vector Networks. *Machine Learning* **1995**, *20* (3), 273–297. <https://doi.org/10.1007/BF00994018>.
- (94) Grömping, U. Variable Importance in Regression Models. *Wiley Interdiscip. Rev. Comput. Stat.* **2015**, *7* (2), 137–152. <https://doi.org/10.1002/wics.1346>.

- (95) Breiman, L. Random Forests. *Machine Learning* **2001**, 45 (1), 5–32. <https://doi.org/10.1023/A:1010933404324>.
- (96) Montavon, G.; Samek, W.; Müller, K.-R. Methods for Interpreting and Understanding Deep Neural Networks. *Digital Signal Processing* **2018**, 73, 1–15. <https://doi.org/10.1016/j.dsp.2017.10.011>.
- (97) 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 (3), 247–278. <https://doi.org/10.1109/JPROC.2021.3060483>.
- (98) 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.
- (99) 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. https://doi.org/10.1007/978-981-10-0557-2_87.
- (100) for the Alzheimer’s Neuroimaging Initiative; Jo, T.; Nho, K.; Risacher, S. L.; Saykin, A. J. Deep Learning Detection of Informative Features in Tau PET for Alzheimer’s Disease Classification. *BMC Bioinformatics* **2020**, 21 (S21), 496. <https://doi.org/10.1186/s12859-020-03848-0>.
- (101) Chereda, H.; Bleckmann, A.; Menck, K.; Perera-Bel, J.; Stegmaier, P.; Auer, F.; Kramer, F.; Leha, A.; Beißbarth, T. Explaining Decisions of Graph Convolutional Neural Networks: Patient-Specific Molecular Subnetworks Responsible for Metastasis Prediction in Breast Cancer. *Genome Med* **2021**, 13 (1), 42. <https://doi.org/10.1186/s13073-021-00845-7>.
- (102) 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 (1), 21155. <https://doi.org/10.1038/s41598-020-78169-6>.
- (103) Karpov, P.; Godin, G.; Tetko, I. V. Transformer-CNN: Swiss Knife for QSAR Modeling and Interpretation. *J Cheminform* **2020**, 12 (1), 17. <https://doi.org/10.1186/s13321-020-00423-w>.
- (104) Hooker, G.; Mentch, L.; Zhou, S. Unrestricted Permutation Forces Extrapolation: Variable Importance Requires at Least One More Model, or There Is No Free Variable Importance. **2019**. <https://doi.org/10.48550/ARXIV.1905.03151>.

- (105) 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* (12), 4454–4465. <https://doi.org/10.1021/ja044210d>.
- (106) 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* (44), 15397–15407. <https://doi.org/10.1021/ja051592u>.
- (107) Lietz, E. J.; Truher, H.; Kahn, D.; Hokenson, M. J.; Fink, A. L. Lysine-73 Is Involved in the Acylation and Deacylation of β -Lactamase. *Biochemistry* **2000**, *39* (17), 4971–4981. <https://doi.org/10.1021/bi992681k>.
- (108) Thomas, V. L.; Golemi-Kotra, D.; Kim, C.; Vakulenko, S. B.; Mobashery, S.; Shoichet, B. K. Structural Consequences of the Inhibitor-Resistant Ser130Gly Substitution in TEM β -Lactamase. *Biochemistry* **2005**, *44* (Copyright (C) 2020 American Chemical Society (ACS). All Rights Reserved.), 9330–9338. <https://doi.org/10.1021/bi0502700>.
- (109) Golemi-Kotra, D.; Meroueh, S. O.; Kim, C.; Vakulenko, S. B.; Bulychev, A.; Stemmler, A. J.; Stemmler, T. L.; Mobashery, Shahriar. The Importance of a Critical Protonation State and the Fate of the Catalytic Steps in Class A β -Lactamases and Penicillin-Binding Proteins. *J. Biol. Chem.* **2004**, *279* (Copyright (C) 2020 American Chemical Society (ACS). All Rights Reserved.), 34665–34673. <https://doi.org/10.1074/jbc.M313143200>.
- (110) Hermann, J. C.; Ridder, L.; Höltje, H.-D.; Mulholland, A. J. Molecular Mechanisms of Antibiotic Resistance: QM/MM Modeling of Deacylation in a Class A β -Lactamase. *Org. Biomol. Chem.* **2006**, *4* (2), 206–210. <https://doi.org/10.1039/B512969A>.
- (111) Pitarch, J.; Pascual-Ahuir, J.-L.; Silla, E.; Tuñón, I. A Quantum Mechanics/Molecular Mechanics Study of the Acylation Reaction of TEM1 β -Lactamase and Penicillanate. *J. Chem. Soc., Perkin Trans. 2* **2000**, No. 4, 761–767.
- (112) Brown, N. G.; Shanker, S.; Prasad, B. V. V.; Palzkill, T. Structural and Biochemical Evidence That a TEM-1 β -Lactamase N170G Active Site Mutant Acts via Substrate-Assisted Catalysis. *J. Biol. Chem.* **2009**, *284* (Copyright (C) 2020 American Chemical Society (ACS). All Rights Reserved.), 33703–33712. <https://doi.org/10.1074/jbc.M109.053819>.
- (113) Imtiaz, U.; Manavathu, E. K.; Lerner, S. A.; Mobashery, S. Critical Hydrogen Bonding by Serine 235 for Cephalosporinase Activity of TEM-1 β -Lactamase. *Antimicrob. Agents Chemother.* **1993**, *37* (Copyright (C) 2020 American Chemical Society (ACS). All Rights Reserved.), 2438–2442. <https://doi.org/10.1128/AAC.37.11.2438>.

- (114) Brooks, B. R.; Brooks III, C. L.; Mackerell Jr., A. D.; 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.; Hodoseck, 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* (10), 1545–1614. <https://doi.org/10.1002/jcc.21287>.
- (115) 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* (2), 569–585. <https://doi.org/10.1021/jp0029109>.
- (116) Woodcock, H. L.; Hodošček, M.; Gilbert, A. T. B.; Gill, P. M. W.; Schaefer III, H. F.; Brooks, B. R. Interfacing Q-Chem and CHARMM to Perform QM/MM Reaction Path Calculations. *J. Comput. Chem.* **2007**, *28* (9), 1485–1502. <https://doi.org/10.1002/jcc.20587>.
- (117) 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.; Begam, 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.; Ladjánszki, I.; Landau, A.; Lawler, K. V.; Lefrancois, 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.; Dunietz, 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* (8), 084801. <https://doi.org/10.1063/5.0055522>.
- (118) 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* (7), e1005659. <https://doi.org/10.1371/journal.pcbi.1005659>.
- (119) Strynadka, N. C. J.; Adachi, H.; Jensen, S. E.; Johns, K.; Sielecki, A.; Betzel, C.; Sutoh, K.; James, M. N. G. Molecular Structure of the Acyl-Enzyme Intermediate in β -Lactam Hydrolysis at 1.7 Å Resolution. *Nature* **1992**, *359* (6397), 700–705. <https://doi.org/10.1038/359700a0>.
- (120) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell Jr., A. D. 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* (4), 671–690. <https://doi.org/10.1002/jcc.21367>.
- (121) 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* (12), 3155–3168. <https://doi.org/10.1021/ci3003649>.
- (122) 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* (12), 3144–3154. <https://doi.org/10.1021/ci300363c>.
- (123) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926–935. <https://doi.org/10.1063/1.445869>.
- (124) 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, É. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
- (125) Baker, E. N.; Hubbard, R. E. Hydrogen Bonding in Globular Proteins. *Progress in Biophysics and Molecular Biology* **1984**, *44* (2), 97–179. [https://doi.org/10.1016/0079-6107\(84\)90007-5](https://doi.org/10.1016/0079-6107(84)90007-5).

- (126) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L.-P.; Lane, T. J.; Pande, V. S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal* **2015**, *109* (8), 1528–1532. <https://doi.org/10.1016/j.bpj.2015.08.015>.
- (127) Lonsdale, R.; Harvey, J. N.; Mulholland, A. J. Inclusion of Dispersion Effects Significantly Improves Accuracy of Calculated Reaction Barriers for Cytochrome P450 Catalyzed Reactions. *J. Phys. Chem. Lett.* **2010**, *1* (21), 3232–3237. <https://doi.org/10.1021/jz101279n>.
- (128) Lonsdale, R.; Harvey, J. N.; Mulholland, A. J. Effects of Dispersion in Density Functional Based Quantum Mechanical/Molecular Mechanical Calculations on Cytochrome P450 Catalyzed Reactions. *J. Chem. Theory Comput.* **2012**, *8* (11), 4637–4645. <https://doi.org/10.1021/ct300329h>.
- (129) Nitandai, Y.; Shimamura, T.; Uchiyama, T.; Ishii, Y.; Takehira, M.; Yutani, K.; Matsuzawa, H.; Miyano, M. The Catalytic Efficiency (Kcat/Km) of the Class A β -Lactamase Toho-1 Correlates with the Thermal Stability of Its Catalytic Intermediate Analog. *Biochim. Biophys. Acta* **2010**, *1804* (4), 684–691. <https://doi.org/10.1016/j.bbapap.2009.10.023>.
- (130) 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* (12), 1170–1177. <https://doi.org/10.1002/2211-5463.12132>.
- (131) Langan, P. S.; Vandavasi, V. G.; Cooper, S. J.; Weiss, K. L.; Ginell, S. L.; Parks, J. M.; Coates, L. Substrate Binding Induces Conformational Changes in a Class A β -Lactamase That Prime It for Catalysis. *ACS Catal.* **2018**, *8*, 2428–2437. <https://doi.org/10.1021/acscatal.7b04114>.
- (132) 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* (2), 785–789. <https://doi.org/10.1103/PhysRevB.37.785>.
- (133) Becke, A. D. A New Mixing of Hartree–Fock and Local Density-Functional Theories. *J. Chem. Phys.* **1993**, *98* (2), 1372–1377. <https://doi.org/10.1063/1.464304>.
- (134) Ditchfield, R.; Hehre, W. J.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1971**, *54* (2), 724–728. <https://doi.org/10.1063/1.1674902>.
- (135) 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* (5), 2257–2261. <https://doi.org/10.1063/1.1677527>.

- (136) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. Self-consistent Molecular Orbital Methods. XXIII. A Polarization-type Basis Set for Second-row Elements. *J. Chem. Phys.* **1982**, *77* (7), 3654–3665. <https://doi.org/10.1063/1.444267>.
- (137) 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* (3), 361–373. <https://doi.org/10.1002/jcc.540110311>.
- (138) Shimizu-Ibuka, A.; Oishi, M.; Yamada, S.; Ishii, Y.; Mura, K.; Sakai, H.; Matsuzawa, Hiroshi. Roles of Residues Cys69, Asn104, Phe160, Gly232, Ser237, and Asp240 in Extended-Spectrum β -Lactamase Toho-1. *Antimicrob. Agents Chemother.* **2011**, *55*, 284–290. <https://doi.org/10.1128/AAC.00098-10>.
- (139) Olmos, J. L.; Pandey, S.; Martin-Garcia, J. M.; Calvey, G.; Katz, A.; Knoska, J.; Kupitz, C.; Hunter, M. S.; Liang, M.; Oberthuer, D.; Yefanov, O.; Wiedorn, M.; Heyman, M.; Holl, M.; Pande, K.; Barty, A.; Miller, M. D.; Stern, S.; Roy-Chowdhury, S.; Coe, J.; Nagarathnam, N.; Zook, J.; Verburt, J.; Norwood, T.; Poudyal, I.; Xu, D.; Koglin, J.; Seaberg, M. H.; Zhao, Y.; Bajt, S.; Grant, T.; Mariani, V.; Nelson, G.; Subramanian, G.; Bae, E.; Fromme, R.; Fung, R.; Schwander, P.; Frank, M.; White, T. A.; Weierstall, U.; Zatsepin, N.; Spence, J.; Fromme, P.; Chapman, H. N.; Pollack, L.; Tremblay, L.; Ourmazd, A.; Phillips, G. N.; Schmidt, M. Enzyme Intermediates Captured “on the Fly” by Mix-and-Inject Serial Crystallography. *BMC Biol.* **2018**, *16* (1), 59. <https://doi.org/10.1186/s12915-018-0524-5>.
- (140) Cheng, H.-T.; Koc, L.; Harmsen, J.; Shaked, T.; Chandra, T.; Aradhye, 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. In *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems*; ACM: Boston, MA, USA, 2016; pp 7–10. <https://doi.org/10.1145/2988450.2988454>.
- (141) Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research* **2014**, *15* (56), 1929–1958.
- (142) Kingma, D. P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980*. 2017.
- (143) Frase, H.; Toth, M.; Champion, M. M.; Antunes, N. T.; Vakulenko, S. B. Importance of Position 170 in the Inhibition of GES-Type β -Lactamases by Clavulanic Acid. *Antimicrob. Agents Chemother.* **2011**, *55* (4), 1556–1562. <https://doi.org/10.1128/AAC.01292-10>.
- (144) Kotsakis, S. D.; Miriagou, V.; Tzelepi, E.; Tzouvelekis, L. S. 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.* **2010**, *54* (11), 4864–4871. <https://doi.org/10.1128/AAC.00771-10>.

- (145) Fonseca, F.; Chudyk, E. I.; van der Kamp, M. W.; Correia, A.; Mulholland, A. J.; Spencer, J. The Basis for Carbapenem Hydrolysis by Class A β -Lactamases: A Combined Investigation Using Crystallography and Simulations. *J. Am. Chem. Soc.* **2012**, *134* (44), 18275–18285. <https://doi.org/10.1021/ja304460j>.
- (146) Chudyk, E. I.; Limb, M. A. L.; Jones, C.; Spencer, J.; van der Kamp, M. W.; Mulholland, A. J. QM/MM Simulations as an Assay for Carbapenemase Activity in Class A β -Lactamases. *Chem. Commun.* **2014**, *50* (94), 14736–14739. <https://doi.org/10.1039/C4CC06495J>.
- (147) Lisa, M.-N.; Palacios, A. R.; Aitha, M.; González, M. M.; Moreno, D. M.; Crowder, M. W.; Bonomo, R. A.; Spencer, J.; Tierney, D. L.; Llarrull, L. I.; Vila, A. J. A General Reaction Mechanism for Carbapenem Hydrolysis by Mononuclear and Binuclear Metallo- β -Lactamases. *Nat Commun* **2017**, *8* (1), 538. <https://doi.org/10.1038/s41467-017-00601-9>.
- (148) Zafaralla, G.; Mobashery, S. Facilitation of the .DELTA.2 .Fwdarw. .DELTA.1 Pyrroline Tautomerization of Carbapenem Antibiotics by the Highly Conserved Arginine-244 of Class A .Beta.-Lactamases during the Course of Turnover. *J. Am. Chem. Soc.* **1992**, *114* (4), 1505–1506. <https://doi.org/10.1021/ja00030a070>.
- (149) Kalp, M.; Carey, P. R. Carbapenems and SHV-1 β -Lactamase Form Different Acyl-Enzyme Populations in Crystals and Solution. *Biochemistry* **2008**, *47* (45), 11830–11837. <https://doi.org/10.1021/bi800833u>.
- (150) Furey, I. M.; Mehta, S. C.; Sankaran, B.; Hu, L.; Prasad, B. V. V.; Palzkill, T. 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.* **2021**, *296*, 100799. <https://doi.org/10.1016/j.jbc.2021.100799>.
- (151) Smith, C. A.; Frase, H.; Toth, M.; Kumarasiri, M.; Wiafe, K.; Munoz, J.; Mobashery, S.; Vakulenko, S. B. Structural Basis for Progression toward the Carbapenemase Activity in the GES Family of β -Lactamases. *J. Am. Chem. Soc.* **2012**, *134* (48), 19512–19515. <https://doi.org/10.1021/ja308197j>.
- (152) Smith, C. A.; Nossoni, Z.; Toth, M.; Stewart, N. K.; Frase, H.; Vakulenko, S. B. Role of the Conserved Disulfide Bridge in Class A Carbapenemases. *J. Biol. Chem.* **2016**, *291* (42), 22196–22206. <https://doi.org/10.1074/jbc.M116.749648>.
- (153) Reed, A. E.; Weinstock, R. B.; Weinhold, F. Natural Population Analysis. *J. Chem. Phys.* **1985**, *83* (2), 735–746. <https://doi.org/10.1063/1.449486>.
- (154) Frase, H.; Shi, Q.; Testero, S. A.; Mobashery, S.; Vakulenko, S. B. Mechanistic Basis for the Emergence of Catalytic Competence against Carbapenem Antibiotics by the GES Family of β -Lactamases. *J. Biol. Chem.* **2009**, *284* (43), 29509–29513. <https://doi.org/10.1074/jbc.M109.011262>.

- (155) McInnes, L.; Healy, J.; Saul, N.; Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* **2018**, *3* (29), 861. <https://doi.org/10.21105/joss.00861>.
- (156) Pulay, P.; Fogarasi, G.; Pang, F.; Boggs, J. E. Systematic Ab Initio Gradient Calculation of Molecular Geometries, Force Constants, and Dipole Moment Derivatives. *J. Am. Chem. Soc.* **1979**, *101* (10), 2550–2560. <https://doi.org/10.1021/ja00504a009>.
- (157) Fogarasi, G.; Zhou, X.; Taylor, P. W.; Pulay, P. The Calculation of Ab Initio Molecular Geometries: Efficient Optimization by Natural Internal Coordinates and Empirical Correction by Offset Forces. *J. Am. Chem. Soc.* **1992**, *114* (21), 8191–8201. <https://doi.org/10.1021/ja00047a032>.

10. APPENDIX

Published and Submitted Manuscripts (First Author Only)

1. **Z. Song**, H. Zhou, H. Tian, X. Wang, P. Tao* Unraveling the Energetic Significance of Chemical Events in Enzyme Catalysis via Machine-Learning based Regression Approach *Commun. Chem.*, 2020, **3**, 134. DOI: 10.1038/s42004-020-00379-w
2. **Z. Song**, F. Trozzi, T. Palzkill, P. Tao* QM/MM Modeling of Class A β -Lactamases Reveals Distinct Acylation Pathways for Ampicillin and Cefalexin *Org. Biomol. Chem.*, 2021, **19**, 9182-9189. DOI: 10.1039/D1OB01593A
3. **Z Song**, F. Trozzi, H. Tian, C. Yin, P. Tao* Mechanistic Insights into Enzyme Catalysis from Explaining Machine-Learned Quantum Mechanical and Molecular Mechanical Minimum Energy Pathways *ACS Phys. Chem. Au*, in press. DOI: 10.1021/acspphyschemau.2c00005
4. **Z Song**, P Tao* Graph-Learning Guided Mechanistic Insights into Imipenem Hydrolysis in GES Carbapenemases. submitted to *Electron. Struct.*

Natural Internal Coordinates for 2-aminothiazole with explanations

```

! Natural Internal Coordinates for 2-aminothiazole with explanations
  1  1  2  0  0  !  1  BOND  C1=N2
  1  2  3  0  0  !  2  BOND  N2-C3
  1  3  4  0  0  !  3  BOND  C3=C4
  1  4  5  0  0  !  4  BOND  C4-S5
  1  1  5  0  0  !  5  BOND  C1-S5
  1  1  6  0  0  !  6  BOND  C1-N6
  1  3  7  0  0  !  7  BOND  C3-H7
  1  4  8  0  0  !  8  BOND  C4-H8
  1  6  9  0  0  !  9  BOND  N6-H9
  1  6 10  0  0  ! 10  BOND  N6-H10
  2  5  1  2  0  ! 11  ANGL  S5-C1=N2      5-member ring a1
  2  1  2  3  0  ! 12  ANGL  C1=N2-C3      a2
  2  2  3  4  0  ! 13  ANGL  N2-C3=C4      a3
  2  3  4  5  0  ! 14  ANGL  C3=C4-S5      a4
  2  4  5  1  0  ! 15  ANGL  C4-S5-C1      a5
  4  5  1  2  3  ! 16  DIHE  S5-C1=N2-C3      t1
  4  1  2  3  4  ! 17  DIHE  C1=N2-C3=C4      t2
  4  2  3  4  5  ! 18  DIHE  N2-C3=C4-S5      t3
  4  3  4  5  1  ! 19  DIHE  C3=C4-S5-C1      t4
  4  4  5  1  2  ! 20  DIHE  C4-S5-C1=N2      t5
  2  2  1  6  0  ! 21  ANGL  N2=C1-N6      C1  b1
  2  5  1  6  0  ! 22  ANGL  S5-C1-N6      b2
  3  6  2  5  1  ! 23  WWAG  N6-C1(=N2)-S5  N6  wagging on N2-C1-S5 plane
  2  2  3  7  0  ! 24  ANGL  N2-C3-H7      C3  b1
  2  4  3  7  0  ! 25  ANGL  C4=C3-H7      b2
  3  7  4  2  3  ! 26  WWAG  H7-C3(-N2)-C4  H7  wagging on N2-C3=C4 plane
  2  3  4  8  0  ! 27  ANGL  C3=C4-H8      C4  b1
  2  5  4  8  0  ! 28  ANGL  S5-C4-H8      b2
  3  8  5  3  4  ! 29  WWAG  H8-C4(-S5)=C3  H8  wagging on S5-C4=C3 plane
  2  9  6 10  0  ! 30  ANGL  H9-N6-H10      N6  a
  2  1  6  9  0  ! 31  ANGL  C1-N6-H9      b1
  2  1  6 10  0  ! 32  ANGL  C1-N6-H10     b2
  3  1  9 10  6  ! 33  WWAG  C1-N6(-H9)-H19  C1  wagging on NH2 plane
  4  2  1  6  9  ! 34  DIHE  N2-C1-N6-H9    dihedral angles of NH2 rolling

! U-matrix that transforms internal coordinates into NICs:
  1  1  1.
  2  2  1.
  3  3  1.
  4  4  1.
  5  5  1.
  6  6  1.
  7  7  1.
  8  8  1.
  9  9  1.
 10 10  1.
 11 11  1.  11 12 -0.81  11 13  0.31  11 14  0.31  11 15 -0.81
 12 12 -1.12  12 13  1.81  12 14 -1.81  12 15  1.12
 13 16  0.31  13 17 -0.81  13 18  1.  13 19 -0.81  13 20  0.31
 14 16 -1.81  14 17 -1.12  14 19  1.12  14 20  1.81
 15 21  1.  15 22 -1.
 16 23  1.
 17 24  1.  17 25 -1.
 18 26  1.
 19 27  1.  19 28 -1.
 20 29  1.
 21 30  2.  21 31 -1.  21 32 -1.
 22 31  1.  22 32 -1.
 23 33  1.
 24 34  1.

```

Force Field for 2-Aminothiazole Compatible with CHARMM36/CGenFF

```

* Toppar stream file for 2-aminothiazole
* Zilin Song, 13 March 2019
*

read rtf card append
* Topologies for 2-aminothiazole
* Zilin Song, 13 March 2019
*

36 1

! Note: This FF has not been extensively tested nor peer-reviewed.

RESI frg5          0.000
GROUP             ! FINAL      R1      INIT      PENALTY
ATOM C1           CG2R53      0.625 !    0.621    0.302      78.615
ATOM N2           NG2R50     -0.440 !   -0.440   -0.620      14.505
ATOM C3           CG2R51     -0.150 !   -0.146    0.209       2.500
ATOM C4           CG2R51     -0.655 !   -0.659   -0.185       2.500
ATOM S5           SG2R50      0.470 !    0.470   -0.053      40.269
ATOM N6           NG321     -0.483 !   -0.479   -0.670      72.686
ATOM H7           HGR52      0.130 !    0.130    0.130       0.000
ATOM H8           HGR52      0.175 !    0.175    0.177       0.000
ATOM H9           HGPAM2     0.164 !    0.164    0.355       9.083
ATOM H10          HGPAM2     0.164 !    0.164    0.355       9.083

BOND  C1  S5  C1  N6  C1  N2  N2  C3
BOND  C3  C4  C3  H7  C4  S5  C4  H8
BOND  N6  H9  N6  H10
IMPR  C1  N2  N6  S5

END

read param card flex append
* Parameters for 2-aminothiazole
* Zilin Song, 13 March 2019
*

BONDS
CG2R53 NG321  370.00    1.3730 ! 330.00    55

ANGLES
NG2R50 CG2R53 NG321      45.00    127.55 ! 45.80      39
NG321  CG2R53 SG2R50      58.00    116.60 ! 25.00      70
CG2R53 NG321  HGPAM2      68.00    117.55 ! 45.00      19

DIHEDRALS
NG321 CG2R53 NG2R50 CG2R51  7.0000  2    180.00 ! 3.0000      70
NG2R50 CG2R53 NG321 HGPAM2  2.5000  2    180.00 ! 0.3200     140.5
SG2R50 CG2R53 NG321 HGPAM2  2.3000  2    180.00 ! 0.3200     224.5
NG321 CG2R53 SG2R50 CG2R51  5.0000  2    180.00 ! 4.0000      70

IMPROPER
CG2R53 NG2R50 NG321 SG2R50  65.000  0      0.00 ! 45.00      209

END

```

Force Field for Cephalothin Compatible with CHARMM36/CGenFF

```
* Force field for cephalothin molecule.
* For use with CGenFF v4.0
* Zilin Song, 21 May 2019
*
```

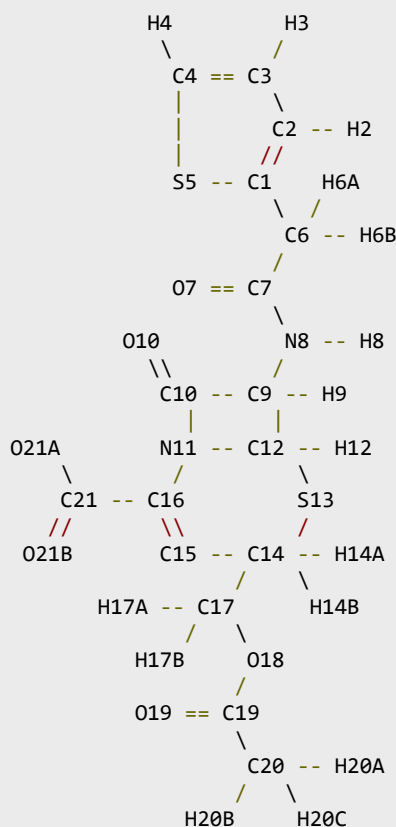
```
READ rtf CARD APPEND
```

```
* Topologies based on CGenFF v4.0
* Zilin Song, 21 May 2019
*
```

```
36 1
```

```
! Note: This FF has not been extensively tested nor peer-reviewed.
```

```
RESI INN      -1.000      ! atom_no
GROUP          ! frg1
ATOM C1        CG2R51    -0.045 ! 1
ATOM C2        CG2R51    -0.210 ! 2
ATOM C3        CG2R51    -0.225 ! 3
ATOM C4        CG2R51    -0.085 ! 4
ATOM S5        SG2R50    -0.015 ! 5
ATOM H2        HGR51      0.160 ! 6
ATOM H3        HGR51      0.170 ! 7
ATOM H4        HGR52      0.180 ! 8
GROUP          ! frg2
ATOM C6        CG321     -0.110 ! 9
ATOM C7        CG201      0.550 ! 10
ATOM N8        NG2S1     -0.535 ! 11
ATOM O7        OG2D1     -0.545 ! 12
ATOM H6A       HGA2       0.090 ! 13
ATOM H6B       HGA2       0.090 ! 14
ATOM H8        HGP1       0.310 ! 15
GROUP          ! frg3
ATOM C9        CG3C41     0.300 ! 16
ATOM C10       CG2R53     0.290 ! 17
ATOM N11       NG2R43    -0.330 ! 18
ATOM C12       CG3RC1     0.070 ! 19
ATOM S13       SG311     0.065 ! 20
ATOM C14       CG321     -0.065 ! 21
ATOM C15       CG2DC1    -0.110 ! 22(-)
ATOM C16       CG2D10    -0.160 ! 23
ATOM H9        HGA1       0.090 ! 24
ATOM O10       OG2D1     -0.470 ! 25
ATOM H12       HGA1       0.090 ! 26
ATOM H14A      HGA2       0.090 ! 27
ATOM H14B      HGA2       0.090 ! 28
GROUP          ! frg4
ATOM C17       CG321     0.135 ! 29
ATOM O18       OG302     -0.490 ! 30
ATOM C19       CG202      0.890 ! 31
ATOM C20       CG331     -0.310 ! 32
ATOM H17A      HGA2       0.090 ! 33
ATOM H17B      HGA2       0.090 ! 34
ATOM O19       OG2D1     -0.615 ! 35
ATOM H20A      HGA3       0.090 ! 36
ATOM H20B      HGA3       0.090 ! 37
ATOM H20C      HGA3       0.090 ! 38
GROUP          ! -COO(-)
ATOM C21       CG203      0.730 ! 39
ATOM O21A      OG2D2     -0.760 ! 40
ATOM O21B      OG2D2     -0.760 ! 41
```



```
BOND  C1  S5  C1  C6  C2  H2  C2  C3
```

```

BOND  C3  H3  C4  H4  C4  S5  C6  H6A
BOND  C6  H6B  C6  C7  C7  N8  N8  H8
BOND  N8  C9  C9  H9  C9  C10  C9  C12
BOND  C10  N11  N11  C12  N11  C16  C12  H12
BOND  S13  C14  C12  S13  C14  C15  C14  H14A
BOND  C15  C17  C14  H14B  C16  C21  C17  H17A
BOND  C17  H17B  C17  O18  O18  C19  C19  C20
BOND  C19  O19  C20  H20A  C20  H20B  C20  H20C
BOND  C21  O21A  C21  O21B
DOUBLE C1  C2
DOUBLE C3  C4
DOUBLE C7  O7
DOUBLE C10  O10
DOUBLE C15  C16
IMPR  C7  C6  N8  O7
IMPR  C10  C9  N11  O10
IMPR  C16  C15  C21  N11
IMPR  C19  C20  O18  O19
IMPR  C21  O21A  O21B  C16
ACCEPTOR O10  C10
ACCEPTOR O21A  C21
ACCEPTOR O21B  C21
ACCEPTOR O19  C19
ACCEPTOR O7  C7
! ICs created based on CHARMM optimized structure.
IC  C6  C1  C2  C3  1.5000  130.00  180.00  107.20  1.3600
IC  C1  C2  C3  C4  1.3600  107.20  0.00  107.20  1.3600
IC  C2  C1  C6  C7  1.3600  130.00  -120.00  112.00  1.4900
IC  C1  C6  C7  N8  1.5000  112.00  180.00  116.50  1.3450
IC  C6  C7  N8  C9  1.4900  116.50  180.00  123.50  1.4550
IC  C7  N8  C9  C12  1.3450  123.50  120.00  101.00  1.5400
IC  N8  C9  C10  N11  1.4550  101.00  -120.00  104.50  1.3800
IC  C9  C12  S13  C14  1.5400  110.20  120.00  98.00  1.8180
IC  C12  S13  C14  C15  1.7930  98.00  0.00  111.82  1.5020
IC  C14  C15  C17  O18  1.5020  122.00  -60.00  109.00  1.4400
IC  C15  C17  O18  C19  1.5020  109.00  180.00  109.60  1.3340
IC  C17  O18  C19  O19  1.4400  109.60  0.00  125.90  1.2200
IC  O18  C19  C20  H20A  1.3340  109.00  180.00  109.50  1.1110
IC  N11  C16  C21  O21A  1.3860  115.50  180.00  111.00  1.2600
IC  S5  C6  *C1  C2  1.7300  125.00  180.00  130.00  1.3600
IC  C3  C1  *C2  H2  1.3600  107.20  180.00  126.40  1.0800
IC  C4  C2  *C3  H3  1.3600  107.20  180.00  126.40  1.0800
IC  S5  C3  *C4  H4  1.7300  109.00  180.00  130.00  1.0830
IC  C7  C1  *C6  H6A  1.4900  112.00  120.00  109.50  1.1110
IC  H6A  C1  *C6  H6B  1.1110  109.50  -120.00  109.50  1.1110
IC  N8  C6  *C7  O7  1.3450  116.50  180.00  121.00  1.2300
IC  C9  C7  *N8  H8  1.4550  123.50  180.00  123.00  0.9970
IC  C12  N8  *C9  C10  1.5400  101.00  120.00  101.00  1.5600
IC  C10  N8  *C9  H9  1.5600  101.00  -120.00  102.00  1.0930
IC  N11  C9  *C10  O10  1.3800  104.50  180.00  135.70  1.2350
IC  N11  C9  *C12  H12  1.4500  104.50  -120.00  110.20  1.1110
IC  N11  C9  *C12  S13  1.4500  104.50  -120.00  110.20  1.7930
IC  C15  S13  *C14  H14A  1.5020  111.82  120.00  111.30  1.1110
IC  C15  S13  *C14  H14B  1.5020  111.82  -120.00  111.30  1.1110
IC  C12  C10  *N11  C16  1.4500  111.50  180.00  113.00  1.3860
IC  C16  C14  *C15  C17  1.3400  126.50  180.00  122.00  1.5020
IC  O18  C15  *C17  H17A  1.4400  109.00  120.00  111.50  1.1110
IC  H17A  C15  *C17  H17B  1.1110  111.50  -120.00  111.50  1.1110
IC  O19  O18  *C19  C20  1.2200  125.90  180.00  109.00  1.5220
IC  H20A  C19  *C20  H20B  1.1110  109.50  120.00  109.50  1.1110
IC  H20A  C19  *C20  H20C  1.1110  109.50  -120.00  109.50  1.1110
IC  C15  N11  *C16  C21  1.3400  111.00  180.00  115.50  1.4890
IC  O21A  C16  *C21  O21B  1.2600  111.00  180.00  111.00  1.2600
END

```

READ param CARD FLEX APPEnd

* Parameters parameterized basing on CGenFF v4.0

* Zilin Song, 21 May 2019

*

BONDS

CG3C41	NG2S1	320.00	1.4550
CG2D10	CG2O3	440.00	1.4890
CG2D10	NG2R43	475.00	1.3860
CG3C41	CG3RC1	270.00	1.5400
CG3C41	HGA1	348.00	1.0930
CG3RC1	NG2R43	245.00	1.4500
CG3RC1	SG311	162.00	1.7930

ANGLES

CG2R53	CG3C41	NG2S1	95.00	101.00		
CG3RC1	CG3C41	NG2S1	95.00	101.00		
NG2S1	CG3C41	HGA1	72.00	102.00		
CG2O1	NG2S1	CG3C41	50.00	123.50		
CG3C41	NG2S1	HGP1	35.00	119.50		
CG321	CG2DC1	CG321	48.00	122.00		
CG2DC1	CG321	OG302	20.00	109.00		
CG2DC1	CG2D10	CG2O3	48.00	132.50		
CG2DC1	CG2D10	NG2R43	60.00	111.00		
CG2O3	CG2D10	NG2R43	95.00	115.50		
CG2D10	CG2O3	OG2D2	40.00	111.00	50.00	2.35300
CG321	CG2R51	SG2R50	25.00	125.00		
CG2O1	CG321	CG2R51	51.80	112.00		
CG321	SG311	CG3RC1	63.00	98.00		
CG3C41	CG3RC1	NG2R43	90.00	104.50		
CG3C41	CG3RC1	SG311	95.00	110.20		
CG3C41	CG3RC1	HGA1	46.00	110.20		
NG2R43	CG3RC1	SG311	30.00	113.80		
NG2R43	CG3RC1	HGA1	51.00	107.00		
SG311	CG3RC1	HGA1	69.00	105.00		
CG2D10	NG2R43	CG2R53	54.00	113.00		
CG2D10	NG2R43	CG3RC1	104.00	119.00		
CG2R53	NG2R43	CG3RC1	85.00	111.50		
CG3RC1	CG3C41	HGA1	46.00	110.50		
CG2DC1	CG321	SG311	65.00	111.82		
CG2R53	CG3C41	CG3RC1	90.00	106.00		
CG2R53	CG3C41	HGA1	46.00	112.30		

DIHEDRALS

CG321	CG2R51	SG2R50	CG2R51	4.0000	2	180.00
CG2R51	CG2R51	CG321	CG2O1	0.3000	1	0.00
CG2R51	CG2R51	CG321	CG2O1	1.2000	2	180.00
SG2R50	CG2R51	CG321	CG2O1	1.5000	3	180.00
SG2R50	CG2R51	CG321	HGA2	1.2000	3	0.00
NG2S1	CG2O1	CG321	CG2R51	3.2000	1	0.00
OG2D1	CG2O1	CG321	CG2R51	1.8000	1	180.00
CG321	CG2O1	NG2S1	CG3C41	1.6000	1	0.00
CG321	CG2O1	NG2S1	CG3C41	2.5000	2	180.00
OG2D1	CG2O1	NG2S1	CG3C41	2.5000	2	180.00
CG2R53	CG3C41	NG2S1	CG2O1	4.0000	3	180.00
NG2R43	CG2R53	CG3C41	NG2S1	3.0000	3	0.00
OG2D1	CG2R53	CG3C41	NG2S1	4.0000	3	0.00
CG3RC1	CG3C41	NG2S1	CG2O1	0.5000	3	180.00
NG2S1	CG3C41	CG3RC1	NG2R43	0.5000	3	0.00
NG2S1	CG3C41	CG3RC1	SG311	3.4000	3	0.00
NG2S1	CG3C41	CG3RC1	HGA1	4.0000	3	180.00
HGA1	CG3C41	NG2S1	CG2O1	1.0000	3	0.00
CG2R53	CG3C41	NG2S1	HGP1	4.0000	1	0.00
CG3RC1	CG3C41	NG2S1	HGP1	3.3000	1	180.00
HGA1	CG3C41	NG2S1	HGP1	0.0000	3	0.00

CG203	CG2D10	CG2DC1	CG321	0.5600	1	180.00
CG203	CG2D10	CG2DC1	CG321	7.0000	2	180.00
NG2R43	CG2D10	CG2DC1	CG321	2.5000	2	180.00
CG2DC1	CG2D10	CG203	OG2D2	1.3000	2	180.00
NG2R43	CG2D10	CG203	OG2D2	0.3000	2	180.00
CG203	CG2D10	NG2R43	CG2R53	1.6000	1	0.00
CG203	CG2D10	NG2R43	CG2R53	2.5000	2	180.00
CG203	CG2D10	NG2R43	CG3RC1	1.6000	1	0.00
CG203	CG2D10	NG2R43	CG3RC1	4.0000	2	180.00
CG2D10	CG2DC1	CG321	OG302	0.9000	3	180.00
CG321	CG2DC1	CG321	OG302	0.1000	3	0.00
CG321	CG2DC1	CG321	SG311	5.4000	3	180.00
CG2DC1	CG321	OG302	CG202	0.5000	3	180.00
CG321	CG2DC1	CG321	HGA2	0.1900	3	0.00
NG2R43	CG2R53	CG3C41	CG3RC1	3.0000	3	0.00
CG2D10	CG2DC1	CG321	SG311	0.5000	2	0.00
CG2D10	CG2DC1	CG321	SG311	0.3000	3	0.00
CG2DC1	CG2D10	NG2R43	CG2R53	2.0000	1	0.00
CG2DC1	CG2D10	NG2R43	CG2R53	3.7000	2	180.00
CG2DC1	CG2D10	NG2R43	CG3RC1	1.2000	1	180.00
CG2DC1	CG2D10	NG2R43	CG3RC1	6.5000	2	180.00
NG2R43	CG2R53	CG3C41	HGA1	0.5700	3	0.00
OG2D1	CG2R53	CG3C41	CG3RC1	0.5700	3	0.00
OG2D1	CG2R53	CG3C41	HGA1	0.5700	3	0.00
CG3C41	CG2R53	NG2R43	CG2D10	1.5000	2	180.00
CG3C41	CG2R53	NG2R43	CG3RC1	1.5000	2	180.00
OG2D1	CG2R53	NG2R43	CG2D10	1.8500	2	0.00
OG2D1	CG2R53	NG2R43	CG3RC1	2.5000	2	180.00
CG2DC1	CG321	SG311	CG3RC1	2.5000	1	180.00
CG2DC1	CG321	SG311	CG3RC1	1.5000	3	180.00
HGA2	CG321	SG311	CG3RC1	0.0000	3	0.00
CG2R53	CG3C41	CG3RC1	NG2R43	3.0000	3	0.00
CG2R53	CG3C41	CG3RC1	SG311	1.6500	3	180.00
CG2R53	CG3C41	CG3RC1	HGA1	0.5700	3	0.00
HGA1	CG3C41	CG3RC1	NG2R43	3.0000	3	0.00
HGA1	CG3C41	CG3RC1	SG311	1.6500	3	0.00
HGA1	CG3C41	CG3RC1	HGA1	0.5700	3	0.00
CG3C41	CG3RC1	NG2R43	CG2D10	3.0000	3	0.00
CG3C41	CG3RC1	NG2R43	CG2R53	3.0000	3	0.00
SG311	CG3RC1	NG2R43	CG2D10	3.7000	3	0.00
SG311	CG3RC1	NG2R43	CG2R53	0.8500	3	0.00
HGA1	CG3RC1	NG2R43	CG2D10	2.0000	3	0.00
HGA1	CG3RC1	NG2R43	CG2R53	0.5700	3	0.00
CG3C41	CG3RC1	SG311	CG321	1.5000	1	180.00
CG3C41	CG3RC1	SG311	CG321	1.0000	3	180.00
NG2R43	CG3RC1	SG311	CG321	5.5000	1	0.00
NG2R43	CG3RC1	SG311	CG321	1.0000	3	0.00
HGA1	CG3RC1	SG311	CG321	0.0000	3	0.00
IMPROPERS						
CG2D10	CG2DC1	CG203	NG2R43	72.0000	0	0.00
CG203	OG2D2	OG2D2	CG2D10	96.0000	0	0.00
CG202	CG331	OG302	OG2D1	62.0000	0	0.00

END