VARIANT EFFECT PREDICTION IN TEM-1 β-LACTAMASES BY INCORPORATING DYNAMICS-BASED FEATURES

Thesis submitted to the SASTRA Deemed to be University
in partial fulfilment of the requirement
for the award of the degree of

M.Sc. Bioinformatics

Submitted by

NAME: S ROSHAN KUMAR
(Reg. No.: 124121010)

June 2024

SCHOOL OF CHEMICAL AND BIOTECHNOLOGY

THANJAVUR, TAMIL NADU, INDIA – 613 401

SASTRA
ENGINEERING · MANAGEMENT · LAW · SCIENCES · HUMANITIES · EDUCATION
DEEMED TO BE UNIVERSITY
(U/S 3 of the UGC Act,1956)
THINK MERIT | THINK TRANSPARENCY | THINK SASTRA

SCHOOL OF CHEMICAL AND BIOTECHNOLOGY

THANJAVUR – 613 401

Bonafide Certificate

This is to certify that the thesis titled "**VARIANT EFFECT PREDICTION OF TEM-1 β-LACATAMASES BY INCORPORATING DYNAMICS-BASED FEATURES**" submitted as a requirement for the award of the degree of M.Sc. Bioinformatics to the SASTRA Deemed to be University, is a bonafide record of the work done by **Mr. ROSHAN KUMAR S, Reg no: 124121010** during the final semester of the academic year 2023-2024, in the School of Chemical and biotechnology, under my supervision. This thesis has not formed the basis for the award of any degree, diploma, associate ship, fellowship or other similar title to any candidate of any university.

Signature of Project Supervisor **:**
Name with Affiliation **:** Dr. Vigneshwar Ramakrishnan, Professor
Date **:**

Project *Viva voce* held on _____

Examiner 1 **Examiner 2** **Examiner 3** **Examiner 4**

SASTRA
ENGINEERING · MANAGEMENT · LAW · SCIENCES · HUMANITIES · EDUCATION
DEEMED TO BE UNIVERSITY
(U/S 3 of the UGC Act,1956)
THINK MERIT | THINK TRANSPARENCY | THINK SASTRA

SCHOOL OF CHEMICAL AND BIOTECHNOLOGY

THANJAVUR – 613 401

Declaration

I declare that the thesis titled "**Variant effect prediction of TEM-1 β-lactamase by incorporating dynamics-based features"** submitted by me is an original work done by me under the guidance of **Dr. Vigneshwar Ramakrishnan, Professor, School of Chemical and Biotechnology, SASTRA Deemed to be University, Thanjavur** during the final semester of the academic year 2023-2024, in the **School of Chemical and Biotechnology**. The work is original and wherever I have used materials from other sources, I have given due credit and cited them in the text of the thesis. This thesis has not formed the basis for the award of any degree, diploma, associate-ship, fellowship or other similar title to any candidate of any University.

Signature of the candidate(s) **:**

Name of the candidate(s) **: Roshan Kumar S**

Date **:**

Acknowledgements

I first thank my guide **Dr.Vigneshwar Ramakrishnan,** Professor, SCBT-SASTRA Deemed to be University whose expertise was valuable in formulating my research work. He consistently steered me in the right direction whenever i needed it. I thank the review panel for their validation of my project and their valuable suggestions.

I would like to thank **Dr.M. Sathish kumar** for his support and guidance.

I would like to thank **SASTRA DEEMED UNIVERSITY** for its infrastructure and facilities.

Table of Contents

List of Figures

List of Tables

ABBREVIATIONS

REMD	Replica Exchange Molecular Dynamics
SVM	Support Vector Machine
MCC	Matthews Correlation Coefficient
AMR	Antimicrobial Resistance
TEM-1	TEM-1 Beta lactamase
φ	Phi (Dihedral angle)
ψ	Psi (Dihedral angle)
PDB	Protein Data Bank
WT	Wild Type

Abstract

TEM-1 beta-lactamases, vital enzymes produced by bacteria, wield significant impact in conferring resistance against beta-lactam antibiotics by catalyzing the hydrolysis of the beta-lactam ring, thereby rendering these antibiotics ineffective. The emergence of point mutations within these enzymes leads to variants that robustly resist the effects of drugs like amoxicillin, a commonly prescribed beta-lactam antibiotic. In this study, we explore into the structural intricacies of TEM-1 beta-lactamases, focusing particularly on the influence of dihedral features in conferring resistance. Notably, mutations such as Q39K, E240K, and N175S denote resistance, while E104K, A237T, and G238S signify susceptibility to amoxicillin. Leveraging dynamic features through Replica Exchange Molecular Dynamics (REMD), we meticulously dissected the structural disparities between resistant and susceptible variants.

Our investigation aims at elucidating the pivotal amino acids that underline drug resistance mechanisms, thereby discerning mutations crucial for resistance from those that are not. Through machine learning analyses employing SVM, Linear Regression, and Random Forest models, we achieved unparalleled accuracy, with Random Forest yielding an impressive accuracy of 0.95. This high predictive performance underscores the robustness of our approach in discerning key structural determinants of drug resistance in TEM-1 beta-lactamases.

By shedding light on these structural nuances, our research offers profound insights into the development of effective therapeutic strategies against beta-lactamase-resistant bacteria. This comprehensive understanding of the molecular basis of resistance paves the way for the design and optimization of novel antibiotics to combat the growing threat of antimicrobial resistance.

CHAPTER 1

INTRODUCTION

1.1 Anti-Microbial-Resistance:

Antimicrobial resistance (AMR) is one of the top global development threats. The misuse of antimicrobials are key drivers for the development of drug-resistant pathogens. Antibiotics, which are small-molecule peptides, widely exist in nature and form components of the innate immunity of almost all living organisms. Antimicrobial resistance (AMR) has become one of the foremost public health challenges of the 21st century, jeopardizing the effective prevention and treatment of a growing array of infections caused by bacteria, parasites, viruses, and fungi that no longer respond to standard medications. The urgency of AMR is particularly pronounced in the context of bacterial antibiotic resistance. Over the past few decades, bacteria responsible for common and severe infections have developed resistance to every new antibiotic introduced to the market, impairing our ability to treat infectious diseases and threatening many other advancements in health and medicine.

Antibiotics play a crucial role in resisting foreign invading microorganisms. They are ubiquitous agents in contemporary healthcare, though historically, this was not the case. From antiquity, individuals endeavored to remedy infections utilizing dyes, molds, and even heavy metals **[5]**. The rampant overutilization and improper deployment of antibiotics are accelerating antimicrobial resistance among pathogenic microorganisms, constituting an escalating global public health predicament. Elevated resistance precipitates severe infections, heightened complications, protracted hospitalizations, and increased mortality rates. Moreover, AMR exerts a profound impact on national economies and health systems, impairing the productivity of patients and caregivers due to extended hospitalizations and the concomitant substantial economic burdens **[2].**

The World Health Organization (WHO) has identified AMR as a critical issue requiring urgent action, emphasizing the need for global collaboration to combat this threat **[3]**. Effective strategies to address AMR include optimizing the use of antibiotics, improving infection prevention and control measures, and investing in research and development of new antimicrobial agents and diagnostic tools **[1]**. The rapid spread of resistant pathogens underscores the importance of monitoring and surveillance, as well as the implementation of

stewardship programs to promote the rational use of antibiotics **[4]**. Addressing AMR is essential to safeguarding the advances of modern medicine and ensuring the continued efficacy of antibiotics for future generations.

In the field of medicine, antibiotics play a pivotal role in combating bacterial infections, saving countless lives since their discovery. Understanding the diverse classes of antibiotics and their mechanisms of action is crucial for effective treatment strategies. Beta-lactam antibiotics, characterized by the presence of a beta-lactam ring, constitute one of the most widely used classes. They include penicillins, cephalosporins, carbapenems, and monobactams, each with unique variations in structure and spectrum of activity **[28].**

Fig 1.1: Antimicrobial resistance **[29],**

The varying effectiveness of antibiotics against different bacterial strains, highlighting the importance of susceptibility testing in guiding effective antibiotic therapy

1.2 β-Lactam antibiotics

Beta-lactam antibiotics represent one of the most prevalently prescribed pharmacological classes, with myriad clinical applications. Their introduction in the 1930s markedly transformed the battle against bacterial infectious diseases. Presently, the annual expenditure for these antibiotics is estimated to be around $15 billion USD, comprising 65% of the total antibiotics market **[6].** However, their extensive utilization is increasingly

undermined by the alarming phenomenon of antimicrobial resistance, a critical global health concern.

From a biochemical perspective, these drugs are characterized by a distinctive structural motif: a highly reactive four-membered ring containing three carbons and one nitrogen, known as the beta-lactam ring. This class encompasses:

- **Penicillins:**

These antibiotics, many of which bear the suffix -cillin, contain a core structure of 6-aminopenicillanic acid (comprising a beta-lactam ring fused with a thiazolidine ring) and various side chains. The penicillin group includes natural penicillins, beta-lactamase-resistant agents, aminopenicillins, carboxypenicillins, and ureidopenicillins. Amoxicillin, a widely utilized penicillin antibiotic, is effective against a variety of bacterial infections, commonly prescribed for respiratory tract infections, such as pneumonia, and dental abscesses. Furthermore, amoxicillin can be used in conjunction with other antibiotics and medications to treat gastric ulcers **[10].** Its mechanism of action involves the inhibition of bacterial cell wall synthesis, ultimately resulting in cell lysis and death. Despite its efficacy, the escalating prevalence of beta-lactamase-producing bacteria, which can hydrolyze the beta-lactam ring of amoxicillin, significantly threatens its continued effectiveness **[11].**

- **Cephalosporins:**

Cephalosporins possess a core structure composed of a 7-aminocephalosporanic acid nucleus and side chains incorporating 3,6-dihydro-2 H-1,3-thiazine rings. Traditionally, cephalosporins are categorized into five classes or generations, although this nomenclature is not universally endorsed **[7].**

- **Carbapenems:**

Carbapenems are characterized by a core structure comprising a carbapenem moiety fused to a beta-lactam ring, which affords substantial protection against a broad spectrum of beta-lactamases. Nonetheless, resistance to carbapenems is a pervasive problem, predominantly encountered among gram-negative pathogens (e.g., Klebsiella pneumoniae, Pseudomonas aeruginosa, and Acinetobacter baumannii), which produce various classes of beta-lactamases known as carbapenemases **[8].**

- **Monobactams:**

Monobactams are distinguished by the presence of a beta-lactam ring that exists independently, not fused to any other ring structure.

1.3 Mechanism of Action of Beta-Lactam Antibiotics

Beta-lactam antibiotics are bactericidal agents that disrupt bacterial cell wall formation by covalently binding to essential penicillin-binding proteins (PBPs). PBPs are enzymes involved in the terminal steps of peptidoglycan cross-linking in both Gram-negative and Gram-positive bacteria. Each bacterial species has a distinctive set of PBPs, ranging from three to eight enzymes per species **[16].** Mechanistically, beta-lactam antibiotics inhibit bacterial peptidoglycan transpeptidation. This process was initially described **[18]**, who noted the structural similarity of penicillin G to the terminal d-Ala-d-Ala dipeptide of nascent peptidoglycan in dividing bacterial cells. Beta-lactams bind to an active site serine in all functional PBPs, forming an inactive acyl-enzyme complex that slowly hydrolyzes the antibiotic, rendering it microbiologically inactive **[14].**

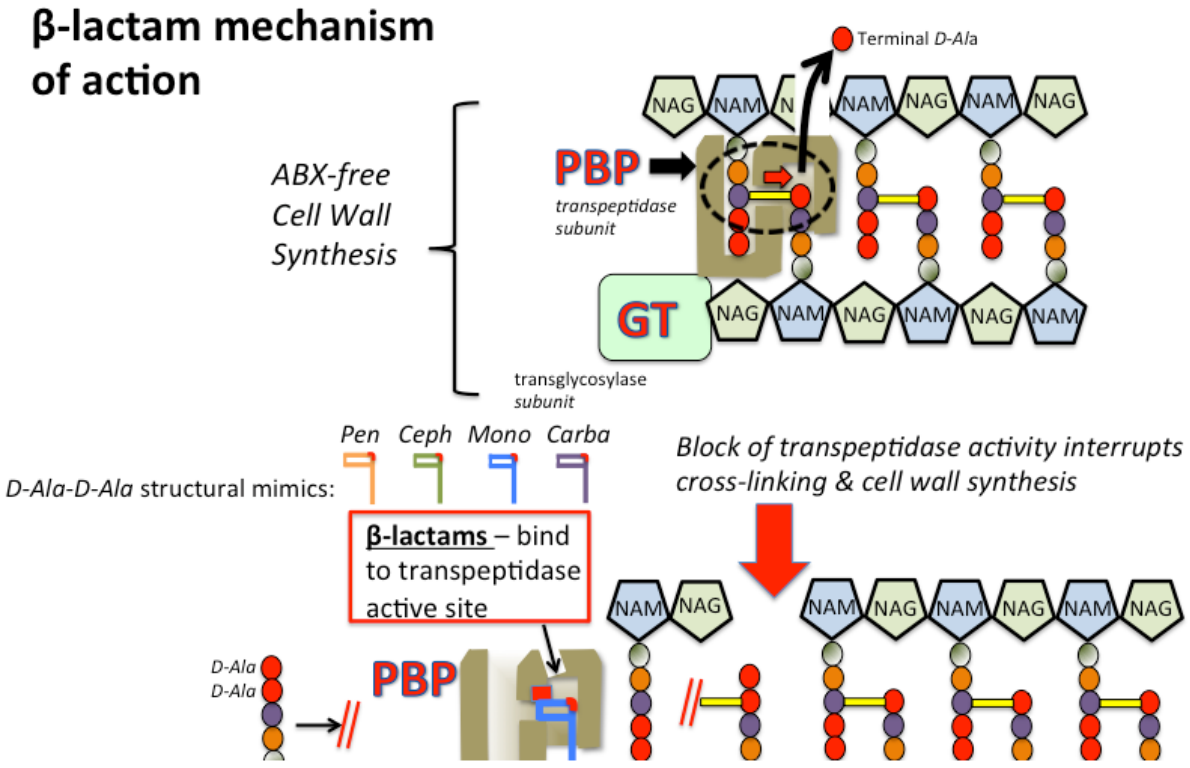

Fig 1.2: β-lactam mechanism of action **[18]**

Recent studies have shown that some beta-lactams, like ceftaroline, can bind to an allosteric site in PBP2a from Staphylococcus aureus, increasing the organism's sensitivity to the antibiotic **[17], [15]**. This allosteric binding induces conformational changes that enhance the drug's efficacy against resistant strains **[13]**.

The unique four-membered ring structure of beta-lactam antibiotics mimics the D-Ala-D-Ala peptide terminus, the natural substrate for transpeptidase activity during cell wall synthesis. This structural mimicry allows tight binding of the drugs to the transpeptidase active site, inhibiting cell wall synthesis and resulting in a weakened cell wall. During periods of cell growth, the high internal osmotic pressure in bacteria causes the compromised cell wall to become susceptible to lysis, leading to bacterial cell death **[12].**

In the realm of antibiotic therapy, a fundamental understanding of the chemical structures of β-lactam antibiotics is paramount. These antibiotics, characterized by the presence of a β-lactam ring, constitute a pivotal class in antimicrobial treatment strategies. Their diverse chemical structures underpin their distinct mechanisms of action, which are essential for combating bacterial infections effectively. Within this class, notable examples include penicillins, cephalosporins, carbapenems, and monobactams, each bearing unique variations in their chemical compositions and molecular arrangements. As such, delving into the intricate chemical structures of β-lactam antibiotics provides a solid foundation for comprehending their mechanisms of action and underscores their indispensable role in modern medicine.

1.4 Beta Lactam Inhibitors

These work primarily by inactivate the beta-lactam ring (especially in gram-negative bacteria). This group includes beta-lactamase inhibitors (clavulanic acid, sulbactam, and tazobactam) **[7]**.

Beta-lactamase enzymes, which confer antimicrobial resistance (AMR) to microorganisms, have been extensively studied. These enzymes hydrolyze lactam ring, rendering antibiotics inactive. Consequently, microorganisms producing these enzymes are resistant to beta-lactam antibiotics **[9].**

1.5 TEM-1 Beta lactamases

β-lactamase enzymes, crucial for conferring antimicrobial resistance (AMR) to microorganisms, have become a focal point of research in recent years. These enzymes play a pivotal role in rendering β-lactam antibiotics ineffective by hydrolyzing the β-lactam ring, a mechanism predominantly observed in Gram-negative bacteria such as Escherichia coli (E. coli). Among the diverse classes of β-lactamases, TEM-1 β-lactamases stand out as a significant contributor to bacterial resistance against a wide spectrum of β-lactam antibiotics.

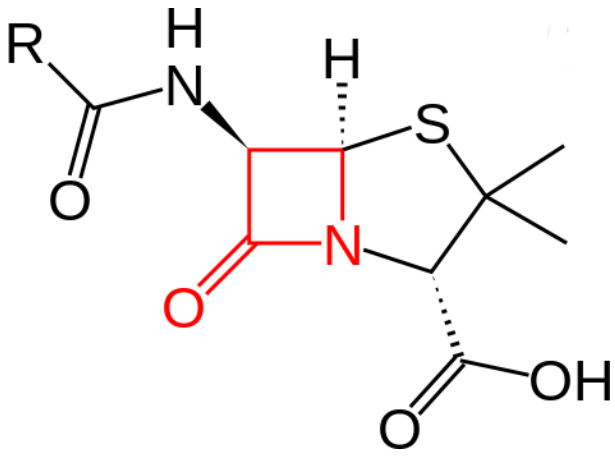

Fig 1.3: β-lactam chemical structure

The TEM-1 β-lactamase, emblematic of β-lactamase-mediated resistance, possesses a complex yet essential structure. At its core lies a catalytic domain adorned with conserved motifs crucial for its enzymatic activity **[19].** This domain harbors an active site responsible for binding and catalyzing the hydrolysis of β-lactam antibiotics. Structural elucidation has unveiled a classic serine-β-lactamase fold, characterized by the presence of a conserved serine residue within the active site, serving as the catalyst for the hydrolysis reaction.

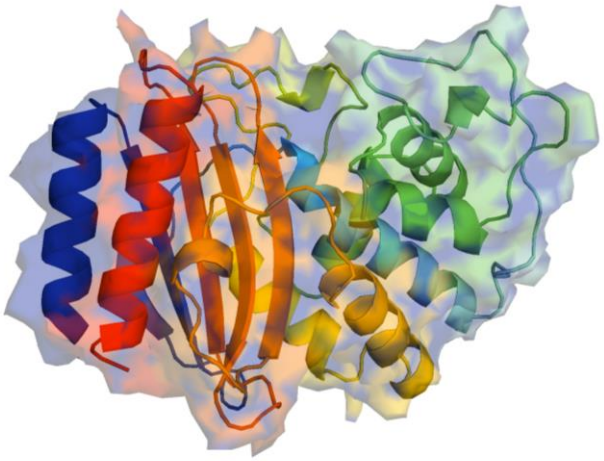

Fig 1.4 3D – Crystal structure of *E. coli* TEM-1 Beta lactamase (PDB 1BTL).

Furthermore, TEM-1 β-lactamases feature peripheral structural elements pivotal for substrate recognition and stabilization of the enzyme-substrate complex. These elements, including loops and helices surrounding the active site, undergo conformational changes upon substrate binding, facilitating efficient catalysis.

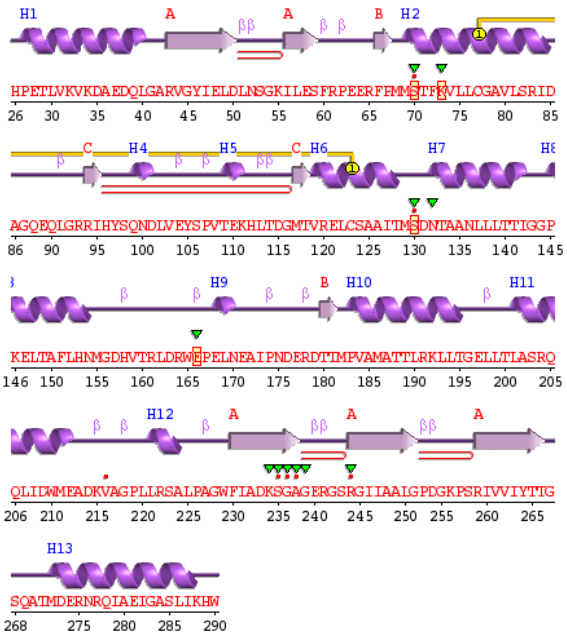

Fig 1.5: Catalytic sites in TEM-1 β-lactamases

Understanding the intricate structural features of TEM-1 β-lactamases is indispensable for devising effective strategies to combat antibiotic resistance. Insights gleaned from structural studies not only inform the design of novel β-lactamase inhibitors but also pave the way for the

development of innovative therapeutic approaches. Additionally, through meticulous analysis utilizing tools like PDBsum **[32]**, critical catalytic sites within TEM-1 β-lactamases have been identified, shedding light on their pivotal role in enzymatic function and offering invaluable guidance for future drug design and enzyme engineering endeavors.

1.6 Existing methods to capture variant effects

Predicting the effects of genetic variants is crucial in understanding their potential impact on gene function, protein structure, and disease association. Various computational tools and methodologies have been developed for this purpose. Here are some of the prominent methods:

1.6.1 Sequence-Based Approaches:

These tools assess variants based on sequence conservation and physicochemical properties.

- **PolyPhen-1:** Predicts the impact of amino acid substitutions using physical and comparative considerations of sequences.
- **PolyPhen-2:** Employs machine learning along with predictive features for improved accuracy **[20].**
- **SIFT (Sorting Intolerant From Tolerant):** Predicts functional impact based on sequence homology and physical properties **[26].**

1.6.2 Structure-Based Approaches:

These methods analyze protein structure to predict variant effects.

- **PredictSNP:**
  A consensus classifier integrating results from multiple tools for unified predictions **[21].**
- **MAPP (Multivariate Analysis of Protein Polymorphism) and PhD-SNP (Predictor of Human Deleterious Single Nucleotide Polymorphisms):**
  Assess the impact of non-synonymous polymorphisms using physicochemical variation and machine learning techniques **[23].**
- **SNAP (Screening for Non-Acceptable Polymorphisms):**
  Predicts functional effects using neural networks and evolutionary information **[22].**

1.6.3 Machine Learning Approaches:

These methods utilize machine learning algorithms for predictive modeling.

- **MutPred:**
  Predicts amino acid substitution impact and identifies molecular mechanisms involved **[25].**
- **DeepSEA (Deep Learning-based Sequence Analyzer):**
  Uses deep learning to predict non-coding variant effects **[31].**
- **EVE (Evolutionary model of Variant Effect):** A deep generative model predicting variant effects from protein sequence alignments **[27].**

1.6.4 Structural Modeling:

These methods involve modeling protein structure to assess variant effects.

- **FoldX:** Predicts mutation effects on stability, binding, and folding using empirical force fields **[24].**

By leveraging these diverse computational approaches, we aim to comprehensively understand the impact of genetic variants on protein function, thereby aiding in disease diagnosis and therapeutic development.

Lacuna:

However, the existing approaches fall short in predicting the specific phenotype—such as resistance or susceptibility—that results from a drug-induced mutation. These methods are primarily based on protein stability-based calculations which may or may not be directly related to resistance/susceptibility phenotypes. Hence, to overcome this limitation, we are employing machine learning techniques incorporating with Molecular Dynamics. By incorporating Replica Exchange Molecular Dynamics (REMD) simulations, we can capture the structural and dynamic properties of proteins, enabling more accurate predictions of how mutations influence antibiotic resistance.

CHAPTER 2
OBJECTIVE

The present study aims to evaluate the prediction of variant effect of TEM-1 beta lactamases by incorporating dynamics-based features

2.1 Specific objectives of the study

Existing sequence analysis, structural modeling, and machine learning approaches fall short in predicting whether a specific mutation will result in a resistant or susceptible phenotype. These methods do not adequately account for the dynamic nature of protein structures and their functional consequences. Therefore, there is a pressing need for an approach that can accurately determine which mutations confer resistance. Our objective is to address this gap by adopting a machine learning strategy that incorporates dynamics-based techniques. Specifically, we will use Replica Exchange Molecular Dynamics (REMD) simulations to capture the structural and dynamic properties of proteins. By integrating these simulations with machine learning, we aim to develop a robust predictive model that can distinguish between resistant and susceptible variants based on mutations. This innovative approach has the potential to significantly enhance our understanding of antibiotic resistance and inform the development of more effective therapeutic strategies.

CHAPTER-3

METHODOLOGY

3.1 TEM-1 beta-lactamase

The crystal structure of TEM-1 beta-lactamase at 1.8 angstroms resolution obtained from *Escherichia coli*, known as the wild type structure (PDB: 1BTL), served as the basis for our study.

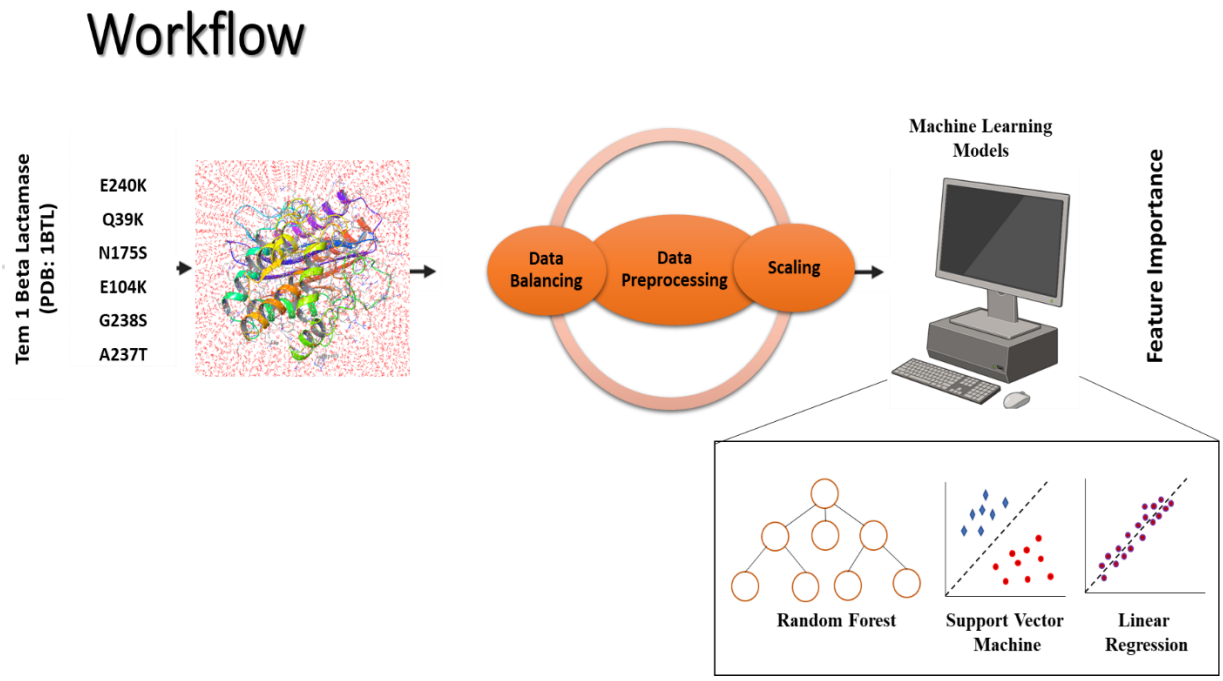

Fig 3.1: Workflow methodology

3.2 Dataset

Our study utilized variants of TEM-1 beta-lactamases and their corresponding Minimum Inhibitory Concentration (MIC) values from the dataset referenced in **[19].** This dataset includes MIC values for 990 variants, which were classified as resistant or susceptible based on their MIC relative to the wild type (WT).

- 54 variants with MIC values higher than WT were classified as Resistant.

- 616 variants with MIC values lower than WT were classified as Susceptible.
- 320 variants with MIC values equivalent to WT were considered to have the same resistance level as WT.

By experimental analysis, resistance is indicated by positive MIC values, while susceptibility is indicated by negative MIC values. The variants were selected based on their experimentally determined MIC values from the referenced dataset. Variants such as A237T, G238S, and E104K, which have significantly lower MIC values than the WT, were classified as susceptible. In contrast, variants like E240K, N175S, and Q39K, with higher MIC values, were classified as resistant. This classification enabled us to create a balanced and representative dataset for our analysis, facilitating the investigation of structural features associated with resistance and susceptibility.

By incorporating these MIC values, we ensured that our dataset accurately reflects the experimental resistance profiles of the TEM-1 beta-lactamase variants. This approach provides a robust basis for our machine learning models to discern the structural determinants underlying antibiotic resistance.

System	MIC value (log)	Category
WT	0	Susceptible
A237T	-2.321928	Susceptible
G238S	-2.321928	Susceptible
E104K	-0.666667	Susceptible
E240K	**0.5**	**Resistant**
N175S	**0.25**	**Resistant**
Q39K	**0.333333**	**Resistant**

Table 3.1: System choosed based on MIC for the study

3.3 Preaparation of variants

In the REMD simulations, six TEM-1 beta-lactamase variants were engineered via mutations introduced into the TEM-1 beta-lactamase structure (PDB ID: 1BTL) using the Schrödinger REMD module. Molecular interactions were modeled using the OPLS4 force field, with solvation employing the TIP3P water model. To maintain system neutrality, counter ions were incorporated. Each variant underwent 10 replicate simulations spanning

temperatures from 310 K to 328 K, each lasting 100 ns with recording intervals of 5 ps. From the resulting 20,000 frames, only the last 50 ns, or 10,000 frames, were utilized, extracting dihedral angles as features. This extensive temperature range and replicates ensured comprehensive sampling of conformational space, facilitating accurate prediction of variant effects. Consequently, this methodology allowed for a robust analysis of structural and dynamic properties under physiological conditions for the TEM-1 variants.

3.4 Acceptance ratio

The acceptance ratio for temperature shifts in the REMD simulations was determined for the six TEM-1 variants and the wild type. According to best practices, an acceptance ratio between 0.2 and 0.3 is considered optimal **[33].**

3.5 REMD simulations

Replica Exchange Molecular Dynamics (REMD) simulations were conducted for each TEM-1 mutation, covering a temperature range from 310 to 328 Kelvin, as recommended by a virtual temperature generator. These simulations offered in-depth insights into the dynamic behavior of the mutated TEM-1 enzyme, highlighting the structural changes linked to resistance or susceptibility to amoxicillin. The results included dihedral angles (phi-psi angles), which were essential for understanding the conformational shifts in the enzyme.

REMD is a sophisticated molecular dynamics simulation method that enhances sampling efficiency in molecular systems, particularly those with complex energy landscapes and multiple conformational states. In REMD, several replicas of the system are run in parallel at different temperatures. Periodically, configuration exchanges between replicas are attempted based on the Metropolis criterion, ensuring that each system maintains the correct Boltzmann distribution at its respective temperature. This process allows higher temperature replicas to overcome energy barriers more readily, promoting the exploration of a wider conformational space. Consequently, replicas at lower temperatures benefit from this enhanced sampling, leading to more accurate thermodynamic and kinetic property estimates.

The choice to use REMD for studying TEM-1 enzyme mutations was motivated by several key factors:

1. Enhanced Sampling Efficiency:

REMD significantly improves the sampling of conformational space compared to conventional molecular dynamics (MD) simulations. This is particularly important for TEM-1, as enzyme activity and stability are closely linked to its conformational dynamics. The enhanced sampling ensures that rare but biologically relevant conformations are adequately captured, leading to a more comprehensive understanding of how mutations influence enzyme behavior.

2. Temperature Control:

By covering a range of temperatures (310 to 328 Kelvin), REMD allows us to investigate the temperature dependence of enzyme dynamics. This range is relevant for biological systems, providing insights into how temperature fluctuations might impact the structural and functional properties of TEM-1 mutants. Such information is critical for understanding enzyme behavior under physiological conditions.

3. Energy Barrier Overcoming:

The structural changes associated with mutations in TEM-1 may involve transitions between different conformational states separated by high energy barriers. REMD facilitates crossing these barriers more efficiently than standard MD, ensuring that the simulations are not trapped in local minima and can explore the global energy landscape more effectively.

4. Detailed Conformational Analysis:

The primary output of REMD simulations includes dihedral angles (phi-psi angles), which are crucial for understanding the conformational shifts in the enzyme. These angles provide detailed information on the backbone flexibility and overall structural changes, which are essential for correlating specific mutations with their effects on enzyme function and antibiotic resistance.

5. Correlation with Experimental Data:

The ability of REMD to sample diverse conformations enhances the predictive power of the simulations, allowing for better correlation with experimental data. This alignment is vital for

validating the simulation results and for gaining confidence in the predicted effects of TEM-1 mutations.

Overall, REMD was chosen as the simulation technique due to its robustness in capturing the dynamic behavior of biomolecules, its efficiency in exploring conformational space, and its ability to provide detailed structural insights necessary for understanding the impact of mutations on the TEM-1 enzyme.

3.6 Data preprocessing

For preprocessing, dihedral angles (φ and ψ) of 263 residues (totaling 526 angles) were extracted from the last 50 ns of REMD simulations, yielding 10,000 frames per variant and a total of 70,000 frames. The dataset included 40,000 records for susceptible variants and 30,000 records for resistant variants. Data balancing techniques were applied to address this discrepancy. Subsequently, center scaling was performed using a Standard Scaler to normalize the data, ensuring consistent scaling across all features for improved model performance.

3.7 Building machine learning models

In our study, we employed machine learning approaches, including Linear Regression, Support Vector Machines (SVM), and Random Forest, utilizing features extracted from REMD dihedral angles. The features, extracted from the structural dynamics of biomolecules via REMD simulations, provide a comprehensive representation of conformational changes induced by genetic variants. Linear Regression model provided us to model the single kernel relationship between dihedral angles and variant effects, providing insights into the directional impact of mutations. SVM, a powerful classifier, enabled us to discern complex patterns within the high-dimensional feature space, facilitating the identification of subtle variant effects. Additionally, Random Forest offered robustness against overfitting and noise, leveraging ensemble learning to predict variant effects based on a multitude of decision trees trained on dihedral angle

features. By harnessing these machine learning methodologies, we aimed to uncover predictive models that elucidate the intricate relationship between structural dynamics and variant effects, advancing our understanding of protein function at the atomic level.

Linear Regression

Linear Regression is a statistical method. It identifies the best-fitting line through the data points by minimizing and by calculating the sum of the squared differences and the observed and predicted values. The equation of this line ($y = mx + b$) represents the relationship, where 'm' denotes the slope (coefficients) and 'b' is the intercept. Linear Regression model is employed in identifying the linear relationship between REMD dihedral angles (features) and variant effects (target), offering insights into how mutations influence the enzyme's function **[34].**

Support Vector Machines (SVM)

SVM seeks the hyperplane that best separates different classes in the feature by maximizing the margin. In regression, it find the hyperplane that best fits the data by minimizing the errors. SVM achieves this by transforming the input data into a high-dimensional using a kernel function, where it identifies the optimal hyperplane. SVM is employed to detect complex patterns within the high-dimensional feature space derived from REMD dihedral angles, enabling the identification of subtle variant effects **[35].**

Random Forest

Random Forest is an ensemble learning technique. Each tree is trained on a random subset of the data and features. During prediction, each tree independently predicts the target variable, and the final prediction is made by averaging (in regression) or voting (in classification) the predictions from all trees. The ensemble nature of Random Forest makes it robust against overfitting and noise. In this study, Random Forest is used to predict variant effects based on numerous decision trees trained on REMD

dihedral angle features, providing a comprehensive understanding of the relationship between structural dynamics and variant effects **[36].**

By utilizing these diverse algorithms, we aimed to capture the intricate relationship between dihedral angles and the resistance or susceptibility of TEM-1 beta-lactamase variants to amoxicillin.

3.7 Feature importance

The identification of important features was conducted utilizing the Gini index **[37]**, resulting in the selection of the top 15 features. Subsequently, machine learning models were constructed using these identified features. This approach enabled the assessment of each feature's influence on the resistance of TEM-1 beta-lactamase variants to amoxicillin. By analyzing the performance of the models built with the top 15 features, we aimed to determine which specific dihedral angles exerted a greater impact on the resistance phenotype. This analysis provided valuable insights into the structural determinants underlying the observed resistance patterns, aiding in the elucidation of crucial molecular mechanisms.

CHAPTER-4
RESULTS AND DISCUSSION

4.1 PredictSNP –results

Mutation	PredictSNP	MAPP	PhD-SNP	PolyPhen-1	PolyPhen-2	SIFT	SNAP
E240K	83%	64%	55%	67%	-	66%	50%
E104K	74%	63%	66%	59%	87%	70%	58%
G238S	63%	80%	82%	67%	68%	68%	58%
Q39K	83%	80%	78%	67%	-	68%	71%
N175S	61%	59%	82%	59%	87%	79%	50%
A237T	74%	72%	58%	67%	87%	45%	77%

Table 4.1: Variant prediction results generated from PredictSNP

In the table, the red color indicates deleterious predictions, while green indicates neutral predictions. Despite these insights, determining whether a deleterious mutation translates to a resistant or susceptible phenotype remains inconclusive due to the tools' inability to comprehensively elucidate the specific role of the protein and the functional impact of the mutation.

The percentages given by various tools indicate the likelihood that a specific mutation is deleterious, meaning it negatively impacts protein function. However, determining whether a deleterious mutation translates to a resistant or susceptible phenotype depends on the specific role of the protein and the functional impact of the mutation, which existing tools do not elucidate comprehensively.

To address this gap, we employed a novel approach, Replica Exchange Molecular Dynamics (REMD). This method captures dihedral angles and, when combined with machine learning, can identify patterns that differentiate resistant variants from susceptible ones. Through REMD simulations, we aim to provide a clearer understanding of the structural and dynamic properties associated with antibiotic resistance, thereby enhancing our ability to predict resistance phenotypes from specific mutations.

4.2 REMD simulations extracted dihedral angles as features

REMD simulations were conducted on resistant variants (E240K, N175S, Q39K), susceptible variants (A237T, G238S, E104K), and the wild type (WT) of TEM-1 beta-lactamase. Dihedral angles (φ and ψ) were extracted from these simulations using Schrödinger software following the preparation of the 1btl molecules. These dihedral angles served as critical features for subsequent machine learning analyses. The extracted angles were pivotal in capturing the structural dynamics nature of the TEM-1 beta-lactamase variants and enabled the identification of key features that influence resistance and susceptibility.

4.3 Acceptance ratio

System	Average Acceptance Ratio
A237T	0.252
G238S	0.252
E104K	0.256
E240K	0.252
N175S	0.275
Q39K	0.252
WT	0.253

Table 4.2: Acceptance ratio

In REMD simulations, the acceptance ratio is calculated to ensure efficient sampling of conformational space. This ratio indicates attempted Monte Carlo moves that are accepted during the simulation. A high acceptance ratio suggests that the simulation is exploring the conformational space effectively, leading to more accurate and reliable results. Conversely, a low acceptance ratio may indicate that the simulation is not adequately sampling the conformational space, potentially leading to biased results. The average values of acceptance ratio typically fall between 0.2 to 0.3, indicating appropriate exploration of conformational space.

4.4 Model performance with all features

ML models	ACCURACY	MCC	Precision	F1 Score	Recall
LINEAR REGRESSION	0.81	0.92	0.96	0.96	0.95
SVM	0.95	0.92	0.93	0.92	0.94
RANDOM FOREST	0.95	0.93	0.95	0.95	0.96

Table 4.3 Model performance with all features

Three machine learning models—Linear Regression, Support Vector Machine (SVM), and Random Forest—were built to classify TEM-1 beta-lactamase variants into resistant and susceptible categories using all dihedral angle features (phi-psi) from the REMD simulations.

The classification report for the Random Forest model showed high precision, recall, and F1-scores:

RES/SUS	PRECISION	RECALL	F1 SCORE
RESISTANT	0.91	0.99	0.95
SUSCETIBLE	0.99	0.93	0.96

Table 4.4: Performance of Random forest in terms of resistance and susceptibility

The high performance of the SVM and Random Forest models suggests that the dihedral angles provide substantial information for accurately predicting resistance. Notably, the Random Forest model achieved the highest recall for resistant variants, indicating its robustness in identifying true resistant cases. These results shows the effectiveness of using REMD-derived features for predicting antibiotic resistance in TEM-1 beta-lactamase variants, providing valuable insights into the molecular dynamics underlying resistance mechanisms.

4.5 Feature importance

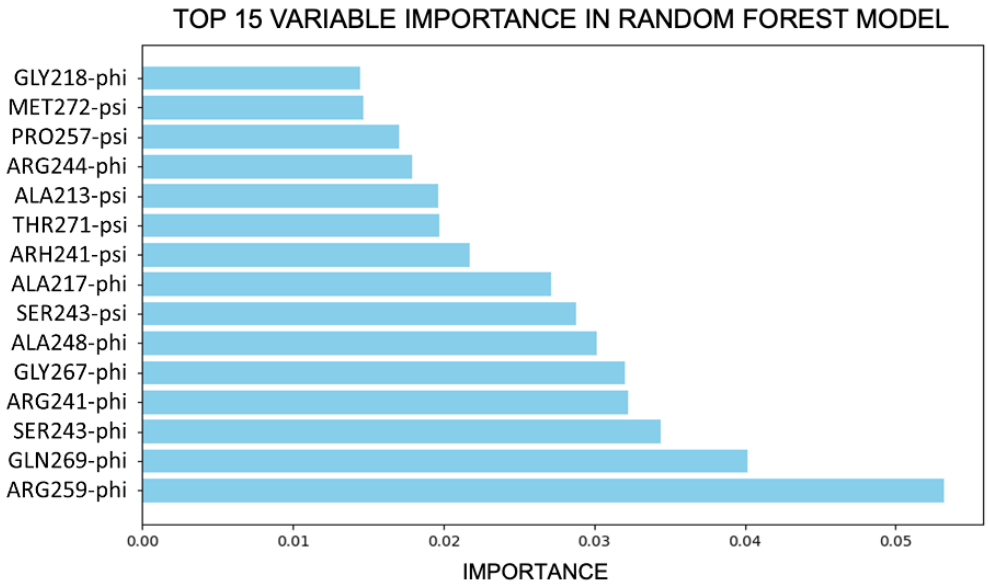

Fig 4.1: Top 15 variable importance in Random forest model

The Random Forest model, which outperformed Linear Regression and SVM in terms of the Matthews Correlation Coefficient (MCC), was employed to pinpoint the 15 most important features. These features were mainly dihedral angles, underscoring their critical impact on the functional phenotype of resistance in TEM-1 beta-lactamase variants.

One of the major findings was that the φ angle of ARG259 emerged as the most influential feature, indicating that ARG259 is pivotal in determining resistance. This suggests that specific residues significantly influence the modulation of resistance. The feature importance analysis reveals that these dihedral angles substantially affect the resistant properties of the protein structure. By identifying the most impactful dihedral angles, we gain valuable insights into the structural dynamics underlying antibiotic resistance in TEM-1 beta-lactamase variants. This understanding is essential for developing targeted strategies to combat resistance and enhance antibiotic effectiveness.valuable insights into the structural dynamics that underpin antibiotic resistance in TEM-1 beta-lactamase variants. This information is crucial for developing targeted strategies to combat resistance and improve antibiotic efficacy.

4.6 Model performance with top 15 features

ML models	ACCURACY	MCC	PRECISION	Recall	F1-score
LINEAR REGRESSION	0.69	0.66	0.71	0.71	0.71
SVM	0.91	0.84	0.92	0.92	0.92
RANDOM FOREST	0.93	0.86	0.93	0.93	0.93

Table 4.6: Model performance with top 15 features

When employing solely the top 15 features, Random Forest maintained its superior performance, achieving an accuracy of 0.93, an MCC of 0.86, and an F1-score of 0.93. In contrast, the efficacy of Linear Regression and SVM models diminished in comparison to their performance when utilizing all features.

The decline in performance observed in Linear Regression and SVM models can be attributed to the truncation of information when the number of features is restricted. These models likely rely on a broader array of features to capture the intricate relationships within the dataset. However, Random Forest, renowned for its robustness and adeptness in managing feature redundancy and interactions, demonstrated sustained proficiency even with only the top 15 features. This underscores Random Forest's capacity to effectively leverage crucial dihedral angles for discriminating between resistant and susceptible variants, thereby yielding more precise predictions despite the reduced feature set. Such findings underscore the significance of selecting appropriate models capable of efficiently utilizing pivotal features for classification tasks.

4.7 Catalytic sites & enzymatic funtional important cites

Catalytic & Active Sites
SER70
LYS73
SER130
ASN132
GLU166
LYS234
SER235
GLY236
ALA237
ARG244

Table 4.7: Catalytic cites identified from PDBsum

C_α Distance of Active Sites to Top 5 Features (1BTL)

	SER70	LYS73	SER130	ASN132	GLU166	LYS234	SER235	GLY236	ALA237	ARG244
ARG259-phi	29.6 Å	26.7 Å	24.1 Å	28.5 Å	30.2 Å	15.1 Å	17.1 Å	20.1 Å	23.2 Å	21 Å
GLN269-phi	3.9 Å	11.8 Å	9.6 Å	24.4 Å	23 Å	23 Å	19.3 Å	16.8 Å	13.5 Å	11.8 Å
SER243-phi	10.1 Å	12.7 Å	5.6 Å	17 Å	15.6 Å	14.6 Å	11.1 Å	8.2 Å	5.6 Å	3.9 Å
ARG241-phi	9.5 Å	14.7 Å	9.7 Å	10.1 Å	16.6 Å	16.4 Å	16.4 Å	13 Å	9.7 Å	9.4 Å
GLY267-phi	9.2 Å	15.5 Å	12 Å	22.8 Å	20.3 Å	20.7 Å	17.3 Å	14.6 Å	12 Å	9.7 Å

Table 4.8: C_α distance of Active sites to Top 5 features identified from Random forest algorithm

The table presents the C_α distances from the active sites of TEM-1 beta-lactamase to the top 5 features identified from our feature importance analysis. These distances were measured to understand the spatial relationship between critical residues and the active sites.

The dihedral angles identified through feature importance analysis (ARG259-φ, GLN269-φ, SER243-φ, ARG241-φ, and GLY267-φ) were visualized in the protein structure, and the distances of these amino acids to the active site residues were calculated. The tabulated data show that these distances are relatively higher, indicating no direct spatial relation between the top features and the active sites.

This lack of direct spatial correlation suggests that the influence of these dihedral angles on resistance may be mediated through complex and possibly long-range interactions rather than through proximity to the catalytic site. These findings provide new insights into the allosteric effects and structural dynamics that has a role in the resistance phenotype of TEM-1 beta-lactamase.

4.8 Machine learning insights show that 259 plays a vital role in functional importance, but it is considerably far from the catalytic sites.

Machine learning insights revealed that residue 259 (ARG259) plays a pivotal role in the functional importance of TEM-1 beta-lactamase variants, particularly in conferring resistance. Notably, ARG259 is situated considerably distant from the catalytic sites, as identified from pdbsum data. This finding raises intriguing questions regarding the relationship

between residue proximity and its contribution to resistance. Despite its spatial separation from the catalytic sites, ARG259 demonstrates significant influence on resistance, suggesting a complex interplay of structural factors in determining antibiotic susceptibility. This discrepancy challenges conventional assumptions about the direct proximity of residues to active sites dictating their functional significance. Further investigation into the molecular mechanisms underlying the involvement of distant residues like ARG259 in resistance could offer novel insights into the intricate dynamics of beta-lactamase function and evolution.

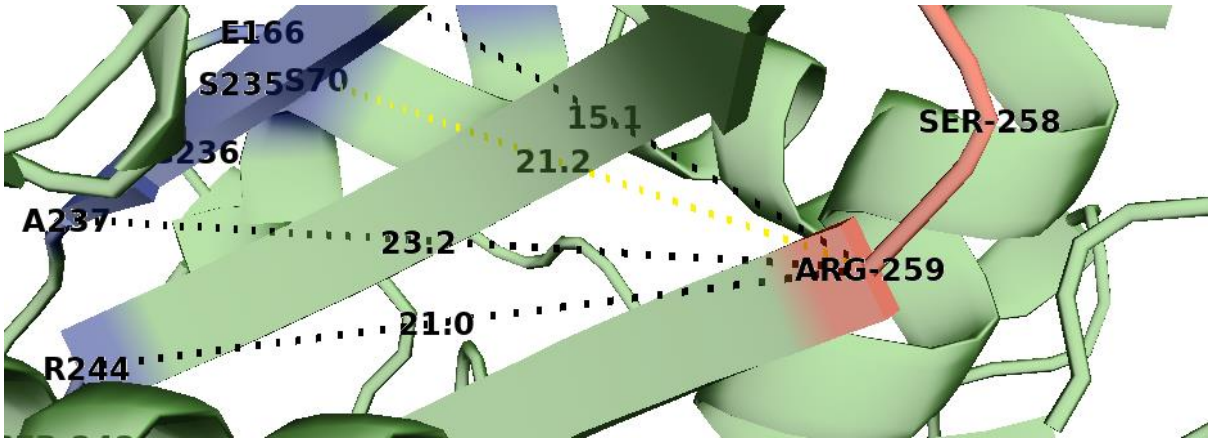

Fig 4.3: Distance Between ARG259 C_α and Catalytic Sites C_α

4.9 244-243 dihedral angles were close to the catalytic site, which is considered to have a important role in resistance.

By the analysis of dihedral angles revealed that residues 244 and 243, situated proximal to the catalytic site, exhibit significant structural importance in conferring resistance. Surprisingly, machine learning results indicate that SER-243, despite its proximity to the catalytic site, demonstrates a comparatively lesser contribution to resistance than residues 259 and 269. This discrepancy challenges conventional assumptions about the direct relationship between proximity to the catalytic site and functional significance.

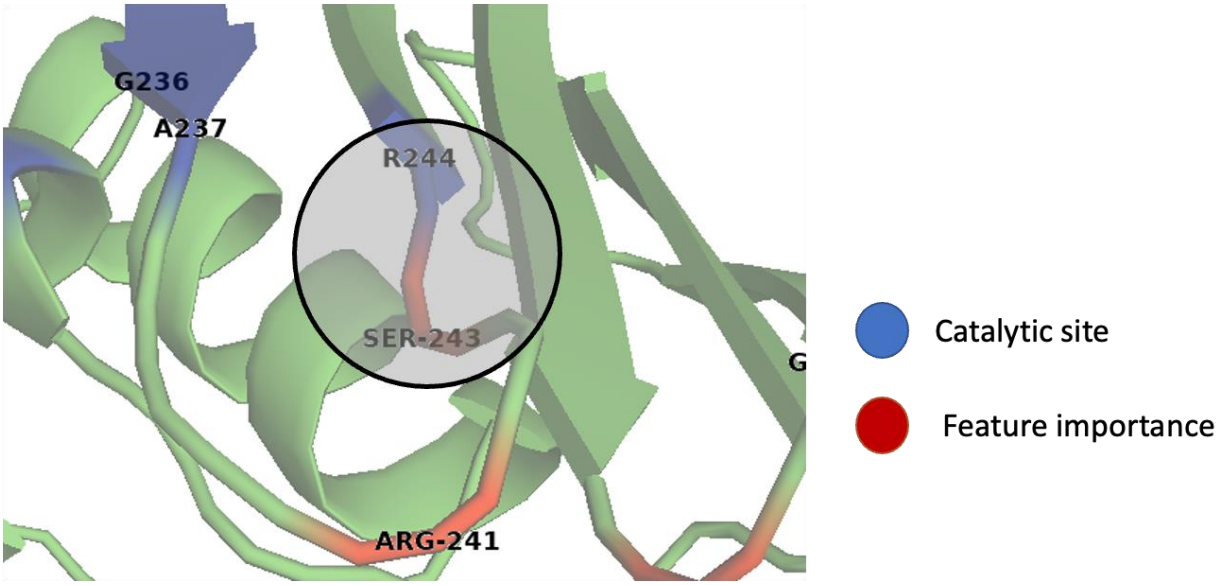

Fig 4.3: Distance Between feature important site SER243 and Catalytic Site ARG244

The unexpected findings underscore the complexity of the molecular determinants governing antibiotic resistance in TEM-1 beta-lactamase variants. Further exploration is warranted to elucidate the intricate interplay of structural factors and molecular dynamics underlying the observed discrepancies, offering valuable insights into the multifaceted mechanisms of beta-lactamase function and resistance.

4.10 Frequency distribution of Q39K and A237T

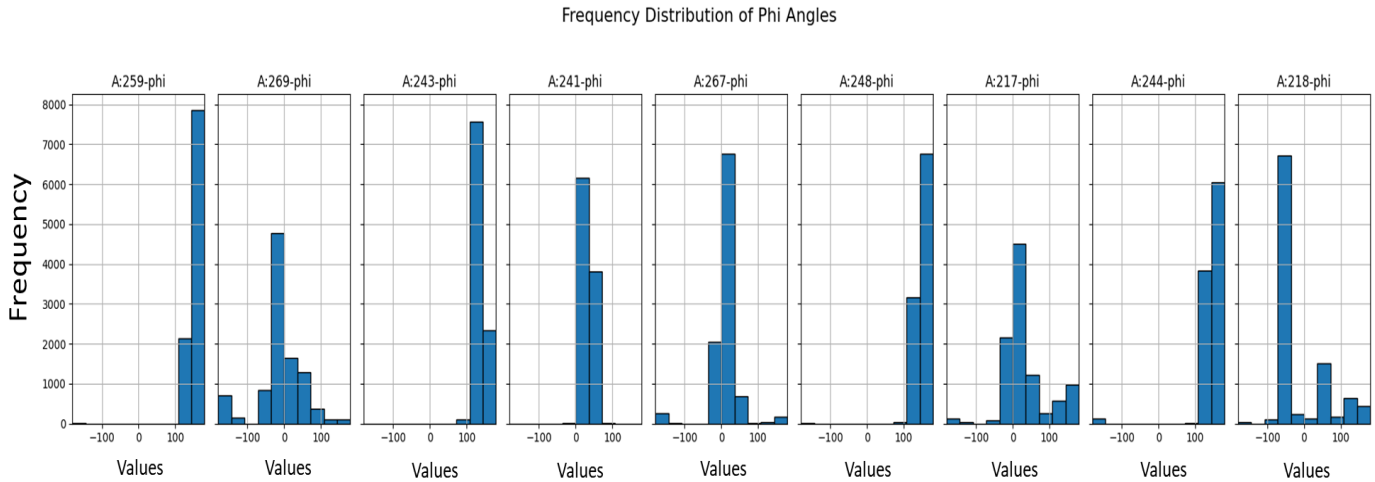

Fig 4.4: Frequency distribution of A237T of important features PHI angles

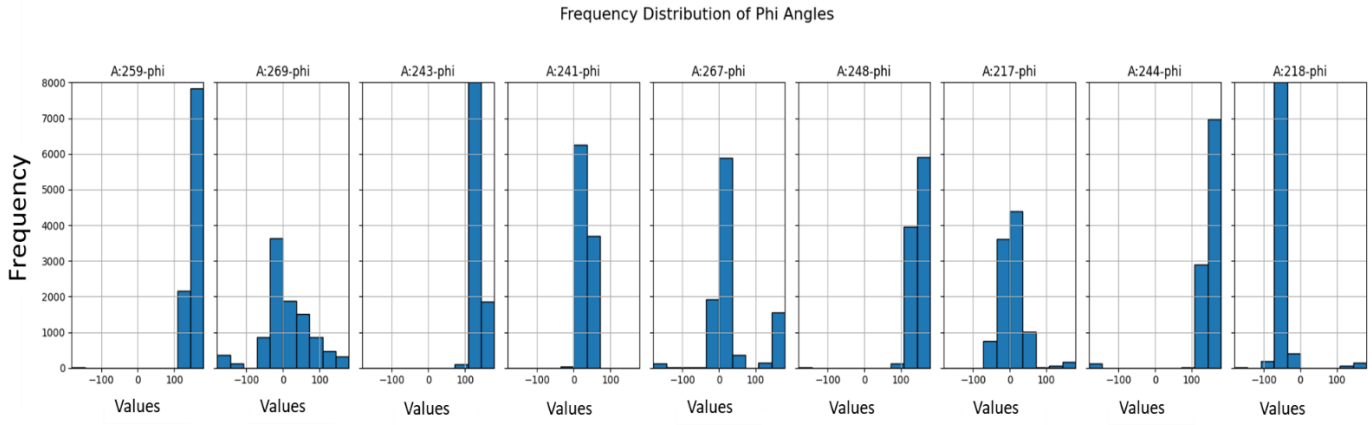

Fig 4.5: Frequency distribution of Q39K of important features PHI angles

The frequency distribution of dihedral angles was visualized to assess feature importance in variants Q39K (resistant) and A237T (susceptible), despite exhibiting distinct phenotypes. Surprisingly, both variants displayed similar frequency distributions, with a slight increase observed in susceptible A237T. This suggests nuanced differences in susceptibility, potentially influenced by structural factors. Given Q39K's distance from the catalytic site, it may affect the frequency distribution. Further investigation is warranted to ascertain if these patterns are linked to catalytic sites or exhibit resistance/susceptibility trends. Machine learning analysis identified important features, but no discernible patterns were evident, warrantingadditional research for conclusive insights.

4.11 Frequency distribution of Psi and Phi angles of all the systems of top 15 important features

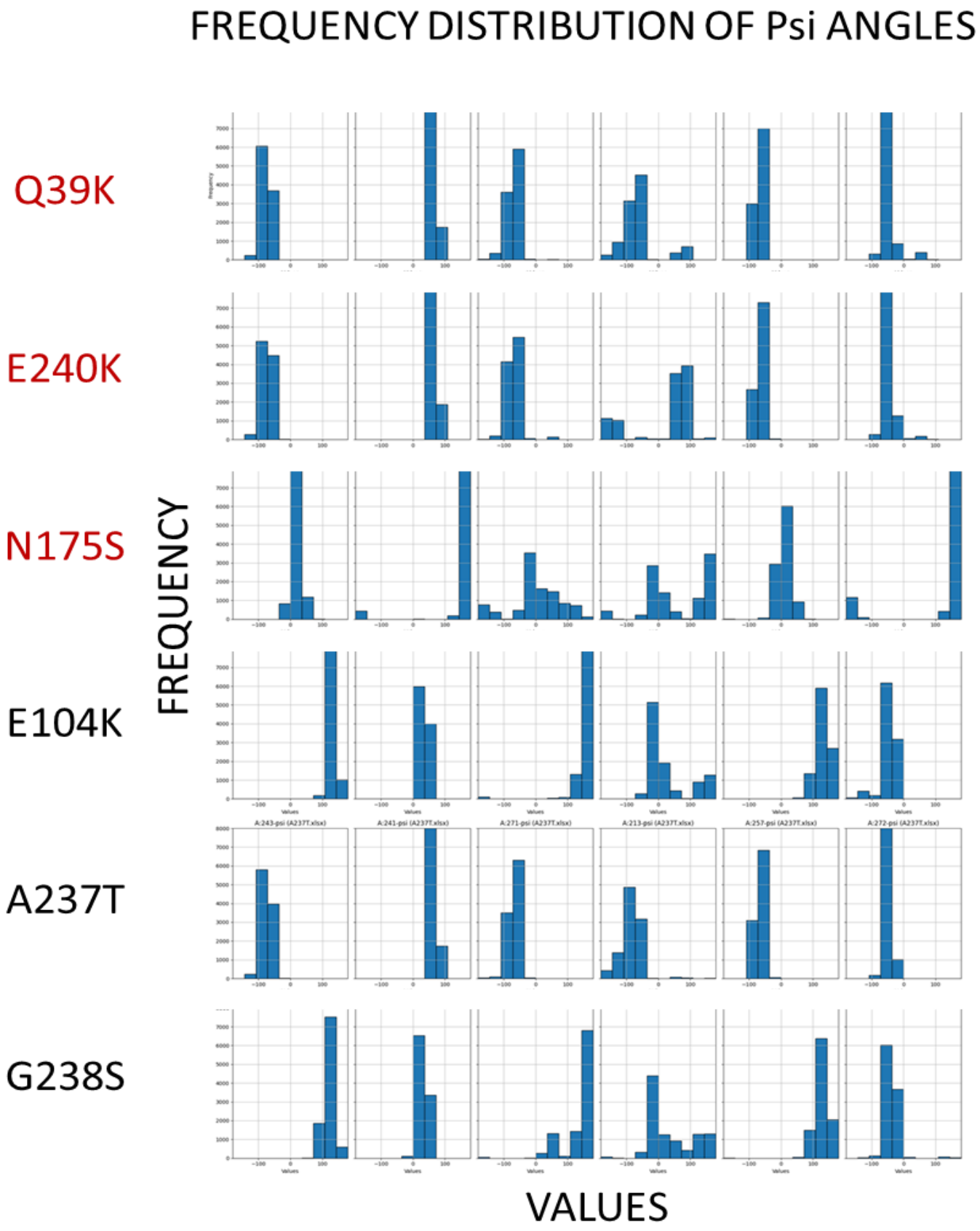

Fig 4.6: Frequency distribution of all the systems of important features Psi angles

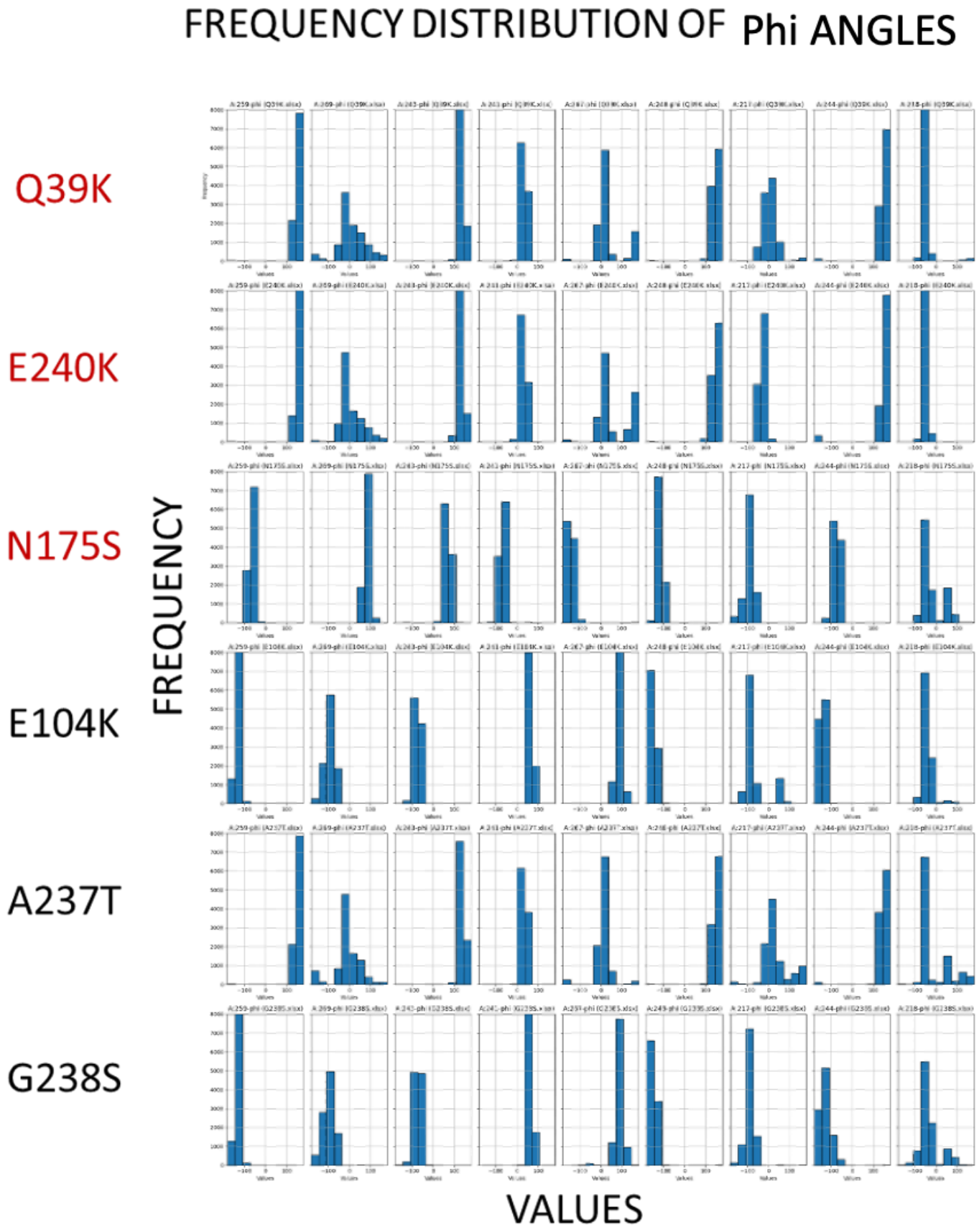

Fig 4.7: Frequency distribution of all the systems of important features Phi angle

Our hypothesis posited distinct patterns in the frequency distribution of Psi and Phi angles for resistant and susceptible variants. However, the analysis of the top 15 important features revealed contradictions. Specifically, the frequency distributions of Psi and Phi angles in resistant variants Q39K and E240K were similar to those in the susceptible variant A237T. Conversely, the N175S variant exhibited a unique pattern that did not align with any other system in terms of dihedral angles. Additionally, while the susceptible variants E104K and G238S followed a discernible pattern, this did not translate into a clear correlation within their frequency distributions (y-axis values). Despite similarities in values, the lack of consistent patterns or correlations suggests that the resistance phenotype may not be solely determined by dihedral angle distributions. This indicates that other structural or dynamic features could be influencing the resistance mechanisms in TEM-1 β-lactamases. These findings underscore the complexity of predicting resistance based on dihedral angles alone, highlighting the need for a multifaceted approach in understanding the structural basis of antibiotic resistance.

4.12 Discussion

The absence of a discernible pattern in the frequency distribution of dihedral angles between resistance and susceptibility in TEM-1 beta-lactamase variants raises intriguing questions about the underlying mechanisms of antibiotic resistance. The feature importance analysis conducted through machine learning revealed a list of dihedral angles unrelated to catalytic sites, suggesting a novel avenue for exploration regarding epistasis. This observation implies that pairwise intragenic epistasis, the phenomenon where the effects of mutations depend on each other's presence, may be pervasive among sequential mutations in TEM-1 lactase. The machine learning models effectively captured these complex interactions, indicating the potential role of epistasis in modulating resistance and susceptibility phenotypes. This insight provides a fresh perspective on the molecular dynamics of beta-lactamase evolution and underscores the need for further investigation into the interplay of genetic mutations in shaping antibiotic resistance profiles.

CHAPTER- 5
CONCLUSION

In conclusion, our study demonstrates the efficacy of a molecular dynamics (MD)-based machine learning framework in accurately classifying resistant and mutant categories of TEM-1 beta-lactamase variants. By leveraging this approach, we identified top features that contribute significantly to the discrimination between resistant and mutant variants, shedding light on crucial residues governing antibiotic resistance. Notably, residue ARG259 emerged as a key determinant of resistance, despite its spatial distance from the catalytic site. This finding underscores the presence of long-range and complex interactions within the protein structure, challenging traditional notions about the direct relationship between residue proximity and functional significance. Our results highlight the utility of integrating MD simulations and machine learning techniques to uncover subtle yet critical molecular determinants of antibiotic resistance. By elucidating the structural and dynamic features underlying resistance, our approach offers valuable insights into the mechanistic basis of beta-lactamase function and evolution. Future research efforts can build upon these findings to develop targeted strategies for combating antibiotic resistance and informing rational drug design initiatives.

CHAPTER-6

REFERENCES

1. Aslam, B., Wang, W., Arshad, M. I., Khurshid, M., Muzammil, S., Rasool, M. H., ... & Baloch, Z. (2018). Antibiotic resistance: a rundown of a global crisis. *Infection and drug resistance*, 1645-1658.

2. Dadgostar, P. (2019). Antimicrobial resistance: implications and costs. *Infection and drug resistance*, 3903-3910.

3. Essack, S. Y., Desta, A. T., Abotsi, R. E., & Agoba, E. E. (2017). Antimicrobial resistance in the WHO African region: current status and roadmap for action. *Journal of public health*, *39*(1), 8-13.

4. Bourdonnais, E., Le Bris, C., Brauge, T., & Midelet, G. (2024). Tracking antimicrobial resistance indicator genes in wild flatfish from the English Channel and the North Sea area: a one health concern. *Environmental Pollution*, *343*, 123274.

5. Ventola, C. L. (2015). The antibiotic resistance crisis: part 1: causes and threats. *Pharmacy and therapeutics*, *40*(4), 277..

6. Kanj, S. S., & Kanafani, Z. A. (2011, March). Current concepts in antimicrobial therapy against resistant gram-negative organisms: extended-spectrum β-lactamase–producing enterobacteriaceae, carbapenem-resistant enterobacteriaceae, and multidrug-resistant Pseudomonas aeruginosa. In *Mayo Clinic Proceedings* (Vol. 86, No. 3, pp. 250-259). Elsevier.

7. Bush, K., & Bradford, P. A. (2016). β-Lactams and β-lactamase inhibitors: an overview. *Cold Spring Harbor perspectives in medicine*, *6*(8), a025247.

8. Nordmann, P., Naas, T., & Poirel, L. (2011). Global spread of carbapenemase-producing Enterobacteriaceae. *Emerging infectious diseases*, *17*(10), 1791.

9. Hasani, K., Sadeghi, H., Vosoughi, M., Sardari, M., Manouchehrifar, M., & Arzanlou, M. (2023). Characterization of beta-lactamase producing Enterobacterales isolated from an urban community wastewater treatment plant in Iran. *Iranian Journal of Microbiology*, *15*(4), 521.

10Shah, M., Barbosa, T. M., Stack, G., & Fleming, A. (2024). Trends in antibiotic prescribing in primary care out-of-hours doctors' services in Ireland. *JAC-Antimicrobial Resistance*, *6*(1), dlae009.

11. Potter, R. F., D'Souza, A. W., & Dantas, G. (2016). The rapid spread of carbapenem-resistant Enterobacteriaceae. *Drug Resistance Updates*, *29*, 30-46.

12. Bonomo, R. A. (2017). β-Lactamases: a focus on current challenges. *Cold Spring Harbor perspectives in medicine*, *7*(1), a025239.

13. Moisan, H., Pruneau, M., & Malouin, F. (2010). Binding of ceftaroline to penicillin-binding proteins of Staphylococcus aureus and Streptococcus pneumoniae. *Journal of antimicrobial chemotherapy*, *65*(4), 713-716.

14. Fisher, J. F., & Mobashery, S. (2020). Constructing and deconstructing the bacterial cell wall. *Protein science*, *29*(3), 629-646.

15. du Plessis, H., & Van Niekerk, A. (2014). A new GISc framework and competency set for curricula development at South African universities. *South African Journal of Geomatics*, *3*(1), 1-12.

16. Lovering, A. L., De Castro, L. H., Lim, D., & Strynadka, N. C. (2007). Structural insight into the transglycosylation step of bacterial cell-wall biosynthesis. *Science*, *315*(5817), 1402-1405.

17. Otero, L. H., Rojas-Altuve, A., Llarrull, L. I., Carrasco-López, C., Kumarasiri, M., Lastochkin, E., ... & Hermoso, J. A. (2013). How allosteric control of Staphylococcus aureus penicillin binding protein 2a enables methicillin resistance and physiological function. *Proceedings of the National Academy of Sciences*, *110*(42), 16808-16813.

18. Tipper, D. J., & Strominger, J. L. (1965). Mechanism of action of penicillins: a proposal based on their structural similarity to acyl-D-alanyl-D-alanine. *Proceedings of the National Academy of Sciences*, *54*(4), 1133-1141.

19. Jacquier, H., Birgy, A., Le Nagard, H., Mechulam, Y., Schmitt, E., Glodt, J., ... & Tenaillon, O. (2013). Capturing the mutational landscape of the beta-lactamase TEM-1. *Proceedings of the National Academy of Sciences*, *110*(32), 13067-13072.

20. Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., ... & Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nature methods*, *7*(4), 248-249.

21. Bendl, J., Stourac, J., Salanda, O., Pavelka, A., Wieben, E. D., Zendulka, J., ... & Damborsky, J. (2014). PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. *PLoS computational biology*, *10*(1), e1003440.

22. Bromberg, Y., & Rost, B. (2007). SNAP: predict effect of non-synonymous polymorphisms on function. *Nucleic acids research*, *35*(11), 3823-3835.

23. Capriotti, E., Calabrese, R., & Casadio, R. (2006). Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics*, *22*(22), 2729-2734.

24. Guerois, R., Nielsen, J. E., & Serrano, L. (2002). Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *Journal of molecular biology*, *320*(2), 369-387.

25. Li, B., Krishnan, V. G., Mort, M. E., Xin, F., Kamati, K. K., Cooper, D. N., ... & Radivojac, P. (2009). Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics*, *25*(21), 2744-2750.

26. Ng, P. C., & Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic acids research*, *31*(13), 3812-3814.

27. Rives, A., Meier, J., Sercu, T., Goyal, S., Lin, Z., Liu, J., ... & Fergus, R. (2021). Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proceedings of the National Academy of Sciences*, *118*(15), e2016239118.
28. Davies, J. E. "Origins, acquisition and dissemination of antibiotic resistance determinants." Ciba Foundation Symposium, vol. 207, 1997, pp. 15-27; discussion 27-35.

29. Dr Graham Beards at en.wikipedia. "Antibiotic resistance tests; the bacteria in the culture on the left are sensitive to the antibiotics contained in the white paper discs. The bacteria on the right are resistant to most of the antibiotics."

30. Akhavan, B. J., Khanna, N. R., & Vijhani, P. (2023). Amoxicillin. In *StatPearls [Internet]*. StatPearls Publishing.

31. Zhou, J., & Troyanskaya, O. G. (2015). Predicting effects of noncoding variants with deep learning–based sequence model. *Nature methods*, *12*(10), 931-934.

32. Laskowski, R. A., Jabłońska, J., Pravda, L., Vařeková, R. S., & Thornton, J. M. (2018). PDBsum: Structural summaries of PDB entries. *Protein science*, *27*(1), 129-134.
33. GitHub repository for REMD. "GitHub - dspoel/remd."

34. Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis*. John Wiley & Sons.

35. Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2, pp. 1-758). New York: springer.

36. Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5-32.

37. Breiman, L. (2017). *Classification and regression trees*. Routledge.

CHAPTER 7

APPENDIX

7.1 Plagiarism report

REPORT OF THE PLAGIARISM CHECK

THIS REPORT CERTIFIES THAT THE ATTACHED WORK
TEM-1 plagiarism thesis
WAS CHECKED WITH THE PLAGIARISM PREVENTION SERVICE
MY.PLAGRAMME.COM AND HAS:
SIMILARITY

3%

RISK OF THE PLAGIARISM

14%

PARAPHRASE

0%

IMPROPER CITATIONS

0%

File name: plagarism.docx
File checked: 2024-05-22
Report generated: 2024-05-22