# Use of Computational Methods To Understand The Pattern Of Antimicrobial Resistance

Thesis submitted to the University of Nottingham for the degree of PhD Bioscience

**Hokin Chio**

**4342366**

Supervised by **Dov Sketel**

Signature: _____

# Use of Computational Methods To Understand The Pattern Of Antimicrobial Resistance

# Acknowledgements

# List of abbreviations

Antimicrobial resistance (AMR)

Antibiotic resistance genes (ARGs)

Minimal inhibitory concentration (MIC)

Liquid chromatography (LC)

Mass spectrometry/ Mass spectrometry (MS/MS)

Liquid chromatography-tandem mass spectrometry (LC-MS/MS)

Solid-phase extraction-liquid chromatography-tandem mass spectrometry (SPE-LC-MS/MS)

Quantitative polymerase chain reaction (qPCR)

Polymerase chain reaction (PCR)

Principal component analysis (PCA)

Metric multidimensional scaling (MDS)

t-distributed stochastic neighbor embedding (t-SNE)

Potential of Heat Diffusion of Affinity-Base Transition Embedding (PHATE)

Molecular dynamics (MD)

*k*-nearest neighbour (KNN)

Genetic optimization for ligand docking (GOLD)

General-purpose graphics processing units (GP-GPU)

Protein data base (PDB)

Penicillin binding proteins (PBPs)

*Pseudomonas aeruginosa (P. aeruginosa)*

High-performance liquid chromatography (HPLC)

# Abstract

Antimicrobial resistance (AMR) as one of the most serious problems in the world is especially urgent with the increase in antibiotic resistance of bacteria across the world. Antibiotics reach the environment via excretions from humans and agriculture, and industrial and hospital waste products. The environmental concentrations of antibiotics are usually much lower than the minimal inhibitory concentrations and most often lower than concentrations predicted to select for resistant strains in the laboratory. However, exposure to low levels of antibiotics has also been shown to increase resistance, resulting in the increase of selective pressure.

The resistance pattern of the AMR in different environments has been identified by many studies but the connection between the antibiotics present in the environment and the resistance pattern remains uncertain. To understand how different patterns of resistance emerge, computational method is essential for processing and analyzing the molecular interaction model to estimate the bioactivity of the metabolites of antibiotics and evaluated methods for visualizing high dimensional resistance data, in order to be able to better ascertain patterns of resistance.

Through molecular docking and molecular dynamics, the metabolites (5R) pseudopenicillin, (5S)-penicilloic acid and 6APA are found to be potentially bioactive towards target protein penicillin binding protein. T-SNE has been suggested to be the most suitable for analyzing AMR data compared with other methods (PCA, MDS, isomap and PHATE) and this helps to have a better understanding of correlative of the AMR development. Therefore, some undetected compounds (metabolites of antibiotics) may cause selective pressure and increase resistance. These compounds may also be involved in developing bacteria resistance within environments. This could have considerable significance for environmental surveillance for antibiotics to reduce antimicrobial resist.

# Chapter 1

## Introduction

This chapter focuses on the development of antimicrobial resistance (AMR) in the environment and agriculture. It provides details of the detection of antibiotics and AMR. This chapter also introduces the difficulties of visualizing different patterns of resistance gene of microbes. This thesis consists of six chapters. Chapter 1 is an introduction, discussing the importance of AMR within the environment and the methods used for its detection. It also highlights and discusses the problem statement. Chapter 2 is the methodological background, containing background and details of the computational methods used in this study, the choice of antibiotic and experimental bacterium and the details of the degradation pathway of penicillin. Chapter 3 investigates methods for analyzing high dimensional data. Analysis of molecular docking and molecular dynamics are presented in chapters 4 and 5, respectively. Chapter 6 summaries the work done in chapter 3, 4 and 5 and discusses the limitations of molecular docking and molecular dynamics used in this study. It also contains discussion, summary, and recommendations resulting from this research.

## 1.1 Introduction

Antibiotics are becoming decreasingly effective as drug resistance spreads with the misuse and overuse of the antibiotics, causing multi- and pan-resistant bacteria (Magiorakos et al., 2012). The emergence of multi- and pan-resistant bacteria has become an urgent global concern in that these can lead to infections not treatable with any existing antibiotics (Murray et al., 2022). As a result, treatments for common illness have become ineffective, and thus potentially life threatening. The emergence of AMR in livestock has been shown to affect human health through contamination of meat, milk, and egg products(Bacanlı & Başaran, 2019; Mărgărita Ghimpet et al., 2022); direct contact with animals(Levy & Fitzgerald, 1978; Van Den Bogaard et al., 2002), and by contamination of the

environment(Kraemer et al., 2019; Van Den Bogaard et al., 2002) through the spread of fertilizer by water flow (Prestinaci et al., 2015). Water and soil with inadequate water, sanitation and hygiene are a major location of AMR development.(WHO,2020) Antibiotics widely used in medicine and agriculture as shown in the most recent report from WHO showing the increasing consumption of antibiotics, from 21.1 billion defined daily doses in 2000 to 34.8 billion defined daily doses in 2015(Klein et al., 2018), the problem of AMR in the environment has received less attention than AMR in humans or animals.

## 1.2 AMR development

Antibiotic resistance can arise both from mutations in the pre-existing genome of a bacterium and from the uptake of foreign DNA. Mutations readily occur and resistance becomes accumulated in patients and animals treated with the antibiotic. Pathogens are rarely exposed to high concentrations of antibiotics outside of clinical treatment, so the external environments are not likely to provide major contribution to mutation-based evolution of resistance for most pathogens. Regarding the uptake of novel resistance factors, the environment, with its highly varied ecological niches, provides an unmatched gene pool with a diversity that exceeds the human and domestic animal microbiota(Larsson & Flach, 2021; Rinke et al., 2013; Schulz et al., 2017). Indeed, the most important feature of the environmental microbiome is its immense diversity, providing a large genetic diversity that can potentially be received and used by pathogens to counteract the effect of antibiotics(Berglund et al., 2017, 2020). All approved antibiotic classes so far, whether they be natural, semi-synthetic or synthetic compounds, have been met by resistance in at least some of the pathogens they target.

### 1.2.1 Resistance mechanisms of bacteria

Resistance to antibiotics can be conveyed via three mechanisms(Walsh, 2000; Wright, 2011). Firstly, antibiotics can be inactivated (Egorov et al., 2018), e.g., beta-lactamases

2

cleaving beta-lactams such as penicillin (X. Zeng & Lin, 2013). e.g., Antibiotic inactivation is one of the resistance mechanisms in bacteria(Egorov et al., 2018). One of the examples is that beta-lactamases deactivate beta-lactam antibiotics such as penicillin. Bacteria produce beta-lactamases to inactivate beta-lactam by cleaving the beta-lactam ring and unable to exert its bactericidal effect. Beta-lactamase recognized the beta-lactam ring structure within the antibiotics. The beta-lactam binds to the active site of the beta-lactamase and the beta-lactamase catalyzes the hydrolysis of the beta-lactam ring by attacking the bond between the nitrogen and the carbonyl carbon which results in the opening of the ring structure.(Bush & Bradford, 2016)

Secondly, antibiotics can be transported outside of the bacterial cell via efflux pumps in the cell membrane to actively expel antibiotic from the bacteria(Soto, 2013), e.g., TetA proteins pumping tetracyclines outside of cells (Grossman, 2016).This lowers the concentration of the antibiotics inside the bacteria. The activation of these efflux pumps is often a part of the bacterial stress response to the presence of antibiotics and is regulated at the genetic level(L. Huang et al., 2022). The bacteria can transfer genes responsible for these pumps to other bacteria, which spreads this form of resistance.(Soto, 2013).

Thirdly, the modification of the antibiotic's target through mutations during cell reprodution, e.g., point mutations in *gyrA* prevent binding by ciprofloxacin(Vinué et al., 2016).As a result, the structure of the active site may change making it harder to bind with the antibiotics(Mendonça & Marana, 2011). Also, the changes in structure may result in reducing the binding affinity, making it less effective in inhibiting the compulsory molecules to maintain the living of bacteria(Sedighpour & Taghizadeh, 2022).

## 1.2.2 Impact of antibiotic pollution

The selective pressures imposed by antibiotic pollution can impact on the evolutionary dynamics observed in microbial populations in diverse ways (Martínez, 2017). Bacterial populations present different levels of tolerance to antibiotics due to changes of gene

expression (El Meouche & Dunlop, 2018). On the other hand, antibiotics can thus reduce diversity in microbial populations, by favoring the growth of resistant or tolerance to microbial lineages under strong selective pressures(Sánchez-Romero & Casadesús, 2014). Moreover, weak selective pressures imposed by lower antibiotic concentrations can selectively favor the growth of bacterial lineages with increased phenotypic and genotypic diversity. Indeed, intermediate concentrations of antibiotics such as amikacin, ciprofloxacin, and streptomycin were found to select for larger colony size in *Staphylococcus* spp. (L. Lee et al., 2018), which was linked to increased genetic diversity and adaptability in several species of bacteria(Andersson & Hughes, 2014; Justice et al., 2008; Llor & Bjerrum, 2014).

Despite the likelihood that the natural production of antibiotic molecules contributes to the evolution of antibiotic resistance genes (ARGs)(J & D, 2010), it is not responsible for the rapid spread of resistance factors across bacteria since the level of antibiotics produced in natural is low (Aminov, 2010). Anthropogenic antibiotics act on a macroscale and associate with selection pressures through the entire microbial community, making a sharp contrast with antibiotics produced by environmental microorganisms which spread widely but act on a concentration gradient around the producing microorganisms (Cycoń et al., 2019).

### 1.2.3 AMR development in environment

Antibiotics enter the environment in three main ways: via excretions (either urine or faeces) from humans and livestock (J. Wang et al., 2020); unused drugs in hospital(Anwar et al., 2020) or agriculture(CABI Agric Biosci et al., 2020); and waste from industrial production of antibiotics (Bielen et al., 2017). (Figure1.1) The spread of antibiotics from excretion, as one of the largest sources, has been restricted by controlling the dosage of antibiotics used in human or domestic animals.
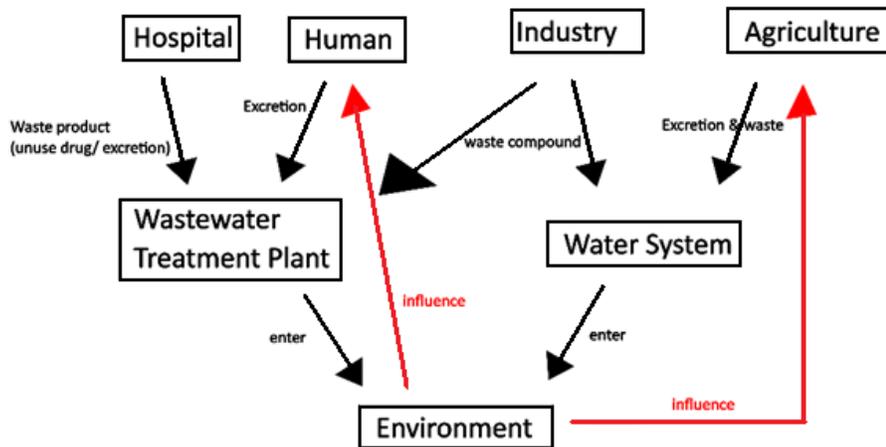
**Figure 1.1** Schematic of flows of antibiotics entering the environment (black arrows) and the feedback from the presence of antibiotics (red arrows). Hospital waste, including patient excreta and unused drugs, domestic human waste and industrial waste enter the wastewater treatment plants for water purification, and ideally degradation of harmful substances. Remaining undegraded antibiotics enter and spread into the environment though underground water systems, rivers, and streams. The antibiotics at this point might be degraded by natural biotic (e.g., microbes) or abiotic (e.g., UV light) factors but remaining antibiotics may select for resistance.

The environmental concentrations of antibiotics are usually much lower than the minimal inhibitory concentration (MIC) and concentrations predicted to select for AMR in the laboratory (Bengtsson-Palme & Larsson, 2016; Gullberg et al., 2011). An exception is the River Ganges in India (Kristiansson et al., 2011), containing a high level of antibiotics with reported highest concentration of beta-lactams13800ng/L, elfamycin 2100ng/L and fluoroquinolone 1400ng/L (Akhter et al., 2023; Bhagat et al., 2020). A wide range of microorganisms, e.g. Nitrospirae, Bacillariophyta, Bacteroidetes containing resistance genes (major ARGs are beta-lactams, elfamycin and aminoglycoside) was discovered in the same river(Reddy & Dubey, 2019). The correlation between concentration of antibiotics and the abundance of ARGs could be explained simply by different levels of pollution produced by human when the pollution level is increasing(Karkman et al., 2019), an increase level of

concentration of antibioics may increase the abundance of ARGs (Rahman et al., 2023). Still, the concentrations suspected to select for resistance are exceeded in many places, particularly in sewage treatment plants.(Y. Yang et al., 2014) The problem of inappropriate use of antibiotics over the world leads to the raise of AMR: unprescribed use, overuse, sometimes underpinned by a lack of knowledge about AMR, are the main factors of inappropriate use of antibiotics (Llor & Bjerrum, 2014; Mallah et al., 2022). When microorganisms are exposed to sublethal concentrations or incomplete antibiotics (Andersson & Hughes, 2014), they have increased opportunities to develop resistance gene.

In addition to changes in the population composition of bacteria, exposure to low levels of antibiotics has also been shown to increase genetic diversity in microbial population through a range of bacterial responses (Liu et al., 2020); this may include increased mutation rate throughout the whole genome (Andersson & Hughes, 2014; Foster, 2007), increased rate of horizontal transfer of genetic material between bacteria through conjugation(Maiques et al., 2006), or increased uptake of extracellular DNA (Muschiol et al., 2015). Most antibiotics were found to affect gene regulation at the transcription level (Goh et al., 2002), with direct binding (Davies et al., 2006) or through riboswitches (Davies et al., 2006) and quorum sensing (Rémy et al., 2018), therefore, phenotypic variability and virulence were increased (Rémy et al., 2018). Together, these mechanisms increase the available pool of genetic and phenotypic diversity in bacterial populations exposed to antibiotics (Bottery et al., 2020). In turn, this may facilitate the further evolution of antibiotic resistance if selective pressure increases, as predicted by Fisher's Fundamental theorem of evolution by natural selection (L. Lee et al., 2018; Li, 1967). The environmental occurrence, toxicity, degradation, and removal of antibiotics are tightly connected to the presence of environmental pollutants.

## 1.2.4 Antibiotics used in agriculture

Antibiotics are not just used to treat human infections; they are also used in livestock farming. Antibiotics can also be used for disease treatment of animals(The Use of Antibiotics on Healthy Farm Animals and Antimicrobial Resistance), the prevention of diseases (Hao et al., 2014), while sub-therapeutic levels in concentrated animal feed can be benefit for growth promotion (L. Lee et al., 2018) or feed conversion efficiency(Hao et al., 2014). The use of antibiotics in livestock animals will result in the composition and diversity of the microbes in the gut of treated animals being changed (Ramirez et al., 2020), including selection for resistance (W. Wang et al., 2021)

Antibiotics can be provided to the livestock population in feed or water (Mărgărita Ghimpet et al., 2022; Van Den Bogaard et al., 2002), rather than targeting only diseased animals. In such systems, development of resistance becomes difficult to avoid(van den Bogaard et al., 2001). Animals treated with antibiotics will excrete a portion of the tetracycline (25%-75%) and macrolide (40%-60%) through their feces and urine(Massé et al., 2014). These excreta will contaminate the soil and water with the antibiotics, resistance genes and resistant organisms (Tian et al., 2021)

## 1.3 Detection of antibiotics

Analytical methods based on liquid chromatography (LC) and mass spectrometry (MS) have been applied to selectively and sensitively detect antibiotics present in the environment. Currently, due to the high sensitivity of tandem mass spectrometry (MS/MS) detection(Hernandez et al., 2006), liquid chromatography-tandem mass spectrometry (LC-MS/MS) detection were widely used in the analysis of antibiotics in different environmental matrices(Y. Zhang et al., 2020). MS/MS involves a two-step process in which ions of interest are selected and fragmented in the first stage, and then the resulting fragments are analyzed in the second stage. This provides information about the molecular weight, sequence, and the structural details of the molecule.(Hernandez et al., 2006) LC-MS/MS combines the

separation capabilities of liquid chromatography with the sensitivity and selectivity of MS/MS. The sample is first separated with liquid chromatography based on its physicochemical properties, such as polarity or size. The separated components are then introduced into the mass spectrometer, where MS/MS performed (Pitt, 2009).

Another approach to detecting antibiotics in the environment is solid-phase extraction-liquid chromatography-tandem mass spectrometry (SPE-LC-MS/MS) (Kairigo et al., 2020). Solid phase extraction uses a soild sorbent material packed in a cartridge to retain the sample while unwanted components are washed away. The extracted samples are then eluted from the sorbent with an appropriated solvent which depends on the polarity or the chemical properties of the analytes of interest and the sorbent; methanol, ethyl acetate hexane and water are common solvents used in solid phase extraction. The eluate is then purified with liquid chromatography based on its physicochemical properties. The purified sample is then analysed by MS/MS.(Deegan et al., 2011)

## 1.4 Detection for AMR

There are three main approaches for detecting antimicrobial resistance genes or phenotypes: culture-based methods, quantitative polymerase chain reaction based methods and metagenomics-based methods.

### 1.4.1 Culture-Based Methods

The main advantage of culture-based methods is that they test for phenotypic resistance in living cells(McLain et al., 2016). The bacterial sample is inoculated into suitable growth media which depends on the selected target. The inoculated media are incubated with controlled environmental conditions. Further, once target colonies have been isolated, they can be subject to further analysis, such as multidrug-resistance testing (Kalokhe et al., 2013), sequence-based typing (Voorter et al., 2014), or whole genome sequencing (Berberich et al., 2018). There are several advantages when using culture-based methods;

this method enables the isolation and the study of specific microbial species. Also, culturing microbes allows researchers to observe the difference in growth rates of the microbes in the presence or absence of antibiotics. The Disk diffusion method is commonly used to determinate the antibiotics sensitivity or resistance. The antibiotics paper disks are placed on the agar plate where the microbes have been transferred. Then the plate is incubated. A zone of inhibition will appear around the disk where bacterial growth has been prevented if the bacteria are sensitive to the antibiotic. Another methods of testing the sensitivity or resistance is broth dilution method. The bacteria are grown in liquid media containing different concentrations of antibiotics. The MIC of the bacteria can be determined in this method. There are also limitations of culture-based methods: only microbes that are already known and culturable can be identified, and culture-based methods rely on viable microbes that can grow and reproduce, and so this method excludes the detection of unknown, non-culturable, non-viable or inactive microbes.

## 1.4.2 Quantitative polymerase chain reaction (qPCR)-Based Methods

The polymerase chain reaction (PCR) (Mullis et al., 1986) is used to amplify DNA or RNA sequences; qPCR, a derived method, can measure the amount of specific DNA or RNA sequence within a sample and it can provide very sensitive detection and quantification in real time (Kralik & Ricchi, 2017). Specific primers are designed to the target DNA (for genomic DNA) or RNA (for expressed genes) sequences. The qPCR reaction undergoes a series of thermal cycles which consists of repeated heating and cooling steps. These involve the denaturation, annealing and extension of the DNA or RNA sequence. During the thermal cycling, the qPCR instrument continuously monitors the fluorescence emitted from the probe. As the DNA synthesis occurs, the probe gets degraded by the DNA polymerase, leading to the separation of the fluorescent dye. This results in the increase in fluorescence signal which is directly proportional to the amount of the target DNA or RNA in the sample (Adams, 2020). Because of its sensitivity, qPCR is useful for identifying ARGs in dilute samples, for example wastewater or water treatments plants, such as the detection of *sul1*, *tetA* and *intl1* within water sample by the use of qPCR.(Keenum et al., 2022). The

development of qPCR arrays, containing a set of designed primers targeting specific ARGs, has enabled detection and quantification of many known resistance genes in parallel. These arrays provide a rapid method for AMR detection compared to traditional culture-based methods and multiple samples can be performed at the same time, making the results highly comparable.

## 1.4.3 Metagenomic-Based Methods

Metagenomic sequencing (L. Zhang et al., 2021) is carried out through direct extraction of genomic DNA(Cheng & Jiang, 2006) and the application of next generation sequencing (Slatko et al., 2018). There are two different approaches for sequencing, short and long read sequencing(Burgess, 2018). Short read sequencing usually produces read from 50 to 300 base pairs and generate millions to billions of short reads in a single sequencing run. It is more cost effective than long read sequencing and normally provides higher quality reads with fewer errors (Pearman et al., 2020), but needs sophisticated algorithms (Delcher et al., 2002; Marschall et al., 2018) to assemble the reads into contigs or genomes. Long read sequencing can produce several thousand base pairs for each read. It is good for assembling complex genomes with repetitive regions and structural variants, and helps in understanding genetic variation(Amarasinghe et al., 2020; Pollard et al., 2018), but the sequences themselves have higher error rates. Metagenomics provides a comprehensive approach to studying AMR. It allows the identification of a wide range of different ARGs present in the sample and monitors the spread of the ARGs in various environments like hospitals, fields, and wastewater. Also, it helps to explore the mechanisms of developing AMR and discovery for unknow ARGs(L. Zhang et al., 2021). Metagenomics has been used to examine river sediments for pharmaceutical (specifically antibiotic) wastewater discharges. High levels of ARGs, transposons, plasmids, and integrons, was found in water(Kristiansson et al., 2011). Metagenomics is now widely used for examining migrate in the resistance through wastewater treating plants (Y. Yang et al., 2014) and detecting the mobile genetic elements in order to observe the horizontal gene transfer. (Majeed et al., 2021).

Metagenomics has the advantage of not relying on the cultivation of microorganisms: it directly extracts and analyzes the DNA or RNA from a complex sample containing a mixture of microorganisms with the presence and abundance of various genes. Thus is it not restricted to culturable and viable cells as is culture-based techniques. By analyzing the arrangement of genes and their functional context, metagenomics helps to identify the metabolic pathways and the antibiotics resistance gene within the microbial community(Chen et al., 2022; Pearman et al., 2020). The main disadvantage of metagenomics is that the data of metagenomics is complex and can be challenging when assembling the genome from the short-read sequencing(Chen et al., 2022). For long read sequencing, it can aid in resolving complex genomic regions but its higher error rates will come with another problems(Amarasinghe et al., 2020). Bioinformatics skills and computational resources are usually required for the analysis, but the outcomes can be highly sensitive to the methods used (Jünemann et al., 2017; Navgire et al., 2022). Moreover, metagenomics sequencing reveals what genes are in the sample, but that does not mean that they are active or expressed.

## 1.5 Difficulty in visualizing patterns of AMR

With large-scale AMR data, patterns of resistance can be used for studying the process of AMR development and form hypotheses about potential selection pressures (Baker et al., 2022a). AMR data often have high dimensions and many variables or features. Indeed, sometimes large-scale patterns of resistances are unclear. For example, the analysis of 150 studies containing 1594 samples of the level of ARGs measured by qPCR(Abramova et al., 2023) revealed a lack of clarity about whether ARG abundances increased over time. Visualizing data in more than three dimensions is hard for human brain so dimensionality reduction methods (Vogelstein et al., 2021) can help to reduce the number features and present the relationships within the data. Thus, improved visualization of AMR data could be useful for exploring and explaining patterns of resistance ARGs or AMR phenotypes in the environment. Visualization of resulting patterns can then relate the development of

antibiotic resistance with, for example, the amount of antibiotic used in different time intervals. Co-selection also drives AMR development.(Baker-Austin et al., 2006), and large-scale visualization can help identify potential coselective conditions, either between antibiotics, or with other non-antibiotic molecules(Arya et al., 2021; Ibrahim et al., 2023). However, these analyses cannot fully explain what drives AMR development for a set of environmental samples because of the presence of undetected molecules, as the detection of environmental pollutants for AMR only detects the presence of antibiotics or other antimicrobials, leaving other molecules not detected as well as with uncertain bioactivity.

## 1.6 Importance of this thesis

This research is going to investigate the following two questions. First, to address the question of how large volumes of the AMR data can be effectively visualized, by comparing different high-dimensional reduction methods to score for the best method and results for further analysis. Second, to address whether selection from antibiotic metabolites, which are not currently considered for environmental detection, could drive AMR development, using computational predictions. In this second question, we will predict whether or not metabolites are bioactive towards bacteria. These results could be important for environmental surveillance for antibiotics.

# Chapter 2

## Methodological background

In this thesis, patterns of antimicrobial resistance are studied though computational methods. This chapter provides the background to these methodologies. Resistance profile visualization was carried out using five high-dimensional reduction methods (Principal component analysis (PCA), metric multidimensional scaling (MDS), t-distributed stochastic neighbor embedding (t-SNE), Isomap, and Potential of Heat Diffusion of Affinity-Base Transition Embedding (PHATE)). Antibiotic metabolite bioactivity is predicted using two computational methods: molecular docking and molecular dynamics (MD) simulations - to predict the binding of antibiotic metabolites to their target binding sites. This chapter also provides a background to the antibiotics studied in Chapters 4 and 5, and specifically the break-down processes and products of penicillins.

## 2.1 Introduction to high-dimensional reduction methods

The increase in both the volume and dimensionality of data often makes them too complex and difficult to understand or process. The aim of dimensionality reduction is to preserve as much of the relational structure of the high-dimensional data as possible in a low-dimensional space. Different techniques have been proposed that differ in the type of structure they preserve: traditional dimensional reduction techniques such as PCA and MDS are linear techniques that focus on keeping dissimilar data far apart in the low dimensional space. Alternatively, dimensional reduction techniques such as Isomap, t-SNE and PHATE are non-linear techniques to transform the high dimensional data to low dimensions.

### 2.1.1 PCA

PCA (Wold et al., 1987) is a widely used unsupervised technique that reduces high dimensionality data to a set of new variables which simplifies the visualization of complex data sets for analysis. The smaller set of new variables can be used with classification techniques that need fewer variables than samples. In PCA, the new variables (principal components) are linear combinations of the original variables extracted. The results of PCA are visualized as projections of multidimensional scores and loadings in the plots.

### 2.1.2 MDS

MDS (Mead, 1992) includes a wide variety of statistical techniques aimed at characterizing structure within a set of data. The most common uses of MDS are to uncover the dimensionality of a set of data and to visually display the placements of features according to their positions on the dimension which is the similarities problem. Such spatial maps can help understand the structure of certain types of decision-making metrics accepted in various settings. MDS is a means of visualizing the level of similarity between data points by finding positions for the data in a low dimensional Cartesian coordinate system that as close as possible preserves the distances between the equivalent data points in the original high dimensional space.

### 2.1.3 Isomap

Isomap (Tenenbaum et al., 2000) is an isometric mapping method that extends metric multidimensional scaling (MDS) by using the geodesic distances imposed by a weighted graph. The employed classical scaling of metric MDS performs low-dimensional embedding based on the pairwise distance between data points, which is measured using straight-line Euclidean distance.

Isomap is distinguished because of the use of the geodesic distance, it was induced by a neighborhood graph embedded in the classical scaling to incorporate manifold structure in

the resulting embedding. Isomap defines the geodesic distance to be the sum of edge weights along the shortest path between two nodes. The top $n$ eigenvectors of the geodesic distance matrix, represent the coordinates in the new $n$-dimensional Euclidean space. The calculation of Isomap is similar to MDS but Isomap does not calculate the Euclidean distances but instead use geodesic distances (H. Yang et al., 2019)

### 2.1.4 t-SNE

In the t-SNE (Maaten & Hinton, 2008) method, Gaussian probability distributions over the high-dimensional space are constructed and used to optimize a t-distribution in low-dimensional space. The low-dimensional embedding descriptors can be obtained by minimizing the Kullback–Leibler divergence between the distributions on high- and low-dimensional spaces using the gradient descent algorithm. Again, similar to MDS, in the t-SNE method, the low-dimensional space maintains the pair-wise similarity to the high-dimensional space, leading to a clustering on the embedding space close to the clustering in the high-dimensional space with minimum loss of structural information.

### 2.1.5 PHATE

PHATE (Moon et al., 2019) provides a denoised, two or three-dimensional visualization of the complete branching trajectory structure in high-dimensional data. It uses heat-diffusion processes to denoise the data and to compute cell-cell affinities. Then, PHATE creates a diffusion-potential geometry by free-energy potentials of these processes. This geometry captures high-dimensional trajectory structures, while enabling a natural embedding of the intrinsic data geometry. This embedding accurately visualizes trajectories and data distances, without requiring strict assumptions.

### 2.1.6 Scoring for the dimensional reduction methods

After using each high-dimensional reduction method to process the data, the high-dimensional dataset was transferred to 2-dimensional plot. The $k$-nearest neighbour (KNN)

algorithm was used to predict the boundaries of the regions in the 2-D space associated with resistance to each of the antibiotics. The KNN algorithm (Z. Zhang, 2016) is a supervised machine learning algorithm used for classification purposes. consideringThe KNN predicts the data classification by considering the features and labels of the training data. Generally, the KNN algorithm can classify datasets using a training model similar to the testing query by considering the *k* nearest training data points (neighbours), which are the closest to the query it is testing. Finally, the training data of the KNN algorithm is used to compare the predicted regions with the actual data and calculate the corresponding score of the dimensional reduction method by the log-likelihood function.

$$\text{Log likelihood function} = \sum \log (Pr(Y\_i=y\_i \mid X\_i = x\_i))$$

## 2.2 Introduction to molecular docking and Genetic optimization for ligand docking (GOLD)

Molecular docking (N. Huang et al., 2006) can be used to model the interaction between a small molecule (ligand) and a protein at the atomic level, which allows for the characterization of the behavior of small molecules in the binding site of target proteins, and description of the resulting complex. Docking can be achieved through two interrelated steps: first by sampling conformations of the ligand in the active site of the protein; then ranking these conformations via a scoring function. The first of these steps has two substeps: prediction of the ligand's position and orientation within the active site, and assessment of the binding affinity.

The first molecular docking algorithm was developed in the 1980s by Kuntz et al.(Kuntz et al., 1982). Subsequently, Fourier transformation was included into the models to describe molecules as a digital model, allowing their interior and exterior parts to be distinguished. Fourier transformation allows faster calculation by determining the surface of contact, overlap, and approximation using the six degrees of freedom (roll, pitch, yaw, surge, heave,

sway)(Duhamel & Vetterli, 1990). In this method, molecules are considered as a rigid body and the changes in structure have degrees of freedom. However, this method proved erratic and ineffective(Ponomareva et al., 2018). To resolve this, flexible docking was developed to offer a more precise technique to describe flexible bodies undergoing rotational conformation, rotation, and translational changes, which mimicking the nature of biological molecules. GOLD uses a genetic algorithm to explore the conformational space of ligands within the active site of the protein (G. Jones et al., 1997). The algorithm involves two steps: prediction of the ligand position and orientation in the active site, followed by assessment of the binding affinity, which enables faster calculations by considering six degrees of freedom. The advantage of using GOLD is its ability to sample ligand conformations and rank with scoring function (G. Jones et al., 1995). Also, flexible docking is used to enhance precision, increasing the inherent flexibility of biological molecules through rotational conformation, rotation, and translational changes.

## 2.3 Introduction and history of molecular dynamics

The field of MD(Hollingsworth & Dror, 2018; Paquet & Viktor, 2015) has undergone significant advancements in recent decades and has become one of the most well-developed areas of theoretical studies in molecular biology. With the use of computer simulations, MD allows investigators to study molecular motions and interactions over time, significantly improving the cost-effectiveness of applied biological research (Şterbuleac, 2021). When used in conjunction with experimental designs, MD can help uncover new features of molecular interactions, providing valuable insights into the complex ways in which chemical compounds interact with biological macromolecules.(Ciccotti et al., 2022; Sedighpour & Taghizadeh, 2022; Şterbuleac, 2021)

The development of general-purpose graphics processing units (GP-GPU) and improved simulation algorithms have been instrumental in advancing MD computational methods(Rapaport, 2022). In contrast to the past, when supercomputers were required to

17

perform MD simulations, a modest workstation with low energy consumption is now sufficient to enable researchers to perform nanosecond simulations in a matter of hours. Moreover, there are now many tools available to aid computational biologists in the preparation, simulation, and analysis of molecular dynamics experiments. With the help of these tools, tasks such as proper membrane placement, ligand parameterization, and full simulation system building can be performed accurately in seconds to minutes, reducing the costs associated with time and computational power while accelerating MD-based studies (Şterbuleac, 2021).

The history of MD can be traced back to the early 1950s, when physicists and chemists began to develop computational methods for studying the behavior of molecules and materials at the atomic level. The first molecular dynamics simulations were performed using simple models of gases and liquids, and were primarily used to study the thermodynamics of these systems.(Hollingsworth & Dror, 2018)

In the 1960s and 1970s, advances in computer technology and the development of new algorithms allowed researchers to perform more complex molecular dynamics simulations. One of the key breakthroughs during this period was the development of the Verlet algorithm (section 2.3.2)(Hairer et al., 2003), which is still widely used in molecular dynamics simulations today.

In the 1980s and 1990s, molecular dynamics simulations became increasingly popular in chemistry and materials science, as researchers began to apply the technique to study more complex systems, such as proteins and materials. During this period, new potential energy functions were developed to model the interactions between atoms and molecules, and new algorithms were developed to simulate the behavior of these systems over longer time periods.(Ciccotti et al., 2022)

## 2.3.1 The equation used in classical MD

In classical MD simulations, Newton's equations of motion are used to describe the interactions between atoms and molecules in a system (Kadupitiya et al., 2022). The positions, velocities, and accelerations of each atom or molecule are calculated at each time step based on the forces acting on them, which are determined by the interactions with neighboring atoms or molecules. This allows researchers to predict the behavior of a system over time and to investigate the properties of materials and molecules at the atomic level.

The first and second derivation of position correspond to the velocity and the acceleration:

$$\frac{dr(t)}{dt} = v(t)$$

$$\frac{d^2r(t)}{dt^2} = a(t)$$

and the acceleration is related to the force of Newton's second law.

$$F = ma$$

The algorithm that predicts positions of atoms at a later point in time in NAND is the Verlet algorithm(Hairer et al., 2003; Spreiter & Walter, 1999). This algorithm approximates the position of particles at by using a Taylor series approximation.

$$r(t + \Delta t) = r(t) + \frac{dr(t)}{dt}(t + \Delta t - t) + \frac{1}{2}\frac{d^2r(t)}{dt^2}(t + \Delta t - t)^2$$

$$r(t + \Delta t) = r(t) + v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2$$

$$r(t - \Delta t) = r(t) + \frac{dr(t)}{dt}(t - \Delta t - t) + \frac{1}{2}\frac{d^2r(t)}{dt^2}(t - \Delta t - t)^2$$

$$r(t - \Delta t) = r(t) - v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2$$

Taking the sum of the positions at previous and next step:

$$r(t + \Delta t) = 2r(t) - r(t - \Delta t) + \frac{d^2r(t)}{dt^2}(\Delta t)^2$$

By substituting $\frac{d^2r(t)}{dt^2} = \frac{F(t)}{m}$

$$r(t + \Delta t) = 2r(t) - r(t - \Delta t) + \frac{F(t)}{m}(\Delta t)^2$$

The initial coordinates of the atoms in the system specify the positions of the atoms within the molecules and state the starting point for the simulation. It is usually obtained from x, y,

z-coordinates of the protein structure. The initial velocities are usually consistent with the temperature of the system which is controlled by using a thermostat. The initial velocities can also be obtained from the Maxwell-Boltzmann distribution which describes the distribution of velocities at a given temperature(Hollingsworth & Dror, 2018).

## 2.3.2 Control for temperature and pressure in MD

The temperature of the system is controlled by the thermostat which to maintain the desired thermodynamic conditions(Yong & Zhang, 2013). A Langevin thermostat is used in this study: this algorithm applies forces to the simulated particles to adjust the velocities, ensuring the system remains at a target temperature throughout the simulation. The Langevin thermostat introduces a damping term and a random force to the equations of motion to control the temperature of the system (Farago, 2019; Jakobsen, 2005)

$$m\frac{d\text{v}}{dt} = -\gamma\text{v} + \sqrt{2\gamma k_B T}\,\text{R(t)}$$

Where $m$ is the mass of the particle, v is the velocity vector of the particle, $\gamma$ is the damping coefficient, $k_B$ is the Boltzmann constant, $T$ is the temperature of the system, R(t) is a random force vector at time t with zero mean and unit variance.

The system maintains the desired pressure with the use of barostats. A Langevin piston barostat is used in this study to control the pressure of the system(Jakobsen, 2005). This algorithm uses Langevin dynamics (Hoover, 1986) to regulate the volume of the system, therefore, maintaining the pressure(Paquet & Viktor, 2015).

$$\frac{d\text{V}}{dt} = \frac{1}{W}[P(t) - P_0] - \gamma\text{V} + \sqrt{\frac{2\gamma k_B T}{W}}\text{R(t)}$$

Where $W$ is the mass of the piston, $P(t)$ is the pressure at time t and $P_0$ is the external pressure acting on the piston, V is the volume of the particles, $\gamma$ is the damping coefficient, $k_B$ is the Boltzmann constant, $T$ is the temperature of the system, R(t) is a random noise vector at time t with zero mean and unit variance.

### 2.3.3 Force field of MD

The force field is the combination of mathematical formula and associated parameters to describe the energy of a protein in related to its atomic coordinates. It consists of a set of the interatomic potential energy (U). The parameters are usually obtained either from *ab initio* or semi-empirical quantum mechanical calculations or by fitting to experimental data. The force field is calculated from the intramolecular total energy (bond stretching, angle bending, dihedral torsion and improper torsion), Van der Waals interactions and the Coulombic interactions. The Lennard-Jones potential is used to describe the attractive and repulsive forces between two molecules. It consists of two terms: a short-range attractive (Van der Waals) and a long-range repulsive term (steric hindrance). The Coulombic interactions describes the interactions between charged particles and is based on the Coulomb law which states the force between two charges is proportional to the product of their charges and inversely proportional to the square of the distance between them. (González, 2011)

$$U = energy\ of\ bonding\ stretching + energy\ of\ angle\ bending$$
$$+\ energy\ of\ dihedral\ torsion + energy\ of\ improper\ torsion$$
$$+\ Van\ der\ Waals\ interactions + Coulombic\ interactions$$

$$U = \sum_{bond\ streching} \frac{1}{2} k_b (r - r_0)^2 + + \sum_{angles\ bending} \frac{1}{2} k_a (\theta - \theta_0)^2$$

$$+ \sum_{dihedral\ torsions} \frac{V_n}{2} [1 + \cos(n\phi - \delta)] + \sum_{improper\ torsions} V_{imp}$$

$$+ \sum_{LJ\ potential} 4\epsilon_{ij} \left( \frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^{6}}{r_{ij}^{6}} \right) + \sum_{Coulombic\ interactions} \frac{q_i q_j}{r_{ij}}$$

CHARMM and Amber are the most widely used computational implementations of force fields in MD. CHAARMM and Amber have their own advantages and some specific functions. CHARMM has includes a large collection of similarly parameterized small molecules which can be directly used for simulation(Vanommeslaeghe et al., 2010). The Amber force field is not a parameter set but rather a software package designed to generate the force field model for the input molecules(Ponder & Case, 2003). In this study, the force field used is CHARMM because of its relevant analysis capability.

## 2.3.4 Water model

Water is a simple but essential molecule involves in a wide range of biological chemical processes and it influences temperature, pressure, and the presence of other molecules(Dill et al., 2005; Finney et al., 2004). The mathematical representation of water molecules is intended to improve reliability of results by capturing the physical properties and behaviour of water molecules in a simulation. When the system contains water, a water model is needed to represent the behaviour of water molecules and their interactions with other molecules within the system.(Finney, 2001; Izadi et al., 2014).

There are several different water models that can be used in MD simulations, for example, TIP3P, TIP4P, CHARMM modified water model. TIP3P is based on the assumption that water molecules can be represented as three points, two hydrogen atoms and one oxygen atom. The TIP3P model includes a Lennard-Jones potential for the interaction between the oxygen and hydrogen atoms of adjacent water molecules and a Coulombic interaction to describe the interaction between charged particles (Mark & Nilsson, 2001). The TIP4P model includes an additional point charge, which is usually massless, at the center of the water molecule(Fuentes-Azcatl & Alejandre, 2014). CHARMM modified water model is part of the CHARMM force field. The CHARMM force field includes a set of parameters to describe the interactions of water molecules.(Gil Pineda et al., 2020)

## 2.3.5 Setting up a typical MD

The simulation is set up through input files that define the system, force field parameters, and simulation settings. The Protein Data Bank (PDB) includes the initial structure of the system. CHARMM is the choice for the force field, and the force field parameters of the ligand also need to be provided. Counterions should be added, and the concentration of ions in the solvent adjusted to reach the desired pH. The corresponding water model (TIP3P) should be added to the system. The SETTLE algorithm is enabled to constrain the water molecules. The system first undergoes minimization until it reaches equilibrium, and then the corresponding ensemble equilibration (NVT and NPT) is used to reach equilibrium for production (Phillips et al., 2005).

## 2.4 Introduction to antibiotics

There are several classes of commonly prescribed antibiotics, including penicillins, cephalosporins, tetracyclines, aminoglycosides, quinolones, and macrolides, each characterized by distinct chemical structures and mechanisms of action that target various bacterial macromolecules. This thesis focuses on penicillins, a class of beta-lactam antibiotics. Beta-lactam antibiotics target penicillin-binding proteins (PBPs), obstructing the binding of D-alanyl-D-alanine, a small peptide involved in bacterial cell wall synthesis, onto the PBP. This inhibition prevents the formation of the cell wall.

Cephalosporins, also belonging to the beta-lactam family, similarly target PBPs and share a binding mechanism with penicillins. Tetracyclines and aminoglycosides both target the 30S ribosome subunit but bind to different positions. Tetracycline binds to the A-site of the 30S ribosome, hindering the binding of aminoacyl-tRNA to the A-site and, consequently, impeding protein formation (Chopra & Roberts, 2001). Aminoglycosides bind to the start codon of the 30S ribosome, forming a complex that prevents the initiation of protein synthesis (Kotra et al., 2000).

Quinolones, classified as chemotherapeutic bactericidal drugs, interfere with DNA replication by impeding bacterial DNA unwinding and duplication. Specifically, they inhibit the ligase activity of type II topoisomerases, DNA gyrase, and topoisomerase IV. This disruption leads to the release of DNA with single- and double-strand breaks, resulting in cell death.

The mechanism of action of macrolides centers on their ability to bind to the bacterial 50S ribosomal subunit, causing the cessation of bacterial protein synthesis. Once bound, the drug prevents the translation of mRNA, specifically the growing peptide chain, by inhibiting the enzyme peptidyltransferase from adding the subsequent amino acid attached to the tRNA.

## 2.5 The choice of antibiotic and experimental bacterium

Penicillins are chosen in this study for three reasons. Firstly, penicillins are important because they are one of the most used antibiotics for human medicine (52% of Defined Daily Doses of total antibiotic used) as well as widely in agriculture. Secondly, the chemical structure of penicillin metabolites is well known because of the well characterized degradation pathways of beta-lactams (Deshpande et al., 2004). Thirdly, the availability of crystal structures of both piperacillin and (5*S*)-penicilloic acid to PBP3 (Van Berkel et al., 2013) with resolution 2.31 Å, providing both an effective starting point for computational investigation of other metabolites within the penicillin family as well as a clear evidence that molecular interactions between metabolites and antibiotic targets are possible.

There are many bacterial phyla including firmicutes, proteobacteria, actinobacteria, bacteroidetes, chlamydiate, spirochaetes, cyanobacteria, tenericutes. The work in this thesis focusses on *Pseudomonas aeruginosa,* which is a Gram-negative, rod-shaped bacterium known for its adaptability, and which is commonly found in different environments including soil, water, and on the surfaces of plants and animals (Radó et al., 2017). *P. aeruginosa* is chosen because it causes disease in plants, animals, humans, and most importantly the serious infections on patients with immunocompromised cancer, cystic fibrosis, or severe burns(Fujii et al., 2014); importantly, *P. aeruginosa* strains can be either sensitive and resistant to beta-lactam antibiotics.

PBPs are classified into high molecular mass and low molecular mass. High molecular mass PBPs are classified by the number of reactions that they can catalyze (Ropy et al., 2015a). The bifunctional enzyme (class A) catalyzes both the glycosyltransfer and transpeptidation, while the monofunctional enzyme (class B) only perform transpeptidation(Haenni et al., 2006). They are serine acyltransferases which catalyze the formation of cross-linked peptidoglycan which is an essential macromolecule surrounding the bacteria. The low molecular mass PBPs (class C) generally function as carboxypeptidases or endopeptidases and are typically genetically deleted without having a significant effect on cell viability or growth(Macheboeuf et al., 2006). The PBPs are

numbered in the order of decreasing molecular weight in a given organism, so there is no relationship between the same numbered PBPs of two unrelated organisms (e.g., PBP-2 of *E. coli* and PBP-2 *of P. Aeruginosa*). In *P. Aeruginosa*, PBPs 1a and 1b are class A; PBPs 2, 3 and 3a are class B; PBPs 4 and 5 are class C (Ropy et al., 2015a). In E. coli, PBPs 1a, 1b and 1c are class A; PBPs 2 and 3 are class B; PBPs 4, 5, 6, 6b and 7 are class C (Kocaoglu & Carlson, 2015). The PBP of *P. Aeruginosa* show close phylogenetic correlation to the PBP of E. coli. (Figure 2.1)
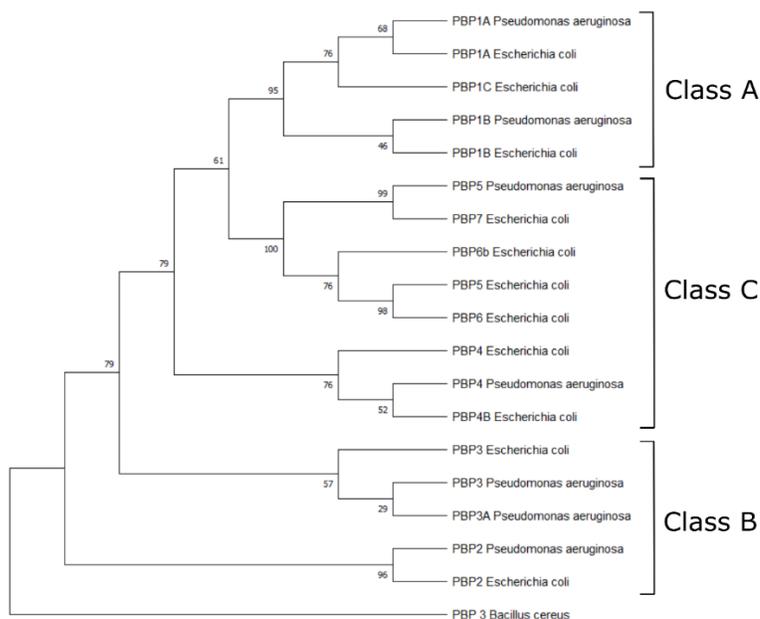


**Figure 2.1** Phylogenetic tree showing the correlation between PBP of *P. Aeruginosa* and *E. coli*. The outgroup is the PBP3 of *Bacillus cereus*. In *P. Aeruginosa*, PBPs 1a and 1b are class A; PBPs 2, 3 and 3a are class B; PBPs 4 and 5 are class C. In *E. coli*, PBPs 1a, 1b and 1c are class A; PBPs 2 and 3 are class B; PBPs 4, 5, 6, 6b and 7 are class C.

## 2.5.1 Mechanism of beta-lactam inhibition of PBPs

PBPs are essential for cell wall biosynthesis. Class A catalyzes both the glycosyltransfer and transpeptidation and class B only performs transpeptidation. For transpeptidation, PBP recognize the D-alanyl-D-alanine on the main stem of the copolymer of N-acetylglucosamine and N-acetylmuramic acid which are polymerized by class A PBP

(glycosyltransfer), and catalyze the attack of the amino group of the adjacent chain to the carboxylic group of D-alanine the forming peptidoglycan (Horcajo et al., 2012). The catalytic mechanism of transpeptidation (Figure 2.2) begins with the attack of the nucleophilic serine of the PBP onto the amide linkage between the two D-alanines (Horcajo et al., 2012). This leads to the breaking of the amide bond, releasing the D-alanine and forming an ester intermediate between the PBP serine and the main peptide chain. The amino group on L-lysine of the adjacent chain attacks the ester linkage of the intermediate, forms a new amide bond, and releases the PBP serine. The two chains are linked together under 3, 4-linkage and formed peptidoglycan.



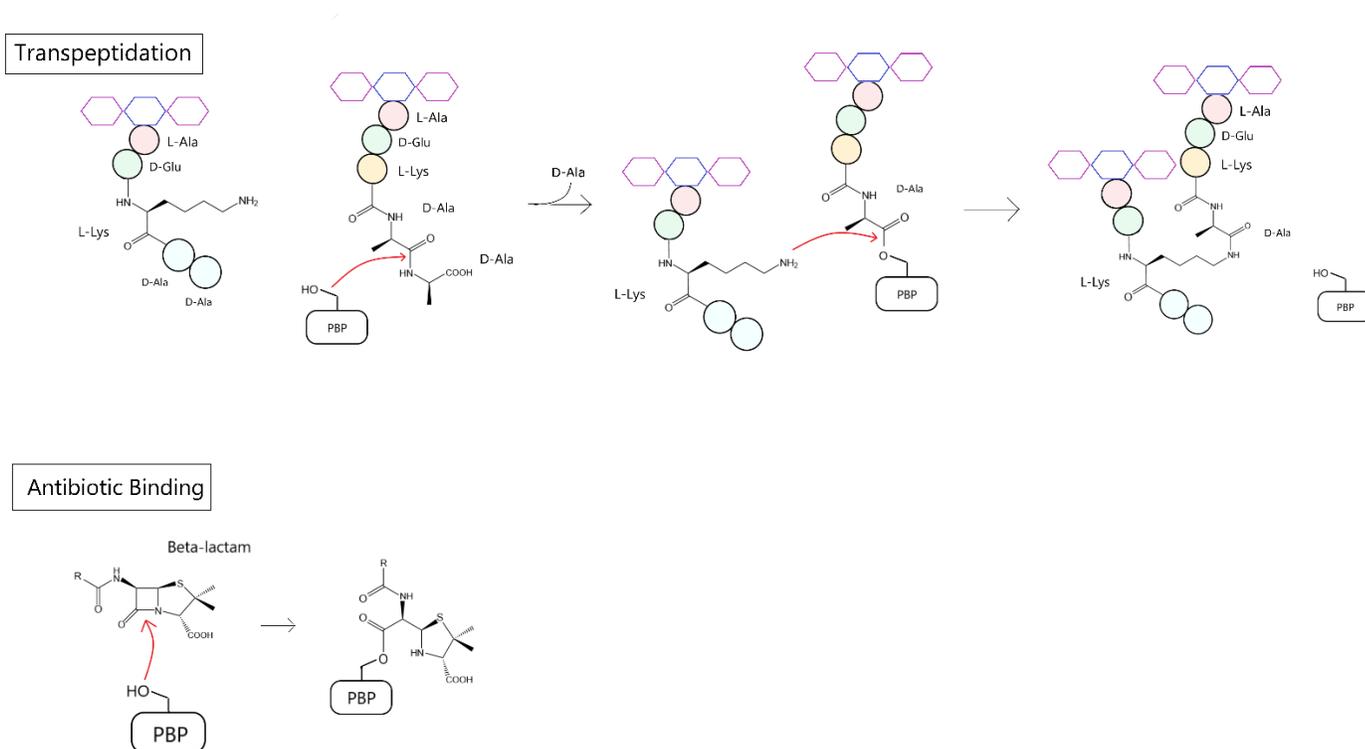**Figure 2.2** The mechanisms of PBP transpeptidation and beta-lactam binding. Only part of the glycan backbone (hexagons) is shown. The purple hexagons are N-acetylglucosamine; the blue hexagon is N-acetylmuramic acid. The red circle is L-alanine, green circle is D-glutamic acid, yellow circle is L-lysine, and blue circle is D-alanine. The -OH group on the PBP is the -OH group on the active serine of the PBP.

Beta-lactam antibiotics target PBPs because their structures are similar to their normal target, D-alanyl-D-alanine. The active serine of the PBP attacks the beta-lactam ring of the beta-lactams, leading to acetylation of the PBP serine by the beta-lactam (Figure 2.2). This covalent complex is stable and prevents the catalytic transpeptidation of the peptidoglycan by the PBP.

## 2.5.2 The choice of PBP used in the study

The PBP used in the study is PBP3 of *P. Aeruginosa*. This PBP3 is chosen as a class B enzyme (Ropy et al., 2015b) that performs transpeptidation and is inhibited by penicillin (M. Lee et al., 2003), in contrast with PBP4 as a class C enzyme which only act as carboxypeptidases or endopeptidases. This lays the ground for further detailed simulations of metabolite binding to other PBP classes (class A or class C), or in other organisms. Importantly, there are four Protein Data Bank (PDB) entries for this protein complexed with antibiotics: piperacillin 4KQO(Van Berkel et al., 2013); amoxicillin 6I1E (Bellini et al., 2019); ceftazidime 3OCL (Sainsbury et al., 2011); and cefoperazone 5DF9(Ren et al., 2016).

The entry 4KQO was chosen as the initial structure of the simulations (Figure 2.3) for four reasons: first, the structure is bound to a penicillin class antibiotic; second, in addition to the antibiotic, there is also a structure of the PBP3 bound to the beta-lactam metabolite (5*S*)-penicilloic acid, providing a useful positive control metabolite; third, the structure is of the wild type protein rather than a mutated version; and fourth, the structure has high resolution of 2.31 Å. In comparison, 3OCL had a similar resolution 2.30 Å to 4KQO, but the structure was bound with ceftazidime which is a cephalosporin and it was not the target antibiotics class; 5DF9 has a lower resolution of 2.70 Å; and while 6I1E had the highest resolution of 1.64 Å, the amoxicillin was bound to a mutant strain. There are also 573 other PDB entries for PBP3 of *P. Aeruginosa* in which the protein is bound with non-antibiotic small molecules, or uncomplexed structures from sensitive or resistant strains.

The function of this PBP is performed by a single domain which is made of two subdomains: a five stranded beta-sheet covered by three alpha-helices and an all-helical domain. The active site sits on the interface of the two subdomains (Sauvage et al., 2008). The five amino acids Ser294, Ser349, Ser485, Thr487 and Tyr503 are known to be the key residues that stabilize the antibiotic-PBP interaction (Figure 2.3b-c) (Mora-Ochomogo & Lohans, 2021; Van Berkel et al., 2013). In particular, Ser294 is the most important residue for the antibiotic-PBP interaction, as this is the residue that forms the covalent bond with the beta-lactam ring(Mora-Ochomogo & Lohans, 2021).



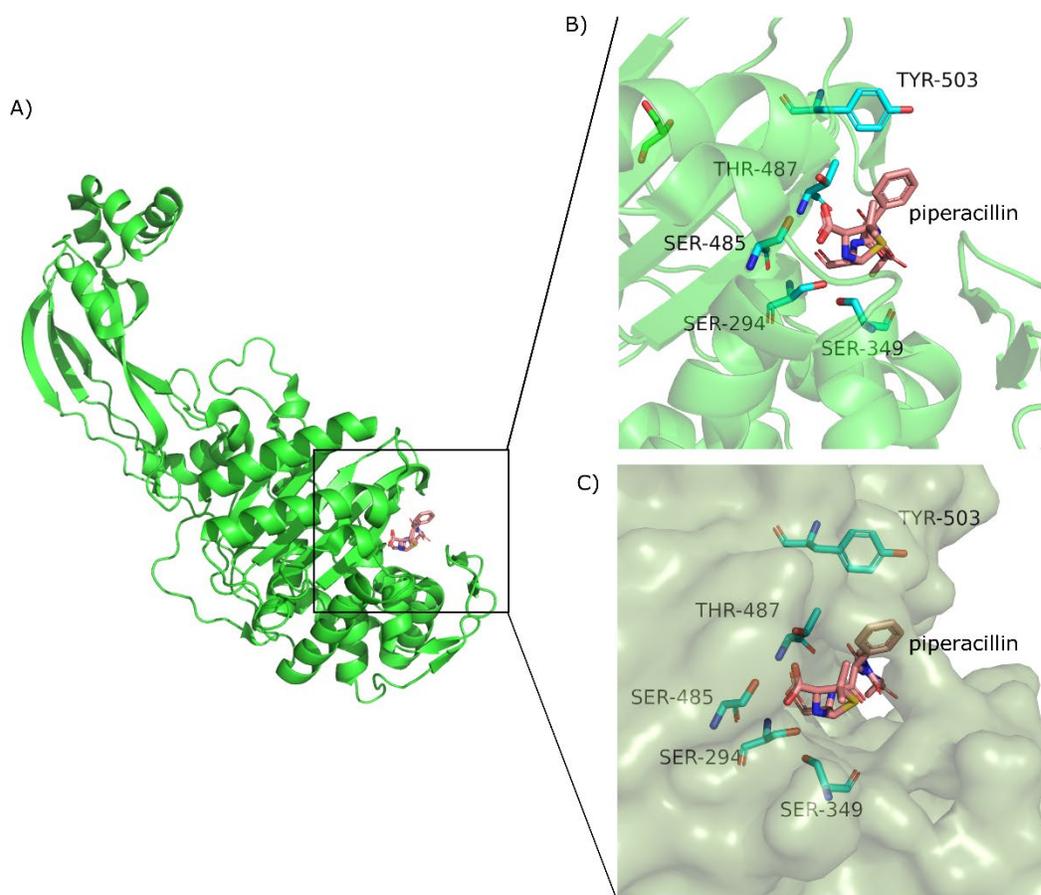**Figure 2.3** a) The structure of PBP3 of *P. Aeruginosa* from PDB 4KQO. The antibiotic piperacillin (pink molecule) is attached to the PBP3 (green molecules) within the active site of the enzyme. b-c) The structure of PBP3 (green molecule) of *P. Aeruginosa* with the antibiotic (pink molecule) with the key residues (light blue stacks), Ser294, Ser349, Ser485, Thr487 and Tyr503 forming the binding site.

## 2.6 Degradation of beta-lactam antibiotics

Degradation of penicillin takes place in a wide range of conditions, both alkaline or acidic (Figure 2.4)(Idsoe et al., 1968; Schwartz, 1965), in the presence or absence of the enzyme b-lactamase, or under the action of weak electrophiles including water and metal ions. Penicillin undergoes further isomerization to penicillenic acid, and the beta-lactam ring and its amide bond break open in the presence of acid giving an array of products, including penilloic acid, penicillamine and penilloaldehyde, through intermediates, namely penillic acid, penicilloic acid and penicillenic acid. (Aldeek et al., 2016; Deshpande et al., 2004).
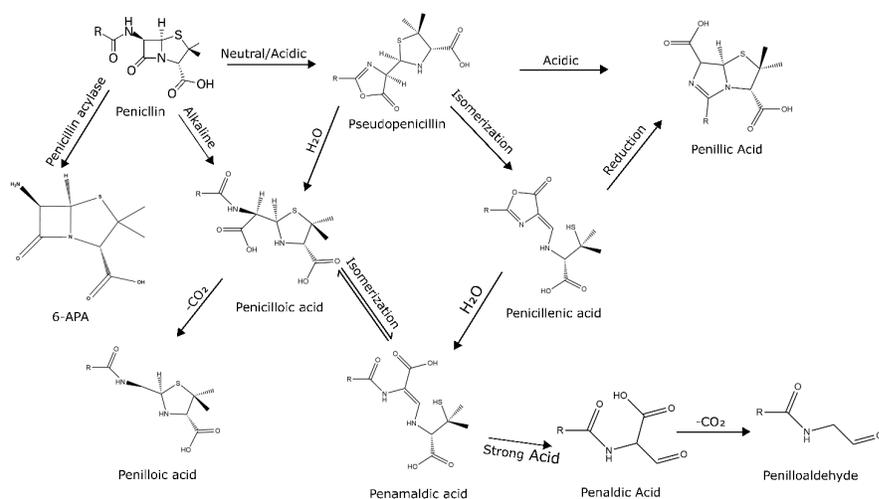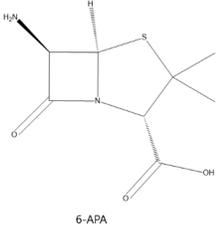


**Figure 2.4**. Degradation pathways of penicillin family beta-lactam antibiotics. The penicillin will degrade into 6-APA (under penicillin acylase), pseudopenicillin (under neutral or acidic conditions), penicilloic acid (under alkaline conditions); pseudopenicillin further degrades to penillic acid (under acidic conditions) and penicillenic acid (isomerization from pseudopenicillin); penamaldic acid is formed from the isomerization of penicilloic acid and hydrolysis of penicillenic acid; penilloic acid is formed from penicilloic acid with the presence of carbon dioxide; penamaldic acid further degrades to penaldic acid then penilloaldhyde under strong acid.

Metabolites (table 2.1) can form in *R* and *S* stereoisomers since the oxidizing agent can attack from both the upper plane or the lower plane of the penicillin. These stereoisomers may have different effects on the affinity of the metabolites towards PBPs(Camacho-Muñoz et al., 2019). Cephalosporins are also beta-lactam antibiotics that target PBPs, and they

have same binding mechanism as penicillin, but they have different degradation pathways from penicillin leading to different metabolite products (Deshpande et al., 2004; Dinner, 1977), principally because the second ring in penicillins is a 5 atom thiazolidine group while in cephalosporins it is a 6 atom thiodine group (in both cases sharing the N with the beta lactam ring); the side chains off the second ring also differ between the two families.

| Common name | Chemical formula | | Citation |
|---|---|---|---|
| 6-APA | C8H12N2O3S | \n6-APA | Deshpande et al., 2004 |
| 5S-Penicilloic acid | C9H13N2O4S-R | \n5S-Penicilloic acid | Deshpande et al., 2004,\nAldeek et al., 2016 |
| 5R-Penicilloic acid | C9H13N2O4S-R | \n5R-Penicilloic acid | Deshpande et al., 2004,\nAldeek et al., 2016 |

| | | | |
|---|---|---|---|
| 5S-Pseudopenicillin | C9H11N2O4S-R | 5S-pseudopenicillin | Deshpande et al., 2004 |
| 5R-Pseudopenicillin | C9H11N2O4S-R | 5R-pseudopenicillin | Deshpande et al., 2004 |
| Penillic acid | C9H11N2O4S-R | Penillic acid | Deshpande et al., 2004, Aldeek et al., 2016 |
| Penicillenic acid | C9H10N2O4S-R | Penicillenic acid | Deshpande et al., 2004 |
| Penilloic acid | C8H13N2O4S-R | Penilloic acid | Deshpande et al., 2004 |

| Penamaldic acid | C9H13N2O4S-R |  Penamaldic acid | Deshpande et al., 2004 |
|---|---|---|---|
| Penaldic acid | C4H4NO4-R |  Penaldic acid | Deshpande et al., 2004 |
| Penilloaldehyde | C3H4NO2-R |  Penilloaldehyde | Deshpande et al., 2004 |
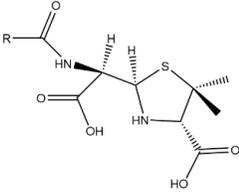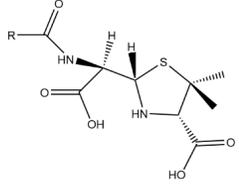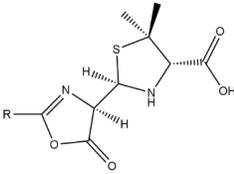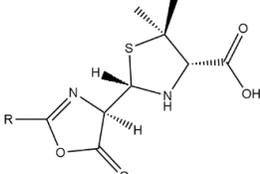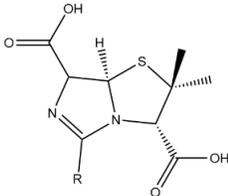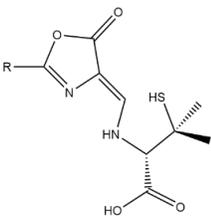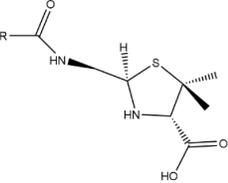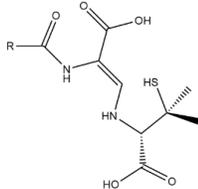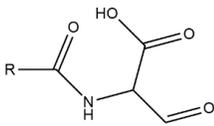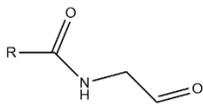
Table 2.1 Table showing the common names, chemical formulae and chemical structure of penicillin family metabolites.

# Chapter 3

## Comparison among dimensional reduction algorithms

Many AMR phenotype data sets are of a size where they are difficult to visualize. Thus, in order to visualize data effectively, some form of processing may be necessary. This chapter compares five different high-dimension reduction methods - PCA, MDS, isomap, t-SNE and PHATE - in order to find the most suitable method for visualizing such data in two-dimensional coordinate space plots. The aim of dimensionality reduction is to preserve as much of the underlying structure of the high-dimensional data as possible in the low-dimensional map. Various techniques for this problem have been proposed that differ in the type of structure they preserve. Traditional dimensional reduction techniques such as PCA and MDS are linear techniques that focus on keeping the low dimensional dissimilar datapoints far apart. In contrast, the dimensional reduction techniques such as Isomap, t-SNE and PHATE are non-linear techniques to transform the high dimensional data to a low dimensional space.

## 3.1 Importance of using high-dimensional reduction methods

Examples of high dimensional AMR phenotype data that are too complex to understand and difficult to process due to their size are shown in Figure 3.1. Humans are visual learners: it is important that these datasets are presented to researchers in intuitive ways to understand both the overall shape and the fine granular structure of the data. This is especially important in biological systems, where structure exists at many different scales and a faithful visualization can lead to hypothesis generation. Data visualization helps to tell stories by curating data into forms that are easier to understand, while highlighting the trends and outliers. A good visualization tells a story, removing the noise from data and highlighting the useful information. Therefore, data visualization requires the data to be simple which needed to reduce the dimensionality of the data.

**Figure 3.1**. Hierarchically-clustered heatmap of three study, A) Dairy slurry study from Baker *et al*. 2022(Baker et al., 2022); B) slurry-amended soil study form Alexander William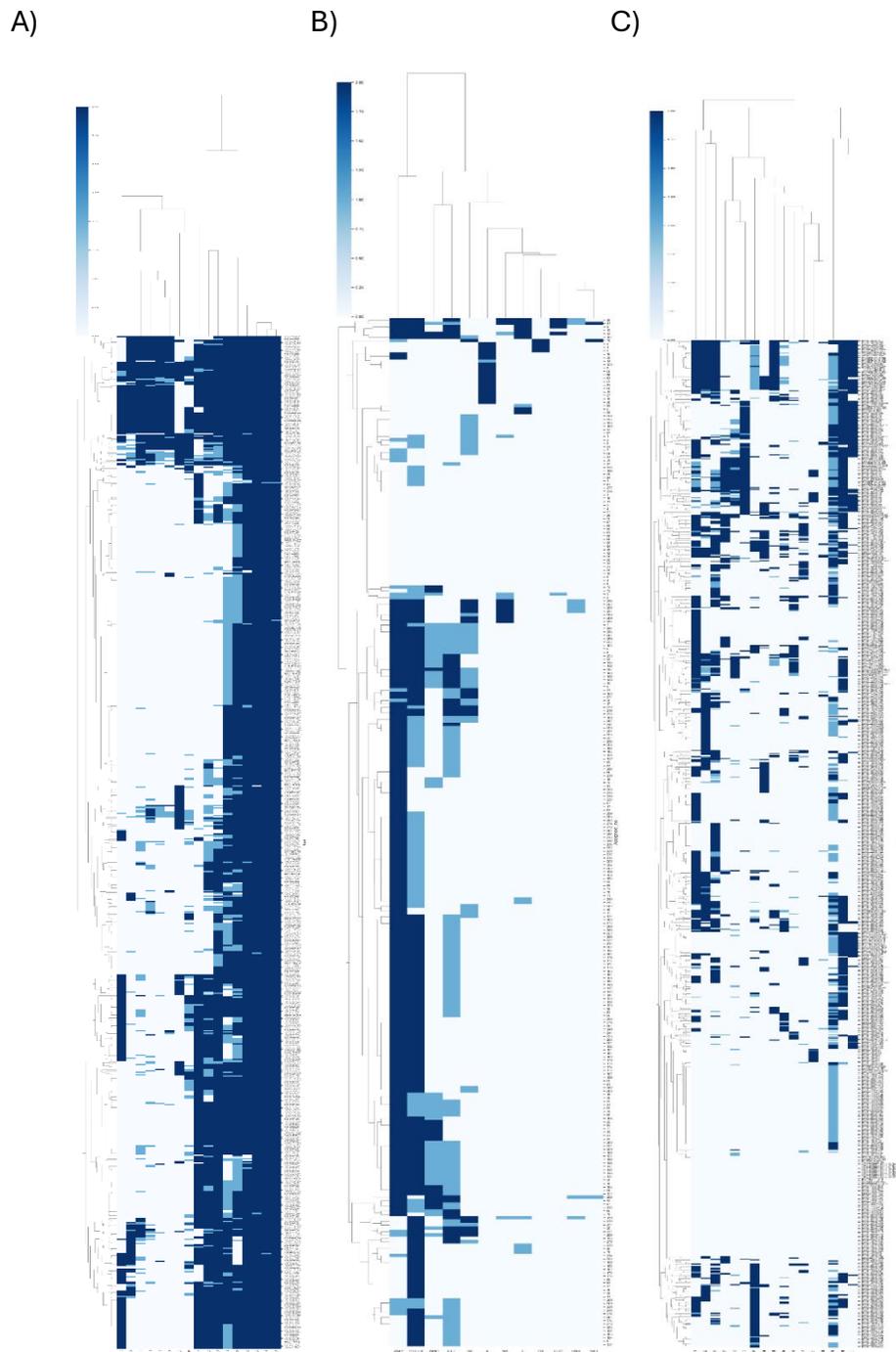s and Helen West; C) Bangladesh ONE-health study from Rousham *et al*. 2018(Rousham et al., 2018). Despite the appropriate use of hierarchical clustering, the size and complexity of the data make the graphs hard to read.

## 3.2 Database and function used

The data used in this chapter were from three different studies. Data set 1 is dairy slurry study from Baker et al., 2022(Baker et al., 2022). Data set 2 is slurry-amended soil study form Alexander Williams and Helen West; Data set 3 is Bangladesh ONE-health study from Rousham et al., 2018(Rousham et al., 2018). The data included the media of the plate, the date, and the place of sampling the type of antibiotic used, and the phenotype of the microbes when antibiotic was added. The datasets were then processed using PCA, MDS, isomap, t-SNE and PHATE with python 3.10.

The three studies studied same bacterial species, *Escherichia coli*, using the same antimicrobial sensitivity testing approach of disk diffusion assays (Baker et al., 2022; Rousham et al., 2018) with three levels of resistant resistance (sensitive to antibiotics, intermediate and resistant to antibiotics), but in each case using a different set of antibiotics, as described below.

The dairy slurry study includes 811 samples with 17 antibiotics (azithromycin, aztreonam, ampicillin, amoxicillin-clavulanic acid, cefpodoxime, chloramphenicol, cefotaxime, cefoxitin, ceftazidime, ciprofloxacin, colistin, imipenem, nalidixic acid, nitrofurantoin, streptomycin, tetracycline, trimethoprim-sulfamethoxazole) with three levels of resistance (sensitive to antibiotics, intermediate and resistant to antibiotics) are shown. The samples were collected from slurry tank on 20 different days (16/05/2017, 22/05/2017, 29/06/2017, 4/07/2017, 11/07/2017, 18/07/2017, 25/07/2017, 1/08/2017, 16/08/2017, 22/08/2017, 5/09/2017, 22/09/2017, 27/09/2017, 10/10/2017, 17/10/2017, 31/10/2017, 14/11/2017, 21/11/2017, 12/12/2017 and 25/01/2018).

The slurry-amended soil study contained 300 samples with 12 antibiotics (amoxicillin, ampicillin, ceftiofur, streptomycin, oxytetracycline, enrofloxacin, nalidixic acid, kanamycin, sulphamethoxazole trimethoprim, florfenicol, nitrofurantoin, cefalexin) with the same three levels of resistance (sensitive to antibiotics, intermediate and resistant to antibiotics). The samples were obtained on 5 different days (15/05/17, 7/09/17, 19/10/17, 10/01/18,

17/05/18) from the farm field (F31) in Sutton Bonington, and 66 samples were obtained from the Nottingham arboretum on 10/01/18 and 17/05/18. The isolates were grown in MacConkey agar plates and TBX agar plates.

The Bangladesh ONE-health study contains 787 samples with 17 different antibiotics (gentamycin, meropenem, ertapenem, imipenem, ceftriaxone, cefotaxime, ceftazimide, cefixime, cefepime, ampicillin, colistin sulphate, ciprofloxacin, nalidixic acid, cefoxitin, sulfamethoxazole, nitrofurantoin, piperacillin-tazobactam) with the same resistance levels as the other two data sets. The day of sampling was not recorded in these data. It contains samples from rural households, commercial farms and urban live bird markets.

Only four types of antibiotics were used commonly among the 12-17 kinds of antibiotics in the three different AMR studies (Figure 3.2). This showed the prevalence and the priority of the choice of antibiotics of different scientists.
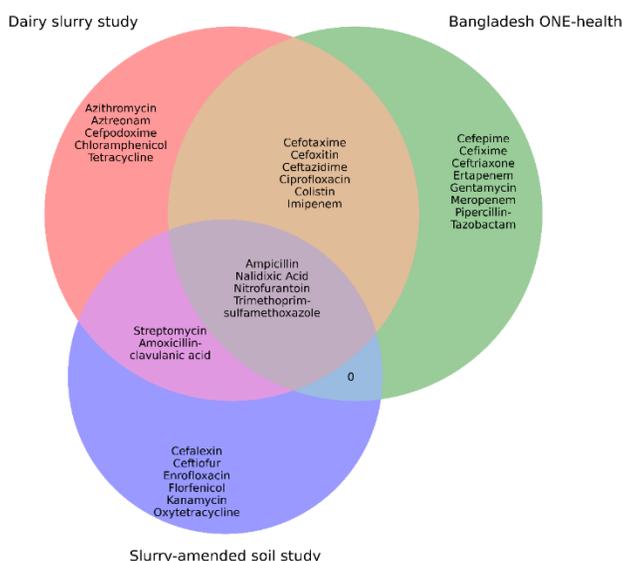


**Figure 3.2** Venn diagram showing all antibiotics used among the three AMR studies whose data are used in this chapter. The common antibiotics used for three study are ampicillin, nalidixic acid, nitrofurantoin, trimethoprim-sulfamethoxazole; 8 antibiotics were used in two studies. Each study had antibiotics unique to its analysis

## 3.2.1 High dimensional reduction methods

The high dimensional reduction methods used PCA, MDS, isomap, t-SNE and PHATE. Details of these methods are mentioned in Chapter 2.1. The package used for PCA, MDS, isomap, t-SNE are scikit-learn 1.3 (Pedregosa et al., 2011); for PHATE the package phate 1.0.11 (Moon et al., 2019) was used.

## 3.2.2 Use of KNN

KNN is a machine learning used for classification and regression. It is an instance-based learning algorithm that makes predictions based on the majority or the average values of the closest K data points to each input datum. KNN does not make any assumptions about the data distribution as it is a non-parametric algorithm.

KNN was used to define the region associated with resistance to each of the antibiotics with the two-dimensional data following data dimensionality reduction as part of the training phase and prediction phase. The data was split into training data (80%) and test data (20%). The best value of K was calculated by cross-validation (Rahim et al., 2022) with the highest average accuracy. The cross-validation randomly split the dataset into 10 groups. One of the groups was used as the test set and the rest were used as the training set. The model was trained on the training set and scored on the test set according to accuracy. Then the process was repeated until each unique group had been used as the test set and recording the accuracy and gave an average accuracy.

In the training phase, it stores all the training data points and their corresponding labels for classification. For prediction phase, the algorithm gives a new input data and calculates its distance to all training data points using a Euclidean distance as a distance metric. Each label's occurrences among the K neighbors are counted, and KNN assigns the class label with the highest count as the predicted class for the input data input.

After the KNN was done, the region was coloured according to the KNN prediction result. Then score could be used to quantify the prediction by log likelihood function. The score considered all data and was calculated according to whether the data (sensitive/ intermediate/ resistance) was located in the correct region. A score equal to zero would mean all data located in the correct regions; more data in incorrect regions would lead to a more negative score. The total score is calculated by the sum of the score of every antiobics.

$$\text{Log likelihood function} = \sum \log (Pr(Y_i = y_i \mid X_i = x_i))$$

## 3.3 Results of the high-dimensional reduction methods and scoring

The datasets processed by PCA, MDS, isomap, t-SNE and PHATE and transformed the data into two-dimensional and had a 2D plot (Figure 3.3, Figure 3.5, Figure 3.6). For each method, KNN was then used on each resulting scatterplot to classify the regions associated with sensitivity or resistance to each of the antibiotics. The score considered all data and was calculated according to whether the data (sensitive/ intermediate/ resistance) were located in the correct regions (as above) and calculated the log-likelihood function as the score of the method (Figure 3.4). From the graph, t-SNE was the most suitable method of the three studies. In the dairy slurry study, MDS and isomap had similar scorse while PCA did the worst clustering. PCA, MDS and isomap had similar scores on the slurry-amended soil study and the Bangladesh ONE-health study.

In PHATE, the same features of the bacteria could not be clustered into the same area. But each antibiotic resistance was clustered in the same region in PCA, MDS, isomap and t-SNE.

A)
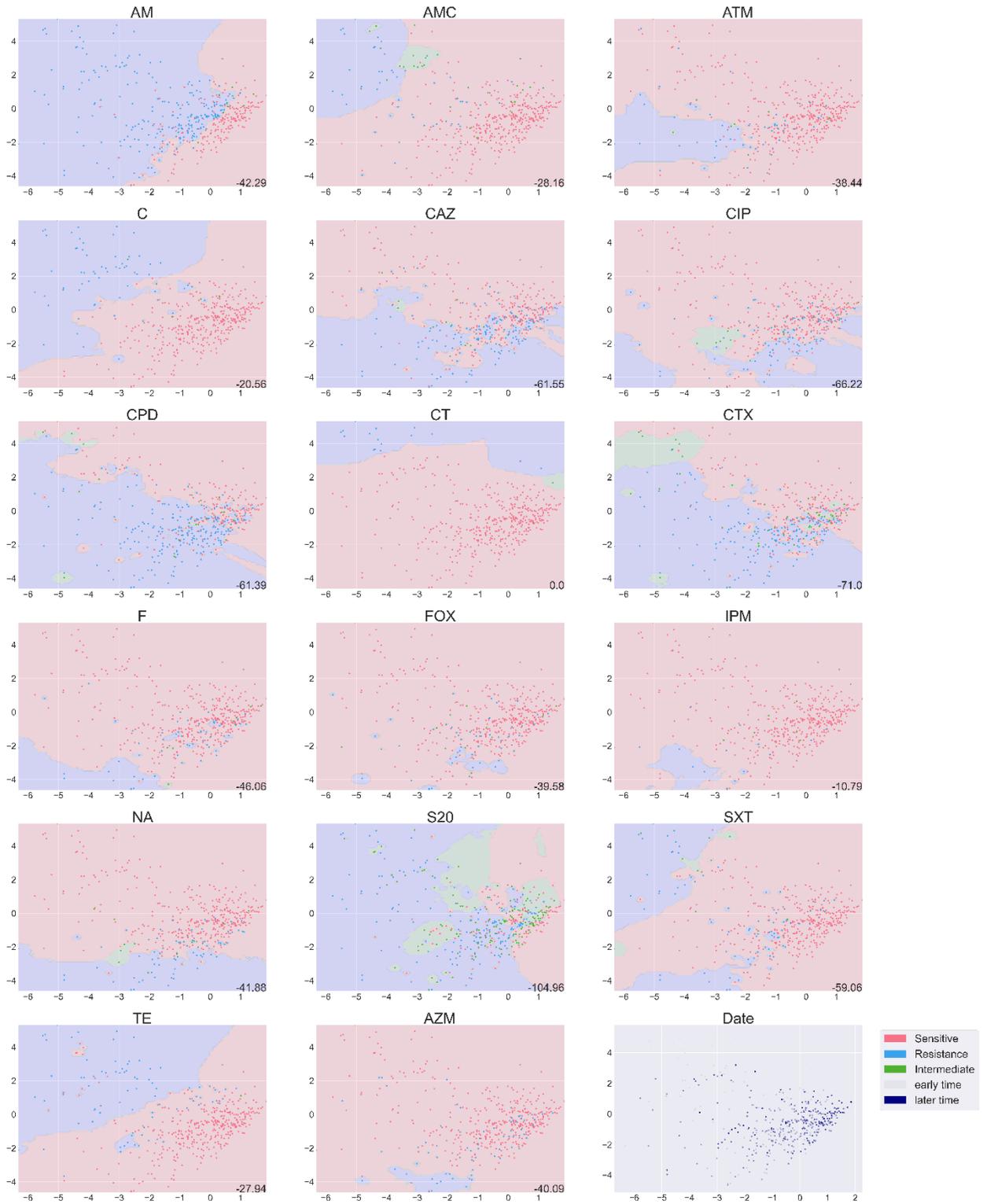
B)
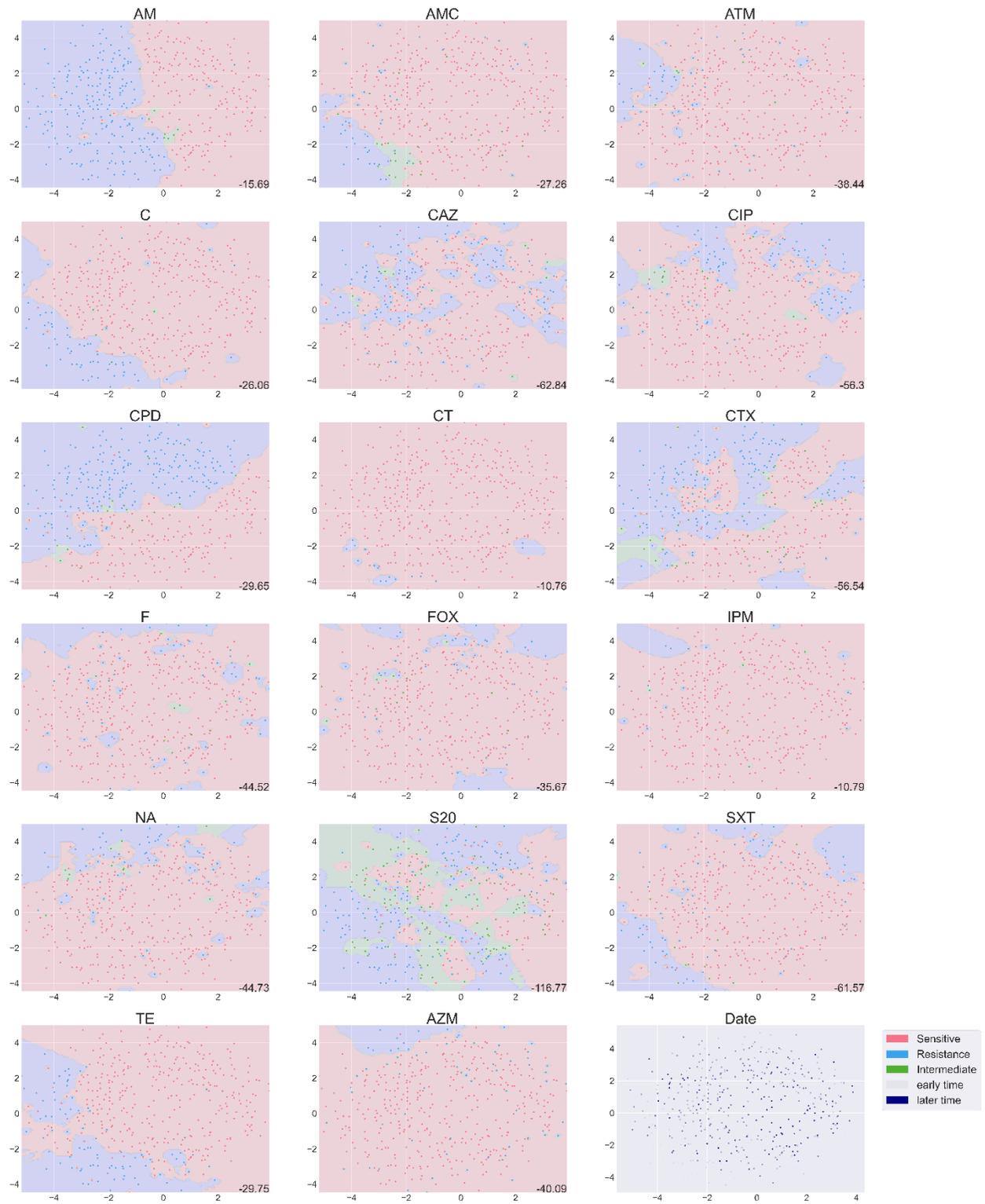
C)
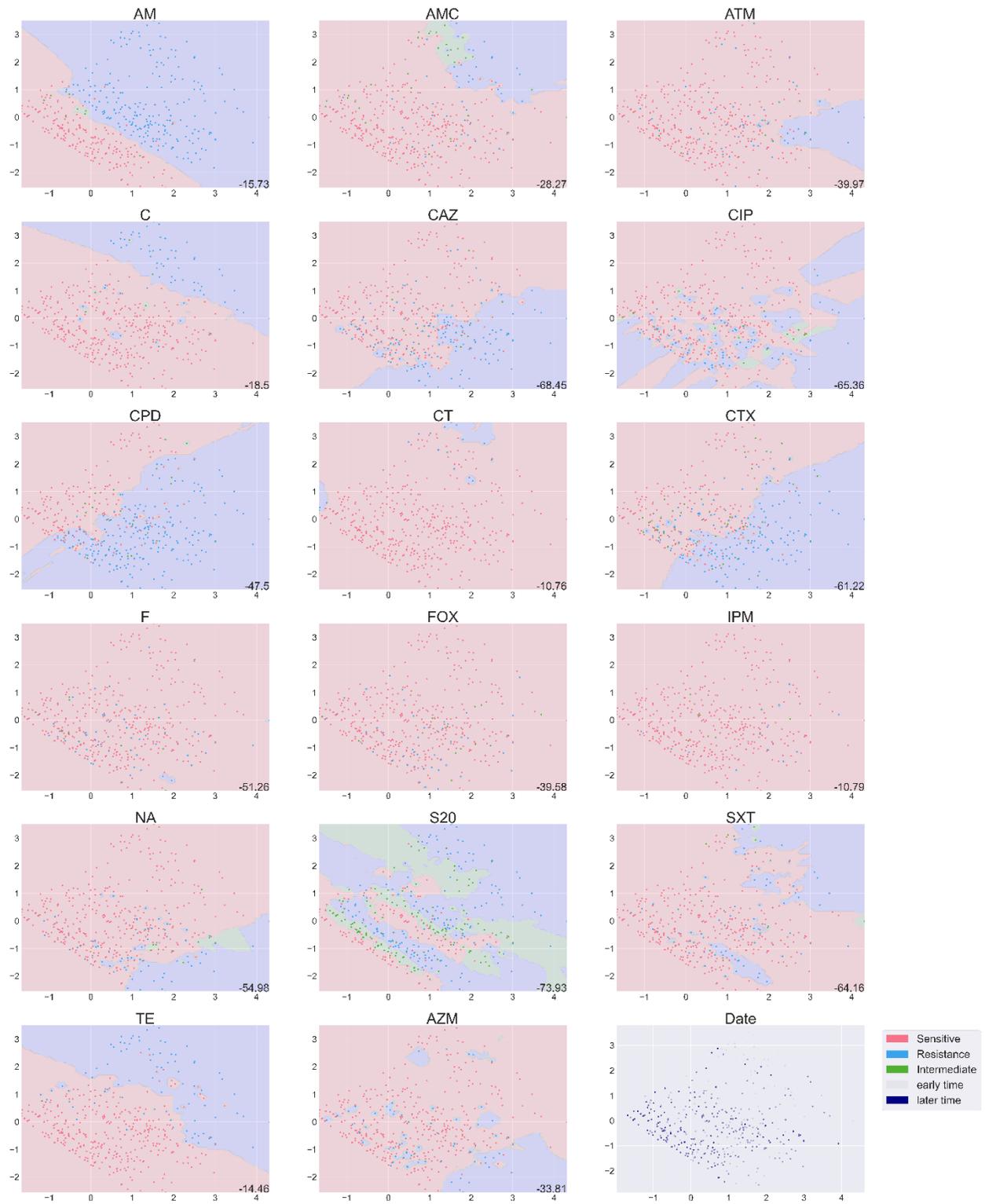
D)

E)
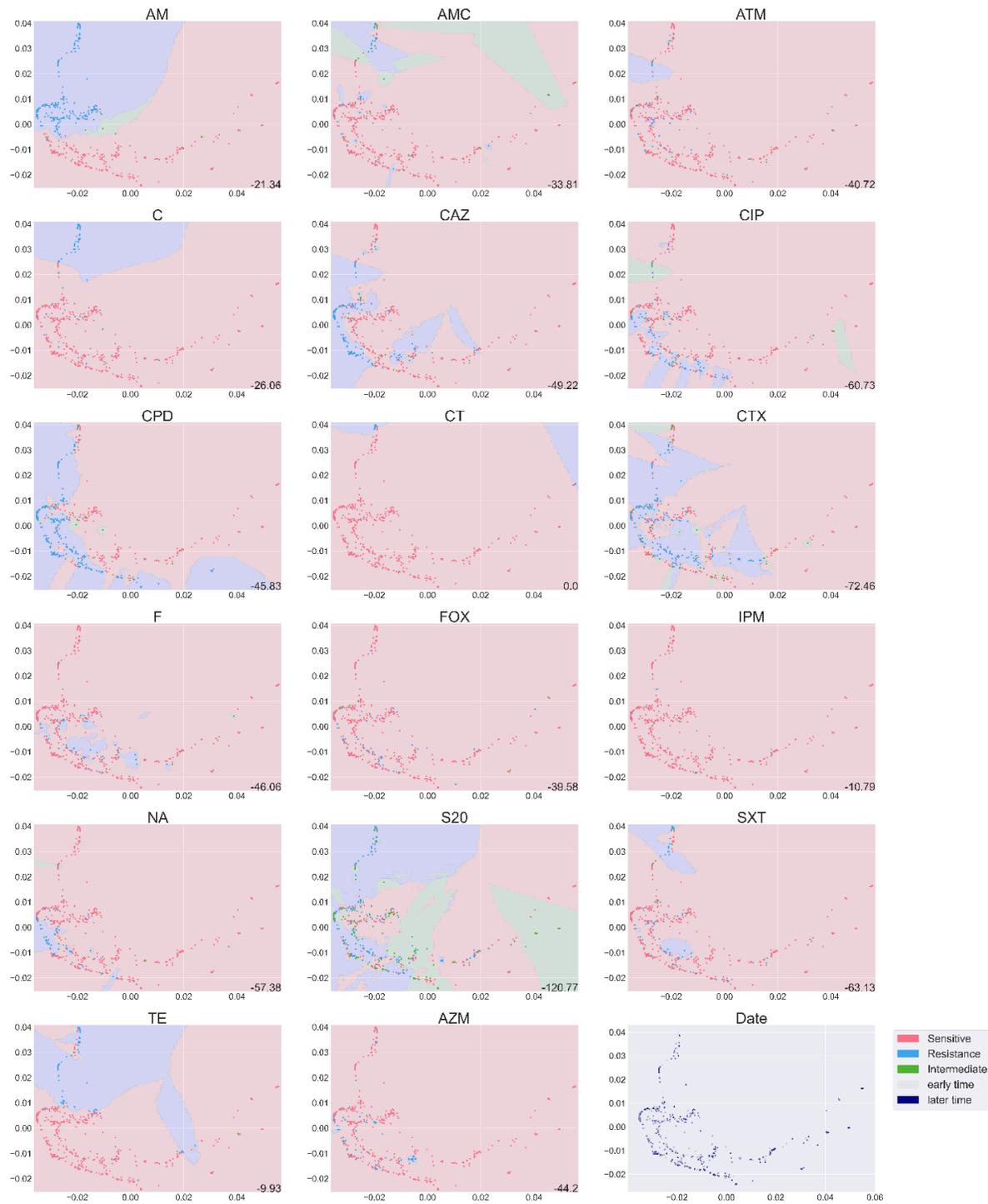


**Figure 3.3** Scatterplot of Data Set 1 (dairy slurry) with a) PCA, b) MDS, c) isomap, d) t-SNE and e) PHATE among each antibiotic. Each subpanel has points highlighted for each different antibiotic: the points coloured in red represent sensitive to that antibiotic, green represent intermediate to that antibiotic and blue represent resistance to that antibiotic. The score at the bottom-right corner represented the log-likelihood for that antibiotic.

43
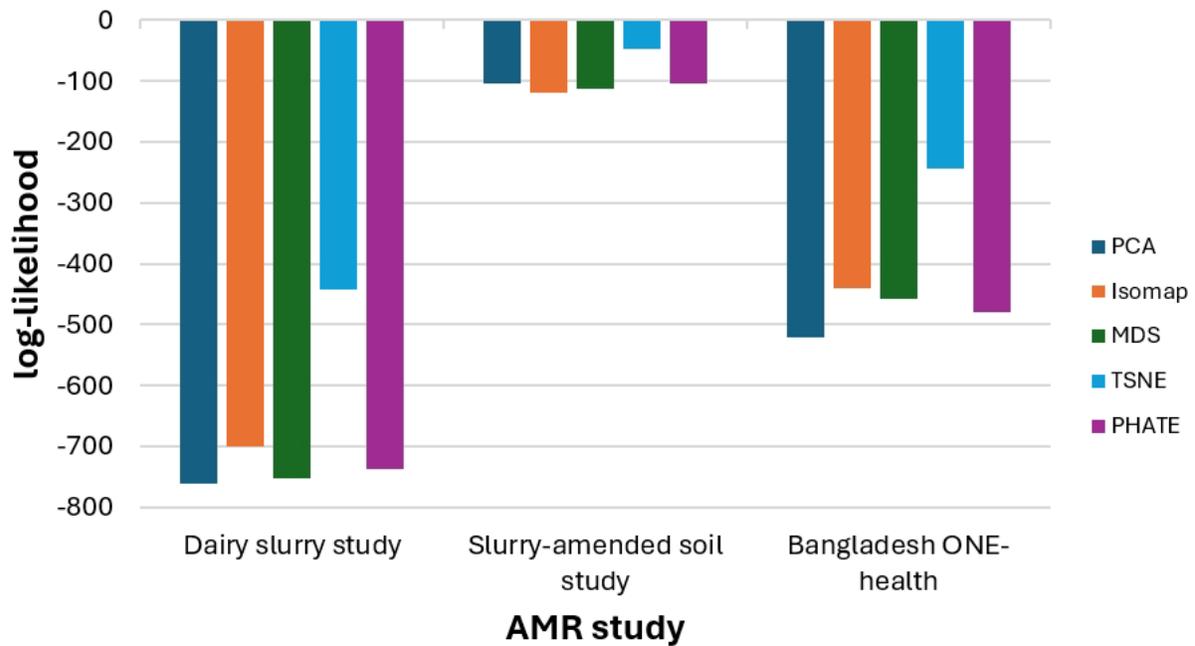
**Figure 3.4** Bar chart showing the log-likelihood of each high-dimensional reduction methods among the three AMR study (dairy slurry study, slurry-amended soil study and Bangladesh ONE-health study). The log-likelihood gives a negative score for the prediction. A score near to zero represented the prediction is close to the actual data. T-SNE has performed best for all three data sets.

A)



AMC      AMP      EFT

-13.32      -12.4      -27.61

S2      OT      ENR

-23.97      0.0      0.0

NA      K      SXT

-1.39      0.0      0.0

FFC      F      CFLEX

0.0      -5.05      -20.09

Date

Sensitive
Resistance
Intermediate
early time
later time

45

B)

C)



AMC    -6.79

AMP    -12.4

EFT    -25.08

S2    -28.01

OT    0.0

ENR    -5.09

NA    -8.73

K    0.0

SXT    0.0

FFC    0.0

F    -13.82

CFLEX    -19.47

Date

Sensitive
Resistance
Intermediate
early time
later time

47
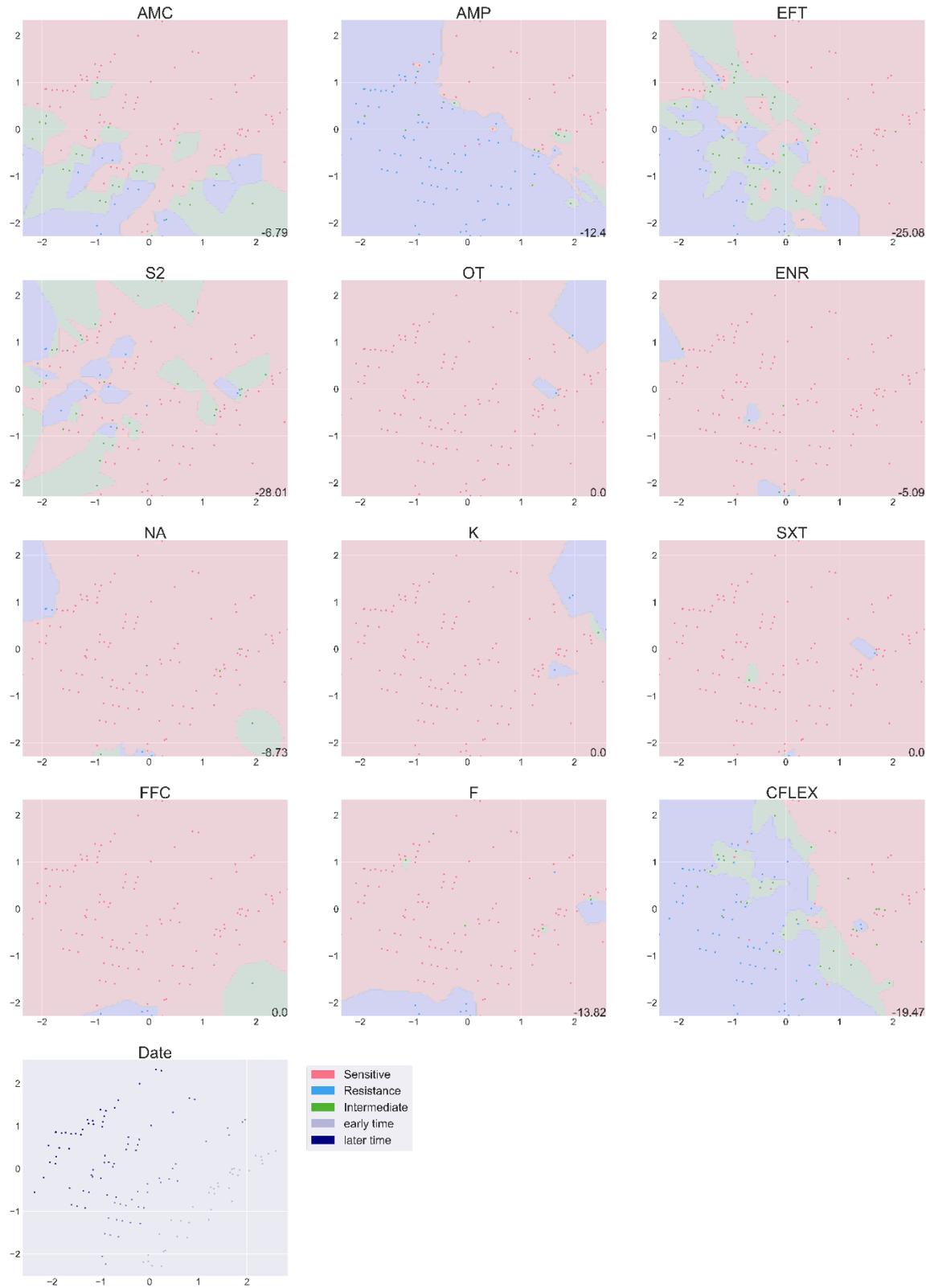
D)
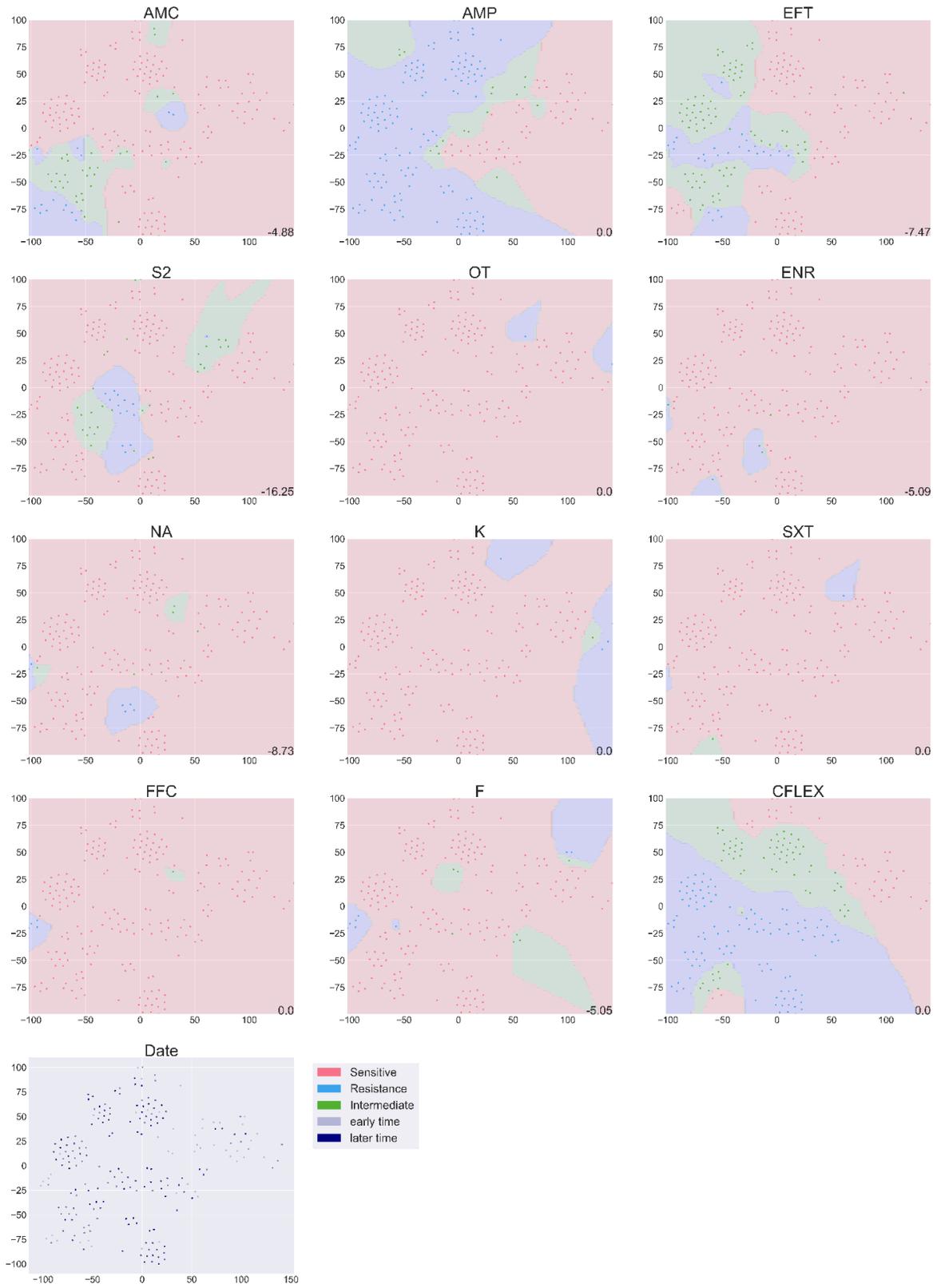
E)
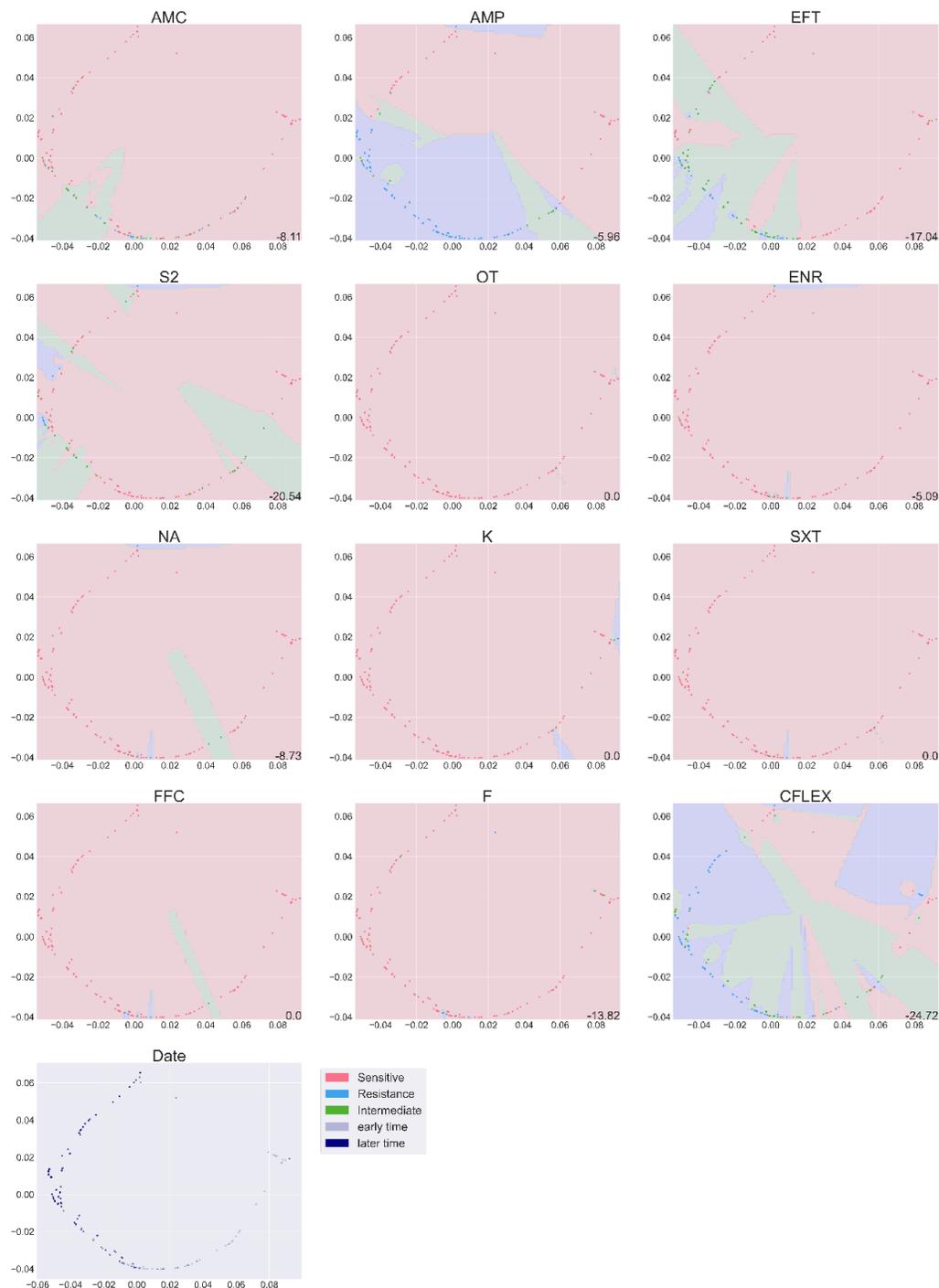


***Figure 3.5*** Scatterplot of Data Set 2 (slurry-amended soil) with a) PCA, b) MDS, c) isomap, d) t-SNE and e) PHATE among each antibiotic. Each subpanel has points highlighted for each different antibiotic: the points coloured in red represent sensitive to that antibiotic, green represent intermediate to that antibiotic and blue represent resistance to that antibiotic. The score at the bottom-right corner represented the log-likelihood for that antibiotic.

A)
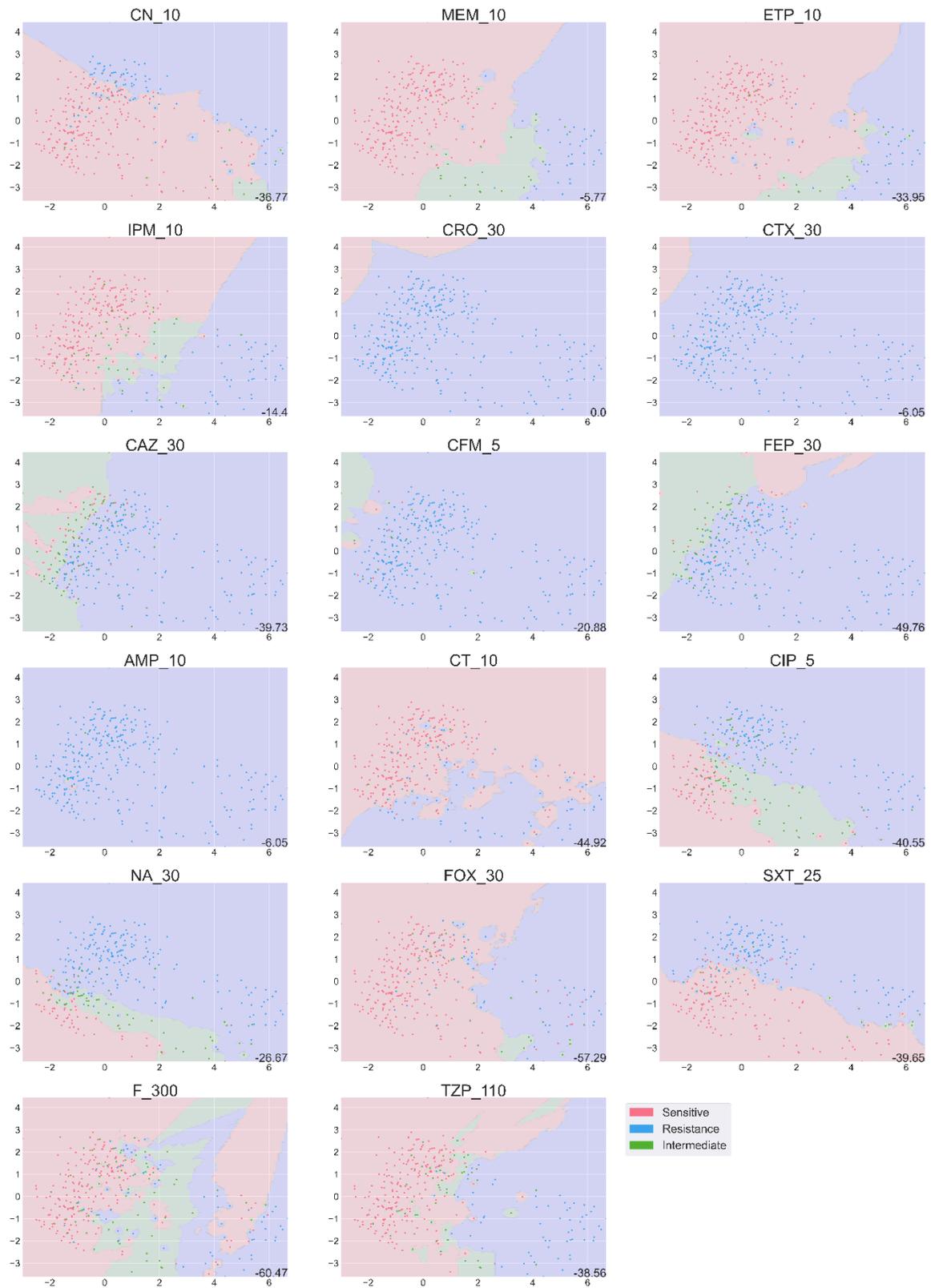
B)
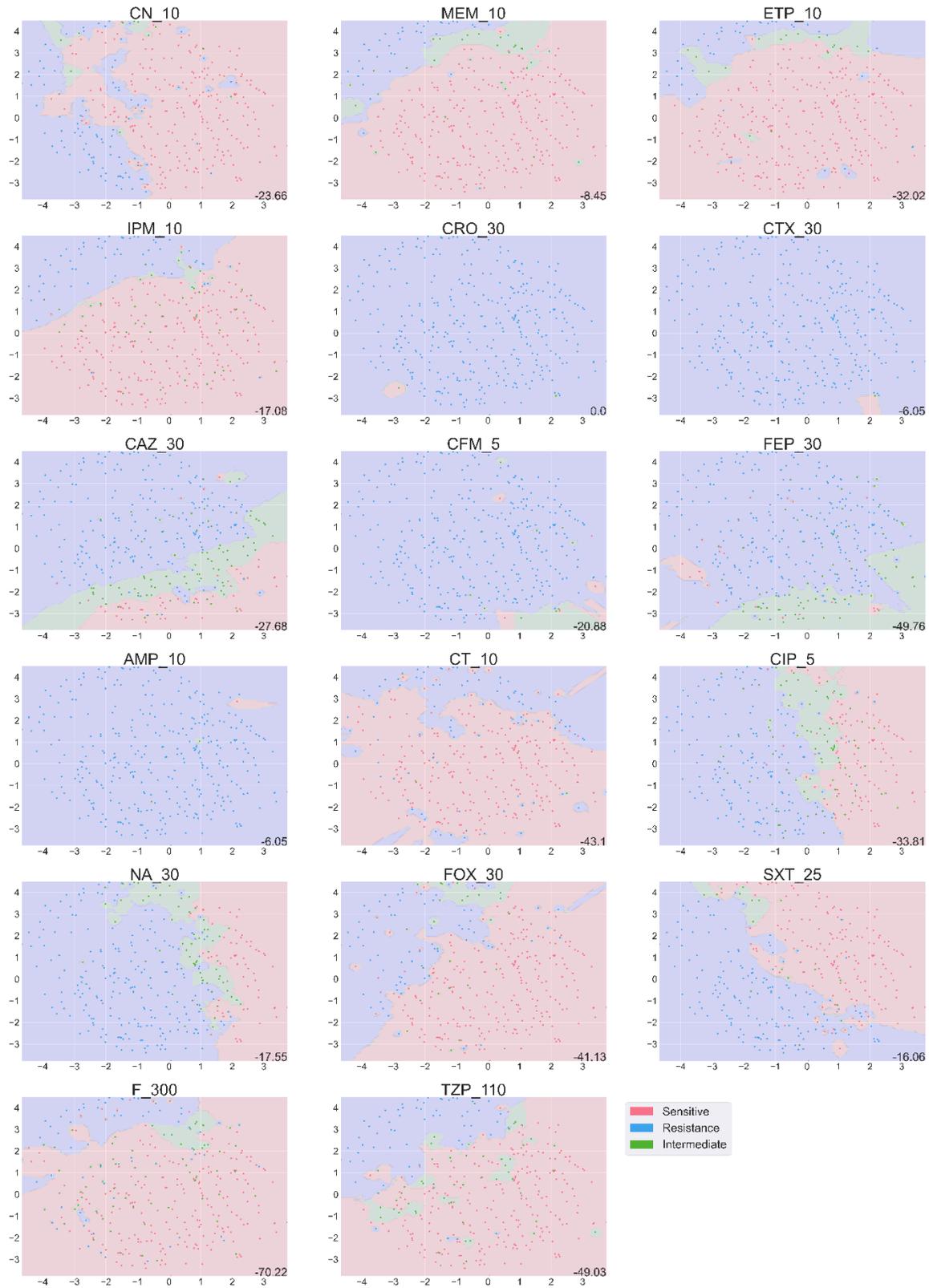
C)
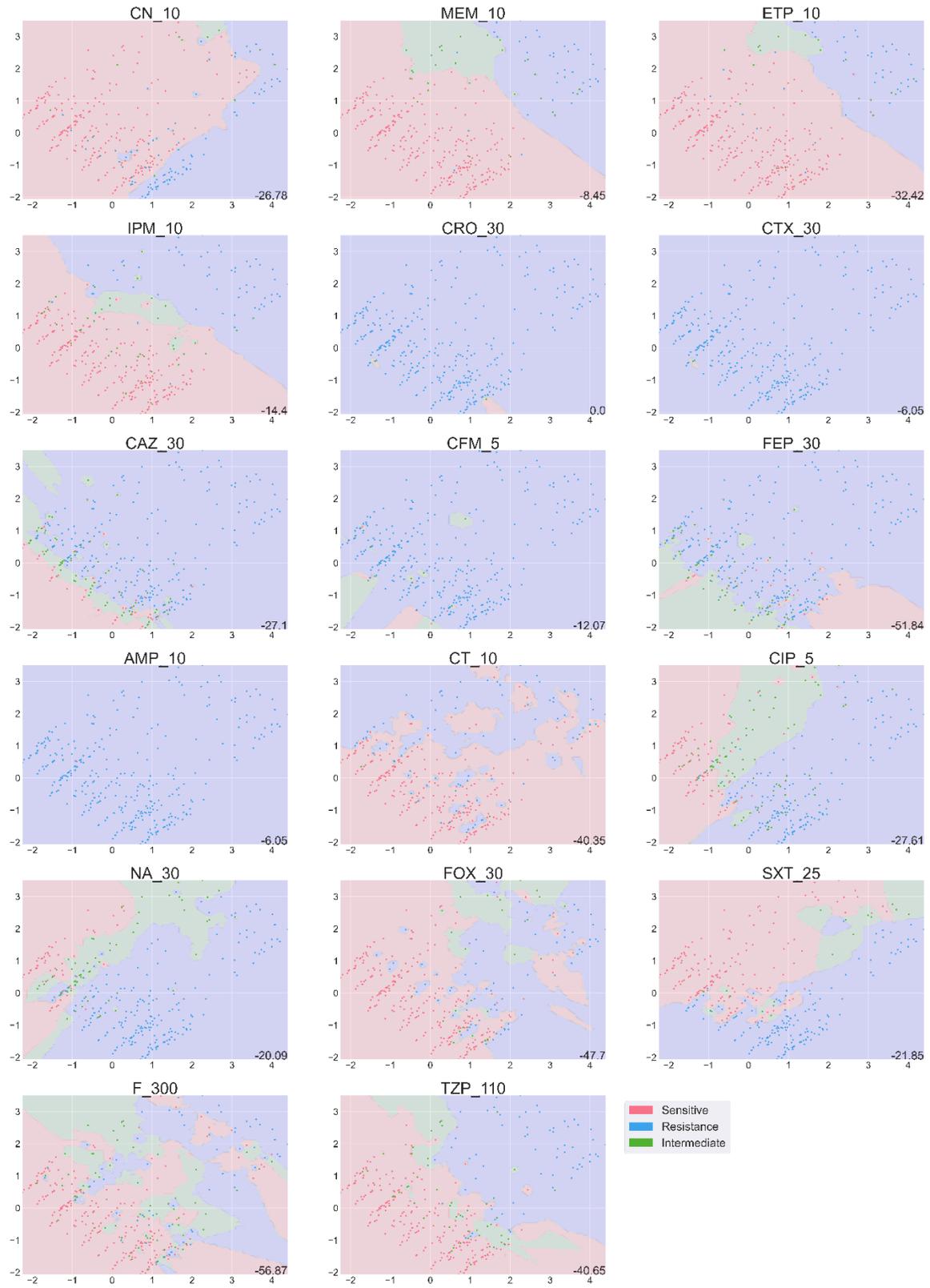
D)
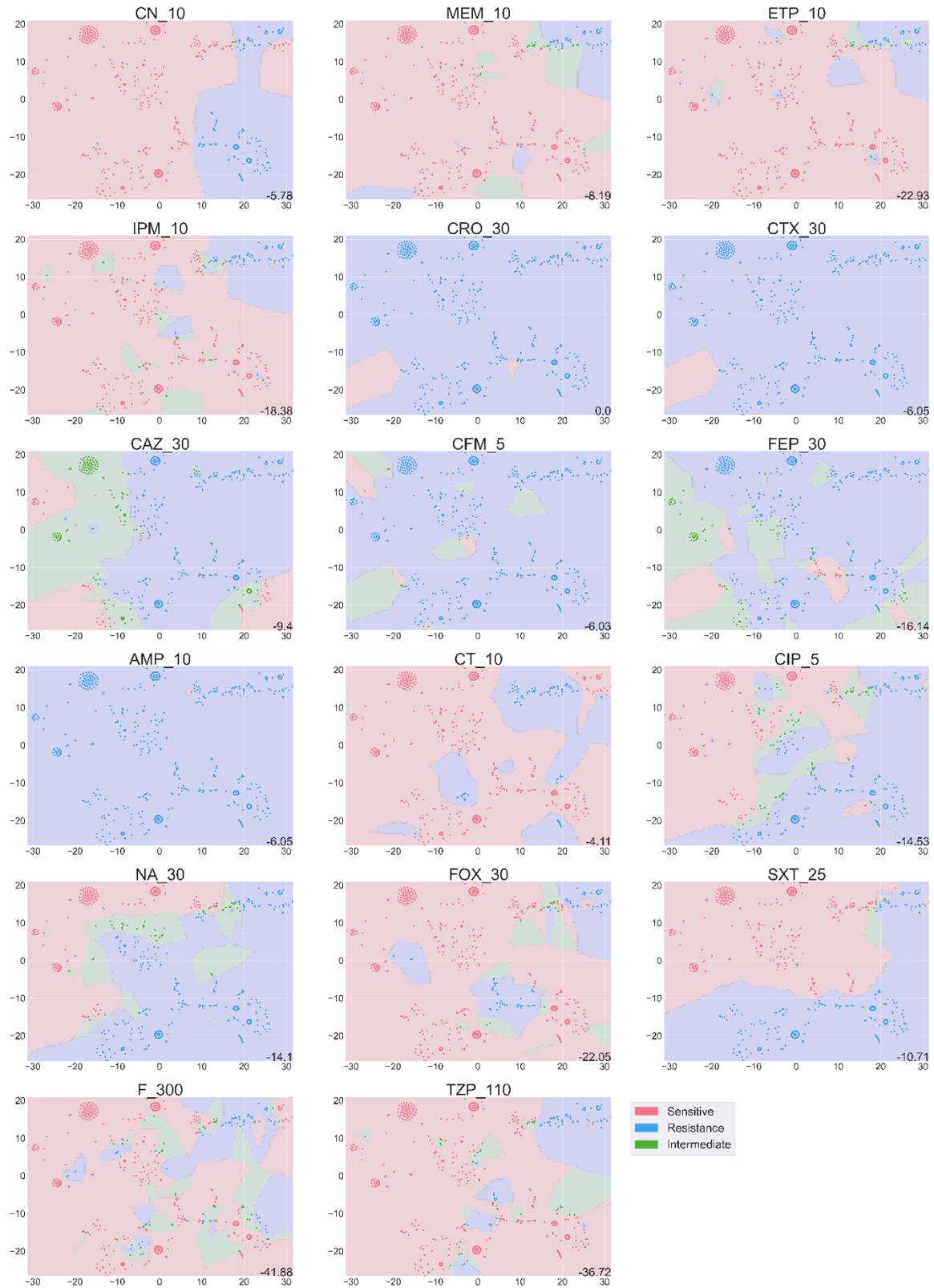
E)
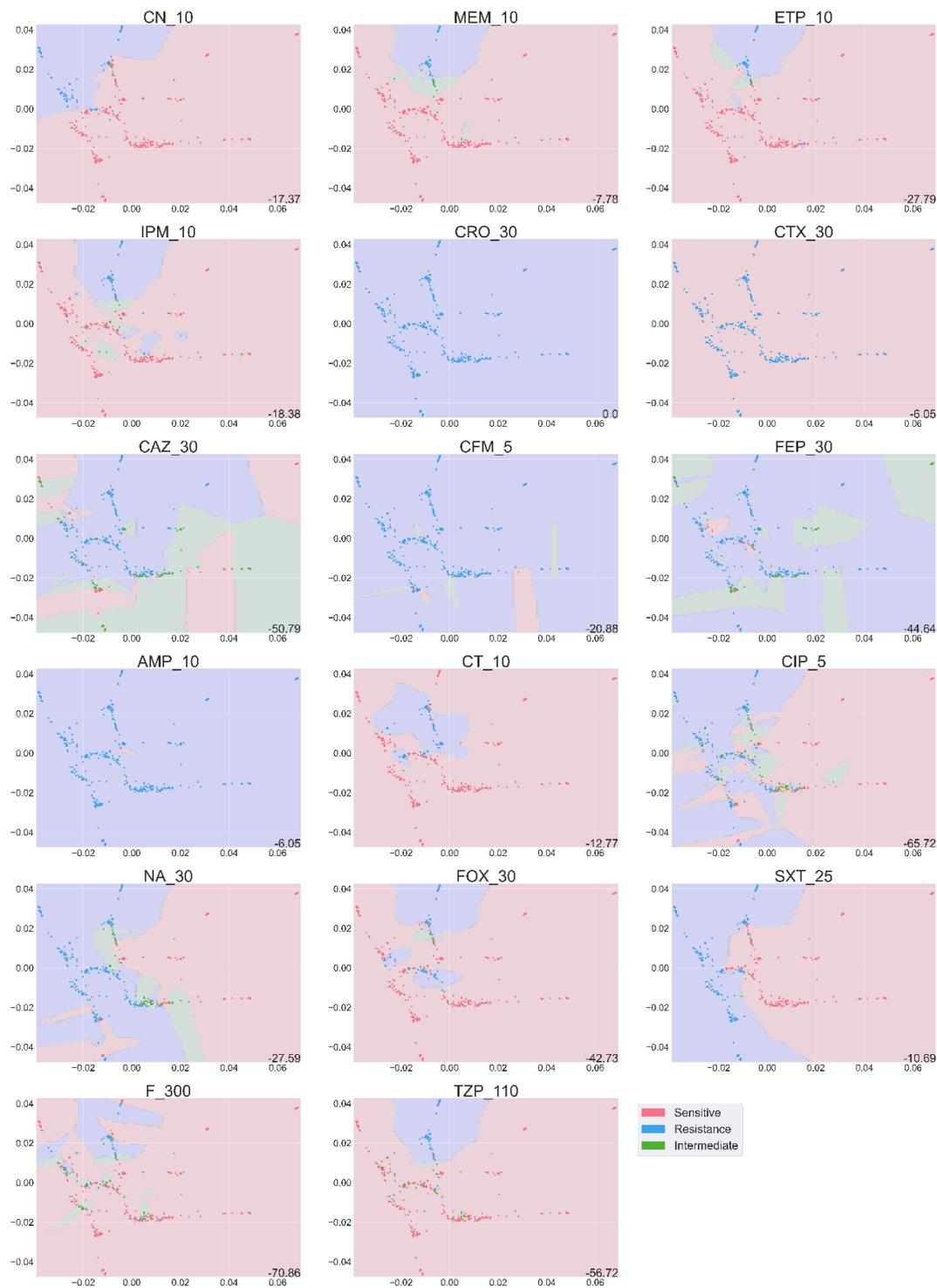


*Figure 3.6* Scatterplot of Data Set 3 (Bangladesh ONE-health) with a) PCA, b) MDS, c) isomap, d) t-SNE and e) PHATE among each antibiotic. Each subpanel has points highlighted for each different antibiotic: the points coloured in red represent sensitive to that antibiotic, green represent intermediate to that antibiotic and blue represent resistance to that antibiotic. The score at the bottom-right corner represented the log-likelihood for that antibiotic.

54

### 3.3.1 The failure of PHATE

The PHATE algorithm could not transform the AMR data successfully mainly because of the discrete nature of the data and lack of relatedness between too many of the data points. The PHATE algorithm calculates the local data distance 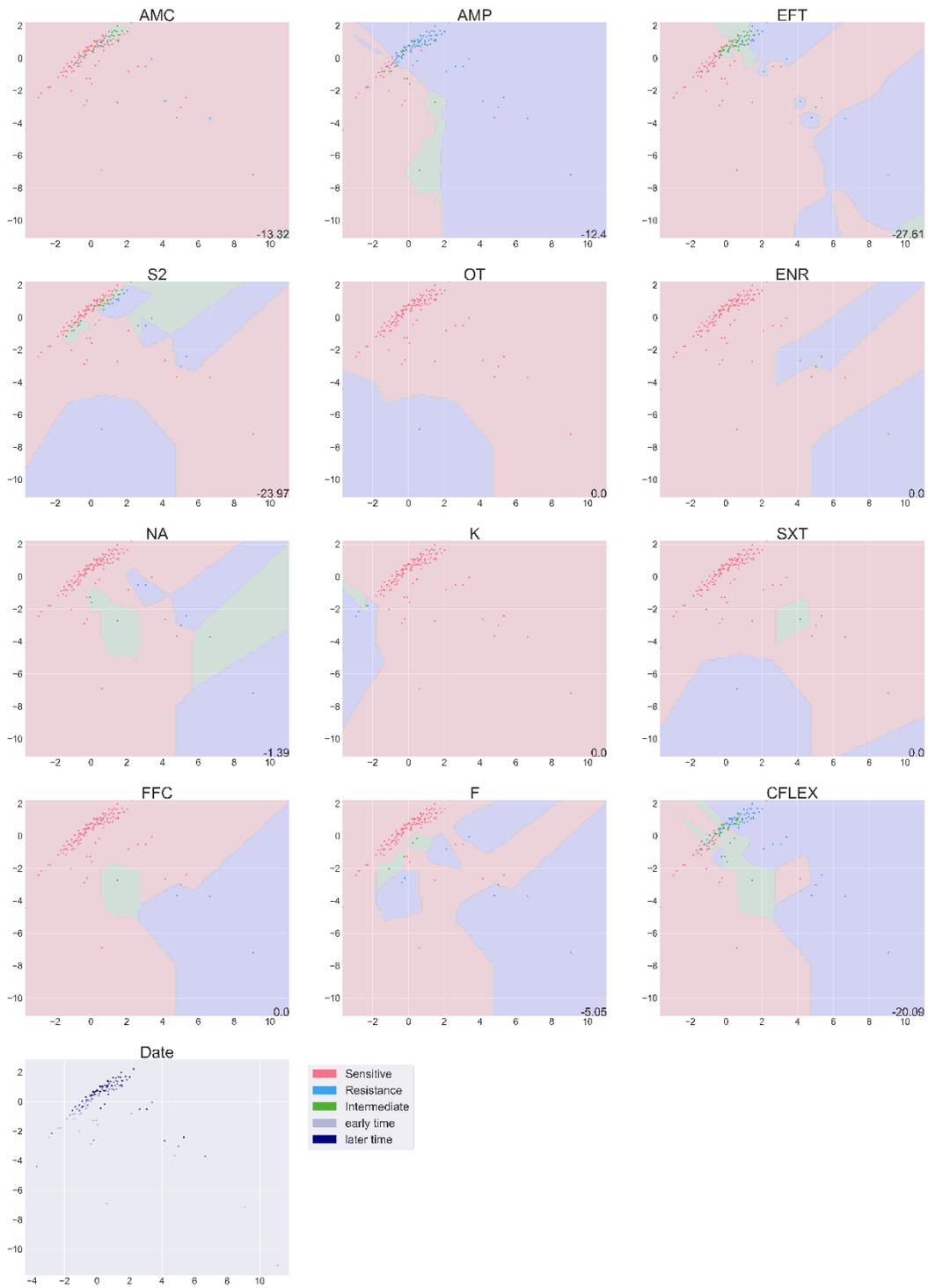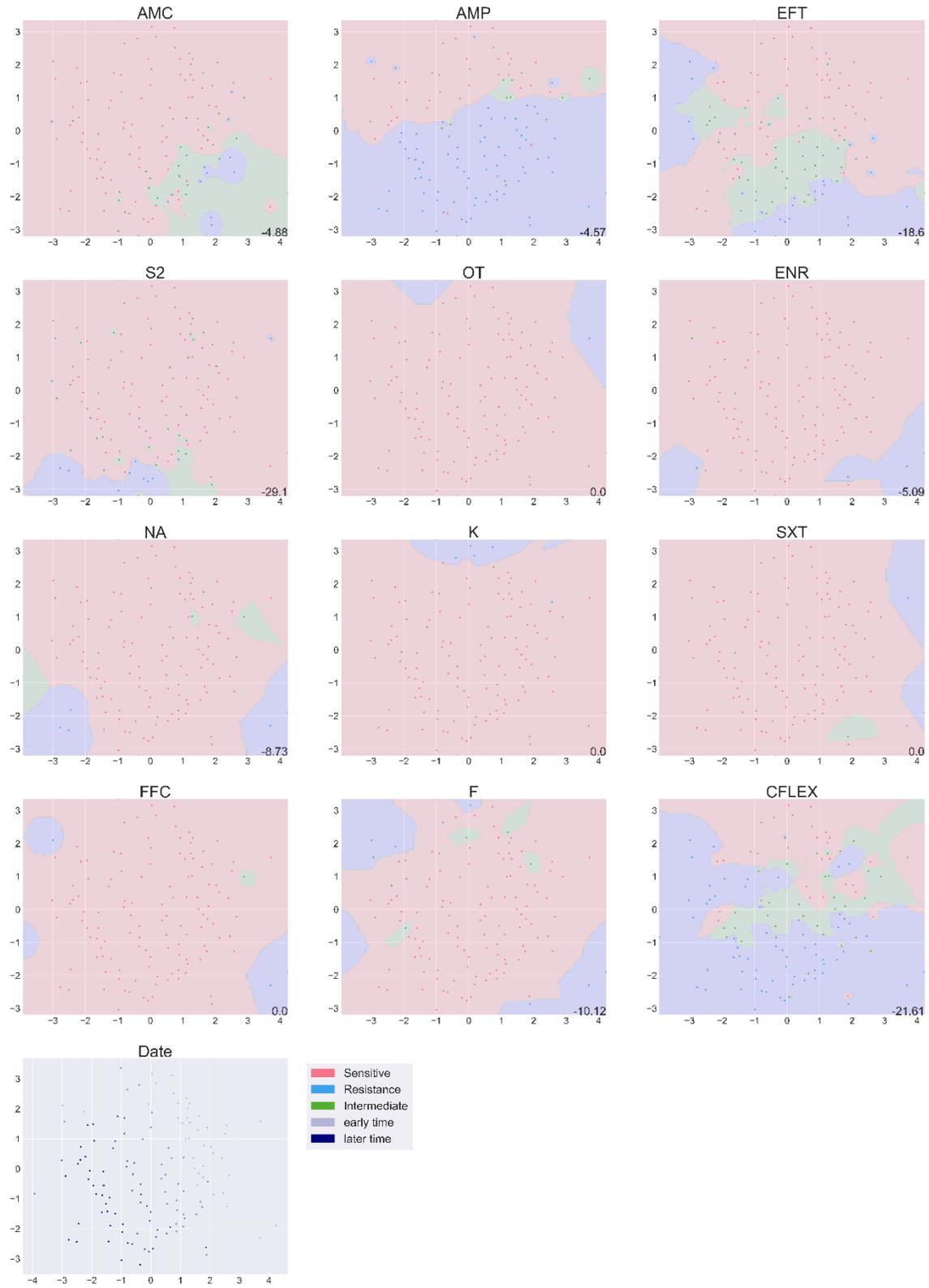between features, and uses the local data distance to calculate the diffusion probability of the data, which was then calculate the informative distance between each of the data. This determined the distance between two points and embedding the closely related points together. As most of the data points are not related to each other, embedding the data in PHATE becomes impossible.

### 3.3.2 Ability of the methods to cluster similar data

As a further test of the effectiveness of four of the methods (PCA, MDS, isomap and t-SNE) to cluster AMR phenotype data, we created a mixed data set consisting of data from all three studies (section 3.2, Figure 3.2), using just the three antibiotics common to the three studies, and tested to see whether these methods would be able to separate out the data. Only PCA could not separate the Bangladesh ONE-health study from dairy slurry study and slurry-amended soil study, and the rest of the methods could separate the Bangladesh ONE-health study from dairy slurry study and slurry-amended soil study (Figure 3.7). PCA decomposes the data based on the maximization of its variance. In some cases, the biological question may not be related to the highest variance in the data. Therefore, MDS, isomap and t-SNE could classifier the difference between the human data and the bacterial data.

A very different test of the methods was to apply them to the antibiotics themselves, to see if they cluster according to antibiotic class based on Dairy Slurry database (Figure3.8). The antibiotics with the same class (eg. Beta-lactam in red) were clustered together in t-SNE. The PCA, MDS and isomap had similar patterns in which antibiotics of different classes were mixed on the plot. As members of the same class of antibiotics, they should share the

same mechanism of action. This implies that they would be countered by the same defensive system of resistance. Consequently, developing resistance to one antibiotic within a class could easily lead to resistance against others in the same class. This suggests that the resistance profile of bacterial individuals belonging to the same class of antibiotics could be similar. t-SNE was able to identify similarities among the input profiles as t-SNE calculated the Gaussian probability contribution of the data in high dimension, while the other methods are based on Euclidean distance. These made the input profile lost the similarity between the similar antibiotic classes.
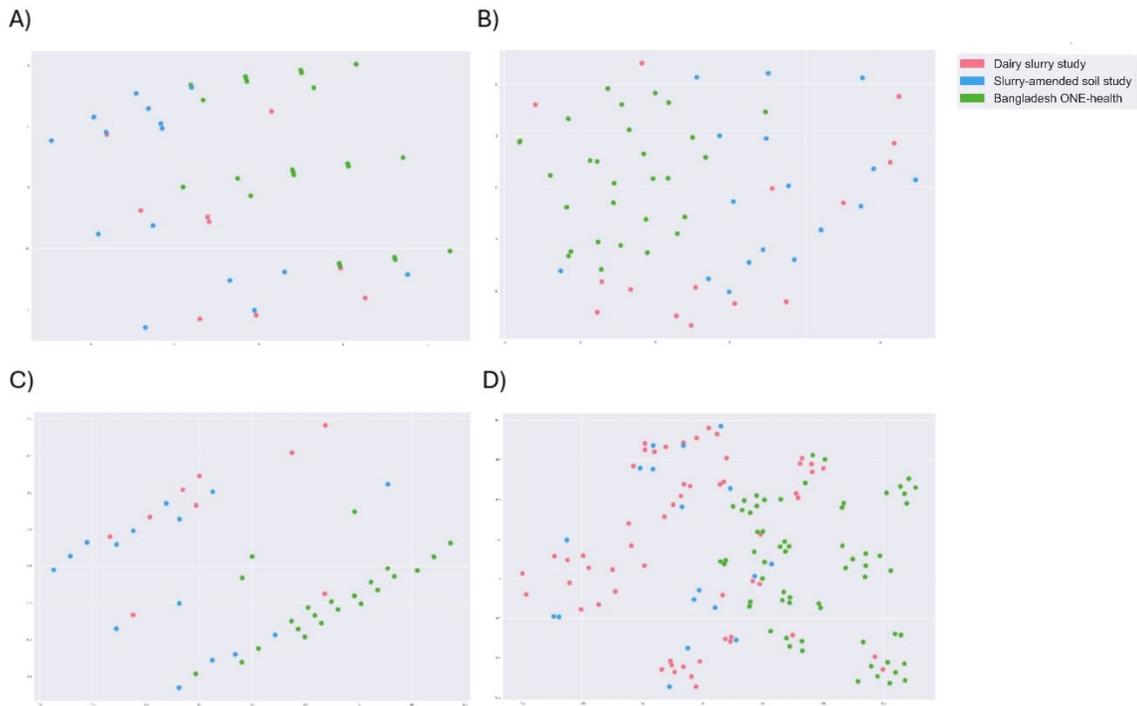


*Figure 3.7* Scatterplot of a) PCA, b) MDS, c) isomap and d) t-SNE when three database was mixed together and used the above method to separate three databases.

A)
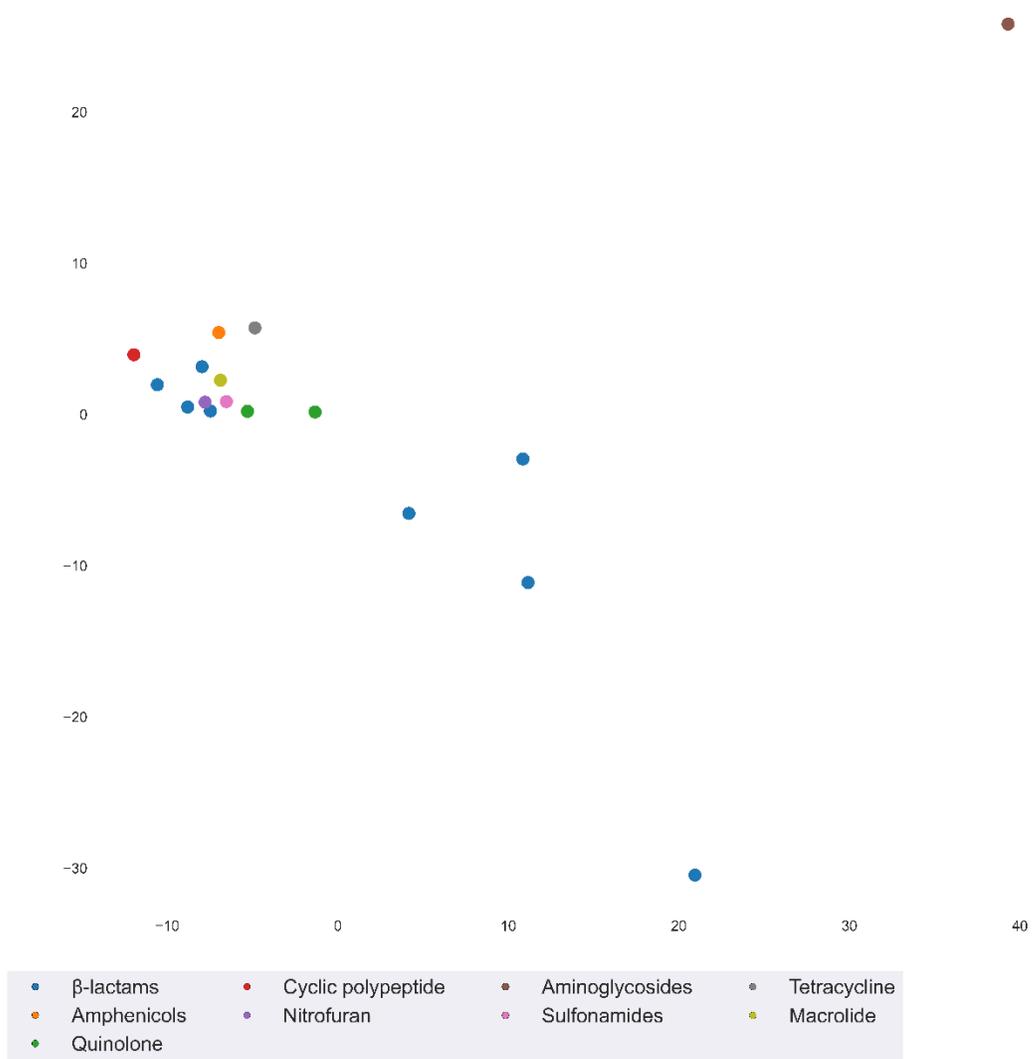
B)
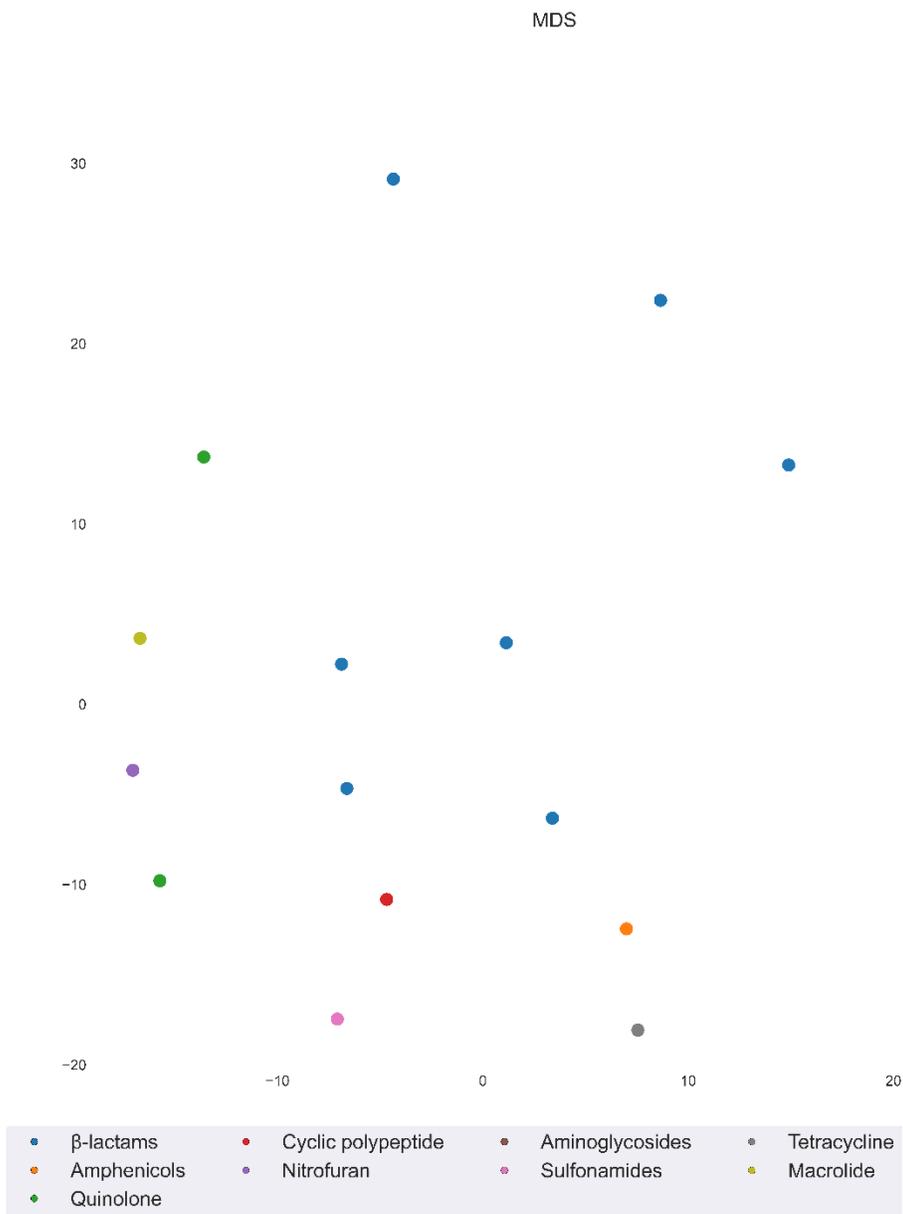
C)
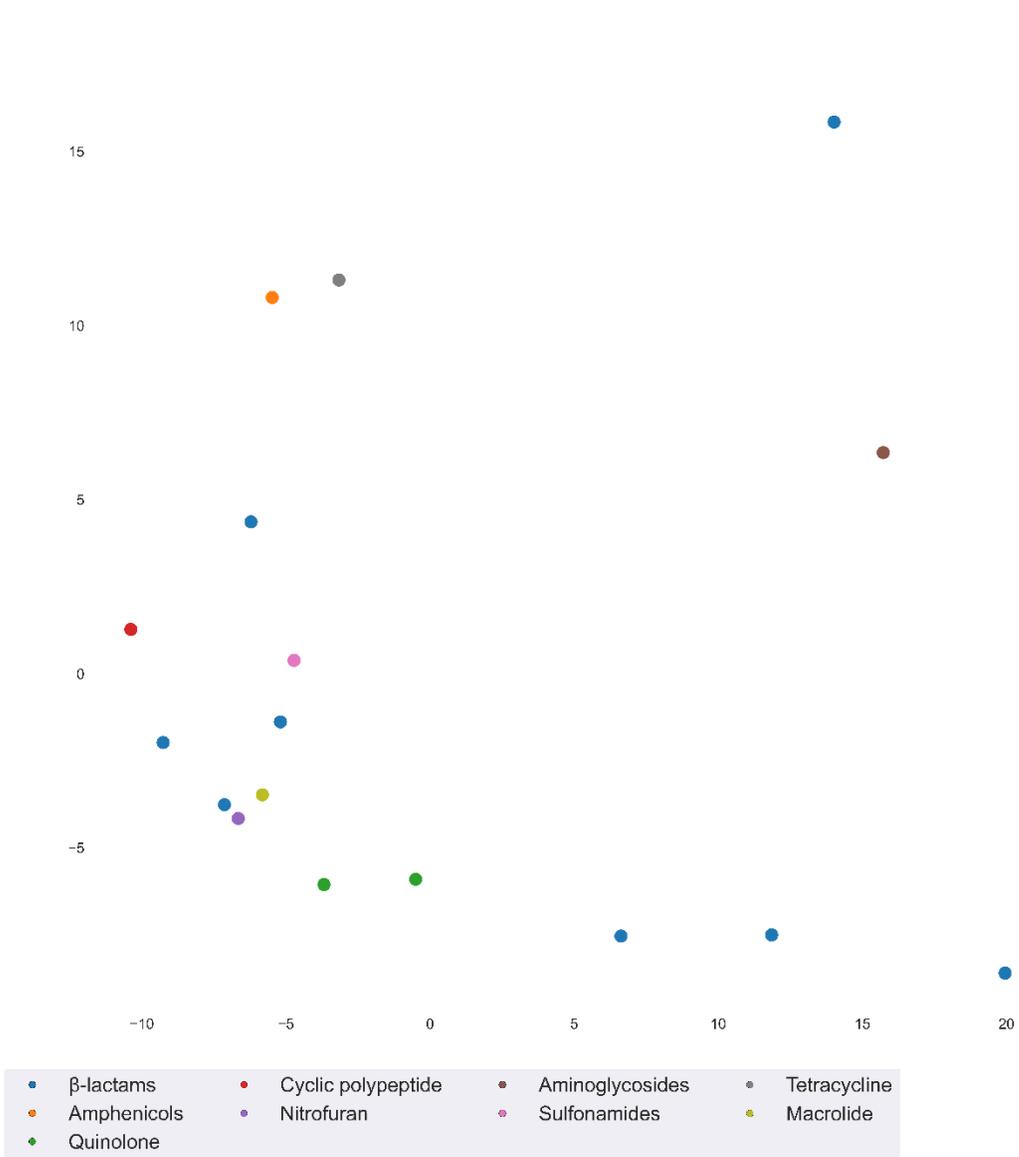


ISOMAP

D)
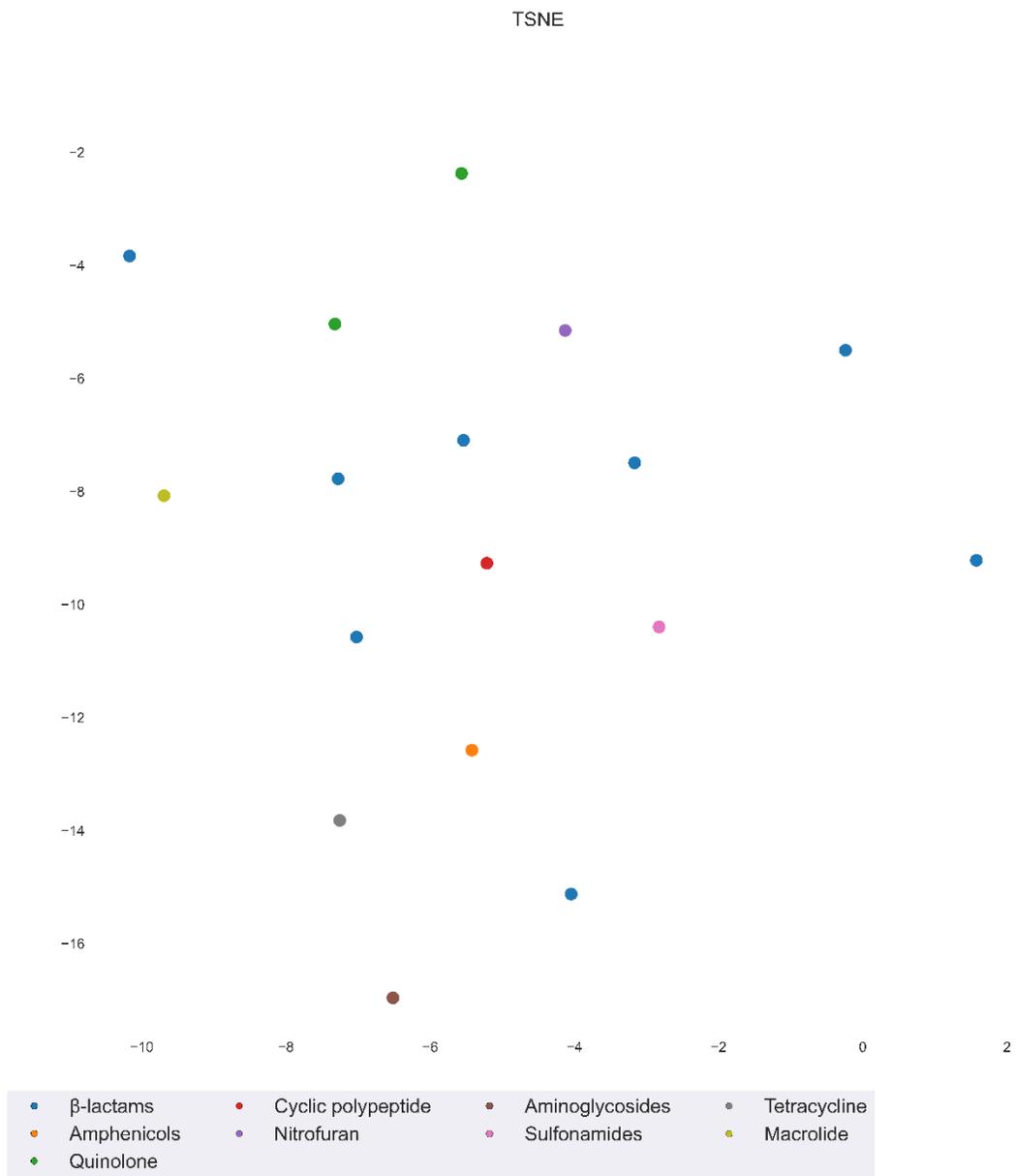


*Figure 3.8* Scatterplot of a) PCA, b) MDS, c) isomap and d) t-SNE when assessing antibiotics from Dairy Slurry database using the strains as the data for each antibiotic.

## 3.4 Using the t-SNE plot to explore patterns of resistance.

The t-SNE plot can be used to identify patterns of resistance among similar bacterial strains. For example, focusing on the upper-right area on the subplot for tetracycline resistance from the dairy slurry (Figure 3.9), it can be seen that strains in this region are also resistant to ampicillin, streptomycin, and chloramphenicol. This may be indicative of co-selection – which could be established through genome sequencing of these strains. The patterns of resistance of these strains can then be assessed through a readable hierarchically-clustered heatmap (Figure 3.9). In this case, the patterns of AMR of tetracycline and streptomycin were closed to one another.  Figure 3.9b shows the time series of tetracycline and streptomycin resistance phenotyes. The data of dairy slurry study (Baker et al., 2022) showed the initial resistance of tetracycline raise from low level to 25% of the population. The supplementary data of dairy slurry study (Baker et al., 2022) also showed high level of streptomycin and tetracycline were used on the farm. Metagenome assemblies also found instances of aminogycloside and tetracycline resistance genes on the same contig. These showed some evidence of the correlation between the developing the resistance of tetracycline and streptomycin, and thus provide validation that the t-SNE method is producing biologically meaningful outcomes.

## 3.5 Summary

In this chapter, t-SNE was identified as the most suitable method to analysis the AMR data from the methods assessed. The advantages and disadvantages of high-dimensional reduction method were shown by using different input features. To exemplify the utility of the dimensionality method, we have used it to provide evidence of correlation between tetracycline and streptomycin resistance in the dairy slurry study. This provides a hypothesis of co-selection which could be assessed through whole genome sequencing of these strains.

A)
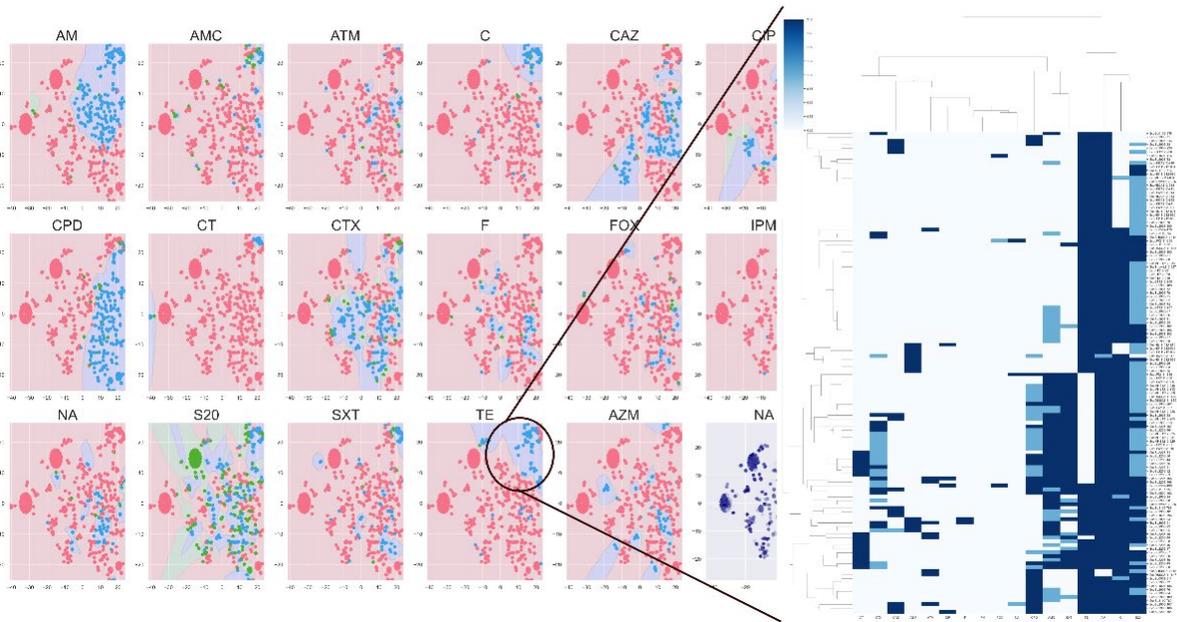


B)
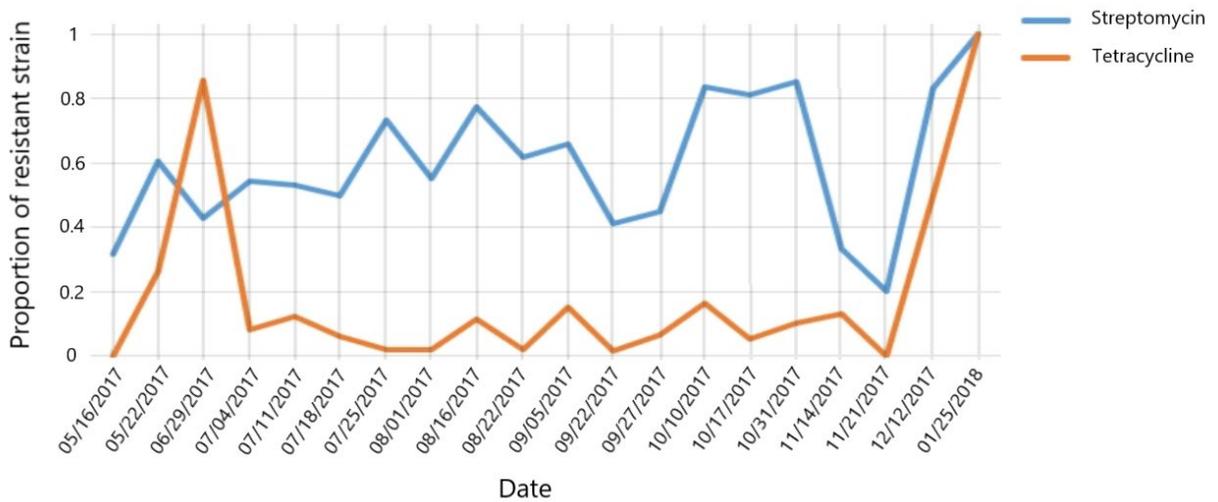


**Figure 3.9** a) The t-SNE scatter plot with KNN classifier with indication of the hierarchically-clustered heatmap showing the pattern of AMR of the particular region, b) the time vs the proportion of the resistance of tetracycline (orange) and streptomycin (blue) within the diversity.

## 3.6 Limitations of high-dimensional reduction methods

PCA is the most commonly used method for dimensionality reduction. For these data, it did not produce results that could accurately reflect our knowledge of biology.  This may be because PCA decomposes data based on the maximization of its variance, which is simply not relevant for this type of data, as the biological question may not be related to the highest variance in the data (Scholz et al., 2004). MDS is similar to PCA and so may suffer from similar limitations.  Isomap and Phate work well only when the sample fits into a low dimensional manifold embedded into a higher dimensional space, which again may not be the case with this type of data. While t-SNE was the most successful method, k=one of the main disadvantages of t-SNE was that the computational time was relatively slow, so it may not scale effectively with much larger data sets.

# Chapter 4

# **Molecular Docking**

This chapter focuses on using molecular docking to predict the affinities of antibiotic metabolites towards the binding pocket of PBP3. Molecular docking is a computational method that models the interaction between a small molecule and a protein at the atomic level, which allows us to characterize the behaviour of small molecules in the binding site of target proteins as well as to elucidate fundamental biochemical processes. The docking process involves two interrelated steps. First, prediction of the ligand conformation as well as its position and orientation within the binding pocket, by sampling conformations of the ligand in the active site of the protein. Second, assessment of the binding affinity, by ranking these conformations using a scoring function. Ideally, sampling algorithms should be able to reproduce any known empirically characterized binding modes, and the scoring function should also rank the highest among all generated conformations.

Chapters 4 and 5 are largely based on our published paper (Chio et al., 2023): I was responsible for all of the simulations and scientific content of the paper, and writing the first draft of the text. My supervisors edited the text and their edits are included into my thesis to improve clarity.

## 4.1 Introduction

Antibiotics are widely used in medicine and agriculture. Moreover, the most recent report from WHO showed that the global antibiotic consumption is increasing, from 21.1 billion defined daily doses in 2000 to 34.8 billion defined daily doses in 2015 (Klein et al., 2018). Orally administered antibiotics, whether in human or veterinary medicine, may be absorbed in the gut; and so may be partly or wholly metabolized prior to excretion, while unabsorbed antibiotics will exit via faeces. Intravenous antibiotics may be subject to similar metabolic fates in serum, faeces and urine (Hoffmann et al., 2007). Most administered antibiotics are

not digested: 50–90% of antibiotic intake is excreted as native antibiotic, while 30% of antibiotics are excreted as closely related metabolites. However, fewer than a third of antibiotic metabolites have yet been tested for bioactivity (Mathers, 2015; Steigbigel et al., 1968; J. Wang & Gardinali, 2014). Antibiotics can also be found in the excreta in farm animals, but metabolite formation within farm animal excreta is largely unknown (Berendsen et al., 2015; Montforts, 1997).

Therefore, antibiotics from medical and veterinary use, and their metabolites, enter the environment and appear as contaminants in wastewater, soil, surface and ground water, sewage, and wastewater treatment plants (Homem & Santos, 2011; Kümmerer, 2009). Antibiotics drive selection for antibiotic resistance; even sub-lethal concentrations of antibiotics can also drive selection (Andersson & Hughes, 2014); the consequence is that antibiotic metabolites, that might be insufficiently potent to have clinical value, might still be able to drive selection for resistance, and so be environmentally important, especially if present in stable forms. While there are studies about how some beta-lactam antibiotics affect specific organisms under controlled photolytic conditions (Timm et al., 2019), there is limited knowledge of the impact of metabolites on bacteria that would normally be affected by the cognate antibiotic. There is also limited knowledge about their environmental stability and the bioactivity of those metabolites. Notably, however, one crystallographic study has found (5$S$)-penicilloic acid complexed to the binding pocket of PBP3 of *P. aeruginosa* (Van Berkel et al., 2013). This raises the important question as to whether a much wider range of antibiotic metabolites may also have this ability. Although wastewater treatment can reduce concentrations of antibiotics, it cannot completely eliminate antibiotics or their metabolites (Abedalwafa et al., 2019; Mahmood et al., 2019; Timm et al., 2019). Moreover, antibiotic and metabolite contamination may be greater in many lower- and middle-income countries (Pokharel et al., 2019; Sartelli et al., 2020; H. Zeng et al., 2022), where wastewater treatment is limited (E. R. Jones et al., 2021).

Chemical detection of antibiotics is an important strand of effective surveillance against antimicrobial resistance. Antibiotics can be detected from the environment through high-performance liquid chromatography (HPLC), mass spectrometry (MALDI-TOF MS and LC-

MS) and colorimetric sensor arrays (Abedalwafa et al., 2019; Mahmood et al., 2019). However, mass spectrometry requires knowledge of the chemical structure of the molecule to be targeted and detected (Welker & van Belkum, 2019); antibiotic metabolites could potentially be tested for in the same ways, but would but would also need to be explicitly targeted in order to be standardized and identified. If metabolites are important, then this is a possible surveillance omission.

In this study, we predict possible bioactivity of antibiotic metabolites to their target binding sites using a two steps computational approach. Molecular docking (N. Huang et al., 2006) is employed as an efficient first step to identify candidate metabolites likely to bind to their cognate target. These candidates were then further assessed using the more accurate but computationally demanding approach of MD simulations (Karplus & McCammon, 2002), which is the subject of Chapter 5. We exemplify this method using the binding of penicillins to PBP3 of *P. aeruginosa* (Van Berkel et al., 2013), recognizing that the approach should be valid for other classes of antibiotics, to other target molecules, and in other organisms.

## 4.2 Preparation of molecular docking

### 4.2.1 Preparation of file for docking

The reference protein structures used for docking were taken from X-ray structures deposited in the Protein Data Bank (www.rcsb.org). 4KQO, the crystal structure of PBP3 from *P. aeruginosa* in complex with piperacillin(Van Berkel et al., 2013), was used. The rationale behind the choice to use this crystal structure was described in Chapter 2. 50 decoy ligands were produced from the Directory of Useful Decoys (DUD)(Huang et al., 2006), based on the structure of 6-APA (as the beta lactam backbone of penicillin antibiotics). The metabolites and the antibiotics used were described in chapter 2 and table 4.1. The ligand files (mol2) were prepared with ChemDraw 12 and the energy minimization was done by Avogadro 1.2.0 (Hanwell et al., 2012). The docking software, GOLD 5.7.1 (Groom et al., 2016), was used.

| Common name | Chemical formula | | Citation |
|---|---|---|---|
| Amoxicillin | C16H19N3O5S |  | Fischer & Robin Ganellin, 2006 |
| Ampicillin | C16H19N3O4S |  | Fischer & Robin Ganellin, 2006 |
| Cloxacillin | C19H18CLN3O5S |  | Fischer & Robin Ganellin, 2006 |
| Dicloxacillin | C19H17CL2N3O5S |  | Fischer & Robin Ganellin, 2006 |
| Flucloxacillin | C19H17CLFN3O5S |  | Fischer & Robin Ganellin, 2006 |

| | | | | |
|---|---|---|---|---|
| Penicillin G | C16H18N2O4S |  | | Fischer & Robin Ganellin, 2006 |
| Penicillin V | C16H18N2O5S |  | | Fischer & Robin Ganellin, 2006 |
| Piperacillin | C23H27N5O7S |  | | Fischer & Robin Ganellin, 2005 |
| Ticarcillin | C15H16N2O6S2 |  | | Fischer & Robin Ganellin, 2006 |
| Oxacillin | C19H19N3O5S |  | | Fischer & Robin Ganellin, 2006 |
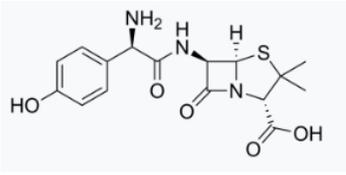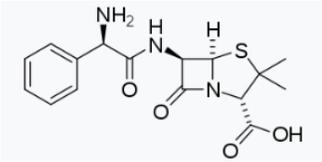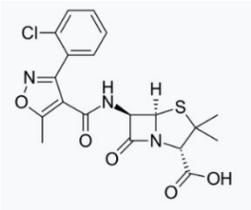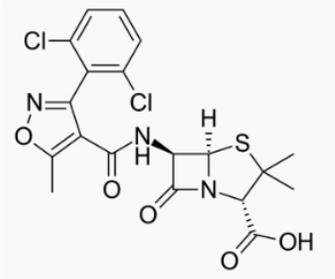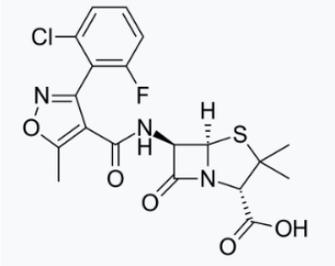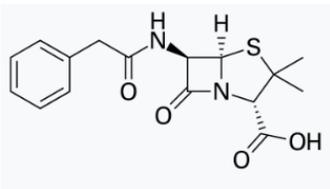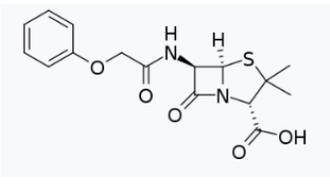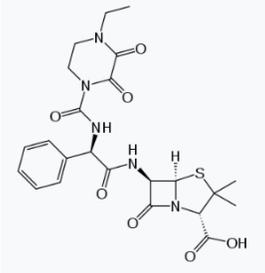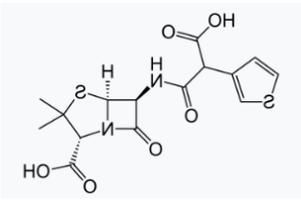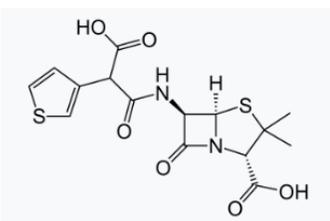
Table 4.1 Table showing the common names, chemical formulae and chemical structure of penicillins used in this study.

## 4.2.2 Setting of molecular docking

All water molecules within the crystallized structure were removed and hydrogen atoms were added when missing from the PDB structure. For each protein target, the active site was defined as the collection of the amino acids enclosed within 8 Å radius sphere which calculated by eBoxSize (Feinstein & Brylinski, 2015) and suggests optimized box size x=y=z=15 Å, centered on the bound antibiotic ligand and flexible docking for 10° of movement freedom of the key residue (Ser294, Ser349, Ser485, Thr487 and Tyr503). The docking used the automatic genetic algorithm setting with crossover weight for point mutation 95, allele mutation weight 95 and migration weight 95 for population size 100 with 100000 of operations with selection pressure1.1 and niche size 2. For the ligand flexibility, the internal hydrogen bonds were detected. The ring conformations, planar amide bond and protonated carboxylic acids were allowed to flip. The torsion angle distributions and postprocess rotatable bonds used the default parameter file. The fitness level was calculated using the Piecewise Linear Potential method.

## 4.2.3 Procedure of molecular docking

The molecular docking process was verified by redocking the ligand back into the crystal structure PBP3 complex with ligand removed. Nine further antibiotics (amoxicillin, ampicillin, cloxacillin, dicloxacillin, flucloxacillin, penicillin G, penicillin V, ticarcillin, oxacillin) were selected to create a reference result towards the docking of the 13 metabolites: (5*R*)- and (5*S*)-penicilloic acid, (5*R*)- and (5*S*)-penillic acid, (5*R*)- and (5*S*)-penilloic acid, (5*R*)- and (5*S*)-pseudopenicillin, penamaldic acid, penicillenic acid, penaldic acid, penilloaldehyde and 6-APA. The first step was to dock the antibiotic itself in order to produce a standard binding pose for that antibiotic. The second step was to dock each metabolite for that antibiotic which is compared with its standard binding pose. 13 metabolites of the antibiotics were screened by docking the molecules into the same binding pocket of the PBP. Ligand interactions were depicted using MOE 2015, while the 3D structures of the ligand with the binding site were visualized using PyMol 2.3.3.

## 4.2.4 Calculation

### 4.2.4.1 Calculation of fitness

The fitness was calculated as the sum of the steric complementarity between protein and ligand ($f_{PLP}$): the score of the ligand consists of the heavy atom clash potential ($f_{clash}$), the torsional potential ($f_{tor}$), covalent docking ($f_{cov}$), flexible sidechains ($f_{side\ chain}$) and water molecules as the constraints ($f_{cons}$). The ChemPLP additionally calculated the distance dependent hydrogen ($f_{chem-hb}$), angle dependent hydrogen ($f_{chem-CHO}$) and metal bonding. (Groom et al., 2016)

$$fitness_{PLP} = f_{PLP} - (f_{clash} + f_{tor} + f_{cov} + f_{side\ chain} + f_{cons})$$

$$fitness_{ChemPLP} = fitness_{PLP} - (f_{chem-hb} + f_{chem-CHO} + f_{chem-metal})$$

$f_{PLP}$ evaluated how the ligand complements with the shape of the protein which focusses on avoiding steric clashes.

$f_{clash}$ calculated the potential for clashes or overlaps between heavy atoms of the ligand and the protein. Higher values indicated more clashes.

$f_{tor}$ calculated the torsional strain associated with rotations around the bonds in the ligand. Low values indicate lower torsional strain and so larger potential energy for the complex.

$f_{cov}$ calculated the potential for covalent interactions between the ligand and the protein. The values were dependent on the force field parameters.

$f_{side\ chain}$ considered the flexibility of protein side chains and the adaptability to bind the ligand. A lower value was more favorable which indicated better binding of the ligand.

$f_{cons}$ related to the presence or absence of water molecules in the binding site. Lower values indicated better adaptation of the water molecules.

$f_{chem-hb}$ related to the strength of hydrogen bonding interacted between the ligand and the protein. Higer values indicated stronger hydrogen bond interactions.

$f_{chem-CHO}$ evaluated the angle dependence of the hydrogen bond interactions between the ligand and the protein. Higher values indicated more optimal angles for hydrogen bond.

$f_{chem-metal}$ considered the interactions involved metal atoms in the ligand and the protein. High values indicated the stronger metal binding interactions (Groom et al., 2016)

### 4.2.4.2 Calculation of RMSD

The RMSDs of the docked metabolite were calculated by comparing the positions of the carbon atoms of the beta-lactam ring of the docked metabolites with their positions in the docked cognate antibiotic. The RMSDs of the decoys were calculated in three steps. Firstly, RMSDs were calculated between unbound 6-APA and 6-APA docked either to the unbound decoy or the decoy in its orientation when bound to the binding pocket. The percentage change of the RMSD of 6-APA aligned to the docked decoy relative to the RMSD of 6-APA aligned to the undocked decoy was calculated. Second, the decoys were replaced in the predicted bound structure by the aligned 6-APA and the RMSD between 6-APA and the carbon atoms of the beta-lactam ring of the docked cognate antibiotic was calculated. Last, the RMSD of the replaced 6-APA was multiplied with the percentage changes of the alignment of 6-APA to the decoy as calculated in the first step of this process.

## 4.3 The docking pose of piperacillin closely matches the crystal structure

The first step in the overall docking process was to verify the integrity of the method by docking an antibiotic ligand back into its cognate binding site from a known crystal structure. We used crystallized piperacillin into 4KQO, PBP3 of *P. aeruginosa*. This produced a benchmark against which the binding of other antibiotics or metabolites to the PBP binding site could be compared. The docking results (Figure 4.1a) showed that the RMSD of the redocking of piperacillin is 0.62 Å with fitness 149. This was confirmed by the proximity of the redocked piperacillin to its position in the crystal structure (Figure 4.1b). This gave confidence that the result was a suitable benchmark for docking the antibiotics and metabolites.
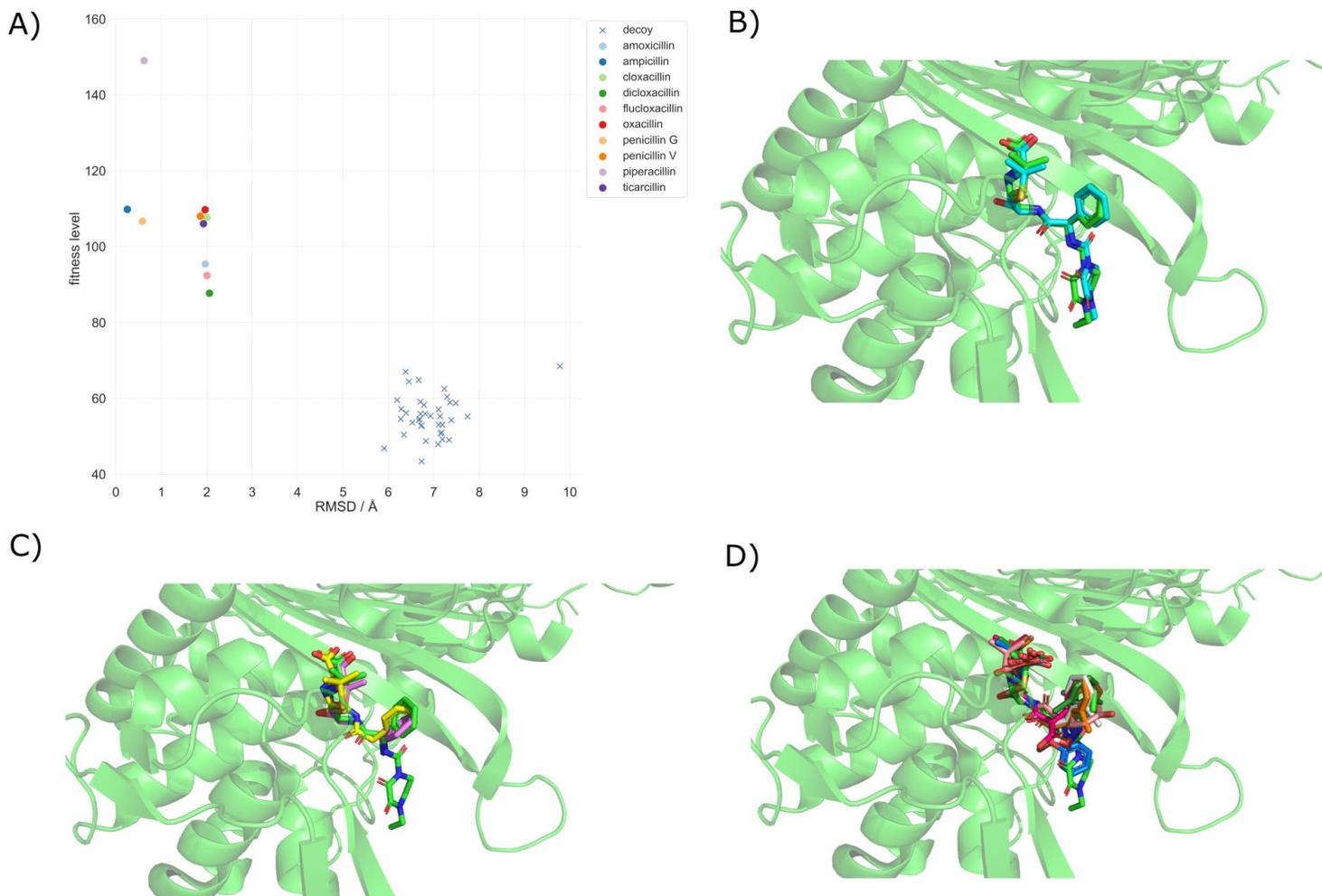
**Figure 4.1** A) Docking results showing the fitness levels and RMSDs of antibiotics and decoys. B-D) Binding poses from docking of b) piperacillin (X-ray structure in green) and docked piperacillin (light blue), c) penicillin G (yellow), ampicillin (purple) which have lower RMSD, c) amoxicillin (pink), cloxacillin (white), dicloxacillin (dark green), flucloxacillin (orange), d) oxacillin (brown), penicillin V (blue), ticarcillin (dark red).

## 4.4 Results

### 4.4.1 Docking of antibiotics

The same docking setup was used to dock the nine antibiotics (amoxicillin, ampicillin, cloxacillin, dicloxacillin, flucloxacillin, penicillin G, penicillin V, ticarcillin, oxacillin) (Fischer

& Robin Ganellin, 2006) into the PBP with the common beta-lactam ring as reference. Ampicillin and Penicillin G showed lower RMSD than piperacillin, with 0.25 Å and 0.58 Å, respectively (Figure 4.1c), but both had lower fitness scores of 109.9 and 106.7. The other penicillins had higher RMSDs (between 1.8 and 2.1 Å). Although these penicillins had a higher RMSDs, they were still close to the crystalized structure of piperacillin (Figure 4.1d).

## 4.4.2 Docking of decoy molecules to penicillin binding protein and metabolites to decoy protein

As controls, 50 decoy ligands were produced from the Directory of Useful Decoys (DUD)(N. Huang et al., 2006). The RMSD of the decoy ligands docked to PBP were between 6 Å to 8 Å (Figure 4.1a, Figure 4.2), considerably higher than any of the antibiotics. The fitness scores of the docking of decoy ligands with the PBPs was between 43 and 68. Taken together, this suggested that the decoy ligands did not bind the PBP and provided useful quantitative controls for metabolite docking below.

## 4.4.3 Determination of the thresholds of the results

Thresholds were estimated for metabolite binding to PBP3 by docking metabolites of piperacillin to the decoy protein, as well as both metabolites and decoy ligands to two decoy proteins to which beta-lactam antibiotics would not be expected to bind: thrombin inhibitor (1BA8) and prohormone-processing carboxypeptidase (1AC5). The RMSD and fitness levels for decoy binding to decoy proteins (1BA8 and 1AC5) were between 5.32 Å and 9.76 Å and 32.1 and 65, respectively (Figure 4.3). The distribution of RMSDs and fitnesses associated with decoy bindings suggested that good metabolite binding could be represented by RMSD less than 5.4 Å and fitness greater than 70. Specifically, antibiotic metabolites of piperacillin docked to the PBP had fitness scores between 77 to 110, while metabolites of docked to decoy proteins (1BA8 and 1AC5) had fitness scores below 74; similar fitness scores were observed for the decoy molecules docked to the decoy proteins (Figure 4.3).
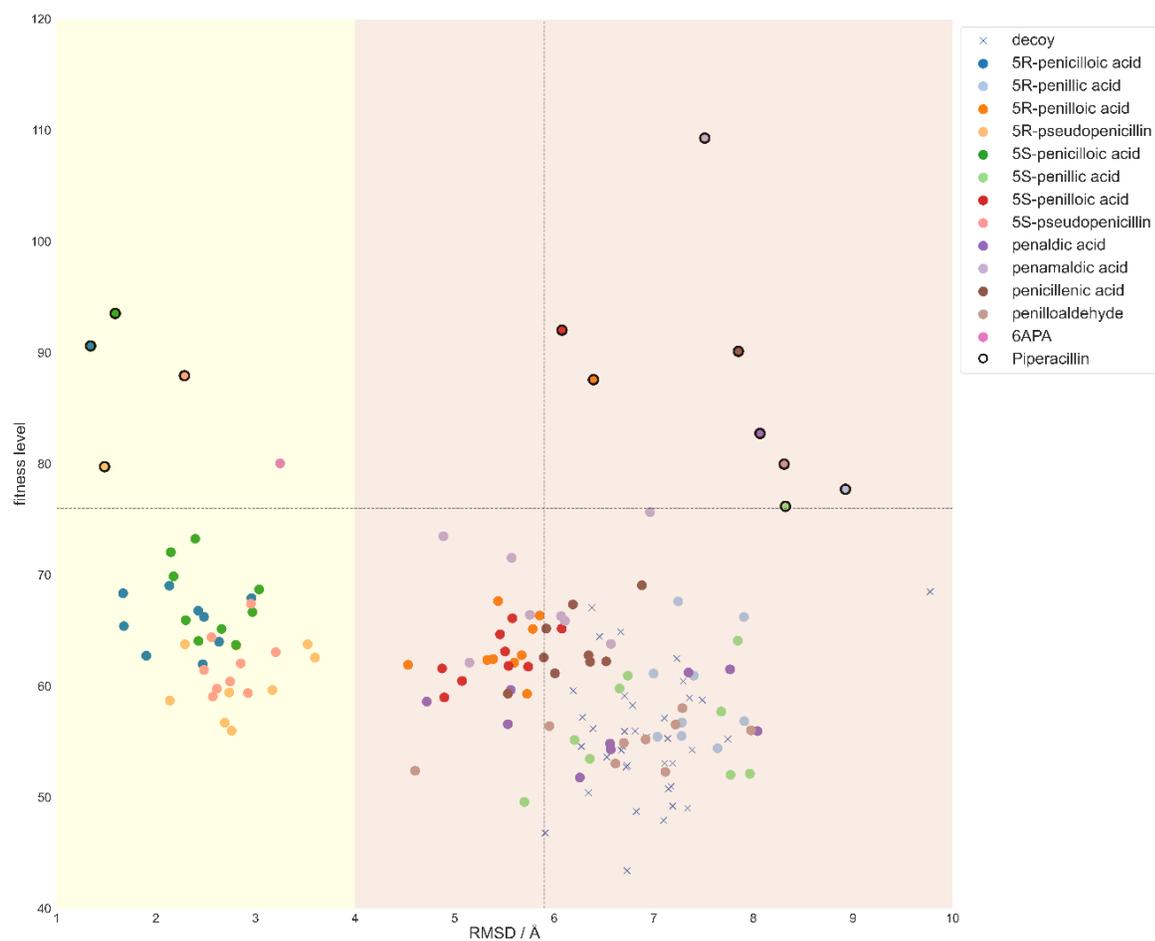
**Figure 4.2** Docking results showing the fitness levels and RMSDs of antibiotic metabolites, together with decoys in order to indicate thresholds for likely docking. The two coloured regions separate the low RMSD (yellow) and high RMSD (red). The horizontal dotted line separates piperacillin metabolites from the metabolites of other antibiotics. As this PBP binds to piperacillin, the R group of piperacillin metabolites will be closer to the empirical crystal structure and thus higher RMSD. The vertical dotted line indicates the decoy ligand with the lowest RMSD and acts as the threshold value to separate the difference between potential binding metabolites and the decoy.
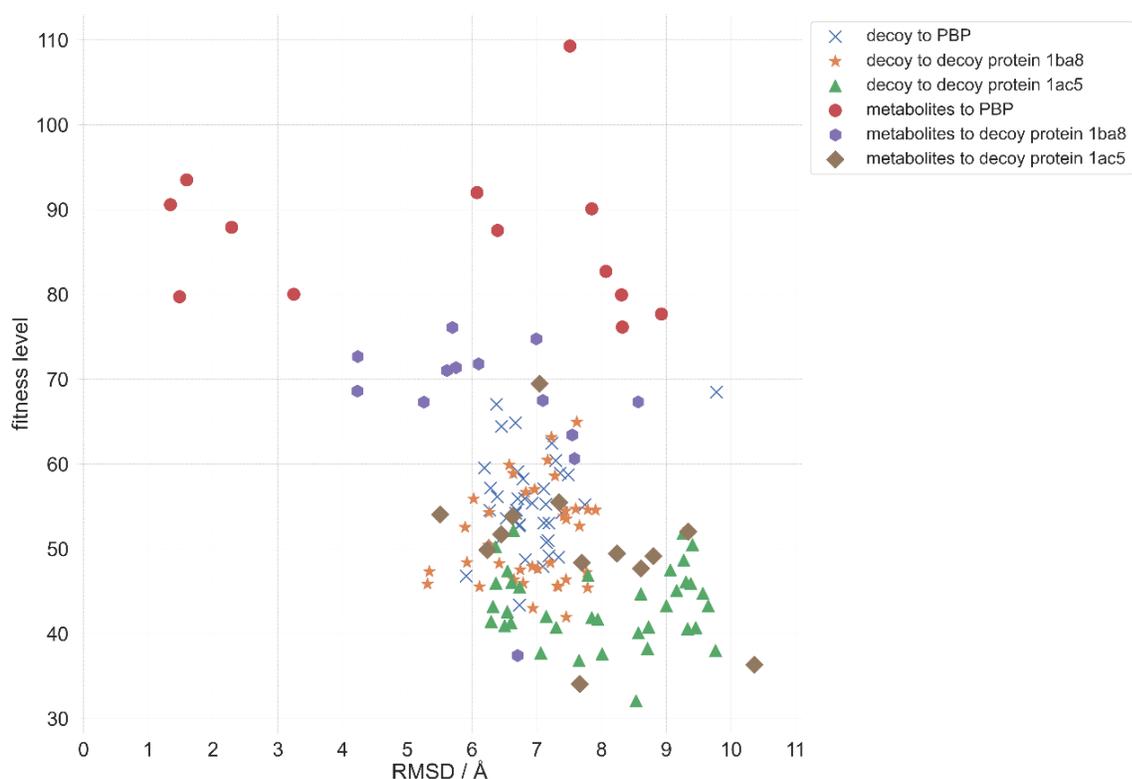
**Figure 4.3** Docking results showing the fitness scores and RMSD of the metabolites and decoy ligands towards PBP and decoy proteins 1BA8 and 1AC5. The decoy ligands towards both PBP and decoy proteins have low fitness and high RMSD. The metabolites towards decoy proteins have slightly higher fitness scores than the decoy ligands and lower fitness scores than metabolites towards PBP and similar RMSD to decoys.

## 4.5 Docking of metabolites with PBPs shows penicilloic acid, pseudopenicillin and 6APA are expected to bind to PBP

When docking the actual penicillin metabolites, (5*R*)- and (5*S*)-penicilloic acid, (5*R*)- and (5*S*)-pseudopenicillin and 6APA had a high fitness (79 to 93) and low RMSD (1.3 Å to 3.2 Å) (Figure 4.2). They were suggested to be more likely to bind to PBP3 compared with the other metabolites. These results are consistent with the X-ray crystal structure for PBP3 complexed with (5*S*)-penicilloic acid (van Berkel et al., 2013), providing confidence for the remaining predictions. (5*R*)- and (5*S*)-penilloic acid, penamaldic acid and penicillenic acid

had a high fitness level (87.6 to 109) and high RMSD (6.1 Å to 7.8 Å) while the decoy molecules bound to PBP have similar RMSD (6 Å to 8 Å) and have a lower fitness (43 to 68) (Figure 4.2). While they might interact with PBPs, the pattern of interaction of the ligand might differ from the antibiotics so they were less likely to bind. (5$R$)- and (5$S$)-penillic acid, penaldic acid, penilloaldyhyde had low fitness level (76 to 82) and high RMSD (8.1 Å to 8.9 Å) and these molecules had a higher RMSD than the decoy molecules (Figure 4.2), so they could not bind to PBP or might easily detach from the protein after binding to it. Because penamaldic acid had the highest fitness and also a high RMSD, it was chosen as a negative control for the MD simulations (Chapter 5). The metabolites docking to the PBP had fitness level between 77 to 110. When metabolites were docked to the decoy proteins, the fitness levels (37 to 74) were similar to the fitness level when the decoy molecules docked to the decoy proteins (Figure 4.3).

## 4.6 T-SNE analysis of the results of molecular docking

With the information of fitness and RMSD of each of the metabolites, t-SNE was performed to analysis the relationship among the metabolites. From the figure drawn by fitness (Figure 4.4a), (5$R$)-penillic acid and (5$R$)-pseudopenicillin were not clustered into low fitness high RMSD and high fitness low RMSD, respectively. The rest of the metabolites clustered well together according to their category. If only fitness was considered, only (5$R$)-penillic acid was not clustered into low fitness group. For RMSD (Figure 4.4b), only metabolites with high fitness and low RMSD clustered together but (5$R$)-pseudopenicillin also leave the cluster. The remaining metabolites were dispersed around the manifold. In terms of RMSD only, all metabolites except (5$R$)-pseudopenicillin clustered together. From Figure 4.4c, (5$R$)-penicilloic acid, penaldic acid and penicillenic acid did not clustered to their own category in terms of fitness and RMSD.
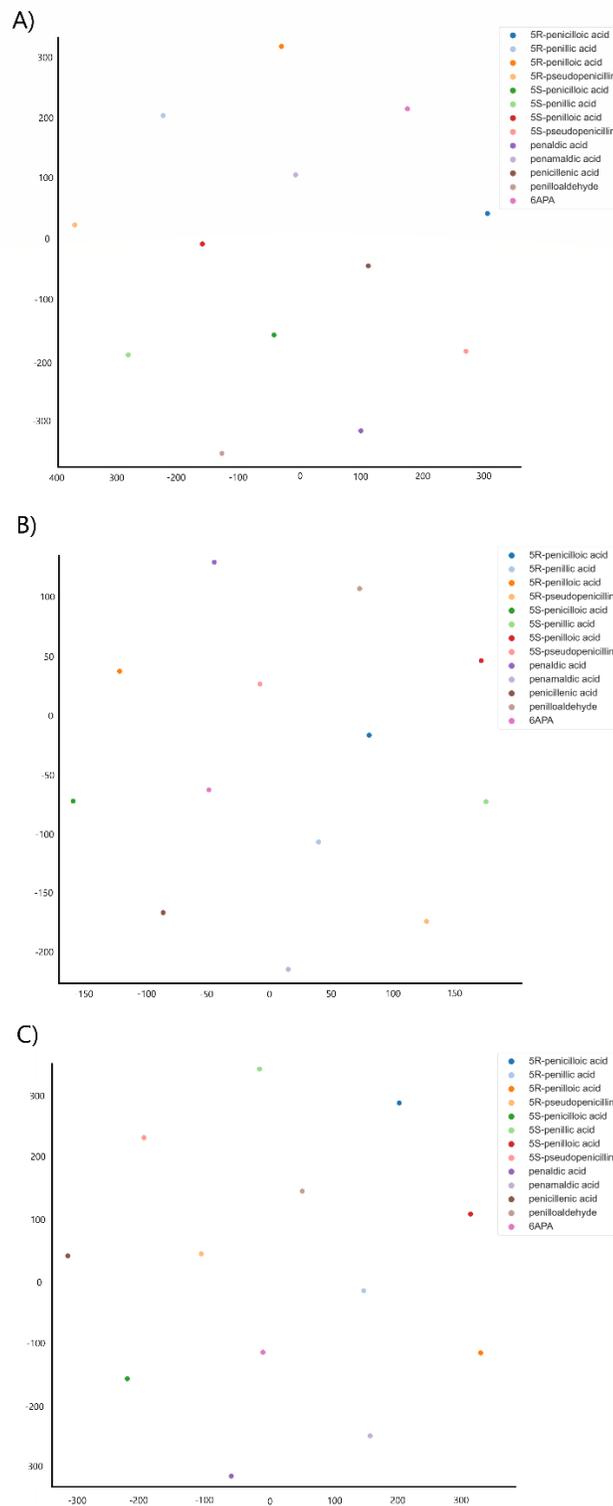
**Figure 4.4** Scatterplot of the t-SNE analysis of the results of molecular docking with data input of a) fitness, b) RMSD, c) fitness and RMSD

## 4.7 Predicted interactions of potential metabolites with key PBP3 residues

In order to verify the structural plausibility of the low RMSD predicting docking poses, we compared the 3D orientation of penicilloic acid and pseudopenicillin of piperacillin, as well as 6-APA, with that of the crystal structure for docked piperacillin. The antibiotic piperacillin interacts with the residue Ser294 with a covalent bond, and with residues Ser349, Ser485, Thr487 and Tyr503 with hydrogen bonds (Figure 4.5). In comparison, (5*R*)-penicilloic acid interacted with Ser485 and Thr487, (5*S*)-penicilloic acid interacted with Ser294, Ser349 and Thr487. (5*R*)-pseudopenicillin interacted with Ser349, Ser485 and Thr48, (5*S*)-penicilloic acid interacted with Ser294. (5*S*)-pseudopenicillin interacted with Phe533, penamaldic acid interacted with Thr487 and 6-APA interacted with Ser294 and Ser349 (Table 4.2); these predicted interactions are all with hydrogen bonds (Figure 4.5). Thus, each of the high fitness and low RMSD metabolites was predicted to interact with PBP3 with at least one residue in common with piperacillin.

| | SER294 | SER349 | SER485 | THR487 | PHE533 |
|---|---|---|---|---|---|
| piperacillin | ■ | ■ | ■ | ■ | ■ |
| (5*R*)-penicilloic acid | | | ■ | ■ | |
| (5*R*)-pseudopenicillin | | ■ | ■ | ■ | |
| (5*S*)-penicilloic acid | ■ | ■ | | ■ | |
| (5*S*)-pseudopenicillin | | | | | ■ |
| penamaldic acid | | ■ | | ■ | |
| 6APA | ■ | ■ | | | |

***Table 4.2***. The filled boxes indicate interactions between residues of the PBP and antibiotics or metabolites based on the structural interaction between the ligand and the protein in Figure 4.5.
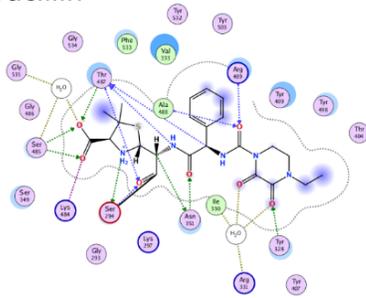
**Figure 4.5** The interaction between the residues of the penicillin binding protein and A) piperacillin, B) (5R)-penicilloic acid C) (5S)-penicilloic acid, D) (5R)- pseudopenicillin, and E) (5S)-pseudopenicillin, F) 6APA and G) penamaldic acid of piperacillin.
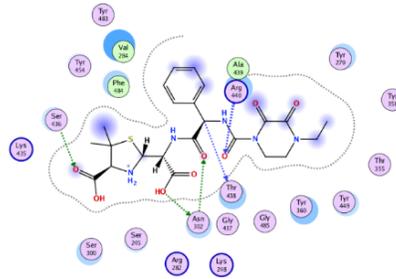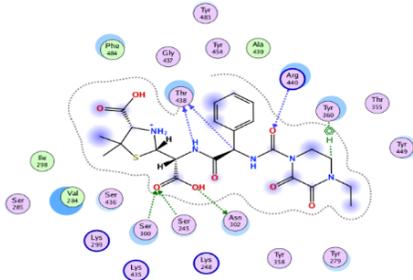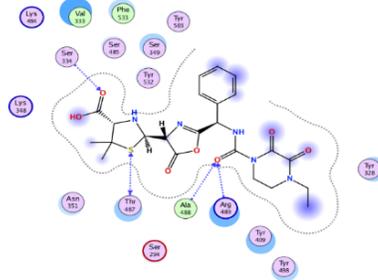
## 4.8 Discussion

In this chapter, docking predictions suggested that the metabolites (5*R*)- and (5*S*)-pseudopenicillin, (5*R*)- and (5*S*)-penicilloic acid and 6APA could bind PBPs. Each of the metabolites were classified as different categories: high fitness (79 to 93) and low RMSD (1.3 Å to 3.2 Å) with (5*R*)- and (5*S*)-penicilloic acid, (5*R*)- and (5*S*)-pseudopenicillin and 6APA, high fitness level (87.6 to 109) and high RMSD (6.1 Å to 7.8 Å) with (5*R*)- and (5*S*)-penilloic acid, penamaldic acid and penicillenic acid and low fitness level (76 to 82) and high RMSD(8.1 Å to 8.9 Å) with (5*R*)- and (5*S*)-penillic acid, penaldic acid, penilloaldyhyde. The molecular docking provided some ideas of which metabolites could have high affinity to the PBP and had the first step screening for metabolites suitable for the next step investigation, molecular dynamics simulations, in Chapter 5. Interestingly, the analysis results of t-SNE showed that some metabolites did not share similar features: (5*R*)-penillic acid in considering fitness only and (5*R*)-pseudopenicillin in considering RMSD only, and (5*R*)-penicilloic acid, penaldic acid and penicillenic acid when both fitness and RMSD were considered.

# Chapter 5

## Molecular Dynamics simulations

In the previous chapter, molecular docking was used to predict possible bioactive metabolites using the fitness level and the structural orientation of the metabolites as predicted by GOLD. In this chapter, a different computational technique, Molecular Dynamics (MD), was used to give more detailed predictions of bioactivity based upon the behavior of the molecules at the atomic level.

## 5.1 Introduction to Molecular Dynamics

MD (Durrant & McCammon, 2011) is a powerful computational technique used to study the behavior of molecules and materials at the atomic level. MD simulations are particularly effective because they use the equations of motion for atoms and molecules in a system, using either classical or quantum mechanics.(Tuckerman & Martyna, 2000). MD is widely employed in chemistry, materials science, and biology, with examples including protein structure, ligand binding, etc. (Wu et al., 2023).Examples of successful use of MD in biomolecular processes, including conformational change (Orellana, 2019), ligand binding(Swegat et al., 2003), and protein folding (Duan et al., 2019). The development of molecular dynamics simulations has revolutionized our understanding of the behavior of materials and molecules. These simulations have become an essential tool for researchers in many fields, enabling them to explore the behavior of complex systems that would be difficult or impossible to study experimentally.

Thus, MD simulations provide a detailed understanding of the dynamics and thermodynamics of materials and molecules, enabling researchers to explore the behavior of molecules and materials at the atomic level, including the motion of individual atoms and the formation of chemical bonds. Moreover, these simulations provide information about the thermodynamic properties of molecules, such as temperature, pressure, and energy.

They predict the movement of every atom in a protein or other molecular system over time, based on a general model of interatomic interactions force field (Duan et al., 2019; MacKerell Jr, 2001; Orellana, 2019; Swegat et al., 2003). Importantly, such simulations can also predict how biomolecules will respond at an atomic level to perturbations such as mutation, phosphorylation, protonation, or the addition or removal of a ligand.

## 5.2 Preparation of Molecular dynamics simulations

### 5.2.1 MD simulations

In order to further assess potential for metabolites to bind into the PBP pocket, MD simulations were run with the highest scoring metabolites from Chapter 4, (5*R*)- and (5*S*)-penicilloic acid, (5*R*)- and (5*S*)- pseudopenicillin and 6APA, with piperacillin as a positive control, and three negative controls: penamaldic acid, a decoy ligand with low fitness score, and a decoy ligand with high fitness score. The MD simulation outcomes were subjected to further analysis: root-mean-square fluctuation (RMSF) to assess whether structure of the system was in equilibrium; interaction frequencies between the metabolite and binding pocket atoms; and the position and structural orientation of the metabolites when they were sitting within the binding pocket.

### 5.2.2 Preparation for MD

The quick MD simulator module in CHARMM-GUI (Brooks et al., 2009) was used to add hydrogen atoms to the crystal structure, LYS297 was protonated to make the system pH7.4 and solvate the system with the standard TIP3P model (Mark & Nilsson, 2001), and apply periodic boundary conditions under the CHARMM36 force field (J. Lee et al., 2016; MacKerell Jr, 2001). The ligand file was first performed energy minimization same as Chapter 4, 4.1.1, and prepared by CHARMM 5.2.0(Lee et al., 2016) with the corresponding coordinate and parameter force field file of the ligand.

## 5.2.3 Periodic boundary conditions

Periodic boundary conditions are a set of boundary conditions often used to simulate systems that extend infinitely or are much larger than the computational domain. The periodic boundary conditions were used to create a simulation cell and replicated to full the whole system. Figure 5.1 shows how the cells extended infinitely. When a particle within the simulation cell exited one side of the cell, it re-entered from the opposite side. The interactions between the particles were calculated as they occurred within the cell, even when the actual distance between them were two neighboring cells. The interaction of a pair of particles in neighboring cells was calculated as if they were in the same cell.
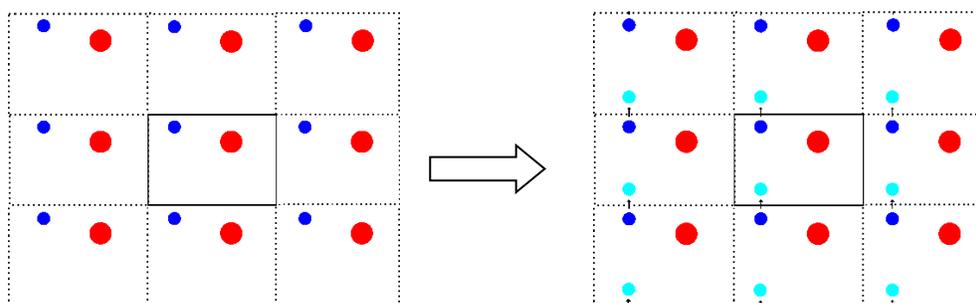


**Figure 5.1** Periodic boundary conditions allow a simulation to extend a single rectangular cell into an infinite array of cells. The rectangular box surrounded by a solid line is the simulation cell. The boxes surrounded by dashed lines represent copies of the simulation cell, extended into infinity. The red dot is a fixed particle and the blue dot is a moving particle. As the blue dot crosses the boundary at the top of the simulation cell, in the simulation it appears at the bottom of the simulation cell (light blue dot), which represents that particle moving to the next cell up in the extended space (dot with another colour).

## 5.2.4 parameters for MD

CHARMM36 is the latest version of the CHARMM potential set for proteins, is the product of a recent reparametrization aimed at improving the balance between the sampling of helical and extended conformations. It was optimized using the standard TIP3P water model. As

one atomic angle parameter for the target protein (CG2O3-CG3C51-NG3C51) was not present in the force field settings, initial estimates for that angle were obtained using the CGenFF atom typing program(Vanommeslaeghe & MacKerell, 2012). The system was solvated in a truncated octahedral periodic boundary cell with edge distances 10 Å from the protein surface. As octahedral cell was isotropic which had equal dimensions along all axes. The symmetry of the octahedron ensures the ions were distributed uniformly in all directions. Also, the octahedral shape helped to minimize edge effect compared to cubic cells which might experience different interactions due to the lack of neighboring molecules in certain directions.

## 5.2.5 Configuration of MD simulation

Energy minimization for each system was performed in NAMD 1.12(Phillips et al., 2020) using the standard conjugate gradient algorithm for 10,000 steps. All heating, equilibration, and production dynamics were performed using NAMD with a time step of 2 fenosceonds (fs), the CHARMM36 force field (J. Lee et al., 2016), and periodic boundary conditions. The parameters for the metabolites were taken from the CHARMM general force field (CGenFF) of drug-like small molecules and the force field file generated by CHARMM 5.2.0(J. Lee et al., 2016).

The system was heated from 0 to 298 K in increments of 5K by temperature reassignment, where the velocities of all the atoms in the system are reassigned so that the entire system is set to the target temperature with Langevin thermostat. The velocities were reassigned every 500 time-steps for 50,000 time-steps in the NVT ensemble. The systems were equilibrated for another 5ns in the NPT ensemble, with Langevin dynamics the pressure set to 1 atm. Production dynamics were run in the NPT ensemble for 30ns for each of the metabolites and the antibiotics. All frames were saved every 1500 steps.

### 5.2.5.1 Choice of timestep

The timestep represents the interval between the calculations of the positions and velocities of the particles in the system. A smaller timestep generally provided more accurate results because it allowed for a more precise representation of the system's dynamics. However, a smaller timestep is required for more computational resources. A 2fs time step was considered a good balance between accuracy and computational efficiency for the systems. As the protein's motion is in scale of picoseconds and the bond vibrations' frequency is more than 2fs. Therefore, 2fs timestep was appropriated.


### 5.2.5.2 Ensemble of MD

There are two ensembles that are used in MD: canonical ensemble (NVT) (Melander, 2021) and isothermal-isobaric ensemble (NPT)(Evans & Morriss, 1983). In NVT, the number of particles (N), volume (V), and temperature (T) are conserved. This ensemble is used to simulate the systems contact with a heat bath, with constant temperature(Melander, 2021). In NPT, the number of particles (N), pressure (P), and temperature (T) are conserved. This ensemble is suitable for simulating the system under constant temperature and pressure(Evans & Morriss, 1983). For these simulations, the initiation was carried out under NVT until the system equilibrated with stable temperature. Then the system was transitioned into NPT during the production phase with the introduction of constant pressure. This shift allowed the system to undergo volume adjustment in response to the change of pressure during the transition. Therefore, this reached a more realistic representation of the system.


### 5.2.5.3 Thermostat and barostat

The temperature of the system is proportional to the average kinetic energy of all particles within the system. The purpose of the thermostat is to ensure the average temperature of the system to be the desired temperature. The Langevin thermostat (described in detail in Section 2.3.2) is used in MD to maintain the temperature. During MD, the positions and the velocities of the atoms are updated at each time step, and the velocities are adjusted using a damping term

and a random force is added to the system. The damping term and the random force combine to ensure that the system remains at the desired temperature.

The pressure within the system is regulated by adjusting the position of the piston according to the difference between the actual pressure and the target pressure. The Langevin Piston (described in detail in Section 2.3.2) is used in MD to maintain the pressure. During MD, the position and the velocity of the piston is updated at each time step, which ensures the pressure remains at the target pressure.

## 5.2.6 Analysis methods of MD simulation results

The analysis of the MD simulation outcomes was performed in VMD 1.94a51 (Humphrey et al., 1996), including use of RMSD, RMSF  and interaction frequencies between the ligand and key residues of the PBP. Time averaged trajectories were used to present the system under long-term equilibrium state. The time averaged trajectories were computed by averaging the positions, velocities of the molecules over the equilibrium state (after frame 400). This could help to remove short-term fluctuations in structure and extract information of the overall system.

$$x = \frac{1}{T - T_0} \int_{T_0}^{T} x(t) dt$$

$x$ represented the time averaged position of the particle.
$x(t)$ represented the instantaneous position of the particle at time $t$.
The integral is taken over the time interval from a particular starting time point $T_0$ to $T$. The position obtained by dividing the integral by the difference in time $T - T_0$.

### 5.2.6.1 RMSD and RMSF

The RMSD trajectory was calculated as the average RMSD of the protein including the ligand in each frame. The RMSD quantifies how the structure differs from the reference over time. The RMSD calculation was based on the backbone atoms of the protein. The equilibration was monitored by root mean square deviation (RMSD) trajectory analysis. Root-mean-square-fluctuation (RMSF) was calculated to quantify flexibility of the individual residues over the simulations, the RSMF can reveal which areas of the system are the most mobile. An area of the structure with high RMSF values indicates high mobility.

### 5.2.6.2 Interaction frequency between the ligand and key residue of PBP

The interaction frequency between the ligand and each target residue was the proportion of 1000 frames in which the distance between the ligand that that residue was less than 3Å. The interaction frequency is calculated for the systems with piperacillin, (5*R*)- and (5*S*)-penicilloic acid, (5*R*)- and (5*S*)- pseudopenicillin, penamaldic acid, 6APA and decoy molecules.

## 5.3 Results

### 5.3.1 RMSD analysis of the MD simulations

During the MD simulations, the structure and the RMSD of the backbone atoms relative to the initial structure of each frame remained stable when the complex is in equilibrium. RMSD trajectory analysis was performed on the binding pocket (between residue 250 and residue 504), as the residues not associated with the binding pocket had very high RMSDs, which adversely bias the average RMSD when calculating RMSD for the whole frame (Figure 5.2). The RMSDs in the binding pocket varied between 0.5 Å to 2 Å indicating that the systems are stable.

**A)**

**B)**

**C)**
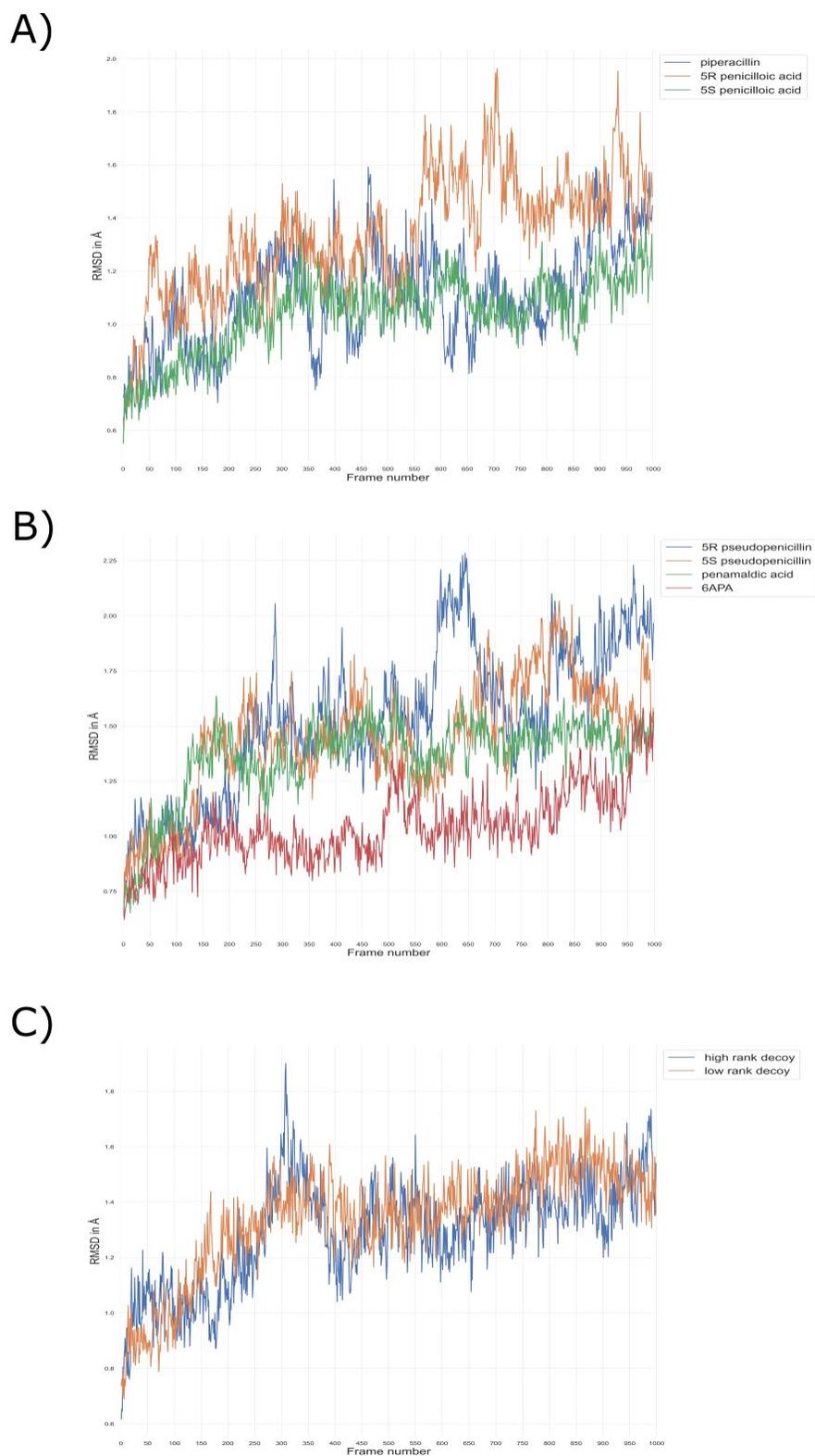
*Figure 5.2* a-d) RMSD of a) piperacillin, (5R)-penicilloic acid and (5S)-penicilloic acid, b) (5R)-pseudopenicillin, (5R)-pseudopenicillin, penamaldic acid and 6APA, c) decoys of the binding pocket (residue 250-477). The RMSD shows the system is relatively stable with their low RMSD.

## 5.3.2 RMSF analysis of the MD simulations

The RMSF (Figure. 5.3a) showed the deviation of the position of an atom with respect to an initial position over all frames. The RMSF describes the flexibility or rigidity of different regions of the system. High RMSF atoms indicate increased flexibility while low RMSF suggest rigidity. This might use to determine the stability of specific regions in a structure. The binding pocket (residues 250-504) was shown to be stable because it had a low RMSF ranging between 0.5 Å to 1.5 Å (Figure. 5.3b). The N- and C- termini of the protein have high RMSF, suggesting that the structure of the N- and C- termini moved considerably across all frames; thus these termini might affect the value in RMSD analysis when determining whether the system was stable. MD simulations of piperacillin showed that the key binding residues (Ser294, Ser349, Ser485, Thr487 and Tyr503) were all close to their positions and orientations in the crystal structure (RMSD of 1 Å) (Figure 5.2), giving confidence in the MD simulations (Figure 5.3c).

## 5.3.3 Interaction frequency of the metabolites with the residues

Piperacillin interacted with the protein and was stabilized by the residues Ser294, Ser349, Ser485, Thr487 and Tyr503, with very high frequency. In order to assess binding of metabolites in more detail, the interactions between these residues and the ligands were assessed, by counting the frequency of distance below 3 Å across all frames (Table 5.1). The interactions of (5*R*)-pseudopenicillin and 6APA with key residues was close to that of piperacillin, with ligand-protein interactions at four of the five residues, Ser294, Ser485, Thr487, Phe533 for (5*R*)-pseudopenicillin and Ser294, Ser349, Ser485, Thr487 for 6APA. (5*S*)- penicilloic acid and (5*S*)- pseudopenicillin were predicted to interact with two residues, Ser294, Thr487 for (5*S*)- penicilloic acid and Thr487, Phe533 for (5*S*)-pseudopenicillin, while (5*R*)-penicilloic acid and penamaldic acid interacted at just one residue, Thr487. The decoy ligands showed relatively weak interactions with the protein, with one residue (Ser294) and two residues (Ser294, Ser349) respectively.
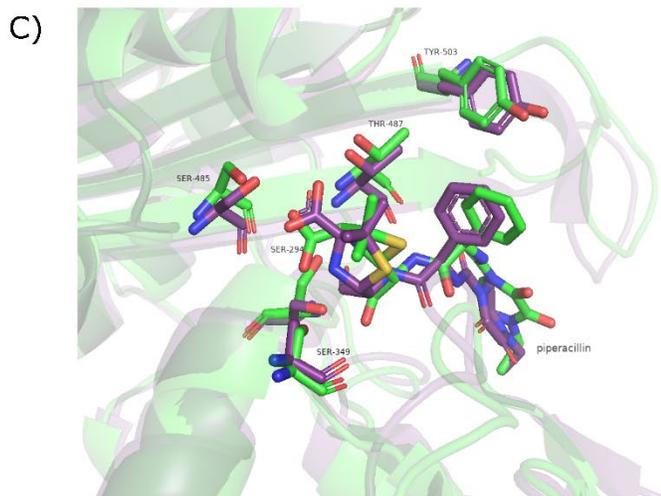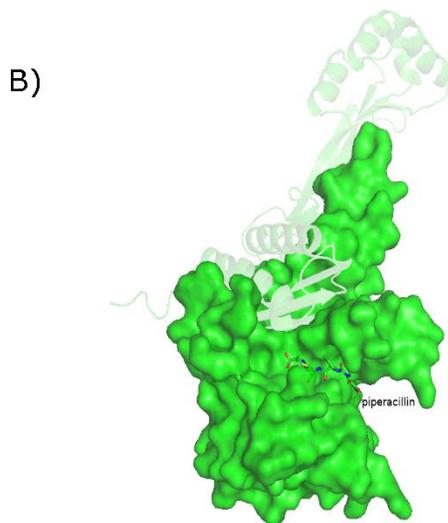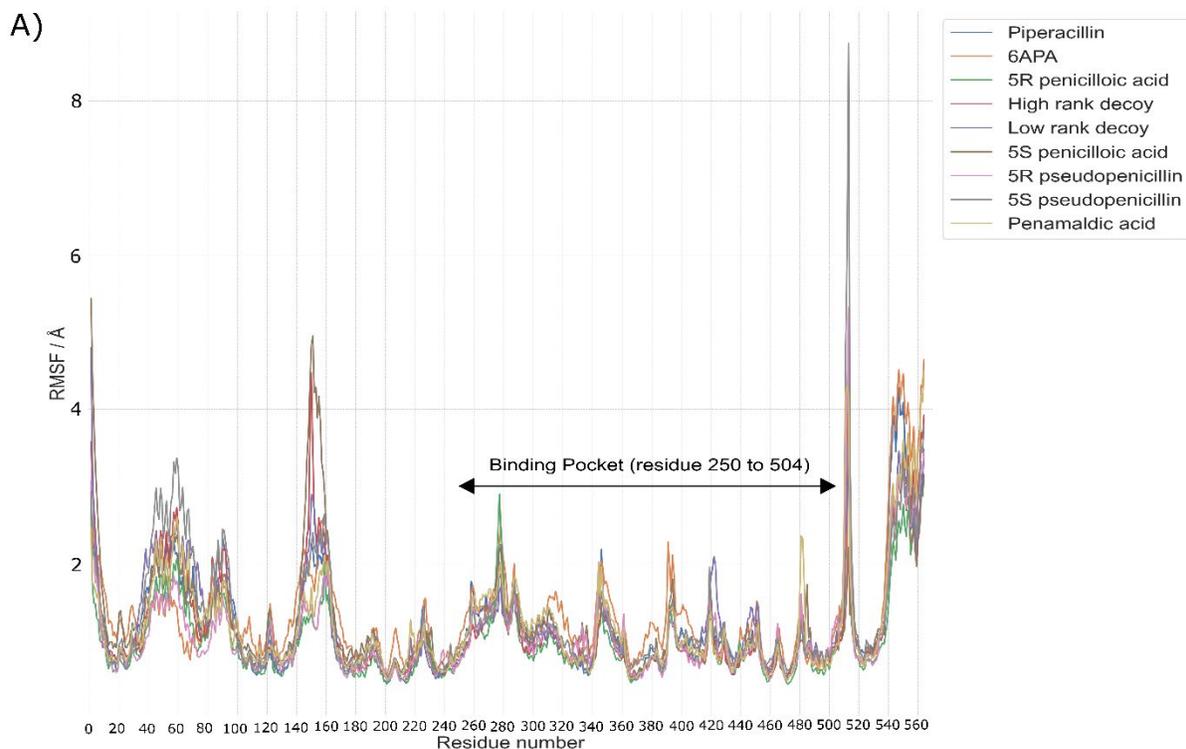
**Figure 5.3** a) RMSF analysis of all MD simulations. It shows the RMSF of each residue among all frames. The binding pocket of the protein (residue 250 to 504) has a relatively fixed 3D structure in the binding pocket (RMSF below 2 Å), with the areas with highest RMSF lying outside of the binding pocket. b) the structure of rigid binding pocket residue 250 to 504 with high stability suggested by RMSF analysis was indicated by the green surface, c) the orientation of the key residues and antibiotics piperacillin of MD simulation (green) and the original pdb file (dark purple) showing good alignment between the MD simulations and crystal structure.

## 5.3.4 Positional orientation of the metabolites when they bound to PBP

(5*S*)- penicilloic acid, (5*R*)- and (5*S*)-pseudopenicillin, 6APA located on the center of the binding pocket (Figure 5.5a) and so had the potential to interact with the key residues. (5*R*)-penicilloic acid (3.1 Å below piperacillin) and penamaldic acid (on right of piperacillin with 6.9 Å) were located on the right side of the binding pocket, while the low rank decoy was in the left (with 4.4 Å); none of these molecules were predicted to enter the binding pocket (Figure 5.4b). The high rank decoy entered the binding pocket but the orientation of the ligand was different from piperacillin (3.5 Å difference between the center of decoy and pipercillin) (Figure 5.4b), so the molecule was not predicted to interact strongly with the protein as they had only one residue and two residue interactions (Table 5.1). (5*R*)-penicilloic acid and penamaldic acid lacked the interaction of the key residue Ser294 (Table 5.1), while the decoys interacted with Ser294 but with different location in the binding pocket relative to piperacillin (Figure 5.4b).

| | Ser294 | | Ser349 | | Ser485 | | Thr487 | | Phe533 | |
|---|---|---|---|---|---|---|---|---|---|---|
| piperacillin | 100% | (blue) | 100% | (blue) | 100% | (blue) | 100.0% | (blue) | 75% | (blue) |
| (5*R*)- penicilloic acid | 36% | | 12% | | 0% | (blue) | 89% | (blue) | 49% | |
| (5*R*)- pseudopenicillin | 100% | | 2% | | 100% | (blue) | 99% | (blue) | 90% | |
| (5*S*)- penicilloic acid | 83% | (blue) | 40% | (blue) | 1% | | 98% | (blue) | 48% | |
| (5*S*)- pseudopenicillin | 68% | | 44% | | 0% | | 91% | | 88% | (blue) |
| penamaldic acid | 31% | | 30% | (blue) | 22% | | 90% | (blue) | 33% | |
| 6APA | 100.0% | (blue) | 100% | (blue) | 100% | | 100% | | 73% | |
| high rank decoy | 98% | | 41% | | 0% | | 32% | | 34% | |
| low rank decoy | 100% | | 100% | | 18% | | 37% | | 13% | |

***Table 5.1***. Interaction frequencies between molecules and PBP residues as predicted by MD simulations and docking. Percentage is the frequency in which residue is within 3 Å from the ligand in MD simulations; interactions below 3 Å are predicted to be strong enough to stable the ligand. Colours represent synthesis of MD and docking results: yellow are interactions predicted from MD; blue are interactions predicted from docking; no colour signifies no predicted interactions. Overall, (5*R*)-pseudopenicillin, (5*S*)-penicilloic acid and 6APA are predicted to bind to the PBP3 binding pocket.
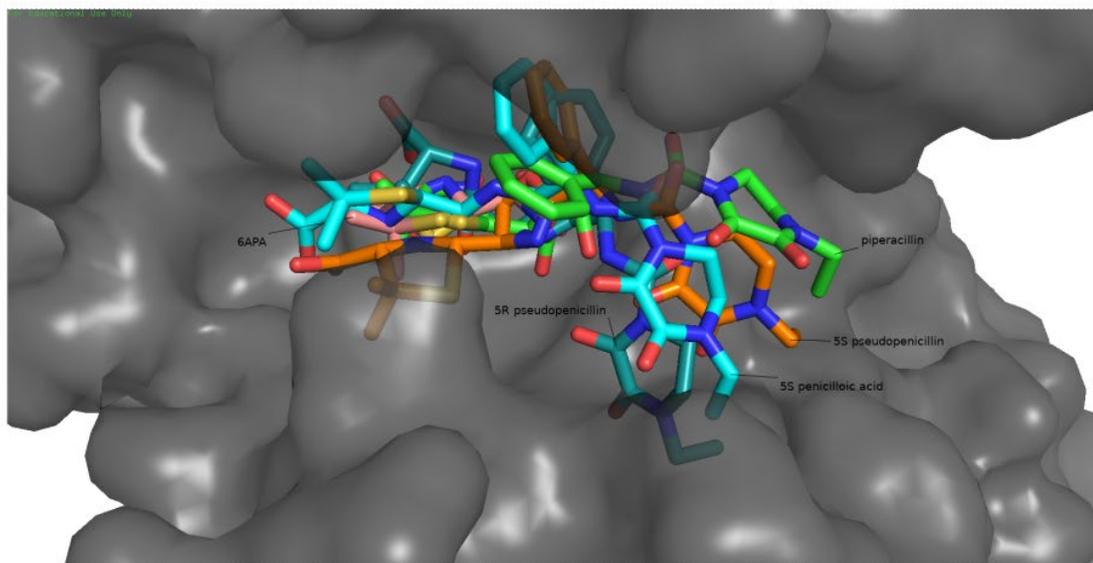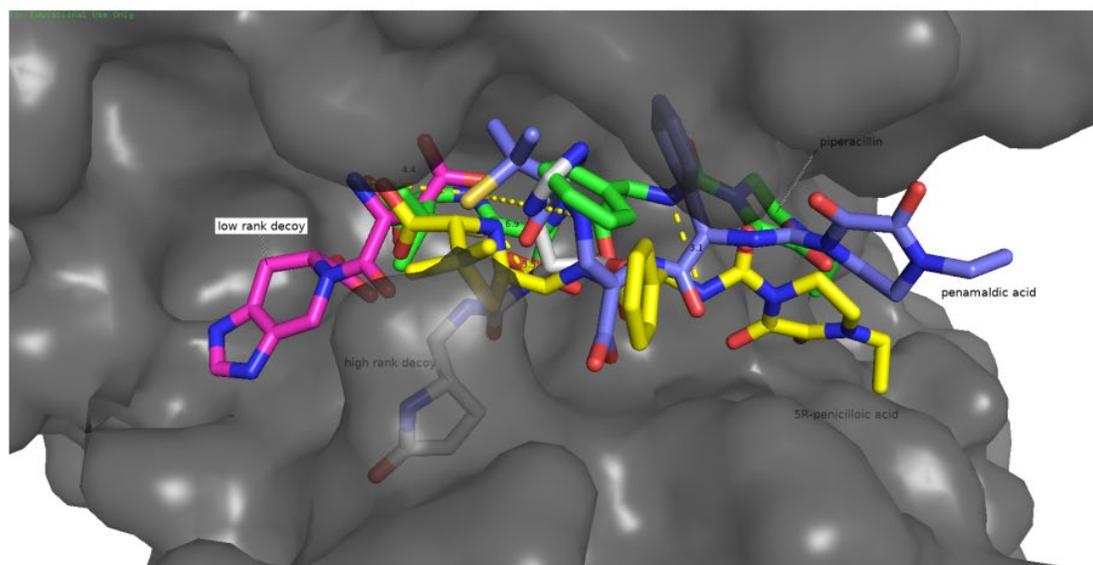
**Figure 5.4** The orientation of the ligand within the protein. A) Piperacillin (green), (5*S*)-penicilloic acid (light blue), (5*R*)-pseudopenicillin (dark blue), (5*S*)-pseudopenicillin (orange) and 6APA (pink) sit in the binding pocket of PBP. B) (5*R*)-penicilloic acid (yellow) and penamaldic acid (blue). (5*R*)-penicilloic acid (3.1 Å below piperacillin) and penamaldic acid (on right of piperacillin with 6.9 Å) locate on the right side of the binding pocket and do not fully enter it. The low rank decoy (purple) is in the left (4.4 Å difference between the center of the decoy and piperacillin) and it is not in the binding pocket. The high rank decoy enters the binding pocket but it binds differently from the piperacillin (3.5 Å difference between the center of decoy and pipercillin).

## 5.3.5 Structural orientation of key residues of PBP when metabolites bound to PBP

The orientation and alpha carbon positions of the protein residues between metabolite and piperacillin binding were compared (Figure 5.5 and Table 5.2 - 5.3). (5*R*)-psudopenicillin, (5*S*)-penicilloic acid and 6APA had a similar pattern of interaction of key residues and torsion angles as piperacillin (Table 5.2), further confirming the prediction that they bound to the PBP. There was a large difference in RMSD of the alpha carbon of (5*R*)- (4.1 Å) and (5*S*)- (2.5 Å) pseudopenicillin on Ser349 (Table 5.3). The residue Ser349 moved outwards, and the binding pocket became larger. These might cause the interaction (2%) between the pseudopenicillin and the protein became weaker but also the movement of the alpha carbon of Ser349 was to adopt difference shape of pesudopenicillin (Table 5.3) and to allow the binding of psudopenicillin towards the PBP. The torsion angles of the residues among systems of (5*R*)-psudopenicillin, (5*S*)-penicilloic acid and 6APA showed they were similar to piperacillin (Figure 5.5).
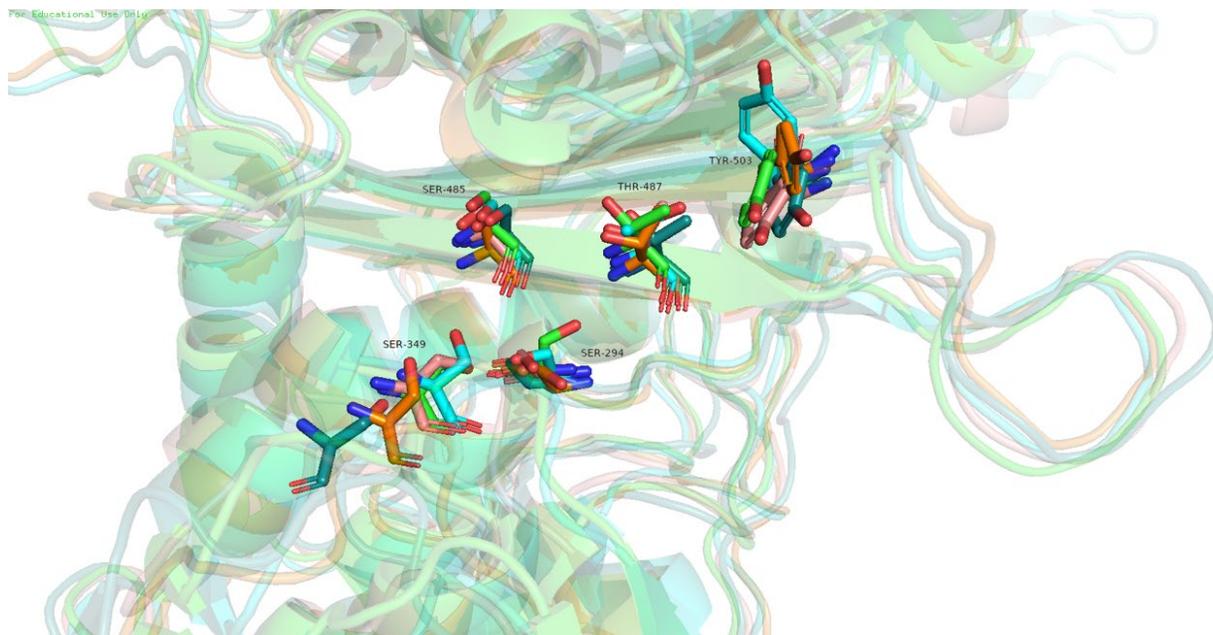


***Figure 5.5*** Structural orientations of the residues Ser294, Ser349, Ser485, Thr487 and Tyr503 in MD simulations of piperacillin (green), (5*R*)- (dark blue) and (5*S*)- (orange) pseudopenicillin, (5*S*)-penicilloic acid (sky blue) and 6APA (pink) showing the movement of the alpha carbon and the torsion angle of the metabolites when comparing with the piperacillin (Table 5.2 and 5.3). The movement of the alpha carbon of the Ser349 of pseudopenicillin makes a larger binding pocket when compared with the binding pocket of piperacillin.

| | Ser294 | | Ser349 | | Ser485 | | Thr487 | | Tyr503 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Psi | Phi | Psi | Phi | Psi | Phi | Psi | Phi | Psi | Phi |
| Piperacillin | -52.2° | -27.6° | 30.3° | 66.5° | 159.4° | -93.5° | 108.7° | -125.7° | 160.7° | -132.6° |
| (5*R*)-pseudopenicillin | -10.9° | -72.2° | 156.3° | 65° | 156° | -81.3° | 133.7° | -141.9° | 165.3° | -141° |
| (5*S*)-penicilloic acid | -12.1° | -63.6° | 17.1° | 60.5° | 166.8° | -99.7° | 143.1° | -144.5° | 154.9° | -159.9° |
| (5*S*)-pseudopenicillin | -3.5° | -76.3° | 48.3° | 42.5° | 152.8° | -96.1° | 153.1° | -122.9° | 152.9° | -135.3° |
| 6APA | -1.4° | -70.2° | 26.2° | 82.9° | 160.9° | -69.6° | 119.6° | -155.9° | 156.6° | -129.8° |

***Table 5.2***. Table of torsion angles (Psi and Phi) of the residues of different system (Piperacillin, (5*R*)-pseudopenicillin, (5*S*)-penicilloic acid, (5*S*)-pseudopenicillin and 6PAP). The box filled with yellow shows there is a large change in the torsion angle.

| | Distance of alpha carbon movement (Å) | | | | |
|---|---|---|---|---|---|
| | Ser294 | Ser349 | Ser485 | Thr487 | Phe533 |
| (5*R*)-pseudopenicillin | 0.8 | 4.1 | 0.9 | 1.5 | 0.8 |
| (5*S*)-penicilloic acid | 0.4 | 0.8 | 0.5 | 0.9 | 0.4 |
| (5*S*)-pseudopenicillin | 0.8 | 2.5 | 1.2 | 0.9 | 0.5 |
| 6APA | 0.3 | 0.4 | 0.4 | 1.0 | 0.9 |

***Table 5.3***. Table of distance moved of the alpha carbons of the metabolites compared with piperacillin. The box filled with yellow shows there is a large change in the alpha carbon atom of the residue and this enlarged the binding pocket making it not likely to interact with the ligand.

## 5.4 Summary

The results of Chapter 4 suggested that (5*R*)- and (5*S*)-pseudopenicillin, (5*R*)- and (5*S*)-penicilloic acid and 6APA could bind to PBP3. In this chapter, these results were refined by using MD to study the dynamic behavior (motion and fluctuations of the molecules) of the resulting protein-ligand complexes, which molecular docking cannot do. The simulation environment could set to the environment we were interested to simulate the reality. The predictions of possible binding of (5*R*)-pseudopenicillin, (5*S*)-penicilloic acid and 6APA into the binding pocket of PBP3 were made from the MD simulation. (5S)-penicilloic acid has already been found complexed to PBP3 in a stable crystal structure (Van Berkel et al., 2013), lending confidence to our predictions for (5R)-pseudopenicllin and 6APA. In the MD simulations of (5*R*)-pseudopenicillin, four of the five main binding residues are in similar positions as compared with the MD simulation of penicillin (Table 5.3), with very high probabilities of interaction (Table 5.2). The most important residue, Ser294, interacting 99.8%, with its carbon atom moving by only 0.8 Å between the simulations of the two ligands. The exception is Ser349, whose alpha carbon moves by 4.9 Å. Although the position of Ser349 moved to adopt the shape of (5*R*)-pseudopenicillin, Ser349 did not interact with (5*R*)-pseudopenicillin (2%), so this may slightly weaken the interaction of the PBP with (5*R*)-pseudopenicillin.

# Chapter 6

## Discussion and Conclusion

### 6.1 Summary of molecular docking and molecular dynamics

In Chapter 4, molecular docking was used to predict that the metabolites (5*R*)- and (5*S*)-pseudopenicillin, (5*R*)- and (5*S*)-penicilloic acid and 6APA should have have high affinity and similar binding poses to their native antibiotics. Among these metabolites, (5*R*)-pseudopenicillin, (5*S*)-penicilloic acid and 6APA were found to have high possibility to be bioactive as they interact PBP similarly to their native antibiotics using Molecular Dynamics described in Chapter 5. This suggests that molecules can bind to the PBP not only though the covalent bonding between serine and the oxygen of the nitrogen quadrilateral ring but also though the non-covalent interaction between the molecules and the penicillin binding protein(Van Berkel et al., 2013). The data also show that the non-covalent binding (5*S*)-penicilloic acid and the antibiotic bind to the same binding site with similar orientation. Taken together, this gives confidence that these metabolites have the potential to bind to the target penicillin binding protein. If so, they may also and have biological effect, specifically providing some (non-clinical) level of antibiotic action, but possibly sufficient to provide selective pressure for resistance. Surprisingly, the results in Chapter5 that (5*R*)-penicilloic acid could not bind to PBP prove the prediction of t-SNE in Chapter 4 that (5*R*)-penicilloic acid does not show common features with other metabolites with high fitness level and low RMSD. Also, the distance of (5*S*)-pseudopenicillin in t-SNE is quite far away from (5*R*)-pseudopenicillin, (5*S*)-penicilloic acid and 6APA and this is also prove in Chapter 5 as (5*S*)-pseudopenicillin is suggested to be not binding.

## 6.2 Limitation for molecular docking and molecular dynamics in this study

An important limitation of our results is that we only had access to one structure of antibiotic complexed to its target protein in a single organism (4KQO). While the predictions have been successful, their generality would be improved with access to further structures of different PBPs, different penicillins, or different organisms. Moreover, molecular docking cannot predict the presence of unknown covalent binding (Groom et al., 2016), which may affect the predictions of how the ligand binds to the target protein. The scoring function used may not accurately represent the true binding affinity (Pantsar & Poso, 2018). Proteins are dynamic and could undergo conformational changes when binding and the docking algorithms might not calculate all range of possible interactions. Moreover, accurate modeling of the solvent is also needed for realistic predictions. (Dill et al., 2005; Finney, 2001)

MD simulations also have limitations (J. Lee et al., 2016), with results that can depend upon the parameter sets used (Guterres et al., 2021), and the molecular mechanics force field chosen. Parameterization can also lead to inaccurate predictions of the structure. The solvent molecules played an important role when doing a realistic simulation. Simple water model could simplify calculations but may not simulate all solvent effects accurately while a complex water models increased computational requirements. Also, simulating a large system or long timescale require more computational resources. The free binding energy was not calculated in this study but it could be calculated though MM/PBSA and MM/GBSA (E. Wang et al., 2019).

The uptake of the metabolites of the organisms also needs to be estimated; antibiotics with cytoplasmic or ribosomal targets need to enter the cell, including crossing the cell wall and membrane, in order to perform their functions (Cinquin et al., 2015). This step is untested in our analysis of metabolite docking and it is important that even if the metabolites are bioactive, it still needs to be entered the cell in order to do their function. The metabolic

state of microbes may influence the antibiotics susceptibility of the microbes (Cabral et al., 2019)and this may affect the amount and effect of uptake of antibiotics (Stokes et al., 2019). Finally, antibiotics in the penicillin family contain an unstable, highly strained, and reactive beta-lactam amide bond(Brouwers et al., 2020). While the beta-lactam bond is essential for the clinical efficacy of penicillins, it remains unknown that whether penicillin metabolites are sufficiently bioactive to have selective impact. Some details of the degradation pathways of penicillin are not yet fully understood; these may produce other metabolites which have not been tested in this study (Ho et al., 2011).

## 6.3 Discussion

In Chapter 3, different high-dimensional reduction methods were compared and t-SNE was identified as the most suitable method to analysis AMR phenotype data. This solved the problem that the raw data of the AMR data was usually too large to view effectively, and t-SNE could transform the high-dimensional data into 2/3-dimensional data without losing the underlying structure of the data in the original high dimensional space. AMR data (Baker et al., 2022; Saravanan & Raveendaran, 2013) could not tell the relationship between development of the resistance and the selective pressure. Computational methods, molecular docking and MD simulation had been used to estimate the affinity of between the metabolites and PBP and the possibility for binding the metabolites towards the PBP in Chapter 4 and 5. Metabolites (5$R$)- and (5$S$)-pseudopenicillin, (5$R$)- and (5$S$)-penicilloic acid and 6APA had been predicted by molecular docking to have high affinity and similar binding pose as the native antibiotics. Among these metabolites, (5$R$)-pseudopenicillin, (5$S$)-penicilloic acid and 6APA were predicted by MD simulations to have a high possibility of being bioactive as they interact with the PBP similarly to their native antibiotics. The results in Chapters 4 and 5 gave confidence that these metabolites had the potential to bind to the target penicillin binding protein and this shows the possibility that the metabolites could be bioactive towards the bacteria. The result in chapter 3 also showed bacteria collected with similar time and antibiotic features was grouped in the same colony as time is not present in the algorithm. This showed the antibiotic

resistance build up was related to the antibiotics used in the farm, which was remained in the animal product and the waste(Mann et al., 2021). Because of the lack of the detection method which only detected the presence of the antibiotics(Hanna et al., 2018; Munteanu et al., 2018), the presence of the small molecules, the metabolites of the antibiotics or compound from the industrial waste product, was not detected.

Therefore, any environmentally stable antibiotic metabolites could be important contaminants, as low antibiotic concentrations may increase genetic variability of microbes and select for resistance(Andersson & Hughes, 2014; Stanczak-Mrozek et al., 2017). The existing metabolites within the environment (Aldeek et al., 2016) could potentially raise the selective pressure on the bacteria which may lead to the selection for AMR(Tello et al., 2012). Meanwhile, a non-antibiotic related compound, 6-trifluoromethyl-3*H*-benzoimidazole-4-carboxylic acid [3-(1*H*-tetrazol-5-yl)-phenyl]-amide, inhibits the beta-lactamase which is used to degrade beta-lactam though catalytic reaction with Ki value 89nM and the clinical bacterial isolates shows the compound performs little effect on bacteria growth(Nichols et al., 2012). The compound enhances the performance of cefotaxime by inhibiting the beta-lactamase, the MIC of cefotaxime is reduced by 64-fold (from 32µg/mL to 0.5µg/mL) with the presence of the compound(Nichols et al., 2012). The compound binds to the active site of the beta-lactamase and blocks the active site, prevents the beta-lactam from entering the active site of beta-lactamase. This increases the possibility of small molecules having biological effects on microorganisms or increasing antibiotic performance.

Also, the structure of the active site of the beta-lactamase where beta-lactam enters is similar to the active site of the PBP(F. Wang et al., 2019). Two enzymes play different roles, beta-lactamase carries out the degradation of beta-lactam and PBP is involved in the synthesis of peptidoglycan. But the antibiotics are capable to bind to both beta-lactamase and PBP as they share similar active site structure(F. Wang et al., 2019). This may lead to the binding of 6-trifluoromethyl-3*H*-benzoimidazole-4-carboxylic acid [3-(1*H*-tetrazol-5-yl)-phenyl]-amide towards the PBP and cause effect on bacterial growth in the clinical bacterial

isolates(Nichols et al., 2012). Importantly, even sub-lethal concentrations of antibiotics can also drive selection(Andersson & Hughes, 2014). This shows 6-trifluoromethyl-3*H*-benzoimidazole-4-carboxylic acid [3-(1*H*-tetrazol-5-yl)-phenyl]-amide can drive selection for resistance and gives the evidence that industrial synthetic compound may also can bind to active site and raise for the selection for resistance.

The metabolites of the antibiotics may also bind to the beta-lactamase and inhibit the function of enzyme as 6-trifluoromethyl-3*H*-benzoimidazole-4-carboxylic acid [3-(1*H*-tetrazol-5-yl)-phenyl]-amide may bind to both beta-lactamase and PBP and the structure of active site of both beta-lactamase and PBP is similar (F. Wang et al., 2019). Also, phenylacetic acid (a plant auxin) and ethane-1,2-diol (an industrial chemical, commonly referred to as ethylene glycol) have been found complexed to penicillin G acylase of *E. coli* (McVey et al., 2001).  The antibiotic metabolites may also increase the performance of the antibiotics (Nichols et al., 2012) and reducing the MIC of the antibiotics as the function of beta-lactamase is inhibited. As the MIC of the antibiotics is reduced, the antibiotics kill the bacteria even the concentration of the antibiotics is much lower than the concentration suggested by estimation of laboratory. Meanwhile, the reduced MIC of the antibiotics may induce the selection of resistance with concentration lower than estimated(Sandegren, 2014). Therefore, the standard of minimum limit of the detection of the antibiotics in the environment may not be suitable with the presence of some potential inhibitor of the beta-lactamase.

The binding of these small molecules demonstrates that antibiotic metabolites and synthetic compounds (Bielen et al., 2017) also can pass through the cell wall of the bacteria and enter the cell. The cell wall of the bacteria prevents the uptake of harmful substances from the environment(Dörr et al., 2019), it is important that these molecules could enter the bacteria. Unfortunately, the mechanism of these molecules entering the bacteria remains unclear. The knowledge of the mechanism may help to improve the determination of what compound could pass though the cell wall and enter the cell. These will help the development for the detection of potentially bioactive compounds.

Also, the presence of the potential bioactive compound may lead to co-selection of the resistance. Resistance genes is linked on the same contigs, such as amoxicillin and trimethoprim resistance genes(Pouwels et al., 2018) and tetracycline and streptomycin resistance genes(Cadena et al., 2018). Therefore, treatment with one antibiotic may be selected for another antibiotic resistance by co-selection(Baker-Austin et al., 2006). Moreover, mutations or horizontal gene transfer of antibiotic resistance may also decrease the resistance of another antibiotic (Baym et al., 2016). Therefore, the presence of the penitential bioactive compound may lead to the raise of other resistance and enter to the genetic diversity which explain the emerge of the different patterns of resistance in the AMR study.

In summary, the use of the high-dimensional reduction method raises a benefit of analysis of the massive and complicated AMR data. As a good high-dimensional reduction could keep required main features, it gives a better understanding of the correlation of different resistance. Also, the prediction of the potential bioactive compounds suggests how the resistance is emerged and selected for. If the potential bioactive compounds exist in the environment, then they could act as selective agents for AMR. Therefore, new detection method of the potential bioactive compound needs to be developed as the existing detection method does not detect the presence of these substances. Also, a new standard of minimum limit of the detection of the antibiotics as the MIC of the antibiotic will be reduced by the presence of the inhibitor (the potential bioactive compounds)(Sandegren, 2014). This could have considerable significance for environmental surveillance for antibiotics to reduce antimicrobial resistance.

## 6.4 Future directions

In this study, only molecular docking and MD are used to determine the bioactivity of the metabolites. A practical experiment (e.g. Disc diffusion test, epsilometer test and colorimetric test) could further be used to test the bioactivity of the metabolites.
The disc diffusion test can simply determine the bioactivity of the metabolite, pre-impregnated paper discs containing the metabolites are placed on the surface of the

inoculated agar plate which the clear around the discs represent the bacterial growth has been prevented and the bioactivity of the predicted metabolites could be tested. Epsilometer test is similar to dis diffusion test but it can also determining the MIC of the metabolites against bacteria. The strips soak with a gradient of concentrations of the metabolites and placed on the surface of the inoculated agar plate. The point at which the border of the inhibition zone intersects the strip is read as the MIC value. For colorimetric test, metabolites are added to the bacteria bath and incubate in suitable temperature. As the bacteria growth, the bacteria bath becomes cloudy and increases in the reading of the colourmeter. If the metabolites have bioactivity towards the bacteria, the growth of the bacteria will be inhibited and the bacteria bath will remain clear.

A further X-ray crystallography can be used to see the exact structure of the complex between the PBP and the predicted metabolites. The complex of the PBP and the metabolite is condensed into crystallite form. The crystal is placed in the path of an X-ray beam. As the X-rays pass through the crystal, they are scattered by the electrons surrounding the atoms. This scattering produces a diffraction pattern on a detector, The pattern depends on the arrangement of the atoms in the crystal. The data can be used to generate a three-dimensional structure of the crystal which can confirm if the predicted metabolites truly bind to PBPs.

More metabolites of different antibiotic classes and bacteria could use for the testing to build up the systematic understanding of the mechanisms of the bioactivity of the metabolites. The PBP used in this study is a class B PBP which only catalyzed transpeptidation. The prediction method can extend to class A PBPs which can catalyze both the glycosltransfer and transpeptidation and it is also important to cell wall biosynthesis. The prediction of the transpeptidation domain have performed in this study, a further prediction of the glycosltransfer domain could be done with the used of the antibiotic meonomycin and its metabolites. Meonomycin (Ostash et al., 2022) inhibits the glycosyltransfer of PBP which catalyze the polymerization of copolymer chain of N-acetylglucosamine and N-acetylmuramic acid. The degradation pathway of the

meonomycin has been suggested (Adachi et al., 2006). Therefore, meonomycin and class A PBPs could be used for further prediction for the bioactive of the metabolites. This will complete the understanding of the effect of the antibiotic metabolites towards the whole process of the cell wall biosynthesis.

# References

Abedalwafa, M. A., Li, Y., Ni, C., & Wang, L. (2019). Colorimetric sensor arrays for the detection and identification of antibiotics. *Analytical Methods*, *11*(22), 2836–2854. https://doi.org/10.1039/C9AY00371A

Abramova, A., Berendonk, T. U., & Bengtsson-Palme, J. (2023). A global baseline for qPCR-determined antimicrobial resistance gene prevalence across environments. *Environment International*, *178*, 108084. https://doi.org/10.1016/J.ENVINT.2023.108084

Adachi, M., Zhang, Y., Leimkuhler, C., Sun, B., LaTour, J. V, & Kahne, D. E. (2006). Degradation and Reconstruction of Moenomycin A and Derivatives: Dissecting the Function of the Isoprenoid Chain. *Journal of the American Chemical Society*, *128*(43), 14012–14013. https://doi.org/10.1021/ja065905c

Adams, G. (2020). A beginner's guide to RT-PCR, qPCR and RT-qPCR. *The Biochemist*, *42*(3), 48–53. https://doi.org/10.1042/BIO20200034

Akhter, S., Bhat, M. A., Ahmed, S., Siddiqi, W. A., Ahmad, S., & Shrimal, H. (2023). Profiling of Antibiotic Residues in Surface Water of River Yamuna Stretch Passing through Delhi, India. *Water (Switzerland)*, *15*(3), 527. https://doi.org/10.3390/W15030527/S1

Aldeek, F., Canzani, D., Standland, M., Crosswhite, M. R., Hammack, W., Gerard, G., & Cook, J. M. (2016). Identification of Penicillin G Metabolites under Various Environmental Conditions Using UHPLC-MS/MS. *Journal of Agricultural and Food Chemistry*, *64*(31), 6100–6107. https://doi.org/10.1021/ACS.JAFC.5B06150

Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., & Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome Biology 2020 21:1*, *21*(1), 1–16. https://doi.org/10.1186/S13059-020-1935-5

Aminov, R. I. (2010). A Brief History of the Antibiotic Era: Lessons Learned and Challenges for the Future. *Frontiers in Microbiology*, *1*(DEC). https://doi.org/10.3389/FMICB.2010.00134

Andersson, D. I., & Hughes, D. (2014). Microbiological effects of sublethal levels of antibiotics. *Nature Reviews Microbiology 2014 12:7*, *12*(7), 465–478. https://doi.org/10.1038/nrmicro3270

Anwar, M., Iqbal, Q., & Saleem, F. (2020). Improper disposal of unused antibiotics: an often overlooked driver of antimicrobial resistance. *Https://Doi.Org/10.1080/14787210.2020.1754797*, *18*(8), 697–699. https://doi.org/10.1080/14787210.2020.1754797

Arya, S., Williams, A., Reina, S. V., Knapp, C. W., Kreft, J. U., Hobman, J. L., & Stekel, D. J. (2021). Towards a general model for predicting minimal metal concentrations co-selecting for antibiotic resistance plasmids. *Environmental Pollution*, *275*, 116602. https://doi.org/10.1016/J.ENVPOL.2021.116602

Bacanlı, M., & Başaran, N. (2019). Importance of antibiotic residues in animal food. *Food and Chemical Toxicology*, *125*, 462–466. https://doi.org/10.1016/J.FCT.2019.01.033

Baker, M., Williams, A. D., Hooton, S. P. T., Helliwell, R., King, E., Dodsworth, T., María Baena-Nogueras, R., Warry, A., Ortori, C. A., Todman, H., Gray-Hammerton, C. J., Pritchard, A. C. W., Iles, E., Cook, R., Emes, R. D., Jones, M. A., Kypraios, T., West, H., Barrett, D. A., … Stekel, D. J. (2022). Antimicrobial resistance in dairy slurry tanks: A

critical point for measurement and control. *Environment International*, *169*, 107516. https://doi.org/10.1016/J.ENVINT.2022.107516

Baker-Austin, C., Wright, M. S., Stepanauskas, R., & McArthur, J. V. (2006). Co-selection of antibiotic and metal resistance. *Trends in Microbiology*, *14*(4), 176–182. https://doi.org/10.1016/J.TIM.2006.02.006

Baym, M., Stone, L. K., & Kishony, R. (2016). Multidrug evolutionary strategies to reverse antibiotic resistance. *Science*, *351*(6268), aad3292. https://doi.org/10.1126/science.aad3292

Bellini, D., Koekemoer, L., Newman, H., & Dowson, C. G. (2019). Novel and Improved Crystal Structures of H. influenzae, E. coli and P. aeruginosa Penicillin-Binding Protein 3 (PBP3) and N. gonorrhoeae PBP2: Toward a Better Understanding of β-Lactam Target-Mediated Resistance. *Journal of Molecular Biology*, *431*(18), 3501–3519. https://doi.org/10.1016/J.JMB.2019.07.010

Bengtsson-Palme, J., & Larsson, D. G. J. (2016). Concentrations of antibiotics predicted to select for resistant bacteria: Proposed limits for environmental regulation. *Environment International*, *86*, 140–149. https://doi.org/10.1016/J.ENVINT.2015.10.015

Berberich, A. J., Ho, R., & Hegele, R. A. (2018). Whole genome sequencing in the clinic: empowerment or too much information? *CMAJ : Canadian Medical Association Journal*, *190*(5), E124. https://doi.org/10.1503/CMAJ.180076

Berendsen, B. J. A., Wegh, R. S., Memelink, J., Zuidema, T., & Stolker, L. A. M. (2015). The analysis of animal faeces as a tool to monitor antibiotic usage. *Talanta*, *132*, 258–268. https://doi.org/10.1016/J.TALANTA.2014.09.022

Berglund, F., Böhm, M. E., Martinsson, A., Ebmeyer, S., Österlund, T., Johnning, A., Joakim Larsson, D. G., & Kristiansson, E. (2020). Comprehensive screening of genomic and metagenomic data reveals a large diversity of tetracycline resistance genes. *Microbial Genomics*, *6*(11), 1–14. https://doi.org/10.1099/MGEN.0.000455

Berglund, F., Marathe, N. P., Österlund, T., Bengtsson-Palme, J., Kotsakis, S., Flach, C. F., Larsson, D. G. J., & Kristiansson, E. (2017). Identification of 76 novel B1 metallo-β-lactamases through large-scale screening of genomic and metagenomic data. *Microbiome*, *5*(1), 134. https://doi.org/10.1186/S40168-017-0353-8

Bhagat, C., Kumar, M., Tyagi, V. K., & Mohapatra, P. K. (2020). Proclivities for prevalence and treatment of antibiotics in the ambient water: a review. *Npj Clean Water*, *3*(1), 42. https://doi.org/10.1038/s41545-020-00087-x

Bielen, A., Šimatović, A., Kosić-Vukšić, J., Senta, I., Ahel, M., Babić, S., Jurina, T., González Plaza, J. J., Milaković, M., & Udiković-Kolić, N. (2017). Negative environmental impacts of antibiotic-contaminated effluents from pharmaceutical industries. *Water Research*, *126*, 79–87. https://doi.org/10.1016/J.WATRES.2017.09.019

Bottery, M. J., Pitchford, J. W., & Friman, V. P. (2020). Ecology and evolution of antimicrobial resistance in bacterial communities. *The ISME Journal 2020 15:4*, *15*(4), 939–948. https://doi.org/10.1038/s41396-020-00832-7

Brooks, B. R., Brooks, C. L., Mackerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., … Karplus, M. (2009). CHARMM: The Biomolecular Simulation Program. *Journal of Computational Chemistry*, *30*(10), 1545. https://doi.org/10.1002/JCC.21287

Brouwers, R., Vass, H., Dawson, A., Squires, T., Tavaddod, S., & Allen, R. J. (2020). Stability of β-lactam antibiotics in bacterial growth media. *PloS One*, *15*(7). https://doi.org/10.1371/JOURNAL.PONE.0236198

Burgess, D. J. (2018). Genomics: next regeneration sequencing for reference genomes. *Nat Rev Genet*, *19*(3), 125. https://doi.org/10.1038/nrg.2018.5

Bush, K., & Bradford, P. A. (2016). β-Lactams and β-Lactamase Inhibitors: An Overview. *Cold Spring Harbor Perspectives in Medicine*, *6*(8). https://doi.org/10.1101/CSHPERSPECT.A025247

CABI Agric Biosci, R., Taylor, P., & Reeder, R. (2020). Antibiotic use on crops in low and middle-income countries based on recommendations made by agricultural advisors. *CABI Agriculture and Bioscience 2020 1:1*, *1*(1), 1–14. https://doi.org/10.1186/S43170-020-00001-Y

Cabral, D. J., Penumutchu, S., Reinhart, E. M., Zhang, C., Korry, B. J., Wurster, J. I., Nilson, R., Guang, A., Sano, W. H., Rowan-Nash, A. D., Li, H., & Belenky, P. (2019). Microbial Metabolism Modulates Antibiotic Susceptibility within the Murine Gut Microbiome. *Cell Metabolism*, *30*(4), 800. https://doi.org/10.1016/J.CMET.2019.08.020

Cadena, M., Durso, L. M., Miller, D. N., Waldrip, H. M., Castleberry, B. L., Drijber, R. A., & Wortmann, C. (2018). Tetracycline and Sulfonamide Antibiotic Resistance Genes in Soils From Nebraska Organic Farming Operations. *Frontiers in Microbiology*, *9*. https://www.frontiersin.org/articles/10.3389/fmicb.2018.01283

Camacho-Muñoz, D., Petrie, B., Lopardo, L., Proctor, K., Rice, J., Youdan, J., Barden, R., & Kasprzyk-Hordern, B. (2019). Stereoisomeric profiling of chiral pharmaceutically active compounds in wastewaters and the receiving environment – A catchment-scale and a laboratory study. *Environment International*, *127*, 558–572. https://doi.org/10.1016/J.ENVINT.2019.03.050

Chen, Y., Fan, L. C., Chai, Y. H., & Xu, J. F. (2022). Advantages and challenges of metagenomic sequencing for the diagnosis of pulmonary infectious diseases. *The Clinical Respiratory Journal*, *16*(10), 646. https://doi.org/10.1111/CRJ.13538

Cheng, H. R., & Jiang, N. (2006). Extremely rapid extraction of DNA from bacteria and yeasts. *Biotechnology Letters*, *28*(1), 55–59. https://doi.org/10.1007/S10529-005-4688-Z

Chio, H., Guest, E. E., Hobman, J. L., Dottorini, T., Hirst, J. D., & Stekel, D. J. (2023). Predicting bioactivity of antibiotic metabolites by molecular docking and dynamics. *Journal of Molecular Graphics and Modelling*, *123*, 108508. https://doi.org/10.1016/J.JMGM.2023.108508

Ciccotti, G., Dellago, C., Ferrario, M., Hernández, E. R., & Tuckerman, M. E. (2022). Molecular simulations: past, present, and future (a Topical Issue in EPJB). *The European Physical Journal B 2021 95:1*, *95*(1), 1–12. https://doi.org/10.1140/EPJB/S10051-021-00249-X

Cinquin, B., Maigre, L., Pinet, E., Chevalier, J., Stavenger, R. A., Mills, S., Réfrégiers, M., & Pagès, J. M. (2015). Microspectrometric insights on the uptake of antibiotics at the single bacterial cell level. *Scientific Reports 2015 5:1*, *5*(1), 1–12. https://doi.org/10.1038/srep17968

Cycoń, M., Mrozik, A., & Piotrowska-Seget, Z. (2019). Antibiotics in the soil environment— degradation and their impact on microbial activity and diversity. *Frontiers in Microbiology*, *10*(MAR), 338. https://doi.org/10.3389/FMICB.2019.00338/BIBTEX

Davies, J., Spiegelman, G. B., & Yim, G. (2006). The world of subinhibitory antibiotic concentrations. *Current Opinion in Microbiology*, 9(5), 445–453. https://doi.org/10.1016/J.MIB.2006.08.006

Deegan, A.-M., Cullen, M., Oelgemöller, M., Nolan, K., Tobin, J., & Morrissey, A. (2011). A SPE-LC-MS/MS Method for the Detection of Low Concentrations of Pharmaceuticals in Industrial Waste Streams. *Analytical Letters*, 44(17), 2808–2820. https://doi.org/10.1080/00032719.2011.565444

Delcher, A. L., Phillippy, A., Carlton, J., & Salzberg, S. L. (2002). Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Research*, 30(11), 2478. https://doi.org/10.1093/NAR/30.11.2478

Deshpande, A., Baheti, K., & Chatterjee, N. R. (2004). Degradation of β-lactam antibiotics. *Current Science*.

Dill, K. A., Truskett, T. M., Vlachy, V., & Hribar-Lee, B. (2005). Modeling water, the hydrophobic effect, and ion solvation. *Annual Review of Biophysics and Biomolecular Structure*, 34, 173–199. https://doi.org/10.1146/ANNUREV.BIOPHYS.34.040204.144517

Dinner, A. (1977). Cephalosporin degradations. *Journal of Medicinal Chemistry*, 20(7), 963–965. https://doi.org/10.1021/JM00217A022

Dörr, T., Moynihan, P. J., & Mayer, C. (2019). Editorial: Bacterial Cell Wall Structure and Dynamics. *Frontiers in Microbiology*, 10, 2051. https://doi.org/10.3389/FMICB.2019.02051/BIBTEX

Duan, L., Guo, X., Cong, Y., Feng, G., Li, Y., & Zhang, J. Z. H. (2019). Accelerated Molecular Dynamics Simulation for Helical Proteins Folding in Explicit Water. *Frontiers in Chemistry*, 7, 540. https://doi.org/10.3389/FCHEM.2019.00540

Duhamel, P., & Vetterli, M. (1990). Fast fourier transforms: A tutorial review and a state of the art. *Signal Processing*, 19(4), 259–299. https://doi.org/10.1016/0165-1684(90)90158-U

Durrant, J. D., & McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. *BMC Biology*, 9(1), 1–9. https://doi.org/10.1186/1741-7007-9-71

Egorov, A. M., Ulyashova, M. M., & Rubtsova, M. Y. (2018). Bacterial Enzymes and Antibiotic Resistance. *Acta Naturae*, 10(4), 33. https://doi.org/10.32607/20758251-2018-10-4-33-48

El Meouche, I., & Dunlop, M. J. (2018). Heterogeneity in efflux pump expression predisposes antibiotic-resistant cells to mutation. *Science (New York, N.Y.)*, 362(6415), 686–690. https://doi.org/10.1126/SCIENCE.AAR7981

Evans, D. J., & Morriss, G. P. (1983). The isothermal/isobaric molecular dynamics ensemble. *Physics Letters A*, 98(8–9), 433–436. https://doi.org/10.1016/0375-9601(83)90256-6

Farago, O. (2019). Langevin thermostat for robust configurational and kinetic sampling. *Physica A: Statistical Mechanics and Its Applications*, 534, 122210. https://doi.org/10.1016/J.PHYSA.2019.122210

Feinstein, W. P., & Brylinski, M. (2015). Calculating an optimal box size for ligand docking and virtual screening against experimental and predicted binding pockets. *Journal of Cheminformatics*, 7(1), 1–10. https://doi.org/10.1186/S13321-015-0067-5

Finney, J. L. (2001). The water molecule and its interactions: The interaction between theory, modelling, and experiment. *Journal of Molecular Liquids*, 90(1–3), 303–312. https://doi.org/10.1016/S0167-7322(01)00134-9

Finney, J. L., Rand, R. P., Franks, F., Pettersson, L., Bowron, D. T., Engberts, J. B. F. N., Ball, P.,

& Zaccai, G. (2004). Water? What's so special about it? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *359*(1448), 1145–1165. https://doi.org/10.1098/RSTB.2004.1495

Fischer, J., & Robin Ganellin, C. (2006). Analogue-based Drug Discovery. *Analogue-Based Drug Discovery*, 1–575. https://doi.org/10.1002/3527608001

Foster, P. L. (2007). Stress-Induced Mutagenesis in Bacteria. *Critical Reviews in Biochemistry and Molecular Biology*, *42*(5), 373. https://doi.org/10.1080/10409230701648494

Fuentes-Azcatl, R., & Alejandre, J. (2014). Non-polarizable force field of water based on the dielectric constant: TIP4P/ε. *Journal of Physical Chemistry B*, *118*(5), 1263–1272. https://doi.org/10.1021/JP410865Y

Fujii, A., Seki, M., Higashiguchi, M., Tachibana, I., Kumanogoh, A., & Tomono, K. (2014). Community-acquired, hospital-acquired, and healthcare-associated pneumonia caused by Pseudomonas aeruginosa. *Respiratory Medicine Case Reports*, *12*, 30. https://doi.org/10.1016/J.RMCR.2014.03.002

Gil Pineda, L. I., Milko, L. N., & He, Y. (2020). Performance of CHARMM36m with modified water model in simulating intrinsically disordered proteins: a case study. *Biophysics Reports 2020 6:2*, *6*(2), 80–87. https://doi.org/10.1007/S41048-020-00107-W

Goh, E. B., Yim, G., Tsui, W., McClure, J. A., Surette, M. G., & Davies, J. (2002). Transcriptional modulation of bacterial gene expression by subinhibitory concentrations of antibiotics. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(26), 17025. https://doi.org/10.1073/PNAS.252607699

González, M. A. (2011). Force fields and molecular dynamics simulations. *École Thématique de La Société Française de La Neutronique*, *12*, 169–200. https://doi.org/10.1051/SFN/201112009

Groom, C. R., Bruno, I. J., Lightfoot, M. P., & Ward, S. C. (2016). The Cambridge Structural Database. *Acta Crystallographica Section B, Structural Science, Crystal Engineering and Materials*, *72*(Pt 2), 171–179. https://doi.org/10.1107/S2052520616003954

Grossman, T. H. (2016). Tetracycline Antibiotics and Resistance. *Cold Spring Harbor Perspectives in Medicine*, *6*(4). https://doi.org/10.1101/CSHPERSPECT.A025387

Gullberg, E., Cao, S., Berg, O. G., Ilbäck, C., Sandegren, L., Hughes, D., & Andersson, D. I. (2011). Selection of Resistant Bacteria at Very Low Antibiotic Concentrations. *PLOS Pathogens*, *7*(7), e1002158. https://doi.org/10.1371/JOURNAL.PPAT.1002158

Guterres, H., Park, S. J., Cao, Y., & Im, W. (2021). CHARMM-GUI Ligand Designer for Template-Based Virtual Ligand Design in a Binding Site. *Journal of Chemical Information and Modeling*, *61*(11), 5336–5342. https://doi.org/10.1021/ACS.JCIM.1C01156

Haenni, M., Majcherczyk, P. A., Barblan, J. L., & Moreillon, P. (2006). Mutational Analysis of Class A and Class B Penicillin-Binding Proteins in Streptococcus gordonii. *Antimicrobial Agents and Chemotherapy*, *50*(12), 4062. https://doi.org/10.1128/AAC.00677-06

Hairer, E., Lubich, C., & Wanner, G. (2003). *Geometric numerical integration illustrated by the Störmer-Verlet method*. 399–450. https://doi.org/10.1017/S0962492902000144

Hanna, N., Sun, P., Sun, Q., Li, X., Yang, X., Ji, X., Zou, H., Ottoson, J., Nilsson, L. E., Berglund, B., Dyar, O. J., Tamhankar, A. J., & Stålsby Lundborg, C. (2018). Presence of

antibiotic residues in various environmental compartments of Shandong province in eastern China: Its potential for resistance development and ecological and human risk. *Environment International*, *114*, 131–142. https://doi.org/10.1016/J.ENVINT.2018.02.003

Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeerschd, T., Zurek, E., & Hutchison, G. R. (2012). Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics*, *4*(8), 1–17. https://doi.org/10.1186/1758-2946-4-17

Hao, H., Cheng, G., Iqbal, Z., Ai, X., Hussain, H. I., Huang, L., Dai, M., Wang, Y., Liu, Z., & Yuan, Z. (2014). Benefits and risks of antimicrobial use in food-producing animals. *Frontiers in Microbiology*, *5*(JUN). https://doi.org/10.3389/FMICB.2014.00288

Hernandez, P., Müller, M., & Appel, R. D. (2006). Automated protein identification by tandem mass spectrometry: Issues and strategies. *Mass Spectrometry Reviews*, *25*(2), 235–254. https://doi.org/https://doi.org/10.1002/mas.20068

Ho, H. P., Lee, R. J., Chen, C. Y., Wang, S. R., Li, Z. G., & Lee, M. R. (2011). Identification of new minor metabolites of penicillin G in human serum by multiple-stage tandem mass spectrometry. *Rapid Communications in Mass Spectrometry : RCM*, *25*(1), 25–32. https://doi.org/10.1002/RCM.4823

Hoffmann, M., DeMaio, W., Jordan, R. A., Talaat, R., Harper, D., Speth, J., & Scatina, J. (2007). Metabolism, Excretion, and Pharmacokinetics of [&lt;sup&gt;14&lt;/sup&gt;C]Tigecycline, a First-In-Class Glycylcycline Antibiotic, after Intravenous Infusion to Healthy Male Subjects. *Drug Metabolism and Disposition*, *35*(9), 1543. https://doi.org/10.1124/dmd.107.015735

Hollingsworth, S. A., & Dror, R. O. (2018). Review Molecular Dynamics Simulation for All. *Neuron*, *99*, 1129–1143. https://doi.org/10.1016/j.neuron.2018.08.011

Homem, V., & Santos, L. (2011). Degradation and removal methods of antibiotics from aqueous matrices – A review. *Journal of Environmental Management*, *92*(10), 2304–2347. https://doi.org/10.1016/J.JENVMAN.2011.05.023

Hoover, W. G. (1986). Constant-pressure equations of motion. *Physical Review A*, *34*(3), 2499–2500. https://doi.org/10.1103/PhysRevA.34.2499

Horcajo, P., De Pedro, M. A., & Cava, F. (2012). Peptidoglycan plasticity in bacteria: Stress-induced peptidoglycan editing by noncanonical D-amino acids. *Microbial Drug Resistance*, *18*(3), 306–313. https://doi.org/10.1089/mdr.2012.0009

Huang, L., Wu, C., Gao, H., Xu, C., Dai, M., Huang, L., Hao, H., Wang, X., & Cheng, G. (2022). Bacterial Multidrug Efflux Pumps at the Frontline of Antimicrobial Resistance: An Overview. *Antibiotics*, *11*(4). https://doi.org/10.3390/ANTIBIOTICS11040520

Huang, N., Shoichet, B. K., & Irwin, J. J. (2006). Benchmarking sets for molecular docking. *Journal of Medicinal Chemistry*, *49*(23), 6789–6801. https://doi.org/10.1021/JM0608356

Ibrahim, D. R., Dodd, C. E. R., Stekel, D. J., Meshioye, R. T., Diggle, M., Lister, M., & Hobman, J. L. (2023). Multidrug-Resistant ESBL-Producing E. coli in Clinical Samples from the UK. *Antibiotics 2023, Vol. 12, Page 169*, *12*(1), 169. https://doi.org/10.3390/ANTIBIOTICS12010169

Idsoe, O., Guthe, T., Willcox, R. R., & De Weck, A. L. (1968). Nature and Extent of Penicillin Side-reactions, with Particular Reference to Fatalities from Anaphylactic Shock. *Bull. Org. Mond. Sante*, *38*, 159–188.

Izadi, S., Anandakrishnan, R., & Onufriev, A. V. (2014). Building water models: A different approach. *Journal of Physical Chemistry Letters*, *5*(21), 3863–3871. https://doi.org/10.1021/JZ501780A

J, D., & D, D. (2010). Origins and evolution of antibiotic resistance. *Microbiology and Molecular Biology Reviews : MMBR*, *74*(3), 9–16. https://doi.org/10.1128/MMBR.00016-10

Jakobsen, A. F. (2005). Constant-pressure and constant-surface tension simulations in dissipative particle dynamics. *Journal of Chemical Physics*, *122*(12). https://doi.org/10.1063/1.1867374/950725

Jones, E. R., Van Vliet, M. T. H., Qadir, M., & Bierkens, M. F. P. (2021). Country-level and gridded estimates of wastewater production, collection, treatment and reuse. *Earth System Science Data*, *13*(2), 237–254. https://doi.org/10.5194/ESSD-13-237-2021

Jones, G., Willett, P., & Glen, R. C. (1995). Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *Journal of Molecular Biology*, *245*(1), 43–53. https://doi.org/10.1016/S0022-2836(95)80037-9

Jones, G., Willett, P., Glen, R. C., Leach, A. R., & Taylor, R. (1997). Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology*, *267*(3), 727–748. https://doi.org/10.1006/JMBI.1996.0897

Jünemann, S., Kleinbölting, N., Jaenicke, S., Henke, C., Hassa, J., Nelkner, J., Stolze, Y., Albaum, S. P., Schlüter, A., Goesmann, A., Sczyrba, A., & Stoye, J. (2017). Bioinformatics for NGS-based metagenomics and the application to biogas research. *Journal of Biotechnology*, *261*, 10–23. https://doi.org/10.1016/J.JBIOTEC.2017.08.012

Justice, S. S., Hunstad, D. A., Cegelski, L., & Hultgren, S. J. (2008). Morphological plasticity as a bacterial survival strategy. *Nature Reviews. Microbiology*, *6*(2), 162–168. https://doi.org/10.1038/NRMICRO1820

Kadupitiya, J. C. S., Fox, G. C., & Jadhao, V. (2022). Solving Newton's equations of motion with large timesteps using recurrent neural networks based operators. *Machine Learning: Science and Technology*, *3*(2), 025002. https://doi.org/10.1088/2632-2153/AC5F60

Kairigo, P., Ngumba, E., Sundberg, L. R., Gachanja, A., & Tuhkanen, T. (2020). Occurrence of antibiotics and risk of antibiotic resistance evolution in selected Kenyan wastewaters, surface waters and sediments. *The Science of the Total Environment*, *720*. https://doi.org/10.1016/J.SCITOTENV.2020.137580

Kalokhe, A. S., Shafiq, M., Lee, J. C., Ray, S. M., Wang, Y. F., Metchock, B., Anderson, A. M., & Ly T. Nguyen, M. (2013). Multidrug-resistant tuberculosis drug susceptibility and molecular diagnostic testing: a review of the literature. *The American Journal of the Medical Sciences*, *345*(2), 143. https://doi.org/10.1097/MAJ.0B013E31825D32C6

Karkman, A., Pärnänen, K., & Larsson, D. G. J. (2019). Fecal pollution can explain antibiotic resistance gene abundances in anthropogenically impacted environments. *Nature Communications 2019 10:1*, *10*(1), 1–8. https://doi.org/10.1038/s41467-018-07992-3

Karplus, M., & McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules. *Nature Structural Biology*, *9*(9), 646–652. https://doi.org/10.1038/NSB0902-646

Keenum, I., Liguori, K., Calarco, J., Davis, B. C., Milligan, E., Harwood, V. J., & Pruden, A. (2022). A framework for standardized qPCR-targets and protocols for quantifying antibiotic resistance in surface water, recycled water and wastewater. *Critical Reviews*

*in Environmental Science and Technology.*
https://doi.org/10.1080/10643389.2021.2024739

Klein, E. Y., Van Boeckel, T. P., Martinez, E. M., Pant, S., Gandra, S., Levin, S. A., Goossens, H., & Laxminarayan, R. (2018). Global increase and geographic convergence in antibiotic consumption between 2000 and 2015. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(15), E3463–E3470. https://doi.org/10.1073/PNAS.1717295115

Kocaoglu, O., & Carlson, E. E. (2015). Profiling of β-Lactam Selectivity for Penicillin-Binding Proteins in Escherichia coli Strain DC2. *Antimicrobial Agents and Chemotherapy*, *59*(5), 2785. https://doi.org/10.1128/AAC.04552-14

Kraemer, S. A., Ramachandran, A., & Perron, G. G. (2019). Antibiotic Pollution in the Environment: From Microbial Ecology to Public Policy. *Microorganisms*, *7*(6). https://doi.org/10.3390/MICROORGANISMS7060180

Kralik, P., & Ricchi, M. (2017). A Basic Guide to Real Time PCR in Microbial Diagnostics: Definitions, Parameters, and Everything. *Frontiers in Microbiology*, *8*. https://www.frontiersin.org/articles/10.3389/fmicb.2017.00108

Kristiansson, E., Fick, J., Janzon, A., Grabic, R., Rutgersson, C., Weijdegård, B., Söderström, H., & Joakim Larsson, D. G. (2011). Pyrosequencing of Antibiotic-Contaminated River Sediments Reveals High Levels of Resistance and Gene Transfer Elements. *PLOS ONE*, *6*(2), e17038. https://doi.org/10.1371/JOURNAL.PONE.0017038

Kümmerer, K. (2009). Antibiotics in the aquatic environment – A review – Part I. *Chemosphere*, *75*(4), 417–434. https://doi.org/https://doi.org/10.1016/j.chemosphere.2008.11.086

Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R., & Ferrin, T. E. (1982). A geometric approach to macromolecule-ligand interactions. *Journal of Molecular Biology*, *161*(2), 269–288. https://doi.org/10.1016/0022-2836(82)90153-X

Larsson, D. G. J., & Flach, C. F. (2021). Antibiotic resistance in the environment. *Nature Reviews Microbiology 2021 20:5*, *20*(5), 257–269. https://doi.org/10.1038/s41579-021-00649-x

Lee, J., Cheng, X., Swails, J. M., Yeom, M. S., Eastman, P. K., Lemkul, J. A., Wei, S., Buckner, J., Jeong, J. C., Qi, Y., Jo, S., Pande, V. S., Case, D. A., Brooks, C. L., MacKerell, A. D., Klauda, J. B., & Im, W. (2016). CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *Journal of Chemical Theory and Computation*, *12*(1), 405–413. https://doi.org/10.1021/ACS.JCTC.5B00935

Lee, L., Savage, V. M., & Yeh, P. J. (2018). Intermediate Levels of Antibiotics May Increase Diversity of Colony Size Phenotype in Bacteria. *Computational and Structural Biotechnology Journal*, *16*, 307–315. https://doi.org/10.1016/J.CSBJ.2018.08.004

Lee, M., Hesek, D., Suvorov, M., Lee, W., Vakulenko, S., & Mobashery, S. (2003). A mechanism-based inhibitor targeting the DD-transpeptidase activity of bacterial penicillin-binding proteins. *Journal of the American Chemical Society*, *125*(52), 16322–16326. https://doi.org/10.1021/JA038445L

Levy, S. B., & Fitzgerald, G. B. (1978). Emergence of Antibiotic-Resistant Bacteria in the Intestinal Flora of Farm Inhabitants. *The Journal of Infectious Diseases*, *137*(5), 688–690. https://doi.org/10.1093/INFDIS/137.5.688

Li, C. C. (1967). Fundamental Theorem of Natural Selection. *Nature 1967 214:5087, 214*(5087), 505–506. https://doi.org/10.1038/214505a0

Liu, Y., Yang, K., Zhang, H., Jia, Y., & Wang, Z. (2020). Combating Antibiotic Tolerance Through Activating Bacterial Metabolism. *Frontiers in Microbiology, 11*, 577564. https://doi.org/10.3389/FMICB.2020.577564

Llor, C., & Bjerrum, L. (2014). Antimicrobial resistance: risk associated with antibiotic overuse and initiatives to reduce the problem. *Therapeutic Advances in Drug Safety, 5*(6), 229. https://doi.org/10.1177/2042098614554919

Maaten, L. van der, & Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research, 9*(86), 2579–2605. http://jmlr.org/papers/v9/vandermaaten08a.html

Macheboeuf, P., Contreras-Martel, C., Job, V., Dideberg, O., & Dessen, A. (2006). Penicillin binding proteins: key players in bacterial cell cycle and drug resistance processes. *FEMS Microbiology Reviews, 30*(5), 673–691. https://doi.org/10.1111/J.1574-6976.2006.00024.X

MacKerell Jr, A. (2001). Atomistic Models and Force Fields. *Computational Biochemistry and Biophysics*. https://doi.org/10.1201/9780203903827.CH2

Magiorakos, A. P., Srinivasan, A., Carey, R. B., Carmeli, Y., Falagas, M. E., Giske, C. G., Harbarth, S., Hindler, J. F., Kahlmeter, G., Olsson-Liljequist, B., Paterson, D. L., Rice, L. B., Stelling, J., Struelens, M. J., Vatopoulos, A., Weber, J. T., & Monnet, D. L. (2012). Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. *Clinical Microbiology and Infection : The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases, 18*(3), 268–281. https://doi.org/10.1111/J.1469-0691.2011.03570.X

Mahmood, A. R., Al-Haideri, H. H., & Hassan, F. M. (2019). Detection of Antibiotics in Drinking Water Treatment Plants in Baghdad City, Iraq. *Advances in Public Health, 2019*. https://doi.org/10.1155/2019/7851354

Maiques, E., Úbeda, C., Campoy, S., Salvador, N., Lasa, Í., Novick, R. P., Barbé, J., & Penadés, J. R. (2006). β-Lactam Antibiotics Induce the SOS Response and Horizontal Transfer of Virulence Factors in Staphylococcus aureus. *Journal of Bacteriology, 188*(7), 2726. https://doi.org/10.1128/JB.188.7.2726-2729.2006

Majeed, H. J., Riquelme, M. V., Davis, B. C., Gupta, S., Angeles, L., Aga, D. S., Garner, E., Pruden, A., & Vikesland, P. J. (2021). Evaluation of Metagenomic-Enabled Antibiotic Resistance Surveillance at a Conventional Wastewater Treatment Plant. *Frontiers in Microbiology, 12*. https://doi.org/10.3389/FMICB.2021.657954

Mallah, N., Orsini, N., Figueiras, A., & Takkouche, B. (2022). Income level and antibiotic misuse: a systematic review and dose–response meta-analysis. *The European Journal of Health Economics, 23*(6), 1015. https://doi.org/10.1007/S10198-021-01416-8

Mann, A., Nehra, K., Rana, J. S., & Dahiya, T. (2021). Antibiotic resistance in agriculture: Perspectives on upcoming strategies to overcome upsurge in resistance. *Current Research in Microbial Sciences, 2*, 100030. https://doi.org/10.1016/J.CRMICR.2021.100030

Mărgărita Ghimpet, O., Narcisa Pogurschi, E., Cătălina Popa, D., Dragomir, N., Diana Mihai, O., & Daniela Petcu, C. (2022). *Antibiotic Use in Livestock and Residues in Food-A*

*Public Health Threat: A Review*. https://doi.org/10.3390/foods11101430

Mark, P., & Nilsson, L. (2001). Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *Journal of Physical Chemistry A*, *105*(43), 9954–9960. https://doi.org/10.1021/JP003020W

Marschall, T., Marz, M., Abeel, T., Dijkstra, L., Dutilh, B. E., Ghaffaari, A., Kersey, P., Kloosterman, W. P., Mäkinen, V., Novak, A. M., Paten, B., Porubsky, D., Rivals, E., Alkan, C., Baaijens, J. A., De Bakker, P. I. W., Boeva, V., Bonnal, R. J. P., Chiaromonte, F., … Schönhuth, A. (2018). Computational pan-genomics: status, promises and challenges. *Briefings in Bioinformatics*, *19*(1), 118–135. https://doi.org/10.1093/BIB/BBW089

Martínez, J. L. (2017). Effect of antibiotics on bacterial populations: A multi-hierachical selection process. *F1000Research*, *6*. https://doi.org/10.12688/F1000RESEARCH.9685.1/DOI

Massé, D. I., Saady, N. M. C., & Gilbert, Y. (2014). Potential of Biological Processes to Eliminate Antibiotics in Livestock Manure: An Overview. *Animals : An Open Access Journal from MDPI*, *4*(2), 146. https://doi.org/10.3390/ANI4020146

Mathers, A. J. (2015). Antibiotics in Laboratory Medicine, 6th Edition. *Clinical Infectious Diseases*, *60*(9), 1446–1447. https://doi.org/10.1093/CID/CIV078

McLain, J. E., Cytryn, E., Durso, L. M., & Young, S. (2016). Culture-based Methods for Detection of Antibiotic Resistance in Agroecosystems: Advantages, Challenges, and Gaps in Knowledge. *Journal of Environmental Quality*, *45*(2), 432–440. https://doi.org/10.2134/JEQ2015.06.0317

McVey, C. E., Walsh, M. A., Dodson, G. G., Wilson, K. S., & Brannigan, J. A. (2001). Crystal structures of penicillin acylase enzyme-substrate complexes: structural insights into the catalytic mechanism. *Journal of Molecular Biology*, *313*(1), 139–150. https://doi.org/10.1006/JMBI.2001.5043

Mead, A. (1992). Review of the Development of Multidimensional Scaling Methods. *The Statistician*, *41*(1), 27. https://doi.org/10.2307/2348634

Melander, M. M. (2021). Grand canonical ensemble approach to electrochemical thermodynamics, kinetics, and model Hamiltonians. *Current Opinion in Electrochemistry*, *29*, 100749. https://doi.org/10.1016/J.COELEC.2021.100749

Mendonça, L. M. F., & Marana, S. R. (2011). Single mutations outside the active site affect the substrate specificity in a β-glycosidase. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1814*(12), 1616–1623. https://doi.org/10.1016/J.BBAPAP.2011.08.012

Montforts, M. (1997). *Environmental risk assessment for veterinary medicinal products. Part 1. Other than GMO-containing and immunological products. First update*.

Moon, K. R., van Dijk, D., Wang, Z., Gigante, S., Burkhardt, D. B., Chen, W. S., Yim, K., Elzen, A. van den, Hirn, M. J., Coifman, R. R., Ivanova, N. B., Wolf, G., & Krishnaswamy, S. (2019). Visualizing structure and transitions in high-dimensional biological data. *Nature Biotechnology 2019 37:12*, *37*(12), 1482–1492. https://doi.org/10.1038/s41587-019-0336-3

Mora-Ochomogo, M., & Lohans, C. T. (2021). β-Lactam antibiotic targets and resistance mechanisms: from covalent inhibitors to substrates. *RSC Medicinal Chemistry*, *12*(10), 1623–1639. https://doi.org/10.1039/D1MD00200G

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., & Erlich, H. (1986). Specific enzymatic

amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor Symposia on Quantitative Biology*, *51 Pt 1*(1), 263–273. https://doi.org/10.1101/SQB.1986.051.01.032

Munteanu, F. D., Titoiu, A. M., Marty, J. L., & Vasilescu, A. (2018). Detection of Antibiotics and Evaluation of Antibacterial Activity with Screen-Printed Electrodes. *Sensors (Basel, Switzerland)*, *18*(3). https://doi.org/10.3390/S18030901

Murray, C. J., Ikuta, K. S., Sharara, F., Swetschinski, L., Robles Aguilar, G., Gray, A., Han, C., Bisignano, C., Rao, P., Wool, E., Johnson, S. C., Browne, A. J., Chipeta, M. G., Fell, F., Hackett, S., Haines-Woodhouse, G., Kashef Hamadani, B. H., Kumaran, E. A. P., McManigal, B., … Naghavi, M. (2022). Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet*, *399*(10325), 629–655. https://doi.org/10.1016/S0140-6736(21)02724-0

Muschiol, S., Balaban, M., Normark, S., & Henriques-Normark, B. (2015). Uptake of extracellular DNA: Competence induced pili in natural transformation of Streptococcus pneumoniae. *Bioessays*, *37*(4), 426. https://doi.org/10.1002/BIES.201400125

Navgire, G. S., Goel, N., Sawhney, G., Sharma, M., Kaushik, P., Mohanta, Y. K., Mohanta, T. K., & Al-Harrasi, A. (2022). Analysis and Interpretation of metagenomics data: an approach. *Biological Procedures Online 2022 24:1*, *24*(1), 1–22. https://doi.org/10.1186/S12575-022-00179-7

Nichols, D. A., Jaishankar, P., Larson, W., Smith, E., Liu, G., Beyrouthy, R., Bonnet, R., Renslo, A. R., & Chen, Y. (2012). Structure-based design of potent and ligand-efficient inhibitors of CTX-M class A β-lactamase. *Journal of Medicinal Chemistry*, *55*(5), 2163–2172. https://doi.org/10.1021/JM2014138

Orellana, L. (2019). Large-Scale Conformational Changes and Protein Function: Breaking the in silico Barrier. *Frontiers in Molecular Biosciences*, *6*, 117. https://doi.org/10.3389/FMOLB.2019.00117

Ostash, B., Makitrynskyy, R., Yushchuk, O., & Fedorenko, V. (2022). Structural diversity, bioactivity, and biosynthesis of phosphoglycolipid family antibiotics: Recent advances. *BBA Advances*, *2*, 100065. https://doi.org/https://doi.org/10.1016/j.bbadva.2022.100065

Pantsar, T., & Poso, A. (2018). Binding Affinity via Docking: Fact and Fiction. *Molecules : A Journal of Synthetic Chemistry and Natural Product Chemistry*, *23*(8), 1DUMMY. https://doi.org/10.3390/MOLECULES23081899

Paquet, E., & Viktor, H. L. (2015). Molecular dynamics, monte carlo simulations, and langevin dynamics: A computational review. *BioMed Research International*, *2015*. https://doi.org/10.1155/2015/183918

Pearman, W. S., Freed, N. E., & Silander, O. K. (2020). Testing the advantages and disadvantages of short- And long- read eukaryotic metagenomics using simulated reads. *BMC Bioinformatics*, *21*(1), 1–15. https://doi.org/10.1186/S12859-020-3528-4

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. (2011). Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.*, *12*(null), 2825–2830.

Phillips, J. C., Hardy, D. J., Maia, J. D. C., Stone, J. E., Ribeiro, J. V., Bernardi, R. C., Buch, R., Fiorin, G., Hénin, J., Jiang, W., McGreevy, R., Melo, M. C. R., Radak, B. K., Skeel, R. D.,

Singharoy, A., Wang, Y., Roux, B., Aksimentiev, A., Luthey-Schulten, Z., … Tajkhorshid, E. (2020). Scalable molecular dynamics on CPU and GPU architectures with NAMD. *Journal of Chemical Physics*, *153*(4), 44130. https://doi.org/10.1063/5.0014475

Pitt, J. J. (2009). Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry. *The Clinical Biochemist. Reviews*, *30*(1), 19–34. https://www.unboundmedicine.com/medline/citation/19224008/Principles_and_appli cations_of_liquid_chromatography_mass_spectrometry_in_clinical_biochemistry_

Pokharel, S., Raut, S., & Adhikari, B. (2019). Tackling antimicrobial resistance in low-income and middle-income countries. *BMJ Global Health*, *4*(6), e002104. https://doi.org/10.1136/BMJGH-2019-002104

Pollard, M. O., Gurdasani, D., Mentzer, A. J., Porter, T., & Sandhu, M. S. (2018). Long reads: their purpose and place. *Hum Mol Genet*, *27*(R2), 234–241. https://doi.org/10.1093/hmg/ddy177

Ponder, J. W., & Case, D. A. (2003). Force fields for protein simulations. *Advances in Protein Chemistry*, *66*, 27–85. https://doi.org/10.1016/S0065-3233(03)66002-X

Ponomareva, O., Ponomarev, A., & Ponomarev, V. (2018). Evolution of Forward and Inverse Discrete Fourier Transform. *Proceedings of 2018 IEEE East-West Design and Test Symposium, EWDTS 2018*. https://doi.org/10.1109/EWDTS.2018.8524820

Pouwels, K. B., Freeman, R., Muller-Pebody, B., Rooney, G., Henderson, K. L., Robotham, J. v, & Smieszek, T. (2018). Association between use of different antibiotics and trimethoprim resistance: going beyond the obvious crude association. *Journal of Antimicrobial Chemotherapy*, *73*(6), 1700–1707. https://doi.org/10.1093/jac/dky031

Prestinaci, F., Pezzotti, P., & Pantosti, A. (2015). Antimicrobial resistance: a global multifaceted phenomenon. *Pathogens and Global Health*, *109*(7), 309. https://doi.org/10.1179/2047773215Y.0000000030

Radó, J., Kaszab, E., Petrovics, T., Pászti, J., Kriszt, B., & Szoboszlay, S. (2017). Characterization of environmental Pseudomonas aeruginosa using multilocus sequence typing scheme. *Journal of Medical Microbiology*, *66*(10), 1457–1466. https://doi.org/10.1099/JMM.0.000589

Rahim, R., Ahmar, A. S., & Hidayat, R. (2022). Cross-Validation and Validation Set Methods for Choosing K in KNN Algorithm for Healthcare Case Study. *JINAV: Journal of Information and Visualization*, *3*(1), 57–61. https://doi.org/10.35877/454RI.JINAV1557

Rahman, S., Kesselheim, A. S., & Hollis, A. (2023). Persistence of resistance: a panel data analysis of the effect of antibiotic usage on the prevalence of resistance. *The Journal of Antibiotics 2023 76:5*, *76*(5), 270–278. https://doi.org/10.1038/s41429-023-00601-6

Ramirez, J., Guarner, F., Bustos Fernandez, L., Maruy, A., Sdepanian, V. L., & Cohen, H. (2020). Antibiotics as Major Disruptors of Gut Microbiota. *Frontiers in Cellular and Infection Microbiology*, *10*. https://www.frontiersin.org/articles/10.3389/fcimb.2020.572912

Rapaport, D. C. (2022). GPU molecular dynamics: Algorithms and performance. *Journal of Physics: Conference Series*, *2241*, 12007. https://doi.org/10.1088/1742-6596/2241/1/012007

Reddy, B., & Dubey, S. K. (2019). River Ganges water as reservoir of microbes with antibiotic and metal ion resistance genes: High throughput metagenomic approach. *Environmental Pollution*, *246*, 443–451.

https://doi.org/https://doi.org/10.1016/j.envpol.2018.12.022

Rémy, B., Mion, S., Plener, L., Elias, M., Chabrière, E., & Daudé, D. (2018). Interference in Bacterial Quorum Sensing: A Biopharmaceutical Perspective. *Frontiers in Pharmacology*, *9*(MAR), 203. https://doi.org/10.3389/FPHAR.2018.00203

Ren, J., Nettleship, J. E., Males, A., Stuart, D. I., & Owens, R. J. (2016). Crystal structures of penicillin-binding protein 3 in complexes with azlocillin and cefoperazone in both acylated and deacylated forms. *FEBS Letters*, *590*(2), 288–297. https://doi.org/10.1002/1873-3468.12054

Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N. N., Anderson, I. J., Cheng, J. F., Darling, A., Malfatti, S., Swan, B. K., Gies, E. A., Dodsworth, J. A., Hedlund, B. P., Tsiamis, G., Sievert, S. M., Liu, W. T., Eisen, J. A., Hallam, S. J., Kyrpides, N. C., Stepanauskas, R., … Woyke, T. (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature 2013 499:7459*, *499*(7459), 431–437. https://doi.org/10.1038/nature12352

Ropy, A., Cabot, G., Sánchez-Diener, I., Aguilera, C., Moya, B., Ayala, J. A., & Oliver, A. (2015a). Role of Pseudomonas aeruginosa low-molecular-mass penicillin-binding proteins in AmpC expression, β-lactam resistance, and peptidoglycan structure. *Antimicrobial Agents and Chemotherapy*, *59*(7), 3925–3934. https://doi.org/10.1128/AAC.05150-14

Ropy, A., Cabot, G., Sánchez-Diener, I., Aguilera, C., Moya, B., Ayala, J. A., & Oliver, A. (2015b). Role of Pseudomonas aeruginosa low-molecular-mass penicillin-binding proteins in AmpC expression, β-lactam resistance, and peptidoglycan structure. *Antimicrobial Agents and Chemotherapy*, *59*(7), 3925–3934. https://doi.org/10.1128/AAC.05150-14

Rousham, E., Unicomb, L., Wood, P., Smith, M., Asaduzzaman, M., & Islam, M. A. (2018). Spatial and temporal variation in the community prevalence of antibiotic resistance in Bangladesh: an integrated surveillance study protocol. *BMJ Open*, *8*(4), e023158. https://doi.org/10.1136/BMJOPEN-2018-023158

Sainsbury, S., Bird, L., Rao, V., Shepherd, S. M., Stuart, D. I., Hunter, W. N., Owens, R. J., & Ren, J. (2011). Crystal structures of penicillin-binding protein 3 from Pseudomonas aeruginosa: comparison of native and antibiotic-bound forms. *Journal of Molecular Biology*, *405*(1), 173–184. https://doi.org/10.1016/J.JMB.2010.10.024

Sánchez-Romero, M. A., & Casadesús, J. (2014). Contribution of phenotypic heterogeneity to adaptive antibiotic resistance. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(1), 355–360. https://doi.org/10.1073/PNAS.1316084111

Sandegren, L. (2014). Selection of antibiotic resistance at very low antibiotic concentrations. *Upsala Journal of Medical Sciences*, *119*(2), 103. https://doi.org/10.3109/03009734.2014.904457

Saravanan, R., & Raveendaran, V. (2013). Antimicrobial resistance pattern in a tertiary care hospital: An observational study. *Journal of Basic and Clinical Pharmacy*, *4*(3), 56. https://doi.org/10.4103/0976-0105.118797

Sartelli, M., Hardcastle, T. C., Catena, F., Chichom-Mefire, A., Coccolini, F., Dhingra, S., Haque, M., Hodonou, A., Iskandar, K., Labricciosa, F. M., Marmorale, C., Sall, I., & Pagani, L. (2020). Antibiotic Use in Low and Middle-Income Countries and the

Challenges of Antimicrobial Resistance in Surgery. *Antibiotics*, *9*(8), 1–12. https://doi.org/10.3390/ANTIBIOTICS9080497

Sauvage, E., Kerff, F., Terrak, M., Ayala, J. A., & Charlier, P. (2008). The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis. *FEMS Microbiology Reviews*, *32*(2), 234–258. https://doi.org/10.1111/J.1574-6976.2008.00105.X

Scholz, M., Gatzek, S., Sterling, A., Fiehn, O., & Selbig, J. (2004). Metabolite fingerprinting: detecting biological features by independent component analysis. *Bioinformatics*, *20*(15), 2447–2454. https://doi.org/10.1093/bioinformatics/bth270

Schulz, F., Eloe-Fadrosh, E. A., Bowers, R. M., Jarett, J., Nielsen, T., Ivanova, N. N., Kyrpides, N. C., & Woyke, T. (2017). Towards a balanced view of the bacterial tree of life. *Microbiome*, *5*(1), 140. https://doi.org/10.1186/S40168-017-0360-9

Schwartz, M. A. (1965). Mechanism of degradation of penicillin G in acidic solution. *Journal of Pharmaceutical Sciences*, *54*(3), 472–473. https://doi.org/10.1002/JPS.2600540336

Sedighpour, D., & Taghizadeh, H. (2022). The effects of mutation on the drug binding affinity of Neuraminidase: case study of Capsaicin using steered molecular dynamics simulation. *Journal of Molecular Modeling*, *28*(2). https://doi.org/10.1007/S00894-021-05005-7

Slatko, B. E., Gardner, A. F., & Ausubel, F. M. (2018). Overview of Next Generation Sequencing Technologies. *Current Protocols in Molecular Biology*, *122*(1), e59. https://doi.org/10.1002/CPMB.59

Soto, S. M. (2013). Role of efflux pumps in the antibiotic resistance of bacteria embedded in a biofilm. *Virulence*, *4*(3), 223. https://doi.org/10.4161/VIRU.23724

Spreiter, Q., & Walter, M. (1999). Classical Molecular Dynamics Simulation with the Velocity Verlet Algorithm at Strong External Magnetic Fields. *Journal of Computational Physics*, *152*(1), 102–119. https://doi.org/10.1006/JCPH.1999.6237

Stanczak-Mrozek, K. I., Laing, K. G., & Lindsay, J. A. (2017). Resistance gene transfer: induction of transducing phage by sub-inhibitory concentrations of antimicrobials is not correlated to induction of lytic phage. *The Journal of Antimicrobial Chemotherapy*, *72*(6), 1624–1631. https://doi.org/10.1093/JAC/DKX056

Steigbigel, N. H., Reed, C. W., & Finland, M. (1968). Absorption and excretion of five tetracycline analogues in normal young men. *American Journal of Medical Sciences*, *255*, 296–312.

Şterbuleac, D. (2021). Molecular dynamics: a powerful tool for studying the medicinal chemistry of ion channel modulators. *RSC Medicinal Chemistry*, *12*(9), 1503. https://doi.org/10.1039/D1MD00140J

Stokes, J. M., Lopatkin, A. J., Lobritz, M. A., & Collins, J. J. (2019). Bacterial Metabolism and Antibiotic Efficacy. *Cell Metabolism*, *30*(2), 251–259. https://doi.org/10.1016/J.CMET.2019.06.009

Swegat, W., Schlitter, J., Krüger, P., & Wollmer, A. (2003). MD Simulation of Protein-Ligand Interaction: Formation and Dissociation of an Insulin-Phenol Complex. *Biophysical Journal*, *84*(3), 1493–1506. https://doi.org/10.1016/S0006-3495(03)74962-5

*Technical brief on water, sanitation, hygiene and wastewater management to prevent infections and reduce the spread of antimicrobial resistance*. (n.d.). Retrieved September 15, 2022, from https://apps.who.int/iris/handle/10665/332243

Tello, A., Austin, B., & Telfer, T. C. (2012). Selective Pressure of Antibiotic Pollution on

Bacteria of Importance to Public Health. *Environmental Health Perspectives*, *120*(8), 1100. https://doi.org/10.1289/EHP.1104650

Tenenbaum, J. B., De Silva, V., & Langford, J. C. (2000). A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science*, *290*(5500), 2319–2323. https://doi.org/10.1126/SCIENCE.290.5500.2319

*The use of antibiotics on healthy farm animals and antimicrobial resistance 2 Commons*. (n.d.).

Tian, M., He, X., Feng, Y., Wang, W., Chen, H., Gong, M., Liu, D., Clarke, J. L., & Van Eerde, A. (2021). Pollution by Antibiotics and Antimicrobial Resistance in LiveStock and Poultry Manure in China, and Countermeasures. *Antibiotics*, *10*(5). https://doi.org/10.3390/ANTIBIOTICS10050539

Timm, A., Borowska, E., Majewsky, M., Merel, S., Zwiener, C., Bräse, S., & Horn, H. (2019). Photolysis of four β-lactam antibiotics under simulated environmental conditions: Degradation, transformation products and antibacterial activity. *Science of The Total Environment*, *651*, 1605–1612. https://doi.org/10.1016/J.SCITOTENV.2018.09.248

Tuckerman, M. E., & Martyna, G. J. (2000). Understanding Modern Molecular Dynamics:  Techniques and Applications. *The Journal of Physical Chemistry B*, *104*(2), 159–178. https://doi.org/10.1021/jp992433y

Van Berkel, S. S., Nettleship, J. E., Leung, I. K. H., Brem, J., Choi, H., Stuart, D. I., Claridge, T. D. W., McDonough, M. A., Owens, R. J., Ren, J., & Schofield, C. J. (2013). Binding of (5S)-penicilloic acid to penicillin binding protein 3. *ACS Chemical Biology*, *8*(10), 2112–2116. https://doi.org/10.1021/CB400200H

van den Bogaard, A. E., London, N., Driessen, C., & Stobberingh, E. E. (2001). Antibiotic resistance of faecal Escherichia coli in poultry, poultry farmers and poultry slaughterers. *The Journal of Antimicrobial Chemotherapy*, *47*(6), 763–771. https://doi.org/10.1093/JAC/47.6.763

Van Den Bogaard, A. E., Willems, R., London, N., Top, J., & Stobberingh, E. E. (2002). Antibiotic resistance of faecal enterococci in poultry, poultry farmers and poultry slaughterers. *Journal of Antimicrobial Chemotherapy*, *49*(3), 497–505. https://doi.org/10.1093/JAC/49.3.497

Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I., & Mackerell, A. D. (2010). CHARMM General Force Field (CGenFF): A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry*, *31*(4), 671. https://doi.org/10.1002/JCC.21367

Vanommeslaeghe, K., & MacKerell, A. D. (2012). Automation of the CHARMM general force field (CGenFF) I: Bond perception and atom typing. *Journal of Chemical Information and Modeling*, *52*(12), 3144–3154. https://doi.org/10.1021/CI300363C

Vinué, L., Corcoran, M. A., Hooper, D. C., & Jacoby, G. A. (2016). Mutations That Enhance the Ciprofloxacin Resistance of Escherichia coli with qnrA1. *Antimicrobial Agents and Chemotherapy*, *60*(3), 1537. https://doi.org/10.1128/AAC.02167-15

Vogelstein, J. T., Bridgeford, E. W., Tang, M., Zheng, D., Douville, C., Burns, R., & Maggioni, M. (2021). Supervised dimensionality reduction for big data. *Nature Communications 2021 12:1*, *12*(1), 1–9. https://doi.org/10.1038/s41467-021-23102-2

Voorter, C. E. M., Palusci, F., & Tilanus, M. G. J. (2014). Sequence-based typing of HLA: an

improved group-specific full-length gene sequencing approach. *Methods in Molecular Biology (Clifton, N.J.)*, *1109*, 101–114. https://doi.org/10.1007/978-1-4614-9437-9_7

Walsh, C. (2000). Molecular mechanisms that confer antibacterial drug resistance. *Nature*, *406*(6797), 775–781. https://doi.org/10.1038/35021219

Wang, E., Sun, H., Wang, J., Wang, Z., Liu, H., Zhang, J. Z. H., & Hou, T. (2019). End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. *Chemical Reviews*, *119*(16), 9478–9508. https://doi.org/10.1021/ACS.CHEMREV.9B00055

Wang, F., Zhou, H., Wang, X., & Tao, P. (2019). Dynamical Behavior of β-Lactamases and Penicillin- Binding Proteins in Different Functional States and Its Potential Role in Evolution. *Entropy 2019, Vol. 21, Page 1130*, *21*(11), 1130. https://doi.org/10.3390/E21111130

Wang, J., Chu, L., Wojnárovits, L., & Takács, E. (2020). Occurrence and fate of antibiotics, antibiotic resistant genes (ARGs) and antibiotic resistant bacteria (ARB) in municipal wastewater treatment plant: An overview. *Science of The Total Environment*, *744*, 140997. https://doi.org/10.1016/J.SCITOTENV.2020.140997

Wang, J., & Gardinali, P. R. (2014). Identification of phase II pharmaceutical metabolites in reclaimed water using high resolution benchtop Orbitrap mass spectrometry. *Chemosphere*, *107*, 65–73. https://doi.org/10.1016/J.CHEMOSPHERE.2014.03.021

Wang, W., Wei, X., Wu, L., Shang, X., Cheng, F., Li, B., Zhou, X., & Zhang, J. (2021). The occurrence of antibiotic resistance genes in the microbiota of yak, beef and dairy cattle characterized by a metagenomic approach. *The Journal of Antibiotics*, *74*(8), 508. https://doi.org/10.1038/S41429-021-00425-2

Welker, M., & van Belkum, A. (2019). One System for All: Is Mass Spectrometry a Future Alternative for Conventional Antibiotic Susceptibility Testing? *Frontiers in Microbiology*, *10*. https://doi.org/10.3389/FMICB.2019.02711

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, *2*(1), 37–52. https://doi.org/https://doi.org/10.1016/0169-7439(87)80084-9

Wright, G. D. (2011). Molecular mechanisms of antibiotic resistance. *Chemical Communications (Cambridge, England)*, *47*(14), 4055–4061. https://doi.org/10.1039/C0CC05111J

Wu, J., Gao, T., Guo, H., Zhao, L., Lv, S., Lv, J., Yao, R., Yu, Y., & Ma, F. (2023). Application of molecular dynamics simulation for exploring the roles of plant biomolecules in promoting environmental health. *Science of The Total Environment*, *869*, 161871. https://doi.org/https://doi.org/10.1016/j.scitotenv.2023.161871

Yang, H., Li, H., Yang, H., & Li, H. (2019). Implementation of Manifold Learning Algorithm Isometric Mapping. *Journal of Computer and Communications*, *7*(12), 11–19. https://doi.org/10.4236/JCC.2019.712002

Yang, Y., Li, B., Zou, S., Fang, H. H. P., & Zhang, T. (2014). Fate of antibiotic resistance genes in sewage treatment plant revealed by metagenomic approach. *Water Research*, *62*, 97–106. https://doi.org/10.1016/J.WATRES.2014.05.019

Yong, X., & Zhang, L. T. (2013). Thermostats and thermostat strategies for molecular dynamics simulations of nanofluidics. *The Journal of Chemical Physics*, *138*(8). https://doi.org/10.1063/1.4792202

Zeng, H., Li, J., Zhao, W., Xu, J., Xu, H., Li, D., & Zhang, J. (2022). The Current Status and Prevention of Antibiotic Pollution in Groundwater in China. *International Journal of Environmental Research and Public Health*, *19*(18). https://doi.org/10.3390/IJERPH191811256

Zeng, X., & Lin, J. (2013). Beta-lactamase induction and cell wall metabolism in Gram-negative bacteria. *Frontiers in Microbiology*, *4*(MAY), 128. https://doi.org/10.3389/FMICB.2013.00128

Zhang, L., Chen, F. X., Zeng, Z., Xu, M., Sun, F., Yang, L., Bi, X., Lin, Y., Gao, Y. J., Hao, H. X., Yi, W., Li, M., & Xie, Y. (2021). Advances in Metagenomics and Its Application in Environmental Microorganisms. *Frontiers in Microbiology*, *12*, 3847. https://doi.org/10.3389/FMICB.2021.766364

Zhang, Y., Liu, L., & Ren, L. (2020). Liquid chromatography-tandem mass spectrometry (LC-MS/MS) determination of cantharidin in biological specimens and application to postmortem interval estimation in cantharidin poisoning. *Scientific Reports 2020 10:1*, *10*(1), 1–8. https://doi.org/10.1038/s41598-020-67278-x

Zhang, Z. (2016). Introduction to machine learning: k-nearest neighbors. *Annals of Translational Medicine*, *4*(11). https://doi.org/10.21037/ATM.2016.03.37