



Molecular Modelling and Machine Learning for the Investigation of 2- Oxazolidinone Ribosomal Antibacterials

McKenna Eleanor Buckley

B.Sc (Chem)/B.IT (InfoSys)

Submitted in fulfilment of the requirements for the degree of
Master of Philosophy

School of Chemistry and Physics

Faculty of Science

Queensland University of Technology

2023

Keywords

Molecular docking, machine learning, oxazolidinone, linezolid, RNA, ribosome

Abstract

The bacterial ribosome, specifically the 23S ribosomal RNA (rRNA) component of the 50S subunit, is the target for antibacterial molecules (antibiotics) like oxazolidinones. These antibiotics bind to the ribosome and interfere with protein synthesis, ultimately leading to bacterial cell death. This thesis uses structure-based and ligand-based molecular modelling approaches to explore molecular recognition and demonstrate the efficacy of two molecular modelling approaches using a large dataset of oxazolidinone derivatives with known 50S ribosome inhibitory activity against *Staphylococcus aureus*.

A total of 5 ribosome adaptable docking programs were used, AutoDock 4, AutoDock Vina, DOCK 6, rDOCK and RLDOCK. The molecular docking study found that AutoDock, AutoDock Vina, and DOCK 6 performed better than rDock and RLDOCK for redocking the native ligands. Structural modifications and the class of MIC activity had minimal to no influence on the performance of the linezolid derivatives when docked against the *S. aureus* crystal structure. The study identified ten derivatives that scored within the top-performing range, with three binding in the native linezolid binding site. However, poor electron density, induced fit mechanisms in the binding pocket for mutation or resistant strains of bacteria, docking programs ignoring the intrinsic flexibility of nucleic acids, and other limitations explain the poor results and low correlation of MIC and docking score for all derivatives.

For the machine learning study, eleven algorithms were fine-tuned, tested, and trained on the pMIC data of an oxazolidinone dataset and their 1440 1D, 2D, and 3D descriptors. Mutual information was used to reduce the number of descriptors, ranging from a cut-off mutual information score of 0.4 to the Top 100, 80, 60, 40 and 20 descriptors. Algorithms include K-Neighbors, Ridge, Lasso, Elastic Net, Gradient Boosting, Random Forest, Ada Boost, Extra Trees, Decision Tree, and Support Vector Regression (SVR). Evaluations on subsets of descriptors from the primary datasets illustrated that the performance of SVR was more stable than other algorithms with an R^2 of 0.632. Using mutual information dramatically improved the performance compared to using all descriptors as is, with the top-performing descriptor set (Top 80) having the highest SVR performance of R^2 of 0.757. While the performance of the algorithms is ideal, it is suggested to incorporate a more

extensive range of classes of antibiotic data to investigate the performance of descriptors and the algorithms thoroughly.

The study suggests that for future structure-based works on RNA-ligand docking, methods such as ensemble docking and polarized molecular mechanics scoring function/force fields can be used to address the high flexibility and negative charges of RNA. For ligand-based studies, incorporating a more extensive and more structurally diverse dataset of antibiotics from a broader range of antibacterial classes and expanding the use of descriptors to incorporate a range of fingerprint techniques can improve the analysis. Additionally, future studies can include biofilm targets and develop tools to differentiate between gram-positive and gram-negative bacteria. Open-source deep learning libraries or in-house graph-based machine learning models can also be used to expand the research.

List of Publications

Buckley, M.E.; Ndukwe, A.R.N.; Nair, P.C.; Rana, S.; Fairfull-Smith, K.E.; Gandhi, N.S. Comparative Assessment of Docking Programs for Docking and Virtual Screening of Ribosomal Oxazolidinone Antibacterial Agents. *Antibiotics* 2023, 12, 463.
<https://doi.org/10.3390/antibiotics12030463>

Buckley, M.E.; Fairfull-Smith, K.E.; Rana, S.; Gandhi, N.S. Molecular descriptor-based machine-learning assisted QSAR model development for ribosomal oxazolidinone activity. Manuscript ready for submission to Elsevier *Computers in Biology and Medicine*.

List of Conference presentations

External

1. **Buckley, M.E.**; Ndukwe A.R.N.; Fairfull-Smith, K.E.; Rana, S.; Gandhi, N.S. *Assessing molecular docking tools for targeted drug discovery of oxazolidinones*. Association of Molecular Modellers (MM) Australasia MM Conference. December 2021. Virtual. Poster.
2. **Buckley, M.E.**; Ndukwe A.R.N.; Fairfull-Smith, K.E.; Rana, S.; Gandhi, N.S. *Modelling ligand docking to RNA in the design of oxazolidinone antibiotics*. Royal Australian Chemical Institute (RACI) 2022 National Congress. July 2022. Poster.
3. **Buckley, M.E.**; Rana, S.; Fairfull-Smith, K.E.; Gandhi, N.S. *Evaluating ML algorithms for antibacterial discovery*. Queensland Annual Chemistry Symposium (QACS) 2022. November 2022. Presentation.
4. **Buckley, M.E.**; Rana, S.; Fairfull-Smith, K.E.; Gandhi, N.S. *ML approaches for the discovery and optimization of oxazolidinones with therapeutic potential*. 2022 Australian Physiological Society (AuPS)/Australian Society for Biophysics (ASB) Scientific meeting. November 2022. Poster.

Internal

1. Jayarathna, D.K.; Schuurs, Z.P.; **Buckley, M.E.**; Barbhuiya. T.K. *Going against the grain - cutting edge approaches to combating diseases*. Centre for Genomics and Personalised Health (CGPH) End of Year Showcase. November 2021. Group Presentation.
2. **Buckley, M.E.**; Rana, S.; Fairfull-Smith, K.E.; Gandhi, N.S. *Assessing ML algorithms for drug discovery*. Centre for Genomics and Personalised Health (CGPH) End of Year Showcase. September 2022. Presentation.
3. **Buckley, M.E.**; Rana, S.; Fairfull-Smith, K.E.; Gandhi, N.S. *Evaluating ML algorithms for antibacterial discovery*. Centre for Genomics and Personalised Health (CGPH) End of Year Showcase. November 2022. Lightning Talk Presentation and Poster.

Contents

Keywords	ii
Abstract	iii
List of Publications	v
List of Conference presentations	vi
Contents	vii
List of Figures	x
List of Tables	xv
List of Supplementary Materials and Appendices	xvi
List of Abbreviations	xvii
Acknowledgements.....	xxi
Chapter 1 — Introduction.....	1
1.1 Background	2
1.2 Aims	3
1.3 Significance.....	3
1.4 Thesis Outline	4
Chapter 2 — Literature Review	6
2.1 Classification of antibacterials	7
2.2 Ribosomal antibiotics	10
2.2.1 Structural characterisation of ribosomal complexes.....	12
2.2.2 Classes of ribosomal antibiotics and their structures.....	12
2.2.3 Chemistry of oxazolidinone antibiotics and modes of action.....	14
2.2.4 Linezolid.....	20
2.2.3 Linezolid mutation and resistance	21
2.3 Oxazolidinone antibiotic molecular studies	24
2.3.1 Structure based methods.....	24
2.3.2 Ligand based methods	26

2.3.3 Molecular dynamics (MD)	27
2.4 Machine Learning and drug discovery studies	28
2.5 Summary	31
Chapter 3 — Methods	32
3.1 Molecular docking.....	33
3.1.1 Search algorithm.....	34
3.1.2 Scoring functions	36
3.2 Quantitative Structural Activity Relationship (QSAR).....	38
3.3 Machine Learning (ML).....	39
3.3.1 k-nearest neighbour (k-NN).....	42
3.3.2 Ridge Regression (RR), Lasso Regression (LR), and Elastic Net (EN).....	42
3.3.3 Adaboost and Gradient Boosting (GB)	43
3.3.4 Decision Tree (DT), Random Forest (RF), and Extra Trees (ET).....	43
3.3.5 Support Vector Machines (SVM).....	44
3.3.6 t-distributed stochastic neighbour embedding (t-SNE)	45
3.4 Neural Networks (NN) and Deep Learning (DL)	46
Chapter 4 — Comparative assessment of docking programs for docking and virtual screening of oxazolidinone antibacterial agents.....	48
4.1 Introduction	49
4.2 Statement of Author Contributions	50
4.3 Published Paper	51
Chapter 5 — Molecular descriptor-based machine-learning assisted QSAR model development for ribosomal oxazolidinone activity	101
5.1 Introduction	102
5.2 Statement of Author Contributions	103
5.3 Paper in Review	104
Chapter 6 — Conclusions and future work.....	126
6.1 Summary and concluding remarks	127

6.2 Contributions to the literature	128
6.3 Research Limitations.....	129
6.3.1 Limitations of current modelling techniques within RNA	129
6.3.2 Challenges and Limitations of ML in drug discovery	132
6.4 Directions for further research	133
6.4.1 Plans for molecular docking	133
6.4.2 Future Directions for machine learning.....	135
6.5 Final conclusion	138
Bibliography	139

List of Figures

Figure 2.1 - Structure of gram-positive and gram-negative bacterial cells, adapted from Kapoor et al., (2017) (Kapoor et al., 2017).....	7
Figure 2.2 - Several classes of antibacterial compounds that are typically classified based on their bacterial target, modelled on a gram-positive bacterium. The image is taken from https://openstax.org/books/microbiology/pages/14-3-mechanisms-of-antibacterial-drugs and accessed on 25 th October 2021.....	9
Figure 2.3 - A simple diagram showing subunit components of ribosomal subunits. The image was obtained from Kumar (2021) (Kumar, 2021)	11
Figure 2.4 - Examples of structures from each bacterial group that specifically target the ribosome.....	13
Figure 2.5 - Oxazolidinones currently on the market, linezolid (6) and tedizolid (7), alongside other in-development oxazolidinones Cadazolid (8), Contezolid (9), Delpazolid (10), Eperezolid (11), Posizolid (12), Radezolid (13), Ranbezolid (14), Sutezolid (15), TBI-223 (16), and T145 (17).....	18
Figure 2.6 - Structure of linezolid, was adapted from Zhao et al. (Q. Zhao et al., 2021).	20
Figure 2.7 - Some examples of antibacterial compounds identified by ML methods.....	29
Figure 3.1 - Illustrates docking a small molecule ligand (grey) to an RNA target (brown). Note: the RNA displayed is a selected area of the complete structure, selected 20Å from the ligand binding site.....	33
Figure 3.2 - Overview of ML model concepts split into 4 stages. (A) Representations of the variety of data used in ML styles, (B) Potential ML methods that can be used, (C) Splitting of data into training, validation, and testing for said algorithms and (D) evaluation methods for algorithm performance. Image adapted from Lulka and Stokes (2022) (Lluka & Stokes, 2022)	40
Figure 3.3 – Example of T-SNE from the Tasic et al. transcriptomic dataset (Tasic et al., 2018), encompassing 23,822 cells from adult mouse cortex's. Image was adapted from Kobak and Berens (2019) (Kobak & Berens, 2019).....	45
Figure 3.4 - Comparing feature extraction and classification between ML and DL. Image adapted from Khan et al. (2021) (Khan et al., 2021)	47

Figure 4.1 - The *S. aureus* (strain NCTC832) 50S ribosome crystal structure (PDB ID 4WFA) with 50S ribosomal proteins (not all proteins are shown for clarity) and 5s and 23s rRNA ribosomes coloured and labelled with (a) rotated side view of the crystal structure with the Linezolid binding site displayed (b) Close-up view of the linezolid at the ribosomal binding site. Linezolid is shown in aqua blue, while the residues that interact in the active site are coloured by element and labelled with their corresponding residue name, with *E. coli* rRNA numbering. The interactions between the *S. aureus* residues and the ligand in the binding pocket are shown via the striped line, coloured by the following interactions: π -stacked (pink), conventional (bright green), carbon (light green), and fluorine (blue). Images were created using Protein Imager (Tomasello et al., 2020) and BIOVIA Discovery Studio (BIOVIA, 2022).54

Figure 4.2 - Oxazolidinones on the market are linezolid (1) and tedizolid (2), with the A ring, B ring, C ring and C-5 side chain indicated on linezolid. Examples of modifications to the linezolid scaffold include morpholine modification (3), C-5 side-chain modifications (4) and (5), and oxazolidinone modification (6) from the oxazolidinone dataset. The image was created using ChemDraw.....55

Figure 4.3 - Comparison of the redocking performance for the linezolid (a) Examples of results produced when redocking linezolid to its native crystal structure (PDB: 4WFA). The native ligand pose is beige, the selected redocked ligand is pink, and the 2 Å mesh of the electron density map is grey. (b) The average scoring for each program is displayed in a heatmap using normalized scoring values, showing all linezolid ligand crystal structures (c) The average RMSD for each program is displayed in a heatmap, showing all linezolid ligand crystal structures. The crystal structures that scored at or below 2.5 Å are indicated by red squares. Figure 3a was visualised through UCSF Chimera (v 1.15) (Pettersen et al., 2004). ..59

Figure 4.4 - The five selected docking programs used to validate crystal structures, using the top-performing values from each crystal structure. Each docking program was ranked by scoring (a) and RMSD (b). Each program's overall success rate was used to calculate the average, using the normalized top-scoring poses of each program. Both graphs display the minimum and maximum values, with the median values indicated with the orange line and the circles being the outliers. (c) A stacked bar chart demonstrating each selected program's success rate is shown. The green, yellow, and red areas correspond to the percentage of models that achieve the RMSD of ≤ 2.5 , ≤ 4.5 , and > 4.5 Å, respectively....61

Figure 4.5 - The following structures show the modification to the oxazolidinones derivatives based on the linezolid structure in red. They are split into the following categories: morpholino modifications (2d), aryl modifications (4c), base modifications (47a), tail modifications (6g), and other oxazolidinone modifications (70e). Note that some structures make have multiple areas of modification. This image is to detail an example of what these modifications may be.....	64
Figure 4.6. - Docking performance of DOCK 6 towards key structural modification of linezolid and correlation with the MIC of against <i>S. aureus</i> . (a) The average performance of structural modifications when docked with DOCK 6. There are some slight differences between each modification, but nothing of significance (b) percentage breakdown of the number of structural modifications. (c) Some derivatives had activity against <i>S. aureus</i> . Similar to the structural modifications, MIC activities had no significant difference.....	65
Figure 4.7 - A brief overview of the RF classifier. The classifier is trained on each derivative's Morgan Fingerprint bit vectors to identify sections of the structure that may contribute to over or underpredictions by the scoring function. Over-predicted derivatives are indicated by green and under-predicted by orange. The top-performing derivatives are outliers in the -90 kcal/mol to -100 kcal/mol range.....	70
Figure 4.8 - Scatter plots showing coefficient of Pearson's correlation (r) between the experimental pMIC and predicted pMIC for test set compounds (a) Absolute DOCK 6 score (b) DOCK 6 rescore.....	72
Figure 5.1 - Histogram of <i>S. aureus</i> pMIC dataset with the spread of pMIC data (A) and grouped by <i>S. aureus</i> pMIC group (B).	107
Figure 5.2 – Dimensionality reduction by t-SNE on fingerprint bit vectors. The colour bar indicates pMIC values from -2.5 (dark purple) to 1 (yellow).....	109
Figure 5.3 – Core structures of molecules in Clusters A, B, C, D, E and F from the t-SNE analysis, shown in (Cluster A-F), respectively. Most core structures, except for Cluster C, have sections showing [X, X]. This indicates that a structure in that group has the same core structure but a slight alteration in the element.	110
Figure 5.4 – Mean performance of the algorithms of predicting the pMIC separated by the pMIC activity groups (left) and clusters determined by the t-SNE (right), evaluated by the coefficient of determination (R^2) test score. The table is shaded by order of performance, the lowest being red and the highest being dark green.	111

Figure 5.5 – Mean performance of the algorithms of predicting the pMIC value of training data and test data, evaluated by the coefficient of determination (R^2) test score. The table is shaded by order of performance, the lowest being red and the highest being dark green.

.....113

Figure 5.6 – Mean performance of SVR on predicting the pMIC value of *S. aureus* using various amounts of descriptors, evaluated by MAE, R^2) and RMSE. The down arrows indicate that a lower score is better, while the up arrow indicates a higher score is better.

.....115

Figure 5.7 – Structural breakdown of virtual screening dataset, with Linezolid (1) shown for baseline structural comparison. The dataset is broken down into the anticoagulant Rivaroxaban (2), 3-(3-pyridyl)-oxazolidinone-5-methyl ester derivatives (3) (Jin, Chen, et al., 2022), 3-(pyridine-3-yl)-2-oxazolidinone derivatives (4) (Jin, Wang, et al., 2022), piperazinylphenyloxazolidinones (5) (Das et al., 2005; Gandhi, 2007), arylpiperazinylphenyloxazolidinones with the diversification of the N-substituents (6) (Gandhi, 2007; Jang et al., 2004), and other derivatives from the Zhao et al. dataset that did not have any recorded MIC values against *S. aureus* (Q. Zhao et al., 2021). All structures are drawn in ChemDraw.....116

Figure 5.8 – *S. aureus* pMIC against the predicted values for the external dataset trained off All the descriptors, 2D and 3D descriptors, top 80 descriptor dataset. Each graph has Pearson's correlation indicated.....117

Figure 6.1 - Examples of antiviral and anticancer oxazolidinone structures136

Supplementary Figure 4.1 - Electron dentistry maps (at binding site) for selected ribosomal crystal structures: 3CPW (2.7 Å), 3CXC (3.0 Å), 3DLL (3.5 Å), 4WFA (3.4 Å), 6DDD (3.1 Å), 6DDG (3.1 Å), 6QUL (3.0 Å), 6WQN (2.9 Å), 6WQQ (3.1 Å), 6WRS (3.2 Å), and 6WRU (3.1 Å). The resolution is shown in the mesh in blue, with the XYZ axis indicator displayed (red, green, blue, respectively). Species type is also listed. Note: the binding site residues are not shown for clarity of the electron density, due to other residue that obstructed the full view of the ligand. ChimeraX (v.1.4) was used for the visualization, in conjunction with the ISOLDE toolkit.....84

Supplementary Figure 4.2 - DOCK 6 scoring of the dataset split into the specific structural modification groups. Many values stayed above -60kcal/mol, while the top performers were all below -90kcal/mol.	86
Supplementary Figure 4.3 - Distribution of molecular features of the derivatives (a) pMIC of <i>S. aureus</i> docked against their respective PDBs (4WFA), and the (b) number of rotatable bonds, (c) molecular weight, and the (d) number of atoms in the molecule vs the absolute docking score for 4WFA. Pearson values are displayed.	87
Supplementary Figure 4.4 - DOCK 6 Top three poses (blue) of each top-performing derivative superimposed against linezolid (beige) in the pocket of 4WFA. Note: the nucleotide residues in the binding site are not shown for clarity of the posing.	89
Supplementary Figure 4.5 - The intermolecular interactions of linezolid and derivatives 31, 29f, and 67 with the active site pockets of <i>S. aureus</i> . Left side (3D representation): the docked residues of the target are shown in the rose stick model (labelled), and compounds are in the grey stick. Right side (2D representation): compounds shown in grey with the residue interactions displayed. E. coli residue numbering is used for both 3D and 2D representations. The dashed lines represent the different interactions, and their colour is the interaction type. The coloured circles represent residues with letter code (A for Adenine, C for Cytosine, G for Guanine, or U for Uracil), chain identifier (Chain X) and residue number. 2D visualization was carried out using BIOVIA Discovery Studio.	90
Supplementary Figure 4.6 - 3D electrostatic-hydrophobic and shape properties. The pharmacophore model generated from linezolid, cadazolid, radezolid, tedizolid, and contezolid presents two distinct and separated regions: a positive electrostatic potential region in red, and a negative electrostatic potential region, in cyan. Image generated using Flare™ from Cresset® (v.3) (Bauer & Mackey, 2019; Cheeseright et al., 2006; Cresset Group, 2022; Kuhn et al., 2020).	99

List of Tables

Table 2.1 - General mechanisms of action of antibiotics and corresponding drug class examples with their natural origin	8
Table 2.2 - The main representative oxazolidinone antibiotics licensed or under development. Known drug names (and any associated names), features of approval, and current stage of approval are listed. The table was partially adapted from Foti et al. 2021 (Foti et al., 2021).....	15
Table 2.3 - 23S rRNA linezolid resistance mutations and related organisms. The table is adapted from (Long & Vester, 2012).....	22
Table 2.4 - List of oxazolidinone molecular docking studies on various bacterial species and strains	25
Table 3.1 - Summary of the target, scoring functions, and search algorithms of the programs tested in the study.	34
Table 4.1 - Summary of the target, scoring functions, search algorithms of the programs tested in the study.....	57
Table 4.2 - List of derivatives determined to be the “top-performing” derivatives, with their similar base structure indicated. The structural differences are indicated by the R, X1 and X2 columns and the docking score and MIC for <i>S. aureus</i>	67
Supplementary Table 4.1 - ABS docking score and descriptors of training set.....	92
Supplementary Table 4.2 - Experimental and predicted pMIC values of the rescores for the test set compounds with descriptors.....	93
Supplementary Table 5.1 – Top descriptors determined from mutual information.....	119
Supplementary Table 5.2 Machine learning parameters used in the study. After some initial testing of algorithms, it was decided to make some adjustments and modifications to the algorithm parameters used in the original study. These additional modifications are indicated in bold.....	124

List of Supplementary Materials and Appendices

Supplementary Discussion 4.1 - Field-template pharmacophore.....	98
Supplementary Discussion 5.1 - Types of descriptors used	121
Supplementary Method 4.1 - Pharmacophore and Field Template Generation.....	100

List of Abbreviations

AATS	Atom-Type E-state
ABSSI	acute bacterial skin and skin structure infections
AD4	AutoDock 4
AI	Artificial intelligence
ANN	artificial neural network
ATCC	American Type Culture Collection
BCUT	Burden-Centered Fragments
CoMFA	comparative molecular field analysis
cSSTI	complicated skin and soft tissue infections
<i>D. radiodurans</i>	<i>Deinococcus radiodurans</i>
DL	deep learning
DMS	Dot molecular surfaces
DNA	Deoxyribonucleic acid
DT	Decision trees
<i>E. coli</i>	<i>Escherichia coli</i>
<i>E. faecalis</i>	<i>Enterococcus faecalis</i>
EN	Elastic Net
ET	Extra tree
GB	gradient boosting
<i>H. marismortui</i>	<i>Haloarcula marismortui</i>
HBA	number of H bond acceptors
HBD	number of H bond acceptors
HOF	heat of formation
HPC	High-performance computing

<i>K. pneumoniae</i>	<i>Klebsiella pneumoniae</i>
k-NN	k-nearest neighbours
LNZ	Linezolid
LR	Lasso regression
LUMO	Lowest Unoccupied Molecular Orbital
MAE	Mean of Absolute value of Errors
MD	Molecular dynamics
MI	Mutual Information
MIC	minimum inhibitory concentration
MIGR	Mutual Information Gain for Regression
MIS	Mutual Info Score
ML	machine learning
MM/GBSA	Molecular Mechanics/Generalized Born Surface Area
MolWt	molecular weight
mRNA	messenger RNA
MRSA	Methicillin-resistant <i>Staphylococcus aureus</i>
MSSA	methicillin-susceptible <i>S. aureus</i>
<i>N. meningitidis</i>	<i>Neisseria meningitidis</i>
NA	Nucleic acid
NCTC	National Collection of Type Cultures
NMR	Nuclear magnetic resonance
NN	neural networks
NumRings	Number of Rings
<i>P. aeruginosa</i>	<i>Pseudomonas aeruginosa</i>
<i>P. meningitis</i>	<i>Pneumococcal meningitis</i>

PCA	Principle component analysis
PDB	Protein Data Bank
pMIC	Log of the MIC
POAP	Parallelized Open Babel & Autodock suite Pipeline
PTC	peptidyl transferase centre
QSAR	quantitative structure-activity relationship
R ²	correlation coefficient
RCSB	Research Collaboratory for Structural Bioinformatics
RF	random forests
RMSD	Root mean square deviation
RMSE	root-mean-square error
RNA	Ribonucleic acid
RP	recursive partitioning
RR	Ridge regression
rRNA	ribosomal RNA
<i>S. agalactiae</i>	<i>Streptococcus agalactiae</i>
<i>S. aureus</i>	<i>Staphylococcus aureus</i>
<i>S. capitis</i>	<i>Staphylococcus capitis</i>
<i>S. epidermidis</i>	<i>Staphylococcus epidermidis</i>
<i>S. faecalis</i>	<i>Streptococcus faecalis</i>
<i>S. pyogenes</i>	<i>Streptococcus pyogenes</i>
SAR	structure-activity relationships
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SF	Scoring Function
SMILES	Simplified molecular-input line-entry system

SNE	Stochastic Neighbor Embedding
SSSIs	skin and skin structure infections
SVM	support vector machine
SVR	Support Vector Regressor
TPSA	topological polar surface area
tRNA	Transfer ribonucleic acid
t-SNE	t-distributed stochastic neighbour embedding
TZD	Tedizolid
VDW	van der Waals
Vina	AutoDock Vina
VRE	vancomycin-resistant enterococci
VREF	vancomycin-resistant <i>Enterococcus faecium</i>
VS	Virtual screening

Acknowledgements

The completion of this thesis would not have been possible without the support and patience of a few individuals.

To my supervisor Dr Neha Gandhi, I am eternally grateful for how you have helped me grow over the years. It started all the way back when I was an undergrad undertaking the Science capstone with you, allowing me to realise that I loved doing research - I just had to find the area that was suited for me. I thank you for fighting for me and encouraging me in every step and misstep. I know my last-minute submissions were not easy for you (and I am very sorry for that!!), so I thank you immensely for all your time and patience.

Thank you to my associate supervisors, Professor Kathryn Fairfull-Smith from QUT and Associate Professor Santu Rana from Deakin University. All your time and expertise have helped me form this thesis into what it is now.

Thank you to the eResearch Office for the computational (and data visualization) resources and services used in this thesis, as most of this thesis would not have occurred without them. To QUT's School of Chemistry and Physics and the Centre of Genomics and Personalised Health for the resources and funding.

Thank you to the current and former members of Dr Gandhi's research group (Dulari, Zach, Tabassum and Lauren) who helped me out along the way and to the other researchers in our area who lent a hand. Also, I would like to thank all the past and present officemates in M209 for all the support and lovely memories. It always helped to have fellow students in the same boat as me to realise that (1) I'm not alone, (2) we all make mistakes, and (3) we're all just doing our best.

To my parents, Jan and John. I know you will most likely never read this thesis in its entirety (I don't blame you), but both of you have always supported me and cheered me on with every milestone I achieved. Even with you having absolutely no clue what I was doing, you were always excited for me. I will always be eternally grateful that both of you raised me in an environment that allowed me to explore what I enjoyed, regardless of the outcome.

To my friends who patiently listened to my frustrations, celebrated my successes, and provided emotional support, thank you for being there for me in both good times and bad. You all provided much-needed breaks from work, shared laughter, and memorable moments, and you helped me maintain a sense of balance in my life.

Finally, it wouldn't be a McKenna style acknowledgement section without me thanking my (probably way too many) pets. You contributed absolutely nothing to this thesis but thank you for your unconditional love and adorable distractions. You have brought so much joy and happiness into my life, and I am grateful for your presence.

What a ride it's been. Enjoy.

Chapter 1 — Introduction

1.1 Background

One of the most straightforward questions that can be asked about molecular diversity is how many usable organic molecules are out there. The usability of organic molecules is up to interpretation, as it depends on the context in which they are being considered. A molecule may be usable as a drug for humans, but not necessarily for animals or plants. Similarly, a molecule may be usable as a pesticide, but not necessarily as a food additive. The chemical space refers to the vast array of all possible chemical compounds that can be synthesised or exist naturally. It is a theoretical concept describing the multidimensional space encompassing all possible combinations of molecular structures, properties, and activities. However, searching the chemical space is far from easy. An upper estimate says it contains 10^{180} compounds, more than twice the magnitude of the number of atoms in the universe (Gorse, 2006). The total number of possible drug-like molecules has been estimated to exceed 10^{63} , with small molecule targets reaching 10^4 in size (Reymond, 2015). An example of these small druggable targets is antibacterials.

Antibacterials are a class of drugs used to treat disease and bacterial infections, in addition to enabling many modern medical procedures that would have historically ended in death, such as cancer treatment, organ transplants, and open-heart surgery (Hutchings et al., 2019). These drugs target bacterial cells by inhibiting their growth or by killing them outright, as well as targeting specific bacterial structures or functions, such as cell wall synthesis, protein synthesis, or DNA replication. Exploring chemical space is essential for discovering new antibacterial agents, and chemists use a variety of approaches to generate new chemical compounds and test their potential antibacterial activity. One approach is synthesising structurally diverse compounds and screening them for antibacterial activity, and another approach is to modifying existing compounds to optimise their antibacterial properties.

The search for new chemical entities in chemical space is an ongoing process. By exploring chemical space, researchers can discover new compounds with novel mechanisms of action or improved efficacy. Traditional ways to explore the chemical space include combinatorial chemistry, high-throughput screening, medicinal chemistry, and empirical modifications of existing compounds. These methods involve systematically synthesising and testing large numbers of chemical compounds to identify those with desired properties or activities, and they have been widely used in drug discovery and other areas of chemical research.

However, these are often expensive and time-consuming, so more computational methods have been used, including structure-based methods such as molecular docking and ligand based methods such as Machine Learning (ML).

Various molecular docking and ML studies have been conducted on smaller subsets of antibiotics, such as oxazolidinones. However, to our knowledge, a docking study has not been conducted with a dataset of this size. The crystal structures examined are RNA molecules targeted by small-molecule oxazolidinone derivatives from Zhao et al.'s large and previously untested dataset (Q. Zhao et al., 2021).

1.2 Aims

This research aims to identify, design, and synthesise novel oxazolidinone-based molecules that inhibit protein synthesis against *Staphylococcus aureus*, with 50S ribosome as the cellular target for these molecules. The availability of 3D structures and target-based 50S ribosome inhibitory data will enable molecular modelling and ML techniques to generate virtual screening models. The main aims are:

- To validate docking programs and eventually virtual screening using a dataset of oxazolidinone class molecules.
- To apply ML algorithms to the oxazolidinone dataset.

1.3 Significance

This project introduces a computation framework to analyse oxazolidinone and linezolid derivatives, a class of antibacterial that inhibits protein synthesis in the 30S and 50S ribosomes. A summary of the significance of this thesis is detailed below.

- 1. Large novel compounds that can treat broad-spectrum bacteria:**
Oxazolidinones are broad-spectrum antibiotics. To our knowledge, a docking study has not been conducted with a dataset of this size that treats broad-spectrum bacteria.
- 2. Understand the molecular recognition of oxazolidinone compounds of ribosomal assembly of *S. aureus*:** The thesis (specifically Chapter 4) focuses on investigating potential interactions of oxazolidinones analogues and *S. aureus* ribosome.

3. **Advancing the knowledge in chemical and molecular sciences:** By conducting a molecular validation and virtual screening study, as well as an ML-focused study, we have determined the potential interactions of the oxazolidinones against *S. aureus*, as well as determining whether the known minimum inhibitory concentration (MIC) values and structural modifications had any influence or impact on the performance on both docking programs and ML algorithms.
4. **Adaptability to other targets or datasets:** These methods can be optimised and applied to study biofilm inhibitors and inhibitors to overcome antimicrobial resistance.

1.4 Thesis Outline

This thesis has been presented as a thesis by publication, and the outline is as follows:

Chapter 2 describes the literature overview, discussing the specific research topics and problems. This chapter introduces the history of the ribosome, ribosomal antibiotics, oxazolidinones and their mode of action, oxazolidinone antibiotic studies, and *in silico* structure and ligand-based design approaches.

Chapter 3 provides a comprehensive description of computational methods used throughout the thesis (corresponding to Chapter 4 and Chapter 5), specifically molecular docking, 2D and 3D quantitative structure-activity relationship (QSAR), and various ML algorithms.

Chapter 4 describes a structure-based study using molecular docking and virtual screening. The study utilizes various molecular docking programs and a large previously untested oxazolidinones dataset with the ribosomal structure of *S. aureus*. This chapter acts as an investigative study on various structure based computational techniques. The corresponding manuscript is published in the MDPI *Antibiotics* special issue “[Ribosomal antibiotics: recent advances](#)”.

Chapter 5 is a ligand-based study detailing the analysis of various ML algorithms and the training and testing of these algorithms on the 2D and 3D descriptors of an extended version of the dataset used in Chapter 4. This chapter acts as an application of various computational approaches and techniques to develop a predictive model for the design of antibacterials. The corresponding manuscript is under review in Elsevier *Computers in Biology and Medicine*.

Chapter 6 concludes the thesis, detailing the contribution to literature, limitations of current modelling techniques, and challenges of ML, as well as directions and plans for further research.

2

Chapter 2 — Literature Review

2.1 Classification of antibacterials

Humans use tonnes of antibiotics per year (Andersson & Hughes, 2010) for medical, veterinary, and agricultural purposes (Levy & Marshall, 2004). Antibiotics are medicines that fight bacterial infections by killing the bacteria or preventing their reproduction. In addition to treating infectious diseases, antibiotics have made many modern medical procedures possible, such as cancer treatment, organ transplants and open-heart surgery (Hutchings et al., 2019). Antibiotics can be categorised based on their target specificity, being either broad or narrow-spectrum. Broad-spectrum antibiotics affect a wide range of bacteria, while narrow-spectrum antibiotics target a specific type of bacteria, such as gram-positive or gram-negative. Gram-positive bacteria like *Bacillus subtilis*, *S. aureus*, and *Streptococcus faecalis* possess a thick cell wall containing many layers of peptidoglycan and teichoic acids (Q. Zhao et al., 2021). Gram-negative bacteria have thin cell walls consisting of various layers of peptidoglycan surrounded by a second lipid membrane containing lipopolysaccharides and lipoproteins. *Escherichia coli*, *K. pneumoniae*, and *Pseudomonas aeruginosa* are examples of gram-negative bacteria (Q. Zhao et al., 2021). Despite their thicker peptidoglycan layer, gram-positive bacteria are more receptive to specific cell wall-targeting antibiotics than gram-negative bacteria due to the absence of the outer membrane (Figure 2.1) (Cunha, 2010).

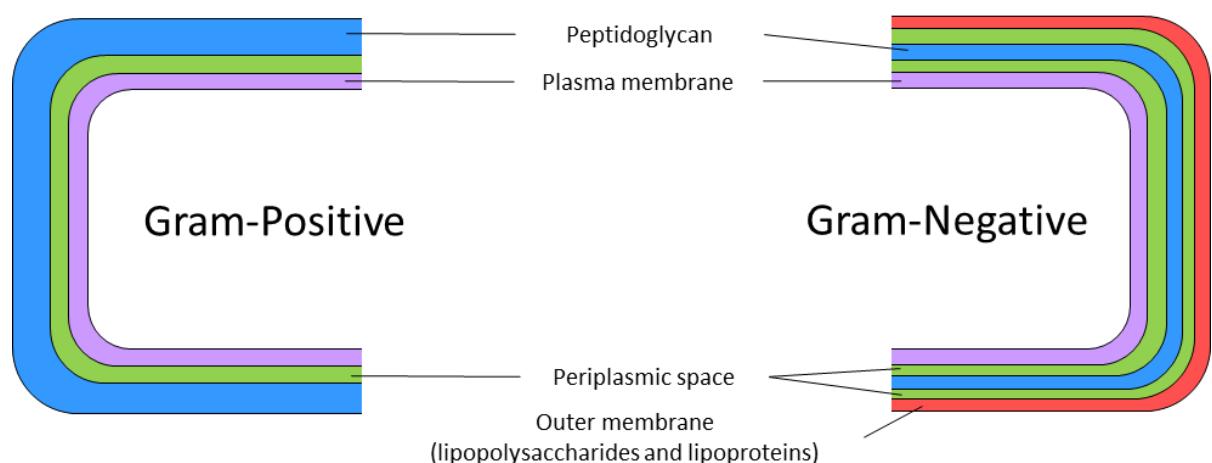


Figure 2.1 - Structure of gram-positive and gram-negative bacterial cells, adapted from Kapoor et al., (2017) (Kapoor et al., 2017).

Antibiotics can also be divided by their product origin or unique mode of action (Table 1). As of 2019, most natural product antibiotics originate from filamentous actinomycetes (Hutchings et al., 2019), which are soil-dwelling, generally gram-positive, anaerobic bacteria with mycelia in a filamentous and branching growth pattern. Some examples of

this type include lincosamides, macrolides and tetracyclines. The remaining naturally product-derived antibiotics are composed of other bacterial natural products and fungal natural products (Hutchings et al., 2019). Other antibiotics have their synthesis inspired by a natural product and are classified as synthetic antibiotics.

Table 2.1 - General mechanisms of action of antibiotics and corresponding drug class examples with their natural origin

Mechanism of action	Drug Class Examples			
	Actinomycetes	Other bacteria	Fungi	Synthetic derivates
Inhibit bacterial cell wall synthesis	Carbapenems, Cycloserines, Phosphonates	Bacitracin, Monobactams, Polypeptides, Polymyxins,	Cephalosporins, Glycopeptides, Penicillins, Enniatins	Pyridinamides, Ethambutol, Thioamides
Inhibit bacterial protein synthesis	Aminoglycosides, Tetracyclines, Macrolides, Lincosamides, Streptogramins, Tuberactinomycins	Mupirocin	Fusidic acid, Pleuromutilins	Oxazolidinones
Inhibit bacterial membrane function	Lipopeptides			
Inhibit bacterial nucleic acid synthesis	Ansamycins, Rifamycin, Fluoroquinolones, Liparmycins			Nitrofurans, Azoles, (Fluoro)quinolones, Phenazines
Inhibit folate synthesis				Sulfonamides, Salicylates, Sulfones,

Alternatively, they are classified based on the bacterial system they affect, in addition to whether they induce cell death (bactericidal) or inhibit cell growth (bacteriostatic). These modes of action are summarised in Figure 2.2. These modes of action include inhibition of the plasma membrane, cell wall, DNA and RNA synthesis, metabolic pathways, or 30S or 50S ribosomes. A brief summary of each target type is as follows below.

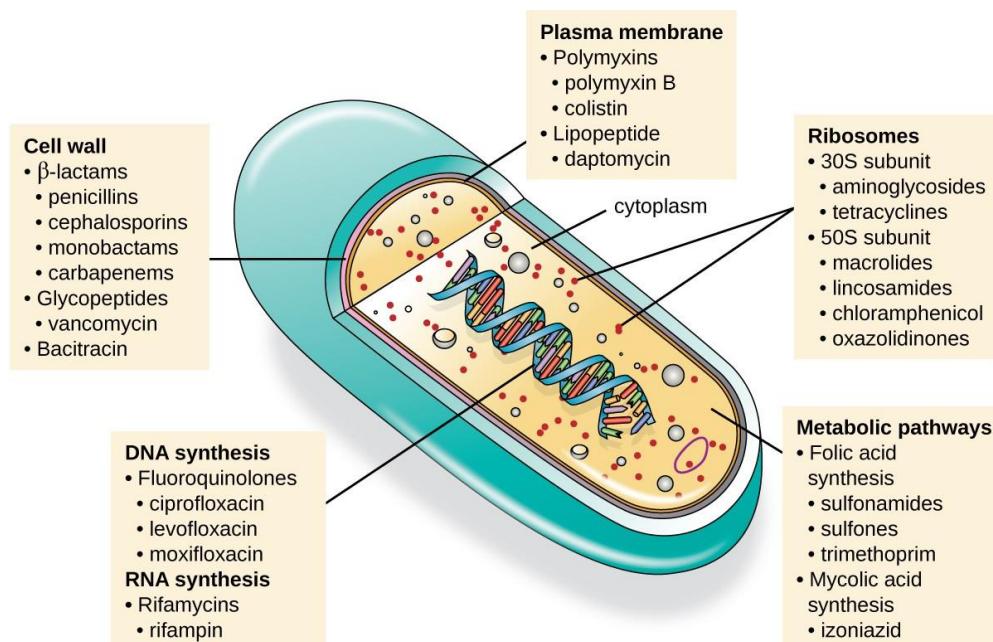


Figure 2.2 - Several classes of antibacterial compounds that are typically classified based on their bacterial target, modelled on a gram-positive bacterium. The image is taken from <https://openstax.org/books/microbiology/pages/14-3-mechanisms-of-antibacterial-drugs> and accessed on 25th October 2021.

β -Lactams and glycopeptides are among the classes of antibiotics that block the biosynthesis of peptidoglycans in the cell wall (Kapoor et al., 2017; Kohanski et al., 2010). This makes cell membranes more susceptible to rupturing due to osmotic movement into the cell wall, known as osmotic lysis. Successful treatment with a cell wall synthesis inhibitor can change cell shape and size and induce cellular stress responses (Chakraverty et al., 2023; Kapoor et al., 2017; Kohanski et al., 2010).

A small number of antibacterials target the bacterial membrane as their modes of action, such as polymyxins and lipopeptides. Polymyxins are lipophilic and interact with the

lipopolysaccharide component of the outer membrane of gram-negative bacteria, disrupting both their outer and inner membranes and killing the bacterial cells. Lipopeptides like daptomycin have a similar mechanism to polymyxins. However, it explicitly targets gram-positive bacteria instead of gram-negative (Chakraverty et al., 2023; Kapoor et al., 2017; Kohanski et al., 2010).

DNA synthesis requires the modification of chromosome coiling through topoisomerase-catalysed strand breakage and re-sealing reactions. These reactions are exploited by the synthetic quinolone class of antibacterial, such as fluoroquinolones, which target DNA-topoisomerase enzymes (Chakraverty et al., 2023; Kapoor et al., 2017; Kohanski et al., 2010). Rifamycins inhibit DNA-dependent transcription by stably binding with high affinity to the β -subunit of a DNA-bound and actively transcribing RNA polymerase (Chakraverty et al., 2023; Kapoor et al., 2017; Kohanski et al., 2010).

Antimetabolites are a class of synthetic drugs that control bacterial infections and are competitive inhibitors for bacterial metabolic enzymes. By inhibiting the enzyme involved in the production of dihydrofolic acid, sulphonamides block bacterial biosynthesis of folic acid. Isoniazid, a mycolic acid antimetabolite, has specific toxicity for mycobacteria and interferes with the synthesis of mycolic acid, essential for mycobacterial cell walls (Chakraverty et al., 2023; Kapoor et al., 2017; Kohanski et al., 2010). The final major class of protein synthesis inhibitors target the 30S or 50S subunits of ribosomes. This class is the focus of this thesis, so each class will be discussed in greater detail in the below sections.

2.2 Ribosomal antibiotics

The simplest ribosomes found in bacteria have a mass of approximately 2.5 million Daltons and are large, intricate enzymes. Both prokaryotic and eukaryotic ribosomes comprise two subunits of varying sizes that are only loosely connected. Figure 2.3 provides a visual representation of this arrangement. The 50S and 30S subunits make up the prokaryotic ribosome, also known as the bacterial ribosome, while the entire ribosome is referred to as 70S based on their respective rates of sedimentation during ultracentrifugation (Schmeing, 2013). The small subunit contains 21 ribosomal proteins (S1 to S21) and a 16S rRNA (rRNA). The large subunit comprises 33 proteins (L1 to L36) and two rRNAs, the 23S and 5S (Lafontaine & Tollervey, 2001; Schmeing, 2013; Shajani et al., 2011; van Bambeke et al., 2017; Wilson, 2014). The 30S subunit is responsible for associating with messenger RNA (mRNA) during translation initiation and mRNA decoding. In contrast, the 50S

subunit contains the peptidyl transferase centre (PTC), the peptide bond formation site (Shajani et al., 2011). Eukaryotic ribosomes exhibit greater complexity due to the composition of their subunits. Specifically, the 40S subunit of eukaryotes is comprised of 18S rRNA and 33 r-proteins, while the 60S subunit is composed of 25S, 5.8S, and 5S rRNAs along with 47 distinct r-proteins (Sahasranaman & Woolford, 2013; Schmeing, 2013). The observed mechanistic differences between prokaryotes and eukaryotes during the initiation and elongation phases of the translation may be attributed to the presence of expansion segments, or insertion elements, in eukaryotic rRNAs as well as 20-30 additional r-proteins. This increased complexity may provide eukaryotes with more flexibility in regulating translation (Sahasranaman & Woolford, 2013).

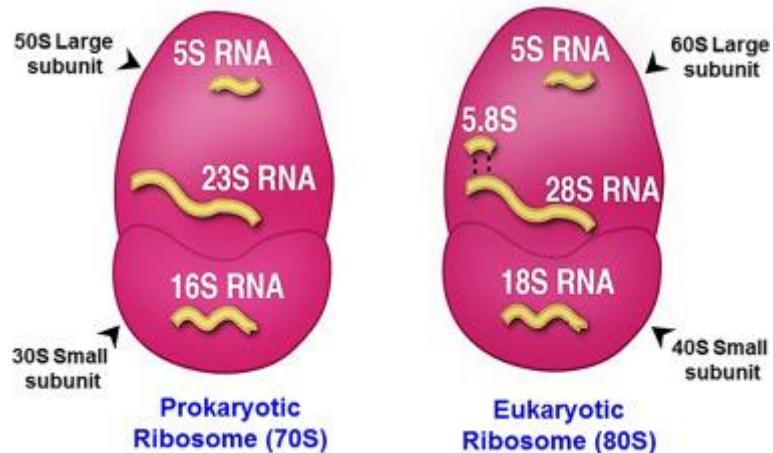


Figure 2.3 - A simple diagram showing subunit components of ribosomal subunits. The image was obtained from Kumar (2021) (*Kumar, 2021*)

The process of ribosome assembly involves several essential stages, including (a) the transcription, processing, and modification of rRNA, (b) the translation and modification of ribosomal proteins, (c) the appropriate folding of both rRNA and ribosomal proteins, (d) the attachment of ribosomal proteins, and (e) the binding and detachment of assembly factors (Shajani et al., 2011). During translation, the 50S subunit of the ribosome is responsible for peptide bond synthesis at the PTC, while the decoding site in the smaller subunit facilitates codon recognition. Transfer ribonucleic acid (tRNA) molecules play a critical role in bridging the two subunits and occupy the A, P, or E (exit) sites of the 50S subunit and the decoding site in the 30S subunit at various times during the synthesis cycle. (Fox, 2010).

2.2.1 Structural characterisation of ribosomal complexes

In 1998 the 9 Å resolution structure of the 50S subunit of the prokaryotic ribosome was solved using crystals of a protein purified from the extremophile archaeon *H. marismortui* (P. Zhao, 2011). This revised resolution revealed numerous features on the ribosome consistent with double-helical RNA and provided evidence confirming that 60% of a ribosome's weight is RNA (Ban et al., 1998). In 1999, the team made further progress toward the 2 Å target, as they reported that they definitively placed protein and RNA structures into a 5 Å resolution map (Ban et al., 1999). The new structure revealed the positions of prominent structural motifs on the ribosome, including the polypeptide exit tunnel, the binding sites for elongation factors G and Tu, and the sarcin-ricin loop. In 2000 they finally reached the goal of obtaining a high-resolution structure of the largest subunit of *Haloarcula marismortui*, publishing the 2.4 Å resolution map in *Science* (Ban et al., 2000). The researchers could definitively place nearly all of the 50S subunit's 3,045 nucleotides and 31 proteins at this resolution (P. Zhao, 2011). The 2009 Nobel Prize in Chemistry was awarded jointly to the team Ramakrishnan, Steitz and Yonath "for studies of the structure and function of the ribosome". There are now numerous crystal structures available for the prokaryotic 50S ribosome. These structures include bacterial ribosomes from various organisms, such as the previously mentioned *H. marismortui*, along with *Deinococcus radiodurans*, *E. coli*, and *S. aureus* (including Methicillin-resistant *S. aureus* (MRSA)). While there are other ribosomal targets in the literature, these bacterial strains were selected for use in chapter 4 of this thesis, so will be discussed in further detail below.

2.2.2 Classes of ribosomal antibiotics and their structures

The significant classes of protein synthesis inhibitors target the 30S (aminoglycosides, tetracyclines) or 50S subunits of ribosomes (macrolides, lincosamides, chloramphenicol, and oxazolidinones).

Aminoglycosides are potent, broad-spectrum antibiotics that act through the inhibition of protein synthesis. The class has been a cornerstone of antibacterial chemotherapy since streptomycin (Figure 2.4, 1) was first isolated from *Streptomyces griseus*. Several other members of the class were introduced over the intervening years, including neomycin (*Streptomyces fradiae*), kanamycin (*Streptomyces kanamyceticus*), gentamicin (*Micromonospora purpurea*), netilmicin (derived from sisomicin), tobramycin (*Streptomyces tenebrarius*), and amikacin (derived from kanamycin) (Krause et al., 2016). Tetracyclines are a class of broad-spectrum antibiotics used to manage and treat various

infectious diseases. Naturally occurring drugs in this class are tetracycline (Figure 2.4, **2**), chlortetracycline, oxytetracycline, and demeclocycline. Semi-synthetic tetracyclines are lymecycline, methacycline, minocycline, rolitetracycline, and doxycycline (Shutter & Akhondi, 2022).

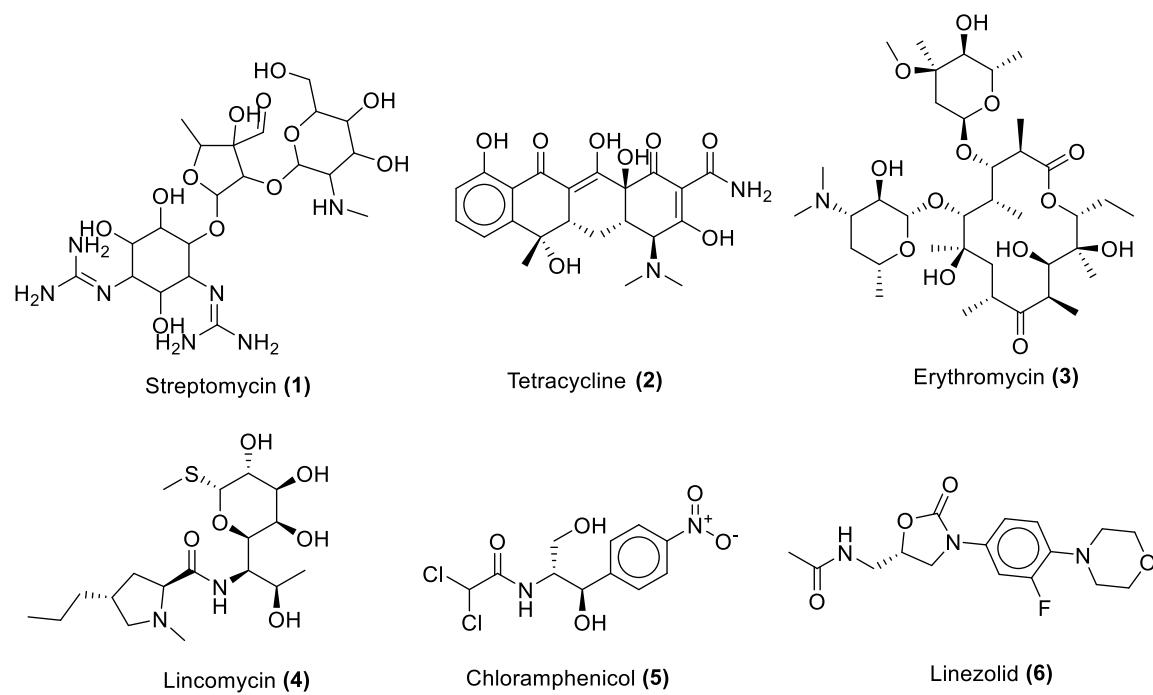


Figure 2.4 - Examples of structures from each bacterial group that specifically target the ribosome

The classes that inhibit the 50S subunit are macrolides, lincosamides, chloramphenicol, and oxazolidinones. Macrolides are naturally occurring compounds comprised of a lactone ring with deoxy sugars attached. Certain macrolides have antibiotic or antifungal properties and are used in pharmaceutical antimicrobial therapy. The first macrolide used in this manner was erythromycin (Figure 2.4, **3**); it was often used for infections in patients who were allergic to penicillin or whose infections were penicillin-resistant (P. H. Patel & Hashmi, 2022). Lincosamides, similar to macrolides, are mainly used to treat gram-positive bacterial infections, which may include β -lactamase-producing gram-positive bacteria, *Staphylococcus spp.*, and *Streptococcus spp.*. In addition to their effectiveness against most gram-positive and anaerobic bacteria, they do not have any activity against enterococci and gram-negative bacteria (Maffioli, 2013). There are three antibiotics in the lincosamides group: lincomycin (Figure 2.4, **4**), pirlimycin and clindamycin (Bal, 2022). Chloramphenicol is a synthetically manufactured broad-spectrum antibiotic (Figure

2.4, 5), initially isolated from the bacteria *Streptomyces venezuelae* in 1948, being the first bulk-produced synthetic antibiotic. Indications for its use include superficial eye infections (bacterial conjunctivitis) and otitis externa. It is also reserved for severe infections, such as rickettsial diseases, meningitis caused by *Haemophilus influenza*, *Neisseria meningitidis*, or *Streptococcus pneumoniae*, or typhoid fever caused by *Salmonella enterica serotype Typhi* (Oong & Tadi, 2022). The final 50S subunit class is oxazolidinone, which is the target class for this thesis. Its chemistry and modes of action will be discussed in further detail below.

2.2.3 Chemistry of oxazolidinone antibiotics and modes of action

Oxazolidinones are a broad-spectrum class of synthetic protein synthesis inhibitors that bind directly to the 50S ribosomal subunit of both gram-positive and gram-negative bacteria. Unlike other 50S inhibitors, which physically block the initiation or translocation of peptide bond formation between amino acids (Kohanski et al., 2010; Stefani et al., 2010), oxazolidinones inhibit protein synthesis at an earlier stage by interfering with the formation of the initiation complex between 50S and 30S subunits (van Bambeke et al., 2017), presenting a unique mode of action. The 5-(S)-configuration of the oxazolidinone ring is essential for activity, which is further improved by its substitution by an N-fluorinated aryl group and a C5 acetamido group (van Bambeke et al., 2017). They can compete for binding to the 50S subunit with other antibiotics (e.g. lincosamides, chloramphenicol) without being antagonistic (van Bambeke et al., 2017). Oxazolidinones mainly prevent the growth of enterococci and staphylococci bacteria (such as *S. aureus*, *Staphylococcus epidermidis*, and *Enterococcus faecalis*) with a short post-antibiotic effect but can also kill certain types of streptococcal species (*S. pneumoniae*, *Streptococcus pyogene*) (van Bambeke et al., 2017; Q. Zhao et al., 2021).

EI DuPont de Nemours & Co. Inc. initially synthesized the first oxazolidinones due to their effectiveness against particular plant pathogens. Two antibacterial oxazolidinones, DuP 721 (*p*-acetylphenyloxooxazolidinylmethylacetamide) and DuP 105 (a methylsulfinyl derivative), were synthesised as antibacterials against human pathogens (Slee et al., 1987, p. 721). The oxazolidinones, a new class of antimicrobial compounds, were initially represented by these clinical candidates. These compounds were derived from an iterative medicinal chemistry project that began with a series of racemic 5-halomethyl-3-phenyl2-oxazolidinones, which had been reported to be effective in treating several plant diseases (Barbachyn & Ford, 2003).

Upjohn Laboratories and Pharmacia, current day Pfizer, continued to study oxazolidinones and developed two nontoxic derivatives of these drugs. These drugs were piperazine derivative U-100592 (*S*-*N*-[[3-[3-fluoro-4-[*N*-1(4-hydroxyacetyl)piperazinyl]phenyl]-2-oxo-5-oxazolidinyl] methyl] acetamide, and a morpholine derivative U-100766 (*S*-*N*-[[3-(3-fluoro-4-morpholinylphenyl) - 2oxo-5-oxazolidinyl] methyl] acetamide, now more commonly known as eperezolid and linezolid respectively (Bozdogan & Appelbaum, 2004; Brickner et al., 1996; Zurenko et al., 1996). Both linezolid and eperezolid demonstrated similar efficacy in preclinical tests as potential drug candidates, with almost identical MIC values, antibacterial spectrum, and pharmacokinetic parameters. Further development of linezolid was chosen due to its prolonged half-life in humans, which emerged after phase I clinical studies for both drugs. A range of oxazolidinone antibiotics are now available or under development and have been tested against various bacterial diseases, as shown in Table 2.2 and illustrated in Figure 2.5.

Table 2.2 - The main representative oxazolidinone antibiotics licensed or under development. Known drug names (and any associated names), features of approval, and current stage of approval are listed. Table 2.2 is adapted from Foti et al. (2021) (*Foti et al., 2021*)

Drug Name	Features	Ref.
Cadazolid (ACT-179811)	Actelion Pharmaceuticals is currently conducting Phase III trials for cadazolid as a treatment for Clostridium difficile infection (CDI). In earlier clinical trials (phases I and II), cadazolid demonstrated safety, good tolerance, and efficacy, indicating its potential as a viable therapeutic option for CDI in the future.	(Endres et al., 2017; Kali et al., 2015)
Contezolid (MRX-1, Youxitai)	Clinical trials in China have completed Phase III, and in June 2021, the National Medical Products Administration of China approved Contezolid for the treatment of complicated skin and soft tissue infections (cSSTI). The drug is currently in clinical development for acute bacterial skin and skin structure infections (ABSSSI) and diabetic foot infections in the USA.	(Foti et al., 2021; Hoy, 2021; Kong & Yang, 2021)

Delpazolid (LCB01-0371)	Clinical trials in Phase I and Phase II are currently ongoing for this drug. Preclinical studies have shown that it is effective against <i>Mycobacterium tuberculosis</i> in <i>in vitro</i> and non-clinical <i>in vivo</i> models.	(Foti et al., 2021; Jeong et al., 2010; Kim et al., n.d.; Kong & Yang, 2021)
Eperezolid	It was previously developed by Pharmacia Corporation for the treatment of various bacterial infections, including MRSA and VRE. However, the development of eperezolid has been discontinued, and it is not currently available on the market.	(Brickner, 2007)
Linezolid (LNZ) (Zyvox)	The first oxazolidinone antibiotic was approved for commercial use in 2000. Used for gram-positive infections and approved for the treatment of bacterial pneumonia, skin and skin structure infections, and VRE infections, including infections complicated by bacteremia	(Azzouz & Preuss, 2023; J. J. Li & Corey, 2013)
Posizolid (AZD2563/ AZD5847) AstraZeneca	Phase II clinical trials were discontinued. The results are not conclusive since the studies for the treatment of TB were discontinued.	(Foti et al., 2021; Tetali et al., 2020)
Radezolid (RX-1741)	The drug is undergoing Phase III clinical trials for the treatment of multidrug-resistant infections, including those caused by LNZ-resistant strains. It has been shown to be eleven times more active than linezolid, but its safety profile has not been established yet.	(Foti et al., 2021; Kong & Yang, 2021)
Ranbezolid (RBx7644)	Ranbaxy Research Labs reported in 2003 that ranbezolid completed Phase I clinical trials conducted in the UK.	(Brickner, 2007)
Sutezolid (NU-100480) Pfizer	The completion of Phase II clinical trials indicated that the medication was secure and well-tolerated.	(Foti et al., 2021; Kong & Yang, 2021)
TBI-223	The current phase of clinical trials (NCT03758612) is focused on assessing the safety, tolerability, and pharmacokinetics of the product.	(Foti et al., 2021; Gordon et al., 2022)

T145	It has the potential therapeutic value of inhibiting the growth of <i>E. faecalis</i> , <i>S. aureus</i> , and <i>Mycobacterium tuberculosis</i> with sub µg/ml potencies.	(Kaushik et al., 2016)
Tedizolid (TZD) formerly Torezolid, (SIVEXTRO)	Tedizolid is approved for the treatment of acute bacterial skin and soft tissue infections by the FDA. Oral and intravenous formulations are available.	(Foti et al., 2021; Jubeh et al., 2020; Principi et al., 2020)

As shown in Table 2.2, to date, only two approved oxazolidinones are on the market - linezolid (6) and tedizolid (7) (Figure 2.5). The pharmacophore consists of an oxazolidinone core (A-ring) decorated with a 5-acetamidomethyl side chain (C-5 side and chain) and 3-fluorophenyl ring (B-ring) attached to a morpholino moiety (C-ring). The 3-fluorophenyl ring at the B-ring has been shown to increase the potency of linezolid by favourably interacting with the residues A2486 and C2487 in a heteroaromatic crevice (Barbachyn & Ford, 2003; Ippolito et al., 2008). The morpholino has little effect on the activity of the oxazolidinone compound as it does not significantly interact with the ribosomal subunits, but its inclusion results in a better safety profile (Brickner et al., 1996). Linezolid also has favourable van der Waals interactions with the base of U2539 (Ippolito et al., 2008) and the *S* configuration at the C-5 position of the oxazolidinone (Barbachyn & Ford, 2003). Tedizolid modifies the linezolid scaffold by replacing the C5-side chain with a hydroxymethyl moiety and incorporating a pyridine moiety in place of the morpholino. Tedizolid extends the linezolid scaffold beyond the C-ring region into a D-ring region by including a methyl tetrazole attached to the pyridine in the 2-position. Tedizolid has been reported to have good activity in a range of gram-positive cocci (methicillin-susceptible *S. aureus* (MSSA), MRSA, *Streptococcus pyogenes*, *Streptococcus agalactiae*, *Streptococcus anginosus*, *Enterococcus faecalis*, and VRE and is approximately 2-4 fold more potent than linezolid (K.-H. Chen et al., 2015). The presence of the tetrazole results in a close lone pair interaction between the moiety and U2584. The hydroxyl can form two hydrogen bonds with phosphate oxygen and the ribose 2'-OH of A2503 (Wright et al., 2020). These changes result in a more potent oxazolidinone while retaining a similar binding pose to linezolid.

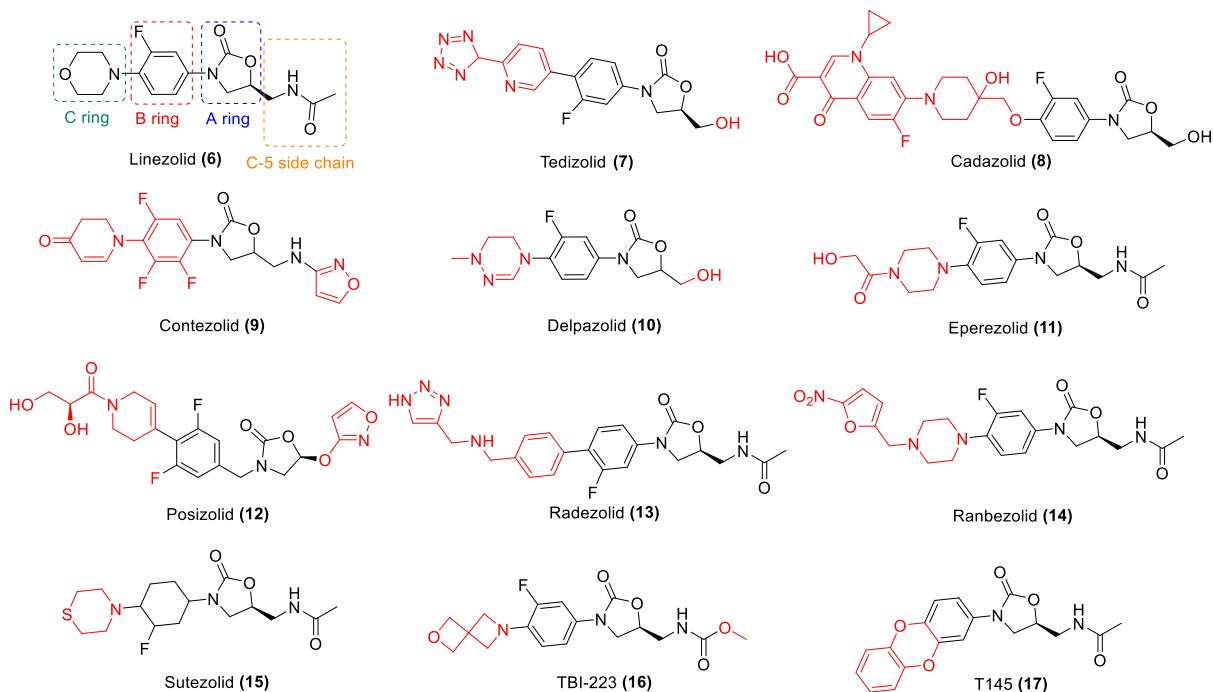


Figure 2.5 - Oxazolidinones currently on the market, linezolid (6) and tedizolid (7), alongside other in-development oxazolidinones Cadazolid (8), Contezolid (9), Delpazolid (10), Eperezolid (11), Posizolid (12), Radezolid (13), Ranbezolid (14), Sutezolid (15), TBI-223 (16), and T145 (17).

There are several structural changes to the oxazolidinone scaffold that have been explored. The modification of the morpholine ring is prevalent in the literature as it is the most tolerant to functionalisation, and vital example of this is ranbezolid, radazolid, sutezolid, cadazolid, and contezolid. There have also been structures exploring the modification to the morpholine ring and C-5 side chain, as shown in tedizolid (7), delpazolid (10), posizolid (12), TBI-223 (16), and T145 (17). Numerous comprehensive reviews have been conducted on the structure-activity relationships of oxazolidinones, all of which discuss the advancements made in oxazolidinone antibacterial agents. These reviews primarily cover oxazolidinone antibacterial drugs currently available in the market, oxazolidinone antibacterial agents that have been modified based on linezolid, or new antibacterial agents that contain oxazolidinone backbones (Q. Zhao et al., 2021; Kong & Yang, 2021; Ford et al., 1997; Chopra, 1998; Genin, 2000; Zhanel et al., 2001; Hutchinson, 2003, 2004; Poce et al., 2008; Michalska et al., 2013). Some of these ligands have been described in ribosomal crystal structures: *H. marismortui*, *D. radiodurans*, *E. coli*, *S. aureus* and MRSA.

Numerous crystal structures have been reported that show linezolid binding to various ribosomes. In the *H. marismortui* structure, the larger A-site pocket observed in the apo

H50S structure is a common trait of eukaryotic ribosomes, which is also utilized by other eukaryotic-specific inhibitors of the PTC (Ippolito et al., 2008). The crystal structure also provided a higher resolution understanding of the structural basis for the mechanism of action of oxazolidinones (Ippolito et al., 2008). Additionally, a linezolid derivative was observed to be bound to the ribosome, confirming that the derivative could take advantage of an energetically favourable π stacking interaction with the ribosome (J. Zhou et al., 2008b, 2008a). In the *D. radiodurans* structure, linezolid was found to stabilise the nucleobase of U2585 in an orientation that is different from when A and P-site tRNA ligands are bound (Stefani et al., 2010; Wilson et al., 2008). In the *S. aureus* structure (Eyal et al., 2015), linezolid was seen to be blocking the A-site in an orientation clearly similar to that observed in other ribosome linezolid complexes with the 50S *D. radiodurans* (Wilson et al., 2008), 50S *H. marismortui* (Ippolito et al., 2008), and 70S *E. coli* linezolid model (Leach et al., 2007). However, in the 50S *S. aureus* complex, the flexible nucleotide U2585 (Bashan et al., 2003) undergoes significant rotation and forms a hydrogen bond with the O4 of the linezolid morpholino ring, leading to a non-productive conformation of the PTC (Eyal et al., 2015). Finally, for the MRSA structures (Belousoff et al., 2019; Wright et al., 2020), a significant conformational change of the rRNA residues U2506 and G2505 was observed in the models. This, in turn, remodelled the linezolid-binding pocket. This, in turn, remodelled the linezolid-binding pocket. A comparative analysis of these mutant ribosomes revealed that most of the rRNA residues had a highly similar conformation, which suggests the possibility of an induced-fit binding mode in the peptidyl transferase pocket (Belousoff et al., 2019).

Other observations have been studied in crystal structures for some of the oxazolidinone structures. Cadazolid was observed to bind with its oxazolidinone moiety in a binding pocket close to the PTC in *E. coli*, similar to linezolid, and extended its unique fluoroquinolone moiety towards the A-site of the PTC (Scaiola et al., 2019). Tedizolid, radezolid, T145, and contezolid have been examined within the PTC of the 50S ribosomal subunit from MRSA (Wright et al., 2020). Conteziolid's C-ring contains a benzyl group trifluorinated that fits snugly into the hydrophobic pocket formed by nucleotides A2451 and C2452. Additionally, the isoxazole in the C5-domain creates new connections with G2061 (a π - π interaction) and A2503 (a van der Waals interaction) (Wright et al., 2020).

2.2.4 Linezolid

Linezolid (Figure 2.6) was introduced in 1996 following intensive examinations at Pfizer, going on to be approved by the US Food and Drug Administration in 2000. (Hashemian et al., 2018) Linezolid prevents the synthesis of bacterial protein via binding to 23S rRNA on both the 30S and 50S ribosomal subunits to form a 70S initiation complex, which is different from the mechanisms of most other protein synthesis inhibitors (Q. Zhao et al., 2021). The presence of the morpholino group plays a crucial role in enhancing the drug's pharmacokinetic properties and water solubility. Additionally, the fluorine atom greatly enhances the drug's activity, the N-aryl group is essential for activity, and maintaining the 5-R configuration is necessary to retain the drug's effectiveness (Stevens et al., 2004; Q. Zhao et al., 2021).

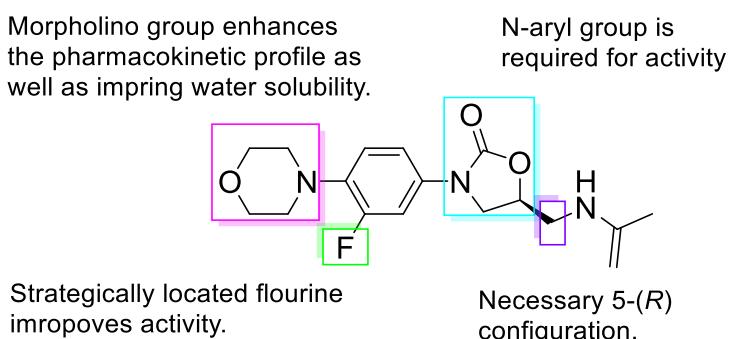


Figure 2.6 - Structure of linezolid, was adapted from Zhao *et al.* (Q. Zhao et al., 2021).

Linezolid demonstrates *in vitro* activity against all the major gram-positive bacteria that are pathogenic in humans, including activity against *Neisseria gonorrhoeae* and *Neisseria meningitidis*, "atypical organisms" such as *Legionella pneumophila*, *Mycoplasma pneumoniae*, and *Chlamydia pneumoniae*, and many Gram-positive anaerobes (Stefani et al., 2010). Linezolid has demonstrated successful treatment activity against many strains of *Mycobacteria*, such as *Mycobacterium abscessus* (Inoue et al., 2017), *Mycobacterium avium* (Peters et al., 1995), *Mycobacterium chelonae* (Brown-Elliott, Wallace, et al., 2001), *Mycobacterium fortuitum*, and *Mycobacterium tuberculosis*. Linezolid also has excellent *in vitro* activity against the *Nocardia* species (Padmanabhan et al., 2005), a weakly staining gram-positive bacteria, including its subspecies of *Nocardia asteroides*, *Nocardia brasiliensis*, *Nocardia farcinica*, *Nocardia nova*, *Nocardia pseudobrasiliensis*, *Nocardia otitidis-cavicularum* and *Nocardia transvalensis* (Brown-Elliott, Ward, et al., 2001).

In 2000 the Food and Drug Administration approved linezolid for the treatment of various infections, such as hospital-acquired and community-acquired pneumonia, complicated skin and skin structure infections (SSSIs), vancomycin-resistant *Enterococcus faecium* (VREF) infections, and *Pneumococcal meningitis* (Hashemian et al., 2018). Hospital-acquired pneumonia remains one of the most lethal and expensive infections in intensive care units worldwide (Kalil et al., 2013). In the United States, one-third of these types of pneumonia are caused by *S. aureus*, including MSSA and MRSA strains, or *S. pneumoniae*, including multidrug-resistant strains (Ament et al., 2002; Klevens et al., 2007). Community-acquired pneumonia is caused by *S. pneumoniae*, including cases with concurrent bacteremia or MSSA (Mc et al., 2003). Linezolid has been used to treat complicated SSSIs, including diabetic foot infections (DFIs) (Lipsky et al., 2004) without concomitant osteomyelitis, are caused by MSSA, MRSA, *S. pyogenes*, or *S. agalactiae* (Ament et al., 2002) and uncomplicated SSSIs caused by MSSA or *S. pyogenes* (Ament et al., 2002). Linezolid has also been used to treat VREF infections, including cases with concurrent bacteremia (Ament et al., 2002), in addition to *P. meningitis* caused by penicillin-resistant *S. pneumoniae* (Faella et al., 2006). Finally, it has been considered for the treatment of patients with endocarditis and bacteriemia (Dailey et al., 2001; Falagas et al., 2006; Muñoz et al., 2007; Stefani et al., 2010), osteomyelitis (Rayner et al., 2004; Senneville et al., 2006), joint infections (Bassetti et al., 2005; Falagas et al., 2007), and tuberculosis (Stefani et al., 2010), pending comparative evaluation for each specific setting. It also inhibits the formation of the initiation complex, which can reduce the length of the developed peptide chains and decrease the rate of translation reaction (Batts, 2000; Hashemian et al., 2018).

2.2.3 Linezolid mutation and resistance

Due to linezolid's entirely synthetic nature, which had allowed an advantage over antibiotics with a natural prototype, it was once claimed that there would be no cross-resistance to linezolid and that resistance would be rare and challenging for the bacteria to develop (Fines & Leclercq, 2000; Long & Vester, 2012). This is true to an extent. It is challenging to induce resistance to linezolid due to its synthetic nature. However, it is possible and has been proven to produce mutations across various residues and bacteria. This is due to the highly conserved PTC. Altering the identity of PTC nucleotides near the antibiotic is unfavourable and standard mechanisms for acquiring resistance are based on altering the conformation and the flexibility of remote nucleotides (Stefani et al., 2010).

Oxazolidinone resistance mechanisms have been classified into three categories: mutations in 23S rRNA genes' domain V region, acquisition of the ribosomal methyltransferase gene cfr, and mutations in genes encoding 50S ribosomal proteins L4 (*rplD*) and L3 (*rplC*) (Stefani et al., 2010). Specific domain V mutations for linezolid and 23S ribosome bacteria will be discussed in the following section.

Mutations that lead to resistance have emerged, primarily attributed to mutations in 23S rRNA, as well as ribosomal proteins L3 and L4. Mutations in 23S rRNA that are associated with resistance to linezolid are typically found in bases near the binding site, including G2061, A2451, C2452, A2503, U2504, G2505, U2506, and U2585 (using *E. coli* numbering). However, mutations can also occur in bases that are further away from the binding site, such as A2062, G2447, A2453, C2499, U2500, and G2576 (Kokkori et al., 2014; Long & Vester, 2012; Stefani et al., 2010). Linezolid resistance can also occur through a less common mechanism that involves mutations in ribosomal proteins L4 and L3, encoded by the *rplD* and *rplC* genes, respectively. The resistance pattern varies between organisms since the mutation sites obtained differ, with some overlap observed among *E. faecalis*, *Enterococcus faecium*, *E. coli*, *H. marismortui*, *Halobacterium halobium*, *Mycobacterium smegmatis*, *Mycobacterium tuberculosis*, *S. aureus*, and *S. pneumoniae* (as shown in Table 2.3) (Long & Vester, 2012; Stefani et al., 2010). Long & Vester's review also provides the *D. radiodurans* position for all 23S RNA mutations (Long & Vester, 2012).

Table 2.3 - 23S rRNA linezolid resistance mutations and related organisms. The table is adapted from Long & Vester (2012) (Long & Vester, 2012)

Area	23s RNA mutation	Organism(s)*	Reference
Within the linezolid binding site	G2061U	Mt	(D et al., 2008)
	C2452U	Hh	(Kloss et al., 1999)
	A2503G	Ms, Sp	(J. Feng et al., 2009; Long et al., 2010)
	A2503U	Ms	(B.-B. Li et al., 2011)
	U2504C	Hh, Sa	(Kloss et al., 1999; Livermore et al., 2007)
	U2504G	Ms	(Long et al., 2010)

	G2505A	Em, Es, Ms	(Bourgeois-Nicolaos et al., 2007; Long et al., 2010; Prystowsky et al., 2001)
Distance from the binding site	A2062C	Hh	(Kloss et al., 1999)
	G2447U	Ec, Ms, Sa, Se	(Bozdogan & Appelbaum, 2004; Livermore et al., 2007; Miller et al., 2008; Sander et al., 2002; Wong et al., 2010; Xiong et al., 2000)
	A2453G	Hh	(Kloss et al., 1999)
	A2453C	Hh	(Kloss et al., 1999)
	C2499U	Hh	(Kloss et al., 1999)
	U2500A	Ec, Sa	(Meka et al., 2004; Miller et al., 2008)
	U2500C	Hh	(Kloss et al., 1999)
	G2576U	Ec, Em, Es, Ms, Mt, Hm [^] Sa, Se, Sh, Sp	(Bozdogan & Appelbaum, 2004; D et al., 2008; J. Feng et al., 2009; Livermore et al., 2007; Long et al., 2010; Marshall et al., 2002; Mazzariol et al., 2012; Miller et al., 2008; Prystowsky et al., 2001; Wong et al., 2010)
*Abbreviations: Ec, <i>E. coli</i>; Em, <i>E. faecium</i>; Es, <i>E. faecalis</i>; Ms, <i>M. smegmatis</i>; Mt, <i>M. tuberculosis</i>; Hh, <i>H. halobium</i>; Hm, <i>H. marismortui</i> Sa, <i>S. aureus</i>; Se, <i>S. epidermidis</i>; Sh, <i>S. haemolyticus</i>; Sp, <i>S. pneumoniae</i>			
^ <i>H. marismortui</i> is listed as G2611U(2576) in original source (Ippolito et al., 2008)			

Although as with most antibiotics, mutations have occurred within various bacterial species. Modifications in 23S rRNA found in linezolid-resistant strains include mutations in *S. aureus* at positions G2576U and G2447U, *E. coli* at positions G2032A and G2447U, and *E. faecium* and *E. faecalis* at G2576U (Bozdogan & Appelbaum, 2004). Studies conducted recently have shown that the binding of linezolid stabilizes the nucleobase U2585 in a way that differs from the orientation seen when A and P-site tRNA ligands are bound, in that linezolid induces a non-productive conformation of PTC (Stefani et al., 2010).

2.3 Oxazolidinone antibiotic molecular studies

The study of oxazolidinones in literature can be split into three distinct areas: Structure-based, Ligand-based, and MD studies. Structure-based molecular studies involve the use of three-dimensional structures of biological macromolecules, such as proteins or nucleic acids, to understand their function, interactions, and mechanisms of action. This can be done using techniques such as X-ray crystallography, NMR spectroscopy, or computational methods such as homology modelling or molecular docking. In structure-based studies, the focus is on the specific structural features of the molecule of interest and how they contribute to its function.

Ligand-based molecular studies, on the other hand, rely on the properties of small molecules that interact with biological macromolecules, such as drugs or ligands. In these studies, the properties of the ligands are analysed to understand their binding interactions and how they affect the behaviour of the macromolecule. Standard techniques in ligand-based studies include QSAR modelling, virtual screening, and comparative molecular field analysis (CoMFA).

MD simulations involve the use of computational methods to simulate the behaviour of molecules over time. MD simulations can be used to study the conformational changes of proteins, the dynamics of protein-ligand interactions, or the behaviour of biomembranes. In MD simulations, the behaviour of the molecule is represented by a set of equations that describe its motion, and the simulation predicts how the molecule will behave under different conditions.

Overall, structure-based studies focus on the structure of the macromolecule, ligand-based studies focus on the properties of ligands that interact with the macromolecule, and MD simulations provide a dynamic view of the behaviour of the macromolecule or the ligand-macromolecule complex. Each approach has its strengths and weaknesses, and choosing the appropriate approach depends on the specific research question and the available resources. In the following sections, there are examples of oxazolidinone studies in the literature.

2.3.1 Structure based methods

In recent years, docking has been applied to targeting DNA oxazolidinones for hit identification in virtual screening studies or predicting the binding pose of a ligand (Deshmukh & Jain, 2018; Fortuna et al., 2014; Kalia et al., 2009; Locke et al., 2010; Orac

et al., 2011). Table 2.4 briefly lists molecular docking studies of various bacterial species with oxazolidinones molecules, such as *S. aureus* (Eyal et al., 2015), MRSA (Belousoff et al., 2019; Wright et al., 2020), *E. coli* (Scaiola et al., 2019; Schuwirth et al., 2005), *H. marismortui* (Ippolito et al., 2008), *Mycobacterium tuberculosis* (X. He et al., 2006) and *D. radiodurans* (Schlüzen et al., 2001; Wilson et al., 2008).

Table 2.4 - List of oxazolidinone molecular docking studies on various bacterial species and strains

Species	Strain	PDB Code	Reference
<i>S. aureus or MRSA</i>	ATCC 6538	4WFA	(Pintilie et al., 2018)
	<i>NCTC 8325</i>	6DDD	(Malik et al., 2023)
		6WRS	(Matsingos et al., 2021)
<i>E. coli</i>	K12	2AW4* or 4V4Q	(Kalia et al., 2009; Panigrahi et al., 2018)
<i>H. marismortui</i>	ATCC 43049	3CPW	(Deshmukh & Jain, 2018; Locke et al., 2010; Matsingos et al., 2021; Romeo et al., 2018)
<i>Mycobacterium tuberculosis</i>	H37Rv	2H7M	(Thomas et al., 2011)
<i>D. radiodurans</i>	ATCC 13939	1K01	(Gandhi et al., 2004)
		3DLL	(Bhattarai et al., 2014; Fortuna et al., 2014; Locke et al., 2010; Malik et al., 2023; L. Yan et al., 2015)
* Entry 2AW4 was removed from the distribution of released PDB entries and is now superseded by 4V4Q			

For example, Orac et al. employed Glide in combination with synthetic and 3D structural methodologies to investigate the binding of 4,5-disubstituted oxazolidinones to T-box riboswitch RNA (Orac et al., 2011). Matsingos and Al-Adhami et al. (2021) utilised the rDock, on a subset of 25 compounds containing various aromatic, heteroaromatic and aliphatic substitutions to the linezolid structure against *S. aureus* (Matsingos et al., 2021). Despite all the ligands occupying the identical binding pocket, their research implied that

the amine groups' presence propelled the ligands towards the more accessible section of the binding site.

Romeo et al. (2018) performed modelling utilising AutoDock and Pyrimidine-1,3-oxazolidin-2-arylimino hybrids against *H. marismortui* (Romeo et al., 2018). The study revealed that certain compounds were strongly bound to the active pocket through a combination of hydrophobic and van der Waals interactions with one or more amino acids in the enzyme's active pocket.

Malik et al. (2023) reported the antibacterial potential of the novel oxazolidinone-sulphonamide/amide conjugates (Malik et al., 2023). Docking studies were conducted on oxazolidinone derivatives, with linezolid as the reference compound, to understand their molecular interactions with the 50S ribosomal subunit and the associated ribosomal peptidyl-transferase protein.

2.3.2 Ligand based methods

QSAR studies provide a valuable tool for drug discovery and chemical design by enabling researchers to predict the biological activity and properties of new compounds before they are synthesised and tested in the lab. QSAR studies can be used to determine a range of things, such as predicting biological activity, optimising lead compounds, identifying potential toxicity, and prioritising compounds for further testing. In recent years, some efforts have been made to understand three-dimensional quantitative structure-activity relationships, 3D QSAR, on oxazolidinone antibacterial agents using CoMFA.

Gandhi et al. (2007) conducted a QSAR based on the HOF and the LUMO energies (Gandhi, 2007). They demonstrated that the calculation of the values of HOF and LUMO energies for compounds containing substitutions could also help to make predictions of antibacterial activities closer to experimental MIC values. It was observed that the activity of the compounds increases when the energy of the LUMO is lower and the HOF is higher. Thus, compounds that have lower HOF become less active and similarly, compounds having low LUMO energies are the most active, and the activity decreases as the LUMO energy increases.

Lohray et al. (2006) conducted a 3D QSAR study for N-4-arylacryloylpiperazin-1-yl-phenyl-oxazolidinones using TSAR 3.3 (Lohray et al., 2006). The authors of the study used a model to calculate the HOF and LUMO values for phenyloxazolidinones with various substitutions, such as CF_3 , COCF_3 , SO_2CH_3 , and SOCHF_2 . According to the model,

compounds with these substitutions have a low HOF (large negative value) and high LUMO (high positive value), leading to poor predicted antibacterial activity based on log (1/C). In contrast, substitutions such as CN or NO₂ groups on the furan or benzene rings result in a high HOF (low positive value) and low LUMO (large negative value), indicating good predicted antibacterial activity. The study found a strong correlation between the in vitro activities (MICs) of the compounds against *S. aureus* ATCC 25923 and the predictions made by the model.

Ding et al. (2019) conducted a study on oxazolidinone-based UDP-3-O-acyl-N-acetylglucosamine deacetylase (LpxC) inhibitors as Gram-negative antibacterial agents (Ding et al., 2019). They found that oxazolidinone-based compounds exhibited a narrower antibacterial spectrum compared with threonine-based compounds. Based on the results, they concluded that the druggability of the oxazolidinone scaffold might be inferior to the classic threonine scaffold in the design of LpxC inhibitors.

Tokuyama et al. (2001) synthesised 5-thiourea and 5-dithiocarbamate oxazolidinones as a continuation of research on 5-thiocarbonyl oxazolidinone antibacterial agents considering the hydrophobic parameters of the molecule (Tokuyama et al., 2001, p. 5). According to their SAR study, the antibacterial efficacy of 5-thiocarbonyl oxazolidinones was considerably influenced by their lipophilicity, specifically the calculated log P value, and the equilibrium between 5-hydrophilic (or hydrophobic) and benzene ring's hydrophobic (or hydrophilic) substituents.

Pae et al. (1999) performed a CoMFA study, concluding that there was a strong correlation between the MIC against MRSA and their steric, electrostatic factors and lipophilicities (Pae et al., 1999). In the same study mentioned earlier, Matsingos and Al-Adhami et al. (2021) synthesized 25 compounds with different aromatic, heteroaromatic, and aliphatic substitutions to investigate the impact of various substituents on the C5-acylaminomethyl moiety of linezolid. The study revealed that smaller and lipophilic substituents are better tolerated at this position than larger and polar substituents. Additionally, the research demonstrated that even minor modifications to the linezolid structure could have negative effects (Matsingos et al., 2021).

2.3.3 Molecular dynamics (MD)

Similarly, the docking results from a linezolid-based oxazolidinones computational study encouraged Malik et al. (2023) to conduct MD in order to predict the binding stabilities

between ligands and ribosomal peptidyl-transferase (Malik et al., 2023). They used a 50 ns MDS experimentation, which demonstrated deviation and the fluctuation of ribosomal peptidyl transferase in water and its complexes with oxazolidinone.

2.4 Machine Learning and drug discovery studies

Traditionally, researchers have used the traditional approaches suggested above to address the challenges of developing efficient and advanced systems for the targeted delivery of therapeutic agents with maximum efficiency and minimum risks (Lipinski et al., 2019). However, these techniques have since imposed challenges such as inaccuracy and inefficiency, as described previously. There has been a surge in the implementation of novel self-sufficient techniques to eliminate the challenges encountered in traditional approaches, such as ML.

In a study by Stokes et al. (2020), a deep NN was utilized to anticipate antibiotic activity in molecules that differ in structure from conventional antibiotics. One of these compounds, Halicin (shown in Figure 2.7), was observed to be effective against a wide range of bacterial infections in mice (Stokes et al., 2020a). One hundred seven million molecular structures were screened and narrowed down to 8 potential candidates with potent activity against a broad range of pathogens.

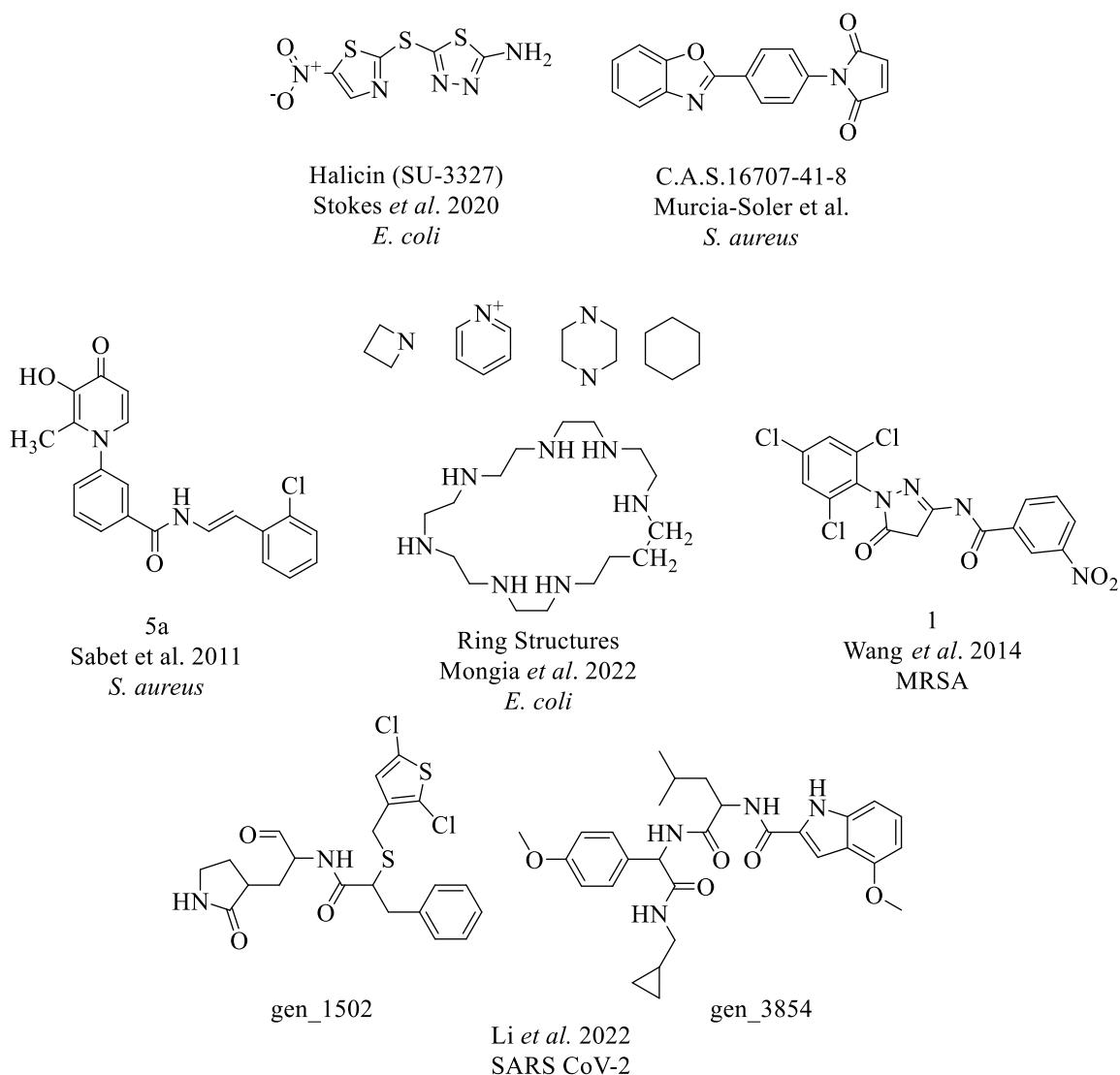


Figure 2.7 - Some examples of antibacterial compounds identified by ML methods.

Murcia-Soler *et al.* (2004) utilised a diverse set of 217 antibiotics spanning multiple classes and targets and 216 additional decoys (Murcia-Soler *et al.*, 2004). A training set of compounds was used to map 62 structure-based molecular descriptors to the biological activity using an artificial neural network (ANN). The top antibiotics predicted by the ANN were then experimentally tested against both Gram-positive and Gram-negative bacteria, namely *E. faecalis*, *S. aureus*, *E. coli*, and *P. aeruginosa*. Four of these antibiotics exhibited low micromolar potency against one or more bacterial strains. Among these, one was found to be more effective against *S. aureus* and *E. faecalis* compared to the two known inhibitors that were included as positive controls, such as cephalosporin C and nalidixic acid.

Sabet *et al.* (2012) utilised a set of thirty-one 3-hydroxypyridine-4-one and 3-hydroxypyran-4-one derivatives, which belong to a class of iron chelators with reported in

vitro antibacterial, antifungal and antimalarial activities (Sabet et al., 2012). An ANN was trained using a bond-based physiochemical molecular map of atom-level properties and molecular descriptors to classify 31 compounds based on their experimentally determined MIC values. This trained network was then employed to prioritize 302 new 3-hydroxypyridine-4-one compounds (Durrant & Amaro, 2015).

Mongia et al. (2022) developed InterPred, an interpretable technique for predicting the bioactivity of small molecules and their mechanism of action, comprised of an ET ensemble classifier and logistic regression model (Mongia et al., 2022). The researchers utilized an FDA-approved library of 800 natural products isolated from plant, animal, and microbial sources that had activity against *E. coli* to train their graph NN model. However, due to the vast number of molecular substructures and the limited training data, the model's interpretability was a challenge. To address this, they focused on simple ring structures and other substructures that are biologically significant. The study identified five essential ring structures, each correlating to a distinct set of modes of action, using InterPred.

Wang et al. (2014) derived in silico models from 5451 cell-based anti-MRSA assay data developed using four ML methods, including naive Bayesian, support vector machine (SVM), recursive partitioning (RP), and k-nearest neighbours (k-NN) (L. Wang et al., 2014). The most effective model was validated by conducting a cell-based assay with three MRSA strains (ST239, ST5, and 252) that are highly resistant to antibiotics. The broth microdilution method was used to confirm the activity of 12 newly discovered anti-MRSA agents, which exhibited MIC values ranging from 4 to 64 mg/L.

Li et al. (2022) used Deep Learning (DL) and transfer learning methods to achieve the *de novo* design of SARS-CoV-2 main protease inhibitors (MproI), which was used in conjunction with molecular modelling and MD simulations (S. Li et al., 2022). Li et al. (2022) started with a potential MproI library (MproIL) 1 containing 6,963 molecules. They used four ML models, including an RF, an SVM, a k-NN classifier, and an XGBoost model, to filter MproIL 1 and obtained PMproIL 2, which had 4,513 molecules that demonstrated activity against both SARS-CoV-2 Mpro and SARS-CoV Mpro. PMproIL 2 was then subjected to molecular docking through AutoDock Vina to filter the compounds further. After several rounds of screening through docking and clustering, the study identified two molecular structures, gen_3854 and gen_1502 (Figure 2.7), as potential inhibitors of SARS-CoV-2 Mpro for further evaluation. However, the authors acknowledged that the

molecular structures designed by MproI-GEN were not diverse and that the ML models used for filtering were trained on SARS-CoV and SARS-CoV-2 inhibitors data due to the limited number of SARS-CoV-2 MproIs.

2.5 Summary

Antibiotics are medicines that fight bacterial infections, categorised as broad or narrow-spectrum. Ribosomes, responsible for protein synthesis, have different subunits, and eukaryotic ribosomes are more complex. Oxazolidinones are synthetic protein synthesis inhibitors that can bind directly to the 50S ribosomal subunit. Linezolid is an oxazolidinone medication that prevents bacterial protein synthesis, approved for several bacterial infections. Linezolid was thought to have no cross-resistance, but mutations have been found leading to resistance.

Various studies have been conducted using different software and methods to understand the molecular interactions of oxazolidinone derivatives with specific receptors. The studies indicate that the compounds have the potential to bind to the receptors through a combination of hydrophobic and van der Waals interactions, with some compounds holding deep into the active pocket. The presence of amine groups can push ligands towards the more open part of the binding site.

QSAR studies predict the biological activity and properties of new compounds, including oxazolidinone antibacterial agents. The studies show that the activity of the compounds can be predicted based on factors such as HOF, LUMO energies, lipophilicity, and various substituents.

The structure-based methods section discusses the limitations of traditional approaches for developing efficient and advanced systems for the targeted delivery of therapeutic agents and the increasing use of self-sufficient techniques such as machine learning to address these challenges. Various studies are presented that have utilised machine learning techniques to identify antibacterial compounds with potent activity against a broad range of pathogens, including deep NNs, ANNs, and transfer learning methods to generate and filter libraries of potential inhibitors.

3

Chapter 3 — Methods

Structure based and ligand based are two different approaches used in computational chemistry and drug discovery. Molecular docking is an example of structure (biological target) based approaches, while methods such as QSAR and ML are ligand based modelling. Each method is discussed in greater detail below.

3.1 Molecular docking

Molecular docking or ‘docking’ is a computational approach aimed to predict the binding poses of a small molecule ligand and receptor such as protein, DNA, or RNA, the latter of which is used through this research. The goal of molecular docking software is to understand and predict molecular recognition structurally, find likely binding modes and energetically predict binding affinity. An example of the process is shown in Figure 3.1, using RNA and linezolid.

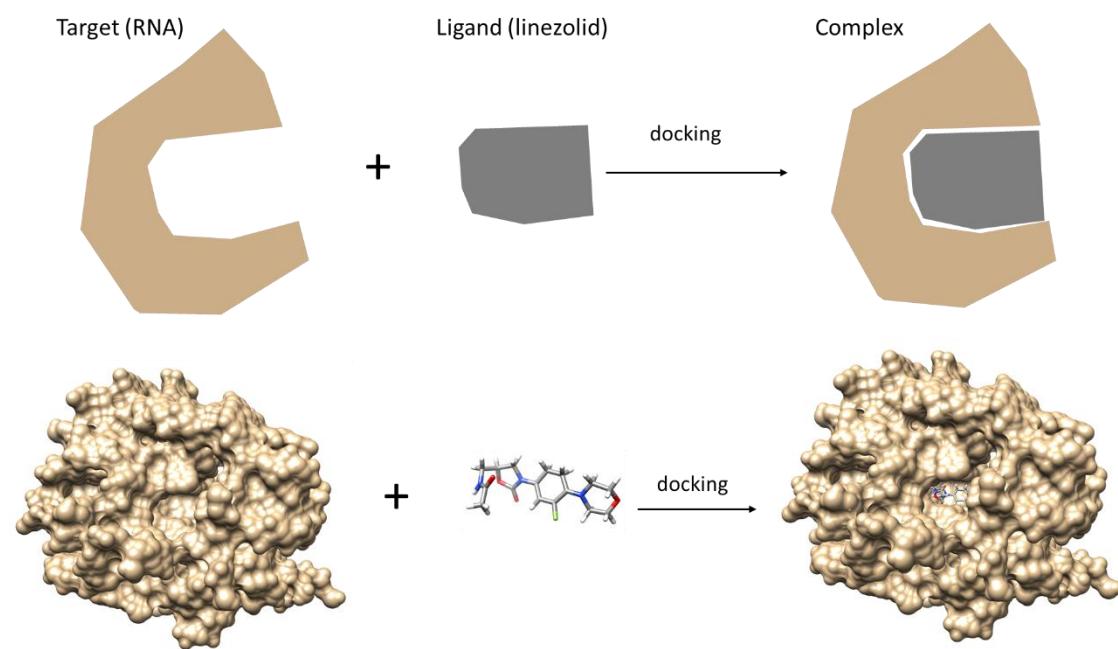


Figure 3.1 - Illustrates docking a small molecule ligand (grey) to an RNA target (brown). Note: the RNA displayed is a selected area of the complete structure, selected 20Å from the ligand binding site.

Traditionally, docking programs, such as AutoDock Vina (Vina) (Trott & Olson, 2010), GOLD (Jones et al., 1995, 1997), and Glide (Friesner et al., 2004; Halgren et al., 2004), were created for small molecule-protein targets. Some programs initially developed for protein-ligand docking have now been optimised for NA targets, such as AutoDock (Morris et al., 2009), DOCK 6 (Lang et al., 2009, p. 6), and FITTED (Corbeil et al., 2007), while other docking programs were developed explicitly for NA targets, including MORDOR

(Guilbert & James, 2008), NLDock (Y. Feng et al., 2021), RiboDock (Morley & Afshar, 2004), rDock (Ruiz-Carmona et al., 2014), and RLDOCK (L.-Z. Sun et al., 2020). Table 3.1 lists the standard or classical docking programs available for ribosomal targets, which are used in the docking section (Chapter 4) of this study. Docking involves two operations, a search algorithm and a scoring function, which are discussed in detail below.

Table 3.1 - Summary of the target, scoring functions, and search algorithms of the programs tested in the study, adapted from Buckley et al. (Buckley et al., 2023).

Program	Target	Scoring Function	Search Algorithm
AutoDock 4	Protein	Physics-based and Empirical	Lamarckian Genetic algorithm
AutoDock Vina	Protein	Physics-based and Empirical	Monte Carlo and quasi-Newton
DOCK 6	RNA	Physics-based and Force Field	Incremental construction
rDOCK	Protein or RNA	Physics-based and Empirical	Genetic algorithm, Monte Carlo and simplex minimisation
RLDOCK	RNA	Physics-based and Empirical	Multiconformer docking

3.1.1 Search algorithm

Algorithms utilized to explore ligand flexibility can be categorized into three main groups based on their employed methodology: systematic, stochastic, and evolutionary algorithms (Brooijmans & Kuntz, 2003; Guedes et al., 2014; Huang & Zou, 2010).

Systematic algorithms explore all degrees of freedom in a ligand during the search. The methods that employ this approach can be further classified as exhaustive, incremental construction and conformational ensemble (Guedes et al., 2014; Huang & Zou, 2010). Exhaustive searches systematically are one the most straightforward search methods, in which flexible-ligand docking is performed by systematically rotating all possible rotatable bonds of the ligand at a given interval (Roy et al., 2015). Therefore, if a ligand is larger and has multiple rotatable bonds, this increases the complexity of the optimisation problem

significantly (Guedes et al., 2014; Huang & Zou, 2010). Two examples of the exhaustive method are Glide (Friesner et al., 2004; Halgren et al., 2004) and FRED (McGann et al., 2003). In the case of fragment methods, also known as incremental construction, the ligand is initially separated into several rigid fragments or parts, and the ligand binding conformation is created incrementally. This is achieved either by placing one fragment at a time in the binding site or by docking all the fragments into the binding site and linking them together covalently (Guedes et al., 2014; Huang & Zou, 2010), DOCK (Ewing & Kuntz, 1997), ADAM (Mizutani et al., 1994), and LUDI (Böhm, 1992) are prime examples of fragmentation methods. Conformational ensemble methods represent ligand flexibility by firmly docking an array of pre-existing ligand conformations. The various ligand binding modes obtained from different docking runs are subsequently gathered and prioritized based on their respective binding energy scores (Guedes et al., 2014; Huang & Zou, 2010). Examples of the conformational ensemble methods for docking include DOCK 3.5 (Lorber & Shoichet, 1998), MS-DOCK (Sauton et al., 2008) and PhDOCK (Joseph-McCarthy et al., 2003).

Stochastic search algorithms introduce random changes in the ligand's degrees of freedom, but they do not guarantee convergence to the optimal solution. To enhance their performance, an iterative process can be employed (Huang & Zou, 2010). There are a variety of algorithms in stochastic, such as Monte Carlo, Evolutionary Algorithms (including genetic), Tabu Search and Swarm Optimisation (Guedes et al., 2014; Huang & Zou, 2010; Torres et al., 2019). In a Monte Carlo method, the probability of accepting an unexpected change is calculated by using the following Boltzmann probability function in Equation 3.1:

$$P \sim \exp \left[\frac{-(E_1 - E_0)}{k_B T} \right] \quad 3.1$$

where E_0 and E_1 stand for the energy scores of the ligand before and after the unexpected change, respectively, k_B is the Boltzmann constant, and T is the absolute temperature of the system.

Evolutionary algorithms utilize the concept of the evolutionary process in biological systems to search for the appropriate ligand binding mode, with the most popular docking

programs such as GOLD (Jones et al., 1995, 1997), AutoDock (Morris et al., 2009), and MolDock (Thomsen & Christensen, 2006). Tabu search algorithms use a history-based strategy to make decisions during the search process. The acceptance probability of a proposed change to the ligand's conformation depends on whether the new solution is in an unexplored region of the conformational space or not. If the Root mean square deviation (RMSD) between the current binding conformation and any of the previously recorded solutions is less than a cutoff, the change is rejected. If the proposed change is in an unexplored region, it is accepted. Examples include PRO_LEADS (Baxter et al., 1998) and PSI-DOCK (Pei et al., 2006). Swarm optimisation algorithms simulate swarm intelligence to find the best solution in a given search space. In this method, the ligand's movements through the search space are guided by information about the best positions of its neighbouring particles. Examples of docking programs that use swarm optimisation algorithms include SODOCK (H.-M. Chen et al., 2007), Tribe-PSO (K. Chen et al., 2006), and PLANTS (Huang & Zou, 2010).

3.1.2 Scoring functions

Scoring functions are crucial in predicting the binding affinity and selecting potential ligands for drug discovery and virtual screening. There are several types of scoring functions used in molecular docking, including Force field-based scoring, Empirical scoring, Knowledge-based scoring, and ML-based scoring functions.

The first category of scoring functions is force-field-based. Force fields have been gradually recognised as a powerful tool for modelling biological macromolecules. They often use simple mathematical equations that consider different types of interactions, such as van der Waals interactions, hydrogen bonding, electrostatic interactions, and solvation effects (Liu & Wang, 2015). The general functional form below (Equation 3.3) thus became more widely adopted:

$$\Delta G_{binding} = \Delta E_{vdw} + \Delta E_{electrostatic} + [\Delta E_{H-bond}] + \Delta G_{desolvation} \quad 3.2$$

Early versions of DOCK (Lang et al., 2009; Meng et al., 1992), AutoDock (Goodsell et al., 1996; Morris et al., 2009) employed energy functions based on the AMBER force field (Cornell et al., 1995; Weiner et al., 1984, 1986) as their internal scoring engine (Liu &

Wang, 2015). Other examples of scoring functions include COMBINE (Ortiz et al., 1995), GoldScore (Jones et al., 1997), and MedusaScore (Yin et al., 2008).

The second type of method is called the empirical scoring function. These scoring functions are based on empirical rules derived from experimental data or known target-ligand complexes. They often use simple mathematical equations that consider different types of interactions, such as van der Waals interactions, hydrogen bonding, electrostatic interactions, and solvation effects. A few examples of popular empirical scoring functions are PLP (Verkhivker et al., 1995), ChemScore (Eldridge et al., 1997; Murray et al., 1998), X-Score (R. Wang et al., 2002), and GlideScore (Friesner et al., 2004, 2006). An example of an empirical scoring function was when ChemScore implemented in the GOLD software uses the following formula shown in Equation 3.3 (Verdonk et al., 2003):

$$\text{ChemScore} = S_{H-bond} + S_{metal} + S_{lipophilic} + P_{rotor} + P_{strain} + P_{clash} \quad 3.3 \\ + [P_{covalent} + P_{constraint}]$$

The third category of methods is commonly known as knowledge-based scoring functions. These functions rely on a statistical analysis of established protein-ligand complexes present in structural databases. They utilize the frequencies and distributions of various protein-ligand interactions observed in the database to approximate the binding affinity of a novel ligand (Liu & Wang, 2015). Although differing in technical aspects, these scoring functions follow the same principle. They sum pairwise statistical potentials between protein and ligand, as shown in Equation 3.4 below:

$$A = \sum_i^{\text{ligand}} \sum_j^{\text{target}} \omega_{ij}(r) \quad 3.4$$

The distance-dependent potential between atom pair i-j, i.e., $\omega_{ij}(r)$, is derived from an inverse Boltzmann analysis as:

$$\omega_{ij}(r) = -k_b T \ln[g_{ij}(r)] = k_b T \ln \left[\frac{\rho_{ij}(r)}{\rho_{ij}^*} \right] \quad 3.5$$

Here, $\rho_{ij}(r)$ is the numeric density of atom pair i-j at distance r and ρ_{ij}^* is the numeric density of the same atom pair in a reference state where interatomic interactions are assumed to be zero (Liu & Wang, 2015). Examples of knowledge-based scoring functions include DrugScore (Gohlke et al., 2000), Potential of Mean Force (Muegge, 2000, 2001;

Muegge & Martin, 1999), ITSscore (Huang & Zou, 2006a, 2006b), and KECSA (Zheng & Merz, 2013).

ML-based scoring functions can capture complex, nonlinear relationships between protein-ligand features and binding affinity. These scoring functions are developed using ML algorithms, such as linear regression, SVMs, or NNs, trained on a dataset of experimentally determined protein-ligand complexes. Examples of ML-based scoring functions include RF-Score (Ballester & Mitchell, 2010), SFCscore^{RF} (Zilian & Sottriffer, 2013), and NNScore (Durrant & McCammon, 2010). ML methods are discussed further below.

It is important to note that the performance and accuracy of scoring functions can vary depending on the specific docking program, the dataset used for training, and the nature of the protein-ligand system being studied. It is recommended to carefully validate and optimise scoring functions for specific applications to ensure reliable results.

3.2 Quantitative Structural Activity Relationship (QSAR)

QSAR is a computational or mathematical modelling method to reveal relationships between biological activities and the structural properties of chemical compounds (Kwon et al., 2019). QSAR, or quantitative structure-activity relationship, continues to be a reliable method for constructing mathematical models that aim to identify a statistically significant relationship between a chemical structure and a continuous (such as pIC₅₀, pEC₅₀, K_i, and MIC) or categorical (such as active vs inactive or toxic vs nontoxic) biological or toxicological property. This is accomplished using regression and classification techniques, respectively (Cherkasov et al., 2014; Neves et al., 2018).

The results for methods such as 2D QSAR, 3D QSAR, and CoMFA are usually expressed in the form of a mathematical equation that describes the correlation between the calculated molecular descriptors selected for establishing the model and the biological toxicity in the series, using the following Equation 3.6:

$$\text{Biological Activity} = c_1d_1 + c_2d_2 + c_3d_3 + \dots c_id_i + \text{constant} \quad 3.6$$

where c_i are coefficients (the fitted parameters), while d_i are molecular descriptors.

QSAR is generally applied to congeneric series of ligands. Congeneric refers to a group of chemical compounds that have similar or identical structures but may have slight structural changes or modifications to their functional group. QSAR typically uses multiple linear regression or partial least square methods to solve the equations. These models are typically

evaluated and validated through a variety of methods: (1) internal validation, using a portion of the dataset that was not used in the model training process, (2) cross validation, such as splitting into testing and training data, k-fold cross-validation or leave-one-out cross-validation, (3) external validation, by using external datasets that were not used in the original training or model development, and (4) statistical methods, such as the correlation coefficient (R^2) of regression, root-mean-square error (RMSE), and the concordance correlation coefficient (Shayanfar & Shayanfar, 2022).

3.3 Machine Learning (ML)

ML is a component of artificial intelligence, although it endeavours to solve problems based on historical or previous examples (Libbrecht & Noble, 2015). Unlike QSAR models, ML algorithms can be trained on a wide variety of data types, including molecular structures, protein sequences, and biochemical assays. ML algorithms can also handle much larger and more complex datasets than QSAR models, making them more suitable for high-throughput screening applications.

Artificial intelligence (AI), including ML and DL algorithms, emerged as a possible solution that can overcome problems and hurdles in the drug design and discovery process. The primary use of AI is in predicting drug properties, which may reduce the need for clinical trials and live study participants, which would be beneficial from both financial and ethical standpoints by significantly reducing cost and time to market (V. Patel & Shah, 2022). Applications of novel approaches of AI in drugs discovery can range from drug design (i.e. prediction of the 3D structure of the target protein, predicting drug-target interaction, and de-novo drug design) and drug screening (prediction of physicochemical properties, prediction of bioactivity and prediction of toxicity) (V. Patel & Shah, 2022). A wide range of tasks in modelling and cheminformatics have been influenced by ML, such as library design (Segler et al., 2018), bioactivities and toxicity prediction (Mayr et al., 2018), and virtual screening (Carpenter et al., 2018; Hofmarcher et al., 2020; Pereira et al., 2016). A general overview of the process of methods ML dataset collection algorithm selection, splitting of data and evaluation of algorithms is shown in Figure 3.2 from Lulka and Stokes (Lluka & Stokes, 2022).

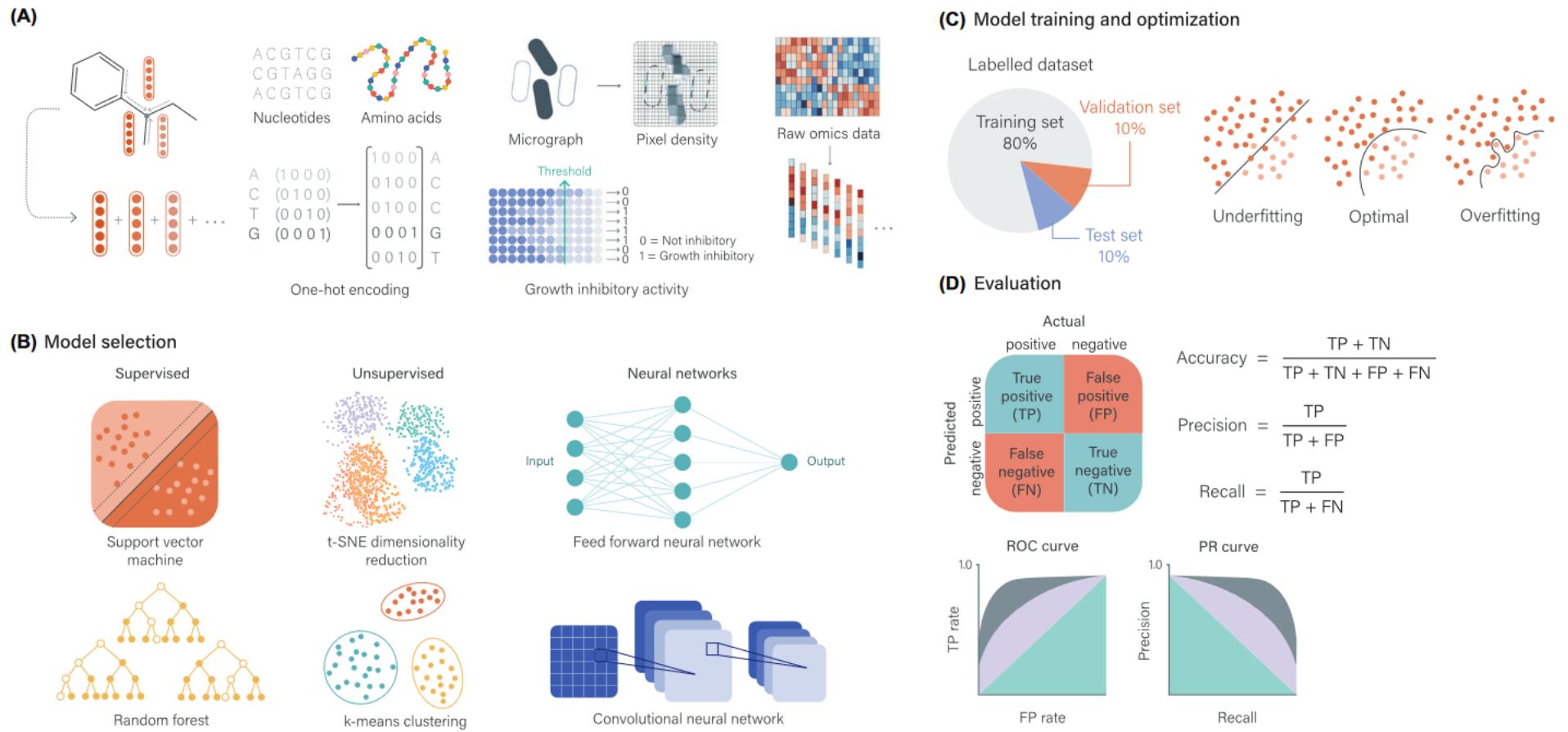


Figure 3.2 - Overview of ML model concepts split into 4 stages. (A) Representations of the variety of data used in ML styles, (B) Potential ML methods that can be used, (C) Splitting of data into training, validation, and testing for said algorithms and (D) evaluation methods for algorithm performance. Image adapted from Lulka and Stokes (2022) (Lluka & Stokes, 2022)

There are four main approaches in ML literature in terms of how the learning process is carried out, hybrid learning, supervised learning, reinforced learning, and unsupervised learning. For this thesis, only supervised and unsupervised learning will be discussed, as these are the only methods used in the research.

Supervised learning can be split into areas - regression or classification, with the former being used in the second part of this thesis. Regression is where the algorithm learns to predict a continuous output variable based on input features. The goal is to learn a mathematical relationship between the input features and the output variable, which can be used to make predictions on new data. Whilst regression problems deal with predicting a numerical score or property value, classification problems deal with categorising into active/inactive or other kinds of category labels. The ML methods differ by the class of models being learnt, and generally, a regression model of any type can be used for classification problems by appropriately transforming the output to emit discrete labels. Broadly ML techniques can be classified into two types: linear and non-linear, depending on whether the output is a linear function of the input or not. Some of the popular linear ML models are linear regression, logistic regression, SVMs, Ridge regression (RR), Lasso regression (LR), and Elastic Net (EN). Whilst the linear models are simple and interpretable, they may not fit the data adequately. Non-linear models such as Decision trees (DT), Random Forest (RF), Gradient Boosting (GB), Neural Networks (NN), k-NN, and SVM's with kernel functions, are complex, offer various degrees of interpretability but often offer better fit than the linear models, making them useful as predictors.

Some prime examples of regression methods are the following algorithms: K Nearest-Neighbors, Linear Regression with different kinds of regularisation such as Ridge, Lasso and EN, DT, Extra Tree, RF, Ada Boost, and GB. Classification is also a type of supervised learning where the algorithm learns to assign input data points to predefined categories or classes. In classification, the goal is to learn a decision boundary that can separate the input data points into different categories.

Unsupervised ML techniques are used to identify patterns in molecular features without any labelled data. These methods are useful when a labelled training set is not available. One such technique is dimensionality reduction, which includes principal component analysis (PCA). Clustering algorithms, such as t-distributed stochastic neighbour embedding (t-SNE) and k-mean clustering, represent another family of unsupervised

algorithms, where the dataset is first divided by predefined distance metrics in high-dimensional space to understand the innate structure and groupings in the data (Lo et al., 2018).

3.3.1 k-nearest neighbour (k-NN)

The k-NN algorithm is among the most commonly employed ML methods due to its low computational costs and predictive accuracy, despite its relative simplicity. This algorithm is predicated on feature similarity, wherein the test sample is categorized based on its nearest neighbours in the training dataset (Z. Zhang, 2016). However, specification of suitable distance metrics may become difficult for complex, e.g., where data lies in non-linear manifold or high-dimensional data where distance measures lose meaning (a.k.a. Curse of dimensionality) (Ivanenkov et al., 2019). Despite that k-NN algorithm is often used as a benchmark for more complex classifiers such as ANNs and SVMs.

3.3.2 Ridge Regression (RR), Lasso Regression (LR), and Elastic Net (EN)

RR is a method that minimizes the residual sum of squares while constraining the L2-norm of the coefficients. By controlling the size of the coefficients, RR balances bias and variance to improve prediction performance. However, RR does not produce a simple model as it includes all predictors (Cohen et al., 2022; Zou & Hastie, 2005).

LR is a type of linear regression that is similar to RR. However, it imposes an L₁ penalty on the regression coefficients as a penalized least squares method. Because of the nature of the L₁ penalty, the lasso regression can simultaneously perform continuous shrinkage and automatic variable selection (Cohen et al., 2022; Zou & Hastie, 2005)

Like Lasso, EN simultaneously does automatic variable selection and continuous shrinkage and can select groups of correlated variables (Zou & Hastie, 2005). It encourages a grouping effect where strongly correlated predictors are selected or excluded from the model together. This feature is particularly useful when dealing with a large number of predictors compared to the number of observations (Zou & Hastie, 2005).

EN combines Ridge and LR, adding a penalty term of the weighted average of L1 and L2 regularisation. A hyperparameter alpha controls the relative weight between the two regularisation terms. EN can strike a balance between Ridge and LR, providing the advantages of both techniques. It can handle highly correlated features, perform feature selection, and produce more stable estimates than LR. RR is best when all features are

relevant, LR is useful for feature selection, and EN can balance the advantages of Ridge and LR.

3.3.3 Adaboost and Gradient Boosting (GB)

Boosting is an ML technique that involves creating an accurate prediction rule by combining multiple weak and inaccurate rules. The AdaBoost algorithm was the first practical implementation of boosting and has since been widely studied and applied in various fields. (Schapire, 2013). It is an iterative algorithm that assigns equal weights to all the training samples. A weak learner is trained on the weighted samples in each iteration, and its error rate is computed. The weights of the misclassified samples are then increased, while the weights of the correctly classified samples are decreased. The process is repeated until a maximum number of iterations is reached or the error rate minimises. The final classifier is a weighted sum of the weak learners, with the weights determined by their error rates.

GB (Friedman, 2001) is one of the most powerful ML methods. GB is an ensemble method that improves the performance of a model by adding new models (DT) sequentially to correct the errors of existing models. GB is resistant to overfitting even when more DT models are added, resulting in better performance. However, increasing the maximum depth of DT models does not always improve prediction quality and can lead to overfitting and increased training time (Ivanenkov et al., 2019).

In summary, Adaboost updates the weights of the samples, while GB updates the weights of the weak learners. Adaboost assigns more weight to the misclassified samples, while GB fits a weak learner to the residuals of the target variable. GB is more flexible than Adaboost, as it can use different loss functions.

3.3.4 Decision Tree (DT), Random Forest (RF), and Extra Trees (ET)

A DT is generated based on the principle of RP, where the feature space is repeatedly split into sections that contain observations with similar response values (Breiman, 2001; Strobl et al., 2009). Each partition is guided by an optimisation criterion to minimise errors in the training data. The resulting tree based model can be used to make predictions for unknown data by following the path from the uppermost node to a terminal node in accordance with the answers devised by chosen feature values at each node (Serafim et al., 2020).

RF (Breiman, 2001) is created from a collection of DT that are trained independently using a random set of data (Ivanenkov et al., 2019; Jukić & Bren, 2022). A combined decision is

obtained either by averaging the response from all the trees (for regression) or taking majority voting (for classification). RF is attractive because it can seamlessly handle different kinds of variables (e.g., continuous, discrete, categorical). The generalization error of a forest of tree classifiers almost surely converges to a limit as the number of trees in the forest increases. The generalization error is affected by the strength of each individual tree in the forest and the correlation between them. A forest with more accurate individual trees and low correlations between them is likely to have better generalization performance (Breiman, 2001).

ET, short for Extra Trees, is an ensemble learning method for classification and regression like RF but with a few key differences. The main idea behind ET is to make the trees in the forest more random and uncorrelated so that the overall error of the forest is lower.

The critical difference between RF and ET is how the tree is built. In ET, the split at each node is chosen randomly from a random subset of the features rather than choosing the best split from all the features, as in RF. This makes the trees in the forest more uncorrelated and less prone to overfitting. Another difference is that in ET, the split points for each feature are chosen at random rather than using the best-split point as in RF. This makes the trees even more random and decorrelated. ET is computationally faster than RF and more random, making it less sensitive to noise and small feature interactions.

3.3.5 Support Vector Machines (SVM)

The SVM (Cortes & Vapnik, 1995) approach consists of generating an optimal hyperplane or decision surface that guarantees the best separation between objects from different classes (Serafim et al., 2020). The key idea behind SVM is to project the data into a very high-dimensional space (even infinite dimension) so that a linear form is enough to model the function accurately. However, instead of actually making the projection and learning a linear function, it uses a ‘kernel trick’ to avoid explicit feature mapping yet perform learning in the high dimensional space. Popular kernels used in ‘kernel trick’ include linear, polynomial, sigmoid, or radial basis kernel functions. The regularisation for learning SVM is achieved either by maximising the margin distance between class separators (for classification) or by maximizing the width of the function (for regression). The hyperplanes that define such margin are called support hyperplanes, and the data points that lie directly on these hyperplanes are called support vectors (Ivanenkov et al., 2019; Serafim et al., 2020).

3.3.6 t-distributed stochastic neighbour embedding (t-SNE)

t-SNE (Maaten & Hinton, 2008) is a statistical method for visualising high-dimensional data by giving each data point a location in a two or three-dimensional map. It is based on Stochastic Neighbor Embedding (SNE) (Hinton & Roweis, 2002), where a *t*-distributed variant is incorporated (Maaten & Hinton, 2008). t-SNE is a type of manifold learning, a nonlinear dimensionality reduction technique that embeds high-dimensional data into a low-dimensional space of two or three dimensions for visualization. It models each high-dimensional object with a two- or three-dimensional point, such that nearby points represent similar objects and distant points represent dissimilar objects with high probability. An example of a t-SNE graph is displayed in Figure 3.3 below.

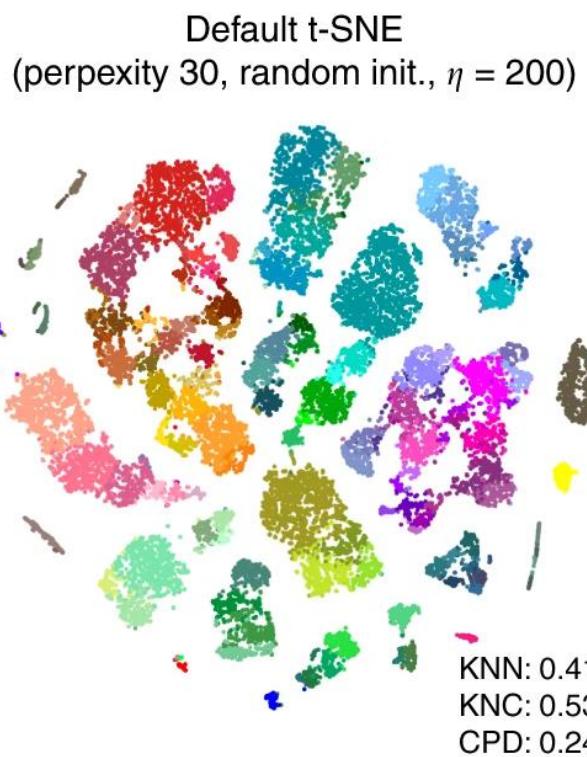


Figure 3.3 – Example of T-SNE from the Tasic et al. transcriptomic dataset (*Tasic et al., 2018*), encompassing 23,822 cells from adult mouse cortex's. Image was adapted from Kobak and Berens (2019) (*Kobak & Berens, 2019*).

The t-SNE algorithm has two primary stages. In the first stage, it creates a probability distribution for pairs of high-dimensional objects. Similar objects are assigned higher probabilities, whereas dissimilar ones are assigned lower probabilities (P. Gupta & Sehgal, 2021). Following this, t-SNE defines a similar probability distribution over the points in

the low-dimensional map, and it minimises the Kullback-Leibler divergence¹ between the two distributions with respect to the locations of the points in the map (P. Gupta & Sehgal, 2021).

3.4 Neural Networks (NN) and Deep Learning (DL)

A NN consists of many connected processors called neurons (inspired by the human nervous system) that produce a sequence of real-valued activations (Schmidhuber, 2015). Input neurons get activated through sensors perceiving the environment, and other neurons get activated through weighted connections from previously active neurons (Schmidhuber, 2015). NNs consist of an input layer of neurons connected to other neurons of one or more hidden layers, which are then connected to an output layer (Serafim et al., 2020).

DL models are NN that can represent complex functions by adding more layers and units within a layer. The term "deep" in deep learning refers to the number of layers in the network, with deep networks containing hundreds of layers compared to classical NN, which has only two or three layers. The development of deep learning was motivated by the limitations of traditional machine learning techniques in processing raw natural data (Serafim et al., 2020). DL reduces the burden on the programmer to select the features explicitly (Figure 3.4).

¹ a measure of how one probability distribution P is different from a second, reference probability distribution Q

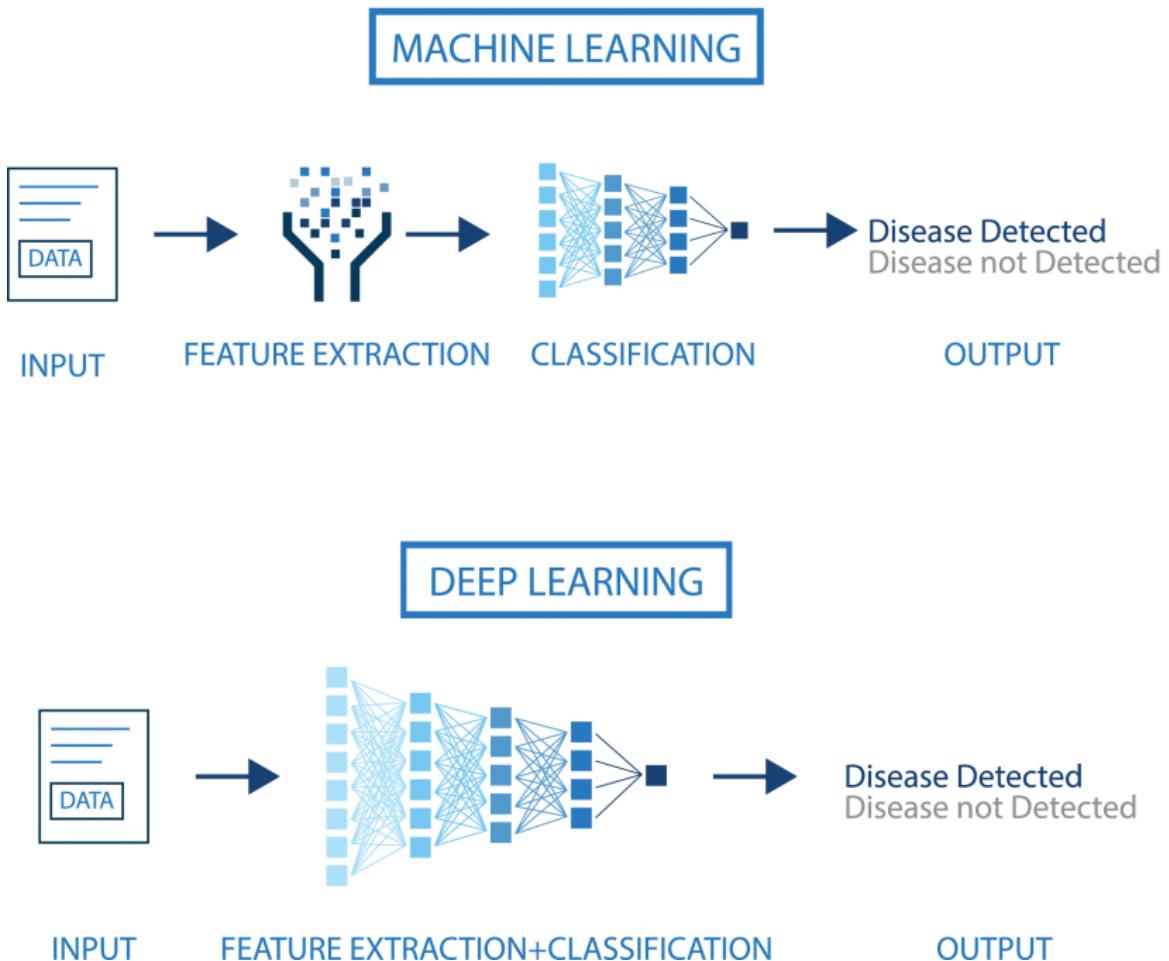


Figure 3.4 - Comparing feature extraction and classification between ML and DL. Image adapted from Khan et al. (2021) (Khan et al., 2021)

So, in essence, NNs are ML algorithms inspired by the human brain. In contrast, DL is a specific application of NNs that uses multiple layers to perform complex tasks. There are a variety of DL algorithms currently available for the drug discovery pipeline, such as AtomNet (Wallach et al., 2015), DeepChem (Ramsundar et al., 2019), TorchDrug (Zhu et al., 2022), and Graph Convolutional Networks (M. Sun et al., 2020). However, DL models tend to be data hungry and often flourishes in the large data regimen.

4

Chapter 4 — Comparative assessment of docking programs for docking and virtual screening of oxazolidinone antibacterial agents

4.1 Introduction

This chapter covered the first aim of my MPhil project (analysis of various molecular docking programs for ribosomal docking and virtual screening of an oxazolidinone dataset) and was published as:

Buckley, M.E.; Ndukwe, A.R.N.; Nair, P.C.; Rana, S.; Fairfull-Smith, K.E.; Gandhi, N.S. Comparative Assessment of Docking Programs for Docking and Virtual Screening of Ribosomal Oxazolidinone Antibacterial Agents. *Antibiotics* **2023**, *12*, 463.
<https://doi.org/10.3390/antibiotics12030463>

Molecular docking is an important tool to predict the poses of ligands dug to potential targets. To present, most of these programs have been solely developed for protein-ligan interactions. However, there have been some developed or adapted to the complexity of RNA-ligand targets. Therefore, in this chapter, I have taken a number of these RNA adaptable docking programs and validated them using a dataset of oxazolidinones. I validate these 5 programs (AutoDock 4, AutoDock Vina, DOCK 6, rDOCK and RLDOCK) against 10 ribosomal crystal structures from a variety of bacteria (*D. radiodurans*, *E. coli*, *H. marismortui*, *S. aureus* and MRSA) that contained oxazolidinones or linezolid derivatives as small molecules. I found that out of all the docking validation, DOCK 6 performed the best. However, overall performance was poor. I further benchmarked the performance of the DOCK 6 docking algorithm and scored in improving virtual screening (VS) enrichment using the dataset of 285 oxazolidinone derivatives against oxazolidinone binding sites in *S. aureus* ribosome. However, I identified that there was no clear trend between the structure and activity of the oxazolidinones in VS. Overall, I found that docking performance indicates that the RNA pocket's high flexibility does not allow for accurate docking prediction, highlighting the need to validate VS protocols for ligand-RNA before future use.

4.2 Statement of Author Contributions



Statement of Contribution of Co-Authors

The following is the suggested format for the required declaration provided at the start of any thesis chapter which includes a co-authored publication, whether published or unpublished. This is a requirement for all Theses by Publication; and for any Theses by Monograph where the relevant published papers are incorporated into the body of the thesis or comprise a chapter within the thesis.

The authors listed below have certified that:

1. they meet the criteria for authorship and that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
2. they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
3. there are no other authors of the publication according to these criteria;
4. potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit, and
5. they agree to the use of the publication in the student's thesis and its publication on the [QUT's ePrints site](#) consistent with any limitations set by publisher requirements.

In the case of this chapter:

Please state the publication title and date of publication or status:

Comparative assessment of docking programs for docking and virtual screening of oxazolidinone antibacterial agents - published MDPI *Antibiotics* 24th Feb 2023

Contributor	Statement of contribution*
McKenna E. Buckley	Conceptualisation, Methodology, Software, Formal analysis, Data curation, Writing — original draft preparation, Writing—review and editing, Visualisation
Audrey R. N. Ndukwue	Conceptualisation, Methodology, Software, Data curation, Writing—review and editing
Pramod C. Nair	Writing—review and editing, Constructive feedback on analysis
Santu Rana	Conceptualisation, Writing—review and editing, Supervision
Kathryn E. Fairfull-Smith	Conceptualisation, Writing—review and editing, Supervision
Neha S. Gandhi	Conceptualisation, Writing—review and editing, Supervision

4.3 Published Paper

Comparative assessment of docking programs for docking and virtual screening of oxazolidinone antibacterial agents

McKenna E. Buckley^{1,3}, Audrey R. N. Ndukwe^{2,3}, Pramod C. Nair^{4,5,6,7}, Santu Rana⁸, Kathryn E. Fairfull-Smith^{2,3} and Neha S. Gandhi^{1,3}

Affiliations

¹ Centre for Genomics and Personalised Health, Queensland University of Technology, Brisbane, QLD, 4059, Australia; (M.E.B.)

² Centre for Materials Science, Queensland University of Technology, Brisbane, QLD, 4000 Australia; (A.R.N.N.); (K.E.F.S.);

³ School of Chemistry and Physics, Queensland University of Technology, Brisbane, QLD, 4000, Australia

⁴ Discipline of Clinical Pharmacology, ⁵ Flinders Health and Medical Research Institute (FHMRI), College of Medicine and Public Health, Flinders University, Adelaide, SA, 5042, Australia; (P.C.N.);

⁶ South Australian Health and Medical Research Institute (SAHMRI), University of Adelaide, Adelaide, SA, 5000, Australia.

⁷ Discipline of Medicine, Adelaide Medical School, The University of Adelaide, Adelaide, SA, 5000, Australia.

⁸ Applied AI Institute (A²I²), Deakin University, Geelong, VIC, 3220, Australia; (S.R.)

* Corresponding author

Abstract: Oxazolidinones are a broad-spectrum class of synthetic antibiotics that bind to the 50S ribosomal subunit of gram-positive and gram-negative bacteria. Many crystal structures of the ribosomes with oxazolidinone ligands have been reported in the literature, facilitating structure-based design using methods like molecular docking. It would be of great interest to know in advance how well docking methods can reproduce the correct ligand binding modes and rank these correctly. We examined the performance of five molecular docking programs (AutoDock 4, AutoDock Vina, DOCK 6, rDock, and RLDock) for their ability to model ribosomal-ligand interactions with oxazolidinones. Eleven ribosomal crystal structures with oxazolidinones as the ligands were docked. The accuracy was evaluated by calculating the docked complexes' root-mean-square deviation (RMSD) and the program's internal scoring function. The rankings for each program based on the median RMSD between the native and predicted were DOCK 6 > AutoDock 4 > Vina > RDOCK >> RLDOCK. Results demonstrate that the top performing DOCK 6 could accurately replicate the ligand binding in only four of the eleven ribosomes due to the poor electron density of said ribosomal structures. In this study, we have further benchmarked the performance of the DOCK 6 docking algorithm and scoring in improving virtual screening (VS) enrichment using the dataset of 285 oxazolidinone derivatives against oxazolidinone binding sites in *S. aureus* ribosome. However, there was no clear trend between the structure and activity of the oxazolidinones in VS. Overall, the docking performance indicates that the RNA pocket's high flexibility does not allow for accurate docking prediction, highlighting the need to validate VS protocols for ligand-RNA before future use. Later, we developed a rescoring method incorporating absolute docking scores and molecular descriptors, and the results indicate that the descriptors greatly improve the correlation of docking scores and pMIC values. Morgan fingerprint analysis was also used, suggesting that DOCK 6 underpredicted molecules with tail modifications with acetamide, n-methylacetamide or n-ethylacetamide and over-predict molecule derivatives with methylamino bits. Alternatively, a ligand-based approach like field template was taken, indicating that each derivative's tail groups have strong positive and negative electrostatic potential contributing to microbial activity. These results indicate that one should perform VS campaigns of ribosomal antibiotics with care and that more comprehensive strategies, including MD simulations and relative free energy calculations, might be necessary in conjunction with VS and docking.

Keywords: molecular docking; oxazolidinones, antibiotics, rRNA, ribosomes, rescoring, DOCK 6, Vina

Introduction

Oxazolidinones are a class of protein synthesis inhibitors that act on the 50S ribosomal subunit across a broad spectrum of gram-positive bacterial strains such as *S. aureus*, *E. faecalis*, and *S. pyogenes* and gram-negative strains such as *E. coli*, *P. aeruginosa*, and *Klebsiella pneumoniae*. Unlike other 50S inhibitors, such as macrolides, lincosamides, and chloramphenicol, which physically block the initiation or translocation of peptide bond formation between amino acids (Kohanski et al., 2010; Stefani et al., 2010), oxazolidinones inhibit protein synthesis at an earlier stage by interfering with the formation of the initiation complex between 50S and 30S subunits (van Bambeke et al., 2017). Linezolid, a synthetic antibiotic based on the oxazolidinone scaffold, binds to the 23S PTC of the 50S ribosomal subunit (Tomasello et al., 2020). It stops bacterial growth and reproduction by preventing mRNA and tRNA's combination with the 50S and 30S ribosomal subunits from forming a 70S initiation complex (Barbachyn & Ford, 2003; van Bambeke et al., 2017; Q. Zhao et al., 2021). Knowing the mechanism of linezolid binding to the PTC is beneficial for developing new oxazolidinone derivatives. This information can indicate the scope for new drug derivatives and suggest sites for other interactions, which can facilitate the prediction of positive and negative interactions. The structures can suggest drug modifications that will allow it to bind to the ribosome despite resistance factors (Long & Vester, 2012). Figure 4.1 demonstrates an example of a detailed three-dimensional ribosomal structure. The 50S ribosome *S. aureus* (PDB 4WFA, strain NCTC832) crystal structure and linezolid binding site, Figures 4.1a display the protein and ribosomal interactions, and Figure 4.1b shows the linezolid binding site and structural interactions with residues G2088, A2478, U2533, and U2612 (*E. coli* numbering).

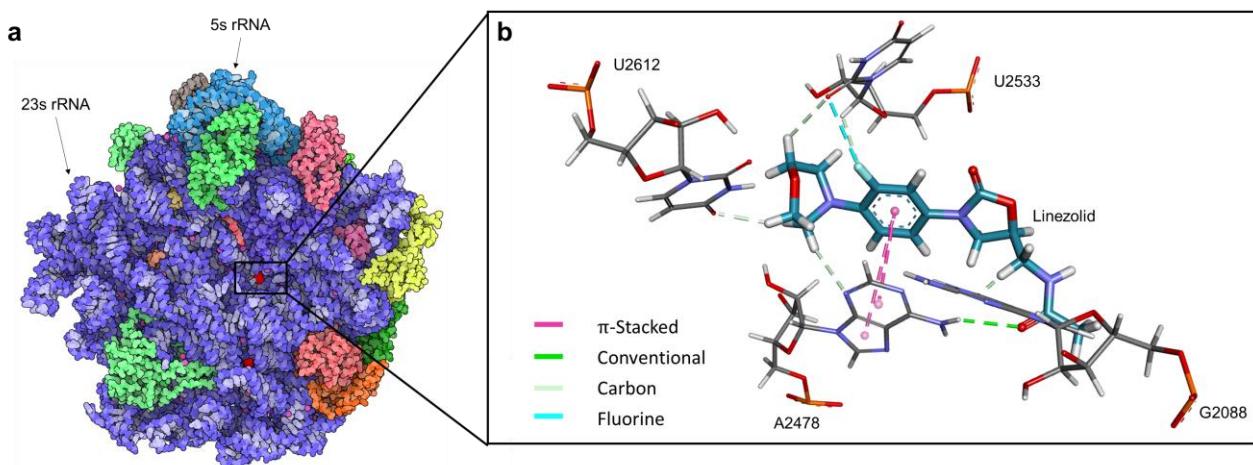


Figure 4.1 - The *S. aureus* (strain NCTC832) 50S ribosome crystal structure (PDB ID 4WFA) with 50S ribosomal proteins (not all proteins are shown for clarity) and 5s and 23s rRNA subunits coloured and labelled with (a) rotated side view of the crystal structure with the Linezolid binding site displayed (b) Close-up view of the linezolid at the ribosomal binding site. Linezolid is shown in aqua blue, while the residues that interact in the active site are coloured by element and labelled with their corresponding residue name, with *E. coli* rRNA numbering. The interactions between the *S. aureus* residues and the ligand in the binding pocket are shown via the striped line, coloured by the following interactions: π-stacked (pink), conventional (bright green), carbon (light green), and fluorine (blue). Images were created using Protein Imager (Tomasello et al., 2020) and BIOVIA Discovery Studio (BIOVIA, 2022).

Many studies have investigated structural modifications to the oxazolidinone scaffold in recent decades. A comprehensive overview of oxazolidinone structure-activity relationships (SAR) is detailed by Zhao et al. (Q. Zhao et al., 2021). This review covers the latest research progress in oxazolidinone antibacterial agents, derivatives with modifications to linezolid and new antibacterial agents containing oxazolidinone backbones (Q. Zhao et al., 2021). Currently, only two approved oxazolidinones exist on the market, linezolid (1) and tedizolid (2) (Figure 4.2). Linezolid consists of an oxazolidinone core (A-ring) with a 5-acetamidomethyl side chain (C-5 side chain) and a 3-fluorophenyl ring (B-ring) attached to a morpholino moiety (C-ring). The key structural features of linezolid critical to its activity are the *N*-aryloxazolidinone moiety and the stereochemistry (*S* configuration) at the C-5 position of the oxazolidinone (Barbachyn & Ford, 2003). The morpholino unit has little effect on the activity of the oxazolidinone compound as it does not significantly interact with the ribosomal subunits, but its inclusion results in a better

safety profile (Brickner et al., 1996). The modification of the morpholine ring is prevalent in the literature as it is most tolerant to functionalisation and a vital example of this is ranbezolid (3). Modification of the C-5 side chain is also common. The replacement of the methyl group in the 5-acetamidomethyl moiety with other small substituents, such as chloromethane (4) and dichloromethane (5), resulted in increased activity (against *S. aureus*, *Staphylococcus capitis*, *S. epidermidis*, and vancomycin-resistant *Enterococcus* (VRE, ATCC700221) strains) making these compounds 2-fold more potent than linezolid (Q. Zhao et al., 2021). While the oxazolidinone core is generally left unchanged, there have been attempts to modify it by developing tricyclic fused oxazolidinones (6) (Q. Zhao et al., 2021). Interestingly, the fluorination of the B-ring in these compounds resulted in reduced activity against *S. aureus* (ATCC 29213), MRSA, methicillin-resistant *S. epidermidis*, penicillin-resistant *Streptococcus pneumoniae*, and *E. faecalis* strains (Xin et al., 2011). In summary, there have been numerous *in vitro* studies investigating the structural modification to oxazolidinones for drug development, and how said modification affect the activity of the ligand. Another approach to expedient these studies is to use computational chemistry methods, such as molecular docking, to predict how well potential drug candidates bind to a target.

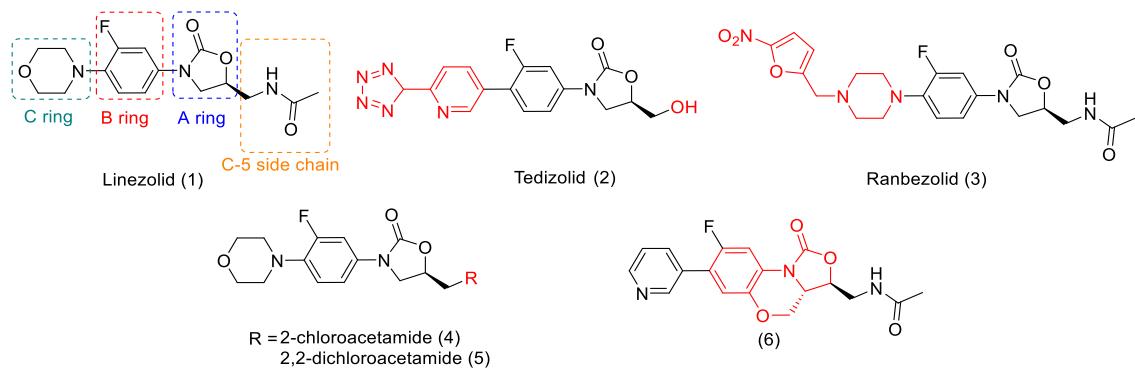


Figure 4.2 - Oxazolidinones on the market are linezolid (1) and tedizolid (2), with the A ring, B ring, C ring and C-5 side chain indicated on linezolid. Examples of modifications to the linezolid scaffold include morpholine modification (3), C-5 side-chain modifications (4) and (5), and oxazolidinone modification (6) from the oxazolidinone dataset. The image was created using ChemDraw.

Small molecule docking or ‘docking’ is a computational approach aimed at predicting the binding poses of a small molecule ligand and biomacromolecules such as protein or nucleic acids (NA) targets such as DNA or RNA. The docking software aims to predict and

understand molecular recognition structurally (by finding likely binding modes) and energetically (by predicting scoring) (Morris & Lim-Wilby, 2008). Traditionally, docking programs, such as AutoDock Vina (Vina) (Trott & Olson, 2010), GOLD (Jones et al., 1997), and Glide (Friesner et al., 2004), were created for protein-ligand targets. Some programs initially developed for protein-ligand docking have now been optimised for NA targets, such as AutoDock4 (AD4) (Morris et al., 2009), DOCK 6 (Lang et al., 2009), and FITTED (Corbeil et al., 2007), while other docking programs were developed explicitly for NA targets, including NLDock (Y. Feng et al., 2021), rDock (Ruiz-Carmona et al., 2014), and RLDOCK (L.-Z. Sun et al., 2020). Scoring functions are mathematical functions used to approximately predict the scoring between two molecules after they have been docked, in addition to removing bias from a docking algorithm's internal scoring function. While developed for protein-ligand interactions, there are various programs explicitly developed for nucleic-acid-ligand interactions that currently exist. Some docking algorithms with internal scoring functions use force field base (DOCK 6 and RLDOCK) or empirical-based scoring functions (AutoDock and rDock), while standard RNA scoring functions are independent of a docking program and utilise knowledge-based functions (ITScore-NL (Y. Feng & Huang, 2020), LigandRNA (Philips et al., 2013), SPA-LN (Z. Yan & Wang, 2017), and DrugScore^{RNA}(Pfeffer & Gohlke, 2007)). In contrast, others utilise ML (ML) scoring functions (AnnapuRNA (Stefaniak & Bujnicki, 2021) and RNAsposers (Chhabra et al., 2020)).

Traditionally, ligand-based approaches like QSAR and pharmacophore modelling have been successfully applied to study structure-activity relationships of oxazolidinone antibacterials against various strains (Gandhi, 2007; Lohray et al., 2006; Pae et al., 1999). In recent years, docking has been applied to targeting -oxazolidinones for hit identification or predicting the interactions and binding pose of a ligand with nucleic acids like DNA and RNA (Deshmukh & Jain, 2018; Fortuna et al., 2014; Jin, Wang, et al., 2022; Kalia et al., 2009; Locke et al., 2010; Orac et al., 2011). For example, Orac *et al.* used Glide combined with synthetic and 3D structural approaches to study 4,5-disubstituted oxazolidinones bound to T-box riboswitch RNA (Orac et al., 2011). To the best of our knowledge, there has been no rigorous benchmarking studying addressing the docking of large oxazolidinone molecule libraries, such as Zhao *et al.*'s review on 285 oxazolidinone scaffolds (Q. Zhao et al., 2021). Thus, our goal in this work was to analyse these derivatives against 23S rRNA molecular structures found in ribosomes by conducting an extensive *in-silico* evaluation.

This evaluation consisted of various computational methods. Molecular re-docking validation was conducted on various docking programs such as AD4 (Morris et al., 2009), Vina (Trott & Olson, 2010), DOCK 6 (Lang et al., 2009, p. 6), rDock (Ruiz-Carmona et al., 2014), and RLDOCK (L.-Z. Sun et al., 2020)) to determine how efficient each program was at reduplicating the poses of a crystal structures native poses, and to select the best performing program to conduct the virtual screening of the dataset. Virtual screening is also performed on of the dataset of derivatives, considering the structural modification and MIC activity analysis against *S. aureus*, to determine whether structural modifications or varying MIC activities have influence on how a derivative performs. Finally, additional methods, such as pharmacophore analysis, re-scoring with external ML scoring functions (AnnapuRNA (Stefaniak & Bujnicki, 2021)), and Morgan fingerprint bit analyses are conducted as supplementary methods, allowing for investigation through ligand based alternatives.

2. Results and Discussion

2.1 Pose prediction using five commonly used RNA docking programs

We examined five docking programs, i.e., AD4, Vina, DOCK 6, rDock, and RLDOCK, to determine how accurately they reproduce the experimental results by self-docking. The scoring functions and search algorithms of the programs used in this study are outlined in Table 4.1.

Table 4.1 - Summary of the target, scoring functions, search algorithms of the programs tested in the study.

Program	Target	Scoring Function	Search Algorithm
AutoDock 4	Protein	Physics-based + Empirical	Lamarckian Genetic algorithm
AutoDock Vina	Protein	Physics-based + Empirical	Monte Carlo and quasi- Newton
DOCK 6	RNA	Physics-based + Force Field	Incremental construction
rDOCK	Protein/RNA	Physics-based + Empirical	Genetic algorithm, Monte Carlo and simplex minimization

RLDOCK	RNA	Physics-based + Empirical	Multiconformer docking
--------	-----	------------------------------	------------------------

Eleven oxazolidinone ligand-based ribosomal crystal structures with resolutions 2.0 Å - 3.5 Å were selected: 3CPW (2.7 Å), 3CXC (3.0 Å), 3DLL (3.5 Å), 4WFA (3.4 Å), 6DDD (3.1 Å), 6DDG (3.1 Å), 6QUL (3.0 Å), 6WQN (2.9 Å), 6WQQ (3.1 Å), 6WRS (3.2 Å), and 6WRU (3.1 Å). The docking performance when predicting the ligand's position compared to the original crystallographic conformations varied across the programs (Figure 4.3a). The performance was ranked based on scoring and root means square deviation (RMSD). Scoring is the strength of the binding interaction between a receptor (protein, RNA, or DNA) and its ligand or small molecule (drug, inhibitor, or derivative). A lower scoring (i.e., a more negative score) indicates a higher performance, displaying better stability of the receptor-ligand complex. RMSD (measured by Å) is an additional measure of redocking success between programs, which gives the average deviation between the corresponding atoms of two targets. The lower the RMSD, the more similar the native and redocked pose. Docking algorithms have their inbuilt scoring functions and can sometimes be biased towards a ligand depending on the size or the structural components. RMSD is typically used as an analysis technique for comparing a redocked ligand to its native ligand. As the virtual screening (VS) of the derivatives utilized a dataset external from the crystal structure, they did not have a native ligand to compare to, so RMSD will not be used for derivatives to compare the whole ligand. The docked poses were considered successful if the ligand RMSD between the docked pose and the structure is less than 2.5 Å (Dickerhoff et al., 2021).

Figure 4.3b and 4.3c demonstrate the average performance of each program across all oxazolidinone based structures for each program and crystal structure, for scoring and RMSD respectively. Figure 4.4a and 4.4b demonstrates the average minimum, medium, maximum of the top-performing values for docking program (the top 5 of each crystal structure averaged) for scoring and RMSD respectively. Separating into overall and top performing was to show an accurate spread of the performance for each docking program, without the separation of specific crystal structures.

Of the eleven crystal structures tested for AD4, the average binding poses were unsuccessful (Figure 4.3c), with none of the crystal structures obtaining a value of less than

2.5 Å, with the overall program performance average at 4.8 Å. Additionally, the overall normalized scoring performance at 0.63, ranking the second lowest of all the docking programs (Figure 4.3b). The overall average experimental binding pose for Vina was only successful in two cases (Figure 4.3c). Like AD4, it had poor scoring performance in comparison to the other programs. The overall normalized average scoring was at 0.63 (Figure 4.3b). DOCK 6 successfully yielded experimental-like poses in four of eleven complexes, displaying the best sampling power for the overall average values for all the docking program (Figure 4.3c). It was the highest performing docking program for overall normalized scoring, being 0.72 on average (Figure 4.3b). rDock had the lowest top-performing pose with no successful poses for any of the eleven complexes (Figure 3c). rDock performed relatively well regarding scoring, with an overall normalised score of 0.69 (Figure 4.3b). Overall, RLDOCK did not successfully yield experimental-like poses in any of the eleven crystal structures (Figure 4.3c). The program had the lowest overall average normalized scoring performance of 0.61 (Figure 4.3b).

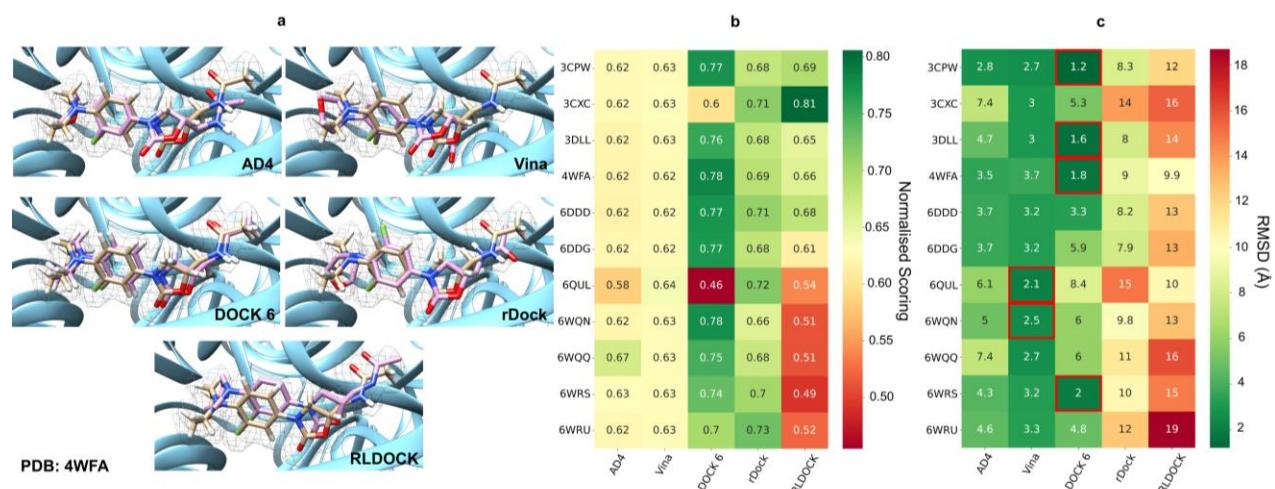


Figure 4.3 - Comparison of the redocking performance for the linezolid (a) Examples of results produced when redocking linezolid to its native crystal structure (PDB: 4WFA). The native ligand pose is beige, the selected redocked ligand is pink, and the 2 Å mesh of the electron density map is grey. (b) The average scoring for each program is displayed in a heatmap using normalized scoring values, showing all linezolid ligand crystal structures (c) The average RMSD for each program is displayed in a heatmap, showing all linezolid ligand crystal structures. The crystal structures that scored at or below 2.5 Å are indicated by red squares. Figure 3a was visualised through UCSF Chimera (v 1.15) (Pettersen et al., 2004).

AD4 scored poorly compared to other programs, with a top-performing median normalized value of 0.63 (Figure 4.4a). AD4 was the second-highest performer for the top-performing median RMSD (Figure 4.4b); however, only 52% of the top-performing values were successful (being below 2.5 Å), being the third best-performing program (Figure 4.4c). Vina had a top-performing median scoring of 0.63 and a top-performing median RMSD of 3.5 Å (Figure 4.4a and 4.4b). Only 37% of the top-performing poses scored under the 2.5 Å success score (Figure 4.4c), with 50% being above 4.5 Å, the highest of all docking programs. DOCK 6 had the highest normalized top-performing median values for both scoring (Figure 4.4a) and RMSD (Figure 4b), of -0.82 and 1.78 Å, respectively. Moreover, 58% of the top-performing scores were under the 2.5 Å success limit (Figure 4.4c). rDOCK had a top-performing normalized median value of 0.74 (Figure 4.4a) and a top-performing median RMSD similar to Vina, 3.8 Å (Figure 4.4b). However, it was the poorest performing in terms of top-performing success rate, with only 10% showing experimental values below 2.5 Å (Figure 4.4c). RLDOCK had the 3rd lowest top-performing normalized score of 0.66. However, of the top-performing values, RLDOCK was successful in almost 85% of ligand binding poses, being by far the best-performing program in terms of success rate (Figure 4.4c). For both scoring and RMSD, RLDOCK had the broadest range of top-performing values (Figures 4.4a and 4.4b), most likely due to its “blind docking” of the ligand. Blind docking refers to the docking of a ligand to the whole surface of a protein or RNA target without defining the ligand’s active site. As RMSD measures the average deviation of the molecule’s structure from the native ligand, most RMSD for RLDOCK scores were high due to redocking poses of the ligand located well outside the linezolid binding site. This failure is due to the vast search space required for blind docking to ribosomal rRNA.

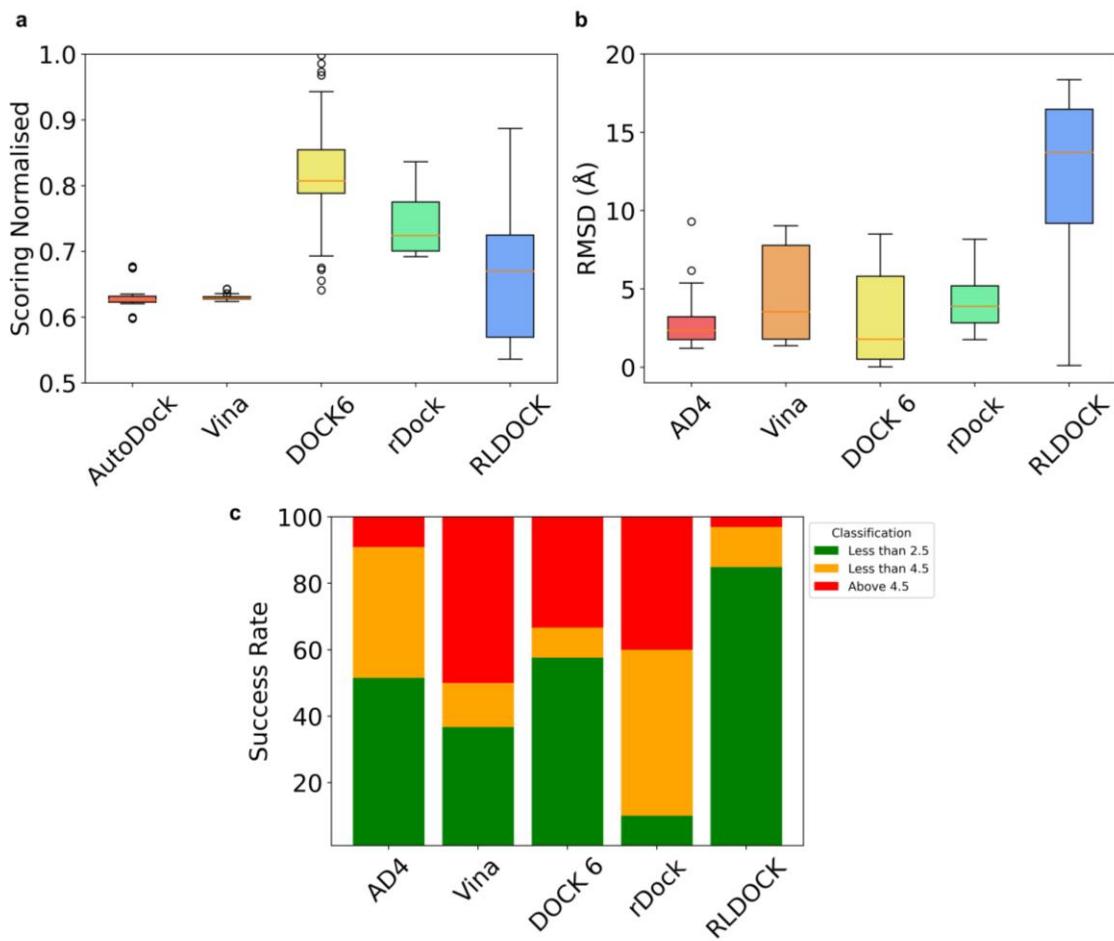


Figure 4.4 - The five selected docking programs used to validate crystal structures, using the top-performing values from each crystal structure. Each docking program was ranked by scoring (a) and RMSD (b). Each program's overall success rate was used to calculate the average, using the normalized top-scoring poses of each program. Both graphs display the minimum and maximum values, with the median values indicated with the orange line and the circles being the outliers. (c) A stacked bar chart demonstrating each selected program's success rate is shown. The green, yellow, and red areas correspond to the percentage of models that achieve the RMSD of ≤ 2.5 , ≤ 4.5 , and $> 4.5 \text{ \AA}$, respectively.

Figure 3 and Figure 4 summarise the re-docking results of eleven small molecule oxazolidinone complexes using five different program programs, for the overall and top performing values respectively. Overall only DOCK 6 and Vina were able to successfully sample at least one structural space of the tested crystal structures with AD4, rDock and RLDOCK. AD4 and Vina consistently had RMSD values over the 2.5 Å cut-off (Figure 4.3c). DOCK 6 was the most consistent program at reduplicating the native pose and having the highest performing scoring. However, while it was the most successful, it still scored

only four of the eleven crystal structures (Figure 4.3c) and only 52% of the top-performing binding poses (Figure 4.4c) were under the experimentally binding value of 2.5 Å. While these docking programs are valuable tools for ribosomal docking, significant improvements are needed to adapt to the high flexibility of the binding pockets. For the above reasons, DOCK 6 was selected to continue the VS section of the study.

2.2 Selection of Ribosomal structure for virtual screening

The availability of X-ray crystal and cryogenic-electron microscopy (cryo-EM) structures of ribosomes has provided an avenue for structure-based drug design (SBDD). SBDD has recently become a suitable and powerful tool for antibacterial drug discovery, mainly due to significant advances in structural biology around the bacterial ribosome (Franceschi & Duffy, 2006). Given the knowledge that minor structural differences between bacterial species can affect drug binding, for the progress of SBDD, it is imperative to have a high-resolution crystal structure of the ribosome and its subunits from the pathogenic bacterial species (Eyal et al., 2015). However, due to the limited number of 50S ribosomes that are bacterial targets, the resolutions ranged from 2.7 Å to 3.5 Å, which is not ideal (Supplementary Figure 4.1). The quality of the structures is poor, as the density is not fully resolved, and does not fully cover the entirety of the ligands. Similar observations for poor electron densities and species variation have been reported for the structures of the erythromycin, azithromycin, telithromycin, and chloramphenicol complexes with ribosomes (Bulkley et al., 2010). However, to our knowledge, there have not large-scale VS studies for a range of ribosome targets. Hence, we aimed to investigate structure-based design despite the low resolutions.

The *S. aureus* crystal structure indicates that linezolid is bound at the PTC, blocking the A-tRNA (aminoacyl binding site, or A-site) in an orientation. This interaction was similarly observed in other ribosome linezolid complexes with 50S *D. radiodurans* (*D. radiodurans*) (Wilson et al., 2008), *H. marismortui* (*H. marismortui*) (Ippolito et al., 2008), or 70S *E. coli* (Leach et al., 2007). However, unlike these other ribosomal complexes, in *S. aureus*, the flexible nucleotide U2585 (Bashan et al., 2003) undergoes significant rotation. It forms a hydrogen bond with the O4 of the linezolid morpholino ring, leading to a non-productive conformation (shape change) of the PTC (Eyal et al., 2015). This difference highlights the interest in the specific interaction of linezolid in *S. aureus*. Of the eleven crystal structures chosen, one was of the bacteria *S. aureus*, and six were the ribosome of

MRSA. While the MRSA structures had slightly higher resolution and better electron density around the ligand than the *S. aureus* structure, mutation or resistant strains of bacteria undergo induced fit mechanisms in the binding pocket (Wilson et al., 2008). Induced fit indicates that the active site changes conformation to allow a better fit between the active site and the ligand. As the residues that the structure interacts with are critical to its efficacy, such as the flexible nucleotide U2585 (Bashan et al., 2003) with the O4 in the morpholino ring, any change to the binding pocket is not preferred. Additionally, most derivatives had literature MIC values against *S. aureus*. Amongst the other crystal structures, for DOCK 6, 4WFA was the third highest-performing structure (Figure 4.3). While 3CPW and 3DLL preceded it, these were crystal structures of *H. marismortui*. Also, 4WFA was the highest-performing crystal structure of *S. aureus* origin (including MRSA structures). For the above reasons, we are investigating the docking of oxazolidinones to the binding pocket of the 50S *S. aureus* crystal structure (4WFA).

2.3 Virtual Screening (VS) of the oxazolidinone dataset

We wanted to determine whether slight modifications in the structures of ligands may influence or impact the docking and scoring of the derivative. For this analysis, the VS and analysis of the derivatives were done against *S. aureus* (PDB:4WFA) using DOCK 6. We grouped each of the 285 derivatives via predetermined specific structure modification groups. The key groupings can be listed as follows: Aryl, Base, Morpholino, Tail, or other oxazolidinones. Descriptions and examples of each modification group are displayed in Figure 5. While Tail was determined as a distinctive modification group, it was always paired with other modifications (such as Aryl-Tail, Morpholino-Tail, or Base-Tail), so we created different groups to address the various forms of structural modifications adequately. We ranked each performance by the top-performing score or Top 1, in addition to the averaged Top 3 and Top 5 scores. Most of the dataset scored above -60 kcal/mol, while the top performers were outliers and performed below -85 kcal/mol (Supplementary Figure 4.2). Only the eleven top-performing derivatives were used for further analysis, displayed in Table 2.

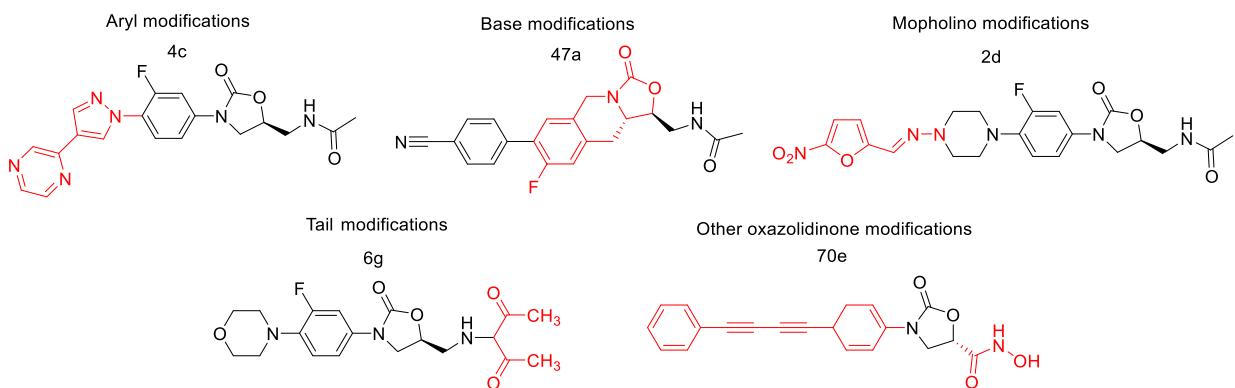


Figure 4.5 - The following structures show the modification to the oxazolidinones derivatives based on the linezolid structure in red. They are split into the following categories: morpholino modifications (2d), aryl modifications (4c), base modifications (47a), tail modifications (6g), and other oxazolidinone modifications (70e). Note that some structures have multiple areas of modification. This image is to detail an example of what these modifications may be.

2.3.1 Structural modification and MIC activity

Figure 6a presents the preliminary analysis of the average top performance when docked with DOCK 6 and separated into structural modification groups. The dataset comprised 32.3% of derivatives with an aryl modification (92 derivatives), 27.0% with base (77 derivatives), 15.4% with morpholino (44 derivatives), and 25.3% with other oxazolidinones (72 derivatives). Figure 6b demonstrates a further percentage breakdown of each leading modification group and the tail groups. Structures with the Base + Tail modification show the highest performance (Figure 6a), followed by Morpholino + Tail, Morpholino, Aryl, Aryl + Tail, Base, and other oxazolidinones. Base + Tail scored an average highest score (Top 1) of -64.1 kcal/mol, while the lowest average classification group, other oxazolidinones, scored 14.4 kcal/mol higher with an average Top 1 score of -49.7 kcal/mol. While there are slight differences between each group of structural modifications, there is no significant correlation between the various derivatives, class of structural modifications and their scoring performance. As some key structural features of linezolid (such as the oxazolidinone core, 3-fluorophenyl ring and morpholino moiety) contribute to its high efficacy as a drug target, it is interesting to observe that many of the derivatives with these core structures did not show good performance.

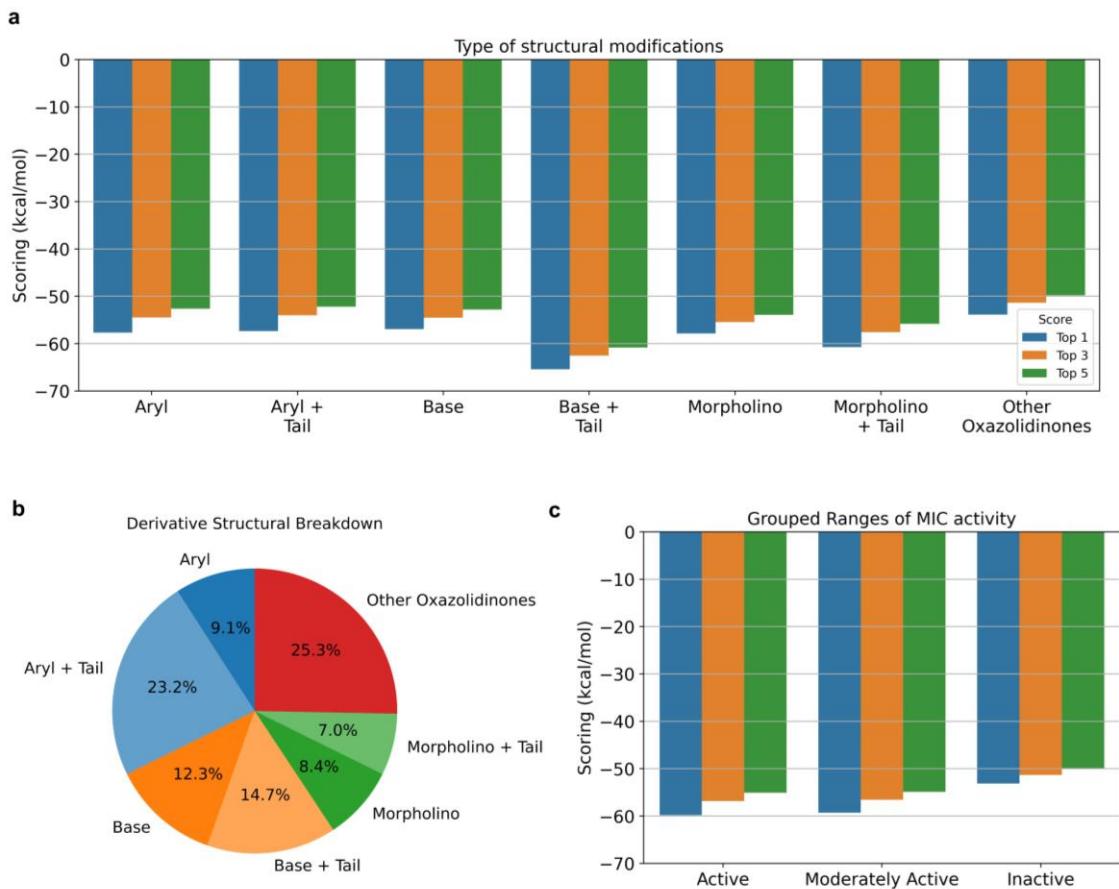


Figure 4.6. - Docking performance of DOCK 6 towards key structural modification of linezolid and correlation with the MIC of against *S. aureus*. **(a)** The average performance of structural modifications when docked with DOCK 6. There are some slight differences between each modification, but nothing of significance **(b)** percentage breakdown of the number of structural modifications. **(c)** Some derivatives had activity against *S. aureus*. Similar to the structural modifications, MIC activities had no significant difference.

We also separated the derivatives by the MIC activity listed in their literature for *S. aureus* (180 out of the total 285 derivatives) to determine whether a change in activity group influenced how well these derivatives performed. The derivatives with MIC values against *S. aureus* were docked against 4WFA, classified as active ($0.01 - <4 \mu\text{g/mL}$) with 99 derivatives, moderately active ($\leq 4 - <16 \mu\text{g/mL}$) with 28 derivatives, or inactive ($> 16 \mu\text{g/mL}$) with 53 derivatives. All values were compiled into their specific group and then averaged to show accurate comparisons (Figure 6c). The active and moderately active derivatives for *S. aureus* show slightly higher performance than the inactive derivatives. While slight differences exist between the activities, these are not significant values to deem any trend or relationship between the *in-vitro* activity and the performance of the *in-*

silico docking. As these derivatives have been reported in the literature, a direct correlation is generally expected between those with lower MIC values and better docking performance. Our current analysis did represent the expected trend, but not substantially. There is a slight difference between active (and moderately active) and inactive; however, this was only an approximate 6 kcal/mol for Top 1. This may suggest inaccuracy in either docking programs/scoring functions or, more likely, due to the high flexibility in the binding pockets of RNA.

The derivatives were separated into groups with or without the O4 position of the morpholino ring (replaced with the R group, Table 2: 1 and 3l). Of the 285 derivatives, 96% had modifications at the O4 position of the morpholino ring. This is an interesting insight, as it has been discussed that in *S. aureus*, the flexible nucleotide U2585 (Bashan et al., 2003) undergoes significant rotation and forms a hydrogen bond with the O4 position of the linezolid morpholino ring, leading to a non-productive conformation of the PTC (Eyal et al., 2015). However, it cannot be solely attributed to the modification of the O4 of the morpholino ring. It has been previously discussed that the poor posing prediction, in consideration of the electron density and the docking program's ability to score with RNA, resulted in poor scoring.

2.3.2 Additional structural analysis

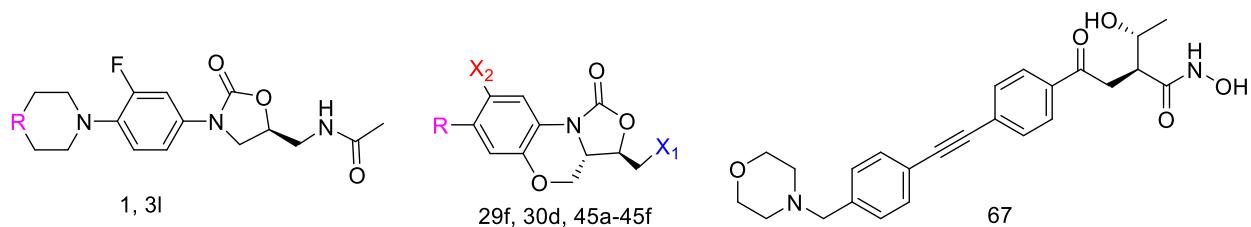
Based on the previous results, it can be assumed that there is no significant relationship between the derivatives and their structural or inhibitory factors besides sharing a similar base structure. However, to confirm our hypothesis, we wanted to do further analysis by comparing other structural factors and their potential linear relationship. Alongside linear regression, we used Pearson's correlation which is the linear relationship between two continuous variables. This statistic ranges from $(-1, 1)$, where coefficients greater than zero represent a positive trend, and coefficients less than zero represent an inverse relationship. Initially, we took the absolute docking score of the derivatives and compared them against the negative log of MIC (pMIC) values of *S. aureus*. As expected, the linear relationship with the derivatives against their respecting pMIC is small for *S. aureus*, with a slight positive correlation of 0.396 for *S. aureus* (Supplementary Figure 4.3a). Using the entire dataset, we further explored structural factors such as the number of rotatable bonds (Supplementary Figure 4.3b), number of atoms (Supplementary Figure c), and molecular weight (Supplementary Figure 4.3d). Supplementary Figure 3b-c reveals a positive

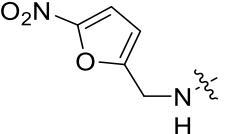
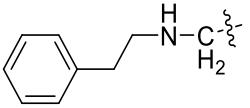
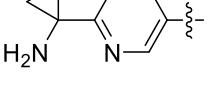
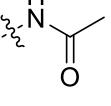
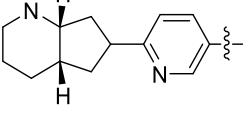
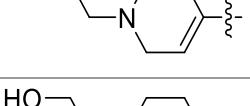
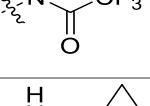
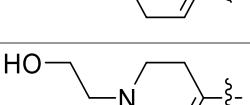
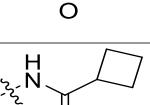
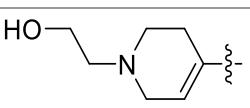
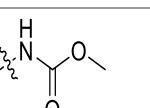
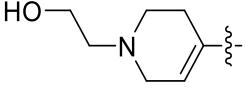
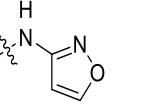
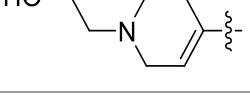
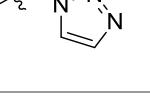
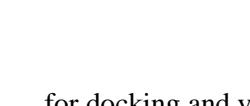
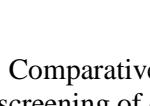
relationship between the absolute values and structural factors. For the number of atoms and molecular weight, removing the derivatives that performed outside the average range (below -80 kcal/mol) improved the correlation (Pearson's correlation coefficient from 0.18 to 0.36 and from 0.18 to 0.41, respectively). A similar trend occurred for the number of rotatable bonds, but to a lesser extent (Pearson's correlation coefficient 0.05 to 0.20). Both numbers of atoms and molecular weight display a medium correlation strength, while rotatable bonds and pMIC all display a weak correlation strength. The number of atoms and molecular weight depends on the size of the molecule, so it was expected that these factors might influence how they perform. These results confirm the limited correlation between derivatives factors and their docking performance. However, these findings may be somewhat limited due to the small sample size of the dataset, so it is suggested to incorporate a more extensive dataset in the future.

2.3.3 Top-performing derivatives and their interactions

We deemed the top-performing values with significant scoring compared to most of the dataset. Out of the 285 derivatives, there were eleven structures determined to be “top-performing” for *S. aureus* (4WFA), those being: **1**, **3l**, **29f**, **30d**, **45a - 45f**, and **67** (Table 4.2). The derivatives are listed with the MIC values reported in their respective literature for both *S. aureus*. The top 3 poses are superimposed against the native linezolid in Supplementary Figure 4.4. Supplementary Figure 4.4 displays that out of the 11 top-performing derivatives, **3l**, **29f**, and **67** were found to dock within the binding site in a similar orientation and residue interactions as linezolid. Like linezolid, which interacts with residues G2088, A2478 U2533 and U2612, derivatives **3l** and **29f** both interact with residues U2533, U2612, and A2478, while derivative **67** interacts with residues G2088 and A2478. Supplementary Figure 4.5 demonstrates each **3l**, **29f**, and **67** residue interaction in the binding pocket, alongside the linezolid.

Table 4.2 - List of derivatives determined to be the “top-performing” derivatives, with their similar base structure indicated. The structural differences are indicated by the R, X1 and X2 columns and the docking score and MIC for *S. aureus*.



Derivative	R Group	X1 Group	X2 Group	Docking Score (kcal/mol)	<i>S. aureus</i> MIC (µg/ml)
1 (Hoellman et al., 2003)		-	-	-94.578	1
3l (Hou et al., 2019)		-	-	-90.462	2
29f (Xin et al., 2011)			H	-96.613	0.25
30d (Guo et al., 2013)			H	-101.091	-
45a (H. Zhao et al., 2017)			F	-85.279	-
45b (H. Zhao et al., 2017)			F	-89.955	1.916
45c (H. Zhao et al., 2017)			F	-93.988	0.473
45d (H. Zhao et al., 2017)			F	-91.024	1.104
45e (H. Zhao et al., 2017)			F	-101.503	2.223
45f (H. Zhao et al., 2017)			F	-99.681	0.98

67 (Barb et al., 2007)	-	-	-	-	-92.438	-
-------------------------------	---	---	---	---	---------	---

2.3.4 Scoring functions

AnnapuRNA is an ML-based scoring function in which the statistical data observed in the experimental RNA-ligand complex is trained to discriminate native structures from decoys (Y. Feng et al., 2022). We re-scored the DOCK 6 scoring function to 35 selected numbers of derivatives in the dataset, with five derivatives from each predetermined structural modification. The scoring function removes any potential bias the DOCK 6 scoring function may have had on the results. As shown in Supplementary Figure 4.6, AnnapuRNA significantly increased the scoring in all structural modifications. These results demonstrate that combining molecular docking and re-scoring using the AnnapuRNA function is a vital scoring function tool and should be utilised further in future studies.

2.3.5 Random Forest to Systematically Classify Bias in Scoring Functions

A typical representation of chemical structures in these models is the molecular fingerprint, such as Morgan fingerprints. Morgan fingerprints work by assigning each atom of a structure with an identifier, updating said identifier based on its neighbour, removing any duplications and then folding this list of identifiers into a 2048-bit vector (Morgan, 1965). This examines the local connectivity of each atom in a structure and creates a unique identifier to represent the atom's local chemical environment (Boittier et al., 2020, p. 38). By doing this, we can perform statistical analyses and ML techniques on the set of molecules to gain new insights that we could not gain as humans.

We trained a RF (RF) model to identify potential problem chemical groups that appear to bias a scoring function to over or underpredict the scoring of a compound. RF is a widely used algorithm designed for large datasets with multiple features, as it simplifies by removing outliers and classifying and designating datasets based on relative features classified for the particular algorithm (L. Patel et al., 2020). The model was trained on the Morgan fingerprints of molecules that produced errors in affinity outside the confidence interval in the regression plot, using molecules with known MIC activity against *S. aureus*. The MIC values were converted to pMIC to reduce the skewing of the data. The models in the data set were classified as over or underpredicted and used in the model's training. Examples of bit-vectors taken from the molecular fingerprints used to train the models

associated with a particular class are shown in Figure 4.7. As expected, it was difficult to ascertain significant differences in bits between the over and under-predicting, as most contributing bits were shared across the two classifications due to the high structural similarity between the derivatives in the dataset. However, DOCK 6 tended to underpredict molecules with tail modifications with acetamide, n-methylacetamide or n-ethylacetamide and over-predict molecule derivatives with methylamino bits. However, due to the size of the dataset and high structural similarities between the derivatives, these are not overly contributing factors, and it is suggested to use them in a larger dataset for future analysis to see if these results persist.

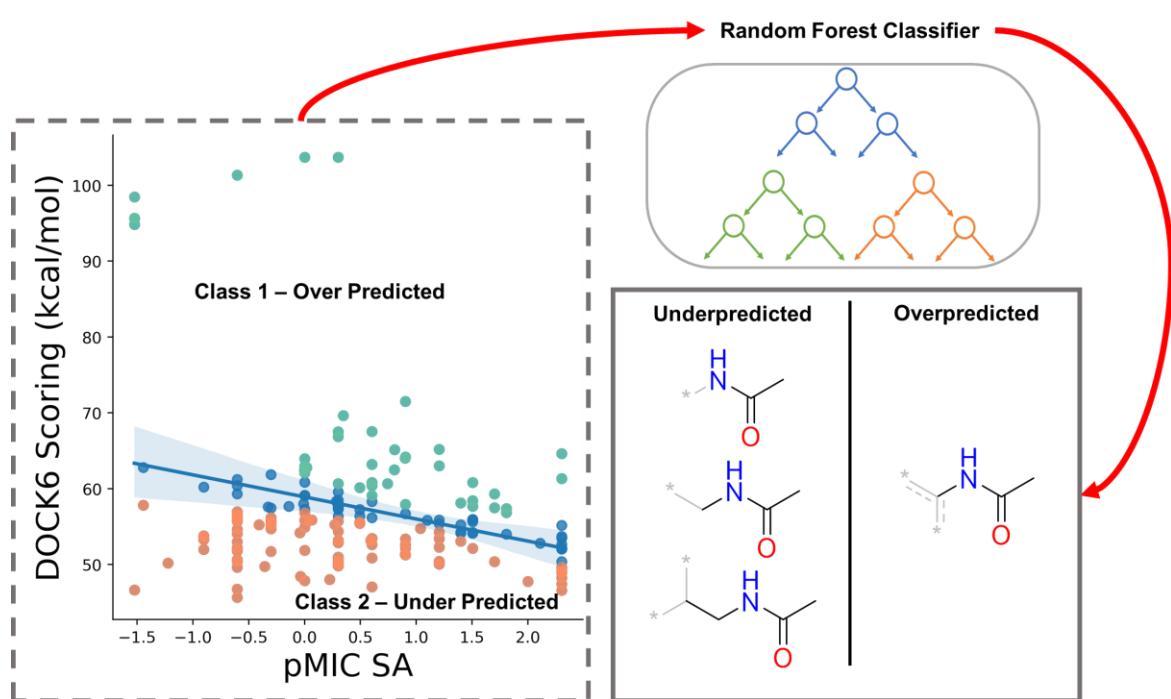


Figure 4.7 - A brief overview of the RF classifier. The classifier is trained on each derivative's Morgan Fingerprint bit vectors to identify sections of the structure that may contribute to over or underpredictions by the scoring function. Over-predicted derivatives are indicated by green and under-predicted by orange. The top-performing derivatives are outliers in the -90 kcal/mol to -100 kcal/mol range.

2.3.6 Tuning the scoring function/Re-scoring function

DOCK 6 employs two physics-based scoring functions, termed Amber score and grid score. The Amber score binding energy ΔG_{bind} is calculated as $E_{\text{Complex}} - (E_{\text{Receptor}} + E_{\text{Ligand}})$, where E_{Complex} , E_{Receptor} , and E_{Ligand} are MM-GB/SA energies as approximated by the Amber force field. Amber score enables all or a part of the ligand-receptor complex to be flexible by

defining a movable region in the DOCK input. Previous study on RNA-ligand complexes suggest that the performance of this scoring approach is dependent on the number of rotatable bonds in a ligand, requires equilibrated structures (Lang et al., 2009) and is computationally expensive for a large scale system like ribosomal rRNA. In this work, we used the grid scoring of DOCK 6 consisting of intermolecular van der Waals (VDW) and Coulombic energies (scaled by a distance-dependent dielectric) between the ligand and receptor. Using the absolute DOCK 6 grid scoring (to allow a positive trendline), the Pearson's correlation between the predicted and experimented pMIC was 0.485 (Figure 4.8a).

We tried to tune the rescoring function using molecular descriptors, to improve the poor docking results discussed above. The Gibbs free energy ΔG_{bind} is dependent on the change in enthalpy and entropy and optimizing these factors can improve affinity of ligand. The molecular descriptors like topological polar surface area (TPSA), number of rotatable bonds (nRotB) can contribute for the enthalpic component, hydrophobic effect and desolvation penalty. The molecular descriptors that considered the overall physical properties of the molecule were LogP, TPSA, nRotB, molecular weight (MolWt), number of H bond donators (HBD), number of H bond acceptors (HBA), LogP, and the number of rings (NumRings). Principal Component Analysis (PCA) was used to determine relevant molecular descriptors and to find a few common factors controlling all variables to study the relationship between the antimicrobial activity and various parameters for the chosen dataset.

The correlation coefficient between the experimental and predicted pMIC values for the test set compounds (30 compounds) were calculated through multiple linear regression.

The formula used for rescoring function for the test set compounds after regression analysis was:

$$\begin{aligned}
 Rescore = & (w1 * \text{Docking Score}) + (w2 * \text{MolWt}) + (w3 * \text{TPSA}) \\
 & + (w4 * \text{nRotB}) + (w5 * \text{HBD}) + (w6 * \text{HBA}) + (w7 * \text{LogP}) \\
 & + (w8 * \text{NumRings}) + c
 \end{aligned} \tag{4.1}$$

Where w1 to w8 = weights obtained after regression analysis on the training set compounds,

Docking score = Absolute DOCK 6 docking scores of the test set compounds,

MolWt = Molecular weight of the test set compounds,

TPSA = Topological Polar Surface Area of the test set compounds,

Nrotb = number of rotatable bonds of the test set compounds,

HBD = Number of H bond donors of the test set compounds,

HBA = Number of H bond acceptors of the test set compounds,

LogP = LogP values of the test set compounds,

NumRing = number of rings of the test set compounds,

c = intercept obtained from regression analysis (-3.02).

This rescore, used to predict pMIC, was a great improvement in correlation of docking scoring and known pMIC, going from a r or Pearson's correlation of 0.485 to 0.779 (Figure 4.8). This improvement in correlation demonstrates that the rescoring function and descriptors can be used as a predictor of pMIC, even when docking results are not ideal. The full list of training set with descriptors and pMIC vs predicted dataset (without the testing set) is shown in Supplementary Table .1 and Supplementary Table 4.2 respectively.

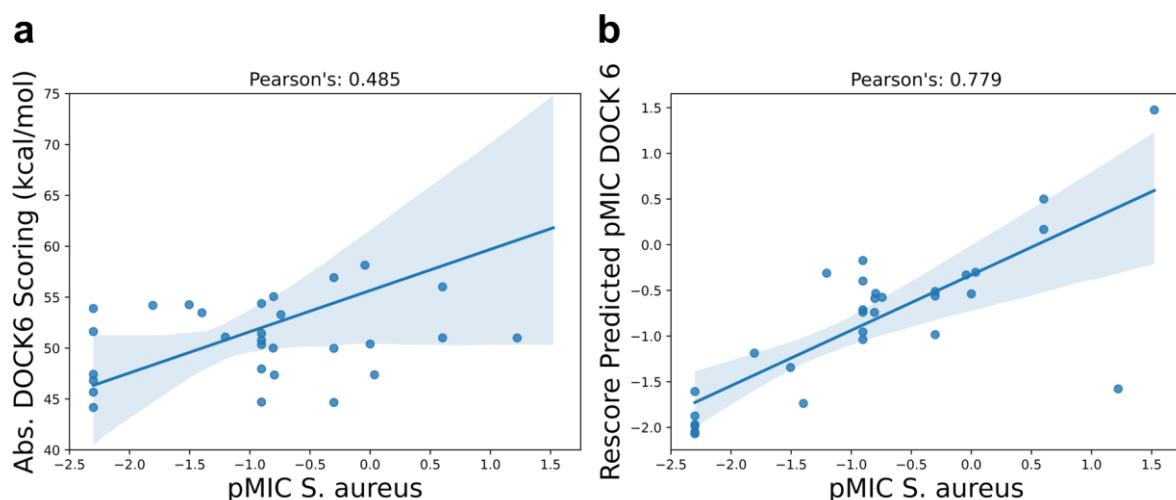


Figure 4.8 - Scatter plots showing coefficient of Pearson's correlation (r) between the experimental pMIC and predicted pMIC for test set compounds (a) Absolute DOCK 6 score (b) DOCK 6 rescore.

2.4 Limitations of study

Computational methods for structure-based docking of small molecules to RNA molecules like the 23S ribosome are not as established as similar methods for protein-ligand docking (Philips et al., 2013). Compared with predicting protein-ligand interaction, modelling bind interaction between RNA-ligand molecules presents some distinctive challenges (Y. Zhou et al., 2022).

rRNA is a large and complex molecule that can adopt a wide range of conformational flexibilities, which can make it challenging to model its binding behaviour accurately and can fold into multiple stable conformations (Luo et al., 2019). Docking programs often overlook the inherent flexibility of nucleic acids. When a ligand binds to RNA, it can induce a conformational change, leading to either an induced-fit effect or a stabilization/destabilization of the tertiary complex structure (Tessaro & Scapozza, 2020). Compared with protein-ligand binding, ligand binding sites on RNA can be shallower, highly polar, solvated, and conformationally flexible, adding further complexity when predicting RNA-ligand interactions.

Many molecular docking studies do not consider the associating effects of water molecules and metal ions. Neglecting such solvent effects can cause a significant impact on water or ion-mediated interactions, leading to inaccurate predictions of RNA-ligand interactions. A solution is to use simulations with explicit waters and ions to improve RNA structures (Moitessier et al., 2006; Park et al., 2011), which allows the results to be sensitive to selecting the critical water molecules. However, achieving complete accuracy can be challenging with this approach, as RNA-ligand interactions are sensitive to the positions and orientations of the water molecules and ions within the cavity space (Belousoff et al., 2019). The prediction of the binding of water molecules and bound metal ions to the RNA before docking, then treating the predicted bound water molecules and ions as part of the receptor for RNA-small molecule docking, is an alternative. Finally, solvation and desolvation effects should also be considered upon ligand binding and water-mediated interactions. While explicit solvation sites within protein-ligand binding pockets can substantially impact affinity and selectivity, similar effects are likely represented in RNA-ligand binding sites as well (Kallert et al., 2022).

Besides these two significant challenges, there are others that limit the performance of RNA docking. The available scoring functions were not trained against rRNA and their ligands; thus, many experimental poses were poorly ranked. There is also a limited number of

experimentally determined RNA structures, so the RNA-ligand complex makes knowledge-based approaches less effective for RNA-ligand predictions (L. Chen et al., 2012; Y. Zhou et al., 2022). Another reason is that rRNA has several functional groups and binding sites that can interact with other molecules. This can make it challenging to identify the specific binding site or sites that a drug candidate would need to bind, a prime example being the overlapping binding sites of chloramphenicol and linezolid. Similar to chloramphenicol, linezolid directly clashes with the placement of the aminoacyl moiety of the aa-tRNA (Marks et al., 2016). Finally, the binding of drugs to rRNA can be affected by pH, temperature, and the presence of other molecules (Bernhardt & Tate, 2012).

Overall, molecular docking can be a valuable tool for predicting the binding of small molecules to protein targets, but it may not be as reliable for predicting the binding of drugs to rRNA. These factors can make it difficult to accurately predict the binding behaviour of a drug candidate using molecular docking. More accurate, ensemble-based free energy scoring methods are routinely used to re-score compounds during VS, such as MD (MD) simulations (Makarov & Makarova, 2020; Makarov & Reshetnikova, 2021; Saini et al., 2013), including Molecular Mechanics with Generalized Born and surface area (MM/GBSA) (Kollman et al., 2000) and Molecular Mechanics Poisson-Boltzmann surface area (MM/PBSA) (Srinivasan et al., 1998) methods. However, it must be noted that these methods are more computationally expensive than docking, usually making them restrictive for extensive-scale compound database screening. Pharmacophore methods are another ligand-based alternative, that uses molecular field points-based similarity method to generate a series of low-energy conformations for each compound. When used in conjunction, docking and pharmacophore can complement each other in revealing critical structural features and could be helpful for the development of highly selective and potency of potential drug molecules, especially when the target is either unknown or poor quality (Supplementary Discussion 4.1, Supplementary Figure 4.7, and Supplementary Methods 4.1). Additionally, several tools like NP Dock (Tuszynska et al., 2015), HNADOCK (J. He et al., 2019) and HDOCK (Y. Yan et al., 2020) have been developed for protein-RNA/DNA docking and RNA/DNA-DNA/RNA docking. However, structure-based VS against RNAs are still rare. Future work will focus on validating these docking tools for its applicability against rRNA-small ligand targets, which is crucial because parameters cannot easily be transferred, and as discussed extensively above, nucleic acids hold their own unique challenges.

In summary, for the reasons explained above, these VS results must be interpreted cautiously and can only guide docking-based virtual screens of oxazolidinone targets.

3. Materials and Methods

3.1 System selection

We obtained available crystal structures from the RCSB Protein Data Bank (PDB), only selecting those which were 50S ribosomal structures that contained oxazolidinones as the ligand. Eleven 50S ribosome PDBs were selected from the following organisms: *D. radiodurans* (3DLL) (Wilson et al., 2008), *E. coli* (6QUL) (Scaiola et al., 2019), *H. marismortui* (3CPW and 3CXC) (Ippolito et al., 2008; J. Zhou et al., 2008a), *S. aureus* (4WFA) (Eyal et al., 2015), and MRSA (6DDD and 6DDG, 6WQN, 6WQQ, 6WRS, and 6WRU) (Belousoff et al., 2019; Wright et al., 2020). To compare the performance of selected docking programs for use with oxazolidinone and linezolid derivates, we performed docking using AD4 (Morris et al., 2009), Vina (Trott & Olson, 2010), DOCK 6 (Lang et al., 2009, p. 6), rDock (Ruiz-Carmona et al., 2014), and RLDOCK (L.-Z. Sun et al., 2020) to the systems described above. We selected programs due to their commercial availability preference and adaptability to RNA and ribosomal targets.

3.2 Pocket location, RNA-ligand preparation, and docking protocols for native ligands

USCF Chimera (Version 1.15) (Pettersen et al., 2004) and Discovery Studio (Version 21.1) (BIOVIA, 2022) were used to visualise the PDB structures and for docking preparations. The proteins, ligands, and additional molecules are removed from the structure so that the RNA structure can be isolated and prepared for docking. The process is repeated, isolating the native ligand instead of the RNA.

The native pocket location for each crystal structure was defined in Chimera, using the Axes/Plane/Centroids method, specifically “Define Centroid”, while the ligand was selected. A 20 X 20 X 20 Å box ($v = 8000 \text{ \AA}^3$) was created around this centroid site to allow the ligand to rotate and flex while not leaving the intended active site. The ligand and RNA files were prepared using Dock prep with hydrogens and charges added in the process, selecting starting residues to be calculated by AMBER ff14SB and other residues AMI-BCC.

3.3 Molecular Docking

3.3.1 Ligand Docking with AutoDock Vina (Version 1.2.0)

Parallelized Open Babel & AutoDock suite Pipeline or POAP (Samdani & Vetrivel, 2018) was used to run Vina and AD4. POAP uses a GNU Parallel to connect the OpenBabel and AutoDock/Vina packages to process ligand preparation, receptor preparation, docking tasks, and result processing via a shell command line. Vina is executed through the POAP pipeline. The exhaustiveness is set to 100, the number of CPUs left at 8, and the number of generated ligand complexes after VS is set to 10.

3.3.2 Ligand Docking with AutoDock4 (Version 4.2.6)

AD4, as described above, was executed through the POAP pipeline. The number of parallel jobs left at the default value of 24 and the number of generated ligand complexes after VS was set to 10.

3.3.3 Ligand Docking with DOCK (Version 6.9)

The molecular surface of the crystal structures was generated using the dot molecular surfaces (DMS) tool from Chimera. The DMS calculation commonly fails with large multi-chain structures such as RNA, so RNA within 20 Å of the ligand was selected to ensure the calculation succeeded. The sampling algorithm, anchor-and-grow, was utilised. The sphgen program generated spheres for the molecular surface, and the grid box and grid were created by Showbox and grid programs, respectively, with a grid spacing of 0.4 Å.

3.3.4 Ligand Docking with rDock

BIOVIA Discovery Studio (BIOVIA, 2022) was used to apply the CHARMM forcefield for all ligands. The reference ligand method in rDOCK was used to define the binding site, and the cavity mapping parameters were optimised for each ligand. The docking results were evaluated using the default scoring function SF3 (without desolvation terms) and SF5 (with desolvation terms). The remaining parameters were left as their default values, and the radius of the large sphere was not defined.

3.3.5 Ligand Docking with RLDOCK

RLDOCK uses global docking for the RNA receptor, so the binding site was not specified. The number of simulation threads was set to 32, and the number of output poses after clustering was set to 100.

3.3.6 Normalisation and rescoring of redocking

As the scoring functions of AD4, Vina, DOCK 6, rDock, and RLDOCK programs use a variety of internal scoring functions (Table 1), resulting in different ranges or scoring, prior to mean calculation and other statistical operations on the docking scores. Data normalization was performed was performed using the formula below to bring all the scores in a notionally common scale from 0 to 1:

$$\text{Normalised score} = 1 - \frac{(x - \min)}{(\max - \min)} \quad 4.2$$

Where x = corresponding score, max = maximum score and min = minimum score of the dataset

3.4 Virtual Screening (VS) with oxazolidinone derivatives

Virtual screening was performed with a dataset comprising 285 oxazolidinone and linezolid derivatives, compiled from multiple studies in Zhao et al.'s review on oxazolidinone scaffolds (Q. Zhao et al., 2021). The derivative structures were taken from the review paper, manually drawn, and placed in a spreadsheet using ChemDraw. Then these were converted into the simplified molecular-input line-entry system (SMILES) with their MIC against *S. aureus* noted, retaining their chirality and stereochemistry. The MIC values were also converted to pMIC for additional analysis using the equation below (See Figure 4.7 and Supplementary Figure 4.3).

$$\text{pMIC} = -\log_{10}[\text{MIC}] \quad 4.3$$

Ligands for virtual screening were prepared as stated above in DOCK 6. Unlike AD4, Vina, RLDOCK, and rDOCK, DOCK 6 defines its binding site using spherical data. Spheres within 8.5 Å of the native ligand binding site allowed the derivatives to move and flex but were also small enough to encapsulate the native binding site. Structural modification and MIC activity analysed each derivative in two distinctive categories. Each derivative was placed into a group according to its modification to the linezolid structure, divided into sub-categories shown in Figure 4.5: Morpholino, Aryl, Base, Tail and other oxazolidinone modifications. Some structures have multiple modifications, such as Aryl-Tail, Morpholino-Tail or Base-Tail, and these modifications were split into separate groups in the analysis. We separated these structures into these groups to determine whether a change in a structural group or MIC activity would positively or negatively alter the performance or score of each derivative. Derivatives with either *S. aureus* values were separated according to the following three different levels of activity: active (0.01 - <4 µg/mL),

moderately active ($\leq 4 - < 16 \mu\text{g/mL}$), or inactive ($> 16 \mu\text{g/mL}$). The crystal structure 4WFA was used for virtual screening of all derivatives.

3.5 Re-scoring of docking results with AnnapuRNA

AnnapuRNA, a knowledge-based scoring method, (Stefaniak & Bujnicki, 2021) was used as a re-scoring function on a sample of the VS oxazolidinone ligands. AnnapuRNA uses a coarse-grained (simplified) representation of both interacting partners. For RNA, only the canonical A, G, C, U residues were considered. For the ligand poses, they applied the concept of modeling pharmacophoric features by Gaussian 3D volumes instead of the more common point or sphere representations (Taminau et al., 2008). The RNA crystal structure and DOCK 6 docked ligand poses were scored using AnnapuRNA's k Nearest Neighbors or “k-NN modern” method (Stefaniak & Bujnicki, 2021). The results were then split into their modification groups for comparative analysis.

3.6 Statistical Analysis and Classifier

Pearson’s correlation calculations were performed using the functions available in the SciPy statistics library in Python 3.6. Several RF classifiers were generated from the data and individually examined, determining whether the docking program over or underpredicted the affinity of the derivative. This was then compared to the linear regression between the pMIC value and the absolute value of the docking score. The RF classifier model was built using the Scikit-learn library in Python 3.6 and trained on the Morgan fingerprint bit vector (radius = 2, encoding = 2048 bits), which was generated using RDKit (v. 2021.09.4) for each of the derivatives from the Zhao et al. dataset (Q. Zhao et al., 2021) that had *S. aureus* pMIC values.

3.7 Principal Component Analysis (PCA) and rescoring function

Molecular descriptors were calculated using RDKit (v. 2021.09.4), being molecular weight (MolWt), topological polar surface area (TPSA), number of rotatable bonds (nRotB), number of H bond donators (HBD), number of H bond acceptors (HBA), LogP, and the number of rings (NumRings), as well as the DOCK 6 scoring of 4WFA. Regression analysis was performed on the training set compounds with *S. aureus* pMIC values. The training set comprised of 30 randomly selected derivatives, 10 from active, 10 from moderately active, and 10 from inactive to provide an even spread of data (Supplementary Table 4.1). PCA or Principal Component Analysis was used through RDKit and Sklearn

python packages, unitizing a K-means clustering of 3. Efficiency of the rescoring equation was determined using r or Pearson's coefficient.

4. Conclusions

In this study, we evaluated the performance of five RNA adaptable docking programs in reproducing the crystallographic pose of ligands located in the native binding site of each target, which in this case, was oxazolidinones. Here, we studied several leading docking programs or algorithms, namely, DOCK 6, AutoDock 4, AutoDock Vina, rDock and RLDOCK, and evaluated their ability to correctly reproduce and score the crystal structure ligand configuration for eleven oxazolidinone crystal ribosomal structures. However, we caution readers about the use of these ribosomes in future studies due to the poor electron density and quality of said structures. Through the validations, we found AutoDock, AutoDock Vina, and DOCK 6 tended to perform better when redocking the native ligands compared to rDock and RLDOCK. RLDOCK had the highest success rate for docking under $\leq 2.5 \text{ \AA}$. However, RLDOCK was excluded due to the program's global docking-only function. DOCK 6 tended to have a much greater scoring than other programs, with the lowest median RMSD and the second highest success rate for docking under $\leq 2.5 \text{ \AA}$.

We reported that the structural modifications, factors, and type of MIC activity of a 285 oxazolidinone dataset had minimal to no influence on how the linezolid derivatives performed when docked. We found ten derivatives that score within the "top-performing" range, with 3 of those binding in the native binding site. However, due to the complexity of rRNA, such as the high flexibility of the binding pocket, which resulted in poor pose prediction and poor posing of the derivatives, results must be taken interpreted with caution and can only guide docking-based virtual screens of oxazolidinone targets. Morgan fingerprint analysis determined that DOCK 6 tended to over-predict derivatives with acetamide containing tail modifications and over-predict those with methylamino bits. However, due to the high similarity between each derivative, and the size of the dataset, these results may not be major contributing factors. We also suggested that scoring function, rescoring equation, and pharmacophore methods can be used in conjunction with docking. The rescoring equation method showed significant improvement in the predictions of training set. Additionally, further RNA-ligand studies should consider water, pH and metal ions, as it affects the stability of the said binding site. Additional methods, like MD, should also be considered. In conclusion, the current results suggest that one

should perform further docking studies and VS campaigns of oxazolidinones, adapt additional analysis methods and utilise a more extensive dataset.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/antibiotics12030463/s1>, Figure S4.1: Electron dentistry maps (at the binding site) for selected ribosomal crystal structures; Figure S4.2: DOCK 6 scoring of the dataset split into the specific structural modification groups; Figure S4.3: Distribution of molecular features of the derivatives (a) pMIC of *S. aureus* docked against their respective PDBs (4WFA), and the (b) number of rotatable bonds, (c) molecular weight, and the (d) number of atoms in the molecule vs the absolute docking score for 4WFA. Pearson values are displayed.; Figure S4.4: DOCK 6 Top three poses (blue) of each top-performing derivative superimposed against the linezolid binding site (beige).; Figure S4.5: The intermolecular interactions of linezolid and derivatives 3l, 29f, and 67 with the active site pockets of *S. aureus*.; Figure S4.6: DOCK 6 scoring compared to the re-scored values from AnnapuRNA.; Table S4.1: Docking score and descriptors of training set compounds.; Table S4.2: Experimental and predicted pMIC values of the rescores for the test set compounds with descriptors.; Discussion S4.1. Field-template pharmacophore.; Figure S4.7: 3D electrostatic-hydrophobic and shape properties. The pharmacophore model generated from linezolid, cadazolid, radezolid, tedizolid, and contezolid presents two distinct and separated regions: a positive electrostatic potential region in red, and a negative electrostatic potential region, in cyan. The Flare module of Cresset soft-ware was used to visualise the figure.; Method S4.1. Pharmacophore and Field Template Generation

Author Contributions: Conceptualisation: M.E.B., A.R.N.N., K.E.F.S., S.R., and N.S.G; Methodology: M.E.B. and A.R.N.N.; Software: M.E.B. and A.R.N.N.; Formal analysis: M.E.B.; Data curation: M.E.B. and A.R.N.N.; Writing - original draft preparation: M.E.B.; Writing-review and editing: M.E.B., A.R.N.N., P.C.R, K.E.F.S., S.R., N.S.G; Constructive feedback on analysis, P.C.R; Visualisation: M.E.B.; Supervision, K.E.F.S., S.R., and N.S.G.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding

Data Availability Statement: Molecular dockings were performed with AutoDock 4, AutoDock Vina, DOCK 6, rDock, and RLDOCK, using the cited versions, all open sources. Pharmacophore performed Cresset Flare was also using cited versions, which can be accessed through a commercial license. Scoring via AnnapuRNA is available through

GitHub via the cited version. All visualisation sources used for the preparation of Figures and Supplementary Figures are listed in captions of figures, with python visualisation toolkits used if not stated otherwise. All prepared oxazolidinone scaffolds from the dataset are available at <https://github.com/gandhi-group-biomolecular/docking-oxazolidinone-dataset>

Acknowledgments: The eResearch Office, Queensland University of Technology, Brisbane, Australia, provided the computational resources and services used in this work. This study was supported by funding from the QUT Centre of Genomics and Personalised Health. We also thank Ms. Tabassum Khair Barbhuiya for her time in proofreading this manuscript.

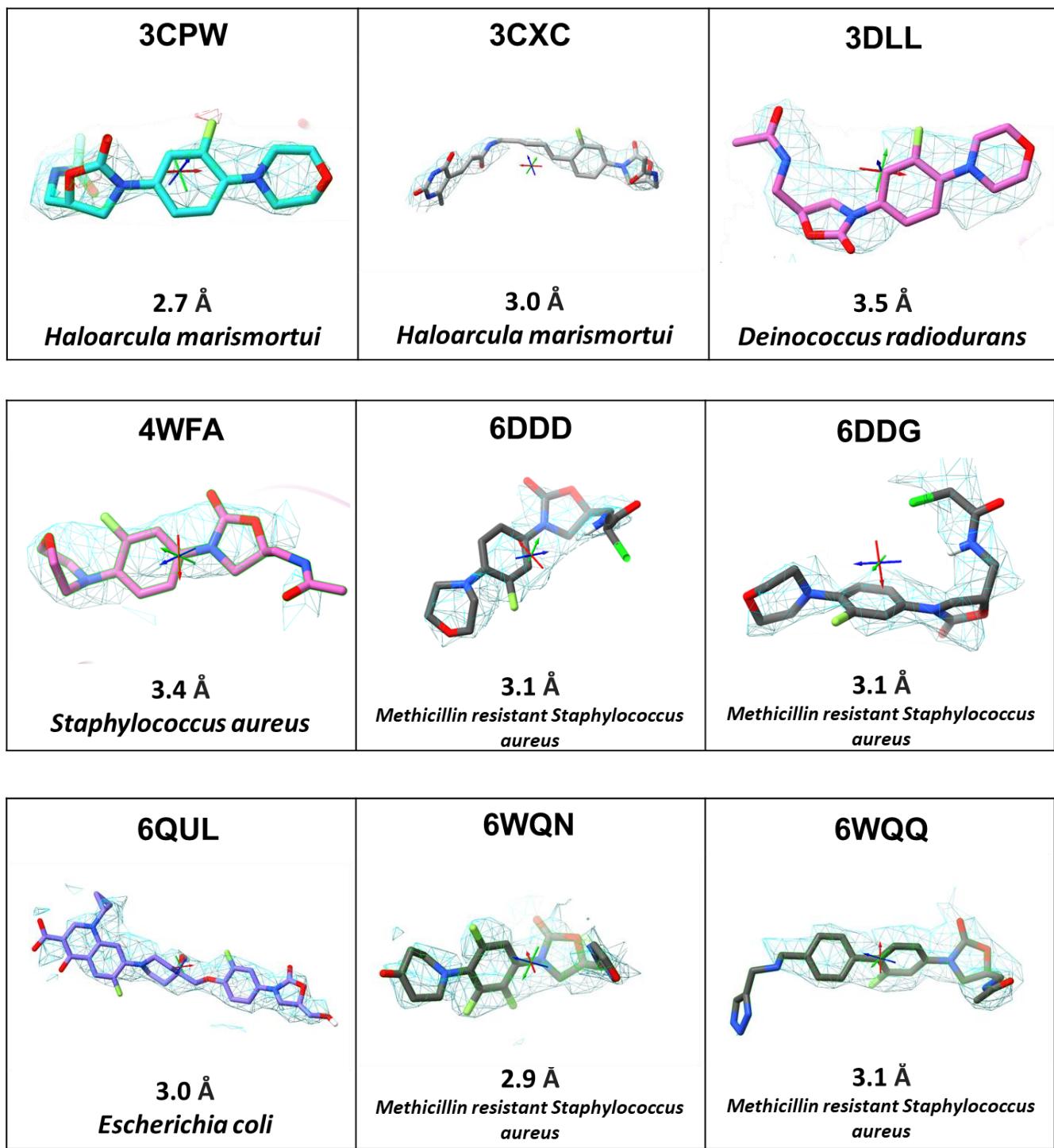
Conflicts of Interest: The authors declare no conflict of interest

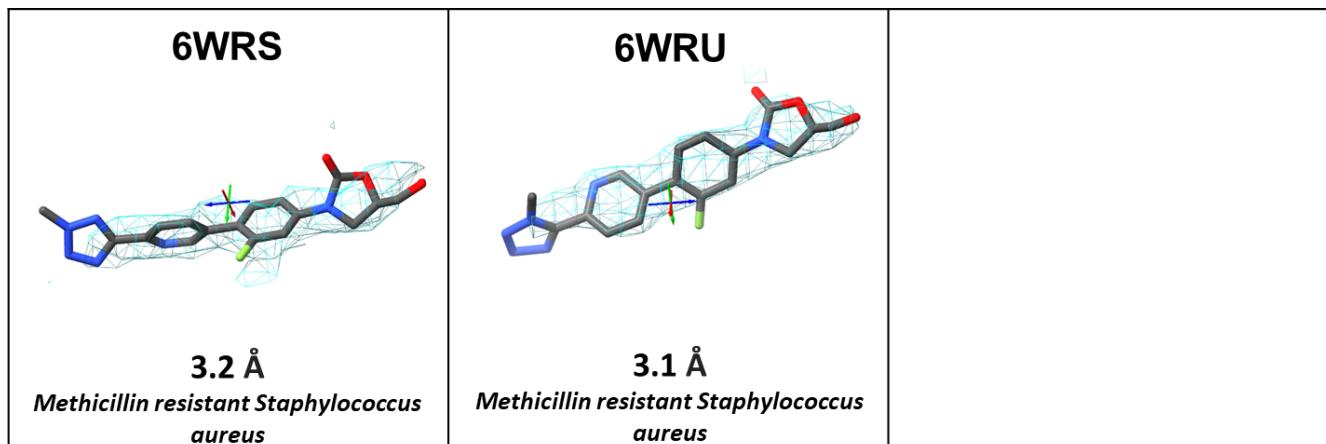
Abbreviations

AD4	AutoDock 4
ATCC	American Type Culture Collection
<i>D. radiodurans</i>	<i>Deinococcus radiodurans</i>
DMS	Dot molecular surfaces
<i>E. coli</i>	<i>Escherichia coli</i>
<i>E. faecalis</i>	<i>Enterococcus faecalis</i>
<i>H. marismortui</i>	<i>Haloarcula marismortui</i>
HBA	number of H bond acceptors
HBD	number of H bond acceptors
HPC	High-performance computing
<i>K. pneumoniae</i>	<i>Klebsiella pneumoniae</i>
kNN	k Nearest Neighbors
MD	Molecular dynamics
MIC	minimum inhibitory concentration
ML	Machine learning
MolWt	molecular weight
mRNA	Messenger Ribonucleic acid
MRSA	Methicillin-resistant <i>Staphylococcus aureus</i>
NA	Nucleic acid
NCTC	National Collection of Type Cultures

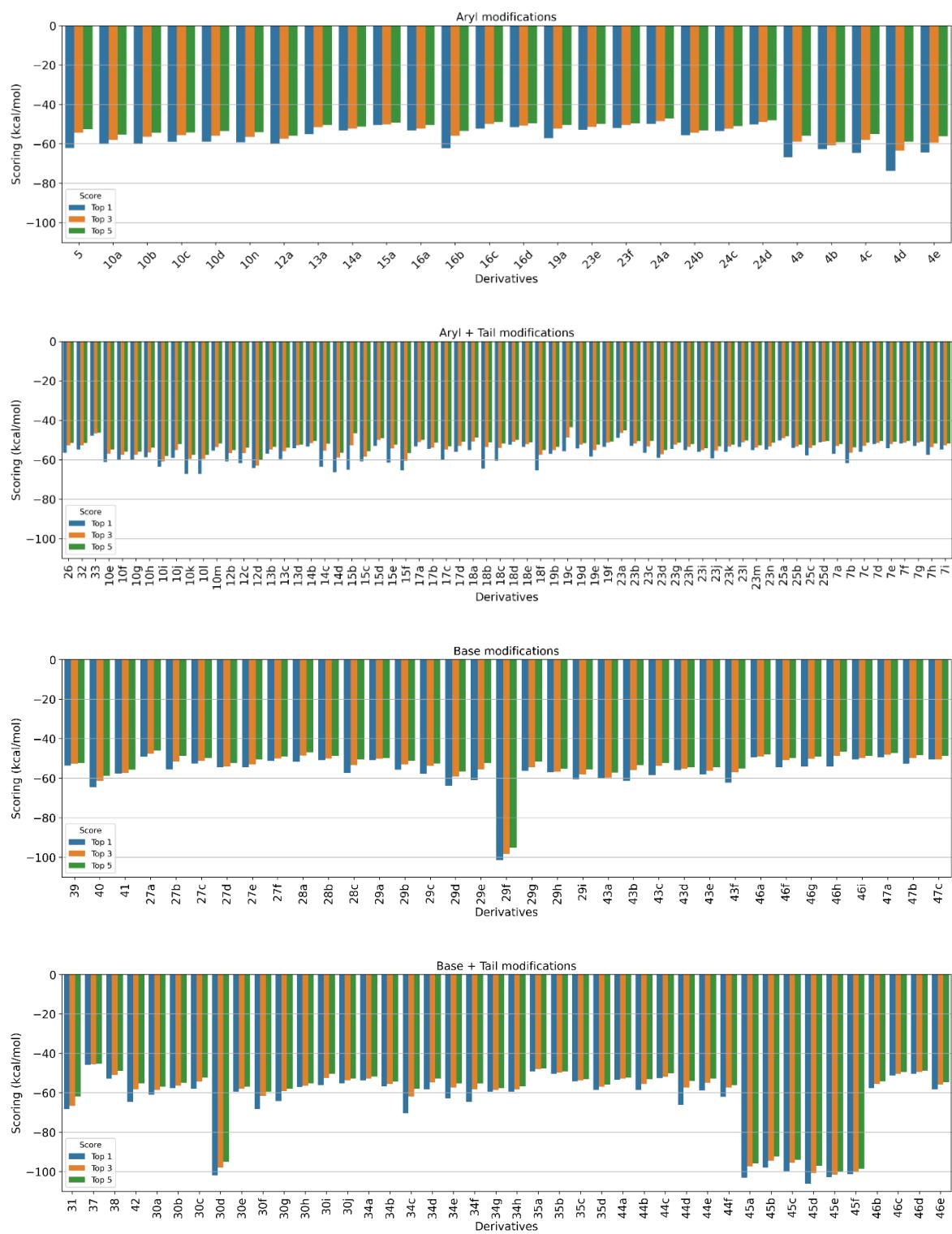
NumRings	Number of Rings
<i>P. aeruginosa</i>	<i>Pseudomonas aeruginosa</i>
PCA	Principle component analysis
PDB	Protein Data Bank
pMIC	Log of the minimum inhibitory concentration
POAP	Parallelized Open Babel & Autodock suite Pipeline
PTC	peptidyl transferase centre
QSAR	Quantitative structure-activity relationship
RCSB	Research Collaboratory for Structural Bioinformatics
RF	Random Forest
RMSD	Root mean square deviation
RNA	Ribonucleic acid
<i>S. aureus</i>	<i>Staphylococcus aureus</i>
<i>S. capitis</i>	<i>Staphylococcus capitis</i>
<i>S. epidermidis</i>	<i>Staphylococcus epidermidis</i>
<i>S. pyogenes</i>	<i>Streptococcus pyogenes</i>
SAR	structure-activity relationships
SF	Scoring Function
SMILE	Simplified molecular-input line-entry system
TPSA	topological polar surface area
tRNA	Transfer ribonucleic acid
VDW	Van der Waals
Vina	AutoDock Vina
VRE	Vancomycin-resistant <i>Enterococci</i>
VS	Virtual screening

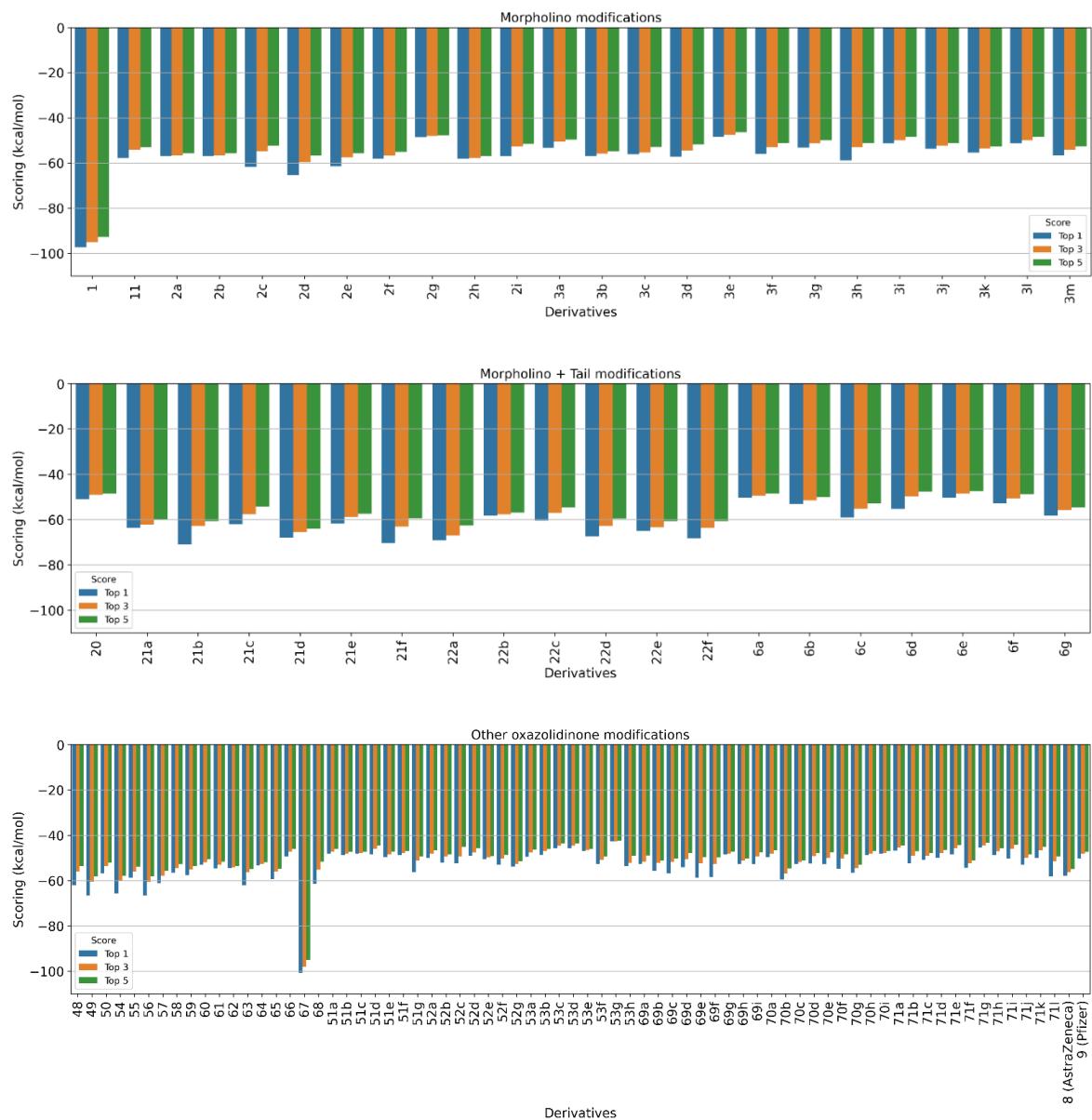
Supplementary Material



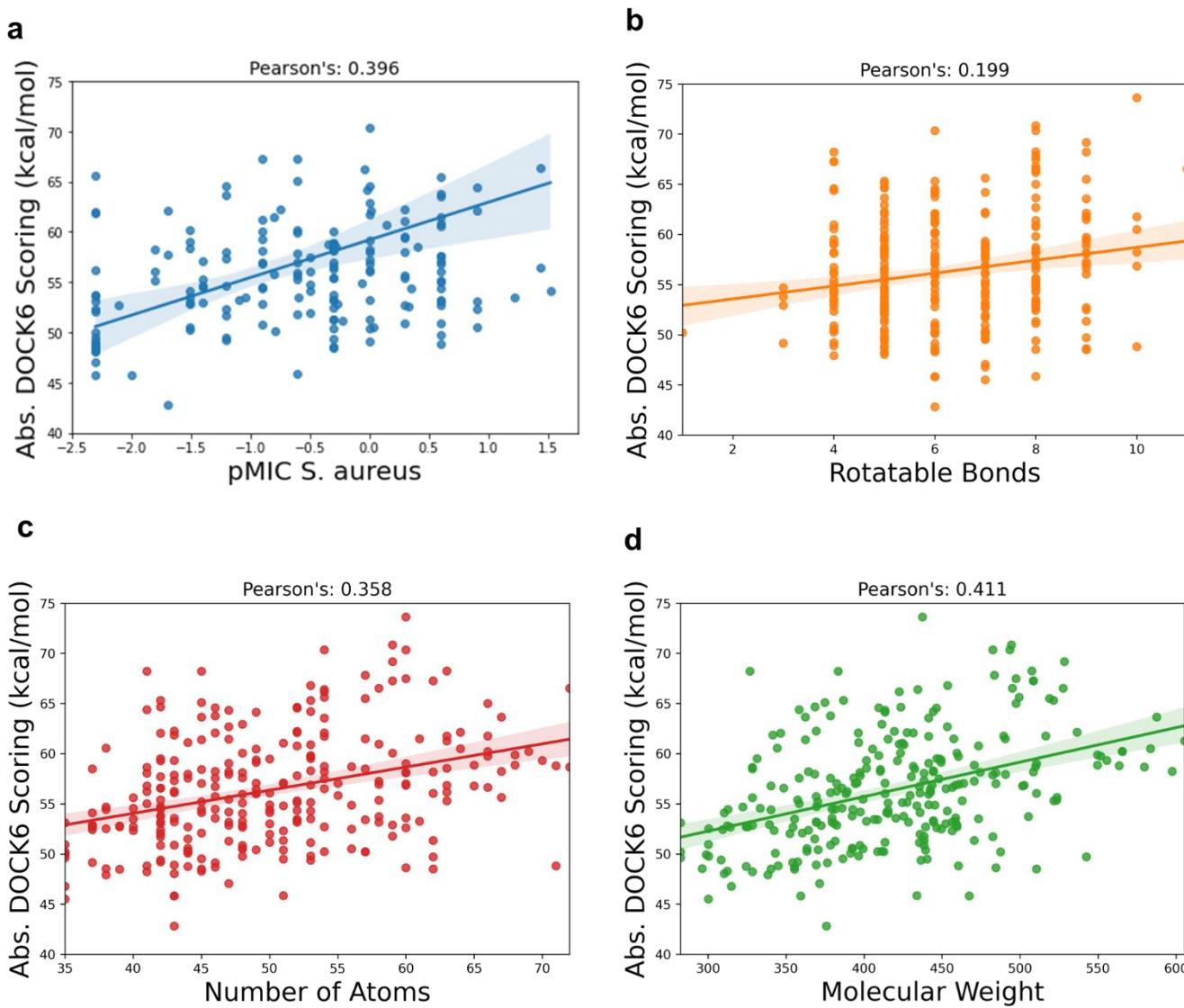


Supplementary Figure 4.1 - Electron dentistry maps (at binding site) for selected ribosomal crystal structures: 3CPW (2.7 Å), 3CXC (3.0 Å), 3DLL (3.5 Å), 4WFA (3.4 Å), 6DDD (3.1 Å), 6DDG (3.1 Å), 6QUL (3.0 Å), 6WQN (2.9 Å), 6WQQ (3.1 Å), 6WRS (3.2 Å), and 6WRU (3.1 Å). The resolution is shown in the mesh in blue, with the XYZ axis indicator displayed (red, green, blue, respectively). Species type is also listed. Note: the binding site residues are not shown for clarity of the electron density, due to other residue that obstructed the full view of the ligand. ChimeraX (v.1.4) was used for the visualization, in conjunction with the ISOLDE toolkit.

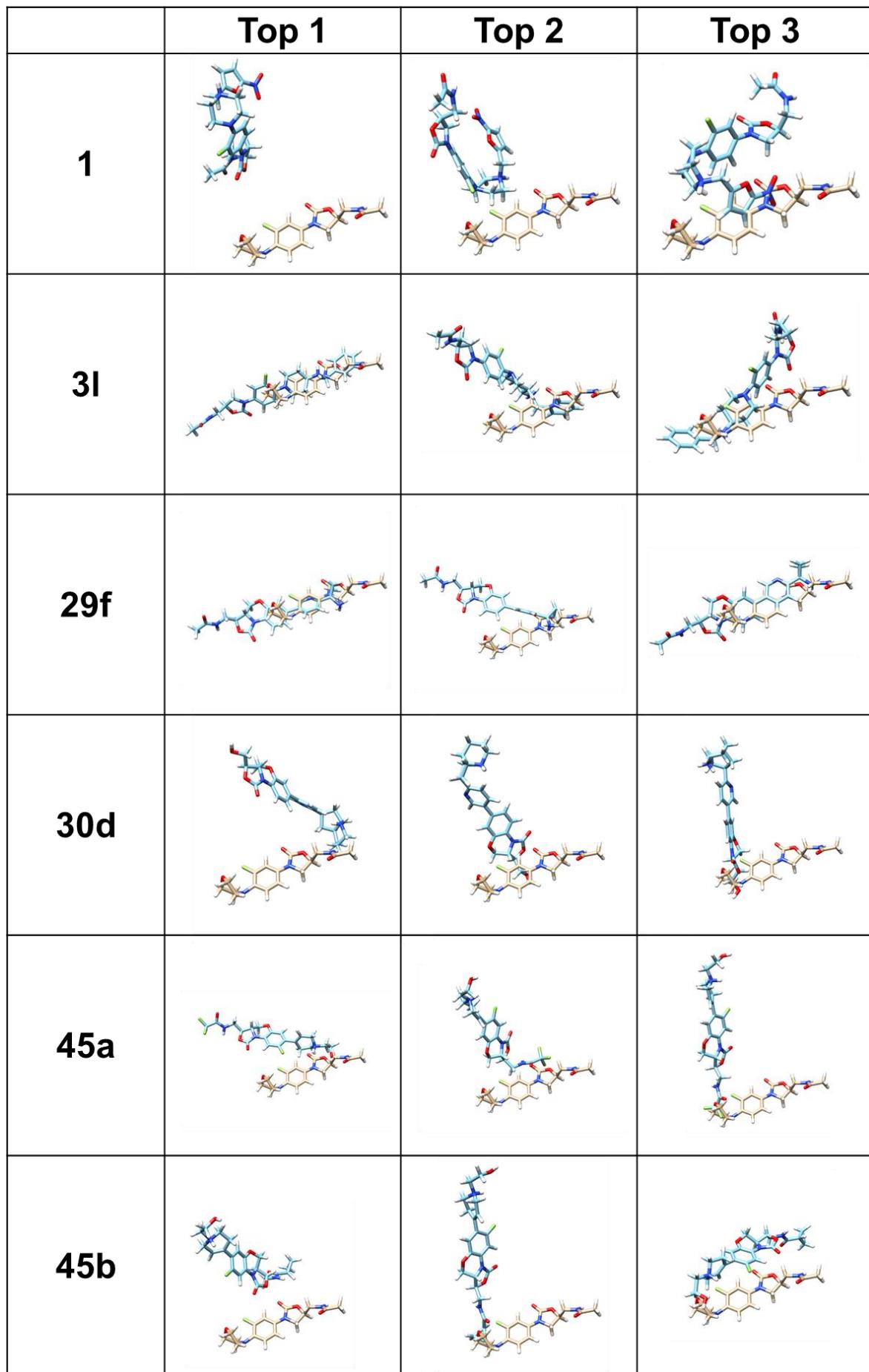


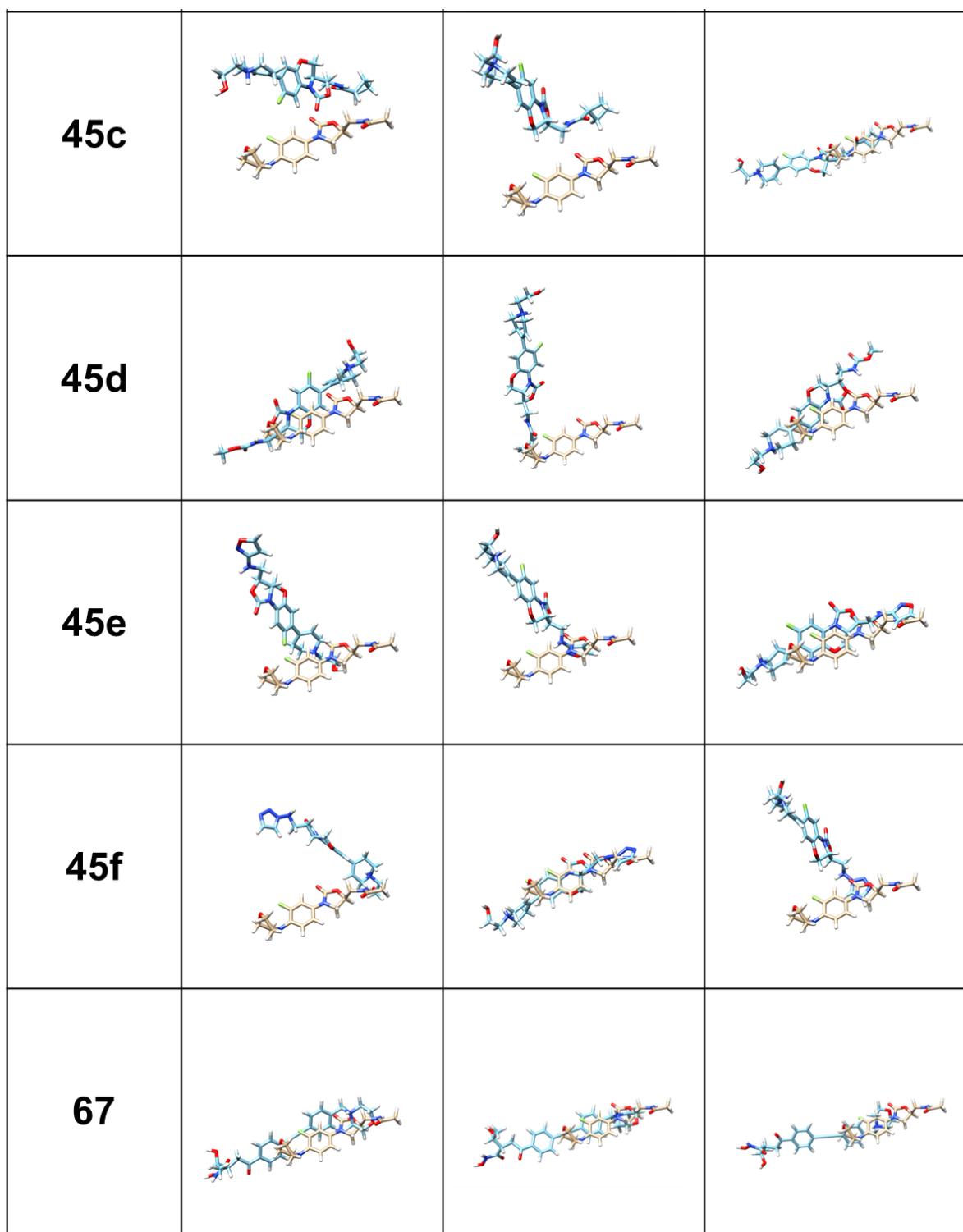


Supplementary Figure 4.2 - DOCK 6 scoring of the dataset split into the specific structural modification groups. Many values stayed above -60kcal/mol, while the top performers were all below -90kcal/mol.

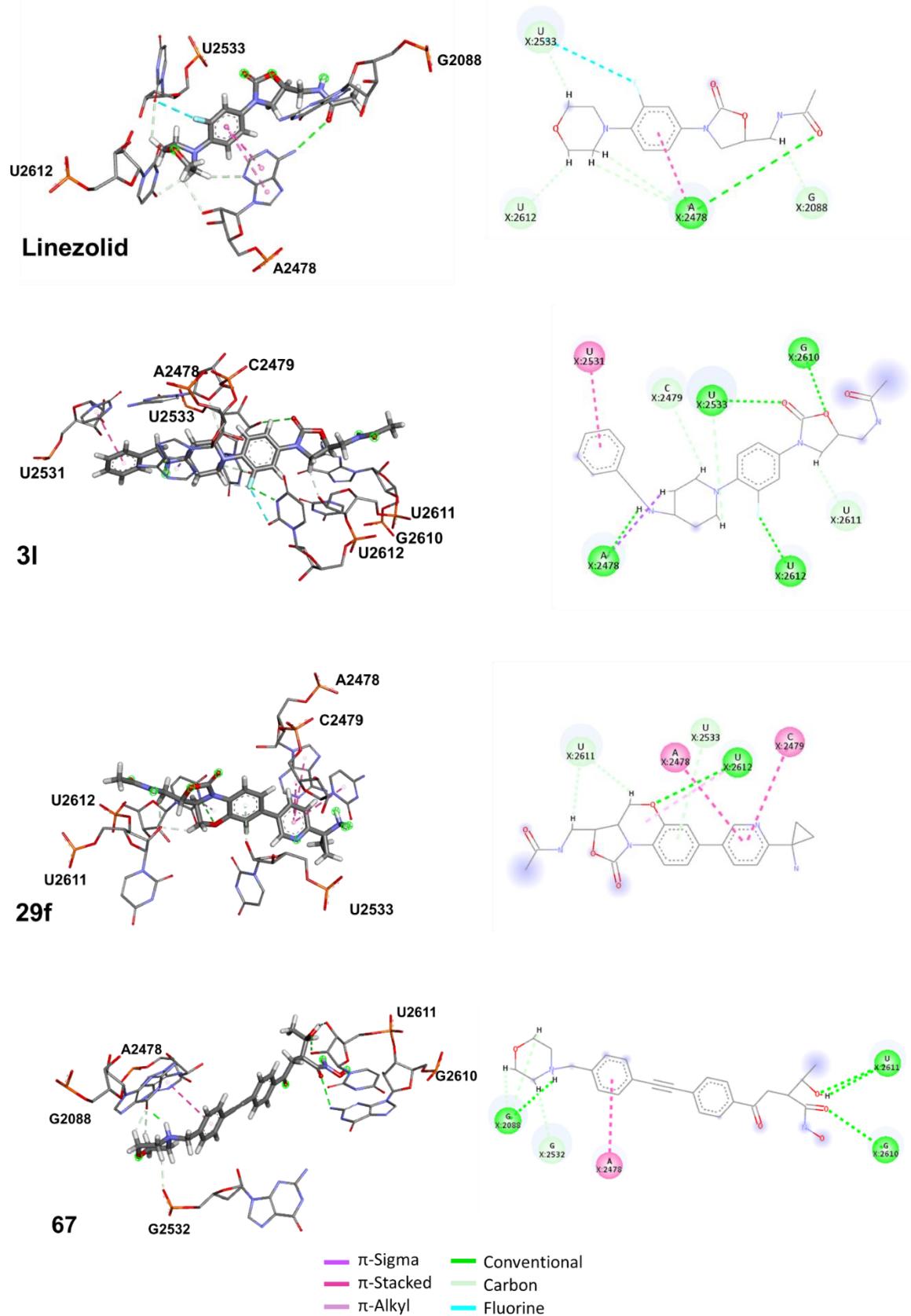


Supplementary Figure 4.3 - Distribution of molecular features of the derivatives (a) pMIC of S. aureus docked against their respective PDBs (4WFA), and the (b) number of rotatable bonds, (c) molecular weight, and the (d) number of atoms in the molecule vs the absolute docking score for 4WFA. Pearson values are displayed.





Supplementary Figure 4.4 - DOCK 6 Top three poses (blue) of each top-performing derivative superimposed against linezolid (beige) in the pocket of 4WFA. Note: the nucleotide residues in the binding site are not shown for clarity of the posing.



Supplementary Figure 4.5 - The intermolecular interactions of linezolid and derivatives 31, 29f, and 67 with the active site pockets of *S. aureus*. Left side (3D representation): the

docked residues of the target are shown in the rose stick model (labelled), and compounds are in the grey stick. Right side (2D representation): compounds shown in grey with the residue interactions displayed. E. coli residue numbering is used for both 3D and 2D representations. The dashed lines represent the different interactions, and their colour is the interaction type. The coloured circles represent residues with letter code (A for Adenine, C for Cytosine, G for Guanine, or U for Uracil), chain identifier (Chain X) and residue number. 2D visualization was carried out using BIOVIA Discovery Studio.

Supplementary Table 4.1 - ABS docking score and descriptors of training set

ID	ABS 4WFA	MolWt	TPSA	nRotB	HBD	HBA	LogP	NumRings	S. <i>aureus</i> pMIC
9	47.36	487.1503	143.31	1	2	9	1.2027	6	-0.796
10e	44.699	591.1377	146.38	4	3	8	1.5676	6	-0.903
13b	51.005	443.1439	97.38	5	3	7	0.3724	3	0.602
14c	56.016	441.1482	106.61	5	3	8	-0.1827	3	0.602
16c	51.088	456.0926	84	4	1	7	3.4416	5	-1.204
18e	50.998	355.0968	95.52	3	2	5	2.0467	4	1.222
19a	51.444	394.1652	102.42	5	2	6	-0.0721	3	-0.903
19f	54.262	406.1641	97.56	5	1	6	3.41706	4	-1.505
23e	50.774	372.1598	84.42	4	1	5	2.66926	3	-0.903
23g	50.352	392.1051	84.42	4	1	5	3.48156	3	-0.903
23m	54.194	424.1547	97.56	5	1	6	3.55616	4	-1.806
2a	54.379	458.219	121.32	6	2	9	0.7588	4	-0.903
2b	50.401	502.2088	138.44	8	1	10	1.50812	4	0.000
2f	47.958	518.1481	120.62	7	1	8	3.0047	4	-0.903
3e	44.662	510.189	83.14	7	2	6	4.2665	4	-0.301
3i	56.918	510.189	83.14	7	2	6	4.2665	4	-0.301
3j	49.982	504.1172	73.91	6	2	5	4.1304	4	-0.301
43b	50.009	457.2013	88.18	4	1	5	2.8539	5	-0.806
43d	55.058	429.17	88.18	4	1	5	2.0737	5	-0.803
43f	53.284	419.1493	108.41	4	2	6	0.656	4	-0.742
44d	58.146	421.1649	100.57	5	2	7	1.3493	4	-0.043
45c	93.988	421.1472	71.11	4	1	6	2.3814	5	1.523
46i	47.383	399.1406	71.11	3	1	5	2.2835	4	0.036
51b	47.424	324.1474	67.43	4	2	3	3.28832	3	-2.301
51c	46.803	328.1223	67.43	4	2	3	3.119	3	-2.301
52d	45.672	376.1093	93.73	6	2	5	2.537	3	-2.301
52f	53.889	380.0598	84.5	5	2	4	3.1818	3	-2.301
53e	44.171	371.1304	71.62	5	3	4	3.5451	3	-2.301
6d	53.47	462.0982	89.75	7	0	5	5.7219	4	-1.398
7b	51.627	341.1198	62.39	4	3	3	3.5365	3	-2.301

Supplementary Table 4.2 - Experimental and predicted pMIC values of the rescores for the test set compounds with descriptors

ID	ABS 4WFA	MolWt	TPSA	nRotB	HBD	HBA	LogP	Num Rings	Experimental pMIC	Predicted pMIC
1	97.353	461.1711	121.4	7	1	8	2.1104	4	0	0.799691
8	57.942	489.0237	71.11	4	1	5	3.3536	3	0	0.851412
11	57.949	368.1285	75.94	4	1	5	2.6017	4	0.30103	-0.7237
20	50.876	378.1241	77.55	4	0	7	2.7573	5	0.60206	-0.55158
26	56.512	452.162	123.41	6	3	9	0.5356	4	1.443697	-0.37488
32	54.685	322.1066	98.48	2	2	5	1.27178	4	-1.20412	-1.43103
37	45.855	433.1081	66.84	6	1	4	5.5193	4	-0.60206	-1.86319
38	52.935	339.0678	86.45	2	1	6	2.41538	4	-1.50515	-0.87498
39	53.558	379.1332	82.43	3	1	4	2.74598	4	-0.59329	-1.06528
40	64.629	423.1543	113.1	4	2	7	1.9606	5	-1.20412	-0.76063
41	57.603	393.1325	100.99	3	2	6	1.3147	4	0.60206	-0.53145
48	62.13	535.1429	62.21	8	0	4	6.4694	5	0.90309	-0.69732
54	65.637	498.0862	92.79	6	2	6	4.6371	5	-2.30103	-0.84279
59	57.708	429.0147	63.25	5	2	5	4.9671	4	-1.69897	-0.92773
61	54.717	369.0947	63.25	5	2	5	4.3437	4	-1.39794	-1.5154
62	54.548	385.0652	63.25	5	2	5	4.858	4	-1.39794	-1.52474
63	62.107	427.1354	63.25	6	2	5	5.8716	5	-1.69897	-1.59757
64	53.264	401.1198	63.25	5	2	5	5.3578	5	-1.09691	-1.85568
10b	59.878	549.2024	137.15	4	3	7	2.2214	6	-0.60206	-0.82068
10c	58.857	549.2024	137.15	4	3	7	2.2214	6	-0.30103	-0.85996
10d	58.792	563.218	137.15	5	3	7	2.6115	6	-0.30103	-1.03962
10f	60.231	565.1973	146.38	4	3	8	2.4413	6	-1.50515	-0.74198
10g	60.231	551.1816	146.38	4	3	8	2.0512	6	-0.60206	-0.76215
10h	58.685	579.2129	146.38	5	3	8	2.8314	6	-0.30103	-0.97862
10i	63.641	587.1486	163.45	5	3	9	0.4808	6	-1.20412	-0.10931
10j	59.013	510.1551	137.51	3	3	8	0.9239	6	-1.50515	-0.40133
10k	67.275	508.1758	128.28	3	3	7	2.0776	6	-0.90309	-0.4923
10l	67.275	494.1602	128.28	3	3	7	1.6875	6	-0.60206	-0.51247
10m	55.29	507.1918	120.08	4	3	6	1.6263	6	-0.60206	-0.98593
10n	59.328	555.2017	123.71	6	1	9	2.0477	5	-0.90309	0.549038
12b	60.898	425.1533	97.38	5	3	7	0.2333	3	0.60206	0.388114

12d	64.199	444.1557	122.05	5	2	9	0.3551	5	-0.0187	0.078089
13a	55.139	427.1667	114.45	5	3	7	-0.4284	3	0.60206	-0.09383
14a	53.173	428.1507	111.65	5	2	7	-0.0014	3	0.60206	0.000876
14d	66.391	434.1714	123.41	6	3	9	0.3965	4	1.443697	-0.15451
15a	50.5	392.1931	58.64	4	1	3	4.18122	3	-1.50515	-0.98059
15b	65.103	374.1146	84.42	4	1	5	3.34246	3	-0.60206	-0.62995
15c	60.765	473.121	108.41	4	2	6	1.1984	4	0.139662	0.144798
15d	52.881	459.1806	108.41	5	2	6	1.4362	5	-0.27138	-0.89247
16a	53.199	369.1237	88.83	4	1	6	1.9967	4	0.60206	-0.82556
16b	62.242	369.1237	88.83	4	1	6	1.9967	4	0.30103	-0.47767
17a	53.157	379.1193	90.44	4	0	8	2.1523	5	0.90309	-0.38295
17b	54.456	379.1193	90.44	4	0	8	2.1523	5	0.60206	-0.33298
17c	60.064	422.0912	84.42	5	1	5	3.31396	3	-0.30103	-0.44751
17d	56.106	378.1241	77.55	4	0	7	2.7573	5	0.60206	-0.35038
18a	55.02	358.0965	88.1	3	2	5	2.1411	4	0.60206	-1.19313
18b	64.396	358.0965	88.1	3	2	5	2.1411	4	0.90309	-0.83243
18c	60.551	344.0808	95.77	3	2	5	2.2447	4	0.60206	-1.42079
18d	52.344	342.1016	78.87	3	2	4	2.3048	4	0.90309	-1.50311
19c	55.651	523.2067	130.61	4	1	8	3.2902	5	-0.60206	-0.49142
19e	58.302	444.1445	117.37	5	2	8	1.62	5	0.000435	-0.71659
21c	62.053	346.0987	84.5	5	2	4	2.5284	3	-2.30103	-1.44943
23a	48.898	369.1125	95.52	3	2	5	2.35512	4	0.60206	-1.60841
23b	52.856	382.1554	86.03	4	0	7	2.82486	4	-1.50515	-0.43075
23d	59.122	449.1248	121.29	4	0	10	2.62402	6	0	-0.25778
23i	55.938	440.0818	84.42	5	1	5	3.45306	3	-0.60206	-0.44655
23k	56.013	388.1547	93.65	4	1	6	2.88916	3	-1.80618	-0.68841
23n	54.867	445.1649	108.41	5	2	6	1.0461	5	-0.06408	-0.83624
24b	55.461	399.123	96.97	4	2	5	2.2849	4	0.30103	-1.21998
25d	51.128	435.1442	117.64	4	2	7	0.8759	4	-0.22531	-0.59359
27a	49.228	357.1125	80.76	3	1	5	2.11	4	-1.20412	-0.95857
27b	55.399	357.1125	80.76	3	1	5	2.11	4	-0.30103	-0.72117
27c	52.478	449.1598	128.64	5	3	7	0.0169	4	-0.30103	-0.82749
27e	54.401	407.1315	71.11	3	1	6	2.1322	5	0.356547	0.077864
27f	51.298	381.1325	97.83	4	1	6	2.1735	4	-0.30103	-1.10199
28a	51.718	339.1219	80.76	3	1	5	1.9709	4	-1.20412	-1.02247

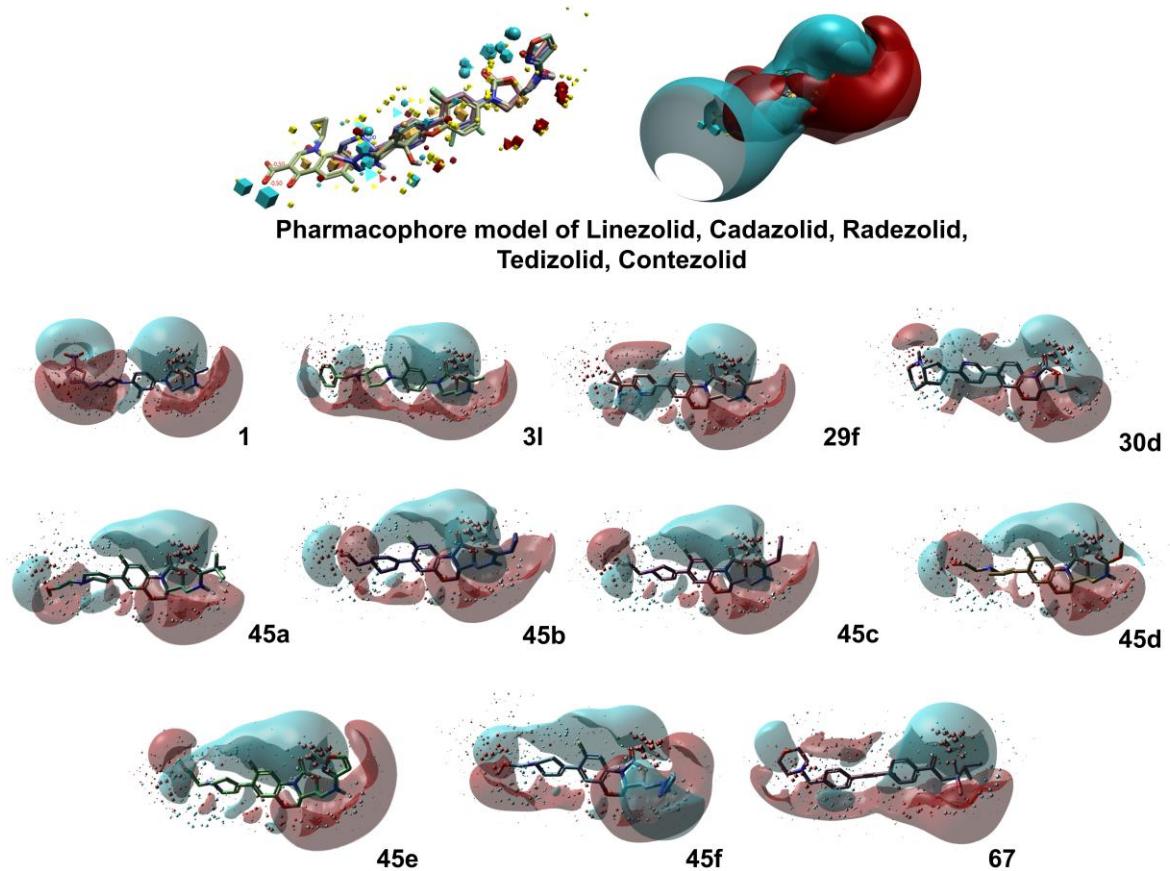
28b	50.842	339.1219	80.76	3	1	5	1.9709	4	-0.90309	-1.05617
28c	57.316	364.1172	104.55	3	1	6	1.84258	4	-1.20412	-0.9586
29a	50.863	397.1274	107.06	4	1	7	1.7575	4	0.30103	-0.81834
29b	55.723	369.1325	89.99	4	1	6	1.9795	4	0.30103	-0.75793
29c	57.867	338.1267	67.87	3	1	4	2.5759	4	-0.60206	-0.86679
29d	63.77	378.1328	104.55	4	1	6	2.03698	4	0.60206	-0.81507
29e	61.053	424.1383	110.3	4	1	7	1.9275	5	0.30103	-0.55764
29f	101.443	421.1499	124.36	4	1	9	1.1614	5	0.60206	1.298353
29g	56.266	382.1077	104.55	3	1	6	1.98168	4	-0.30103	-0.83931
29h	57.098	394.1641	106.78	4	2	6	1.9187	5	0.60206	-1.42399
2c	61.764	516.2245	138.44	8	1	10	1.89822	4	0	-0.08064
2d	65.502	474.1663	133.76	7	1	9	1.9443	4	0.60206	-0.34988
2e	61.484	490.1435	120.62	7	1	9	2.4128	4	0.60206	-0.08319
2g	48.624	484.187	120.62	7	1	8	2.3513	4	-0.30103	-0.86977
2h	58.09	484.187	120.62	7	1	8	2.3513	4	-0.30103	-0.50561
2i	57.016	484.187	120.62	7	1	8	2.3513	4	-0.60206	-0.54693
30b	57.565	358.1165	112.35	4	3	7	0.8528	4	0.60206	-1.30758
34c	70.348	482.135	132.14	5	1	10	1.5831	6	0	0.183135
34d	58.207	364.1648	86.03	4	0	7	2.68576	4	-1.80618	-0.38458
34e	62.936	411.1343	102.6	4	1	8	2.3189	5	0	-0.29939
34f	64.572	409.1186	99.44	4	0	8	2.4682	5	0	0.012577
34g	59.486	448.1408	113.08	5	0	10	1.9054	6	0.30103	0.067558
34h	59.388	449.136	125.97	5	0	11	1.3004	6	0.30103	0.144662
35a	49.15	354.0764	118.54	2	1	6	1.33088	4	0	-1.43736
35b	50.274	411.0979	133.67	3	1	8	1.81762	5	-1.50515	-1.36669
35c	54.136	414.0976	124.29	3	1	7	1.4158	5	-1.50515	-0.98835
35d	58.431	410.1139	125.46	4	1	8	1.099	5	-1.50515	-0.74155
3a	53.255	458.2129	73.91	7	2	5	3.1847	4	-0.30103	-0.82182
3b	56.935	460.1677	73.91	6	2	5	4.0213	4	-0.30103	-0.76949
3c	56.163	442.238	73.91	6	2	5	4.004	4	0	-1.00322
3d	57.243	504.1172	73.91	6	2	5	4.1304	4	0	-0.2822
3f	56.058	510.189	83.14	7	2	6	4.2665	4	0	-0.54546
3g	53.177	444.1973	73.91	6	2	5	3.507	4	0	-0.91123
3h	58.861	470.2329	94.14	7	3	6	3.4212	4	-0.30103	-1.16463
3k	55.374	522.1078	73.91	6	2	5	4.2695	4	-0.30103	-0.19442

3l	51.338	438.1452	132.44	6	2	7	1.6336	4	0	-1.68378
3m	56.62	456.2537	73.91	7	2	5	3.6817	4	-0.30103	-0.89924
43a	60.073	459.1806	97.41	4	1	6	1.8427	5	-0.89982	0.001942
43c	58.561	401.1587	108.41	4	2	6	0.5169	4	0.407823	-0.53244
43e	58.053	340.1172	93.65	3	1	6	1.3659	4	-1.20412	-0.69788
44a	53.503	459.1417	91.34	5	2	6	1.6718	4	-1.03743	-0.14228
44b	58.496	431.1856	91.34	6	2	6	1.5195	5	-0.2824	-0.70434
44c	52.627	445.2013	91.34	6	2	6	1.9096	5	0.325139	-0.90995
44e	58.728	430.1652	100.3	6	2	8	2.0934	5	-0.34694	-0.70791
44f	62.062	430.1765	104.98	6	2	9	0.8285	5	0.008774	-0.01183
45b	97.893	407.1315	71.11	4	1	6	1.9913	5	1.522879	1.605823
45d	106.038	397.1108	80.34	3	1	7	1.8211	4	1.522879	2.305655
45e	102.765	406.1111	80.07	4	1	8	2.5652	5	1.522879	1.780758
46a	49.482	440.1154	101.07	4	1	7	2.6408	5	-1.20412	-0.79322
46d	50.432	370.1641	93.65	4	1	6	2.75006	3	-0.30103	-1.06279
46f	54.464	381.1159	71.11	3	1	6	1.6012	4	-0.90069	0.256464
46g	54.358	377.1387	80.34	3	1	6	0.8846	5	0.040959	-2.19135
46h	54.191	397.1108	88.18	3	1	6	0.6167	4	1.522879	0.269091
47a	49.382	436.1659	102.24	4	1	7	2.0648	5	-0.30103	-0.66656
47b	52.667	354.1692	84.42	4	1	5	2.53016	3	-2.10721	-1.04193
47c	50.534	363.1594	71.11	3	1	5	1.0439	4	0.90309	-0.14398
51a	48.044	355.1168	110.57	5	2	5	2.8881	3	-2.30103	-2.57856
51d	48.506	340.1423	76.66	5	2	4	2.9885	3	-2.30103	-1.96784
51f	48.814	344.0928	67.43	4	2	3	3.6333	3	-2.30103	-1.9096
51g	56.231	388.0423	67.43	4	2	3	3.7424	3	-2.30103	-1.14882
52a	50.071	391.0838	127.64	6	2	6	2.4366	3	-2.30103	-2.39287
52b	52.089	360.1144	84.5	5	2	4	2.83682	3	-2.30103	-1.78236
52c	52.373	364.0893	84.5	5	2	4	2.6675	3	-2.30103	-1.66214
52g	53.759	424.0092	84.5	5	2	4	3.2909	3	-2.30103	-1.13622
53a	49.513	386.1049	105.53	5	3	5	3.4447	3	-2.30103	-2.43816
53b	48.695	355.1354	62.39	4	3	3	3.84492	3	-2.30103	-1.93675
53c	45.827	359.1104	62.39	4	3	3	3.6756	3	-2	-1.93779
53d	45.827	467.0164	62.39	4	3	3	4.1411	3	-2.30103	-0.84358
53f	52.704	459.1388	106.61	5	3	8	-0.0436	3	0.60206	0.532949
53g	42.806	375.0808	62.39	4	3	3	4.1899	3	-1.69897	-2.05685

53h	53.585	419.0303	62.39	4	3	3	4.299	3	-2.30103	-1.16673
6b	53.184	455.0626	71.11	5	1	5	2.7871	3	-0.30103	0.28205
6e	50.401	439.1355	84.25	5	1	6	2.8902	4	0	-0.49752
6f	52.71	407.086	71.11	4	1	5	2.8156	3	0.30103	-0.11248
6g	58.198	444.1621	91.42	5	1	7	1.7668	3	-0.30103	0.584672
70a	49.763	399.1406	71.11	5	1	5	2.1694	3	0.60206	-0.27735
7a	57.039	463.1187	76.07	7	1	5	5.5279	4	-1.39794	-1.32248
7c	56.074	435.1037	46.61	6	0	3	5.9528	4	-0.49485	-0.99497
7d	51.96	435.1037	46.61	6	0	3	5.9528	4	-0.49485	-1.15324
7e	54.12	435.1037	46.61	6	0	3	5.9528	4	-0.49485	-1.07014
7f	51.787	438.218	86.03	7	0	7	4.16746	4	-0.60206	-0.9026
7g	53.017	451.0742	46.61	6	0	3	6.4671	4	-1.39794	-1.11541
7h	57.499	447.1237	55.84	7	0	4	5.8223	4	-0.49485	-0.98937
7i	54.815	433.1081	66.84	6	1	4	5.5193	4	-0.49485	-1.5185

Supplementary Discussion 4.1 - Field-template pharmacophore

To expand the molecular docking results, we performed ligand-based reverse pharmacophore modelling of the active compounds of known oxazolidinone-based antibacterial drugs. The approach uses a molecular field points-based similarity method to generate a series of low-energy conformations for each compound. The ligands from the chosen crystal structure (linezolid (Bulkley et al., 2010; Eyal et al., 2015; Wilson et al., 2008), cadazolid (Scalio et al., 2019), radezolid (Wright et al., 2020), tedizolid (Wright et al., 2020), and contezolid (Wright et al., 2020)) were imported and utilised to identify structural regions of the ligands crucial for bioactivity with the help of the Flare field template (Cresset Group, 2022). After the conformation hunt, the 2D similarity, shape similarity, and field similarity scores corresponding to the generated template were found to be 0.664, 0.609, and 0.720, respectively. Figure S7 shows the resulting pharmacophore model, displaying important features: positive electrostatic potential (red), negative electrostatic potential (cyan), hydrophobicity (tan), and van der Waals descriptors (yellow). Two well-distinct and separated regions can be identified in the native ligands: a positive electrostatic region and a negative electrostatic region. The top-performing derivatives are displayed for comparison. The orientation of the positive and negative electrostatic regions is consistent amongst all the derivatives, with variations between the width and segmentation of the parts.



Supplementary Figure 4.6 - 3D electrostatic-hydrophobic and shape properties. The pharmacophore model generated from linezolid, cadazolid, radezolid, tedizolid, and contezolid presents two distinct and separated regions: a positive electrostatic potential region in red, and a negative electrostatic potential region, in cyan. Image generated using Flare™ from Cresset® (v.3) (*Bauer & Mackey, 2019; Cheeseright et al., 2006; Cresset Group, 2022; Kuhn et al., 2020*).

In addition to the ligand, electrostatics of the binding site can show the complementarity to the ligand and enable the prioritisation of new designs. The electrostatic potential measures the strength of the nearby charges, nuclei, and electrons at a particular position (Rahuman et al., 2020). The positive electrostatic potential is associated with the repulsion of the proton by the atomic nuclei in regions with low electron density and when the nuclear charge is incompletely shielded. At the same time, negative electrostatic potential corresponds to an attraction of the proton by the concentrated electron density in the molecules (from lone pairs and pi-bonds) (Chami Khazraji & Robert, 2013). Figure S7 demonstrates strong positive and negative electrostatic potential surrounding the varying

tail groups attached to the oxazolidinone rings. The more significant areas of cyan and red colour indicate a higher value of negative and positive electrostatics, respectively, leading to potentially higher activity.

When used in conjunction, docking and pharmacophore can complement each other in revealing critical structural features and could be helpful for the development of highly selective and potency of potential drug molecules. An example is the pharmacophore filtering method in which post-processing docking reduces poses and molecules not chemically compatible with the binding site. This includes poses that do not fill the site or leave unpaired buried hydrogen bond donors or acceptors (Peach & Nicklaus, 2009). This method could be viewed as implementing a fundamental principle of structure-based drug design; ligands that bind well must be chemically complementary to their receptors.

Supplementary Method 4.1 - Pharmacophore and Field Template Generation

We performed pharmacophore modelling through a field-based approach using the align molecule wizard in the Flare module of Cresset software (Cresset Group, 2022). For template generation, the maximum number of conformations was set as 100, conformation hunt and templating calculation methods were set as “Normal” gradient cut-off for conformer minimisation kept at 0.100 kcal/mol/A, and the other parameters were set at default settings. The native ligands of each crystal structure selected earlier were imported from their crystal structure, except for 3CPW, 3DLL, 6DDG, and 6WRU, due to program constraints and high similarity of ligands to other crystal structures (e.g., Isomers). These imported ligands were used as the reference molecules to allow the original positioning of each ligand in its native crystal structure. The template from the FieldTemplater was transferred to Flare and utilised to align the training and test set while building the 3D-QSAR (Quantitative structure-activity relationship) model. For building the 3D-QSAR models, parameters for the conformation hunt were set like that of FieldTemplater. The “very accurate and slow” option was selected for alignment, while the rest of the options remained at the default settings.

5

Chapter 5 — Molecular descriptor-based machine-learning assisted QSAR model development for ribosomal oxazolidinone activity

5.1 Introduction

This chapter covers the second aim of my MPhil project (finetuning, testing and training of ten machine learning algorithms against a dataset of oxazolidinones). The research work included in this chapter is in review for Elsevier *Computers in Biology and Medicine*.

As demonstrated in the previous chapter, structure-based virtual screening methods, such as molecular docking, have provided a cost-effective and convenient in-silico solution in the early stages of RNA drug discovery. However, due to the limitations of performance, due to various factors discussed previously, Ligand-based virtual screening methods, such as ML or QSAR, have become increasingly popular. I have conducted a study that finetuned, trained and tested ten machine learning algorithms against a dataset of oxazolidinones and their descriptors (1875 descriptors (1444 2D and 431 3D)) to measure their accuracy in predicting the pMIC values. Additionally, I analysed the usefulness of Mutual Information based feature selection as a pre-processing technique, using a cut of mutual information score of 0.4 to Top 100, 80, 60, 40 and 20 descriptors. I found that when using all descriptors as a baseline, SVR was the best performing algorithm, with an $R^2 = 0.632$. When repeating the analysis using the Top performing descriptors resulted in the top 80 having the highest SVR score, $R^2 = 0.757$, suggesting that mutual information is a vital pre-processing tool. Nevertheless, incorporating a larger set of more structurally diverse antibiotic datasets will allow for a more thorough investigation of the algorithms.

5.2 Statement of Author Contributions



Statement of Contribution of Co-Authors

The following is the suggested format for the required declaration provided at the start of any thesis chapter which includes a co-authored publication, whether published or unpublished. This is a requirement for all Theses by Publication; and for any Theses by Monograph where the relevant published papers are incorporated into the body of the thesis or comprise a chapter within the thesis.

The authors listed below have certified that:

1. they meet the criteria for authorship and that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
2. they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
3. there are no other authors of the publication according to these criteria;
4. potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit, and
5. they agree to the use of the publication in the student's thesis and its publication on the [QUT's ePrints site](#) consistent with any limitations set by publisher requirements.

In the case of this chapter:

Please state the publication title and date of publication or status:

Molecular descriptor-based machine-learning assisted QSAR model development for prediction of ribosomal oxazolidinone antibiotics – in Review in Elsevier *Computers in Biology and Medicine*.

Contributor	Statement of contribution*
McKenna E. Buckley	Conceptualisation, Methodology, Software, Formal analysis, Data curation, Writing — original draft preparation, Writing—review and editing, Visualisation
Santu Rana	Conceptualisation, Writing—review and editing, Supervision
Kathryn E. Fairfull-Smith	Conceptualisation, Writing—review and editing, Supervision
Neha S. Gandhi	Conceptualisation, Writing—review and editing, Supervision

5.3 Paper in Review

Molecular descriptor-based machine-learning assisted QSAR model development for prediction of ribosomal oxazolidinone antibacterials

McKenna E. Buckley^{1,2}, Santu Rana^{3*}, Kathryn E. Fairfull-Smith^{2,4}, Neha S. Gandhi^{1,2*}

¹ Centre for Genomics and Personalised Health, Queensland University of Technology, 60 Musk Ave, Kelvin Grove, Brisbane, Qld 4059, Australia; (M.E.B.)

² School of Chemistry and Physics, Queensland University of Technology (QUT), Brisbane, QLD, 4000, Australia

³ Applied AIInstitute (A²I²), Deakin University, Geelong, VIC, 3220, Australia

⁴ Centre for Materials Science, Queensland University of Technology, Brisbane, QLD, 4000 Australia, (K.E.F.S.);

* Corresponding authors

Abstract

An ML based approach was developed to predict pMIC activity score for oxazolidinones, a class of synthetic antibacterial agents that act as protein synthesis inhibitors. A range of 2D and 3D descriptors were first extracted based on the SMILES of the derivatives. A Support Vector Regressor (SVR) was chosen as the best-performing predictor out of several choices of regressors and descriptor sets, using the top 80 descriptors. The model was able to achieve a test R² score of 0.76, indicating that it can explain 76% of the variance in the pMIC activity score. Our approach in this work provides a promising tool for predicting the antibacterial activity of oxazolidinones based on their molecular structure, which could be useful in drug discovery and development.

1.0 Introduction

Molecular docking, as a structure-based virtual screening method, offers a cost-effective and convenient approach in the initial stages of RNA drug discovery. By utilizing structural information, molecular docking can predict receptor-ligand recognition and quickly provide useful insights into large chemical libraries (Varela-Rial et al., 2022). However, as

discussed in an earlier molecular docking study (Buckley et al., 2023), there are many limitations to structure-based docking of small molecules to rRNA molecules. Ligand binding sites on RNA can be shallow, more polar, solvated, and conformationally flexible, as RNA is a large, complex, and flexible molecule that can adopt various conformations, which adds further complexity when predicting RNA-ligand interactions (Luo et al., 2019). Many molecular docking studies neglect the effects of water molecules and metal ions, and the binding of the ligand to the corresponding RNA can result in an induced-fit effect or a conformational stabilisation/destabilisation of the tertiary complex structure (Buckley et al., 2023; Tessaro & Scapozza, 2020). Additionally, currently available ribosomal crystal structures have very poor resolution, meaning the electron density of the ligand in the ribosomal binding site is problematic. In general, molecular docking is a useful method for predicting small molecule binding to protein targets, but it may not be as accurate in predicting antibacterial binding to rRNA (Buckley et al., 2023).

QSAR models are computational methods that can be used to uncover relationships between the structural properties of molecules or their derivatives and their biological activities. These models are helpful in prioritizing a large number of chemicals based on their desired biological activities, serving as an *in silico* methodology that significantly reduces the number of candidate chemicals that need to be tested with *in vivo* experiments (Kwon et al., 2019). The QSAR datasets often consist of a vast number of chemicals, sometimes exceeding hundreds of thousands, each represented by various descriptors. Sparse fingerprints are commonly used, resulting in a lot of zero values, and certain characteristics may have a high correlation. Given that relationships are based on *in situ* experiments, it is generally assumed that the dataset contains errors (Kwon et al., 2019). Due to these constraints, it has become difficult for QSAR-based model prediction to achieve a reliable prediction score.

Ligand-based virtual screening methods, such as ML, have become an increasingly popular alternative to QSAR models. ML has evolved as a useful technology in small-molecule drug discovery, with techniques based on relative molecular similarity analysis of compounds with known and unknown activity. ML can be used to aid in the discovery and development of antibiotics in several ways. ML algorithms have been used to analyse large chemical and biological data sets to identify new compounds with antibacterial activity.(Kurczyk et al., 2015; Stokes et al., 2020b) ML was used to predict the activity (Sabet et al., 2012) and toxicity (Khabbaz et al., 2021) of new compounds, which can help

to prioritise which compounds should be synthesised and tested in *in vitro* studies. The identification of proteomic (A. s. Chowdhury et al., 2019) and genomic (Nguyen et al., 2020; Ren et al., 2021) data for antibacterial drug development can also be found through ML. Finally, ML can be used to predict and analyse the resistance of bacteria to certain drugs, which can help to identify new strategies to overcome resistance (Anahtar et al., 2021; A. S. Chowdhury et al., 2020; Pearcy et al., n.d.).

This study aimed to develop an ML based approach to predict pMIC activity scores for the oxazolidinone class of antibiotics. A dataset of derivatives containing structure modifications to the antibiotic linezolid (Buckley et al., 2023; Q. Zhao et al., 2021) from the antibiotic class of protein synthesis inhibitors (oxazolidinones) was used against various algorithms. The dataset was trained and evaluated against various 2D and 3D descriptors, calculated using the PaDEL-descriptor software, generating 1875 descriptors (1444 2D and 431 3D) (Yap, 2011). Additionally, we analysed the usefulness of Mutual Information (MI) based feature selection as a pre-processing technique.

2. Materials and methods

2.1 Data collection

A dataset containing oxazolidinone antibacterial compounds was assembled from the review by Zhao et al., comprising over 30 studies of oxazolidinone scaffolds from the last decade. The data is in MIC values, representing the lowest concentration of an antibiotic at which bacterial growth is completely inhibited, expressed in mg/L (μ g/mL). The molecules were represented by the SMILES. The activity measurements against *S. aureus* were utilised directly in our model and treated as our regression target.

The pMIC values were used to respond to the skewness of the data regarding larger values. MIC values of the dataset ranged from 0.125 μ g/mL to over 200 μ g/mL, with the transformed pMIC being 0.9 to -2.3, respectively. pMIC values were calculated using the equation below.

$$pMIC = -\log_{10}(MIC \text{ value}) \quad 5.1$$

A histogram of pMIC values of the dataset is shown below in Figure 5.1.

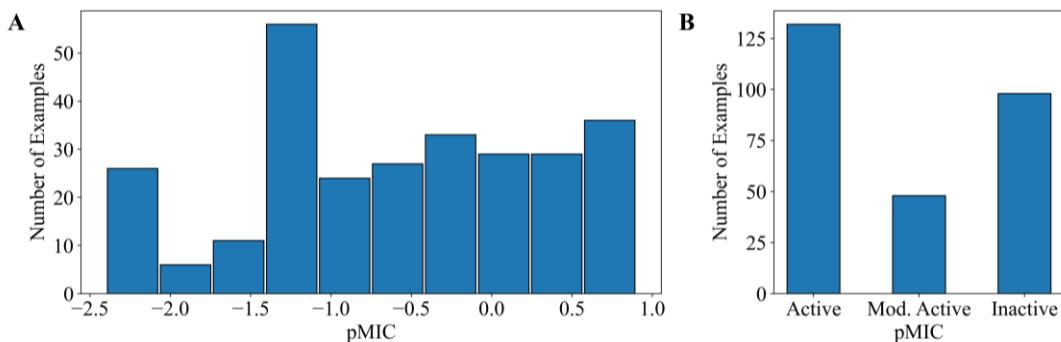


Figure 5.1 - Histogram of *S. aureus* pMIC dataset with the spread of pMIC data (A) and grouped by *S. aureus* pMIC group (B).

2.2 Descriptors

The samples were used for feature extraction in the PaDEL (Yap, 2011) software using the dataset described earlier, calculateing a total of 1875 descriptors (1444 2D descriptors and 431 3D descriptors). The 3D descriptors were calculated using the PaDEL software, utilising the MMFF94 force field.

The testing and training were conducted with stratification to allow an even spread of active, moderately active, and inactive molecules. The final dataset containing the 2D and 3D features for 278 molecules present was divided randomly into 80:20% ratios, from which 222 molecules (80% part) were used for the training set, and the remaining 20% consisted of 56 molecules that were labelled as testing sets. The descriptors were then scaled from -1 to 1, with the descriptors with the highest explained variance ratio used for further testing.

Mutual Information Gain for Regression (MIGR) was performed on the dataset using the scikit-learn package (version 0.23.2). SelectKBest, for regression, was used to determine the Mutual Info Score (MIS) cut-off of 0.4, as well as determining the Top 100, 80, 60, 40, and 20 descriptors. The list of descriptors determined by MIGR are located in Supplementary Table 5.1.

2.3 t-distributed stochastic neighbour embedding (t-SNE) analysis

t-SNE was performed on the dataset using the scikit-learn package (version 0.23.2), specifically sklearn.manifold package, using calculated normalised descriptors and pMIC values of the molecules. Data normalization was performed using the formula below to

bring all the scores in a common scale from -1 to 1, utilising the `sklearn.preprocessing` module.

$$z = (x - u) / s \quad 5.2$$

where:

z is the standardized value.

x is the original value.

u is the mean of the feature values in the training data.

s is the standard deviation of the feature values in the training data.

2.4 Model Construction

Several ML algorithms (K-Neighbors, Ridge, Lasso, Elastic Net, Gradient Boosting, Random Forest, Ada Boost, Extra Trees, Decision Tree, and SVR) were modified from a study conducted by Zhou *et al.* (J. Zhou et al., 2021) in the scikit-learn package (version 0.23.2). A multi-layer perceptron (MLP) was also trained using the PyTorch package (version 1.8.1) with CUDA (version 10.1). MLP input and output size was set at 1, hidden layer 1 and layer 2 were set at 256 and 64, respectively, the learning rate was set to 0.02, the epsilon was left at its default of $1e^{-8}$, and the weight decay was set at 0. The dataset was randomly divided into a training set and a test set with a split of about 80:20 (744:187), stratified by their antibacterial activity (active, moderately active and inactive), using a fixed random seed value before the split to allow for reproducibility. Based on the training set, for each algorithm, a specified pool of hyperparameters was optimised with a five-fold cross-validation strategy to find the combination that achieved minimal loss, maximising the use of a limited dataset in a relatively small dataset (J. Zhou et al., 2021). The algorithms predict continuous variables, so performance was evaluated by the coefficient of determination score (R^2) and the Root of the Mean of the Square of Errors (RMSE). The list of specific algorithms parameters used in the study can be found in Supplementary Table 5.2.

2.5 Prediction

Pearson's correlation calculations were performed using the functions available in the SciPy statistics library in Python 3.6, used for a measurement of linear association between literature value and predicted value. The evaluation of all the descriptors and the Top 100

descriptors was compared by the linear regression between the predicted and known pMIC values of an external dataset.

3. Results and discussion

3.1 Characterisation of the initial dataset

The molecules in the dataset covered a considerable range of *S. aureus* pMIC activity (Figure 5.1), which is essential for ML algorithms to model the QSAR well. Based on the pMIC values for *S. aureus*, 154 compounds (55.4 %) had an activity below 0.6, while 124 compounds (44.6 %) were above 0.6.

To represent the relationship between the structural features and oxazolidinone inhibition in a two-dimensional space, t-SNE was utilized due to its capability of preserving local data structures while providing clustering information. As it can capture a large portion of the local structure of high-dimensional data and reveal global structures like the presence of clusters at various scales, t-SNE is an effective tool for visualizing the mapping between the two variables (Maaten & Hinton, 2008).

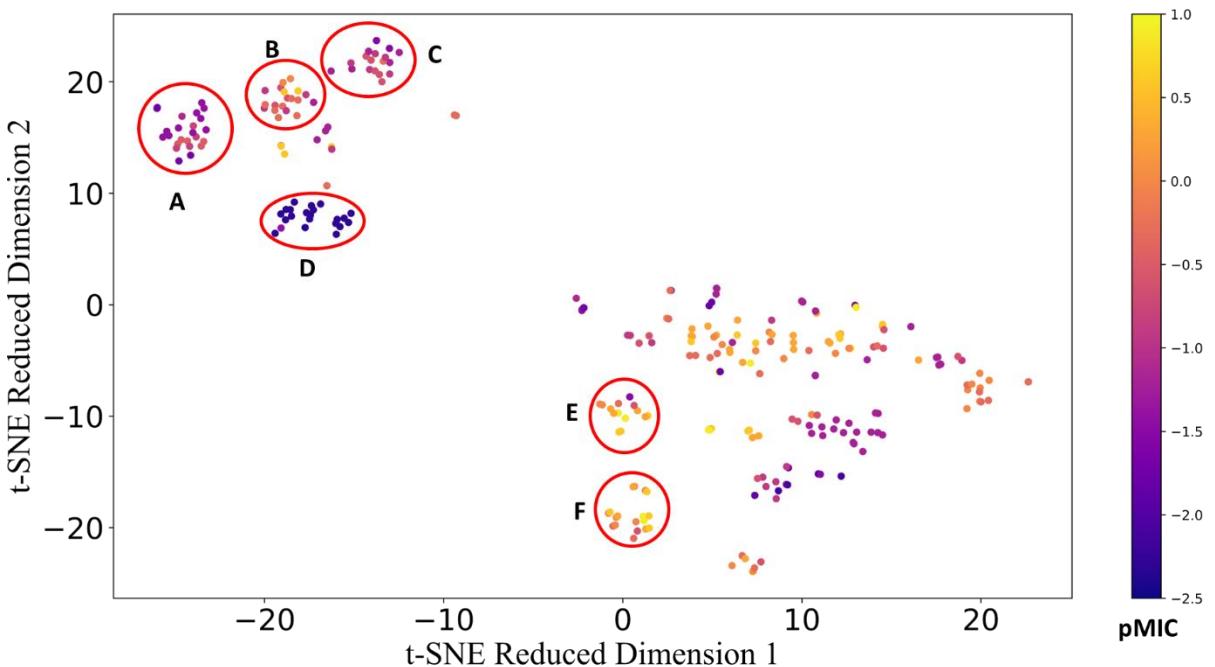


Figure 5.2 – Dimensionality reduction by t-SNE on fingerprint bit vectors. The colour bar indicates pMIC values from -2.5 (dark purple) to 1 (yellow).

As ML heavily relies on the dataset, a more diverse dataset will create a more accurate model. The presence of several clusters of molecules with almost identical core structures and similar activities, as shown by the groupings of similar colours in Figure 5.2, provides

compelling evidence for a QSAR. In Figure 5.2, A – F are marked as clusters with a clear separation from other data points. While there is a common substructure between all the molecules (oxazolidinone), there is good structure diversity, as shown in Figure 5.3. Cluster B is the only cluster structure that doesn't alter the side chain; Clusters D and E both share the 3-fluorophenyl ring, shown in linezolid. Due to the differences, we decided to conduct further ML analysis on said clusters to determine whether the minute changes would affect the performance.

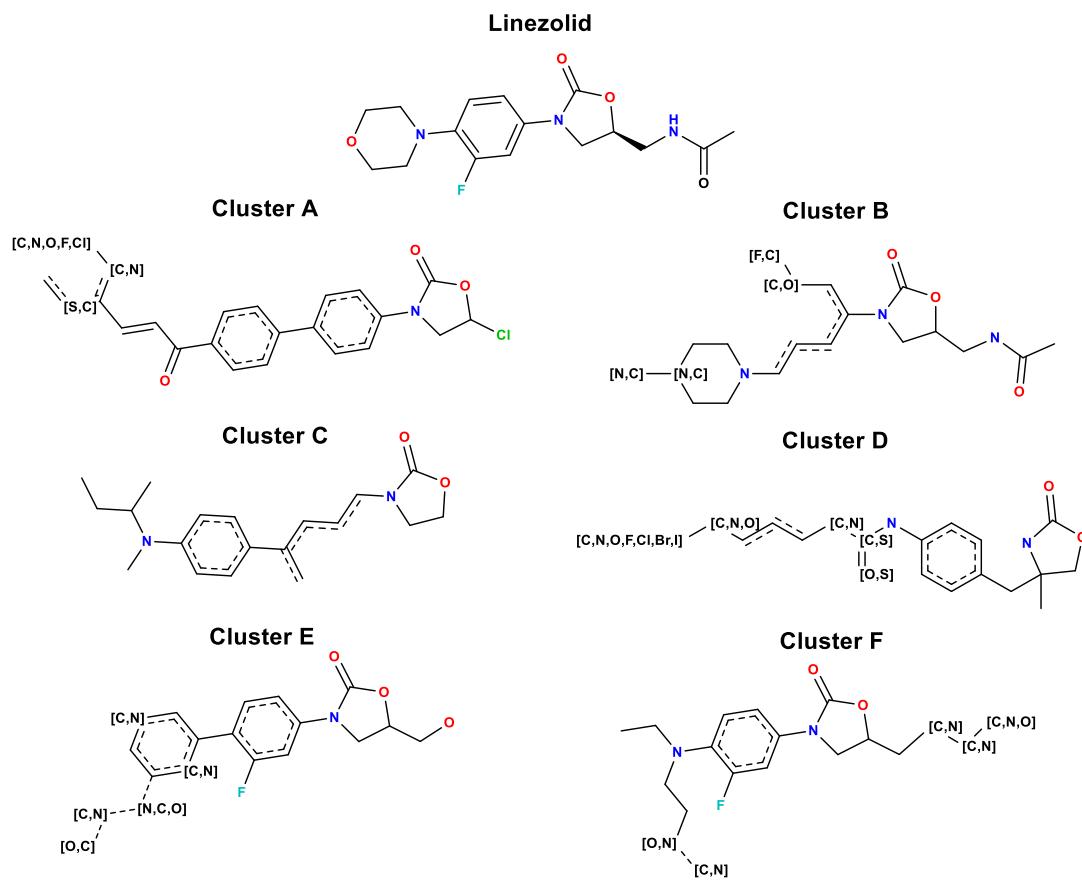


Figure 5.3 – Core structures of molecules in Clusters A, B, C, D, E and F from the t-SNE analysis, shown in (Cluster A-F), respectively. Most core structures, except for Cluster C, have sections showing [X, X]. This indicates that a structure in that group has the same core structure but a slight alteration in the element.

3.2 Performance of the ML algorithms

Several commonly used ML algorithms were applied for comparison and evaluation, with the models adapted from the Zhou et al. (2021) database (J. Zhou et al., 2021). The ML algorithms were trained on the descriptors of the structures. Principal component analysis (PCA) was performed to calculate the variance between antibacterial activity (active,

moderately active, and inactive) molecules with the help of descriptors as input features. However, PCA results performed worse than just the descriptor values, so it was not explored further as a prediction method.

3.2.1 pMIC and Cluster group analysis

The ML algorithms were analysed by being separated into the pMIC groups: active (132 derivatives), moderately active (48 derivatives), and inactive (98 derivatives), to determine whether an actively only pMIC group would allow the algorithm to perform better than the moderately active and inactive. Inactive was, by far, the best performing (Figure 5.4), with an average $R^2 = 0.705$, followed by active ($R^2 = 0.086$) and moderately active ($R^2 = -0.351$). It would be expected that those with activity pMIC would perform better. However, the opposite was observed. Both active and moderately active performed quite poorly, with an R^2 score below 0.25 or even in negative values. These results state that the pMIC values do indeed affect how well the ML algorithm performs. However, due to the small dataset, this would have to be explored further to determine whether this is a consistent trend with other bacteria strains.

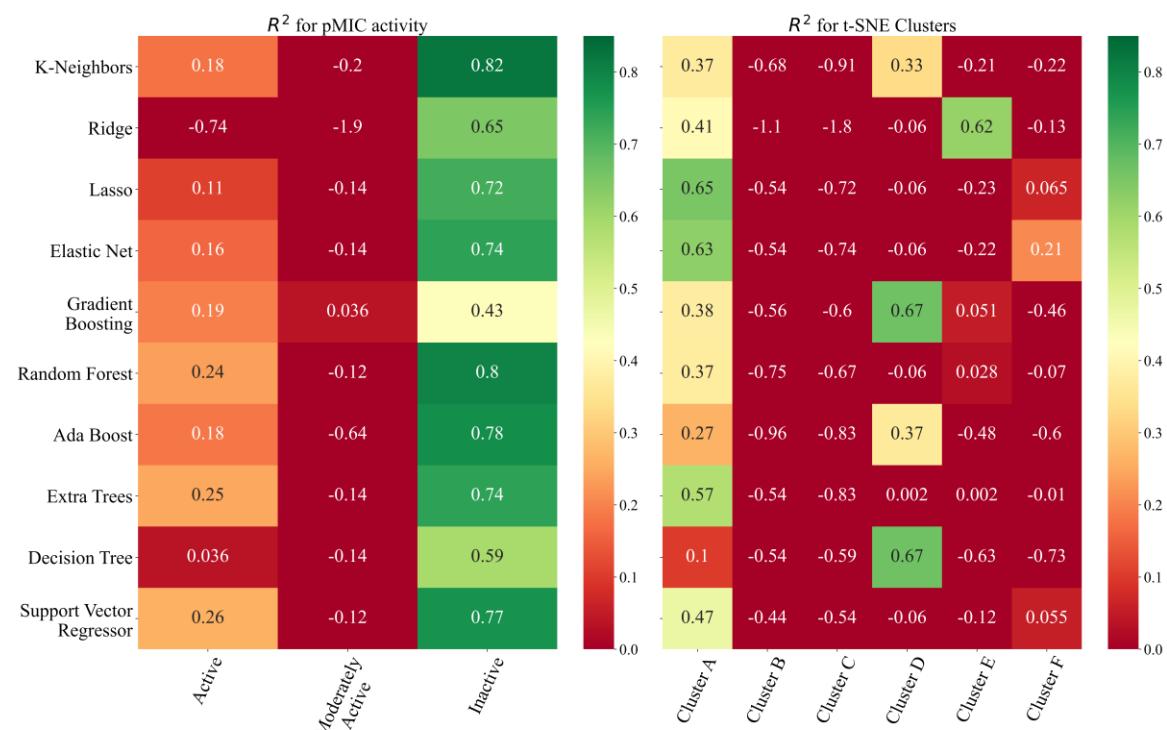


Figure 5.4 – Mean performance of the algorithms of predicting the pMIC separated by the pMIC activity groups (left) and clusters determined by the t-SNE (right), evaluated by the

coefficient of determination (R^2) test score. The table is shaded by order of performance, the lowest being red and the highest being dark green.

Each ML algorithm was tested on the cluster groups determined by the t-SNE analysis above Cluster A (23 derivatives), Cluster B (19 derivatives), Cluster C (17 derivatives), Cluster D (22 derivatives), Cluster E (14 derivatives), and Cluster F (18 derivatives). As each group have relatively different largest common substructures (as much as possible within the same antibiotic class), this analysis was done to determine whether specific structural modifications would alter the results. Cluster A had the highest average performance, with $R^2 = 0.42$, followed by Cluster D ($R^2 = 0.17$), Cluster E ($R^2 = -0.12$), Cluster F ($R^2 = -0.19$), Cluster B ($R^2 = -0.67$), and Cluster C ($R^2 = -0.82$). Cluster A, Cluster D, and Cluster E were the only cluster to produce results over $R^2 = 0.55$.

The majority of the results were incredibly poor, with many having negative R^2 scores. This suggests that reducing the results down to clusters, resulting in much smaller datasets, drastically reduces the performance of the algorithms across the board. Additionally, even though there is some structural diversity, the clusters have very similar core structures due to them being in the same class of molecules. In future, it is suggested to incorporate more classes of antibiotics that have different core structures to increase the diversity and size of the dataset to improve the results.

3.2.2 Top descriptors

Using all of the calculated descriptors, MIGR was used to determine the “top performing” descriptors, split into Top 20, 40, 60, 80 and 100. A cut-off MIS value of 0.4 was determined based on the results of the MIGR as a comparison, resulting in 116 descriptors used. Each of these descriptor’s algorithms was applied for comparison and evaluation, with the data stratified using the ranges of pMIC activity (active, moderately active, and inactive).

All algorithms were initially tested with a variety of parameters to determine which would perform best for each descriptor set. Many of the parameters were not “one size fits all” or consistently the same across each descriptor set, and so each had to be adjusted and fine-tuned with the different Top descriptor datasets. The fine-tuned algorithm with the highest Train R^2 on the test set was deployed for each algorithm and descriptor set chosen. The predictive ability of each algorithm and each descriptor dataset, shown in Figure 5.5, was evaluated by the test and train R^2 .

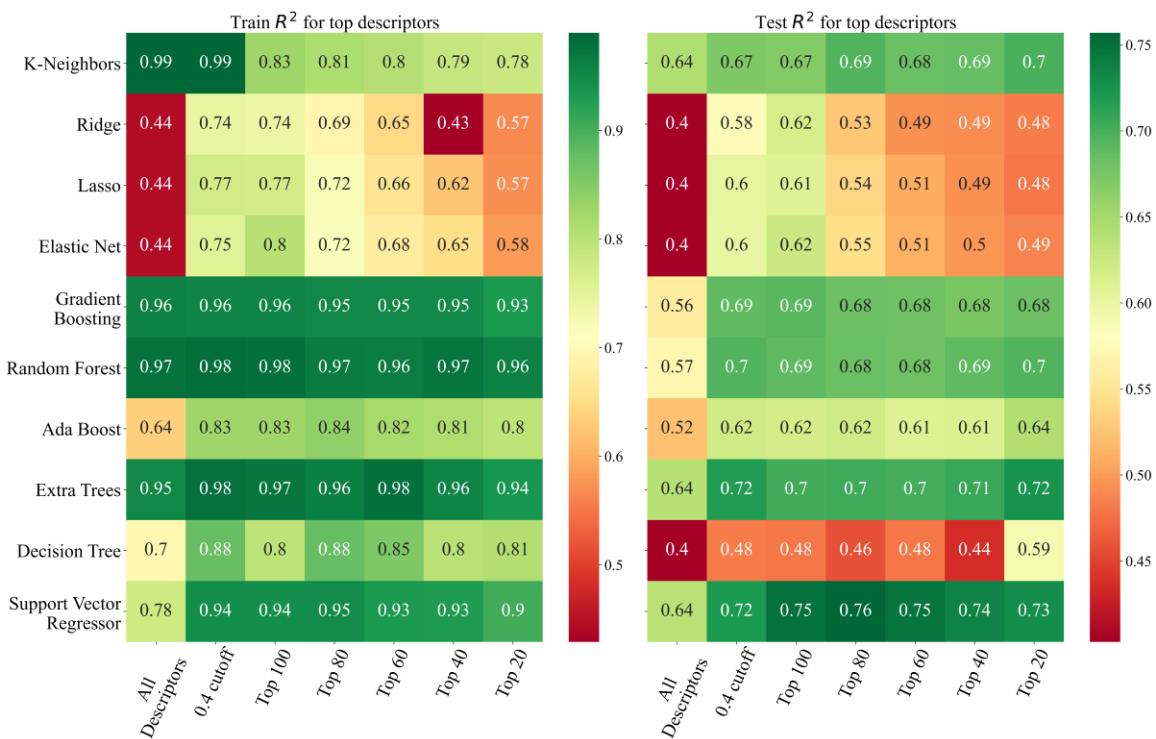


Figure 5.5 – Mean performance of the algorithms of predicting the pMIC value of training data and test data, evaluated by the coefficient of determination (R^2) test score. The table is shaded by order of performance, the lowest being red and the highest being dark green.

Using all the descriptors as a baseline of descriptor performance resulted in a relatively poor test R^2 score, an average for test $R^2 = 0.518$. The poor score indicates that some descriptors may not be necessary and counteract the performance, which could explain the train R^2 overfitting in the more “complex” models (k-NN, GB, RF, and ET) and underfitting with the generally considered “simpler” methods (Ridge, Lasso, and EN).

While the SVR performed relatively well with test $R^2 = 0.64$ when using all the descriptors, we wanted to see if we could improve this score. Therefore, the descriptors were narrowed down to the top contributing, using mutual information cut-off value of 0.4 and the top 100, 80, 60, 40, and 20 descriptors (as shown in Figure 5.5). The 0.4 mutual information cut-off was used as an initial value to determine whether reducing the number of descriptors would increase the performance of the algorithms. When reduced to 0.4 MIS, Ridge, Lasso, and EN train and R^2 jump up significantly. This proved true, as the average performance was substantially increased compared to using all descriptors, $R^2 = 0.638$.

The Top 100 descriptors gave the best performance with the test R^2 value above 0.6 for all algorithms (besides the DT, at $R^2 = 0.48$). While it did not have the highest overall R^2 , it

had the highest overall performance of test $R^2 = 0.645$. This indicates that the Top 100 would be the best dataset of all descriptors to use well, regardless of the algorithm used. The Top 80 produced the highest test R^2 , with an SVR $R^2 = 0.757$, with an overall average of test $R^2 = 0.621$. The Top 60 has the second lowest overall score, with $R^2 = 0.609$. The Top 40 had the lowest average performance, with a test $R^2 = 0.602$. Lastly, the Top 20 had a similar average performance to the top 80, with an overall test average of $R^2 = 0.621$. k-NN, GB regression, RF regressor, ET and SVR consistently performed well, regardless of the number of descriptors used, with the majority test score above $R^2 = 0.69$.

The predictive ability of each algorithm, shown in Figure 5.6, was evaluated by R^2 , the Mean of Absolute value of Errors (MAE), and the RMSE. The SVR algorithm, was used as the comparison algorithm, as it consistently performed the best regardless of the descriptor type. The RMSE, MAE, and mean values of R^2 and their standard deviations were calculated to assess the stability of the algorithms over different train-test splits of the dataset. For an ideal model, RMSE/MAE should be close to 0 and R^2 score close to 1. Figure 5.6 displays the SVR scores for each top descriptor, compared to the SVR when using Morgan Bits and all the descriptors. Using all the descriptors performed the lowest, with $R^2 = 0.637$. In regard to RMSE, the 0.4 MIS had the highest of 0.452, while the lowest was the Top 80 descriptors at 0.421. While there are differences, they are all similar in performance and have minimal variation across all the descriptors. This is a further indication that using Mutual Information to determine the highest-performing descriptors is a very useful tool. For this study, the Top 80 dataset was used for further comparison due to its highest performance.

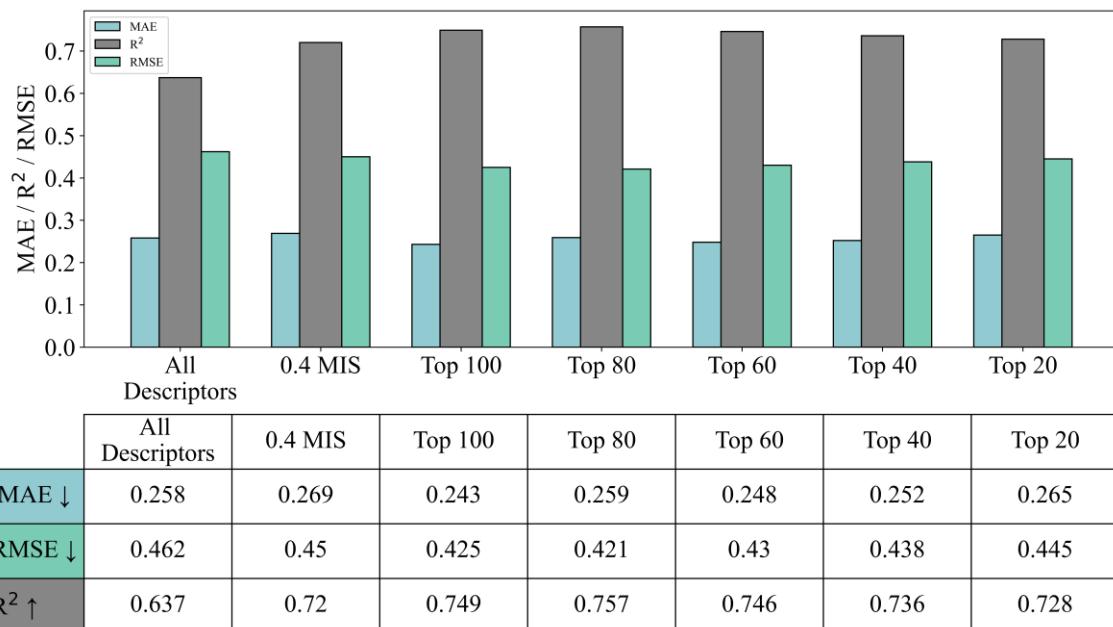


Figure 5.6 – Mean performance of SVR on predicting the pMIC value of *S. aureus* using various amounts of descriptors, evaluated by MAE, R^2) and RMSE. The down arrows indicate that a lower score is better, while the up arrow indicates a higher score is better.

3.3 Validation with the external dataset

To test the algorithms, we used oxazolidinone molecules that had known activity (whether active or inactive) that were not used in the training and testing. Rivaroxaban (Figure 7, 2) is similar to the antibiotic linezolid (Figure 5.7, 1), with both drugs sharing the same oxazolidinone-derived core structure. Rivaroxaban was studied for any possible antimicrobial effects and the possibility of mitochondrial toxicity, a known complication of long-term linezolid use (A. K. Singh et al., 2020). However, no antibiotic activity was shown, so it is currently used as an anticoagulant medication (blood thinner) used to treat and prevent blood clots. We placed this molecule into the dataset as a baseline for the prediction. As we know, it has no activity, so the algorithm should predict it to have poor activity. The second part of the virtual screening dataset was from a study conducted by Jin et al. (2022) on 3-(3-pyridyl)-oxazolidinone-5-methyl ester derivatives (Figure 5.7, 3).(Jin, Chen, et al., 2022) Additionally, Jin et al.(2022) designed a set of 3-(pyridine-3-yl)-2-oxazolidinone derivatives in a separate study (Figure 5.7, 4) (Jin, Wang, et al., 2022). These two datasets of molecules have known inactivity against *S. aureus* (between 16 to above 256 MIC), allowing for a good range of diversity in the data. The third part of the virtual screening dataset was a set of piperazinylphenyloxazolidinone's (Figure 5.7, 5)(Das et al.,

2005) and arylpiperazinyloxazolidinones with the diversification of the N-substituents (Figure 5.7, 6) (Jang et al., 2004), collated in a QSAR study by Gandhi (2007) (Gandhi, 2007).

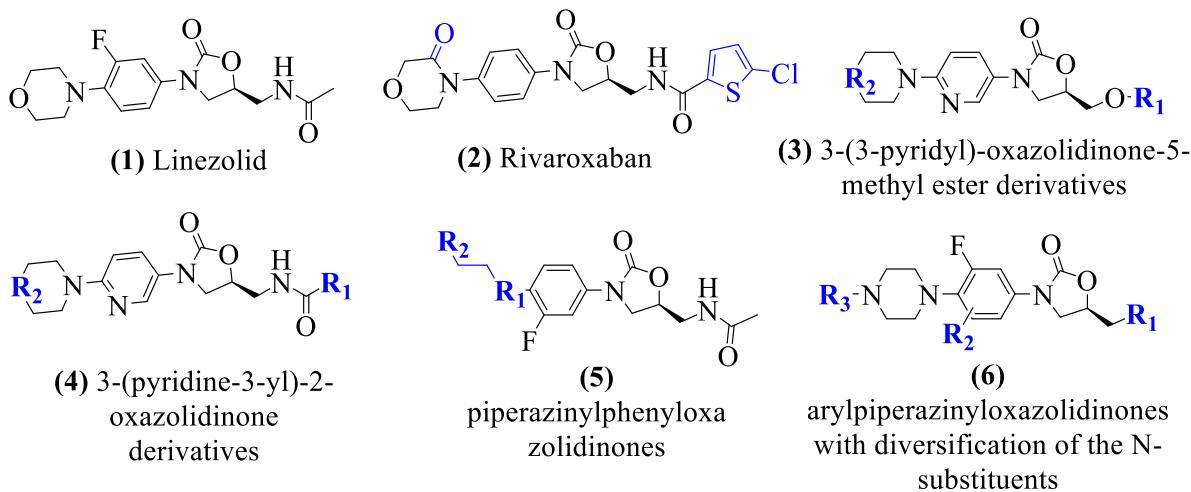


Figure 5.7 – Structural breakdown of virtual screening dataset, with Linezolid (1) shown for baseline structural comparison. The dataset is broken down into the anticoagulant Rivaroxaban (2), 3-(3-pyridyl)-oxazolidinone-5-methyl ester derivatives (3) (Jin, Chen, et al., 2022), 3-(pyridine-3-yl)-2-oxazolidinone derivatives (4) (Jin, Wang, et al., 2022), piperazinylphenyloxazolidinones (5) (Das et al., 2005; Gandhi, 2007), arylpiperazinyloxazolidinones with the diversification of the N-substituents (6) (Gandhi, 2007; Jang et al., 2004), and other derivatives from the Zhao et al. dataset that did not have any recorded MIC values against *S. aureus* (Q. Zhao et al., 2021). All structures are drawn in ChemDraw.

The deployed model was based on the SVR, the best-performing algorithm. We then compared the prediction of SVR, comparing the prediction between using all the descriptors and the Top 80 descriptors (Figure 5.8) for predicting pMICs outside the primary dataset. Both are measured using Pearson's correlation. This correlation is the linear relationship between two continuous variables ranging from (-1, 1). Coefficients greater than zero represent a positive trend, and coefficients less than zero represent an inverse relationship. Using all the descriptors resulted in a negative correlation of -0.333 and a p value of 0.004, while the Top 80 had a positive correlation of 0.356 and a p value of 0.002. Rivaroxaban, an oxazolidinone known to not have any antibacterial activity, was predicted to have a predicted pMIC value of -0.940 by all descriptors and -0.830 by the top 80 (equivalent to MIC values of approximately 8 and 7, respectively). This suggested that

both descriptor sets can only be used as an indicator of activity range, not a specific activity performance.

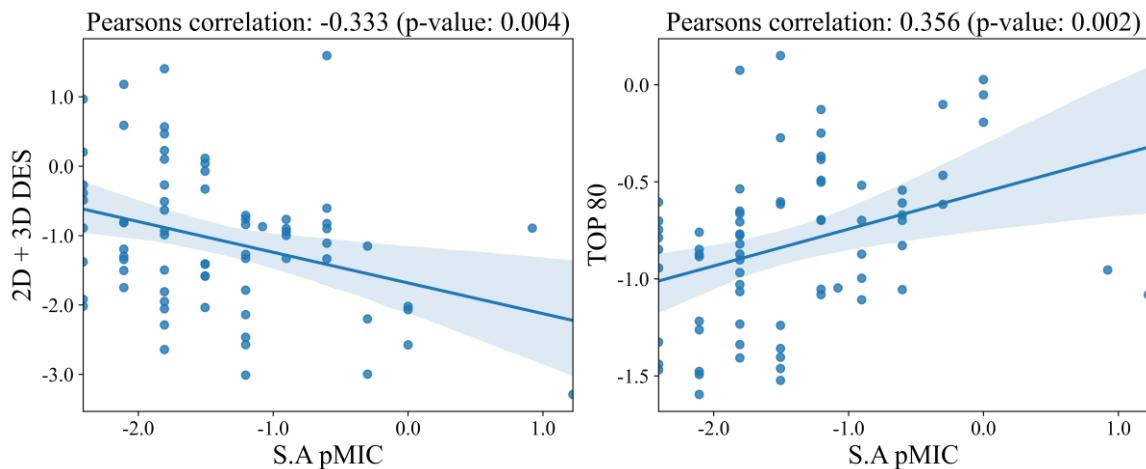


Figure 5.8 – *S. aureus* pMIC against the predicted values for the external dataset trained off All the descriptors, 2D and 3D descriptors, top 80 descriptor dataset. Each graph has Pearson's correlation indicated.

4. Conclusion

In this study, an ML model was analysed using an oxazolidinone dataset from Zhao et al. (2021) to analyse a set of algorithms and the usage of 2D and 3D descriptors. The best-performing algorithm, SVR, reached an average coefficient of determination (R^2) on the test set of 0.632 on the primary dataset using all 1440 descriptors. MIGR was used to calculate and reduce the number of descriptors used, ranging from the cut of mutual information score of 0.4 to Top 100, 80, 60, 40 and 20 descriptors. Evaluations on subsets of descriptors from the primary datasets illustrated that the performance of SVR was more stable than other algorithms. Repeating the analysis using the Top performing descriptors resulted in the top 80 having the highest SVR score, $R^2 = 0.757$. This also suggests that using mutual information greatly improves the performance compared to using all descriptors as is. However, each dataset will have different performances, so consideration must be taken when factoring in the number of descriptors used. In future works, we plan to continue evaluating these methods with a more extensive and structurally diverse dataset.

Author Contributions: Conceptualisation: M.E.B., K.E.F.S., S.R., and N.S.G; Methodology: M.E.B.; Software: M.E.B.; Formal analysis: M.E.B.; Data curation: M.E.B.; Writing – original draft preparation: M.E.B.; Writing-review and editing: M.E.B., K.E.F.S.,

S.R., N.S.G; Visualisation: M.E.B.; Supervision, K.E.F.S., S.R., and N.S.G.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding

Data Availability Statement: All machine learning algorithms are available from <https://github.com/JiajunZhou96/ML-for-LSD1>. All prepared oxazolidinone scaffolds from the dataset are available at <https://github.com/gandhi-group-biomolecular/machine-learning-oxazolidinone-dataset>

Acknowledgments: The eResearch Office, Queensland University of Technology, Brisbane, Australia, provided the computational resources and services used in this work. This study was supported by funding from the QUT Centre of Genomics and Personalised Health.

Conflicts of Interest: The authors declare no conflict of interest

Abbreviations Used

AATS	Atom-Type E-state
BCUT	Burden-Centered Fragments
DT	decision tree
MAE	Mean of Absolute value of Errors
MIC	minimum inhibitory concentration
MIGR	Mutual Information Gain for Regression
MIS	Mutual Info Score
ML	machine learning
PC	principal components
PCA	principal component analysis
pMIC	log of minimum inhibitory concentration values
QSAR	Quantitative structure activity relationship
R ²	coefficient of determination score
RMSE	The root of the Mean of the Square of Errors
RNA	Ribonucleic acid
<i>S. aureus</i>	<i>Staphylococcus aureus</i>
SMILES	Simplified Molecular-Input Line-Entry System
SNE	Stochastic Neighbor Embedding

SVR support vector regressor
t-SNE t-distributed Stochastic Neighbour Embedding
VS Virtual screening

Supporting Information

Supplementary Table 5.1 – Top descriptors determined from mutual information

				SpMax1_Bhi, SpMax3_Bhi, SpMin1_Bhi, SpMin2_Bhi, SpMin3_Bhi, SpMin3_Bhs, SpMin4_Bhs, SpMin8_Bhs	SpMin4_Bhp, SpMax1_Bhi, SpMax3_Bhi, SpMin1_Bhi, SpMin2_Bhi, SpMin3_Bhi, SpMin4_Bhi, SpMin3_Bhs, SpMin4_Bhs, SpMin8_Bhs	SpMin3_Bhp, SpMin4_Bhp, SpMax1_Bhi, SpMax3_Bhi, SpMin1_Bhi, SpMin2_Bhi, SpMin3_Bhi, SpMin4_Bhi, SpMin3_Bhs, SpMin4_Bhs, SpMin3_Bhs
Connectivity			C2SP2, nTRing	C2SP2, nTRing	C2SP2, nTRing, nT5Ring	C1SP2, C2SP2, nTRing, nT5Ring
Substructure count	SCH-5, ASP-4, ASP-5, ASP-7, VP-7	SCH-5, VCH-5, VCH-6, VC-5, ASP-4, ASP-5, ASP-7, VP-7	SCH-5, SCH-7, VCH-6, VC-5, ASP-4, ASP-5, ASP-6, ASP-7, VP-7	SCH-5, SCH-7, VCH-5, VCH-6, VC-5, ASP-4, ASP-5, ASP-6, ASP-7, VP-6, VP-7	SCH-5, SCH-7, VCH-5, VCH-6, VC-5, ASP-3, ASP-4, ASP-5, ASP-6, ASP-7, VP-6, VP-7	SCH-5, SCH-7, VCH-5, VCH-6, VC-5, ASP-3, ASP-4, ASP-5, ASP-6, ASP-7, VP-6, VP-7
Crippen's logP		CrippenLogP	CrippenLogP	CrippenLogP	CrippenLogP	CrippenLogP
Constitutional	nssssCH, SHCsats, SdsCH, SsssCH, minHBa, minsssCH, hmax, hmin, nTRing, SRW9	nHBint4, nssssCH, SHCsats, SdsCH, SsssCH, minHBa, minsssCH, hmax, hmin, ETA_BetaP_s, ZMIC1, nAtomP, MDEN-23, R_TpiPCTPC, nTRing, SRW9	nHdsCH, nHother, ndsCH, nssssCH, SHBint4, SHCsats, SdsCH, SsssN, minHBa, minsssCH, ETA_BetaP_s, ZMIC1, nAtomP, MDEN-23, R_TpiPCTPC, nTRing, SRW9	nwHBa, nwHBint4, nHCsats, nssssCH, nsssN, SHBint4, SHBs, SHBint5, SHssNH, SHCsats, SdsCH, SsssCH, SssNH, SsssN, nHBAcc, nHBAcc_Lipi, nski, ZMIC1, nAtomP, MDEN-23, MPC8, piPC8, piPC9	nwHBa, nwHBint4, nHother, nssssCH, nsssN, SHBint4, SHBs, SHBint5, SHssNH, SHCsats, SdsCH, SsssCH, SssNH, SsssN, nHBAcc, nHBAcc_Lipi, nski, ZMIC1, nAtomP, MPC8, piPC8, piPC9	nwHBa, nHBint4, nHCsats, ndsCH, nssssCH, SHBint4, SHBs, SHBint5, SHssNH, SHCsats, SdsCH, SsssCH, SssNH, SsssN, nHBAcc, nHBAcc_Lipi, nski, ZMIC1, nAtomP, MPC8, piPC8, piPC9

			R_TpiPCTPC, nTRing, SRW9			nski, ZMIC1, nAtomP, MPC8, piPC8, piPC9	
Topologic al				minHBa, minHBint4, minHCsats, minssCH2, minssCH, minssN, maxHBint4, maxssCH2, hmax, hmin, gmin, ETA_BetaP_s, R_TpiPCTPC, TopoPSA, SRW9, WTPT-3, TDB6s	minHBa, minHBint4, minHCsats, minssCH2, minssCH, minssN, maxHBint4, maxssCH2, hmax, hmin, gmin, ETA_BetaP_s, R_TpiPCTPC, TopoPSA, TDB5p, TDB6s	minHBa, minHBint4, minHCsats, minssCH2, minssCH, minssN, maxHBa, maxHBint4, maxssCH2, hmax, gmax, hmin, gmin, MAXDP, ETA_BetaP_s, MDEC-22, MDEN-23, R_TpiPCTPC, TopoPSA, TDB5p, TDB6s	minHBa, minHBint4, minHCsats, minssCH2, minssCH, minssN, maxHBa, maxHBint4, maxssCH2, hmax, gmax, hmin, gmin, MAXDP, ETA_BetaP_s, MDEC-22, MDEN-23, R_TpiPCTPC, TopoPSA, TDB5p, TDB6s
Hydrogen bond acceptor			nHBAcc_Lipi nski	nHBAcc, nHBAcc_Lipi nski	nHBAcc, nHBAcc2, nHBAcc_Lipi nski	nHBAcc, nHBAcc_Lipi nski	
Hybridiza tion ratio				HybRatio	HybRatio	HybRatio	
Radial distributio n function				RDF55m	RDF55m	RDF55m	
Walk and path count					WTPT-3	WTPT-3	

Supplementary Discussion 5.1 - Types of descriptors used

The different types of descriptors discussed herein collectively contribute to the understanding of how they impact the performance of antibiotics. This discussion will provide a detailed exploration of the following descriptor types: ALOGP, AATS, BCUT, Burden modified eigenvalue, Connectivity, Substructure count, Crippen's logP,

Constitutional, Topological, Hydrogen bond acceptor, Hybridization ratio, Radial distribution function, and Walk and path count.

ALOGP is a measure of the partition coefficient of a molecule between a hydrophobic and hydrophilic phase. It provides information about the molecule's lipophilicity, which is crucial for antibiotic absorption and distribution in the body. High ALOGP values indicate higher affinity for lipid-based environments and can potentially affect antibiotic penetration through cell membranes.

Atom-Type E-state or AATS descriptors quantify the electronic properties of atoms in a molecule. They can provide insights into the distribution of electronic charges and the reactivity of different atom types. AATS descriptors may help in understanding the molecular interactions of antibiotics with their target sites or with enzymes involved in antibiotic resistance.

Burden-Centered Fragments or BCUT descriptors are used to characterize molecular fragments based on atom-centered properties. These descriptors can provide information about the molecular shape, size, and distribution of different functional groups. In the context of antibiotics, BCUT descriptors can aid in understanding the structural features associated with antibiotic activity or the recognition of specific binding sites.

Burden modified eigenvalue descriptor is related to the eigenvalues of the Burden matrix, which represents the molecular graph. It provides information about the topological properties of a molecule, such as the connectivity and branching patterns. In the context of antibiotics, Burden modified eigenvalues can help identify molecular features associated with antibacterial activity or structural motifs important for drug-target interactions.

Connectivity descriptors quantify the connectivity or degree of interaction between atoms in a molecule. They provide information about the molecular framework and can help identify key atoms or functional groups involved in antibacterial activity or interactions with biological targets.

Substructure count descriptors involve counting the occurrence of specific substructures or patterns within a molecule. They can help identify molecular motifs or pharmacophores associated with antibacterial activity or structural features important for antibiotic potency.

Crippen's logP is a measure of the partition coefficient of a molecule between octanol and water. It provides information about the molecule's hydrophobicity and its ability to cross

biological membranes. Hydrophobicity is an important factor in the absorption, distribution, and bioavailability of antibiotics.

Constitutional descriptors describe the molecular constitution or composition, including the number and types of atoms, bonds, and functional groups. They can provide insights into the chemical diversity and complexity of antibiotic molecules.

Topological descriptors quantify the molecular topology, such as the number of rings, cycles, or branches. They can help identify structural features related to antibiotic activity or predict the behavior of a molecule in biological systems.

Hydrogen bond acceptor descriptor counts the number of hydrogen bond acceptor sites in a molecule. Hydrogen bonding is an essential molecular interaction involved in antibiotic binding to target sites or in the recognition of antibiotic-resistant enzymes.

Hybridization ratio descriptors provide information about the hybridization state of atoms in a molecule. They can help predict the reactivity and stability of different functional groups. In the context of antibiotics, hybridization ratios can aid in understanding the chemical reactivity and potential metabolic transformations.

The radial distribution function characterizes the distribution of atoms around a central atom. It provides information about the spatial arrangement and organization of atoms in a molecule. In the context of antibiotics, this descriptor can help analyze the local environment of key atoms or functional groups important for drug-target interactions.

Walk and path count descriptors involve counting the number of specific types of walks or paths in a molecular graph. They provide information about the connectivity patterns and can aid in the identification of structural features associated with antibiotic activity or molecular recognition.

Supplementary Table 5.2 Machine learning parameters used in the study. After some initial testing of algorithms, it was decided to make some adjustments and modifications to the algorithm parameters used in the original study. These additional modifications are indicated in bold.

Algorithm	Parameters
Kneighbors Regressor	<ul style="list-style-type: none"> ‘n_neighbors’: 2, 3, 4, 5, 6, 7, 8, 9, 10 ‘weights’: ‘uniform’, ‘distance’ ‘algorithm’: ‘auto’, ‘ball_tree’, ‘kd_tree’, ‘brute’ ‘leaf_size’: 20, 30, 40 ‘p’: 1, 2
Ridge	<ul style="list-style-type: none"> ‘alpha’: 0.001, 0.01, 0.1, 1, 10 ‘solver’: ‘auto’ ‘max_iter’: 200, 500 (from the original 100000)
Lasso	<ul style="list-style-type: none"> ‘alpha’: 0.001, 0.01, 0.1, 1, 10 ‘selection’: ‘cyclic’, ‘random’ ‘max_iter’: 200, 500 (from the original 100000)
ElasticNet	<ul style="list-style-type: none"> ‘alpha’: 0.001, 0.01, 0.1, 1, 10 ‘l1_ratio’: 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 ‘max_iter’: 100000
Gradient Boosting Regressor	<ul style="list-style-type: none"> ‘learning_rate’: 0.001, 0.01, 0.1, 1 ‘min_samples_split’: 2, 3, 4, 5, 6, 7, 8, 9 ‘loss’: ‘ls’, ‘lad’, ‘huber’, ‘quantile’ ‘criterion’: ‘mse’, ‘friedman_mse’
Ada Boost Regressor	<ul style="list-style-type: none"> ‘learning_rate’: 0.001, 0.01, 0.1, 1 ‘loss’: ‘linear’, ‘square’, ‘exponential’
Extra Trees Regressor	<ul style="list-style-type: none"> ‘bootstrap’: True, False ‘min_samples_split’: 2, 3, 4, 5, 6, 7, 8, 9, 20, 30, 40, 50, 60, 70, 80, 90 (Increased the sample split values to address overfitting issues)
Random Forest Regressor	<ul style="list-style-type: none"> ‘bootstrap’: True, False ‘max_features’: ‘auto’, ‘log2’, ‘sqrt’ ‘min_samples_split’: 2, 3, 4, 5, 6, 7, 8, 9, 20, 30, 40, 50, 60, 70, 80, 90 (Increased the sample split values to address overfitting issues)

Decision Tree Regressor	<p>'criterion': 'mse', 'friedman_mse'</p> <p>'min_samples_split': 2, 3, 4, 5, 6, 7, 8, 9, 20, 30, 40, 50, 60, 70, 80, 90</p> <p>(Increased the sample split values to address overfitting issues)</p>
SVR	<p>'kernel': 'rbf', 'linear', 'poly', 'sigmoid'</p> <p>'gamma': 'scale', 'auto'</p> <p>'C': 0.001, 0.01, 0.1, 1, 10, 20, 50, 100, 100, 200, 500, 1000, 2000, and 5000 (Increased the C value by a factor of 10, to see if the R² values would improve)</p>

6

Chapter 6 — Conclusions and future work

6.1 Summary and concluding remarks

This thesis presented a computational framework to identify and analyse ribosomal antibacterial with oxazolidinone scaffold through two main methods, molecular modelling and ML. The first part of the thesis (Chapter 4) serves as a two-part study: (1) as a benchmarking study for a variety of RNA-developed molecular modelling programs and (2) as a virtual screening study on a novel dataset of linezolid-based oxazolidinone scaffolds and derivatives. The second part (Chapter 5) is an ML study analysis of the performance of a variety of ML algorithms as well as the potential of using scaled descriptors. The conclusions of the two studies are summarised below:

Chapter 4: Comparative Assessment of Docking Programs for Docking and Virtual Screening of Ribosomal Oxazolidinone Antibacterial Agents

- Out of the five programs tested, AutoDock, AutoDock Vina, and DOCK 6 tended to perform better when redocking the native ligands, compared to rDock and RLDOCK.
- The structural modifications and class of MIC activity of the 285 oxazolidinone dataset had minimal to no influence on how the linezolid derivatives performed when docked against the *S. aureus* crystal structure.
- We found ten derivatives that score within the “top-performing” range, with 3 binding in the native linezolid binding site.
- Due to the complexity of rRNA, such as the high flexibility of the binding pocket resulting in poor pose prediction and poor posing of the derivatives, results must be interpreted with caution. For this reason, the study can only guide docking-based virtual screens of oxazolidinone targets, and all limitations should be considered when conducting similar studies.

Chapter 5: Molecular descriptor-based machine-learning assisted QSAR model development for ribosomal oxazolidinone activity.

- ML models were analysed using an oxazolidinone dataset from Zhao et al. (2021) to analyse a set of algorithms. A total of 1440 1D, 2D and 3D descriptors were calculated for each compound and were narrowed down using Mutual Information.

- Mutual information was used to calculate and reduce the number of descriptors used, ranging from the cut of mutual information score of 0.4 to Top 100, 80, 60, 40 and 20 descriptors.
- Using mutual information dramatically improves the performance compared to using all descriptors as is, with the Top 80 descriptors having the highest performance.
- Evaluations on subsets of descriptors from the primary datasets illustrated that the performance of Support Vector Regression was more stable than other algorithms: K-Neighbors, Ridge, Lasso, EN, GB, RF, Ada Boost, ET, and DT.

6.2 Contributions to the literature

The contribution of the thesis towards oxazolidinone literature can be classified into four aspects, (1) the validation of ribosomal adaptable molecular docking programs, (2) the evaluation of eleven ribosomal crystal structures, (3) the virtual screening of a dataset of linezolid scaffold-based oxazolidinones, and (4) linezolid scaffold descriptor-based ML algorithm testing.

The thesis describes the use of a novel antibacterial oxazolidinone dataset to validate and test various molecular docking programs (in conjunction with eleven 50S ribosomal crystal structures) and ML algorithms. In past studies, oxazolidinones have been described and reviewed through these methods.

However, this study has filled some voids: (1) used a large oxazolidinone dataset and its corresponding antibacterial MIC data, (2) tested a multitude of ribosomal adaptable docking programs with the said dataset against eleven ribosomal 50S ribosomal crystal structures, and (3) used an extended version of the dataset (to incorporate more data) and their 2D and 3D descriptors and tested against ten ML algorithms.

In summary, the molecular docking study utilised an extensive range of ribosomal crystal structures and oxazolidinone ligands against a multitude of ribosomal docking programs. The ML study again utilised the oxazolidinone dataset, in conjunction with 1440 descriptors of the said dataset, training and testing ten ML algorithms.

Overall, this MPhil project has significantly contributed to the analysis of molecular docking programs for ribosomal targets, in addition to the analysis of an oxazolidinone target through molecular docking and ML. A similar approach can be applied to a more

extensive range of antibiotic datasets or applied to other antibacterial classes such as aminoglycosides, beta-lactams, tetracyclines, sulphonamides, quinolones, peptides, and macrolides.

6.3 Research Limitations

6.3.1 Limitations of current modelling techniques within RNA

Structure based docking of small molecules to RNA molecules like the 23S ribosome is not as established as similar methods for protein-ligand docking (Philips et al., 2013). Compared with predicting protein-ligand interaction, modelling binds interaction between RNA-ligand molecules presents five distinctive challenges; adaptation of molecular docking methods from proteins to RNA, scoring functions, RNA and DNA flexibility, water molecules, and metal ions, highly charged macromolecules, and availability of experimentally determined RNA structures (Y. Zhou et al., 2022).

Firstly, one of the significant issues is the adaptation of molecular docking methods from protein targets to RNA targets. In 2008, Moitessier et al. reported that over 60 molecular docking methods were successfully developed for proteins, with many more being developed since then (Luo et al., 2019; Moitessier et al., 2008). Traditionally, docking programs, such as AutoDock Vina (Vina) (Trott & Olson, 2010), GOLD (Jones et al., 1997), and Glide (Friesner et al., 2004), were created for protein-ligand targets. Some programs initially developed for protein-ligand docking have now been optimised for NA targets, such as AutoDock (Morris et al., 2009), DOCK 6 (Lang et al., 2009), and FITTED (Corbeil et al., 2007), while other docking programs were developed explicitly for NA targets, including rDock (Ruiz-Carmona et al., 2014), RLDOCK (L.-Z. Sun et al., 2020), and NLDOCK (Y. Feng et al., 2021). RNA molecules are highly flexible and can fold into multiple stable conformations. Predicting RNA-ligand interactions is more complex than protein-ligand binding because the binding sites on RNA may be shallower, highly polar, solvated, and conformationally flexible (Hermann, 2016; Luo et al., 2019; Tessaro & Scapozza, 2020). Docking programs often overlook the inherent flexibility of nucleic acids. When a ligand binds to its corresponding RNA, it can cause an induced-fit effect, as well as a conformational stabilization or destabilization of the tertiary complex structure, as noted previously (Tessaro & Scapozza, 2020). MD-based methods are not appropriate for large-scale virtual screening due to their significant computational requirements. Although rigid docking is rapid, it results in low accuracy. Soft docking and ensemble docking are a middle ground between these extremes, as they balance accuracy and computational

efficiency. However, both methods sacrifice the ability to thoroughly sample conformations to minimize the computational time (Y. Zhou et al., 2022). At this current time, a versatile approach to accurately treat receptor flexibility still needs to be developed (Y. Zhou et al., 2022, p.).

Secondly, existing scoring functions are primarily developed to predict protein-ligand affinities, which poses challenges for identifying the native ligand binding pose from a set of candidates in RNA-ligand docking. To overcome this limitation, reliable scoring functions based on physics, knowledge, or machine learning-based approaches are necessary. While physics-based force field scoring functions can provide valuable insights into the physical mechanism of RNA-small molecule interactions, they are computationally intensive and require expertise in simulating a specific system, which limits their applicability in large-scale virtual screening for drug discovery (Y. Zhou et al., 2022). The empirical energy function methods use simple, functional forms to model RNA-small molecule interactions, which reduces the computational burden. However, there are two main limitations to this approach. First, it neglects the correlation between different interactions. Second, it may be difficult to transfer weight coefficients between different RNA-ligand systems (Y. Zhou et al., 2022). The potential statistical approach is associated with much higher computational efficiency than atomistic force field methods. However, the choice of reference state places obstacles to the accurate modelling of RNA-small molecule interactions. Several tools like NP Dock (Tuszynska et al., 2015), HNADOCK (J. He et al., 2019), and HDOCK (Y. Yan et al., 2020) have been developed for protein-RNA/DNA docking and RNA/DNA-DNA/RNA docking. Even though iterative approaches have been developed to bypass the problem, constructing diverse and complete decoy datasets for training remains challenging for RNA-ligand complexes (Y. Zhou et al., 2022). While ML models have been successful in predicting protein folding and protein-small molecule interactions, applying these models to RNA-small molecule interactions is a recent development. However, there are limitations to using ML for this purpose. Due to the large number of trainable parameters involved in the model, overfitting can be an issue, particularly when there is limited training data available (Y. Zhou et al., 2022).

The next issue is that many molecular docking studies do not consider the associated effects of water molecules and metal ions. Neglecting such solvent effects is a significant drawback that can cause inaccurate predictions for RNA-ligand interactions. One possible solution to address the neglect of solvent effects in molecular docking studies is to perform

simulations with explicit water molecules and metal ions, which can improve the accuracy of RNA structures by taking into account the critical role of solvent molecules. However, achieving high accuracy with this approach can be challenging, as RNA-ligand interactions depend on the precise positions and orientations of water molecules and ions within the binding site (Belousoff et al., 2019; Moitessier et al., 2006; Park et al., 2011). A feasible approach is to predict the binding of metal ions and water molecules to RNA prior to docking and incorporate them as part of the RNA-small molecule docking receptor. It is crucial to consider the solvation and desolvation effects during ligand binding and water-mediated interactions. The influence of explicit solvation sites within protein-ligand binding pockets on affinity and selectivity is well-established, and such effects are expected to occur in RNA-ligand binding sites as well (Kallert et al., 2022).

A significant challenge in RNA-targeted drug discovery is the high charge density of RNA molecules, which makes scoring the free energy of ligand-RNA binding difficult, especially considering the presence of metal ions. While nucleic acids can be suitable for target-based drug discovery from a structural perspective, their high charge density presents a challenge for accurate scoring of ligand-RNA interactions (Tessaro & Scapozza, 2020). One of the significant issues when docking or screening positively charged ligands is the ability to discriminate specific interactions, given the prevalence of backbone negative charges. All the phosphate-negative formal charges are neutralised by counterions in solutions. Due to their unfixed location, these objects are frequently mobile and cannot be easily included in docking considerations. An additional aspect is the adjustment of the polarisation of water molecules when hydrogen bonds to various groups, such as phosphates. Each water is expected to have a different polarisation and to interact more or less with ligands, while the waters are considered equivalent by docking programs (Luo et al., 2019).

The final challenge for RNA docking is that there are a limited number of experimentally determined RNA structures. The RNA-ligand complex makes knowledge-based approaches less effective for RNA-ligand predictions (L. Chen et al., 2012; Y. Zhou et al., 2022). Up until recently, when Fromm et al. described the 1.55 Å *E. coli* crystal structure (Fromm et al., 2023), the quality of ribosomal crystal structures was not ideal for high-quality molecular docking studies - ranging from 2.5 Å to 3.7 Å. Although protein-focused libraries, such as PDB and PDBbind, can contain data for RNA-small molecule complexes, a more comprehensive, dedicated NDB-like database (Coimbatore Narayanan et al., 2014) for RNA-ligand complexes is needed.

There are several minor factors that need to be considered as well. Firstly, rRNA contains various functional groups and binding sites that can interact with other molecules, making it difficult to identify the precise site or sites that a potential drug candidate should bind to, particularly when the chloramphenicol and linezolid sites overlap. Moreover, various factors like pH, temperature, and the presence of other molecules can influence drug binding to rRNA, which can make it challenging to accurately forecast the binding behaviour of a drug candidate using molecular docking (Buckley et al., 2023).

6.3.2 Challenges and Limitations of ML in drug discovery

While ML has allowed for immense improvement in various areas, like medicine and drug discovery, currently, there are several challenges: lack of data, heterogeneity of data, and noisy data.

Firstly is the lack of data. Due to Lipinski's rule of 5, there is a limited amount of data on druggable targets for drug discovery (Doak & Kihlberg, 2017). As machine learning models thrive on large datasets, this makes it a challenging application scenario. In drug discovery, high-quality experimental data is often expensive and time-consuming to generate, especially for complex biological systems like proteins or RNA (Vamathevan et al., 2019). Additionally, pharmaceutical companies and research institutions often hold valuable data and intellectual property rights, which may limit the sharing of data with others. The complexity of drug discovery is also another major limitation. The complexity of biological systems can make it challenging to obtain comprehensive data that captures all relevant aspects of a particular drug target or disease (Vamathevan et al., 2019).

The second problem is the heterogeneity of data. The data available for drug discovery is often heterogeneous, being diverse in character or content. Data often comes from a variety of sources, such as medical records, literature, and preclinical studies. Furthermore, as the quantity and diversity of data produced and saved in extensive databases continue to grow, the issue of maintaining and preserving data accuracy has become increasingly significant. The creation and maintenance of vast volumes of data necessitate substantial investments in IT infrastructure (David et al., 2019). This makes it challenging to integrate and analyse the data.

The ultimate concern is the potential presence of noisy data, which has been demonstrated by several empirical studies to significantly impair prediction accuracy. Machine learning datasets can contain two types of noise: attribute noise, which affects the predictive

attributes, and class noise, which affects the target attribute (S. Gupta & Gupta, 2019). The presence of noise in a data set can increase the model complexity and learning time, degrading the performance of learning algorithms. It is essential to pre-process and clean the data before feeding it into machine learning models to mitigate the effects of noisy data. This can involve techniques such as data imputation, outlier detection and removal, and normalisation. Additionally, it is crucial to choose appropriate machine learning models and evaluation metrics that can handle noisy data effectively.

6.4 Directions for further research

As there are two sections to this thesis, the discussion of the future directions will be split into two parts: molecular docking and ML. Some of these future directions will address the limitations mentioned in previous sections, while others are methods that have not been addressed in the thesis but are potential future steps that could be implemented into any further research to help improve results.

6.4.1 Plans for molecular docking

There are various plans I have to address the molecular docking limitations, summarised as follows: (1) Utilisation of MD methods such as ensemble docking, (2) developing an ML scoring function, (3) addressing negative charges of RNA using polarised molecular mechanics, (4) the use of advanced MD simulations to overcome induce fit or improving the quality of 3D structures using Cryo-EM, and (5) inclusion of alternative datasets.

The first issue is that molecular docking program algorithms are not catered for RNA targets, so they do not incorporate the high flexibility of RNA targets. A potential avenue is to use what is called ensemble docking. Ensemble docking is the “ensemble” of drug target conformations in computational structure-based drug discovery, often obtained by using MD simulation (MD) (Amaro et al., 2018).

As previously mentioned, most scoring functions are made explicitly for protein-ligand functions. While there are some RNA-developed scoring functions out there, such as AnnapuRNA (Stefaniak & Bujnicki, 2021) (used in Chapter 3), LigandRNA (Philips et al., 2013), and SPA-LN (Z. Yan & Wang 2017), these are all knowledge-based scoring functions. This means that these scoring functions rely on current experiments on RNA-ligand complexes, and as previously mentioned, current RNA crystal structures are not at the same standard as protein-ligand structures. A potential plan would be to develop an in-house ML-based scoring function for RNA, as ML aims to circumvent these limitations by

learning the functional form directly from the training data (H. Li et al., 2021). There is also the potential of using Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) scoring functions. This method has been utilised a number of times for protein targets (Wichapong et al., 2014; X. Zhang et al., 2017), and protein-RNA (F. Chen et al., 2018) has the potential to be used for RNA-based targets.

Another potential avenue is to address the high negative charges of RNA by using polarised molecular mechanics scoring function/force fields. Illingworth et al. (2008) utilised a method of ligand and protein polarisation in docking that is based on the conversion of induced dipoles to induced charges. To avoid issues related to parametrization, the ligand was treated using quantum mechanics, while the target protein was considered as a set of point charges and used to polarize the ligand. The resulting induced dipole at a specific target atom was expressed as induced charges at the atom and its bonded neighbours, which could repolarize the ligand iteratively (Illingworth et al., 2008). A similar method could be adapted and applied to RNA-ligand targets.

Using advanced to address the poor quality of current RNA crystal structures is another alternative. Various MD (MD) simulation studies (J. Chen et al., 2019; Makarov & Makarova, 2020) have been used to overcome induce fit or improve the quality of 3D structures using Cryo-EM. A fast MD software called ACEMD (Harvey et al., 2009) and the Supervised MD (SuMD) algorithm (Bissaro et al., 2020) could be used. ACEMD maximises performance by running the whole computation on the GPU rather than offloading only selected computationally expensive parts (Harvey et al., 2009), while SuMD is an all-atoms MD protocol that allows accelerating the sampling of molecular recognition events on a nanosecond timescale to ribonucleotide targets of pharmaceutical interest (Bissaro et al., 2020). MD programs like these, amongst many others, can be used to account for the poor quality of structures and find potential binding modes that may have been missed in the original docking study.

Including an alternate dataset in molecular docking studies could provide insights into the extrapolation of findings by assessing the robustness and generalizability of the results. The inclusion of an alternate dataset in molecular docking studies allows for extrapolations by evaluating the generalizability, validating binding affinities, and testing the method's performance on diverse compounds. This approach strengthens the reliability and

applicability of the docking results, enabling confident extrapolation to novel compounds and expanding the scope of the findings.

6.4.2 Future Directions for machine learning

Future research directions mainly lie in the machine learning portion of the thesis. I have various plans to expand this thesis further: (1) expanding the dataset, (2) exploring a different way to describe the characteristics of a derivative, such as fingerprints, (3) predicting the qualitative antibiofilm activity of antibiotics, and the distinction between gram-positive and gram-negative bacteria, (4) utilisation of open source deep learning libraries and development of in-house graph-based methods, (5) utilising additional feature selection methods, and (6) using a variation of training and testing splits.

As this thesis was solely focused on oxazolidinones, the future is to extend our methods to encompass a much larger, more structurally diverse dataset. An excellent example of a potential dataset comes from the 2012 paper by Singh et al., which utilises antibacterial compounds from multiple classes: aminoglycosides, beta-lactams, tetracyclines, sulphonamides, quinolones, peptides, macrolides in addition to oxazolidinones (N. Singh et al., 2012). Additionally, we plan to include oxazolidinones that are not antibiotics, such as using antiviral (Ghosh et al., 2018; Parai et al., 2012) or anticancer (Cho et al., 2017; Xie et al., 2017) oxazolidinone structures as an alternate dataset for prediction using ML (Fernandes et al., 2023). Some examples are shown in Figure 6.1. An additional way to expand the dataset would be to include bound-conformation information obtained from the crystal structure, such as demonstrated in molecular docking studies. This inclusion of structure-based information (residue interactions) could allow for additional information to improve the performance of the ligand-based machine learning methods.

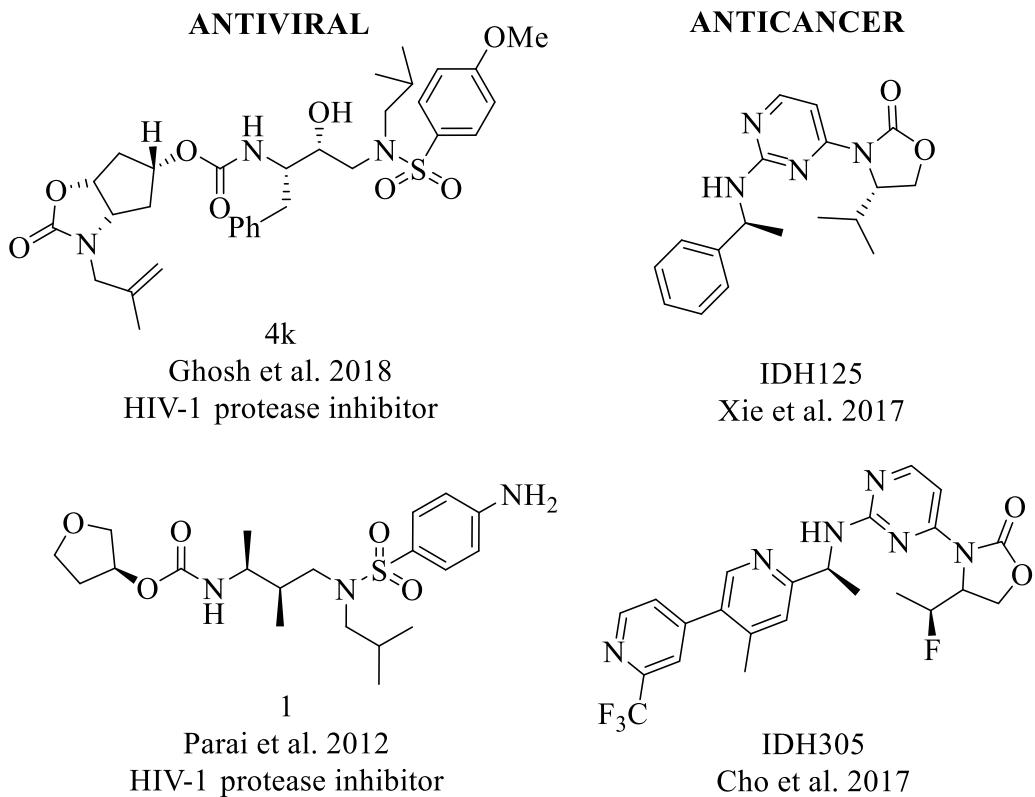


Figure 6.1 - Examples of antiviral and anticancer oxazolidinone structures

As Chapter 5 was also only focused on using the 2D and 3D descriptors, we also plan to incorporate fingerprints. Molecular descriptors are typically generated from basic molecular information, such as the SMILES used throughout this thesis. However, there are alternative ways to use these SMILES, such as fingerprints. Fingerprints tend to be more compact and informationally rich, and more computationally efficient. Descriptors in drug discovery are numerical representations that capture the chemical properties of compounds, while fingerprints are binary or bit-vector representations that encode the presence or absence of specific molecular features or substructures in compounds. There are numerous types of fingerprints in the field of drug discovery and cheminformatics due to various factors, including the diverse nature of molecular structures, the specific requirements of different applications, and the evolution of computational techniques. There have been various antibiotic studies done using various types of fingerprints, including but not limited to Extended-Connectivity Fingerprint (ECFP) (Choi et al., 2020; Myint et al., 2012), Functional-Class Fingerprint (Choi et al., 2020; Myint et al., 2012), Molecular ACCess System (MACCS) fingerprint (Myint et al., 2012), Atom-centred fragment descriptors (Kandel et al., 2014), and Morgan/Circular Fingerprint (H. Wang et

al., 2021). These different types can be used to train and test ML methods to determine what may be a better predictor of antibiotic activity.

An additional approach would be to predict the qualitative antibiofilm activity of antibiotics using ML techniques (Shaban & Alkawareek, 2022), as many of these biofilm targets are unknown. Additionally, the avenue of distinction between gram-positive and gram-negative activity in antibiotic targets could be explored. A comprehensive, multidisciplinary predictive computational tool could also be developed to identify such molecules by carrying out large-scale, quantitative analyses of the accumulation of diverse small molecules in Gram-negative bacteria/positive bacteria. These tools will then enable medicinal chemistry campaigns to develop novel antibiotics.

Open-source deep learning libraries like DeepChem (Ramsundar et al., 2019) and TorchDrug (Zhu et al., 2022) have been used for property prediction, prediction of synthetic feasibility, molecule generation, retrosynthesis, and knowledge graphs. These can all be applied to antibacterial targets. Graph-based methods that aim to capture graph structure in addition to regular node features can be developed in-house further to expand the ML aspect of this research. Graph-based methods are a common way to represent complicated relationships within a complex network, allowing for the inclusion of more complex data.

Another method to further expand and improve the quality of the research would be to expand the method of feature selection. This thesis only used one method of feature selection, the filter-based method of mutual information. Other methods, such as wrapper-based methods (e.g. ensemble feature selection) or dimensionality reduction techniques (e.g. PCA), can be used as a comparison tool (Cai et al., 2018).

The final method that can be used to improve the results of the studies is the variation in the splitting of testing and training to the dataset. This study utilised the commonly used two-way 80:20 split method, with 80% of the dataset used for training and 20% used for testing. This split indicated that the model was not evaluated with an independent set after checking the performance with 20% of the remaining dataset. An alternative splitting method could be a three-way split, such as a 60:20:20 split. With this method, 60% of training data is used to train the model's parameters, 20% of validation data is used for tuning hyperparameters and selecting the best model, and the final 20% of testing data is used to assess the model's performance on unseen data. This extra splitting allows for a more comprehensive evaluation by having both a validation set and a dedicated testing set.

6.5 Final conclusion

Molecular docking studies for RNA-ligand are not at the stage of current protein ligands, so for future structure-based works, there are a few methods that can be used for the improvement of simulations. Methods such as ensemble docking can address the high flexibility of RNA. Additionally, an in-house ML scoring function can be developed to address the issues with scoring functions for RNA-based targets in addition to MM/GBSA scoring functions. Polarised molecular mechanics scoring function/force fields can be utilised to address RNA's negative charges, while MD simulations can be used to counteract the poor quality of currently available crystal structures.

Future works, specifically ligand-based studies, should focus on the incorporation of a larger, more structurally diverse dataset of antibiotics from a more extensive range of antibacterial classes. Additionally, the use of descriptors should be expanded further to incorporate and analyse the uses of a wide range of fingerprints. The extension can also include biofilm targets and develop a tool to differentiate between gram-positive and gram-negative bacteria. Open-source DL libraries or in-house graph-based ML models can be used extensively in future studies to expand the research further. Additional feature selection methods could be used alongside mutual information, such as ensemble feature selection or PCA. Lastly, a variety of data training-test splitting techniques could be used to include an independent set of data for comparison and performance improvement.

In summary, there are many methods (for both molecular docking and ML) and datasets (antibiotic and non-antibiotic) that can be either developed in-house or incorporated, respectively, to explore and expand on ML research in antibiotics or biofilms.

Bibliography

- Amaro, R. E., Baudry, J., Chodera, J., Demir, Ö., McCammon, J. A., Miao, Y., & Smith, J. C. (2018). Ensemble Docking in Drug Discovery. *Biophysical Journal*, 114(10), 2271–2278. <https://doi.org/10.1016/j.bpj.2018.02.038>
- Ament, P. W., Jamshed, N., & Horne, J. P. (2002). Linezolid: Its Role in the Treatment of Gram-Positive, Drug-Resistant Bacterial Infections. *American Family Physician*, 65(4), 663–671. <https://www.aafp.org/pubs/afp/issues/2002/0215/p663.html>
- Anahtar, M. N., Yang, J. H., & Kanjilal, S. (2021). Applications of Machine Learning to the Problem of Antimicrobial Resistance: An Emerging Model for Translational Research. *Journal of Clinical Microbiology*, 59(7), e01260-20. <https://doi.org/10.1128/JCM.01260-20>
- Andersson, D. I., & Hughes, D. (2010). Antibiotic resistance and its cost: Is it possible to reverse resistance? *Nature Reviews. Microbiology*, 8(4), 260–271. <https://doi.org/10.1038/nrmicro2319>
- Azzouz, A., & Preuss, C. V. (2023). Linezolid. In *StatPearls*. StatPearls Publishing. <http://www.ncbi.nlm.nih.gov/books/NBK539793/>
- Bal, A. M. (2022). 7.10—Lincosamide Antibiotics. In T. Kenakin (Ed.), *Comprehensive Pharmacology* (pp. 185–200). Elsevier. <https://doi.org/10.1016/B978-0-12-820472-6.00111-0>
- Ballester, P. J., & Mitchell, J. B. O. (2010). A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking. *Bioinformatics*, 26(9), 1169–1175. <https://doi.org/10.1093/bioinformatics/btq112>
- Ban, N., Freeborn, B., Nissen, P., Penczek, P., Grassucci, R. A., Sweet, R., Frank, J., Moore, P. B., & Steitz, T. A. (1998). A 9 Å Resolution X-Ray Crystallographic Map of the Large Ribosomal Subunit. *Cell*, 93(7), 1105–1115. [https://doi.org/10.1016/S0092-8674\(00\)81455-5](https://doi.org/10.1016/S0092-8674(00)81455-5)
- Ban, N., Nissen, P., Hansen, J., Capel, M., Moore, P. B., & Steitz, T. A. (1999). Placement of protein and RNA structures into a 5 Å-resolution map of the 50S ribosomal subunit. *Nature*, 400(6747), 841–847. <https://doi.org/10.1038/23641>

- Ban, N., Nissen, P., Hansen, J., Moore, P. B., & Steitz, T. A. (2000). The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 Å Resolution. *Science*, 289(5481), 905–920. <https://doi.org/10.1126/science.289.5481.905>
- Barb, A. W., Jiang, L., Raetz, C. R. H., & Zhou, P. (2007). Structure of the deacetylase LpxC bound to the antibiotic CHIR-090: Time-dependent inhibition and specificity in ligand binding. *Proceedings of the National Academy of Sciences*, 104(47), 18433–18438. <https://doi.org/10.1073/pnas.0709412104>
- Barbachyn, M. R., & Ford, C. W. (2003). Oxazolidinone Structure–Activity Relationships Leading to Linezolid. *Angewandte Chemie International Edition*, 42(18), 2010–2023. <https://doi.org/10.1002/anie.200200528>
- Bashan, A., Agmon, I., Zarivach, R., Schluzen, F., Harms, J., Berisio, R., Bartels, H., Franceschi, F., Auerbach, T., Hansen, H. A. S., Kossoy, E., Kessler, M., & Yonath, A. (2003). Structural Basis of the Ribosomal Machinery for Peptide Bond Formation, Translocation, and Nascent Chain Progression. *Molecular Cell*, 11(1), 91–102. [https://doi.org/10.1016/S1097-2765\(03\)00009-1](https://doi.org/10.1016/S1097-2765(03)00009-1)
- Bassetti, M., Vitale, F., Melica, G., Righi, E., Di Biagio, A., Molfetta, L., Pipino, F., Cruciani, M., & Bassetti, D. (2005). Linezolid in the treatment of Gram-positive prosthetic joint infections. *Journal of Antimicrobial Chemotherapy*, 55(3), 387–390. <https://doi.org/10.1093/jac/dki016>
- Batts, D. H. (2000). Linezolid—A new option for treating gram-positive infections. *Oncology (Williston Park, N.Y.)*, 14(8 Suppl 6), 23–29.
- Bauer, M. R., & Mackey, M. D. (2019). Electrostatic Complementarity as a Fast and Effective Tool to Optimize Binding and Selectivity of Protein–Ligand Complexes. *Journal of Medicinal Chemistry*, 62(6), 3036–3050. <https://doi.org/10.1021/acs.jmedchem.8b01925>
- Baxter, C. A., Murray, C. W., Clark, D. E., Westhead, D. R., & Eldridge, M. D. (1998). Flexible docking using tabu search and an empirical estimate of binding affinity. *Proteins: Structure, Function, and Bioinformatics*, 33(3), 367–382. [https://doi.org/10.1002/\(SICI\)1097-0134\(19981115\)33:3<367::AID-PROT6>3.0.CO;2-W](https://doi.org/10.1002/(SICI)1097-0134(19981115)33:3<367::AID-PROT6>3.0.CO;2-W)
- Belousoff, M. J., Venugopal, H., Wright, A., Seoner, S., Stuart, I., Stubenrauch, C., Bamert, R. S., Lupton, D. W., & Lithgow, T. (2019). CryoEM-Guided Development of Antibiotics for Drug-Resistant Bacteria. *ChemMedChem*, 14(5), 527–531. <https://doi.org/10.1002/cmdc.201900042>

- Bernhardt, H. S., & Tate, W. P. (2012). Primordial soup or vinaigrette: Did the RNA world evolve at acidic pH? *Biology Direct*, 7, 4. <https://doi.org/10.1186/1745-6150-7-4>
- Bhattarai, D., Lee, J., Seo, S. H., Nam, G., Choo, H., Kang, S. B., Kwak, J.-H., Oh, T., Cho, S.-N., Pae, A. N., Kim, E. E., Jeong, N., & Keum, G. (2014). Synthesis and *in Vitro* Evaluation of the Antitubercular and Antibacterial Activity of Novel Oxazolidinones Bearing Octahydrocyclopenta[c]pyrrol-2-yl Moieties. *Chemical and Pharmaceutical Bulletin*, 62(12), 1214–1224. <https://doi.org/10.1248/cpb.c14-00510>
- BIOVIA. (2022). *Discovery Studio* (v21.1.0.20298) [Computer software].
- Bissaro, M., Sturlese, M., & Moro, S. (2020). Exploring the RNA-Recognition Mechanism Using Supervised Molecular Dynamics (SuMD) Simulations: Toward a Rational Design for Ribonucleic-Targeting Molecules? *Frontiers in Chemistry*, 8. <https://www.frontiersin.org/articles/10.3389/fchem.2020.00107>
- Böhm, H. J. (1992). The computer program LUDI: A new method for the de novo design of enzyme inhibitors. *Journal of Computer-Aided Molecular Design*, 6(1), 61–78. <https://doi.org/10.1007/BF00124387>
- Boittier, E. D., Tang, Y. Y., Buckley, M. E., Schuurs, Z. P., Richard, D. J., & Gandhi, N. S. (2020). Assessing Molecular Docking Tools to Guide Targeted Drug Discovery of CD38 Inhibitors. *International Journal of Molecular Sciences*, 21(15), E5183. <https://doi.org/10.3390/ijms21155183>
- Bourgeois-Nicolaos, N., Massias, L., Couson, B., Butel, M.-J., Andremont, A., & Doucet-Populaire, F. (2007). Dose dependence of emergence of resistance to linezolid in *Enterococcus faecalis* *in vivo*. *The Journal of Infectious Diseases*, 195(10), 1480–1488. <https://doi.org/10.1086/513876>
- Bozdogan, B., & Appelbaum, P. C. (2004). Oxazolidinones: Activity, mode of action, and mechanism of resistance. *International Journal of Antimicrobial Agents*, 23(2), 113–119. <https://doi.org/10.1016/j.ijantimicag.2003.11.003>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Brickner, S. J. (2007). 7.23—Oxazolidinone Antibiotics. In J. B. Taylor & D. J. Triggle (Eds.), *Comprehensive Medicinal Chemistry II* (pp. 673–698). Elsevier. <https://doi.org/10.1016/B0-08-045044-X/00223-6>

- Brickner, S. J., Hutchinson, D. K., Barbachyn, M. R., Manninen, P. R., Ulanowicz, D. A., Garmon, S. A., Grega, K. C., Hedges, S. K., Toops, D. S., Ford, C. W., & Zurenko, G. E. (1996). Synthesis and antibacterial activity of U-100592 and U-100766, two oxazolidinone antibacterial agents for the potential treatment of multidrug-resistant gram-positive bacterial infections. *Journal of Medicinal Chemistry*, 39(3), 673–679. <https://doi.org/10.1021/jm9509556>
- Brooijmans, N., & Kuntz, I. D. (2003). Molecular Recognition and Docking Algorithms. *Annual Review of Biophysics and Biomolecular Structure*, 32(1), 335–373. <https://doi.org/10.1146/annurev.biophys.32.110601.142532>
- Brown-Elliott, B. A., Wallace, R. J., Jr., Blinkhorn, R., Crist, C. J., & Mann, L. B. (2001). Successful Treatment of Disseminated *Mycobacterium chelonae* Infection with Linezolid. *Clinical Infectious Diseases*, 33(8), 1433–1434. <https://doi.org/10.1086/322523>
- Brown-Elliott, B. A., Ward, S. C., Crist, C. J., Mann, L. B., Wilson, R. W., & Wallace, R. J. (2001). In Vitro Activities of Linezolid against Multiple Nocardia Species. *Antimicrobial Agents and Chemotherapy*, 45(4), 1295–1297. <https://doi.org/10.1128/AAC.45.4.1295-1297.2001>
- Buckley, M. E., Ndukwe, A. R. N., Nair, P. C., Rana, S., Fairfull-Smith, K. E., & Gandhi, N. S. (2023). Comparative Assessment of Docking Programs for Docking and Virtual Screening of Ribosomal Oxazolidinone Antibacterial Agents. *Antibiotics*, 12(3), Article 3. <https://doi.org/10.3390/antibiotics12030463>
- Bulkley, D., Innis, C. A., Blaha, G., & Steitz, T. A. (2010). Revisiting the structures of several antibiotics bound to the bacterial ribosome. *Proceedings of the National Academy of Sciences*, 107(40), 17158–17163. <https://doi.org/10.1073/pnas.1008685107>
- Cai, J., Luo, J., Wang, S., & Yang, S. (2018). Feature selection in machine learning: A new perspective. *Neurocomputing*, 300, 70–79. <https://doi.org/10.1016/j.neucom.2017.11.077>
- Carpenter, K. A., Cohen, D. S., Jarrell, J. T., & Huang, X. (2018). Deep learning and virtual drug screening. *Future Medicinal Chemistry*, 10(21), 2557–2567. <https://doi.org/10.4155/fmc-2018-0314>
- Chakraverty, R., Samanta, K., Mandal, P., Karmakar, S., & Karmakar, S. (2023). Chapter 18—Mechanisms of action of antibacterial agents (AMA). In I. Kazmi, S. Karmakar, Md. A. Shaharyar, M. Afzal, & F. A. Al-Abbas (Eds.), *How Synthetic*

- Drugs Work* (pp. 421–429). Academic Press. <https://doi.org/10.1016/B978-0-323-99855-0.00018-X>
- Chami Khazraji, A., & Robert, S. (2013). Self-Assembly and Intermolecular Forces When Cellulose and Water Interact Using Molecular Modeling. *Journal of Nanomaterials*, 2013, e745979. <https://doi.org/10.1155/2013/745979>
- Cheeseright, T., Mackey, M., Rose, S., & Vinter, A. (2006). Molecular Field Extrema as Descriptors of Biological Activity: Definition and Validation. *Journal of Chemical Information and Modeling*, 46(2), 665–676.
<https://doi.org/10.1021/ci050357s>
- Chen, F., Sun, H., Wang, J., Zhu, F., Liu, H., Wang, Z., Lei, T., Li, Y., & Hou, T. (2018). Assessing the performance of MM/PBSA and MM/GBSA methods. 8. Predicting binding free energies and poses of protein–RNA complexes. *RNA*, 24(9), 1183–1194. <https://doi.org/10.1261/rna.065896.118>
- Chen, H.-M., Liu, B.-F., Huang, H.-L., Hwang, S.-F., & Ho, S.-Y. (2007). SODOCK: Swarm optimization for highly flexible protein–ligand docking. *Journal of Computational Chemistry*, 28(2), 612–623. <https://doi.org/10.1002/jcc.20542>
- Chen, J., Wang, X., Pang, L., Zhang, J. Z. H., & Zhu, T. (2019). Effect of mutations on binding of ligands to guanine riboswitch probed by free energy perturbation and molecular dynamics simulations. *Nucleic Acids Research*, 47(13), 6618–6631.
<https://doi.org/10.1093/nar/gkz499>
- Chen, K., Li, T., & Cao, T. (2006). Tribe-PSO: A novel global optimization algorithm and its application in molecular docking. *Chemometrics and Intelligent Laboratory Systems*, 82(1), 248–259.
<https://doi.org/10.1016/j.chemolab.2005.06.017>
- Chen, K.-H., Huang, Y.-T., Liao, C.-H., Sheng, W.-H., & Hsueh, P.-R. (2015). In Vitro Activities of Tedizolid and Linezolid against Gram-Positive Cocci Associated with Acute Bacterial Skin and Skin Structure Infections and Pneumonia. *Antimicrobial Agents and Chemotherapy*, 59(10), 6262–6265.
<https://doi.org/10.1128/AAC.00390-15>
- Chen, L., Calin, G. A., & Zhang, S. (2012). Novel insights of structure-based modeling for RNA-targeted drug discovery. *Journal of Chemical Information and Modeling*, 52(10), 2741–2753. <https://doi.org/10.1021/ci300320t>
- Cherkasov, A., Muratov, E. N., Fourches, D., Varnek, A., Baskin, I. I., Cronin, M., Dearden, J., Gramatica, P., Martin, Y. C., Todeschini, R., Consonni, V., Kuz'min,

- V. E., Cramer, R., Benigni, R., Yang, C., Rathman, J., Terfloth, L., Gasteiger, J., Richard, A., & Tropsha, A. (2014). QSAR Modeling: Where Have You Been? Where Are You Going To? *Journal of Medicinal Chemistry*, 57(12), 4977–5010. <https://doi.org/10.1021/jm4004285>
- Chhabra, S., Xie, J., & Frank, A. T. (2020). RNAPosers: Machine Learning Classifiers for Ribonucleic Acid–Ligand Poses. *The Journal of Physical Chemistry B*, 124(22), 4436–4445. <https://doi.org/10.1021/acs.jpcb.0c02322>
- Cho, Y. S., Levell, J. R., Liu, G., Caferro, T., Sutton, J., Shafer, C. M., Costales, A., Manning, J. R., Zhao, Q., Sendzik, M., Shultz, M., Chenail, G., Dooley, J., Villalba, B., Farsidjani, A., Chen, J., Kulathila, R., Xie, X., Dodd, S., ... Gowney, J. D. (2017). Discovery and Evaluation of Clinical Candidate IDH305, a Brain Penetrant Mutant IDH1 Inhibitor. *ACS Medicinal Chemistry Letters*, 8(10), 1116–1121. <https://doi.org/10.1021/acsmedchemlett.7b00342>
- Choi, K.-E., Balupuri, A., & Kang, N. S. (2020). The Study on the hERG Blocker Prediction Using Chemical Fingerprint Analysis. *Molecules*, 25(11), Article 11. <https://doi.org/10.3390/molecules25112615>
- Chopra, I. (1998). Research and development of antibacterial agents. *Current Opinion in Microbiology*, 1(5), 495–501. [https://doi.org/10.1016/S1369-5274\(98\)80080-5](https://doi.org/10.1016/S1369-5274(98)80080-5)
- Chowdhury, A. S., Call, D. R., & Broschat, S. L. (2020). PARGT: A software tool for predicting antimicrobial resistance in bacteria. *Scientific Reports*, 10(1), Article 1. <https://doi.org/10.1038/s41598-020-67949-9>
- Chowdhury, A. s., Khaledian, E., & Broschat, S. l. (2019). Capreomycin resistance prediction in two species of Mycobacterium using a stacked ensemble method. *Journal of Applied Microbiology*, 127(6), 1656–1664. <https://doi.org/10.1111/jam.14413>
- Cohen, T. A., Patel, V. L., & Shortliffe, E. H. (2022). Introducing AI in Medicine. In T. A. Cohen, V. L. Patel, & E. H. Shortliffe (Eds.), *Intelligent Systems in Medicine and Health: The Role of AI* (pp. 3–20). Springer International Publishing. https://doi.org/10.1007/978-3-031-09108-7_1
- Coimbatore Narayanan, B., Westbrook, J., Ghosh, S., Petrov, A. I., Sweeney, B., Zirbel, C. L., Leontis, N. B., & Berman, H. M. (2014). The Nucleic Acid Database: New features and capabilities. *Nucleic Acids Research*, 42(D1), D114–D122. <https://doi.org/10.1093/nar/gkt980>

- Corbeil, C. R., Englebienne, P., & Moitessier, N. (2007). Docking Ligands into Flexible and Solvated Macromolecules. 1. Development and Validation of FITTED 1.0. *Journal of Chemical Information and Modeling*, 47(2), 435–449.
<https://doi.org/10.1021/ci6002637>
- Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., & Kollman, P. A. (1995). A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society*, 117(19), 5179–5197.
<https://doi.org/10.1021/ja00124a002>
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297. <https://doi.org/10.1007/BF00994018>
- Cresset Group. (2022). *FlareTM* (V.3) (V.3) [Computer software]. Cresset Group.
<http://www.cresset-group.com/flare/>
- Cunha, B. A. (2010). *Antibiotic Essentials 2010*. Jones & Bartlett Publishers.
- D, H., S, R.-G., & E, R. (2008). In vitro-selected linezolid-resistant Mycobacterium tuberculosis mutants. *Antimicrobial Agents and Chemotherapy*, 52(2).
<https://doi.org/10.1128/AAC.01189-07>
- Dailey, C. F., Dileto-Fang, C. L., Buchanan, L. V., Oramas-Shirey, M. P., Batts, D. H., Ford, C. W., & Gibson, J. K. (2001). Efficacy of Linezolid in Treatment of Experimental Endocarditis Caused by Methicillin-Resistant *Staphylococcus aureus*. *Antimicrobial Agents and Chemotherapy*, 45(8), 2304–2308.
<https://doi.org/10.1128/AAC.45.8.2304-2308.2001>
- Das, B., Rudra, S., Yadav, A., Ray, A., Rao, A. V. S. R., Srinivas, A. S. S. V., Soni, A., Saini, S., Shukla, S., Pandya, M., Bhateja, P., Malhotra, S., Mathur, T., Arora, S. K., Rattan, A., & Mehta, A. (2005). Synthesis and SAR of novel oxazolidinones: Discovery of ranbezolid. *Bioorganic & Medicinal Chemistry Letters*, 15(19), 4261–4267. <https://doi.org/10.1016/j.bmcl.2005.06.063>
- David, L., Arús-Pous, J., Karlsson, J., Engkvist, O., Bjerrum, E. J., Kogej, T., Kriegl, J. M., Beck, B., & Chen, H. (2019). Applications of Deep-Learning in Exploiting Large-Scale and Heterogeneous Compound Data in Industrial Pharmaceutical Research. *Frontiers in Pharmacology*, 10.
<https://www.frontiersin.org/articles/10.3389/fphar.2019.01303>
- Deshmukh, M. S., & Jain, N. (2018). Correction to “Design, Synthesis, and Antibacterial Evaluation of Oxazolidinones with Fused Heterocyclic C-Ring Substructure.”

ACS Medicinal Chemistry Letters, 9(5), 512–512.

<https://doi.org/10.1021/acsmmedchemlett.8b00164>

Dickerhoff, J., Warnecke, K. R., Wang, K., Deng, N., & Yang, D. (2021). Evaluating Molecular Docking Software for Small Molecule Binding to G-Quadruplex DNA.

International Journal of Molecular Sciences, 22(19), Article 19.

<https://doi.org/10.3390/ijms221910801>

Ding, S., Ji, J.-C., Zhang, M.-J., Yang, Y.-S., Wang, R., Zhu, X.-L., Wang, L.-H., Zhong, Y., Gao, L., Lu, M., Liu, J., & Chen, Y. (2019). Exploration of the structure–activity relationship and druggability of novel oxazolidinone-based compounds as Gram-negative antibacterial agents. *Archiv Der Pharmazie*, 352(11), 1900129.

<https://doi.org/10.1002/ardp.201900129>

Doak, B. C., & Kihlberg, J. (2017). Drug discovery beyond the rule of 5—Opportunities and challenges. *Expert Opinion on Drug Discovery*, 12(2), 115–119.

<https://doi.org/10.1080/17460441.2017.1264385>

Durrant, J. D., & Amaro, R. E. (2015). Machine-Learning Techniques Applied to Antibacterial Drug Discovery. *Chemical Biology & Drug Design*, 85(1), 14–21.

<https://doi.org/10.1111/cbdd.12423>

Durrant, J. D., & McCammon, J. A. (2010). NNScore: A Neural-Network-Based Scoring Function for the Characterization of Protein–Ligand Complexes. *Journal of Chemical Information and Modeling*, 50(10), 1865–1871.

<https://doi.org/10.1021/ci100244v>

Eldridge, M. D., Murray, C. W., Auton, T. R., Paolini, G. V., & Mee, R. P. (1997). Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *Journal of Computer-Aided Molecular Design*, 11(5), 425–445.

<https://doi.org/10.1023/a:1007996124545>

Endres, B., Bassères, E., Alam, P., M. Jahangir, & Garey, K. (2017). Cadazolid for the treatment of Clostridium difficile. *Expert Opinion on Investigational Drugs*, 26.

<https://doi.org/10.1080/13543784.2017.1304538>

Ewing, T. J. A., & Kuntz, I. D. (1997). Critical evaluation of search algorithms for automated molecular docking and database screening. *Journal of Computational Chemistry*, 18(9), 1175–1189. [https://doi.org/10.1002/\(SICI\)1096-987X\(19970715\)18:9<1175::AID-JCC6>3.0.CO;2-O](https://doi.org/10.1002/(SICI)1096-987X(19970715)18:9<1175::AID-JCC6>3.0.CO;2-O)

- Eyal, Z., Matzov, D., Krupkin, M., Wekselman, I., Paukner, S., Zimmerman, E., Rozenberg, H., Bashan, A., & Yonath, A. (2015). Structural insights into species-specific features of the ribosome from the pathogen *Staphylococcus aureus*. *Proceedings of the National Academy of Sciences of the United States of America*, 112(43), E5805–5814. <https://doi.org/10.1073/pnas.1517952112>
- Faella, F., Pagliano, P., Fusco, U., Attanasio, V., & Conte, M. (2006). Combined treatment with ceftriaxone and linezolid of pneumococcal meningitis: A case series including penicillin-resistant strains. *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*, 12(4), 391–394. <https://doi.org/10.1111/j.1469-0691.2006.01352.x>
- Falagas, M. E., Manta, K. G., Ntziora, F., & Vardakas, K. Z. (2006). Linezolid for the treatment of patients with endocarditis: A systematic review of the published evidence. *Journal of Antimicrobial Chemotherapy*, 58(2), 273–280. <https://doi.org/10.1093/jac/dkl219>
- Falagas, M. E., Siempos, I. I., Papagelopoulos, P. J., & Vardakas, K. Z. (2007). Linezolid for the treatment of adults with bone and joint infections. *International Journal of Antimicrobial Agents*, 29(3), 233–239. <https://doi.org/10.1016/j.ijantimicag.2006.08.030>
- Feng, J., Lupien, A., Gingras, H., Wasserscheid, J., Dewar, K., Légaré, D., & Ouellette, M. (2009). Genome sequencing of linezolid-resistant *Streptococcus pneumoniae* mutants reveals novel mechanisms of resistance. *Genome Research*, 19(7), 1214–1223. <https://doi.org/10.1101/gr.089342.108>
- Feng, Y., & Huang, S.-Y. (2020). ITScore-NL: An Iterative Knowledge-Based Scoring Function for Nucleic Acid–Ligand Interactions. *Journal of Chemical Information and Modeling*, 60(12), 6698–6708. <https://doi.org/10.1021/acs.jcim.0c00974>
- Feng, Y., Yan, Y., He, J., Tao, H., Wu, Q., & Huang, S.-Y. (2022). Docking and scoring for nucleic acid-ligand interactions: Principles and current status. *Drug Discovery Today*, 27(3), 838–847. <https://doi.org/10.1016/j.drudis.2021.10.013>
- Feng, Y., Zhang, K., Wu, Q., & Huang, S.-Y. (2021). NLDock: A Fast Nucleic Acid–Ligand Docking Algorithm for Modeling RNA/DNA–Ligand Complexes. *Journal of Chemical Information and Modeling*, 61(9), 4771–4782. <https://doi.org/10.1021/acs.jcim.1c00341>

- Fernandes, G., Scarim, C. B., Kim, S. H., Wu, J., & Castagnolo, D. (2023). Oxazolidinones as Versatile Scaffolds in Medicinal Chemistry. *RSC Medicinal Chemistry*, 10.1039.D2MD00415A. <https://doi.org/10.1039/D2MD00415A>
- Fines, M., & Leclercq, R. (2000). Activity of linezolid against Gram-positive cocci possessing genes conferring resistance to protein synthesis inhibitors. *Journal of Antimicrobial Chemotherapy*, 45(6), 797–802.
<https://doi.org/10.1093/jac/45.6.797>
- Ford, C. W., Hamel, J. C., Stapert, D., Moerman, J. K., Hutchinson, D. K., Barbachyn, M. R., & Zurenko, G. E. (1997). Oxazolidinones: New antibacterial agents. *Trends in Microbiology*, 5(5), 196–200. [https://doi.org/10.1016/S0966-842X\(97\)01032-9](https://doi.org/10.1016/S0966-842X(97)01032-9)
- Fortuna, C. G., Berardozzi, R., Bonaccorso, C., Caltabiano, G., Di Bari, L., Goracci, L., Guarcello, A., Pace, A., Palumbo Piccionello, A., Pescitelli, G., Pierro, P., Lonati, E., Bulbarelli, A., Cocuzza, C. E. A., Musumarra, G., & Musumeci, R. (2014). New potent antibacterials against Gram-positive multiresistant pathogens: Effects of side chain modification and chirality in linezolid-like 1,2,4-oxadiazoles. *Bioorganic & Medicinal Chemistry*, 22(24), 6814–6825.
<https://doi.org/10.1016/j.bmc.2014.10.037>
- Foti, C., Piperno, A., Scala, A., & Giuffrè, O. (2021). Oxazolidinone Antibiotics: Chemical, Biological and Analytical Aspects. *Molecules (Basel, Switzerland)*, 26(14), 4280. <https://doi.org/10.3390/molecules26144280>
- Fox, G. E. (2010). Origin and Evolution of the Ribosome. *Cold Spring Harbor Perspectives in Biology*, 2(9), a003483.
<https://doi.org/10.1101/cshperspect.a003483>
- Franceschi, F., & Duffy, E. M. (2006). Structure-based drug design meets the ribosome. *Biochemical Pharmacology*, 71(7), 1016–1025.
<https://doi.org/10.1016/j.bcp.2005.12.026>
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189–1232.
<https://doi.org/10.1214/aos/1013203451>
- Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., & Shenkin, P. S. (2004). Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of Medicinal Chemistry*, 47(7), 1739–1749. <https://doi.org/10.1021/jm0306430>

- Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., Sanschagrin, P. C., & Mainz, D. T. (2006). Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein–Ligand Complexes. *Journal of Medicinal Chemistry*, 49(21), 6177–6196.
<https://doi.org/10.1021/jm051256o>
- Fromm, S. A., O'Connor, K. M., Purdy, M., Bhatt, P. R., Loughran, G., Atkins, J. F., Jomaa, A., & Mattei, S. (2023). The translating bacterial ribosome at 1.55 Å resolution generated by cryo-EM imaging services. *Nature Communications*, 14(1), Article 1. <https://doi.org/10.1038/s41467-023-36742-3>
- Gandhi, N. (2007). 3D QSAR analysis of oxazolidinone antibacterials: Can we predict? *Arkivoc*, 2006(16), 109–121. <https://doi.org/10.3998/ark.5550190.0007.g12>
- Gandhi, N., Srivastava, B. K., Lohray, V. B., & Lohray, B. B. (2004). Oxazolidine-2-thiones: A molecular modeling study. *Tetrahedron Letters*, 45(33), 6269–6272.
<https://doi.org/10.1016/j.tetlet.2004.06.090>
- Genin, M. J. (2000). Recent progress with oxazolidinone antibacterial agents. *Expert Opinion on Therapeutic Patents*, 10(9), 1405–1414.
<https://doi.org/10.1517/13543776.10.9.1405>
- Ghosh, A. K., Williams, J. N., Ho, R. Y., Simpson, H. M., Hattori, S., Hayashi, H., Agniswamy, J., Wang, Y.-F., Weber, I. T., & Mitsuya, H. (2018). Design and Synthesis of Potent HIV-1 Protease Inhibitors Containing Bicyclic Oxazolidinone Scaffold as the P2 Ligands: Structure–Activity Studies and Biological and X-ray Structural Studies. *Journal of Medicinal Chemistry*, 61(21), 9722–9737.
<https://doi.org/10.1021/acs.jmedchem.8b01227>
- Gohlke, H., Hendlich, M., & Klebe, G. (2000). Knowledge-based scoring function to predict protein-ligand interactions. *Journal of Molecular Biology*, 295(2), 337–356. <https://doi.org/10.1006/jmbi.1999.3371>
- Goodsell, D. S., Morris, G. M., & Olson, A. J. (1996). Automated docking of flexible ligands: Applications of autodock. *Journal of Molecular Recognition*, 9(1), 1–5.
[https://doi.org/10.1002/\(SICI\)1099-1352\(199601\)9:1<1::AID-JMR241>3.0.CO;2-6](https://doi.org/10.1002/(SICI)1099-1352(199601)9:1<1::AID-JMR241>3.0.CO;2-6)
- Gordon, O., Dikeman, D. A., Ortines, R. V., Wang, Y., Youn, C., Mumtaz, M., Orlando, N., Zhang, J., Patel, A. M., Gough, E., Kaushik, A., Nuernberger, E. L., Upton, A. M., Fotouhi, N., Miller, L. S., & Archer, N. K. (2022). The Novel Oxazolidinone TBI-223 Is Effective in Three Preclinical Mouse Models of

- Methicillin-Resistant *Staphylococcus aureus* Infection. *Microbiology Spectrum*, 10(5), e0245121. <https://doi.org/10.1128/spectrum.02451-21>
- Gorse, A.-D. (2006). Diversity in medicinal chemistry space. *Current Topics in Medicinal Chemistry*, 6(1), 3–18. <https://doi.org/10.2174/156802606775193310>
- Guedes, I. A., de Magalhães, C. S., & Dardenne, L. E. (2014). Receptor-ligand molecular docking. *Biophysical Reviews*, 6(1), 75–87. <https://doi.org/10.1007/s12551-013-0130-2>
- Guilbert, C., & James, T. L. (2008). Docking to RNA via root-mean-square-deviation-driven energy minimization with flexible ligands and flexible targets. *Journal of Chemical Information and Modeling*, 48(6), 1257–1268.
<https://doi.org/10.1021/ci8000327>
- Guo, B., Fan, H., Xin, Q., Chu, W., Wang, H., Huang, Y., Chen, X., & Yang, Y. (2013). Solubility-Driven Optimization of (Pyridin-3-yl) Benzoxazinyl-oxazolidinones Leading to a Promising Antibacterial Agent. *Journal of Medicinal Chemistry*, 56(6), 2642–2650. <https://doi.org/10.1021/jm4000598>
- Gupta, P., & Sehgal, N. K. (2021). Practical Aspects in Machine Learning. In P. Gupta & N. K. Sehgal (Eds.), *Introduction to Machine Learning in the Cloud with Python: Concepts and Practices* (pp. 115–142). Springer International Publishing.
https://doi.org/10.1007/978-3-030-71270-9_5
- Gupta, S., & Gupta, A. (2019). Dealing with Noise Problem in Machine Learning Datasets: A Systematic Review. *Procedia Computer Science*, 161, 466–474.
<https://doi.org/10.1016/j.procs.2019.11.146>
- Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., & Banks, J. L. (2004). Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *Journal of Medicinal Chemistry*, 47(7), 1750–1759. <https://doi.org/10.1021/jm030644s>
- Harvey, M. J., Giupponi, G., & Fabritiis, G. D. (2009). ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. *Journal of Chemical Theory and Computation*, 5(6), 1632–1639. <https://doi.org/10.1021/ct9000685>
- Hashemian, S. M. R., Farhadi, T., & Ganjparvar, M. (2018). Linezolid: A review of its properties, function, and use in critical care. *Drug Design, Development and Therapy*, 12, 1759–1767. <https://doi.org/10.2147/DDDT.S164515>

- He, J., Wang, J., Tao, H., Xiao, Y., & Huang, S.-Y. (2019). HNADOCK: A nucleic acid docking server for modeling RNA/DNA–RNA/DNA 3D complex structures. *Nucleic Acids Research*, 47(W1), W35–W42. <https://doi.org/10.1093/nar/gkz412>
- He, X., Alian, A., Stroud, R., & Ortiz de Montellano, P. R. (2006). Pyrrolidine Carboxamides as a Novel Class of Inhibitors of Enoyl Acyl Carrier Protein Reductase from *Mycobacterium tuberculosis*. *Journal of Medicinal Chemistry*, 49(21), 6308–6323. <https://doi.org/10.1021/jm060715y>
- Hermann, T. (2016). Small molecules targeting viral RNA. *Wiley Interdisciplinary Reviews. RNA*, 7(6), 726–743. <https://doi.org/10.1002/wrna.1373>
- Hinton, G. E., & Roweis, S. (2002). Stochastic Neighbor Embedding. *Advances in Neural Information Processing Systems*, 15. <https://papers.nips.cc/paper/2002/hash/6150ccc6069bea6b5716254057a194ef-Abstract.html>
- Hoellman, D. B., Lin, G., Ednie, L. M., Rattan, A., Jacobs, M. R., & Appelbaum, P. C. (2003). Antipneumococcal and antistaphylococcal activities of ranbezolid (RBX 7644), a new oxazolidinone, compared to those of other agents. *Antimicrobial Agents and Chemotherapy*, 47(3), 1148–1150. <https://doi.org/10.1128/AAC.47.3.1148-1150.2003>
- Hofmarcher, M., Mayr, A., Rumetshofer, E., Ruch, P., Renz, P., Schimunek, J., Seidl, P., Vall, A., Widrich, M., Hochreiter, S., & Klambauer, G. (2020). Large-Scale Ligand-Based Virtual Screening for SARS-CoV-2 Inhibitors Using Deep Neural Networks. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3561442>
- Hou, Y., Dong, Y., Ye, T., Jiang, J., Ding, L., Qin, M., Ding, X., & Zhao, Y. (2019). Synthesis and antibacterial evaluation of novel oxazolidinone derivatives containing a piperidinyl moiety. *Bioorganic & Medicinal Chemistry Letters*, 29(23), 126746. <https://doi.org/10.1016/j.bmcl.2019.126746>
- Hoy, S. M. (2021). Contezolid: First Approval. *Drugs*, 81(13), 1587–1591. <https://doi.org/10.1007/s40265-021-01576-0>
- Huang, S.-Y., & Zou, X. (2006a). An iterative knowledge-based scoring function to predict protein–ligand interactions: I. Derivation of interaction potentials. *Journal of Computational Chemistry*, 27(15), 1866–1875. <https://doi.org/10.1002/jcc.20504>
- Huang, S.-Y., & Zou, X. (2006b). An iterative knowledge-based scoring function to predict protein-ligand interactions: II. Validation of the scoring function. *Journal*

of Computational Chemistry, 27(15), 1876–1882.

<https://doi.org/10.1002/jcc.20505>

Huang, S.-Y., & Zou, X. (2010). Advances and Challenges in Protein-Ligand Docking.

International Journal of Molecular Sciences, 11(8), Article 8.

<https://doi.org/10.3390/ijms11083016>

Hutchings, M. I., Truman, A. W., & Wilkinson, B. (2019). Antibiotics: Past, present and future. *Current Opinion in Microbiology*, 51, 72–80.

<https://doi.org/10.1016/j.mib.2019.10.008>

Hutchinson, D. K. (2003). Oxazolidinone antibacterial agents: A critical review. *Current Topics in Medicinal Chemistry*, 3(9), 1021–1042.

<https://doi.org/10.2174/1568026033452195>

Hutchinson, D. K. (2004). Recent advances in oxazolidinone antibacterial agent research.

Expert Opinion on Therapeutic Patents, 14(9), 1309–1328.

<https://doi.org/10.1517/13543776.14.9.1309>

Illingworth, C. J. R., Morris, G. M., Parkes, K. E. B., Snell, C. R., & Reynolds, C. A. (2008). Assessing the Role of Polarization in Docking. *The Journal of Physical Chemistry A*, 112(47), 12157–12163. <https://doi.org/10.1021/jp710169m>

Inoue, T., Tsunoda, A., Nishimoto, E., Nishida, K., Komatsubara, Y., Onoe, R., Saji, J., & Mineshita, M. (2017). Successful use of linezolid for refractory Mycobacterium abcessus infection: A case report. *Respiratory Medicine Case Reports*, 23, 43–45. <https://doi.org/10.1016/j.rmcr.2017.11.007>

Ippolito, J. A., Kanyo, Z. F., Wang, D., Franceschi, F. J., Moore, P. B., Steitz, T. A., & Duffy, E. M. (2008). Crystal Structure of the Oxazolidinone Antibiotic Linezolid Bound to the 50S Ribosomal Subunit. *Journal of Medicinal Chemistry*, 51(12), 3353–3356. <https://doi.org/10.1021/jm800379d>

Ivanenkov, Y. A., Zhavoronkov, A., Yamidanov, R. S., Osterman, I. A., Sergiev, P. V., Aladinskiy, V. A., Aladinskaya, A. V., Terentiev, V. A., Veselov, M. S., Ayginin, A. A., Kartsev, V. G., Skvortsov, D. A., Chemeris, A. V., Baimiev, A. Kh., Sofronova, A. A., Malyshev, A. S., Filkov, G. I., Bezrukov, D. S., Zagribelnyy, B. A., ... Dontsova, O. A. (2019). Identification of Novel Antibacterials Using Machine Learning Techniques. *Frontiers in Pharmacology*, 10.

<https://www.frontiersin.org/articles/10.3389/fphar.2019.00913>

Jang, S. Y., Ha, Y. H., Ko, S. W., Lee, W., Lee, J., Kim, S., Kim, Y. W., Lee, W. K., &

Ha, H. J. (2004). Synthesis and antibacterial activity of arylpiperazinyl

- oxazolidinones with diversification of the N-substituents. *Bioorganic and Medicinal Chemistry Letters*, 14(15), 3881–3883.
<https://doi.org/10.1016/j.bmcl.2004.05.066>
- Jeong, J.-W., Jung, S.-J., Lee, H.-H., Kim, Y.-Z., Park, T.-K., Cho, Y.-L., Chae, S.-E., Baek, S.-Y., Woo, S.-H., Lee, H.-S., & Kwak, J.-H. (2010). In vitro and in vivo activities of LCB01-0371, a new oxazolidinone. *Antimicrobial Agents and Chemotherapy*, 54(12), 5359–5362. <https://doi.org/10.1128/AAC.00723-10>
- Jin, B., Chen, J.-Y., Sheng, Z.-L., Sun, M.-Q., & Yang, H.-L. (2022). Synthesis, Antibacterial and Anthelmintic Activity of Novel 3-(3-Pyridyl)-oxazolidinone-5-methyl Ester Derivatives. *Molecules (Basel, Switzerland)*, 27(3), 1103.
<https://doi.org/10.3390/molecules27031103>
- Jin, B., Wang, T., Chen, J., Liu, X., Zhang, Y., Zhang, X., Sheng, Z., & Yang, H.-L. (2022). Synthesis and Biological Evaluation of 3-(Pyridine-3-yl)-2-Oxazolidinone Derivatives as Antibacterial Agents. *Frontiers in Chemistry*, 10.
<https://www.frontiersin.org/articles/10.3389/fchem.2022.949813>
- 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](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>
- Joseph-McCarthy, D., Thomas, B. E., Belmarsh, M., Moustakas, D., & Alvarez, J. C. (2003). Pharmacophore-based molecular docking to account for ligand flexibility. *Proteins*, 51(2), 172–188. <https://doi.org/10.1002/prot.10266>
- Jubeh, B., Breijeh, Z., & Karaman, R. (2020). Antibacterial Prodrugs to Overcome Bacterial Resistance. *Molecules*, 25(7), Article 7.
<https://doi.org/10.3390/molecules25071543>
- Jukić, M., & Bren, U. (2022). Machine Learning in Antibacterial Drug Design. *Frontiers in Pharmacology*, 13, 864412. <https://doi.org/10.3389/fphar.2022.864412>
- Kali, A., Charles, M. V. P., & Srirangaraj, S. (2015). Cadazolid: A new hope in the treatment of Clostridium difficile infection. *The Australasian Medical Journal*, 8(8), 253–262. <https://doi.org/10.4066/AMJ.2015.2441>
- Kalia, V., Miglani, R., Purnapatre, K. P., Mathur, T., Singhal, S., Khan, S., Voleti, S. R., Upadhyay, D. J., Saini, K. S., Rattan, A., & Raj, V. S. (2009). Mode of action of

- Ranbezolid against staphylococci and structural modeling studies of its interaction with ribosomes. *Antimicrobial Agents and Chemotherapy*, 53(4), 1427–1433.
<https://doi.org/10.1128/AAC.00887-08>
- Kalil, A. C., Klompas, M., Haynatzki, G., & Rupp, M. E. (2013). Treatment of hospital-acquired pneumonia with linezolid or vancomycin: A systematic review and meta-analysis. *BMJ Open*, 3(10). <https://doi.org/10.1136/bmjopen-2013-003912>
- Kallert, E., Fischer, T. R., Schneider, S., Grimm, M., Helm, M., & Kersten, C. (2022). Protein-Based Virtual Screening Tools Applied for RNA–Ligand Docking Identify New Binders of the preQ1-Riboswitch. *Journal of Chemical Information and Modeling*, 62(17), 4134–4148. <https://doi.org/10.1021/acs.jcim.2c00751>
- Kandel, D. D., Raychaudhury, C., & Pal, D. (2014). Two new atom centered fragment descriptors and scoring function enhance classification of antibacterial activity. *Journal of Molecular Modeling*, 20(4), 2164. <https://doi.org/10.1007/s00894-014-2164-1>
- Kapoor, G., Saigal, S., & Elongavan, A. (2017). Action and resistance mechanisms of antibiotics: A guide for clinicians. *Journal of Anaesthesiology, Clinical Pharmacology*, 33(3), 300–305. https://doi.org/10.4103/joacp.JOACP_349_15
- Kaushik, A., Heuer, A. M., Bell, D. T., Culhane, J. C., Ebner, D. C., Parrish, N., Ippoliti, J. T., & Lamichhane, G. (2016). An evolved oxazolidinone with selective potency against Mycobacterium tuberculosis and gram positive bacteria. *Bioorganic & Medicinal Chemistry Letters*, 26(15), 3572–3576.
<https://doi.org/10.1016/j.bmcl.2016.06.019>
- Khabbaz, H., Karimi-Jafari, M. H., Saboury, A. A., & BabaAli, B. (2021). Prediction of antimicrobial peptides toxicity based on their physico-chemical properties using machine learning techniques. *BMC Bioinformatics*, 22(1), 549.
<https://doi.org/10.1186/s12859-021-04468-y>
- Khan, P., Kader, Md. F., Islam, S. M. R., Rahman, A. B., Kamal, Md. S., Toha, M. U., & Kwak, K.-S. (2021). Machine Learning and Deep Learning Approaches for Brain Disease Diagnosis: Principles and Recent Advances. *IEEE Access*, 9, 37622–37655. <https://doi.org/10.1109/ACCESS.2021.3062484>
- Kim, J. S., Kim, Y., Lee, S. H., Kim, Y. H., Kim, J., Kang, J. Y., Kim, S. K., Kim, S. J., Kang, Y.-S., Kim, T., Mok, J., Byun, M. K., Park, H. J., Joh, J., Park, Y. B., Lim, H.-S., Choi, H., Lee, S. H., Kim, H., ... Shim, T. S. (n.d.). Early Bactericidal Activity of Delpazolid (LCB01-0371) in Patients with Pulmonary Tuberculosis.

- Antimicrobial Agents and Chemotherapy*, 66(2), e01684-21.
<https://doi.org/10.1128/aac.01684-21>
- Klevens, R. M., Morrison, M. A., Nadle, J., Petit, S., Gershman, K., Ray, S., Harrison, L. H., Lynfield, R., Dumyati, G., Townes, J. M., Craig, A. S., Zell, E. R., Fosheim, G. E., McDougal, L. K., Carey, R. B., Fridkin, S. K., & Active Bacterial Core surveillance (ABCs) MRSA Investigators. (2007). Invasive methicillin-resistant *Staphylococcus aureus* infections in the United States. *JAMA*, 298(15), 1763–1771. <https://doi.org/10.1001/jama.298.15.1763>
- Kloss, P., Xiong, L., Shinabarger, D. L., & Mankin, A. S. (1999). Resistance mutations in 23 S rRNA identify the site of action of the protein synthesis inhibitor linezolid in the ribosomal peptidyl transferase center. *Journal of Molecular Biology*, 294(1), 93–101. <https://doi.org/10.1006/jmbi.1999.3247>
- Kobak, D., & Berens, P. (2019). The art of using t-SNE for single-cell transcriptomics. *Nature Communications*, 10(1), Article 1. <https://doi.org/10.1038/s41467-019-13056-x>
- Kohanski, M. A., Dwyer, D. J., & Collins, J. J. (2010). How antibiotics kill bacteria: From targets to networks. *Nature Reviews Microbiology*, 8(6), Article 6. <https://doi.org/10.1038/nrmicro2333>
- Kokkori, S., Apostolidi, M., Tsakris, A., Pournaras, S., Stathopoulos, C., & Dinos, G. (2014). Linezolid-Dependent Function and Structure Adaptation of Ribosomes in a *Staphylococcus epidermidis* Strain Exhibiting Linezolid Dependence. *Antimicrobial Agents and Chemotherapy*, 58(8), 4651–4656. <https://doi.org/10.1128/AAC.02835-14>
- Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D. A., & Cheatham, T. E. (2000). Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Accounts of Chemical Research*, 33(12), 889–897. <https://doi.org/10.1021/ar000033j>
- Kong, Q., & Yang, Y. (2021). Recent advances in antibacterial agents. *Bioorganic & Medicinal Chemistry Letters*, 35, 127799. <https://doi.org/10.1016/j.bmcl.2021.127799>
- Krause, K. M., Serio, A. W., Kane, T. R., & Connolly, L. E. (2016). Aminoglycosides: An Overview. *Cold Spring Harbor Perspectives in Medicine*, 6(6), a027029. <https://doi.org/10.1101/cshperspect.a027029>

- Kuhn, M., Firth-Clark, S., Tosco, P., Mey, A. S. J. S., Mackey, M., & Michel, J. (2020). Assessment of Binding Affinity via Alchemical Free-Energy Calculations. *Journal of Chemical Information and Modeling*, 60(6), 3120–3130. <https://doi.org/10.1021/acs.jcim.0c00165>
- Kumar, V. (2021). Chapter 1—Introduction to ribosome factory, origin, and evolution of translation. In V. Kumar (Ed.), *Emerging Concepts in Ribosome Structure, Biogenesis, and Function* (pp. 1–13). Academic Press. <https://doi.org/10.1016/B978-0-12-816364-1.00002-0>
- Kurczyk, A., Warszycki, D., Musiol, R., Kafel, R., Bojarski, A. J., & Polanski, J. (2015). Ligand-Based Virtual Screening in a Search for Novel Anti-HIV-1 Chemotypes. *Journal of Chemical Information and Modeling*, 55(10), 2168–2177. <https://doi.org/10.1021/acs.jcim.5b00295>
- Kwon, S., Bae, H., Jo, J., & Yoon, S. (2019). Comprehensive ensemble in QSAR prediction for drug discovery. *BMC Bioinformatics*, 20(1), 521. <https://doi.org/10.1186/s12859-019-3135-4>
- Lafontaine, D. L. J., & Tollervey, D. (2001). The function and synthesis of ribosomes. *Nature Reviews Molecular Cell Biology*, 2(7), Article 7. <https://doi.org/10.1038/35080045>
- Lang, P. T., Brozell, S. R., Mukherjee, S., Pettersen, E. F., Meng, E. C., Thomas, V., Rizzo, R. C., Case, D. A., James, T. L., & Kuntz, I. D. (2009). DOCK 6: Combining techniques to model RNA-small molecule complexes. *RNA (New York, N.Y.)*, 15(6), 1219–1230. <https://doi.org/10.1261/rna.1563609>
- Leach, K. L., Swaney, S. M., Colca, J. R., McDonald, W. G., Blinn, J. R., Thomasco, L. M., Gadwood, R. C., Shinabarger, D., Xiong, L., & Mankin, A. S. (2007). The Site of Action of Oxazolidinone Antibiotics in Living Bacteria and in Human Mitochondria. *Molecular Cell*, 26(3), 393–402. <https://doi.org/10.1016/j.molcel.2007.04.005>
- Levy, S. B., & Marshall, B. (2004). Antibacterial resistance worldwide: Causes, challenges and responses. *Nature Medicine*, 10(12 Suppl), S122-129. <https://doi.org/10.1038/nm1145>
- Li, B.-B., Wu, C.-M., Wang, Y., & Shen, J.-Z. (2011). Single and dual mutations at positions 2058, 2503 and 2504 of 23S rRNA and their relationship to resistance to antibiotics that target the large ribosomal subunit. *The Journal of Antimicrobial Chemotherapy*, 66(9), 1983–1986. <https://doi.org/10.1093/jac/dkr268>

- Li, H., Sze, K.-H., Lu, G., & Ballester, P. J. (2021). Machine-learning scoring functions for structure-based virtual screening. *WIREs Computational Molecular Science*, 11(1), e1478. <https://doi.org/10.1002/wcms.1478>
- Li, J. J., & Corey, E. J. (2013). *Drug Discovery: Practices, Processes, and Perspectives*. John Wiley & Sons.
- Li, S., Wang, L., Meng, J., Zhao, Q., Zhang, L., & Liu, H. (2022). De Novo design of potential inhibitors against SARS-CoV-2 Mpro. *Computers in Biology and Medicine*, 147, 105728. <https://doi.org/10.1016/j.combiomed.2022.105728>
- Libbrecht, M. W., & Noble, W. S. (2015). Machine learning applications in genetics and genomics. *Nature Reviews Genetics*, 16(6), Article 6. <https://doi.org/10.1038/nrg3920>
- Lipinski, C. F., Matarollo, V. G., Oliveira, P. R., da Silva, A. B. F., & Honorio, K. M. (2019). Advances and Perspectives in Applying Deep Learning for Drug Design and Discovery. *Frontiers in Robotics and AI*, 6. <https://www.frontiersin.org/articles/10.3389/frobt.2019.00108>
- Lipsky, B. A., Itani, K., Norden, C., & Linezolid Diabetic Foot Infections Study Group. (2004). Treating Foot Infections in Diabetic Patients: A Randomized, Multicenter, Open-Label Trial of Linezolid versus Ampicillin-Sulbactam/Amoxicillin-Clavulanate. *Clinical Infectious Diseases*, 38(1), 17–24. <https://doi.org/10.1086/380449>
- Liu, J., & Wang, R. (2015). Classification of Current Scoring Functions. *Journal of Chemical Information and Modeling*, 55(3), 475–482. <https://doi.org/10.1021/ci500731a>
- Livermore, D. M., Warner, M., Mushtaq, S., North, S., & Woodford, N. (2007). In vitro activity of the oxazolidinone RWJ-416457 against linezolid-resistant and -susceptible staphylococci and enterococci. *Antimicrobial Agents and Chemotherapy*, 51(3), 1112–1114. <https://doi.org/10.1128/AAC.01347-06>
- Lluka, T., & Stokes, J. M. (2022). Antibiotic discovery in the artificial intelligence era. *Annals of the New York Academy of Sciences*, n/a(n/a). <https://doi.org/10.1111/nyas.14930>
- Lo, Y.-C., Rensi, S. E., Torng, W., & Altman, R. B. (2018). Machine learning in chemoinformatics and drug discovery. *Drug Discovery Today*, 23(8), 1538–1546. Scopus. <https://doi.org/10.1016/j.drudis.2018.05.010>

- Locke, J. B., Finn, J., Hilgers, M., Morales, G., Rahawi, S., G C, K., Picazo, J. J., Im, W., Shaw, K. J., & Stein, J. L. (2010). Structure-activity relationships of diverse oxazolidinones for linezolid-resistant *Staphylococcus aureus* strains possessing the cfr methyltransferase gene or ribosomal mutations. *Antimicrobial Agents and Chemotherapy*, 54(12), 5337–5343. <https://doi.org/10.1128/AAC.00663-10>
- Lohray, B. B., Gandhi, N., Srivastava, B. K., & Lohray, V. B. (2006). 3D QSAR studies of N-4-arylacryloylpiperazin-1-yl-phenyl-oxazolidinones: A novel class of antibacterial agents. *Bioorganic & Medicinal Chemistry Letters*, 16(14), 3817–3823. <https://doi.org/10.1016/j.bmcl.2006.04.023>
- Long, K. S., Munck, C., Andersen, T. M. B., Schaub, M. A., Hobbie, S. N., Böttger, E. C., & Vester, B. (2010). Mutations in 23S rRNA at the peptidyl transferase center and their relationship to linezolid binding and cross-resistance. *Antimicrobial Agents and Chemotherapy*, 54(11), 4705–4713.
<https://doi.org/10.1128/AAC.00644-10>
- Long, K. S., & Vester, B. (2012). Resistance to Linezolid Caused by Modifications at Its Binding Site on the Ribosome. *Antimicrobial Agents and Chemotherapy*, 56(2), 603–612. <https://doi.org/10.1128/AAC.05702-11>
- Lorber, D. M., & Shoichet, B. K. (1998). Flexible ligand docking using conformational ensembles. *Protein Science*, 7(4), 938–950.
<https://doi.org/10.1002/pro.5560070411>
- Luo, J., Wei, W., Waldspühl, J., & Moitessier, N. (2019). Challenges and current status of computational methods for docking small molecules to nucleic acids. *European Journal of Medicinal Chemistry*, 168, 414–425.
<https://doi.org/10.1016/j.ejmech.2019.02.046>
- 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>
- Maffioli, S. I. (2013). A Chemist's Survey of Different Antibiotic Classes. In *Antibiotics* (pp. 1–22). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9783527659685.ch1>
- Makarov, G. I., & Makarova, T. M. (2020). A noncanonical binding site of linezolid revealed via molecular dynamics simulations. *Journal of Computer-Aided Molecular Design*, 34(3), 281–291. <https://doi.org/10.1007/s10822-019-00269-x>
- Makarov, G. I., & Reshetnikova, R. V. (2021). Investigation of radezolid interaction with non-canonical chloramphenicol binding site by molecular dynamics simulations.

- Journal of Molecular Graphics and Modelling*, 105, 107902.
<https://doi.org/10.1016/j.jmgm.2021.107902>
- Malik, M. S., Faazil, S., Alsharif, M. A., Sajid Jamal, Q. M., Al-Fahemi, J. H., Banerjee, A., Chattopadhyay, A., Pal, S. K., Kamal, A., & Ahmed, S. A. (2023). Antibacterial Properties and Computational Insights of Potent Novel Linezolid-Based Oxazolidinones. *Pharmaceuticals*, 16(4), Article 4.
<https://doi.org/10.3390/ph16040516>
- Marks, J., Kannan, K., Roncase, E. J., Klepacki, D., Kefi, A., Orelle, C., Vázquez-Laslop, N., & Mankin, A. S. (2016). Context-specific inhibition of translation by ribosomal antibiotics targeting the peptidyl transferase center. *Proceedings of the National Academy of Sciences*, 113(43), 12150–12155.
<https://doi.org/10.1073/pnas.1613055113>
- Marshall, S. H., Donskey, C. J., Hutton-Thomas, R., Salata, R. A., & Rice, L. B. (2002). Gene dosage and linezolid resistance in Enterococcus faecium and Enterococcus faecalis. *Antimicrobial Agents and Chemotherapy*, 46(10), 3334–3336.
<https://doi.org/10.1128/AAC.46.10.3334-3336.2002>
- Matsingos, C., Al-Adhami, T., Jamshidi, S., Hind, C., Clifford, M., Mark Sutton, J., & Rahman, K. M. (2021). Synthesis, microbiological evaluation and structure activity relationship analysis of linezolid analogues with different C5-acylamino substituents. *Bioorganic & Medicinal Chemistry*, 49, 116397.
<https://doi.org/10.1016/j.bmc.2021.116397>
- Mayr, A., Klambauer, G., Unterthiner, T., Steijaert, M., Wegner, J. K., Ceulemans, H., Clevert, D.-A., & Hochreiter, S. (2018). Large-scale comparison of machine learning methods for drug target prediction on ChEMBL. *Chemical Science*, 9(24), 5441–5451. <https://doi.org/10.1039/C8SC00148K>
- Mazzariol, A., Lo Cascio, G., Kocsis, E., Maccacaro, L., Fontana, R., & Cornaglia, G. (2012). Outbreak of linezolid-resistant *Staphylococcus haemolyticus* in an Italian intensive care unit. *European Journal of Clinical Microbiology & Infectious Diseases: Official Publication of the European Society of Clinical Microbiology*, 31(4), 523–527. <https://doi.org/10.1007/s10096-011-1343-6>
- Mc, B., Cr, R., Ak, M., Sm, F., Dh, B., & Jj, S. (2003). Linezolid for the treatment of multidrug-resistant, gram-positive infections: Experience from a compassionate-use program. *Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America*, 36(2). <https://doi.org/10.1086/345744>

- McGann, M. R., Almond, H. R., Nicholls, A., Grant, J. A., & Brown, F. K. (2003). Gaussian docking functions. *Biopolymers*, 68(1), 76–90.
<https://doi.org/10.1002/bip.10207>
- Meka, V. G., Pillai, S. K., Sakoulas, G., Wennersten, C., Venkataraman, L., DeGirolami, P. C., Eliopoulos, G. M., Moellering, R. C., & Gold, H. S. (2004). Linezolid resistance in sequential *Staphylococcus aureus* isolates associated with a T2500A mutation in the 23S rRNA gene and loss of a single copy of rRNA. *The Journal of Infectious Diseases*, 190(2), 311–317. <https://doi.org/10.1086/421471>
- Meng, E. C., Shoichet, B. K., & Kuntz, I. D. (1992). Automated docking with grid-based energy evaluation. *Journal of Computational Chemistry*, 13(4), 505–524.
<https://doi.org/10.1002/jcc.540130412>
- Michalska, K., Karpiuk, I., Król, M., & Tyski, S. (2013). Recent development of potent analogues of oxazolidinone antibacterial agents. *Bioorganic & Medicinal Chemistry*, 21(3), 577–591. <https://doi.org/10.1016/j.bmc.2012.11.036>
- Miller, K., Dunsmore, C. J., Fishwick, C. W. G., & Chopra, I. (2008). Linezolid and tiamulin cross-resistance in *Staphylococcus aureus* mediated by point mutations in the peptidyl transferase center. *Antimicrobial Agents and Chemotherapy*, 52(5), 1737–1742. <https://doi.org/10.1128/AAC.01015-07>
- Mizutani, M. Y., Tomioka, N., & Itai, A. (1994). Rational Automatic Search Method for Stable Docking Models of Protein and Ligand. *Journal of Molecular Biology*, 243(2), 310–326. <https://doi.org/10.1006/jmbi.1994.1656>
- Moitessier, N., Englebienne, P., Lee, D., Lawandi, J., & Corbeil, C. R. (2008). Towards the development of universal, fast and highly accurate docking/scoring methods: A long way to go. *British Journal of Pharmacology*, 153(S1), S7–S26.
<https://doi.org/10.1038/sj.bjp.0707515>
- Moitessier, N., Westhof, E., & Hanessian, S. (2006). Docking of Aminoglycosides to Hydrated and Flexible RNA. *Journal of Medicinal Chemistry*, 49(3), 1023–1033.
<https://doi.org/10.1021/jm0508437>
- Mongia, M., Guler, M., & Mohimani, H. (2022). An interpretable machine learning approach to identify mechanism of action of antibiotics. *Scientific Reports*, 12(1), Article 1. <https://doi.org/10.1038/s41598-022-14229-3>
- Morgan, H. L. (1965). The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service. *Journal of Chemical Documentation*, 5(2), 107–113. <https://doi.org/10.1021/c160017a018>

- Morley, S. D., & Afshar, M. (2004). Validation of an empirical RNA-ligand scoring function for fast flexible docking using Ribodock. *Journal of Computer-Aided Molecular Design*, 18(3), 189–208.
<https://doi.org/10.1023/b:jcam.0000035199.48747.1e>
- Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, 30(16), 2785–2791. <https://doi.org/10.1002/jcc.21256>
- Morris, G. M., & Lim-Wilby, M. (2008). Molecular docking. *Methods in Molecular Biology* (Clifton, N.J.), 443, 365–382. https://doi.org/10.1007/978-1-59745-177-2_19
- Muegge, I. (2000). A knowledge-based scoring function for protein-ligand interactions: Probing the reference state. *Perspectives in Drug Discovery and Design*, 20(1), 99–114. <https://doi.org/10.1023/A:1008729005958>
- Muegge, I. (2001). Effect of ligand volume correction on PMF scoring. *Journal of Computational Chemistry*, 22(4), 418–425. [https://doi.org/10.1002/1096-987X\(200103\)22:4<418::AID-JCC1012>3.0.CO;2-3](https://doi.org/10.1002/1096-987X(200103)22:4<418::AID-JCC1012>3.0.CO;2-3)
- Muegge, I., & Martin, Y. C. (1999). A General and Fast Scoring Function for Protein–Ligand Interactions: A Simplified Potential Approach. *Journal of Medicinal Chemistry*, 42(5), 791–804. <https://doi.org/10.1021/jm980536j>
- Muñoz, P., Rodríguez-Creixéms, M., Moreno, M., Marín, M., Ramallo, V., Bouza, E., & GAME Study Group. (2007). Linezolid therapy for infective endocarditis. *Clinical Microbiology and Infection*, 13(2), 211–215. <https://doi.org/10.1111/j.1469-0691.2006.01585.x>
- Murcia-Soler, M., Pérez-Giménez, F., García-March, F. J., Salabert-Salvador, M. T., Díaz-Villanueva, W., Castro-Bleda, M. J., & Villanueva-Pareja, A. (2004). Artificial neural networks and linear discriminant analysis: A valuable combination in the selection of new antibacterial compounds. *Journal of Chemical Information and Computer Sciences*, 44(3), 1031–1041.
<https://doi.org/10.1021/ci030340e>
- Murray, C. W., Auton, T. R., & Eldridge, M. D. (1998). Empirical scoring functions. II. The testing of an empirical scoring function for the prediction of ligand-receptor binding affinities and the use of Bayesian regression to improve the quality of the

- model. *Journal of Computer-Aided Molecular Design*, 12(5), 503–519.
<https://doi.org/10.1023/a:1008040323669>
- Myint, K.-Z., Wang, L., Tong, Q., & Xie, X.-Q. (2012). Molecular Fingerprint-Based Artificial Neural Networks QSAR for Ligand Biological Activity Predictions. *Molecular Pharmaceutics*, 9(10), 2912–2923. <https://doi.org/10.1021/mp300237z>
- Neves, B. J., Braga, R. C., Melo-Filho, C. C., Moreira-Filho, J. T., Muratov, E. N., & Andrade, C. H. (2018). QSAR-Based Virtual Screening: Advances and Applications in Drug Discovery. *Frontiers in Pharmacology*, 9.
<https://www.frontiersin.org/articles/10.3389/fphar.2018.01275>
- Nguyen, M., Olson, R., Shukla, M., VanOeffelen, M., & Davis, J. J. (2020). Predicting antimicrobial resistance using conserved genes. *PLOS Computational Biology*, 16(10), e1008319. <https://doi.org/10.1371/journal.pcbi.1008319>
- Oong, G. C., & Tadi, P. (2022). Chloramphenicol. In *StatPearls*. StatPearls Publishing.
<http://www.ncbi.nlm.nih.gov/books/NBK555966/>
- Orac, C. M., Zhou, S., Means, J. A., Boehm, D., Bergmeier, S. C., & Hines, J. V. (2011). Synthesis and Stereospecificity of 4,5-Disubstituted Oxazolidinone Ligands Binding to T-box Riboswitch RNA. *Journal of Medicinal Chemistry*, 54(19), 6786–6795. <https://doi.org/10.1021/jm2006904>
- Ortiz, A. R., Pisabarro, M. T., Gago, F., & Wade, R. C. (1995). Prediction of Drug Binding Affinities by Comparative Binding Energy Analysis. *Journal of Medicinal Chemistry*, 38(14), 2681–2691. <https://doi.org/10.1021/jm00014a020>
- Padmanabhan, R. A., LaRosa, S. P., & Tomecki, K. J. (2005). What's New in Antibiotics? *Dermatologic Clinics*, 23(2), 301–312.
<https://doi.org/10.1016/j.det.2005.01.003>
- Pae, A. N., Kim, S. Y., Kim, H. Y., Joo, H. J., Cho, Y. S., Choi, K. I., Choi, J. H., & Koh, H. Y. (1999). 3D QSAR studies on new oxazolidinone antibacterial agents by comparative molecular field analysis. *Bioorganic & Medicinal Chemistry Letters*, 9(18), 2685–2690. [https://doi.org/10.1016/S0960-894X\(99\)00474-6](https://doi.org/10.1016/S0960-894X(99)00474-6)
- Panigrahi, N., Ganguly, S., & Panda, J. (2018). Synthesis, Antimicrobial Evaluation and Molecular Docking Studies of Novel Oxazolidinone-Thiophene Chalcone Hybrid Derivatives. *Research Journal of Pharmacy and Technology*, 11(12), 5611–5622.
<https://doi.org/10.5958/0974-360X.2018.01019.3>
- Parai, M. K., Huggins, D. J., Cao, H., Nalam, M. N. L., Ali, A., Schiffer, C. A., Tidor, B., & Rana, T. M. (2012). Design, Synthesis, and Biological and Structural

- Evaluations of Novel HIV-1 Protease Inhibitors To Combat Drug Resistance. *Journal of Medicinal Chemistry*, 55(14), 6328–6341. <https://doi.org/10.1021/jm300238h>
- Park, S.-J., Kim, Y.-G., & Park, H.-J. (2011). Identification of RNA Pseudoknot-Binding Ligand That Inhibits the –1 Ribosomal Frameshifting of SARS-Coronavirus by Structure-Based Virtual Screening. *Journal of the American Chemical Society*, 133(26), 10094–10100. <https://doi.org/10.1021/ja1098325>
- Patel, L., Shukla, T., Huang, X., Ussery, D. W., & Wang, S. (2020). Machine Learning Methods in Drug Discovery. *Molecules (Basel, Switzerland)*, 25(22), E5277. <https://doi.org/10.3390/molecules25225277>
- Patel, P. H., & Hashmi, M. F. (2022). Macrolides. In *StatPearls*. StatPearls Publishing. <http://www.ncbi.nlm.nih.gov/books/NBK551495/>
- Patel, V., & Shah, M. (2022). Artificial intelligence and machine learning in drug discovery and development. *Intelligent Medicine*, 2(3), 134–140. Scopus. <https://doi.org/10.1016/j.imed.2021.10.001>
- Peach, M. L., & Nicklaus, M. C. (2009). Combining docking with pharmacophore filtering for improved virtual screening. *Journal of Cheminformatics*, 1, 6. <https://doi.org/10.1186/1758-2946-1-6>
- Pearcy, N., Hu, Y., Baker, M., Maciel-Guerra, A., Xue, N., Wang, W., Kaler, J., Peng, Z., Li, F., & Dottorini, T. (n.d.). Genome-Scale Metabolic Models and Machine Learning Reveal Genetic Determinants of Antibiotic Resistance in Escherichia coli and Unravel the Underlying Metabolic Adaptation Mechanisms. *MSystems*, 6(4), e00913-20. <https://doi.org/10.1128/mSystems.00913-20>
- Pei, J., Wang, Q., Liu, Z., Li, Q., Yang, K., & Lai, L. (2006). PSI-DOCK: Towards highly efficient and accurate flexible ligand docking. *Proteins: Structure, Function, and Bioinformatics*, 62(4), 934–946. <https://doi.org/10.1002/prot.20790>
- Pereira, J. C., Caffarena, E. R., & dos Santos, C. N. (2016). Boosting Docking-Based Virtual Screening with Deep Learning. *Journal of Chemical Information and Modeling*, 56(12), 2495–2506. <https://doi.org/10.1021/acs.jcim.6b00355>
- Peters, J., Kondo, K. L., Lee, R. K., Lin, C. K., & Inderlied, C. B. (1995). In-vitro activity of oxazolidinones against Mycobacterium avium complex. *The Journal of Antimicrobial Chemotherapy*, 35(5), 675–679. <https://doi.org/10.1093/jac/35.5.675>

- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13), 1605–1612. <https://doi.org/10.1002/jcc.20084>
- Pfeffer, P., & Gohlke, H. (2007). DrugScoreRNA--knowledge-based scoring function to predict RNA-ligand interactions. *Journal of Chemical Information and Modeling*, 47(5), 1868–1876. <https://doi.org/10.1021/ci700134p>
- Philips, A., Milanowska, K., Lach, G., & Bujnicki, J. M. (2013). LigandRNA: Computational predictor of RNA-ligand interactions. *RNA (New York, N.Y.)*, 19(12), 1605–1616. <https://doi.org/10.1261/rna.039834.113>
- Pintilie, L., Stefanici, A., Nicu, A. I., Negut, C., Tanase, C., & Caproiu, M. T. (2018). Design, Synthesis and Molecular Docking of some Oxazolidinone Compounds. *Revista de Chimie*, 69(11), 2981–2986. <https://doi.org/10.37358/RC.18.11.6666>
- Poce, G., Zappia, G., Porretta, G. C., Botta, B., & Biava, M. (2008). New oxazolidinone derivatives as antibacterial agents with improved activity. *Expert Opinion on Therapeutic Patents*, 18(2), 97–121. <https://doi.org/10.1517/13543776.18.2.97>
- Principi, N., Argentiero, A., Neglia, C., Gramegna, A., & Esposito, S. (2020). New Antibiotics for the Treatment of Acute Bacterial Skin and Soft Tissue Infections in Pediatrics. *Pharmaceuticals*, 13(11), Article 11. <https://doi.org/10.3390/ph13110333>
- Prystowsky, J., Siddiqui, F., Chosay, J., Shinabarger, D. L., Millichap, J., Peterson, L. R., & Noskin, G. A. (2001). Resistance to linezolid: Characterization of mutations in rRNA and comparison of their occurrences in vancomycin-resistant enterococci. *Antimicrobial Agents and Chemotherapy*, 45(7), 2154–2156. <https://doi.org/10.1128/AAC.45.7.2154-2156.2001>
- Rahuman, M. H., Muthu, S., Raajaraman, B. R., Raja, M., & Umamahesvari, H. (2020). Investigations on 2-(4-Cyanophenylamino) acetic acid by FT-IR,FT-Raman, NMR and UV-Vis spectroscopy, DFT (NBO, HOMO-LUMO, MEP and Fukui function) and molecular docking studies. *Heliyon*, 6(9). <https://doi.org/10.1016/j.heliyon.2020.e04976>
- Ramsundar, B., Eastman, P., Walters, P., & Pande, V. (2019). *Deep Learning for the Life Sciences: Applying Deep Learning to Genomics, Microscopy, Drug Discovery, and More*. O'Reilly Media.

- Rayner, C. R., Baddour, L. M., Birmingham, M. C., Norden, C., Meagher, A. K., & Schentag, J. J. (2004). Linezolid in the treatment of osteomyelitis: Results of compassionate use experience. *Infection*, 32(1), 8–14.
<https://doi.org/10.1007/s15010-004-3029-9>
- Ren, Y., Chakraborty, T., Doijad, S., Falgenhauer, L., Falgenhauer, J., Goesmann, A., Hauschild, A.-C., Schwengers, O., & Heider, D. (2021). Prediction of antimicrobial resistance based on whole-genome sequencing and machine learning. *Bioinformatics*, 38(2), 325–334.
<https://doi.org/10.1093/bioinformatics/btab681>
- Reymond, J.-L. (2015). The Chemical Space Project. *Accounts of Chemical Research*, 48(3), 722–730. <https://doi.org/10.1021/ar500432k>
- Romeo, R., Chiacchio, M. A., Campisi, A., Monciino, G., Veltri, L., Iannazzo, D., Broggini, G., & Giofrè, S. V. (2018). Synthesis and Biological Evaluation of Pyrimidine-oxazolidin-2-arylimino Hybrid Molecules as Antibacterial Agents. *Molecules*, 23(7), Article 7. <https://doi.org/10.3390/molecules23071754>
- Roy, K., Kar, S., & Das, R. N. (2015). Chapter 10—Other Related Techniques. In K. Roy, S. Kar, & R. N. Das (Eds.), *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment* (pp. 357–425). Academic Press. <https://doi.org/10.1016/B978-0-12-801505-6.00010-7>
- Ruiz-Carmona, S., Alvarez-Garcia, D., Foloppe, N., Garmendia-Doval, A. B., Juhos, S., Schmidtke, P., Barril, X., Hubbard, R. E., & Morley, S. D. (2014). rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids. *PLOS Computational Biology*, 10(4), e1003571.
<https://doi.org/10.1371/journal.pcbi.1003571>
- Sabet, R., Fassihi, A., Hemmateenejad, B., Saghaei, L., Miri, R., & Gholami, M. (2012). Computer-aided design of novel antibacterial 3-hydroxypyridine-4-ones: Application of QSAR methods based on the MOLMAP approach. *Journal of Computer-Aided Molecular Design*, 26(3), 349–361.
<https://doi.org/10.1007/s10822-012-9561-2>
- Sahasranaman, A., & Woolford, J. L. (2013). Ribosome Assembly. In W. J. Lennarz & M. D. Lane (Eds.), *Encyclopedia of Biological Chemistry (Second Edition)* (pp. 116–121). Academic Press. <https://doi.org/10.1016/B978-0-12-378630-2.00259-0>
- Saini, J. S., Homeyer, N., Fulle, S., & Gohlke, H. (2013). Determinants of the species selectivity of oxazolidinone antibiotics targeting the large ribosomal subunit.

Biological Chemistry, 394(11), 1529–1541. <https://doi.org/10.1515/hsz-2013-0188>

Samdani, A., & Vetrivel, U. (2018). POAP: A GNU parallel based multithreaded pipeline of open babel and AutoDock suite for boosted high throughput virtual screening.

Computational Biology and Chemistry, 74, 39–48.

<https://doi.org/10.1016/j.compbiochem.2018.02.012>

Sander, P., Belova, L., Kidan, Y. G., Pfister, P., Mankin, A. S., & Böttger, E. C. (2002). Ribosomal and non-ribosomal resistance to oxazolidinones: Species-specific idiosyncrasy of ribosomal alterations. *Molecular Microbiology*, 46(5), 1295–1304. <https://doi.org/10.1046/j.1365-2958.2002.03242.x>

Sauton, N., Lagorce, D., Villoutreix, B. O., & Miteva, M. A. (2008). MS-DOCK: Accurate multiple conformation generator and rigid docking protocol for multi-step virtual ligand screening. *BMC Bioinformatics*, 9(1), 184. <https://doi.org/10.1186/1471-2105-9-184>

Scaiola, A., Leibundgut, M., Boehringer, D., Caspers, P., Bur, D., Locher, H. H., Rueedi, G., & Ritz, D. (2019). Structural basis of translation inhibition by cadazolid, a novel quinoxolidinone antibiotic. *Scientific Reports*, 9(1), 5634. <https://doi.org/10.1038/s41598-019-42155-4>

Schapire, R. E. (2013). Explaining AdaBoost. In B. Schölkopf, Z. Luo, & V. Vovk (Eds.), *Empirical Inference* (pp. 37–52). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-41136-6_5

Schlünzen, F., Zarivach, R., Harms, J., Bashan, A., Tocilj, A., Albrecht, R., Yonath, A., & Franceschi, F. (2001). Structural basis for the interaction of antibiotics with the peptidyl transferase centre in eubacteria. *Nature*, 413(6858), Article 6858. <https://doi.org/10.1038/35101544>

Schmeing, T. M. (2013). Ribosome Structure. In W. J. Lennarz & M. D. Lane (Eds.), *Encyclopedia of Biological Chemistry (Second Edition)* (pp. 128–135). Academic Press. <https://doi.org/10.1016/B978-0-12-378630-2.00400-X>

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, 85–117. <https://doi.org/10.1016/j.neunet.2014.09.003>

Schuwirth, B. S., Borovinskaya, M. A., Hau, C. W., Zhang, W., Vila-Sanjurjo, A., Holton, J. M., & Cate, J. H. D. (2005). Structures of the Bacterial Ribosome at 3.5 Å Resolution. *Science*, 310(5749), 827–834. <https://doi.org/10.1126/science.1117230>

- Segler, M. H. S., Kogej, T., Tyrchan, C., & Waller, M. P. (2018). Generating Focused Molecule Libraries for Drug Discovery with Recurrent Neural Networks. *ACS Central Science*, 4(1), 120–131. <https://doi.org/10.1021/acscentsci.7b00512>
- Senneville, E., Legout, L., Valette, M., Yazdanpanah, Y., Beltrand, E., Caillaux, M., Migaud, H., & Mouton, Y. (2006). Effectiveness and tolerability of prolonged linezolid treatment for chronic osteomyelitis: A retrospective study. *Clinical Therapeutics*, 28(8), 1155–1163. <https://doi.org/10.1016/j.clinthera.2006.08.001>
- Serafim, M. S. M., Kronenberger, T., Oliveira, P. R., Poso, A., Honório, K. M., Mota, B. E. F., & Maltarollo, V. G. (2020). The application of machine learning techniques to innovative antibacterial discovery and development. *Expert Opinion on Drug Discovery*, 15(10), 1165–1180. <https://doi.org/10.1080/17460441.2020.1776696>
- Shaban, T. F., & Alkawareek, M. Y. (2022). Prediction of qualitative antibiofilm activity of antibiotics using supervised machine learning techniques. *Computers in Biology and Medicine*, 140, 105065.
<https://doi.org/10.1016/j.combiomed.2021.105065>
- Shajani, Z., Sykes, M. T., & Williamson, J. R. (2011). Assembly of Bacterial Ribosomes. *Annual Review of Biochemistry*, 80(1), 501–526. <https://doi.org/10.1146/annurev-biochem-062608-160432>
- Shayanfar, S., & Shayanfar, A. (2022). Comparison of various methods for validity evaluation of QSAR models. *BMC Chemistry*, 16(1), 63.
<https://doi.org/10.1186/s13065-022-00856-4>
- Shutter, M. C., & Akhondi, H. (2022). Tetracycline. In *StatPearls*. StatPearls Publishing.
<http://www.ncbi.nlm.nih.gov/books/NBK549905/>
- Singh, A. K., Noronha, V., Gupta, A., Singh, D., Singh, P., Singh, A., & Singh, A. (2020). Rivaroxaban: Drug review. *Cancer Research, Statistics, and Treatment*, 3(2), 264. https://doi.org/10.4103/CRST.CRST_122_19
- Singh, N., Chaudhury, S., Liu, R., AbdulHameed, M. D. M., Tawa, G., & Wallqvist, A. (2012). QSAR Classification Model for Antibacterial Compounds and Its Use in Virtual Screening. *Journal of Chemical Information and Modeling*, 52(10), 2559–2569. <https://doi.org/10.1021/ci300336v>
- Slee, A. M., Wuonola, M. A., McRipley, R. J., Zajac, I., Zawada, M. J., Bartholomew, P. T., Gregory, W. A., & Forbes, M. (1987). Oxazolidinones, a new class of synthetic antibacterial agents: In vitro and in vivo activities of DuP 105 and DuP

721. *Antimicrobial Agents and Chemotherapy*, 31(11), 1791–1797.

<https://doi.org/10.1128/AAC.31.11.1791>

Srinivasan, J., Cheatham, T. E., Cieplak, P., Kollman, P. A., & Case, D. A. (1998).

Continuum Solvent Studies of the Stability of DNA, RNA, and

Phosphoramidate–DNA Helices. *Journal of the American Chemical Society*, 120(37), 9401–9409. <https://doi.org/10.1021/ja981844+>

Stefani, S., Bongiorno, D., Mongelli, G., & Campanile, F. (2010). Linezolid Resistance in Staphylococci. *Pharmaceuticals*, 3(7), Article 7.

<https://doi.org/10.3390/ph3071988>

Stefaniak, F., & Bujnicki, J. M. (2021). AnnapuRNA: A scoring function for predicting RNA-small molecule binding poses. *PLOS Computational Biology*, 17(2), e1008309. <https://doi.org/10.1371/journal.pcbi.1008309>

Stevens, D. L., Dotter, B., & Madaras-Kelly, K. (2004). A review of linezolid: The first oxazolidinone antibiotic. *Expert Review of Anti-Infective Therapy*, 2(1), 51–59. <https://doi.org/10.1586/14787210.2.1.51>

Stokes, J. M., Yang, K., Swanson, K., Jin, W., Cubillos-Ruiz, A., Donghia, N. M., MacNair, C. R., French, S., Carfrae, L. A., Bloom-Ackermann, Z., Tran, V. M., Chiappino-Pepe, A., Badran, A. H., Andrews, I. W., Chory, E. J., Church, G. M., Brown, E. D., Jaakkola, T. S., Barzilay, R., & Collins, J. J. (2020a). A Deep Learning Approach to Antibiotic Discovery. *Cell*, 180(4), 688–702.e13. <https://doi.org/10.1016/j.cell.2020.01.021>

Stokes, J. M., Yang, K., Swanson, K., Jin, W., Cubillos-Ruiz, A., Donghia, N. M., MacNair, C. R., French, S., Carfrae, L. A., Bloom-Ackermann, Z., Tran, V. M., Chiappino-Pepe, A., Badran, A. H., Andrews, I. W., Chory, E. J., Church, G. M., Brown, E. D., Jaakkola, T. S., Barzilay, R., & Collins, J. J. (2020b). A Deep Learning Approach to Antibiotic Discovery. *Cell*, 181(2), 475–483. <https://doi.org/10.1016/j.cell.2020.04.001>

Strobl, C., Malley, J., & Tutz, G. (2009). An Introduction to Recursive Partitioning: Rationale, Application and Characteristics of Classification and Regression Trees, Bagging and Random Forests. *Psychological Methods*, 14(4), 323–348. <https://doi.org/10.1037/a0016973>

Sun, L.-Z., Jiang, Y., Zhou, Y., & Chen, S.-J. (2020). RLDOCK: A New Method for Predicting RNA–Ligand Interactions. *Journal of Chemical Theory and Computation*, 16(11), 7173–7183. <https://doi.org/10.1021/acs.jctc.0c00798>

- Sun, M., Zhao, S., Gilvary, C., Elemento, O., Zhou, J., & Wang, F. (2020). Graph convolutional networks for computational drug development and discovery. *Briefings in Bioinformatics*, 21(3), 919–935. <https://doi.org/10.1093/bib/bbz042>
- Taminau, J., Thijs, G., & De Winter, H. (2008). Pharao: Pharmacophore alignment and optimization. *Journal of Molecular Graphics & Modelling*, 27(2), 161–169. <https://doi.org/10.1016/j.jmgm.2008.04.003>
- Tasic, B., Yao, Z., Graybuck, L. T., Smith, K. A., Nguyen, T. N., Bertagnolli, D., Goldy, J., Garren, E., Economo, M. N., Viswanathan, S., Penn, O., Bakken, T., Menon, V., Miller, J., Fong, O., Hirokawa, K. E., Lathia, K., Rimorin, C., Tieu, M., ... Zeng, H. (2018). Shared and distinct transcriptomic cell types across neocortical areas. *Nature*, 563(7729), Article 7729. <https://doi.org/10.1038/s41586-018-0654-5>
- Tessaro, F., & Scapozza, L. (2020). How ‘Protein-Docking’ Translates into the New Emerging Field of Docking Small Molecules to Nucleic Acids? *Molecules*, 25(12), 2749. <https://doi.org/10.3390/molecules25122749>
- Tetali, S. R., Kunapaeddi, E., Mailavaram, R. P., Singh, V., Borah, P., Deb, P. K., Venugopala, K. N., Hourani, W., & Tekade, R. K. (2020). Current advances in the clinical development of anti-tubercular agents. *Tuberculosis*, 125, 101989. <https://doi.org/10.1016/j.tube.2020.101989>
- Thomas, K. D., Adhikari, A. V., Chowdhury, I. H., Sandeep, T., Mahmood, R., Bhattacharya, B., & Sumesh, E. (2011). Design, synthesis and docking studies of quinoline-oxazolidinone hybrid molecules and their antitubercular properties. *European Journal of Medicinal Chemistry*, 46(10), 4834–4845. <https://doi.org/10.1016/j.ejmech.2011.07.049>
- Thomsen, R., & Christensen, M. H. (2006). MolDock: A New Technique for High-Accuracy Molecular Docking. *Journal of Medicinal Chemistry*, 49(11), 3315–3321. <https://doi.org/10.1021/jm051197e>
- Tokuyama, R., Takahashi, Y., Tomita, Y., Tsubouchi, M., Yoshida, T., Iwasaki, N., Kado, N., Okezaki, E., & Nagata, O. (2001). Structure-Activity Relationship (SAR) Studies on Oxazolidinone Antibacterial Agents. 2.¹ Relationship between Lipophilicity and Antibacterial Activity in 5-Thiocarbonyl Oxazolidinones. *Chemical and Pharmaceutical Bulletin*, 49(4), 353–360. <https://doi.org/10.1248/cpb.49.353>

- Tomasello, G., Armenia, I., & Molla, G. (2020). The Protein Imager: A full-featured online molecular viewer interface with server-side HQ-rendering capabilities. *Bioinformatics*, 36(9), 2909–2911. <https://doi.org/10.1093/bioinformatics/btaa009>
- Torres, P. H. M., Sodero, A. C. R., Jofily, P., & Silva-Jr, F. P. (2019). Key Topics in Molecular Docking for Drug Design. *International Journal of Molecular Sciences*, 20(18), 4574. <https://doi.org/10.3390/ijms20184574>
- Trott, O., & Olson, A. J. (2010). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *Journal of Computational Chemistry*, 31(2), 455–461.
<https://doi.org/10.1002/jcc.21334>
- Tuszynska, I., Magnus, M., Jonak, K., Dawson, W., & Bujnicki, J. M. (2015). NP Dock: A web server for protein–nucleic acid docking. *Nucleic Acids Research*, 43(Web Server issue), W425–W430. <https://doi.org/10.1093/nar/gkv493>
- Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., Li, B., Madabhushi, A., Shah, P., Spitzer, M., & Zhao, S. (2019). Applications of machine learning in drug discovery and development. *Nature Reviews Drug Discovery*, 18(6), Article 6. <https://doi.org/10.1038/s41573-019-0024-5>
- van Bambeke, F., Mingeot-Leclercq, M.-P., Glupczynski, Y., & Tulkens, P. M. (2017). 137—Mechanisms of Action. In J. Cohen, W. G. Powderly, & S. M. Opal (Eds.), *Infectious Diseases (Fourth Edition)* (pp. 1162-1180.e1). Elsevier.
<https://doi.org/10.1016/B978-0-7020-6285-8.00137-4>
- Varela-Rial, A., Majewski, M., & De Fabritiis, G. (2022). Structure based virtual screening: Fast and slow. *WIREs Computational Molecular Science*, 12(2), e1544.
<https://doi.org/10.1002/wcms.1544>
- Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W., & Taylor, R. D. (2003). Improved protein–ligand docking using GOLD. *Proteins: Structure, Function, and Bioinformatics*, 52(4), 609–623. <https://doi.org/10.1002/prot.10465>
- Verkhivker, G., Appelt, K., Freer, S. T., & Villafranca, J. E. (1995). Empirical free energy calculations of ligand-protein crystallographic complexes. I. Knowledge-based ligand-protein interaction potentials applied to the prediction of human immunodeficiency virus 1 protease binding affinity. *Protein Engineering*, 8(7), 677–691. <https://doi.org/10.1093/protein/8.7.677>

- Wallach, I., Dzamba, M., & Heifets, A. (2015). *AtomNet: A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-based Drug Discovery* (arXiv:1510.02855). arXiv. <https://doi.org/10.48550/arXiv.1510.02855>
- Wang, H., Qin, Z., & Yan, A. (2021). Classification models and SAR analysis on CysLT1 receptor antagonists using machine learning algorithms. *Molecular Diversity*, 25(3), 1597–1616. <https://doi.org/10.1007/s11030-020-10165-4>
- Wang, L., Le, X., Li, L., Ju, Y., Lin, Z., Gu, Q., & Xu, J. (2014). Discovering New Agents Active against Methicillin-Resistant *Staphylococcus aureus* with Ligand-Based Approaches. *Journal of Chemical Information and Modeling*, 54(11), 3186–3197. <https://doi.org/10.1021/ci500253q>
- Wang, R., Lai, L., & Wang, S. (2002). Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *Journal of Computer-Aided Molecular Design*, 16(1), 11–26. <https://doi.org/10.1023/a:1016357811882>
- Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., & Weiner, P. (1984). A new force field for molecular mechanical simulation of nucleic acids and proteins. *Journal of the American Chemical Society*, 106(3), 765–784. <https://doi.org/10.1021/ja00315a051>
- Weiner, S. J., Kollman, P. A., Nguyen, D. T., & Case, D. A. (1986). An all atom force field for simulations of proteins and nucleic acids. *Journal of Computational Chemistry*, 7(2), 230–252. <https://doi.org/10.1002/jcc.540070216>
- Wichapong, K., Rohe, A., Platzer, C., Slyko, I., Erdmann, F., Schmidt, M., & Sippl, W. (2014). Application of Docking and QM/MM-GBSA Rescoring to Screen for Novel Myt1 Kinase Inhibitors. *Journal of Chemical Information and Modeling*, 54(3), 881–893. <https://doi.org/10.1021/ci4007326>
- Wilson, D. N. (2014). Ribosome-targeting antibiotics and mechanisms of bacterial resistance. *Nature Reviews Microbiology*, 12(1), Article 1. <https://doi.org/10.1038/nrmicro3155>
- Wilson, D. N., Schluenzen, F., Harms, J. M., Starosta, A. L., Connell, S. R., & Fucini, P. (2008). The oxazolidinone antibiotics perturb the ribosomal peptidyl-transferase center and effect tRNA positioning. *Proceedings of the National Academy of Sciences of the United States of America*, 105(36), 13339–13344. <https://doi.org/10.1073/pnas.0804276105>

- Wong, A., Reddy, S. P., Smyth, D. S., Aguero-Rosenfeld, M. E., Sakoulas, G., & Robinson, D. A. (2010). Polyphyletic emergence of linezolid-resistant staphylococci in the United States. *Antimicrobial Agents and Chemotherapy*, 54(2), 742–748. <https://doi.org/10.1128/AAC.00621-09>
- Wright, A., Deane-Alder, K., Marschall, E., Bamert, R., Venugopal, H., Lithgow, T., Lupton, D. W., & Belousoff, M. J. (2020). Characterization of the Core Ribosomal Binding Region for the Oxazolidone Family of Antibiotics Using Cryo-EM. *ACS Pharmacology & Translational Science*, 3(3), 425–432. <https://doi.org/10.1021/acsptsci.0c00041>
- Xie, X., Baird, D., Bowen, K., Capka, V., Chen, J., Chenail, G., Cho, Y., Dooley, J., Farsidjani, A., Fortin, P., Kohls, D., Kulathila, R., Lin, F., McKay, D., Rodrigues, L., Sage, D., Touré, B. B., Plas, S. van der, Wright, K., ... Pagliarini, R. A. (2017). Allosteric Mutant IDH1 Inhibitors Reveal Mechanisms for IDH1 Mutant and Isoform Selectivity. *Structure*, 25(3), 506–513. <https://doi.org/10.1016/j.str.2016.12.017>
- Xin, Q., Fan, H., Guo, B., He, H., Gao, S., Wang, H., Huang, Y., & Yang, Y. (2011). Design, synthesis, and structure-activity relationship studies of highly potent novel benzoxazinyl-oxazolidinone antibacterial agents. *Journal of Medicinal Chemistry*, 54(21), 7493–7502. <https://doi.org/10.1021/jm200614t>
- Xiong, L., Kloss, P., Douthwaite, S., Andersen, N. M., Swaney, S., Shinabarger, D. L., & Mankin, A. S. (2000). Oxazolidinone resistance mutations in 23S rRNA of *Escherichia coli* reveal the central region of domain V as the primary site of drug action. *Journal of Bacteriology*, 182(19), 5325–5331. <https://doi.org/10.1128/JB.182.19.5325-5331.2000>
- Yan, L., Wu, J., Chen, H., Zhang, S., Wang, Z., Wang, H., & Wu, F. (2015). Synthesis and in vitro antibacterial activity of novel fluoroalkyl-substituted pyrazolyl oxazolidinones. *RSC Advances*, 5(90), 73660–73669. <https://doi.org/10.1039/C5RA11782H>
- Yan, Y., Tao, H., He, J., & Huang, S.-Y. (2020). The HDOCK server for integrated protein–protein docking. *Nature Protocols*, 15(5), Article 5. <https://doi.org/10.1038/s41596-020-0312-x>
- Yan, Z., & Wang, J. (2017). SPA-LN: A scoring function of ligand–nucleic acid interactions via optimizing both specificity and affinity. *Nucleic Acids Research*, 45(12), e110. <https://doi.org/10.1093/nar/gkx255>

- Yap, C. W. (2011). PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *Journal of Computational Chemistry*, 32(7), 1466–1474. <https://doi.org/10.1002/jcc.21707>
- Yin, S., Biedermannova, L., Vondrasek, J., & Dokholyan, N. V. (2008). MedusaScore: An Accurate Force Field-Based Scoring Function for Virtual Drug Screening. *Journal of Chemical Information and Modeling*, 48(8), 1656–1662. <https://doi.org/10.1021/ci8001167>
- Zhanel, G. G., Shroeder, C., Vercaigne, L., S Gin, A., Embil, J., & J Hoban, D. (2001). A critical review of oxazolidinones: An alternative or replacement for glycopeptides and streptogramins? *The Canadian Journal of Infectious Diseases*, 12(6), 379–390. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2094842/>
- Zhang, X., Perez-Sanchez, H., & Lightstone, F. C. (2017). A Comprehensive Docking and MM/GBSA Rescoring Study of Ligand Recognition upon Binding Antithrombin. *Current Topics in Medicinal Chemistry*, 17(14), 1631–1639. <https://doi.org/10.2174/1568026616666161117112604>
- Zhang, Z. (2016). Introduction to machine learning: K-nearest neighbors. *Annals of Translational Medicine*, 4(11), Article 11. <https://doi.org/10.21037/atm.2016.03.37>
- Zhao, H., Lu, Y., Sheng, L., Yuan, Z., Wang, B., Wang, W., Li, Y., Ma, C., Wang, X., Zhang, D., & Huang, H. (2017). Discovery of Fluorine-Containing Benzoxazinyl-oxazolidinones for the Treatment of Multidrug Resistant Tuberculosis. *ACS Medicinal Chemistry Letters*, 8(5), 533–537. <https://doi.org/10.1021/acsmmedchemlett.7b00068>
- Zhao, P. (2011). The 2009 Nobel Prize in Chemistry: Thomas A. Steitz and the structure of the ribosome. *The Yale Journal of Biology and Medicine*, 84(2), 125–129.
- Zhao, Q., Xin, L., Liu, Y., Liang, C., Li, J., Jian, Y., Li, H., Shi, Z., Liu, H., & Cao, W. (2021). Current Landscape and Future Perspective of Oxazolidinone Scaffolds Containing Antibacterial Drugs. *Journal of Medicinal Chemistry*, 64(15), 10557–10580. <https://doi.org/10.1021/acs.jmedchem.1c00480>
- Zheng, Z., & Merz, K. M. Jr. (2013). Development of the Knowledge-Based and Empirical Combined Scoring Algorithm (KECSA) To Score Protein–Ligand Interactions. *Journal of Chemical Information and Modeling*, 53(5), 1073–1083. <https://doi.org/10.1021/ci300619x>

- Zhou, J., Bhattacharjee, A., Chen, S., Chen, Y., Duffy, E., Farmer, J., Goldberg, J., Hanselmann, R., Ippolito, J. A., Lou, R., Orbin, A., Oyelere, A., Salvino, J., Springer, D., Tran, J., Wang, D., Wu, Y., & Johnson, G. (2008a). Design at the atomic level: Design of biarylloxazolidinones as potent orally active antibiotics. *Bioorganic & Medicinal Chemistry Letters*, 18(23), 6175–6178.
<https://doi.org/10.1016/j.bmcl.2008.10.011>
- Zhou, J., Bhattacharjee, A., Chen, S., Chen, Y., Duffy, E., Farmer, J., Goldberg, J., Hanselmann, R., Ippolito, J. A., Lou, R., Orbin, A., Oyelere, A., Salvino, J., Springer, D., Tran, J., Wang, D., Wu, Y., & Johnson, G. (2008b). Design at the atomic level: Generation of novel hybrid biarylloxazolidinones as promising new antibiotics. *Bioorganic & Medicinal Chemistry Letters*, 18(23), 6179–6183.
<https://doi.org/10.1016/j.bmcl.2008.10.014>
- Zhou, J., Wu, S., Lee, B. G., Chen, T., He, Z., Lei, Y., Tang, B., & Hirst, J. D. (2021). Machine-Learning-Enabled Virtual Screening for Inhibitors of Lysine-Specific Histone Demethylase 1. *Molecules*, 26(24), Article 24.
<https://doi.org/10.3390/molecules26247492>
- Zhou, Y., Jiang, Y., & Chen, S.-J. (2022). RNA–ligand molecular docking: Advances and challenges. *WIREs Computational Molecular Science*, 12(3), e1571.
<https://doi.org/10.1002/wcms.1571>
- Zhu, Z., Shi, C., Zhang, Z., Liu, S., Xu, M., Yuan, X., Zhang, Y., Chen, J., Cai, H., Lu, J., Ma, C., Liu, R., Xhonouex, L.-P., Qu, M., & Tang, J. (2022). *TorchDrug: A Powerful and Flexible Machine Learning Platform for Drug Discovery* (arXiv:2202.08320). arXiv. <https://doi.org/10.48550/arXiv.2202.08320>
- Zilian, D., & Sottriffer, C. A. (2013). SFCscoreRF: A Random Forest-Based Scoring Function for Improved Affinity Prediction of Protein–Ligand Complexes. *Journal of Chemical Information and Modeling*, 53(8), 1923–1933.
<https://doi.org/10.1021/ci400120b>
- Zou, H., & Hastie, T. (2005). Regularization and Variable Selection Via the Elastic Net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67(2), 301–320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>
- Zurenko, G. E., Yagi, B. H., Schaadt, R. D., Allison, J. W., Kilburn, J. O., Glickman, S. E., Hutchinson, D. K., Barbachyn, M. R., & Brickner, S. J. (1996). In vitro activities of U-100592 and U-100766, novel oxazolidinone antibacterial agents.

Antimicrobial Agents and Chemotherapy, 40(4), 839–845.

<https://doi.org/10.1128/AAC.40.4.839>