A MOLECULAR-DYNAMIC-SIMULATION-BASED DRUG DISCOVERY PLATFORM

FOR PATHOGENIC BACTERIA

A Dissertation

by

HAOQI WANG

Submitted to the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Sandun Fernando
Committee Members,	Arul Jayaraman
	Kranthi Mandadi
	Lisa Pérez
	Janie Moore
Head of Department,	John Tracy

December 2021

Major Subject: Biological and Agricultural Engineering

ABSTRACT

This dissertation presents an *in-silico* molecular-dynamic-simulation-based drug discovery tool and its utility to find inhibitors for pathogenic microorganisms. Specifically, the applicability of the tool has been demonstrated using two pathogenic bacteria, enterohemorrhagic *E. coli* subtype O104:H4 and *Candidatus* Liberibacter spp.

First, a pharmacophore modeling and refinement tool ELIXIR-A (Enhanced Ligand Exploration and Interaction Recognition Algorithm) was developed using Python programming language. The tool helps refine pharmacophores generated from multiple ligand-receptor interaction points using the Iterative Closest Point (ICP) variant algorithm.

ELIXIR-A identified six potential inhibitory compounds for the *E. coli* O104: H4 β-lactamase receptor protein. One non-β-lactam compound showed good inhibitory activity of *E. coli* O104: H4 on Kirby Bauer disk diffusion susceptibility testing. These results suggested that this novel non-β-lactam compound could be used as a lead compound to develop potent drugs targeting β-lactam-resistant Gram-negative bacterial strains.

Then the tool was applied to screen inhibitors targetting serine tyrosine phosphatase, a putative virulence protein of *Candidatus* Liberibacter spp. The *in silico* analysis followed by *in vitro* binding kinetic studies resulted in two small molecules (G6P3510 and G6P6373) that were further verified by *in planta* studies (performed elsewhere).

Finally, a computational modeling strategy was employed to identify antimicrobial peptides (AMPs) that could be used as potential bacterial inhibitors via blocking TolC, an essential protein in the efflux pump of the Type 1 Secretion System (T1SS) of *Candidatus* Liberibacter asiaticus (*C*Las). Multiple *in silico* approaches such as homology modeling, molecular docking, molecular dynamics simulations, Molecular Mechanics/Generalized Born

Surface Area (MM-GBSA) calculations, and Principal Component Analysis (PCA) were used to identify potential AMPs against the outer membrane protein TolC. The results (including *in vitro* studies with surrogate bacteria conducted elsewhere) suggested that the screened AMPs can be used as inhibitors targeting the TolC receptor.

DEDICATION

I would like to dedicate my work to my family, Dad and Mom, who have supported and encouraged me over the years to complete this dissertation.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a thesis (or) dissertation committee consisting of Dr. Sandun Fernando and Dr. Janie Moore of the Department of Biological and Agricultural Engineering, Dr. Arul Jayaraman of the Department of Chemical Engineering, Dr. Kranthi Mandadi of the Department of.Plant Pathology and Microbiology, and Dr. Lisa Pérez of the Texas A&M High Performance Research Computing.

All the work conducted for the dissertation was completed by the student independently.

Funding sources

ACKNOWLEDGEMENTS

First, I would like to thank my major professor, Dr. Sandun Fernando. This dissertation work would not have been completed without his guidance and support. His constant encouragement and advice helped me conquer and solve research challenges throughout the years at Texas A&M University. I hope I can continue to apply the lessons I learned from him to my future career.

Second, I want to thank my committee members Dr. Arul Jayaraman, Dr. Kranthi Mandadi, Dr. Lisa Pérez, Dr. Janie Moore, for their helpful discussions, encouragement, and supports.

Third, I truly appreciate Dr. R Karthikeyan for his encouragement support, who taught me research and teaching. I would like to thank Dr. Rani Menon for all her help.

Last, I want to thank my colleagues and friends for their encouragement and support all these years.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

GENERAL INTRODUCTION

This dissertation includes four distinct studies with the shared objective of developing computational tools to predict and screen active lead compounds for specific bacterial receptors and applying *in vitro* models to validate the results. This first chapter introduces the objectives of each study in the dissertation. The computer-aided drug design methods relevant to this dissertation, including pharmacophore modeling, molecular docking, molecular dynamics (MD) simulations, and validation results, are given in chapters 2 through 5.

Chapter 2 presents a novel method of pharmacophore refinement resulting from ligand-receptor binding from multiple targets based on the ICP variant algorithm. Chapter 3 and 4 present the application of the above approach for two bacterial enzyme systems, including β-lactamase and serine/tyrosine phosphatase, respectively. In Chapters 4 and 5, the computational methods are further applied to screen small molecule and antimicrobial peptide inhibitors targetting *C*Las.

CHAPTER II

ELIXIR-A: AN INTERACTIVE VISUALIZATION TOOL FOR MULTI-TARGET

PHARMACOPHORE REFINEMENT

(This Chapter is under review by the Journal of Computer-Aided Molecular Design)

Introduction

Drug discovery and development is a complex, time-consuming and expensive process. Computer-aided drug design approaches have the potential to accelerate this process in a cost-effective manner when compared to the laborious traditional compound screening methods. The existing computational techniques include quantitative structure-activity relationship (QSAR), molecular docking-based high-throughput, and pharmacophore-based virtual compound screening. In the usual virtual drug discovery process, molecular docking, pharmacophore models, and 3D QSAR models are often used in combination [3-5]. The existing QSAR or docking approaches do not have the capability of pharmacophore refinement, and there is a need for a technique for refining pharmacophores to identify the best set of pharmacophores for the inhibitor screening step.

The pharmacophore concept was introduced by Ehrlich in early 1909 [6]. A pharmacophore is an ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response [7]. Pharmacophores can describe specific ligand-receptor interactions as a generalized pattern. The first 3D pharmacophore screening software was developed by Gund in 1977 [8]. Pharmacophore modeling approaches can be broadly classified as ligand-based or receptor-based [9]. Software such as DISCO [10], GASP [11], GALAHAD [12], HipHop [13], HypoGen [14], MOE (Chemical Computing Group,

2

https://www.chemcomp.com) , PharmaGist [15], MolAlign [16], and PHASE [17] have been developed to construct ligand-based pharmacophore models. Their performance differs based on the efficiency of the algorithm to handle ligand flexibility and alignment and requires a set of pharmacologically active ligands [9]. On the contrary, receptor-based methods, such as GBPM [18], LigandScout [19], Pharmit [20-22], PyRod [23] and ZINCPharmer [24] analyze the receptor-ligand complex structures to isolate essential pharmacophoric features [25]. In situations where ligands are not known for the target receptor, methods such as CavityPlus [26], GRID [27], HS-Pharm [28], Pocket version 4.0 [21, 22], Shaper2 [29], and SuperStar [30] have been developed to identify hotspots (highly probable ligand binding sites) on the receptor. Druggability simulations (molecular dynamics simulations conducted in the presence of drug-like organic molecules) assess ligand hotspots while maintaining receptor flexibility. Tools like SILCS-Pharm from the Mackerell lab [31, 32] and Pharmmaker from the Bahar lab [33] have been developed to extract pharmacophore features from druggability simulations. However, there is a need for a systematic tool that analyzes and compares multiple pharmacophore models irrespective of their method of construction.

Here, we present ELIXIR-A, an open-source, user-friendly application that serves the purpose of both pharmacophore modeling as well as pharmacophore mapping. ELIXIR-A is a Python-based application that can import pharmacophore models created in visual molecular dynamics (VMD) [34], as well as manual coordinate input from any other platform. In addition, the output files from ELIXIR-A can be easily visualized in VMD [35] and can be exported to the pharmacophore-based virtual screening platforms like Pharmit [20].

Methodology
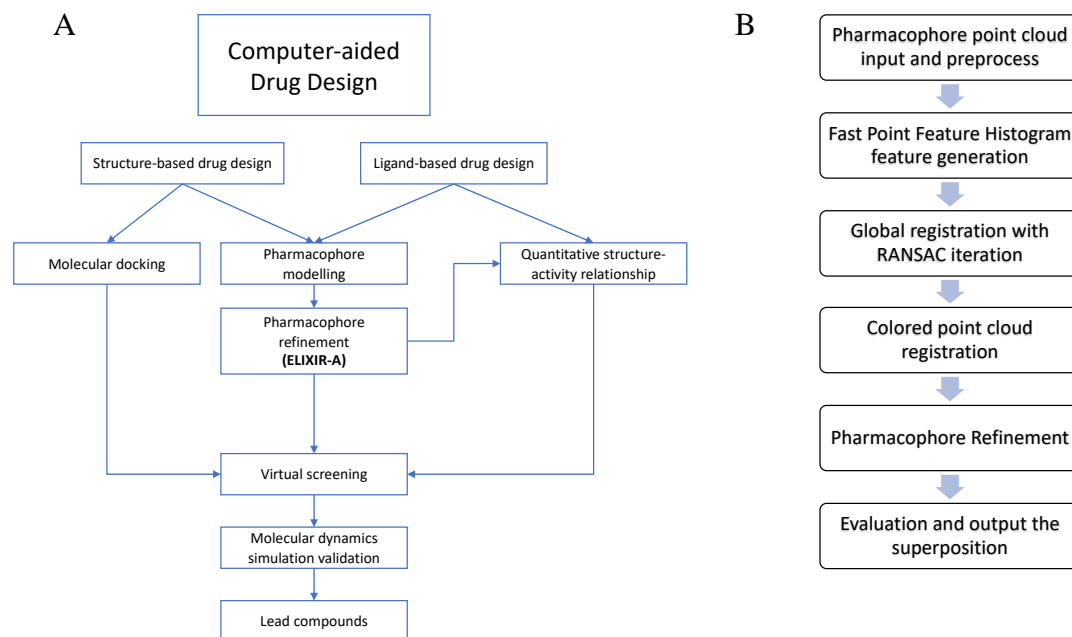
Figure 1 Representation of an example of (A) computer-aided drug design (CAAD) flowchart and (B) ELIXIR-A.

ELIXIR-A is developed to unify and simplify the interaction data from multiple pharmacophore models (Figure 1. left). This tool can accept two sets of pharmacophore models created directly by a pharmacophore generating platform. For example, the platform can accept ligand-base pharmacophore models created directly by Pharmit as the input. ELIXIR-A is developed on a modern 3D data processing library Open3D [36] to include the algorithms of fast global registration [37] and colored point cloud registration (colored ICP) [38, 39] (Figure 1. Right). ELIXIR-A initially prepares each pharmacophore point represented as a point cloud with its radius, and the pharmacophores are color-coded into different types. Then, ELIXIR-A aligns two structures and calculates the initial transformation matrix based on their geometric properties only. Finally, the two pharmacophore clouds are superpositioned with their geometric and

4

pharmacophore type information, and the overlaid pharmacophore point is refined with matched points.

Pharmacophore point clouds

As a shape similarity approach, ELIXIR-A takes pharmacophore points and applies each of them to a three-dimensional volume with a point cloud. Each pharmacophore type consists of 1000 uniformly distributed points in a sphere. The radius of the pharmacophore cloud is defined in the occupancy category of the Protein Data Bank (PDB) [40] file. The pharmacophore type is marked with three characters in the residue name category (S. Table 1). Throughout the complete analysis, the Biopandas framework is used to handle the point cloud data [41].

Global registration with RANSAC iteration

The first alignment is global alignment with features. Two pharmacophore point clouds would be treated with the same prepress to calculate a vector of descriptors called the Fast Point Feature Histogram (FPFH) [42]. FPFH is a 33-dimensional vector that describes the geometric characteristics and principles for a point. The FPFH vector can search the points with similar features [42]. As ELIXIR-A attempts to align pharmacophore sites from multiple ligands or receptors, there could be distinct differences between some areas of the clouds considered as "noise". It is known from experience that substantial noise levels can affect the alignment of point clouds with FPFH matching. To enhance the search of the FPFH matching algorithm, the random sample consensus (RANSAC) algorithm, proposed by Fischler and Bolles [43] is included in the registration process. RANSAC can estimate parameters of a mathematical model from a set of observed data with "noise". The global registration process calculates a preliminary rigid rotation and transformation matrix with their geometric characteristics. A fitness score will be given to evaluate this initial transformation.

Colored ICP

The second alignment is local ICP alignment with pharmacophore features. The pharmacophore alignment utilizing colored features is inspired by the alignment of the red, green, blue model (RGB) image and its corresponding depth image by Park. In general, the colored ICP algorithm extends the color vector as the extra dimension of the point coordinate data. The standard ICP algorithm will repeat the transformation of matrices to find the minimum square distance between two clouds. In comparison, the colored ICP algorithm will find the best fit of the extended paired matrices with pharmacophore "color" information for each point. ELIXIR-A uses a slightly different number of iterations, fitness values, and root means square deviation (RMSD) values from the default to optimize pharmacophore alignment. ELIXIR-A also supports the user adjusting these critical parameters to obtain a more appropriate fit for the transformation.

Pharmacophore Refinement

After applying preliminary global transformation and local colored ICP transformation from the point cloud superposition to the first pharmacophore input, two pharmacophore points are partially overlapped. It is necessary to remove the non-overlapped pharmacophore using the refinement algorithm. The algorithm first creates two $3 \times N$ matrices A and B for two groups of pharmacophore datasets. A is defined as the transforming source group, and B is defined as the fixed target. The Euclidean distance for each point in group A to find the corresponding point in group B is calculated. Then, the Euclidean distance for each point in group B to find the corresponding point in group A is evaluated. In each group, if there is no corresponding point within the threshold distance, this point will be considered irrelevant and removed from the group. According to the definition of the pharmacophore, it is possible for different types of

pharmacophores to exist in superposition; therefore, this refinement only calculates the geometric properties and does not consider the color parameters of the pharmacophore. The alignment of the two refined pharmacophore points will be visualized as Van der Waals surface (VdW) in VMD. And these files will be saved in the same path as the two input pharmacophore files.

$$Euclidean\ distance = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2 + (z_A - z_B)^2}$$

Evaluation function

The ELIXIR-A uses the fitness value to evaluate the effectiveness of the transformation for both alignment algorithms. Fitness is calculated using the formula below to find the volume ratio of overlap. The higher the fitness value, the better the alignment performance.

$$Fitness = \frac{pairs\ of\ corresponded\ points}{\#\ of\ points\ in\ the\ second\ pharmacophore\ point\ clouds}$$

Fitness Score function

The ELIXIR-A uses the fitness value to evaluate the effectiveness of the transformation for both alignment algorithms. Fitness is calculated using the formula below to find the volume ratio of overlap. The higher the fitness value, the better the alignment performance.

$$Fitness = \frac{pairs\ of\ corresponded\ points}{\#\ of\ points\ in\ the\ second\ pharmacophore\ point\ clouds}$$

Benchmark compound validation

To validate the efficiency of the pharmacophore refinement algorithm, a molecular dataset consisting of active inhibitors and inactive decoys targeting specific protein receptors was designed to reduce testing bias. The pharmacophore models were screened on the Pharmit platform and the Directory of Useful Decoys (DUD-e) dataset was used as the small molecule

library [44]. The benchmark receptor was human immunodeficiency virus type 1 protease (HIVPR), including 536 active inhibitors and 35750 inactive decoys. The feature "Shared pharmacophores" in LigandScout 4.4 demo was used as the comparison pharmacophore refinement algorithm in this study.

Results

Configuration

ELIXIR-A is designed as a graphical user interface (GUI) plugin based on VMD (Figure 2). Default parameter values are displayed in the GUI. These parameter values can be optimized according to the actual pharmacophores for better fit. The first step is to install the VMD and then edit the startup files in the VMD folder. The physical system requirements for ELXIR-A will not exceed the requirements of the VMD software itself. Once the setup files have been updated, the ELXIR-A folder must be placed in the VMD TCL plugin directory. The ELIXIR-A option can be found under the top menu. Python environment is required to run the script of ELIXIR-A. The package has been tested in cPython 3.8.8 with the following packages. (Numpy==1.20.1; open3d==0.13.0; biopandas==0.2.9.dev0; matplotlib==3.3.4) [36, 41, 45, 46].

Figure 2 ELIXIR-A user interface. Computations can be performed by providing default values for the fields and then allowing the user to modify the default information if necessary.

Execution of pharmacophores alignment jobs

ELIXIR-A provides two ways to introduce the pharmacophore information to the GUI (Figure 3). The user can choose to pre-edit the text file with ".pdb" extension via "Import pdb file:", or directly use the pharmit saved session file ".json" as same as the pre-edited pdb file. The guideline for the text files are described in S. Table 2.
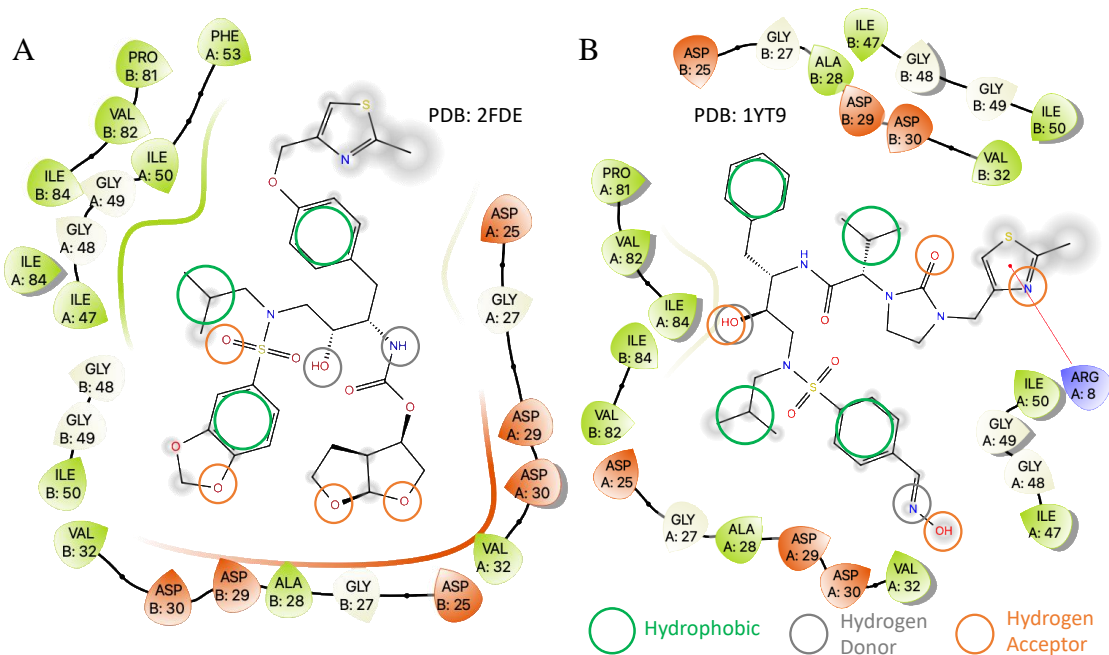
Figure 3 Pharmacological description of two HIVPR complexes (A and B) with inhibitors in crystal structures. These models were prepared in Pharmit and visualized in maestro [45].
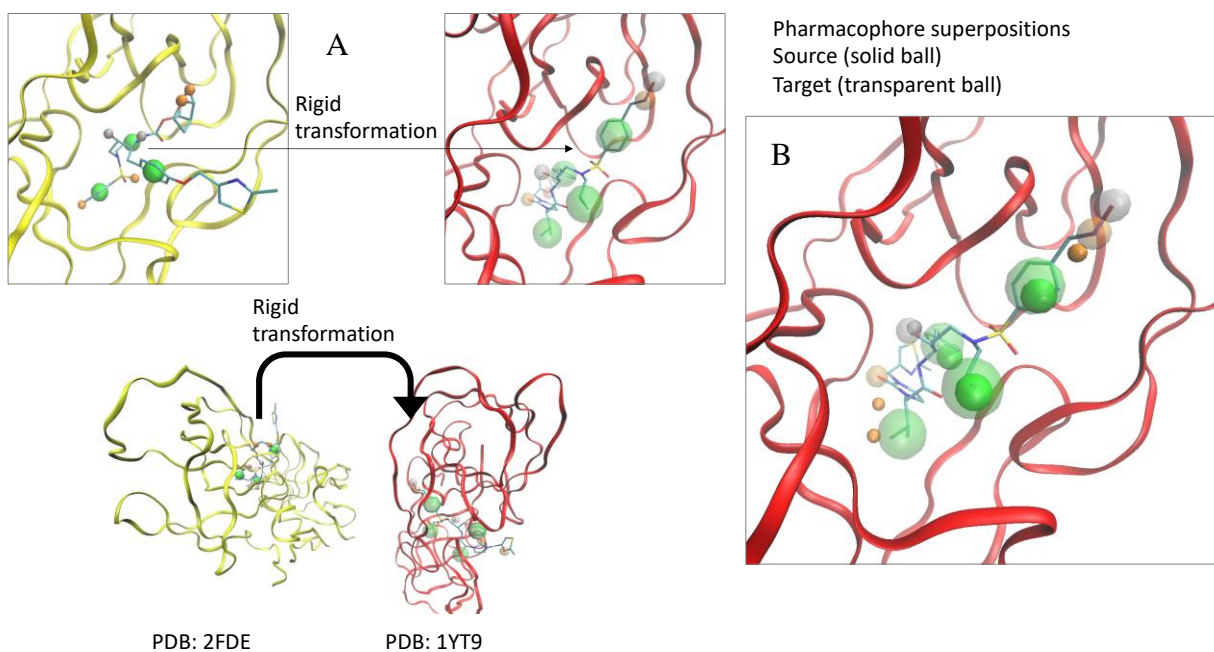
Figure 4 A. Structure-based pharmacophore refinement between HIVPR-inhibitor complexes utilizing ELIXIR-A. B. The transformed pharmacophores are presented with solid spheres. The fixed pharmacophores are presented using transparent spheres.

Visualization of results

The first pharmacophore cluster is recommended as the first primary site. And the second cluster would be interactively rotated and transformed to retrieve the conformation that fitted best with the primary site. Once all the data are prepared in the GUI, click the "submit" button to analyze the pharmacophore points' alignment. This demonstration used ELIXIR-A to find refined structure-based pharmacophores from the HIVPR (Figure 3). The receptor HIVPR is an essential enzyme of HIV replication. The active site of HIVPR was at the core of the dimerization interface (Figure 4). The pharmacophores of the ligand(s) were extracted from Pharmit Engine (Figure 3) [47]. In the VMD OpenGL Display window, two groups of VdW spheres are represented as pharmacophores after the ELXIR-A alignment. The transformed pharmacophores were presented with solid spheres. And the fixed pharmacophores were presented as transparent spheres. The ligand regarding the fixed pharmacophores was also shown in Figure 4 left. The initial transformation has 15.39% of point cloud overlap and the colored ICP has improved the overlay up to 65.66%. After refinement, two separate pharmacophore spheres in each group did not overlap and were therefore removed. A total of eight pharmacophores shared the superposition space, and this result can be used for further drug discovery. In addition to the HIVPR, the use of similarity refinement for the analysis of two other types of proteins was also applied in Table 1. The overlaid structures and fitness scores suggested the good similarity of pharmacophores between the active site of these proteins (S. Figure 2).

Table 1 Selection of ELIXIR-A aligned pharmacophore clusters.

PDB ID[a]	Protein Description	Ligand ID[b]	Initial points	Refined points	RMSD (Å)	Best fitness Score		Ref
						Global Registration	Colored ICP	
1YT9	HIV-1 Protease	OIS	11	8	0.267	15.39%	65.66%	[48]
2FDE		385	9	8				[49]
5I3P	Dengue Virus NS5 Polymerase	68T	9	7	0.219	57.40%	84.60%	[50]
5I3Q		68E	9	7				[50]
3GFW	Dual specificity protein kinase TTK	S22	5	4	0.255	38.10%	80.50%	[51]
6N6O		KE7	4	3				[52]
3PKJ	NAD-dependent deacetylase sirtuin-6	A2N	22	21	0.271	25.0%	74.4%	[53]
6QCN	NAD-dependent protein deacetylase sirtuin-2	AR6	22	22				[54]

a. The protein complexes are downloaded from https://www.rcsb.org.

b. These ligands are the local substrates corresponding to the protein complexes. Detailed information on the ligands is available at https://www.rcsb.org/ligand/(ligand ID).

Benchmark compound validation

The benchmark compound study was used to validate the effectiveness of the module in generating enriched screened molecules using refined pharmacophores. As a comparison, the same analysis of benchmark compound was done using LigandScout.

The ELIXIR-A refined pharmacophores were extracted from the overlaid points (Figure 4 right). Five features with three different types were built from twenty initial features. The HIVPR pharmacophore model prepared directly by Pharmit could not generate any screened results because the initial model contained too many pharmacophores – confirming the need for a

tool to refine pharmacophores leading to effective virtual screening. The original ELIXIR-A

model screened 19 molecules with 3 active molecules (Table 2). The ELIXIR-A enrichment was

higher than LigandScout, because of the extra hydrogen acceptor feature (Figure 5). If the

hydrogen acceptor feature was removed, ELIXIR-A would have similar performance to that of

LigandScout model with around 4 EF. This benchmark compound validation indicated that

ELIXIR-A could be a valuable tool to refine pharmacophores for better enrichment during

virtual screening.

Table 2 The enrichment analysis for two refined pharmacophore models.

	Refined (shared) pharmacophores			Virtual Screening results		
	Hydrogen Donor	Hydrogen Acceptor	Hydrophobic	Active	Decoy	EF
ELIXIR-A	1	1	3	3	16	**10.7**
ELIXIR-A (reduced)[a]	1	0	3	238	3491	**4.3**
LigandScout	1	0	3	253	3236	**4.9**

[a.] In the reduced ELIXIR-A model, the hydrogen acceptor feature was removed to have the same number of features comparing to the LigandScout model.
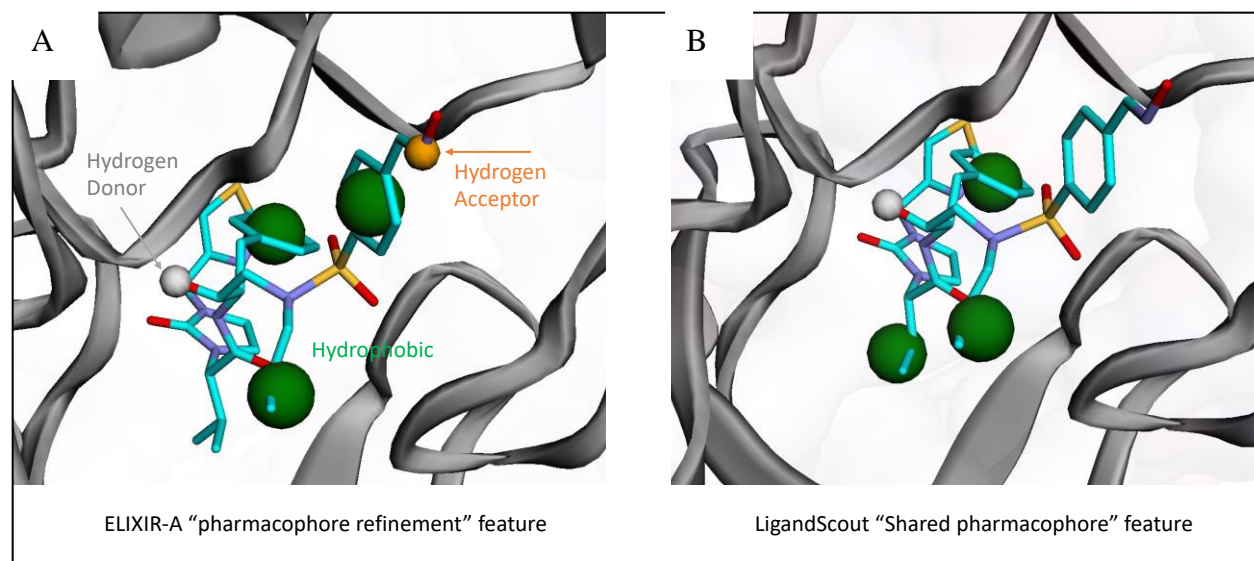
Figure 5 Refined pharmacophore models of HIVPR generated by ELIXIR-A (A) and LigandScout (B). The ELIXIR-A model had one hydrogen acceptor. The two models each used one hydrophobic group at a different position.

Discussion

Pharmacophores are a set of steric and electronic features that recognize the optimal supramolecular interactions. Structure-based manner with protein-ligand interactions and ligand-based manner with common chemical features from a set of active/inactive ligands are the two common approaches for building 3D pharmacophore models [55, 56]. ELIXIR-A is not a typical pharmacophore modeling software; however, it is a pharmacophore refinement algorithm that uses sets of pharmacophores as input and aligns these pharmacophores in 3D space to identify any overlap. ELIXIR-A uses a computer vision-inspired ICP variant algorithm to align multiple pharmacophore models with similar geometric and physicochemical properties as point clouds into a refined point cloud. Although researchers have applied this ICP algorithm for drug discovery to match three-dimensional protein structures [57] as well as the alignment of protein binding cavities [58, 59], it has not been used for pharmacophore refinement. ELIXIR-A fills this need.

14

ELIXIR-A uses a geometry-based approach to find the pharmacophore similarity between two protein-ligand binding pockets before calculating the binding energy between the receptor and small ligand molecules using colored ICP. ICP requires a good initial transformation to ensure that the point cloud converges to a minimum acceptable value [60]. Also, the presence of outliers (non-uniform points) can affect the alignment in ICP. ELIXIR-A utilized FPFH matching and RANSAC iterations to solve the global fitting and outlier problems. Another algorithm that can perform similar calculations is the Kabsch algorithm. Kabsch algorithm is widely used in bioinformatics and can calculate the RMSD between two 3D protein or pharmacophore structures via rotation and translation [61, 62]. When the point correspondences are known, Kabsch can be applied. Since point correspondences are not known, the Kabsch algorithm cannot be applied here and thus ICP was used. When using colored ICP, the whole cloud will be paired to figure out the corresponding point, and the points in one group can be mapped with multiple corresponding points in the other group. It is also possible to find some unpaired important pharmacophore points when comparing the output results. Another algorithm that is commonly used for ligand-based pharmacophore generation is LAMDA (linear assignment for molecular dataset alignment) that finds globally optimal pairing between objects [56]. LAMDA was not considered for ELIXIR-A due to its limitation for ligand-based pharmacophore generation. Genetic algorithms (GA) used to calculate the initial transformation position before ICP transformation [63-66]. For example, GA is used for pharmacophore generation via molecular docking-based data set such as Autodock vina [67]. While ELIXIR-A used FPFH matching with RANSAC iteration for global alignment. Compared with GA, RANSAC iteration requires less parameters for the initial model. And the RANSAC iterations

are included in the same library as the colored ICP. Therefore, in this case, the RANSAC iteration is more applicable to ELIXIR-A than GA.

For the initial pharmacophore analysis, ELIXIR-A and LigandScout used the same method to find the rigid superposition of two pharmacophore clusters, while ELIXIR-A used the ICP variant algorithm and LigandScout used the Hungarian algorithm. In general, the time complexity of the ICP algorithm in ELIXIR-A was $O(N^2)$ and the time complexity of the Hungarian algorithm ranged from $O(N^4)$ to $O(N^3)$. ELIXIR-A provided better time complexity with the potential to efficiently solve large pharmacophore alignment problems.

Conclusion

ELIXIR-A is a python-based VMD plugin to help refine pharmacophore models in situations where many pharmacophores are present from multiple models due to various ligand-receptor interactions and many conformations a ligand may pose on a receptor binding site. ELIXIR-A GUI can help refine the pharmacophores generated from multiple ligands from multiple receptors using the ICP variant algorithm. The output from ELIXIR-A can be used in the virtual pharmacophore-based platforms for compound screening.

Requirements and availability

The ELIXIR-A package was developed in python3 for the data operations and tcl/tk for the user interface. The package has been tested on the Ubuntu 20.04, macOS Big Sur, and Windows 10 systems. The python versions can be 3.7.4 or later. The latest version of ELXIR-A is available for download on GitHub. (https://github.com/sfernando-BAEN/ELIXIR-A/releases). VMD could be downloaded from their official websites. A 64-bit version of VMD is recommended for this package. ELIXIR-A follows the Apache-2.0 license and is open-source on GitHub.

Availability of data

The datasets generated during and/or analysed during the current study are available in

the GitHub repository, https://github.com/sfernando-BAEN/ELIXIR-A

CHAPTER III

A NON-BETA-LACTAM ANTIBIOTIC INHIBITOR FOR ENTEROHEMORRHAGIC

ESCHERICHIA COLI O104:H4

This chapter is published in the *Journal of Molecular Medicine*

Introduction

β-lactam antibiotics are a class of broad-spectrum antibiotics for treating bacterial

infections [68]. β-lactam antibiotics have been successfully used to treat staphylococcal and non-

enterococcal streptococcal infections [69, 70]; however, overuse of antibiotics has resulted in an

increasing prevalence of drug-resistant bacteria. The most widely used broad-spectrum class of

antibiotics that contain a β-lactam ring in their molecular structures includes

penicillin derivatives such as methicillin, cephalosporins, monobactams, and carbapenems [71]

and bacteria often develop resistance to β-lactam antibiotics by synthesizing β-lactamases,

enzymes that attack the β-lactam ring. β-lactam antibiotics work by inhibiting cell

wall biosynthesis in the organism [72].

E. coli O104: H4 has multi-drug resistance to most known β-lactam antibiotics and other

antibiotics [73, 74]. β-lactamases from *E. coli* O104: H4 plays an essential role in the hydrolysis

of antibiotics. Antibiotics like benzylpenicillin inhibit bacterial growth by binding on to

penicillin-binding protein (PBP), DD-transpeptidase, thus inhibiting its cross-linking activity and

preventing it from forming a cell wall for the new cells [75]. β-lactamases destroy the β-lactam

ring of penicillin and make the antibiotic ineffective. From Shiga-toxin-producing *E. coli* O104:

H4, a β-lactam antibiotic, penicillin G first gets trapped with extended-spectrum β-lactamase

protein followed by the β-lactam ring on penicillin G getting tightly bound to the enzyme. Then

hydrolysis of the enzyme opens the β-lactam ring on penicillin G (Figure 6) [75]. To retain the

18

efficacy of penicillin, β-lactamases need to be blocked from binding on to the antibiotic. This can be achieved by introducing another molecule in addition to penicillin that can bind to β-lactamases with greater affinity as compared to penicillin and thus making β-lactamases ineffective at destroying penicillin.
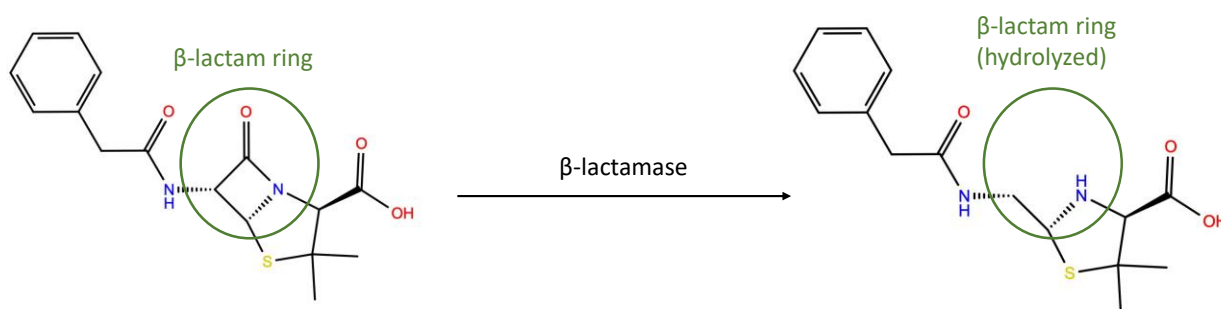
Figure 6 β-lactam antibiotic penicillin G is hydrolyzed by β-lactamase.

Numerous techniques have been used for drug discovery for β-lactamase inhibitors, including *in silico* and high throughput experimental screening[76]. Although experimentation is a critical element in discovering inhibitor drugs, experimental techniques alone do not provide any direct information on the nature and quantitative details of binding phenomena like the computational techniques. It has been established in recent studies [77, 78] that MD is an effective method to explore the relationship between equilibrium dynamics and structural changes of proteins involved in the function. This is because MD can handle multiple ligands in a simulated natural environment while protein strands are dynamically moving [79, 80]. MD simulations enable real-time visualization of ligand-receptor interactions [81] that is critical in deciphering interaction phenomena for drug development. In this work, we present an *in vitro* experimental methodology based on the computational design for identifying inhibitors targeting β-lactamases of a highly drug-resistant *E. coli* O104: H4 strain.

Materials and methods

Screening studies were conducted by first identifying the primary and any allosteric active sites of β-lactamase TEM-1 using a combination of docking studies via AutoDock Vina [82] and Nanoscale Molecular Dynamics (NAMD). Once the binding region of penicillin on β-lactamases was identified, pharmacophores were deciphered via the DruGUI plugin[34] in combination with Enhanced Ligand Exploration and Interaction Recognition Algorithm (ELIXIR-A). Potential inhibitors were screened using ZincPharmer[20, 24] and Pharmit online platforms, which were verified *in silico* via Autodock Vina and NAMD before experimental verification.

The homology models of the two β-lactamase (TEM-1 and CTX-M-15) used in this study were built using gene sequences from GenBank (CP003289.1 and CP003297.1) [83]. The 3D homology model templates of β-lactamase were searched via the SWISS-MODEL server [84]. The template from PDB [40] (PDB id:1zg4 [85]) served as the model for β-lactamase TEM-1, and PDB id: 5t66 chain A was the model for β-lactamase CTX-M-15 [86]. The final homology models were isolated based on the scoring functions available in the SWISS-MODEL server. PDB ID: 1hvb [87] was used as the Penicillin-Binding protein (PBP).

Identification of primary and allosteric active sites:

The first step in screening novel inhibitory ligands is identifying the primary penicillin-binding pocket(s) of β-lactamase. Initial approximations were made via docking analysis (Autodock Vina) followed by in-depth analyses via NAMD simulations using penicillin G and clavulanic acid as the probes.

For identification of sites initially, the crystal structures of proteins were obtained from the Protein Data Bank [40], and any ligands and ions were removed/balanced using VMD. The

receptor proteins, i.e., PBP and β-lactamases, were uploaded to Autodock Tools (ADT), and hydrogen atoms were added. The 3D-optimize, penicillin, or clavulanic acid were uploaded to Autodock as the ligand. The root detection, identification, and selection of the active number of torsions were made to the ligand as appropriate. The search space grid for the macromolecule was set initially to the entire molecule (and, during subsequent iterations, were confined to identified pockets to obtain the cluster of docking confirmations that are most stable). Then, docking simulations were done via AutoDock Vina. Subsequent to running Autodock, the ligand conformations, ranked by energy (largest negative values having the highest binding affinity), were stored as PDB files. The locations of the predicted binding pockets were verified using NAMD simulations using penicillin G and clavulanic acid as ligands and by overlaying ligand-bound crystal structures with the docking output(s).

The location where a significant number of ligand conformations overlapped in the Autodock Vina output and where many binding hotspots aggregated in the NAMD output was selected as the primary site. This information was further verified by comparing it with ligand-bound crystal structured deposited in PDB. Any of the other sites that had a high affinity to penicillin G and clavulanic acid were assumed as allosteric sites.

Identification of binding features (i.e., pharmacophores)

To accurately identify binding features in an environment that closely resembles what the enzyme would encounter in the real world, a simulation system that allows the capability of sending multiple types and quantities of ligands to the dynamically changing receptor was used. To emulate cytosol, the protein was immersed in a water sphere in the presence of different concentrations of select ligand(s). Each system was minimized and simulated to discern the hotspots where ligand(s) preferentially bound while the protein was still undergoing structural

21

motions (i.e., the "jiggling and wiggling" motions). Once the hotspots were identified, the features of the binding site(s) and locations (in 3D space) were recorded as pharmacophore points. This information was used during the inhibitor screening step. The steps involved in the NAMD simulations are presented below:

Creation of Protein Structure Files and force-field parameter files for proteins

To perform MD simulations using NAMD, four distinct files are needed: A Protein Data Bank (PDB or. PDB) file that stores atomic coordinates and/or velocities for the system; a Protein Structure File (PSF or .psf) that stores structural information of the protein, such as various types of bonding interactions; a force field parameter file that contains the mathematical expressions of the potentials experienced by the atoms in the system that defines bond strengths, equilibrium lengths, and other variables; and a configuration file in which the user specifies all the options that NAMD should use in running a simulation. The configuration file tells NAMD how the simulation is to be run.

Select concentrations of isopropanol, isobutene, acetamide, acetate, isopropylamine, penicillin G, and clavulanic acid were used as the ligand(s). The original PDB files of proteins were used to build the PSF files. The procedure used to build the PSF files is given elsewhere [88]. The PSF file is developed based on the information in the PDB file while taking into consideration any chains that are missing from the original crystal structure. The final PSF file is built after patches are applied. The force field files of the ligands must be developed to populate the ligands in the vicinity of the receptor protein and simulate their interactions. Chemistry at HARvard Macromolecular Mechanics (CHARMM) force fields for larger proteins and the CHARMM General Force Fields (CGenFF) for smaller ligands were used for the simulations. The procedure used to develop force field files is given elsewhere [89, 90].

Initial system setup: Once the structure (.psf) and coordinate (.pdb) files are available, the protein was solvated by building a water box over a water box. The *solvate* plugin available in VMD was used for this purpose. Note that a Tcl script also could be written for this. Any counter ions will be added using the *autoionize* plugin. Ligands will be added using the DruGUI plugin. Next, the configuration (.conf) files required to run the NAMD simulation will be prepared. Figure 7 shows the output solvation box after the system has been set up to run the NAMD simulations for a model protein.
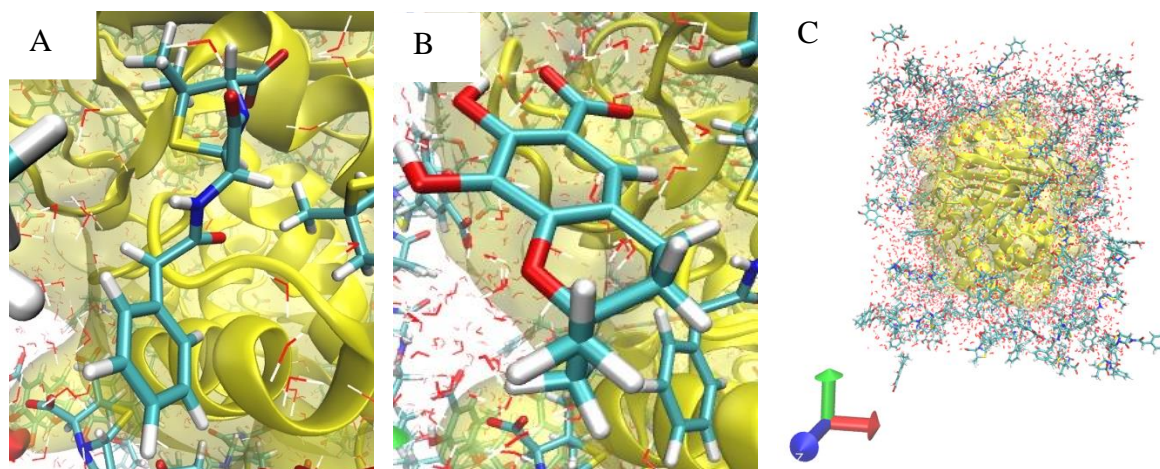
Figure 7 (A) Probes penicillin G (green), (B) screened ligand (green), and solvation box (C) after preparing for NAMD simulations. The solvation box includes probe molecules (like penicillin G in A. and screened ligands in B), water molecules (red "dots") and beta-lactamase (yellow). The system was neutralized with counter sodium chloride ions.

NAMD simulations: NAMD simulations were performed in the High-Performance Computing Center (HPCC) at Texas A&M University. The simulations were run in the Ada cluster, an Intel x86-64 Linux cluster with 852 compute nodes (17,340 total cores) and eight login nodes.

Site Analyses

Probe grid calculation: The coordinate file, structure file, and coordinate trajectory file were used to calculate the probe grid. The DruGUI plugin in VMD allows setting up grid

resolution (Å); contact distance (Å); hotspot free energy, dG, (kcal/mol); the number of hotspots to the cluster; lowest affinity (μm); and charge (e)[34]. Grids were calculated for different types of probes using their central carbon atoms. These grids were merged in the hotspot analysis described below. Custom probes were added to the simulation by modifying appropriate files in DruGUI. The output files subsequent to the probe grid calculation were .dx files that store occupancy grids of each probe/ligand.

Protein surface analysis for binding hotspots and hotspot clusters: The binding hotspots based on occupancy grids were calculated using the DruGUI plugin in VMD. This step involved selecting high-affinity probe binding spots, clustering them, and then merging them to assess high-affinity sites.

Pharmacophore screening

For the identification of proper pharmacophores of the β-lactamase enzyme, it is critical to identify those hotspots that partake in binding phenomena at the primary site. This is because a simulation outputs many hotspots scattered around the protein that do not participate in the binding of the primary substrate at the active site. The selection of active pharmacophores is made by overlapping the hotspots identified from MD simulations with docking results and crystallographic data and isolating hotspots that belong to primary and allosteric sites. The coordinates of interaction points in 3D space and the nature of the interaction(s) (i.e., hydrogen accepting, hydrogen-donating, aromatic, and hydrophobic) were recorded for each of the binding sites.

For MD and docking simulations, whenever unsupported metal atoms were involved, respective atoms were replaced with those with close structural similarity.

In silico screening and verification of ligand druggability

Pharmacophore screening was done via ELIXIR-A, which is under development in our laboratory. The algorithm consists of a computer-guided routine that recognizes pharmacophore points, i.e., the ensemble of steric, electrostatic, and hydrophobic properties, which is essential for optimal supramolecular interactions with the receptor to inhibit its biological effect. Then potential ligands were screened via Zincpharmer using: a combination of the location of the functional groups (e.g., proton donor/acceptor, hydrophobic groups, positive/negative ion, and exclusion spheres); stabilization of the most effective conformation; Lipinski's rule of five that defines properties necessary for good permeation (i.e., the molecule having less than five proton donors, the molecular weight is less than 500 Da, log P smaller than 5, the molecule having <10 acceptors, and the molecule using biological transporters so that the ligand does not attach too strong); and having at least three minimum pharmacophore points [91].

To further improve the accuracy of the hits, enzyme inhibitor scoring criteria was employed via Molinspiration's Fragment-based Virtual Screening Engine (http://www.molinspiration.com). The enzyme inhibitor score of an unknown drug-like ligand was given by the comparison of the known active drug ligands to the unknown ligand with sophisticated Bayesian statistics. The peak of the score with the best activity is 0.5. Octanol-water partition coefficient (LogP), number of hydrogen bond acceptors/donors, number of rotate bonds were predicted with Molinspiration as well.

In silico verification of the potential ligands was done via AutoDock Vina, NAMD, and ZDOCK by comparing the binding affinities, binding interactions, and binding specificity to the primary site of the ligands with penicillin and clavulanic acid controls.

Antimicrobial susceptibility testing

Escherichia coli serotype O104: H4 (ATCC® BAA-2326™) from VWR was used for validating molecular dynamics predictions. Amoxicillin and clavulanic acid were used as a β-lactam antibiotic and β-lactamase inhibitor controls, respectively, and kanamycin was used as a positive non-β-lactam antibiotic control. All the screened compounds and antibiotics were purchased from Molport and Sigma-Aldrich.

Antimicrobial susceptibility testing was performed via the minimum inhibitory concentration (MIC) method. Disk diffusion methodology and broth dilution method were performed to determine the MIC. Kirby Bauer disk diffusion susceptibility testing methodology as defined by the standard protocol from The European Committee on Antimicrobial Susceptibility Testing (EUCAST) was used[92]. The MIC was determined by broth microdilution assay with LB broth [93]. The plate was incubated at 37˚C with a 180 rpm reciprocating shaker for 16 hours. The MIC was defined as the lowest concentration of the wells where the broth was not turbid, and there were no sediments at the bottom of the wells. The ligand compounds were solved in either sterile DI water or DMSO.

Results and discussion

Identification of active site and pharmacophores.

Using the combination of docking and MD simulations with model probes and known drug molecules a clear primary site was revealed with multiple binding hot spots (Figure 8) along with some allosteric sites. It should be noted that aforementioned methods alone may not be sufficient to identify all allosteric sites. For example, newly developed methods such as Hamiltonian replica-exchange-based method[94] and site-directed mutagenesis & enhanced sampling computational methods[95] had to be used to reveal hidden allosteric sites of multi type β-lactamases. However, in this case, since the primary site was found to be druggable, allosteric phenomena was not pursued.

The binding hotspot distribution for small molecular probes is given in (Figure 8A) while those for penicillin G and the clavulanic acid mixture is presented in Figure 8B. The images below the hotspot distribution give the affinity map of distribution of each of the probes ranging from red, white to blue depicting, high, medium and low affinities respectively.

The binding free energies coupled with the occupancy grid distribution will help identify: 1) the existence of any hotspots where these probes tend to linger for longer times, i.e., specific pharmacophore points (i.e., hydrogen accepting, hydrogen-donating, aromatic, and hydrophobic); 2) if any hotspot clusters where several of the hotspots tend to conglomerate, suggesting the existence of an active (ligandable) site; and 3) the location of the pharmacophores in 3D space that assists in inhibitor screening .

(A1)

Binding free energy: (kcal/mol)
-1.00
-1.36
-1.71
-2.07
-2.42
-2.78

(A2)

(B1)

3Clav:-2.94
1Peni:-3.01
2Peni: -2.96

Binding free energy (kcal/mol)
-1.01
-1.41
-1.81
-2.21
-2.61
-3.01
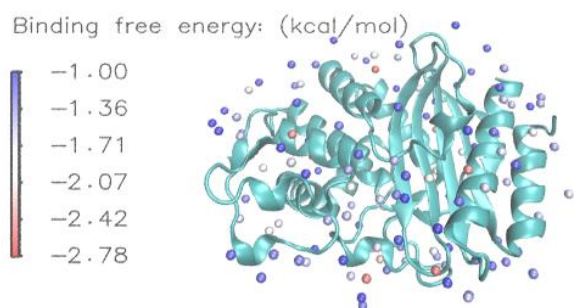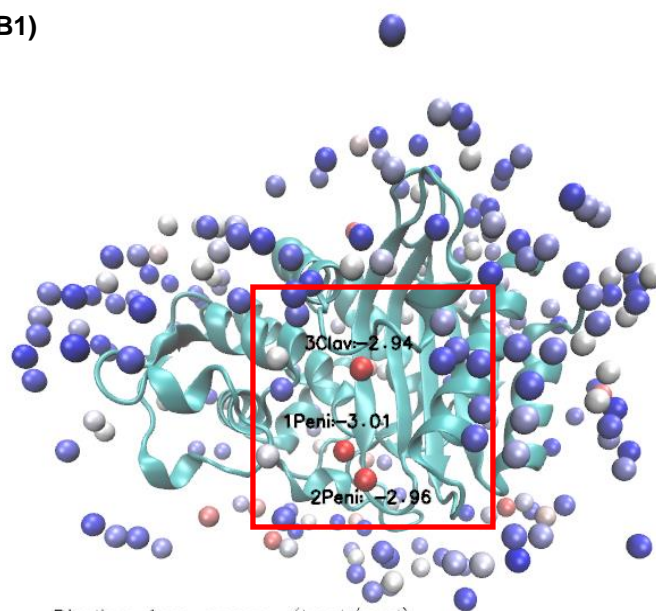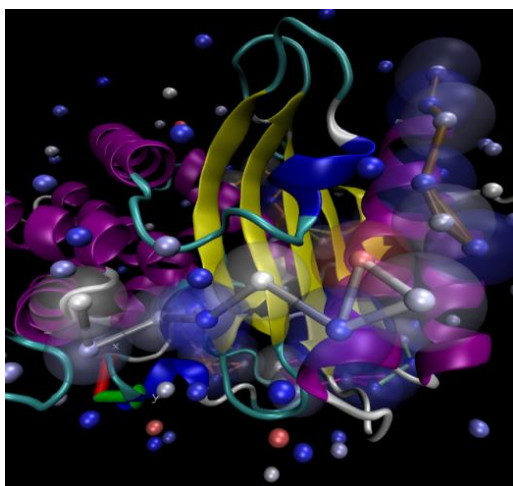
NAMD simulations with penicillin and clavulanic acid clearly delineates the active site with red-spots depicting binding hotspots with highest affinity. The area demarcated by red box depicts the primary binding site.

(A3)

isopropanol
acetate
isopropylamine
acetamide
isobutene

(B2)

penicillin-G
clavulanic acid
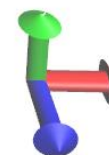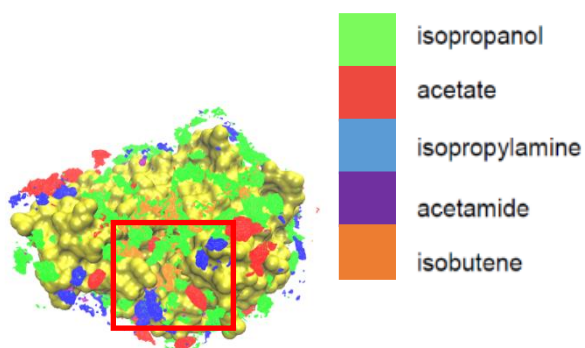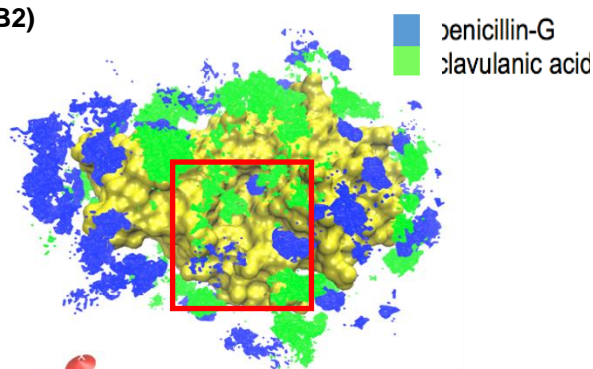
Figure 8 A1**)** High binding energy hotspots distribution with the molecular probes; A2) pharmacophore cluster at the active site; A3) Probe occupancy grid and binding free energy distribution; B1**)** High binding energy hotspots distribution with penicillin G and clavulanic acid along with the binding free energy distribution. B2**)** Pharmacophore occupancy grid distribution of penicillin G and clavulanic acid ligands on β-lactamase protein's active domain.

The drug-like probe distribution around the active site and the pharmacophore occupancy grid distribution is presented in Figure 8A1 and Figure 8A3 respectively. Hotspots 19, 35 and 119 situated at the active site. Hotspot 19 (binding free energy: -1.81 kcal/mol) had probabilities of 65.5% IBUT, 28.8% IPRO and 5.7% ACAM binding onto the hotspot respectively. On hotspot 35 (binding energy: -1.60 kcal/mol), 99.7% IPAM and 0.3% ACET were bound. On hotspot 119 (binding energy: -1.04 kcal/mol), 74.5% IPRO, 25.2% IBUT and 0.4% ACET were bound. These probe compositions were used to determine the properties of the functional group(s) at the active residue(s) on the receptor.

It was interesting that when penicillin G and clavulanic acid were used, individual hotspots occupied only one type of the molecules at 100% probability. It is likely that the size of these two probes restricted another molecule finding the active site once bound. Hotspots 1, 2 and 3 were all located on the active site with binding free energies of -3.01 kcal/mol Penicillin G, -2.96 kcal/mol Penicillin G and -2.94 kcal/mol Clavulanic acid respectively.

Isolation of Pharmacophores

Isolation of pharmacophores for compound screening was done based on hotspot analysis via model probes, penicillin, and clavulanic acid along with properties of the residues that interacted with these ligands. Pharmacophores include the location coordinates of potential functional groups (e.g., proton donor/acceptor, hydrophobic groups, positive/negative ion, and exclusion spheres) of the target compound(s). Hydrogen bonding information between ligands and specific residues are a key part of pharmacophore analysis. Figure 9 and Table 3 depict hydrogen bonding information deciphered from docking analysis between penicillin G and clavulanic acid. Penicillin G and clavulanic acid were bound to active sites with binding affinities of -7.5 and -6.3 kcal/mol, respectively via hydrogen bonding. The region with the

highest binding energy from these simulations was found to be in close proximity to that identified via laboratory experiments *[86]*.

The analysis with most stable conformations suggests that residues SER 70 and ARG244 form hydrogen bonds with β-lactamase TEM-1 with penicillin while SER 70, SER 130, ASN 132, SER 235 forms hydrogen bonds with clavulanic acid. Crystal structure showed that nucleophilic Ser 70 and Ser 130 bound to the β-lactam rings [1]. Experimental verification by others confirm these findings [96]. However, it should also be noted that in other β-lactamases SER130 can be shifted to SER132 based on mutations. Residues SER 130 and TYR105 on β-lactamase TEM-1 were attracted to penicillin while LYS234, GLY236, and ALA237 were attracted to clavulanic acid via van der Waal's forces. It has been reported that the surrounding residues Glu166 and LYS73 may catalyze the β-lactam ring hydrolysis process[1]. These interactions were used to screen pharmacophores for ligand screening.

Table 3 Interactions between Clavulanic Acid or Penicillin G to the residues on the active site of β-lactamase TEM-1

LIGAND NAME	RECEPTOR AMINO ACID NUMBER	TYPE OF INTERACTIONS	MAXIMUM BINDING AFFINITY (KCAL/MOL)
PENICILLIN G	SER 70, ARG244	Hydrogen bond	-6.73 ± 0.23
	SER130, TYR105	Van der Waals force	
CLAVULANIC ACID	SER 70, SER 130, ASN 132, SER 235	Hydrogen bond	-5.97 ± 0.06
	LYS234, GLY236, ALA237	Van der Waals force	
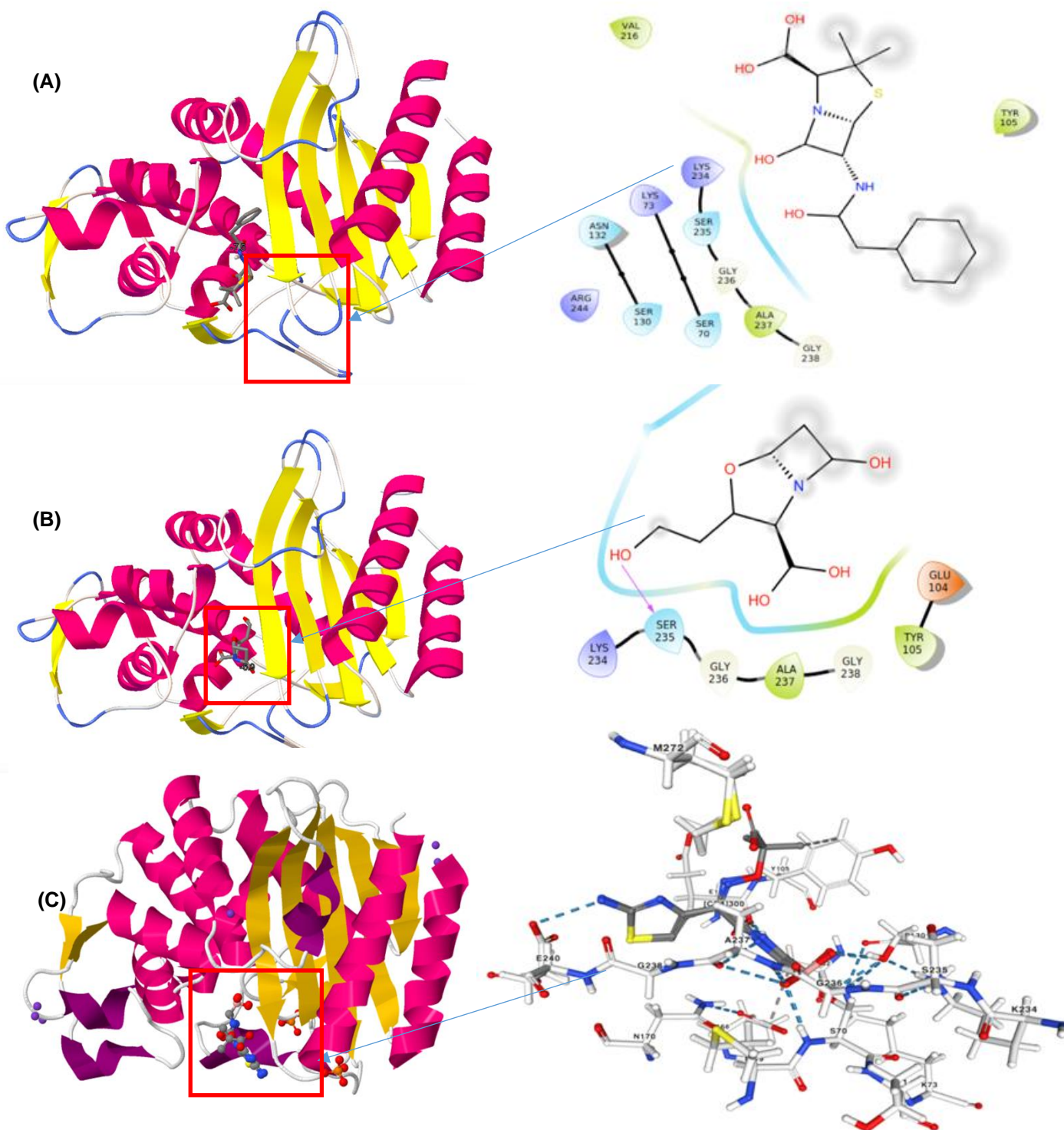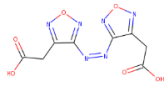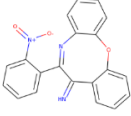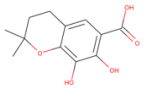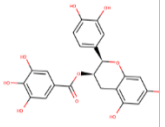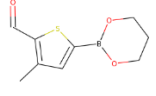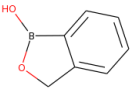
Figure 9 A) Interaction diagram of penicillin on 1ZG4; B) Interaction diagram of clavulanic acid on 1ZG4; C) Crystal structure of β-lactamase TEM-1 bound with an active ligand (pdb 1M40) showing that the simulations were able to accurately predict the active site [1]. Note: Active site interaction plots were made with Maestro (Schrödinger, 2018-2). The plot of interaction was made with NGL[2] and the protein was made with Jmol (http://www.jmol.org).

Table 4 Structural and functional information on potential inhibitor ligands.

Ligand Id	1	2	3	4	5	6
Id	ZINC13777605	ZINC75280048	ZINC333270	ZINC3978503	BAPE	BACM
Ligand skeletal formula						
IUPAC Name	2-[4-[[4-(carboxymethyl)-1,2,5-oxadiazol-3-yl]diazenyl]-1,2,5-oxadiazol-3-yl]acetic acid	11-(2-nitrophenyl)benzo[b][1,6]benzoxazocin-12-imine	7,8-dihydroxy-2,2-dimethyl-3,4-dihydrochromene-6-carboxylate	[(3R)-2-(3,4-dihydroxyphenyl)-5,7-dihydroxy-3,4-dihydro-2H-chromen-3-yl] 3,4,5-trihydroxybenzoate	5-(1,3,2-dioxaborinan-2-yl)-3-methylthiophene-2-carbaldehyde (henceforth called as BAPE)	1-hydroxy-3H-2,1-benzoxaborole (henceforth called as BACM)
Octanol-water partition coefficient logP	-1.94	3.52	2.70	2.54	1.17	1.04
molecular weight	280.16	343.34	238.24	442.38	170	133.94
number of hydrogen bond acceptors	12	6	5	10	3	2
number of hydrogen bond donors	0	1	3	7	2	1
"Rule of five" violations	1	0	0	1	0	0
Number of Rotatable Bonds	6	2	1	4	2	0
Enzyme inhibitor score[a]	-0.22	-0.16	0.21	0.25	0.73	1.92

Average docking affinity kcal/mol	-5.31±0.54	-7.32±0.07	-6.41±0.10	-7.52±0.06	-5.20±0.08	-5.42±0.03
Highest free binding energy ΔG in MD, kcal/mol	-3.07	-2.80	-2.84	-2.59	-2.18	-2.45

a. The enzyme inhibitor score >0.25 is good and >0.50 is excellent.

Ligand screening and *in silico* verification

Then potential ligands were screened via Zincpharmer [24] using: a combination of the location of the functional groups; stabilization of the most effective conformation; Lipinski's rule of five that defines properties necessary for good permeation [91] (i.e., the molecule having less than five proton donors , the molecular weight is less than 500 Da, log P smaller than 5, the molecule having <10 acceptors, and the molecule using biological transporters so that the ligand does not attach too strong); and having at least three minimum pharmacophoric points.

Initial screening resulted in several dozen potential compounds. However, based on above criteria, six potential inhibitor molecules were isolated. The criteria used and information on the ligands are shown in Table 4. The two boronic containing compounds have good inhibitor scores on enzymes.

In silico **verification**

Before moving to experimental verification, the potential efficacy of the molecules was verified *in silico* via NAMD/DruGUI, Autodock Vina, and ZDock. Figure 10 depicts simulation results when β-lactamase TEM-1 was impinged with the six isolated ligands mixed with 50% (molar concentrations) of penicillin. Results indicate that the free binding energy ΔG varied from -2.18 to -3.07 kcal/mol under simulated conditions. The value for penicillin was varied from -2.76 to -3.14 kcal/mol indicating that there is a high possibility for competitive inhibition. In all

instances, the potential candidates bound at the vicinity of the active site. Nevertheless, it was clear that penicillin preferentially bound to the active site. The binding domain coincided with experimental results using other inhibitors [97]. It should also be noted that no allosteric sites were evident. In order to further confirm how the compounds interacted with the beta-lactamase, docking simulations were performed with penicillin once each of the inhibitors was bound to the active site. Figure 11 depicts binding maps using Autodock Vina. It is noted that when BACM or clavulanic acid is bound on the active site of β-lactamase TEM-1, the penicillin molecules docked around the active site with much lower incidence (Figure 11). This suggested that BACM or clavulanic acid were the potent inhibitors of β-lactamase.

Figure 10 A-F: Pharmacophore point distribution of β-lactamase TEM-1 (PDB id: 1zg4) with the
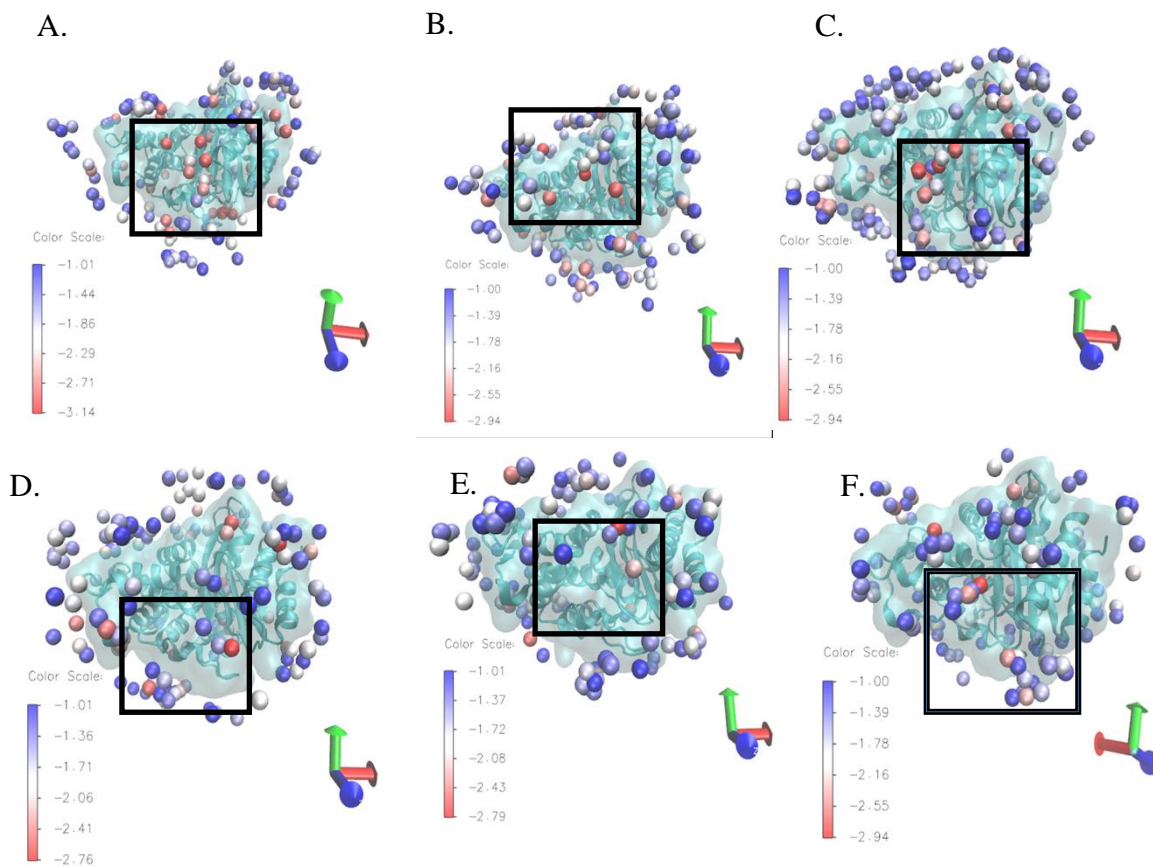
probes ZINC13777605, ZINC75280048, ZINC337270, ZINC3978503, BAPE, and BACM together with Penicillin G, respectively. The free binding energy ΔG (kcal/mol) is illustrated with

34

the red or blue colored van der Waals force-represented balls. The active site is highlighted on each of the complexes with a black rectangle.

Antimicrobial susceptibility testing via experimental verification

In vitro tests were performed to investigate the antimicrobial susceptibility of screened ligands with β-lactam antibiotics. The inhibition of *E. coli* O104: H4 was measured using disk diffusion methodology on the particular β-lactamase inhibitor - β-lactam combination (Table 5A). From the EUCAST clinical zone diameter breakpoint tables, the susceptibility breakpoint for amoxicillin (20 µg) and clavulanic acid (10 µg) is 19 mm [92]. In our studies, the strain was susceptible to the same amoxicillin and clavulanic acid combination with a diameter of 22.3 ± 4.7 (mean +/- std. dev. n=4) mm which was in line with EUCAST protocol. The negative control amoxicillin (20 µg) did not inhibit *E. coli* O104:H4 indicating its resistance to the antibiotic. For the combination of BACM (10 µg) and amoxicillin (20 µg), the diameter was 18.8 ± 0.5 mm (mean +/- std. dev. n=3) indicating that the strain was susceptible to the non-β-lactam compound. None of the other compounds tested when mixed with amoxicillin showed any potency. The non-β-lactam antibiotic kanamycin (10 and 30 µg) inhibited the strain with 16.0 ± 1.0 (mean +/- std. dev. n=3) and 26.3 ± 1.2 (mean +/- std. dev. n=3) mm respectively.

MIC testing also showed good antimicrobial activity of BACM on *E. coli* O104:H4 (Table 5B). Amoxicillin and BACM showed MIC of 4 mg/L and 2 mg/L respectively while BACM alone showed a MIC of 2 mg/L. This is an interesting finding since the experiments proved that BACM alone can inhibit the bacterial strain. The amoxicillin control did not show any inhibitory action as observed earlier. Amoxicillin and clavulanic acid resulted in a MIC similar to that of BACM. It was also observed that clavulanic acid alone was able to inhibit the strain at the same concentration as BACM likely because both molecules have an affinity toward PBP which was confirmed via docking analysis.

Table 5 A. Disk diffusion values of select β-lactamase inhibitor - β-lactam combination

β-lactam/ non β-lactam	Inhibitor	Disk content (Antibiotic or Antibiotic – Inhibitor) (µg)	Zone Diameter (mm) (mean +/- std. dev.)	Resisted (R) / Susceptible (S)
Amoxicillin	--	20	0.0±0.0	R
Kanamycin[a]	--	30	26.3±1.2	S
Kanamycin	--	10	16.0±1.0	S
Amoxicillin	Clavulanic acid	20-10	22.3±4.7	S
Amoxicillin	ZINC13777605	20-10	0.0±0.0	R
Amoxicillin	ZINC75280048	20-10	0.0±0.0	R
Amoxicillin	ZINC337270	20-10	0.0±0.0	R
Amoxicillin	ZINC3978503	20-10	0.0±0.0	R
Amoxicillin	BAPE	20-10	0.0±0.0	R
Amoxicillin	BACM	20-10	18.8±0.5	S

a. kanamycin is not a β-lactam antibiotic. It was used as the non β-lactam positive control in this testing.

Table 5 B. Disk diffusion values of select β-lactamase inhibitor and β-lactam combination

Strain	MIC (mg/L)					
E. coli O104:H4	Amoxicillin	Amoxicillin and Clavulanic acid (2:1)	Amoxicillin and BACM (2:1)	Clavulanic acid	BACM
	R[b] (128)	S (2)[d]	S (4)	S (2)	S (2)

b. R means resistant, and S means susceptible.
c. The minimum inhibitory concentration of this compound was 1 mg/L of Amoxicillin and 0.5 mg/L of Clavulanic acid. The ratio of antibiotic to inhibitor was 2:1.

In order to get some insights on what occurs at the active site when these compound combinations were introduced to the active site, docking simulations were performed on β-lactamases and penicillin-binding protein (PBP). The simulation revealed that BACM showed similar binding interactions to amoxicillin and clavulanic acid on PBP as well as two types of β-

37

lactamases, TEM-1 and CTX-M-15, which were found in the strain. Experimental studies suggested that amoxicillin-clavulanic acid combination was as potent as the amoxicillin-BACM combination. This observation is verified by docking simulations which suggests clavulanic acid binding to the β-lactamases TEM-1 and CTX-M-15 with affinities of 5.8 and 6.0 kcal/mol as compared to 5.9 and 5.8 kcal/mol with BACM to the active site as depicted in Figure **11** A and B. Also clavulanic acid formed hydrogen bonding with SER 235 on TEM-1, SER73 and THR238 on CTX-M-15 and ASN161 on PBP while BACM form hydrogen bonding with ASN170 on TEM-1, ASN135 on CTX-M-15 and ASN161 and THR310 on PBP. It was interesting that amoxicillin, Clavulanic acid, and BACM bound to the PBP of this strain with affinities of 8.7 kcal/mol, 7.2 kcal/mol and 6.2 kcal/mol respectively. This explains why clavulanic acid and BACM alone were able to inhibit this strain. However, it should be noted that BACM is a non-β-lactam compound which thwarts any attempts of the β-lactamase to hydrolyze the β-lactam ring (i.e., clavulanic acid being a suicidal inhibitor allowing its own β-lactam ring to hydrolyze instead of that of the antibiotic) and thus is a superior candidate to clavulanic acid (by not depending on ring hydrolysis that could potentially lead to resistance).

It should be noted that the transporter of antibiotics and screened molecules from the outer membrane into the bacteria plays a key role in the potency of the drugs. In previous studies, the *E. coli* was found to transport ions, acids and small molecules via the diffusion of outer membrane porins [98-100]. Several OMC/OMF porins such as OmpC20, OmpC26, OmpC28, and OmpC33 in multi-β-lactam –resistant *E. coli* strain were found to influence the permeability of polar antibiotics [101]. Since the size of the molecules were low, porins could affect the potency of these inhibitors. Of the two cyclic boronates that progressed based on *in silico* screening, only BACM was able to inhibit this particular strain. In the previous study, it

38

showed that the molecules with amine, amphiphilic, rigidity, and low globularity properties were

likely to accumulate in the Gram-negative bacteria[102]. Accordingly, the rigidity and low

globularity of BACM may have improved the accumulation of the molecules in the cell

contributing to its potency.
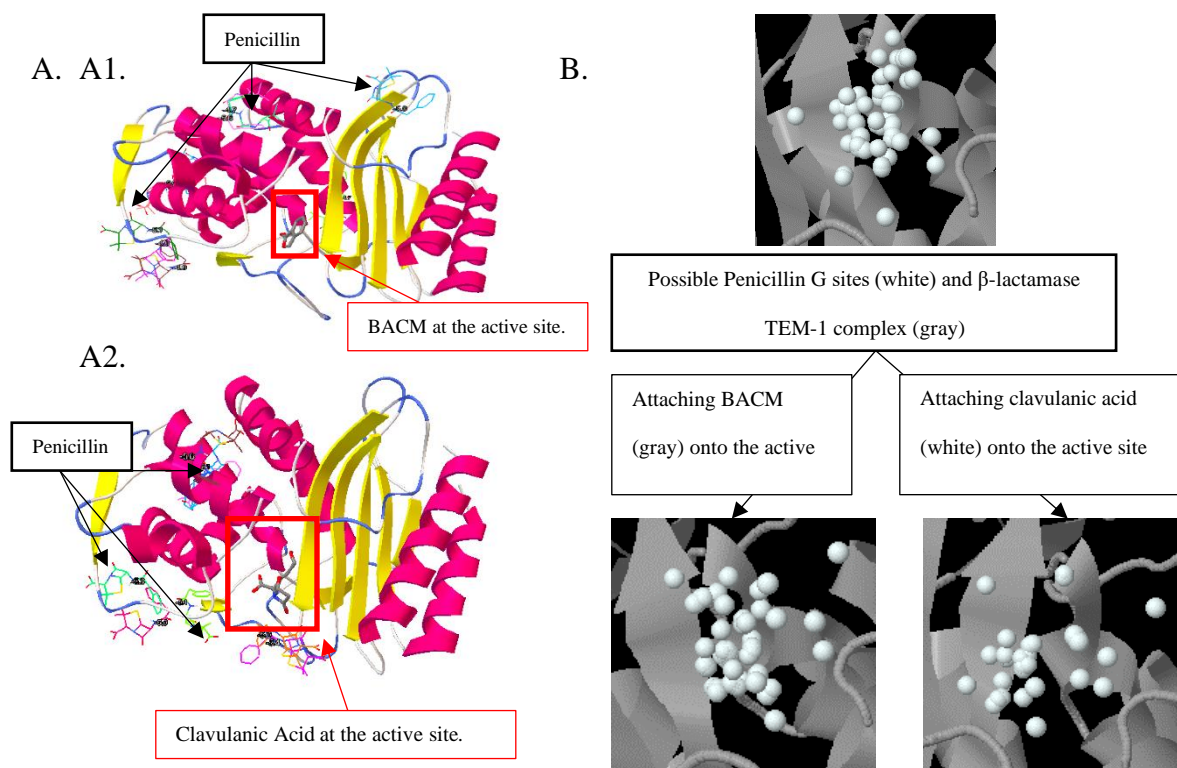
Figure 11 A) BACM (A1.) and Clavulanic Acid (A2.)- β-lactamase TEM-1 complex with
penicillin G molecules docking around. BACM and Clavulanic acid molecule were in the red
box. Penicillin G molecules were represented in colorful sticks. All the penicillin G molecules
were not able to bind to the active site because of the inhibitor. B) The ZDOCK simulation of
penicillin G- β-lactamase TEM-1 complex with and without inhibitors in 500 hundred attempts.
The binding confirmations of penicillin G reduced when inhibitors were bound on the active site.

Cyclic boronates have been successfully tested as antibiotics for several bacterial strains

[86, 103-105]. The first cyclic boronate drug, KERYDIN (tavaborole)[106] was approved by

FDA to treat fungal infection onychomycosis. Vaborbactam [107] was found to have an

inhibitory effect on multi-β-lactam-resistant strains. The potency of vaborbactam was found to

be impacted by the outer membrane porins OmpK35 and OmpK36, which controlled the influx

39

and efflux of small molecules into the starin. VNRX-5133[108] was an injectable cylic boronate beta-lactamase inhibitor on the phase 1 clinical stage developed by VenatoRx Pharmaceuticals, Inc. Cefepime/VNRX-5133 combination had high activity on serine-active site β-lactamases (Ser-BL) and emerging VIM/NDM metallo-β-lactamases (MBL). These works show the high efficacy of cyclic boronates on inhibiting the β-lactam-resistant strains. Nevertheless, this class of antibiotics has not been successfully tested on the antibiotic-resistant strain O104: H4.
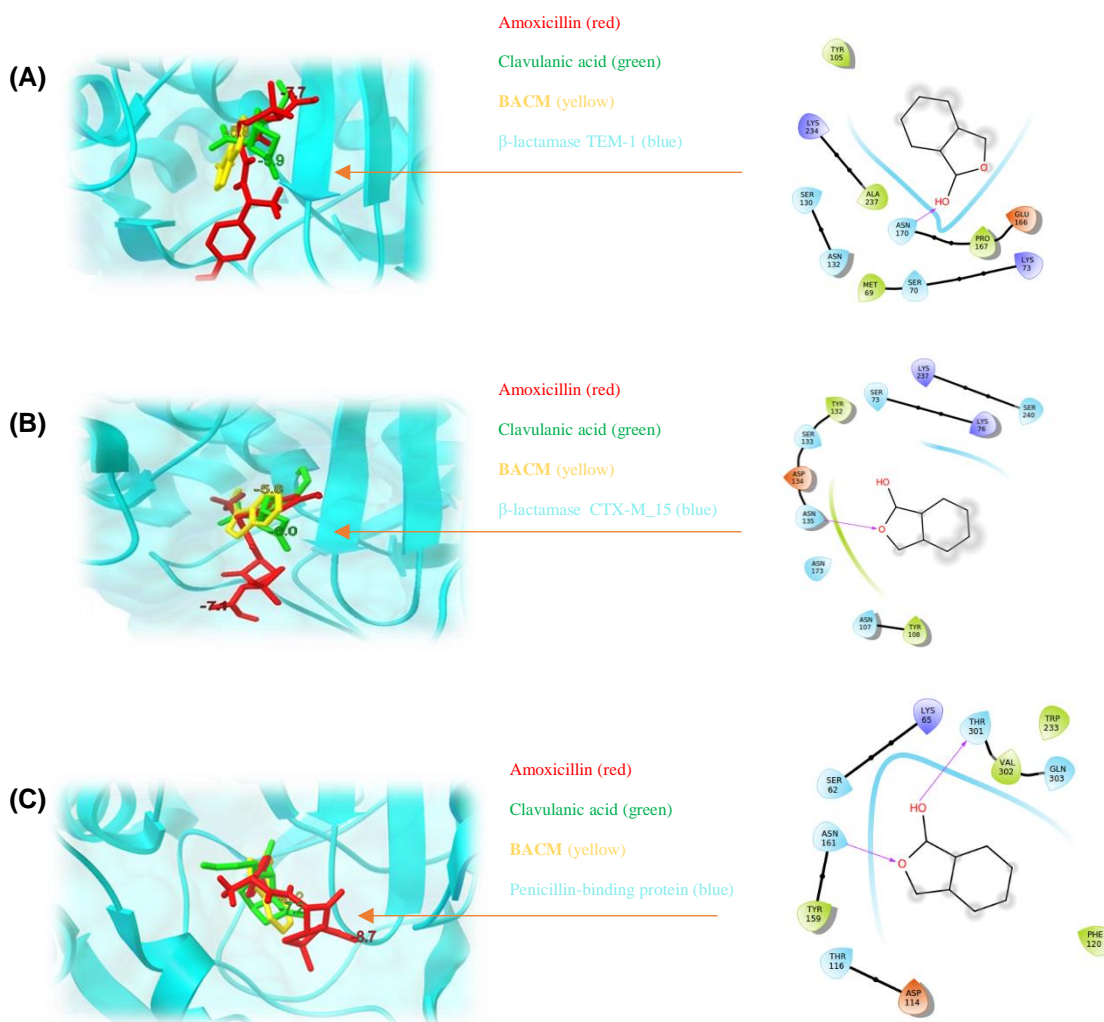
Figure 12 A) The conformations of ligands in complex with β-lactamase TEM-1 (pdb: 1zg4), amoxicillin, clavulanic acid, and BACM – ligands bound with affinities of 7.7 kcal/mol, 5.8 kcal/mol and 5.9 kcal/mol respectively on the active site; B) The conformations of ligands in complex with β-lactamase CTX-15 (pdb: 5t66, chain A) -amoxicillin, clavulanic acid and BACM bound with affinities of 7.1 kcal/mol, 6.0 kcal/mol and 5.8 kcal/mol respectively on the active site; C) The conformations of ligands in complex with PBP (pdb: 1hvb) - amoxicillin, clavulanic acid and BACM have binding affinities of 8.7 kcal/mol, 7.2 kcal/mol and 6.2 kcal/mol on the active site, respectively. Hydrogen bonding interactions were found between the compounds and ASN residues.

Conclusion

The primary active site of *E. coli* O104: H4 β-lactamase was able to be accurately identified using computational techniques. NAMD combined docking simulations with probes IPRO, IPAM, IBUT, ACAM and ACET, penicillin (substrate), and clavulanic acid (positive control) were able to identify interaction points on the receptor as well as on the ligands which in turn helped isolate pharmacophores. Using these pharmacophores, six potential inhibitory compounds, ZINC13777605, ZINC75280048, ZINC337270, ZINC3978503, BAPE, and BACM, were initially screened and verified *in silico* via MD simulations combined docking simulations, Lipinski's rule of five, and bioactivity scores. Experimental verification studies based on Kirby Bauer disk diffusion susceptibility testing indicated that *E. coli* O104: H4 was susceptible to amoxicillin and clavulanic acid (control) combination while the negative control of amoxicillin was not indicating any resistance to the antibiotic. The non-β-lactam compound BACM and amoxicillin combination was able to inhibit the strain. None of the other compounds tested when combined with amoxicillin showed any potency. Non-B-Lactam antibiotic kanamycin that was used as a positive control was also able to inhibit the *E. coli* strain. The amoxicillin and BACM combination showed a MIC of 2 mg/L to and 1 mg/L while BACM alone showed MIC of 1 mg/L indicating that BACM alone can inhibit the bacterial strain (without a β-lactam antibiotic). The amoxicillin control did not show any inhibitory action. Amoxicillin and clavulanic acid resulted in a MIC similar to that of BACM. It was also observed that clavulanic acid alone was able to inhibit the strain at the same concentration as BACM. BACM and clavulanic acid may be inhibiting the strain not only via β-lactamase inhibition but also via direct inhibition of PBP which plays an integral role in cell wall biosynthesis.

Data analysis and statistics

All *in-vitro* experiments were conducted as a completely randomized design (CRD) with three replicates. Data analysis was done using Microsoft Excel software and R-studio. The standard deviation was calculated by the STDEV() function in Microsoft Excel. Data between different ligands were tested by two-sample unpaired Student's t-test. T test was given by the t.test() function in R-studio. The significance level was $\alpha = 0.05$.

CHAPTER IV

TARGET-BASED MOLECULAR MODELING AND LIGAND SCREENING TO IDENTIFY

INHIBITORS FOR *CANDIDATUS* LIBERIBACTER SPP.

Introduction

Several devastating plant diseases are caused by fastidious (unculturable) *'Candidatus* Liberibacter spp.' For instance, potato zebra chip disease, caused by *'Candidatus* Liberibacter solanacearum' (*C*Lso), is causing millions of dollars in losses to the potato industry[109]. Similarly, citrus greening or Huanglongbing (HLB) disease, caused by *'Candidatus* Liberibacter asiaticus' (*C*Las) is the most devastating disease of citrus today and threatens citrus production worldwide. HLB is transmitted by an insect vector, the Asian citrus psyllid (*Diaphorina citri*). Since its inception in 2006, in Florida alone, HLB caused ≥50% reduction in citrus acreage and economic losses upwards of US $8 billion [110, 111].

Multiple antibiotic-based therapies are currently being evaluated by the industry to control *'Candidatus* Liberibacter spp.' Between the 1970's and 1980's, tetracycline, penicillin and other antibiotics were continuously tested for HLB suppression in citrus [112-114]. Trunk applications of penicillin in symptomatic sweet orange trees provided some tolerance and tree recovery [112]. Recently, in a drastic measure to control HLB, the Florida citrus industry has obtained a section 18 emergency exemption from the Environmental Protection Agency (EPA) to test the efficacy of commercial antibiotic/bactericide products. Some products (e.g., FirewallTM and FireLineTM) have already been tested in Florida over the last two years and approval has been granted for a third year [115]. California is following suit with an emergency registration of its own for streptomycin and oxytetracycline. Although promising, the antibiotics/bactericide therapies have variable results in the studies, are also being used in animal and human

medicine[116], and their extensive deployment in agriculture can pose threats for the emergence of antibiotic-resistance[117]. Thus, there is a need to explore alternative therapies for disease management.

Target-based molecular modeling and ligand screening are widely employed in biomedical and pharmaceutical research to discover new drugs and inhibitors[118]. Few studies have employed this approach in plant-microbial interactions and limited to studies of plant proteins[119, 120]. Here, we employed target-based molecular modeling and *in silico* ligand screening approach to identify two competitive inhibitors of a conserved *'Candidatus* Liberibacter spp.' virulence protein, serine/tyrosine phosphatase (STP)[121].

Materials and methods

Homology modeling, pharmacophore identification and ligand screening

The inhibitor screening was performed in several steps beginning with the development of the STP homology model, identification of active sites and pharmacophores, followed by ligand/inhibitor screening. Simulation platforms including Autodock Vina[122], VMD[123], Drugui[124], Enhanced Ligand Exploration and Interaction Recognition Algorithm (ELIXIR-A)[125, 126], NAMD [127] were used during *in silico* analysis. ELIXIR-A, the algorithm described in Chapter II for pharmacophore refinement, is also deposited in GitHub [https://github.com/sfernando-BAEN/ELIXIR-A]. NAMD simulations were performed in the Texas A&M High Performance Research Computing. To perform MD simulations using NAMD, four distinct files were used: A Protein Data Bank (PDB or.pdb) file that stores atomic coordinates and/or velocities for the system; A Protein Structure File (PSF or .psf) that stores structural information of the protein, such as various types of bonding interactions; A force field parameter file that contains the mathematical expressions of the potentials experienced by the

45

atoms in the system; and a configuration file in which the user specifies all the options that NAMD should use in running a simulation. These files were created by substrate ligands as well as any probe molecules (small molecule alcohols, amines, amides, and esters) that were used to interrogate the enzymes' active sites. Small probes were selected so they can diffuse rapidly and explore small and even transient pockets in the simulations. This property also helps to sample many binding events while enabling reaching equilibrium in reasonable timeframes. The binding hotspots (pharmacophores) were isolated by initially selecting high-affinity probe binding spots, clustering them, and then merging them to assess high-affinity sites [124]. All the interaction diagrams were developed using Schrodinger Maestro software. ELIXIR-A [125, 126] and the zincpharmer algorithm was used to screen and match the identified pharmacophores with potential inhibitors[128]. The top hits were selected based on the following criteria: i) a combination of location of the functional groups (e.g., proton donor/acceptor, hydrophobic groups, positive/negative ion, and exclusion spheres); ii) stabilization of the most effective conformation; iii) Lipinski's rule of five that defines properties necessary for good permeation[129] (i.e., the molecule having less than five proton-donators, the molecular weight is less than 500 Daltons, log P smaller than 5, the molecule having < 10 acceptors, and the molecule using biological transporters so that the ligand does not attach too strong); iv) Molinspiration drug-likeness & bioactivity score (www.molinspiration.com), and v) having at least three pharmacophoric points.

Molecular dynamics simulations

A 100 ns MD simulation was performed using the Schrödinger-Desmond platform in order to analyze the conformational and interaction stability of STP proteins in complex with screened ligands[130]. The *C*Las STP protein was prepared by Maestro's Protein Preparation

46

Wizard[131, 132]. Both C and N termini were capped to reduce the flexibility of the termini. For the screened ligands, Schrödinger's Glide extra-precision mode was used to reproduce the docking poses from the previous Vina docking studies [47]. The ligand-protein complexes were minimized under the OPLS3e force field [133]. The system was solvated using a simple point-charge (SPC) solvent model with orthogonal boundary conditions maintaining a buffer distance of 5 Å. The system charge was neutralized using chloride or sodium ions. Each system was minimized under the OPLS3e force field using Desmond's default relaxation protocol. The system was then simulated under isothermal–isobaric conditions at 300 K and 1.01325 bar pressure. A thousand frames were recorded at every 100 ps following the initial frame. The Simulation Interaction Diagram (SID) module was used to assess the quality of the entire simulation over time, including protein-ligand interactions, complex RMSD, and complex root mean square fluctuations (RMSF).

Bio-layer interferometry (BLI) and binding kinetics assays

Bio-layer interferometry (BLI) experiments were conducted using the BLItz system with Anti-Penta-HIS (HIS1K) Biosensors (ForteBio Corp., Menlo Park, CA). BLI is an optical detection technique that can determine the affinities of protein-ligand interactions on the sensor by the reflection of a light beam[134]. STP protein with the hexahistidine (His6) tags was attached to the fiber optic sensor tip with a selective tag immobilization method. The sensor tip was dipped into the small-molecule solution medium with a known concentration to load the small molecule ligands onto the sensor. The association of small molecules to the sensor represented by the association constant, ka, causes the white light to shift in the interference pattern. The unbinding phenomenon, which is represented by the dissociation constant, kd, will cause the white light to shift in the opposite direction on the sensor. The dissociation/association

47

ratio defines the affinity constant for the binding interactions (KD), and is represented by the equation as follows:

$$K_D = \frac{[A] \cdot [B]}{[AB]} = \frac{k_d}{k_a}$$

Where KD represents the affinity constants (M, or mol·L-1), kd represents the dissociation rate (s-1), ka represents the association rate (M·s-1). Here, [A], [B] and [AB] represent the concentrations (M) of the unbound STP protein, small molecule ligands, and protein-ligand complex, respectively. The affinity constants will reveal the overall binding kinetics of the protein-ligand complex, which helps to evaluate the performance of the screened molecules.

The binding kinetics of two STP proteins, from '*Candidatus* Liberibacter asiaticus' (*C*Las) and '*Candidatus* Liberibacter solanacearum' (*C*Lso), were validated with screened molecules G6P6373 and G6P3510. Proteins were stored in a storage buffer 50 mM Tris-HCl, 500 mM NaCl, 10% Glycerol, two mM DTT, pH 8.0. The HIS1K sensors were hydrated in a 250 μL assay buffer for 30 minutes. The assay buffer was prepared with 9.95% 10X Pall Forte Bio kinetics buffer (ForteBio Corp., Menlo Park, CA) and 89.55% PBS (Thermo Fisher Scientific, Waltham, MA) and 0.5% Dimethyl sulfoxide (DMSO) (Sigma-Aldrich, St. Louis, MO). The BLI test used a cycle of 30s assay buffer baseline, 300s STP protein loading, 120s baseline, 120s analyte association, and 120s analyte dissociation. The assay buffer was used in the two baseline phases and the last dissociation phase. The concentration of protein and analytes were optimized for significant and measurable binding signals. For the *C*Las STP protein assay, 4 μL of *C*Las STP protein with 0.34 mg/ml concentration was loaded, and the testing analyte 1 μM glucose-6-phosphate (G6P), 50 μM G6P6373 or 50 μM G6P3510 was associated in a 250 μL assay buffer, respectively. For the *C*Lso STP protein assay, 4 μL of *C*Las STP protein with 0.30 mg/ml

concentration was loaded, and the testing analyte G6P, G6P6373 or G6P3510 was associated

with 1 μM in a 250 μL assay buffer, respectively. All the BLItz experiments were conducted at

room temperature. The Kinetics energies were analyzed in BLItz Pro 1.3.0.5 software

(Fortiebio), and the interaction curves were exported from BLItz Pro and replotted with R studio.

Statistical Analysis:

All *in silico* experiments were conducted as a CRD with three replicates.

<div align="center">Results and discussion</div>

**Molecular dynamic simulation revealed key binding hotspots (pharmacophores) of *C*Lso
and *C*Las STPs**

To identify binding hotspots (pharmacophores), we first generated homology models of

*C*Lso and *C*Las STPs, using experimentally determined structures of a protein template in PDB

[40]. The homology model was built based on the template protein (PDB: 4RKK[135]) using

SWISS-MODEL via the HHblits protein sequence search package while enabling the highest

sequence coverage and identity [136]. This selected template was a phosphatase with a

carbohydrate-binding domain and a dual-specificity phosphatase domain. Small molecules of

G6P were used as substrates to represent the phosphate ions and α-D-glucose native substrates.

Both *C*Lso and *C*Las STP structures overlapped on residue 40 to 191 with 0.72 coverage rate and

0.29 sequence identity rate (Figure 13a). VMD [123] was used to depict the structural differences

between the homology model *C*Lso STP and *C*Las STP (Figure 13). Sequence alignment showed

a homology value (QH) of 0.9563, RMSD of 0.6389 Å, and percent identity of 62.5 which

suggested that the two proteins had a high degree of structural similarity, in line with the

multiple sequence alignment results. The two homology models had good overlap on various

substrate-binding domains with the template. Next, we identified hotspots using the *C*Lso STP

<div align="center">49</div>

protein model. Identification of primary and allosteric (if present) substrate-binding site(s) of

STP were initially approximated via Autodock Vina [122] followed by in-depth analyses via

NAMD [127] simulations and Drugui[124]. The goal was to accurately identify binding features

in an environment that closely resembles what the enzyme would encounter *in vivo*. NAMD

combined DruGUI were used to identify the binding features by sending multiple types and

quantities of select ligands to the dynamically changing receptor. To emulate the cytosol, the

protein was immersed in a water sphere in the presence of different concentrations of the

ligand(s). Each system was minimized and simulated to discern the hotspots where ligand(s)

preferentially bound while the protein was still undergoing structural motions. Once the hotspots

were identified, the binding features and locations (in 3D space) were recorded as

pharmacophore points. From this, we identified at least sixty-seven potential binding hotspots

(Figure 13b), with binding energies varying between -1.01 to -2.62 kcal/mol. The acetate drug-

like probe had the highest binding affinity of -2.62 kcal/mol, and the site where acetate was

bound was considered as an active site. The interaction points that participate in substrate

binding were further isolated by identifying clusters where a number of high-affinity hotspots lie

closer to each other suggesting an active site (Figure 13c) (S. Table 6). To determine if the

pharmacophores generated stituated with in the active site, the *C*Las STP pharmacophore

clusters, the original subtrate, and the template/homology model (PDB: 4RKK[137]) were

overlapped. The protein-ligand structures after alignment showed that the selected *C*Las STP

active site overlapped with the dual specific phosphatase (DSP) domain on laforin (Figure 13d).
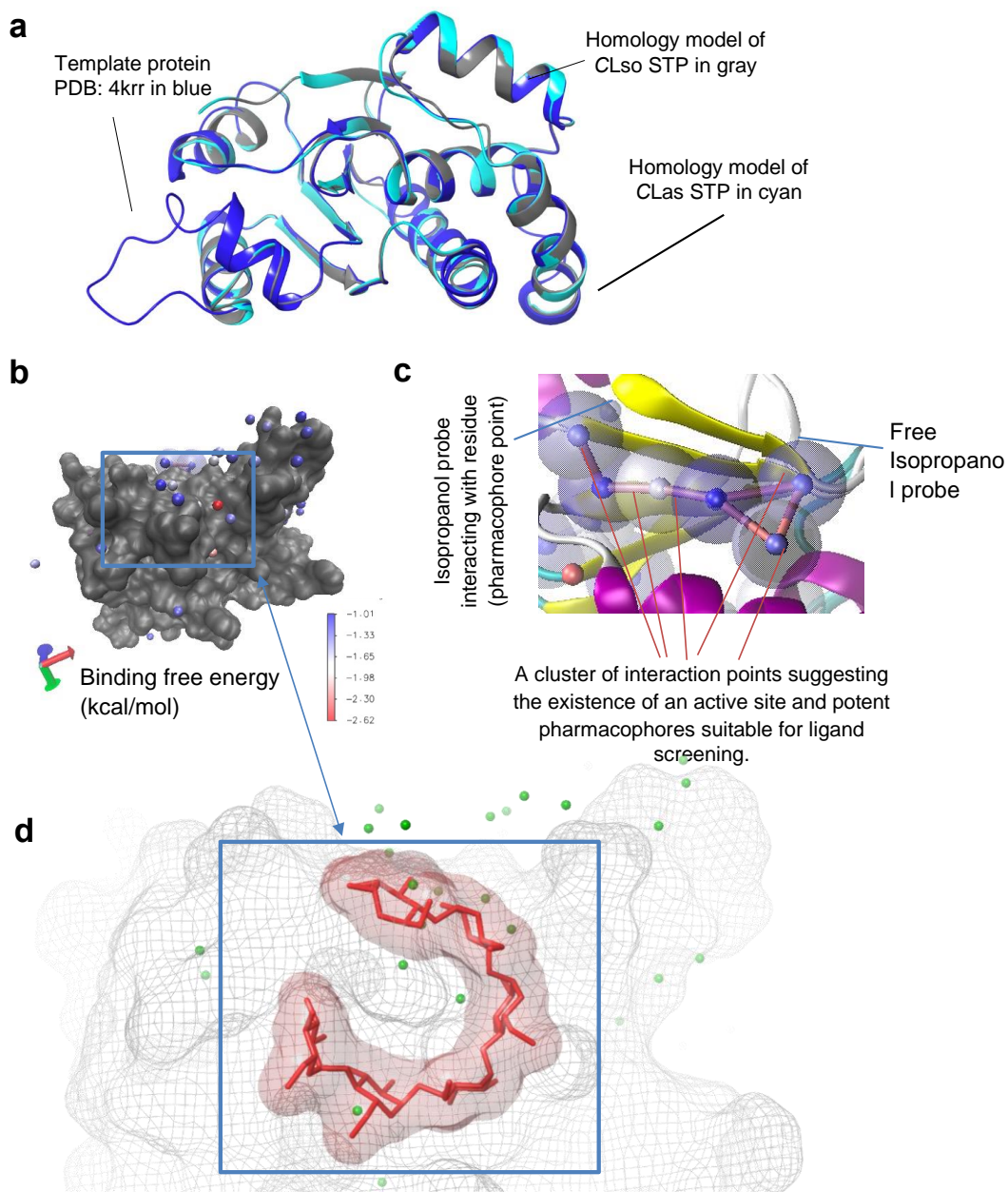
Figure 13 Homology modeling of *CL*so and *CL*as STP and identification of binding hotspots. (a) The homology model of *CL*so STP (gray) and *CL*as STP (cyan) overlaid with the template (PDB ID: 4RKK in blue). (b) Hotspot distribution when NAMD simulations were performed with select probes on the *CL*so STP enzyme. The red spot depicts a binding hotspot with highest affinity. The area demarcated by the red box depicts pharmacophore distribution at the primary site. (c) Multiple pharmacophores were clustered around the *CL*so STP enzyme implying potential binding site(s). (d) The selected pharmacophores (green) occupied a binding pocket in the *CL*as STP (gray mesh), consistent with the experimental substrate analog alpha-maltohexaose (red).
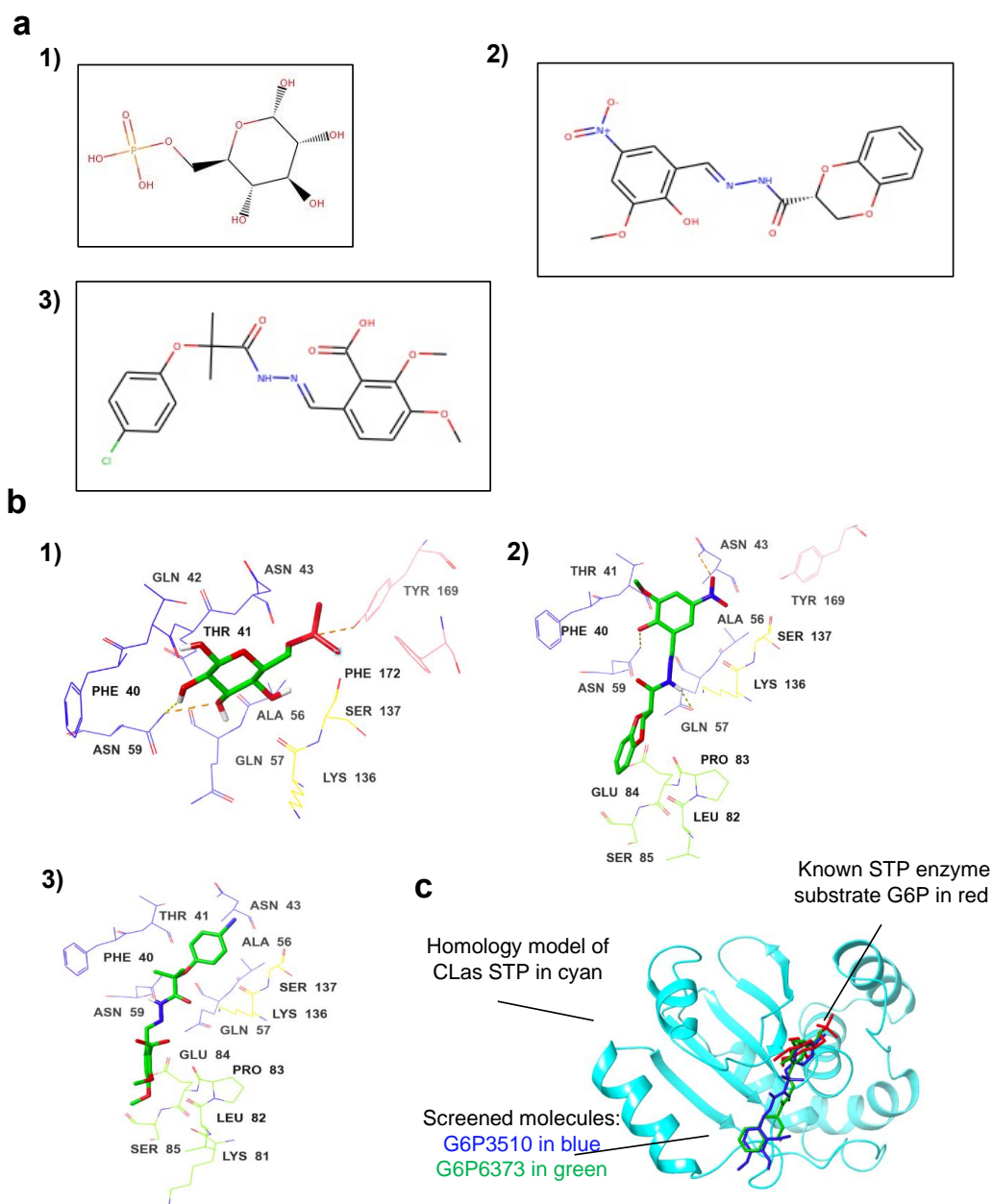
Figure 14 Identification of competitive inhibitors (G6P3510 and G6P6373) that bind to *C*Las STP. (a) structural information of the screened compounds; 1) Known STP enzyme substrate G6P; 2) Screened inhibitor G6P6373; 3) Screened inhibitor G6P3510. (b) The interactions of STP with G6P, G6P6373 and G6P3510, respectively. (c) G6P (red) occupying the active site and an isolated ligand (blue) occupying the same site as G6P on *C*Las STP (cyan).

Table 6 Structural and functional information of potential inhibitor ligands with G6P control.

Compound name	G6P	G6P3510	G6P6373
InChI Key	NBSCHQHZLSJFNQ-DVKNGEFBSA-N	FNDOMYALHUMMDS-SSDVNMTOSA-N	GKLIPLBJOGUXHG-RIEVCORGSA-N
Octanol-water partition coefficient logP	-3.31	3.90	-0.87
molecular weight (Daltons)	260.13	420.85	373.321
number of hydrogen bond acceptors	9	8	10
number of hydrogen bond donors	6	2	1
"Rule of five" violations	1	0	0
Number of Rotatable Bonds	7	8	5
Enzyme inhibitor score[a]	1.23	-0.19	-0.54
Average docking affinity on *CL*so STP, kcal/mol $\pm$ std (n=10)	-4.74 ± 0.196	-5.38 ± 0.379	-5.41 ± 0.438
Average docking affinity on *CL*as STP, kcal/mol $\pm$ std (n=10)	-4.38 ± 0.379	-5.30 ± 0.283	-5.27 ± 0.320

[a]The enzyme inhibitor score is one of the drug-likeness & bioactivity score provided by Molinspiration.

***In silico* ligand screening identified two competitive inhibitors (G6P3510 and G6P6373) of *C*Lso and *C*Las STPs**

Ligand screening was performed by pooling the features of both the ligand (via

Autodock) and the receptor pocket (NAMD/Drugui) using ELIXIR-A[125, 126]. This hybrid

approach allowed screening a pool of pharmacophore features orders of magnitude larger than

those that would have been possible when using either the ligand or pocket alone. Above

approach in turn, would allow deciphering more effective and site-specific inhibitors. The

zincpharmer algorithm was used to sift through ≥250 million compounds for possible hits that

match the identified pharmacophore with potential inhibitors [128]. Lipinski's rule of five along

with high pharmacophore scores in Zincpharmer was used to narrow down the hits [138]. The

screening efforts isolated two potential candidates with higher affinities (G6P3510, -5.38 ± 0.379

kcal/mol and G6P6373, -5.41 ± 0.438 kcal/mol) to the active binding site on *C*Lso STP than a

native substrate, Glucose-6-Phopsphate (G6P, -4.74 ± 0.196 kcal/mol) (Figure 14a. 1-3, S.

Figure 5a. 1-3). G6P is a primary metabolite in the cell fundamental to numerous biochemical

processes and is a substrate for the phosphatases. Importantly, both the inhibitors conformed

with the Lipinski's rule of five that defines properties necessary for good permeation[129] (Table

6). Based on the analysis of octanol-water partition coefficient (logP), G6P was likely to be

hydrophilic, while G6P3510 and G6P6373 were hydrophobic due to the presence of aromatic

groups. Furthermore, the G6P substrate appeared to have a higher bioactivity score (1.23), when

compared to the two screened compounds (-0.19 and -0.54) (Table 6). Interestingly, the two

novel compounds had distinct structures and functional groups, although they were predicted

based on similar pharmacophores compared to the original substrate G6P. This underscores the

value of large-scale screening of *in silco* chemical databases to identify potentially novel and

unique chemistries or interactions.

For any ligand to be an effective drug (i.e., an inhibitor), it is essential that, after binding, the ligand blocks any possibilities for the enzyme substrate to bind to the enzyme. This could happen only if the ligand binds more strongly to the pocket than the substrate, and if the ligand occupies a vast majority of possible binding sites that would otherwise be available for the substrate to bind. Thus, it is critically important that the substrate-binding site(s) and inhibitor-binding site(s) overlap. Accordingly, the nature of interactions were deciphered via Autodock Vina docking simulation, this time, in the presence of substrate(s) and potential inhibitor(s) on respective enzymes. Autodock Vina binding confirmations showed that the substrate and inhibitors had similar contact residue groups on CLso STP (S. Figure 5. 2b). G6P formed Van der Waals interactions with residues Asn 43 while forming hydrogen bonding with Tyr 160 and Thr 41 (with a distance of 1 Å). G6P3510 formed Van der Waals interactions with residues Tyr 40, Asn 43, Gln 57 while forming hydrogen bonding with Thr 41 (with a distance of 1 Å). G6P6373 formed Van der Waals interactions with residue Thr 59 while forming hydrogen bonding with Tyr 40 and Thr 41 (with a distance of 1 Å). In a manner similar to the CLso STP interactions, the two inhibitors, G6P3510 and G6P6373 also displayed high binding affinities of - 5.30 ± 0.283 kcal/mol and -5.27± 0.320 kcal/mol respectively vs. -4.38 ± 0.379 kcal/mol for G6P to the active binding site on CLas STP (Figure 14b. 1-3). On CLas STP (Figure 14c), the substrate and inhibitors showed similar contact residue groups as well. G6P formed Van der Waals interactions with residues Tyr 169 and Phe 40 while forming hydrogen bonding with Thr 41, Asn 43 and Asn 59 (with a distance of 1 Å). G6P3510 formed Van der Waals interactions with residues Lys136, Asn 43, Glu 84 and Ser 85 while forming hydrogen bonding with Asn 59 (with a distance of 1 Å). G6P6373 formed Van der Waals interactions with residue Phe 40, Gln 57, Asn 59 Pro 83, Ser 85, Lys 136 while forming hydrogen bonding with Asn 43 and Thr 41

55

(with a distance of 1 Å). In summary, the *in silico* simulations identified multiple potential interactions between inhibitors and the STP proteins.

Structure stability and intermolecular interaction analysis

MD simulations were performed to validate the performance and stability of molecules in the binding pocket. The G6P/CLas STP, G6P3510/CLas STP and G6P6373/CLas STP complexes were simulated for 100 ns to find the conformational changes within the active site. RMSD and RMSF values were used to determine whether the system had reached equilibrium during the simulation.

The RMSD changes of G6P/CLas STP, G6P3510/CLas STP and G6P6373/CLas STP complexes were analyzed to evaluate the stability of each complex during 100 ns MD simulations (S. Figure 6a). All three complexes plateaued after 20 ns of simulation and remained in the range of 2.5 Å to 4 Å. The final RMSDs were 3.082 Å, 3.683 Å, and 2.972 Å for G6P/CLas STP, G6P3510/CLas STP and G6P6373/CLas STP, respectively, which reached equilibrium.

The RMSF variances of the α-carbon residue of each residue in three complexes were calculated to find the flexibility of the protein structures. A higher RMSF value represents a much more flexible protein structure. Overall, the RMSF values of the three complexes ranged from 0.46 Å to 7.7 Å (S. Figure 6b). The outliers were from the tails. Although the tails were covered by the N-terminal peptide, they were still more flexible than the rest of the structure. No protein-ligand interactions were found at the last two amino acid positions, so the last two amino acids were excluded in subsequent RMSF analyses. After excluding these outliers, the RMSF ranged from 0.46 Å to 4.1 Å. Then, the average RMSF values were 1.33 Å, 1.32 Å and 1.22 Å for G6P/CLas STP, G6P3510/CLas STP and G6P6373/CLas STP, respectively. The two

molecules screened, G6P3510 and G6P6373, caused lower RMSF values than G6P, indicating

interactions between compounds and the protein (e.g., hydrogen bonding, hydrophobic

interactions, ionic and water bridge) had some impact on stabilizing the protein structure.

Intermolecular interaction analysis helps to reveal the binding properties of the ligands at

the active site. In the G6P/*C*Las STP complex, the protein-ligand interaction diagram indicated

the formation of three strong hydrogen bonds during a 100ns simulation (S. Figure 7a). The

residues Lys 73, Lys 129, and Lys 67 formed hydrogen bonds with G6P, which lasted 110.5%,

62.1%, and 61.9% of the simulation time, respectively. A ratio greater than 100% indicates that a

single residue forms multiple hydrogen bonds with the ligand at the same time. Meanwhile,

water bridge interactions were also detected on these three residues. In the last 1ns frame of the

simulation, G6P also formed three salt bridges with the residue Lys 67 and Lys 73 at the active

site, which increased the affinity for binding (S. Figure 8a).

In the complex of G6P3510/*C*Las STP, multiple strong hydrogen bonds on residue

Phe40, Thr41, Gln57, Asn 59, Lys 136 and Ser 137 were detected during the simulation,

accounting for 71.9%, 124.8%, 15.1%, 2%, 72.2% and 20.6% of binding contribution during the

MD simulation, respectively. (S. Figure 7b). Due to the presence of an aromatic ring, strong

hydrophobic interactions were detected at residues Phe 40, Ala 56, Tyr 169, Phe 172 and Pro

173, accounting for 8.1%, 7.6%, 14.9%, 46% and 14.0%, respectively. Even though the small

molecule remained in a fixed active site, more strong bonds were formed during the MD

simulation. Also, the constant presence of hydrogen bonds indicated that the molecule bound

stably at the active target site.

In G6P6373/*C*Las STP complex, five main hydrogen bonds were detected at residue Gln

57 (32.6%), Gly 60 (28%), Lys 136 (70.4%), Thr 188 (21.2%) and Gln 189 (25.2%).

Additionally, all these hydrogen bonds could switch to water-mediated hydrogen bonds during the simulation. Similar to the G6P3510 complex, this complex also detected more non-covalent bond formation during the simulation.

Throughout the simulations, all three molecules formed at least one or more stable non-covalent contacts with the protein. The two screened compounds had better binding properties compared to G6P due to the observation of more protein-ligand contacts. The simulation results confirmed the stability of the protein-ligand complexes.

In vitro **binding kinetic analysis on CLso and *C*Las STPs**

To validate the binding and determine kinetics of the potential inhibitors of STPs , we performed ligand-binding assays using Bio-layer Interferometry (BLI)[139]. Briefly, recombinant *C*Lso and *C*Las STPs were produced in *Escherichia coli* with C-terminus 6xHis tags, and purified by one-step purification using Ni column. Based on BLI binding kinetics data, the two test molecules, G6P3510 and G6P6373 along with the positive control G6P interacted on both STPs of *C*Lso and *C*Las generating association and dissociation signals (Figure 15a-b,). For *C*Lso STP, the affinity constants, K_D, for G6P3510 and G6P6373 were 1.381×10^{-5} M and 1.42×10^{-2} M respectively. The K_D for the native substrate G6P was 6.688×10^{-6} M. The K_D values indicated that G6P3510 had an affinity closer to that of G6P. Neither of the ligands showed significant K_D value comparing to the G6P native substrate ($\alpha < 0.05$). A molecule having an affinity closer to but higher than that of the native substrate suggests competitive inhibition. In comparison to G6P, G6P6373 dissociated faster on *C*Lso STP, leading to comparatively higher K_D values leading to lower affinity. In comparison to G6P, G6P3510 associated and dissociated faster; however, the affinity of the compound was slightly lower than that of G6P.
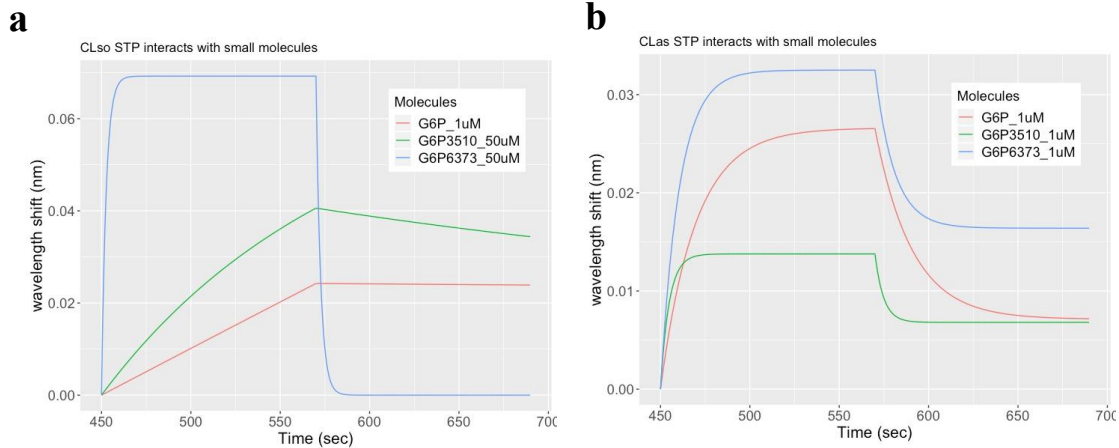
Figure 15 Biolayer interferometry data for *C*Lso and *C*Las STPs interacting with G6P, G6P3510 and G6P6373. (a) The BLI sensorgram shows the binding data of *C*Lso with 1 μM G6P, 50 μM G6P6373 and 50 μM G6P3510, respectively. (b) The BLI sensorgram shows the binding data of *C*Las with 1 μM G6P, G6P6373 and G6P3510, respectively. A known substrate G6P was used as the positive control in this assay. All experiments were conducted in three replicates; each of the curves were globally fitted using three replicates.

For CLas STP, the K_D for G6P3510 and G6P6373 was 8.144×10^{-5} M and 2.401×10^{-5}

M, respectively, while it was 2.361×10^{-5} M for the native substrate G6P indicating both

compounds had a comparable affinity to that of G6P (Table 7). It was evident that both the

compounds had a comparable association rate to that of G6P; however, G6P3510 dissociated

faster as compared to the other two. Neither of the ligands showed significant K_D value

comparing to the G6P native substrate ($\alpha < 0.05$). Overall, the BLI kinetics data indicate that

G6P6373 and G6P3510 had similar affinity comparing to G6P native substrate, suggesting

competitive inhibition.

Table 7 The kinetic energy analysis of the protein-ligand interactions. A known substrate G6P
was used as the positive control in this assay.

Ligand	Targeted protein	K_D (M)	K_a (1/Ms)	K_d (1/s)	R^2	K_D p-value[a]
G6P	CLso STP	6.688×10^{-6}	1.758×10^{1}	1.176×10^{-4}	0.4832	--
G6P3510	CLso STP	1.381×10^{-5}	1.064×10^{2}	1.469×10^{-3}	0.8156	0.3739
G6P6373	CLso STP	1.42×10^{-2}	2.691×10^{1}	3.82×10^{-1}	0.8876	0.05259
G6P	CLas STP	2.361×10^{-5}	2.166×10^{3}	5.114×10^{-2}	0.7641	--
G6P3510	CLas STP	8.144×10^{-5}	1.285×10^{3}	1.047×10^{-1}	0.5428	0.3778
G6P6373	CLas STP	2.401×10^{-5}	1.924×10^{3}	4.618×10^{-2}	0.6449	0.5864

a. The null hypothesis of two sample unpaired student t-test was $\mu_{control = \mu_{ligands}}$ and the
alternative hypothesis was $\mu_{control\_G6P} \neq \mu_{ligands}$.

Conclusion

In conclusion, this study used an *in silico* targeted pharmacophore screening strategy to screen and identify two small molecules (G6P3510 and G6P6373) with high binding affinity to a putative virulence protein of '*Candidatus* Liberibacter spp.'. The docking and molecular dynamics simulations confirmed the stability of the protein-ligand complexes. BLI assays suggested that the mechanism of inhibition may be competitive inhibition between the screened active compounds and the native substrates. Further *in vitro* and *in planta* experiments are needed to reveal the efficacy of the screened molecules. We suggest the two inhibitors, either singly or in combination, could be further deployed for citrus greening and zebra chip disease management.

Data availability

The datasets generated and analyzed during the current study are available from the corresponding author upon request.

Data analysis and statistics

All *in-vitro* experiments were conducted as a CRD with three replicates. Data analysis was done using Microsoft Excel software and R-studio. The standard deviation was calculated by the STDEV() function in Microsoft Excel. Data between different ligands were tested by two-sample unpaired Student's t-test. T test was given by the t.test() function in R-studio. The significance level was $\alpha = 0.05$.

CHAPTER V

IN SILICO SIMULATION OF *CANDIDATUS* LIBERIBACTER ASIATICUS EFFLUX PUMP

AND INHIBITION BY ANTIMICROBIAL PEPTIDES.

Introduction

Citrus greening or HLB disease is the deadliest disease threatening citrus production worldwide. The disease is associated with the unculturable bacterium *Candidatus* Liberibacter asiaticus (*C*Las), which is transmitted between trees by the Asian citrus psyllid *Diaphorina citri* [140]. There is no treatment for the disease, and citrus trees become unproductive and die within a few years. After its emergence in the 1920s in China, HLB widely spread to other countries in Africa and Asia. In 2005, the disease was first reported in the U.S. in South Florida, and within a year, it led to an approximate 50 percent reduction in Florida's citrus production, resulting in economic losses of more than $8 billion [141]. HLB has now spread to many citrus-producing areas in the US, including California, Florida, Georgia, Louisiana, Puerto Rico, South Carolina, Texas, and the U.S. Virgin Islands [142]. Current management regimes are focused on early detection and removal of infected trees and insect control. Few treatment strategies include a combination of broad-spectrum antibiotics, insecticides, nutritional supplements, and thermotherapy [143, 144]. The U.S. Environmental Protection Agency recently approved the use of oxytetracycline [145] and the emergency use of streptomycin [146] on citrus. However, health experts have expressed concerns about the use of medically important antibiotics in agriculture due to the increased risk of antibiotic resistance in humans[147-150].

As with most Gram-negative bacteria, *C*Las utilizes secretion systems to transport proteins and nutrients across the membrane to manipulate host cells [151]. Some studies have shown that *C*Las secreted proteins can regulate *C*Las multiplication and colonization of the citrus

phloem cells by altering host cellular functions [152]. Blocking bacterial secretory pathways can

be a promising approach for novel antimicrobial discovery [153]. Whole-genome sequencing

analyses revealed that *C*Las lacks the type III and type IV secretory systems as well as the type II

plant cell-wall degrading enzymes, which are known to play a key role in the pathogenicity of

gram-negative bacteria [140]. *C*Las expresses all genes of type I secretory system (T1SS). The

primary functions of T1SS are multidrug efflux and effector protein secretion. The T1SS is a

tripartite machinery composed of an inner-membrane protein (IMP) ABC transporter, a

periplasmic membrane fusion protein (MFP), and TolC, an outer membrane protein (OMP)

[154]. The substrates (proteins or drugs) bind to the IMP and are transported across the inner

membrane from the bacterial cytoplasm into the periplasm of the MFP, using the energy

produced from ATP hydrolysis. Upon substrate binding, the IMP-MFP complex associates with

TolC, which induces the opening of TolC and the subsequent release of the substrate outside the

bacterial cell [153]. Inhibiting the activity of the tripartite multidrug efflux pump [155] can

terminate the secretion of virulent effector proteins and can prevent drug efflux, providing some

protection against the pathogen (Figure 16).

The mechanism of how HLB *C*Las secreted effector proteins damage the host phloem

cells remains elusive [156]. However, one *C*Las effector CsACD2 has been reported to interact

with citrus papain-like cysteine proteases that break down the citrus defense system, and hence

enhances the virulence of the pathogen [157, 158], and the sec-dependent system can be targeted

for the control of *C*Las virulence and pathogenicity. AMPs are short sequences of amino acids.

Some AMPs are known to activate the innate immune response in the host [159]. Several studies

have shown that AMPs have good activity against pathogenic bacteria in plant systems [160,

161]. While studies exist on the use of AMPs and antimicrobial small molecules against CLas,

currently, there are no studies regarding the use of AMPs for the treatment of *C*Las through inhibition of the bacterial secretion system [162-164]. In this study, we screened AMPs that interact with the efflux pump TolC of the T1SS system to suppress the pathogenicity of *C*Las. A combination of *in silico* techniques was used to virtually screen AMPs targeting channel-tunnel protein. In addition, molecular dynamics simulations and binding free energy calculations of the interactions of the complexes were performed to evaluate the effectiveness of the screened AMPs. Using this approach, we aimed to identify novel AMPs that have the potential to be used for HLB management.
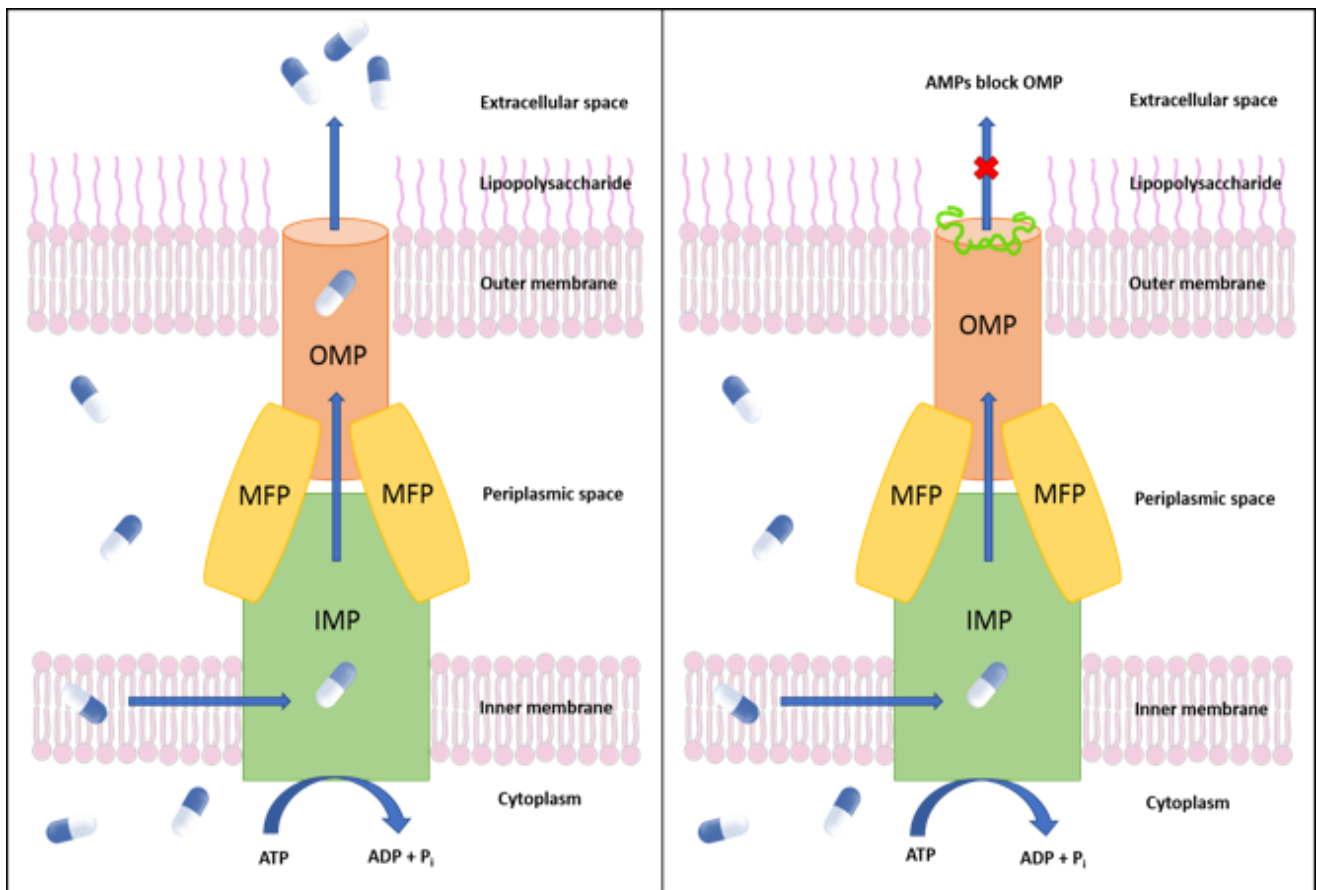
Figure 16. Left: Schematic representation of an efflux pump in *Candidatus* Liberibacter asiaticus. Right: Inhibition of the *C*Las TolC outer membrane protein (OMP) by AMPs. MFP: membrane fusion protein; IMP: inner membrane protein.

Materials and methods

Sequence analysis and homology modeling

Computational predictions of the three-dimensional structure of proteins were constructed using comparative homology modeling techniques. The *C*Las efflux pump protein is identified by the NCBI unique identifier CLIBASIA_04145 in *C*Las [165]. Using the SWISS-MODEL server [84], the crystal structure of the efflux pump component OprN from *Pseudomonas aeruginosa* (PDB: 5iuy) [166] was used as a template to construct a structural model of the *C*Las TolC efflux protein. The Ramachandran plots were visualized by the PROCHECK [167] server as energetically allowed regions of skeletal dihedral angles ψ to amino acid residues φ in the protein structure.

The virtual screening of AMPs

The initial set of AMPs were screened from the Antimicrobial Peptide Database (APD3) [168]. Several criteria were considered for screening AMPs targeting the membrane protein tripartite efflux pump in *C*Las. First, AMPs should be active against Gram-negative bacteria. Second, the availability of crystal structures in the PDB. Third, the sequence length of the AMPs was around 15 amino acids to allow stability and bioavailability. Additionally, the AMPs with known membrane protein activities were also included.

Molecular docking

The binding mode prediction of AMPs with the *C*Las efflux pump receptor complex was performed on the Schrödinger Glide platform [47]. Both the protein and AMPs were prepared by Schrödinger's Protein Preparation Wizard [131]. The AMPs were docked on the rigid protein receptor to find potential binding poses with the highest docking scores. The docking mode was peptide standard precision with top 50 possible docking poses. The ligand-receptor interactions

were visualized using Schrödinger Maestro [132], and those AMPs with multiple and stable non-covalent interactions were selected for MD simulations.

Molecular Dynamics simulations

Some promising AMPs were selected to study the possible mode of action (MOA) on the receptor after the docking simulation [130]. The MD simulation was performed on Schrödinger Desmond platform [130]. The system charge was neutralized with Na or Cl ions and a 0.1M concentration of NaCl. The system was built with the Simple Point-Charge (SPC) solvent model. The boundary conditions were set as orthorhombic box with a buffer distance of 10 Å. The model membrane of the system was the bilayer 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) membrane. The protein-ligand complex was prepared by Schrödinger's Protein Preparation Wizard [131]. The docked AMPs were exported from Glide. The C and N termini of the protein were capped for stabilizing the protein structure. All the missing hydrogens were added and hydrogen bonds were optimized. The strained minimization was performed with the OPLS3e force field [133].

Each simulation was performed with Desmond's default relax protocol. The force field for the simulation was also OPLS3e. The system heavy atoms were first minimized with restrains under 10K, then increasing the temperature to 300K with restrains and the final relaxation step under 300K Normal Pressure and Temperature (NPT) ensemble to get the equilibrium status for the system. After relaxation, the simulations were performed under a 300K NPT ensemble at 1.01325 bar pressure for 100 ns. The recording interval was 200ps and 500 frames were saved. Post simulation trajectory analyses including complex RMSD, and ligand/protein RMSF, complex interactions were performed by Schrödinger Simulation Interactions Diagram.

Calculation of the binding free energy

Binding interactions of peptide-protein complexes can be calculated using the MM-GBSA binding energy method [169]. The primary MM-GBSA uses the VSGB 2.0 dissolution model with the OPLS3e force field.

Docking poses of the AMPs on receptor were evaluated by the calculation of the total binding free energy using Schrödinger Prime. Prime MM-GBSA calculations gave the complex binding energies which validated the performance of the current binding conformation. The Schrödinger Prime calculates the energy of the AMP-protein system via MM-GBSA using Desmond simulation trajectory. From the entire 100 ns simulation, the last 50 ns trajectory was chosen for energy calculation. The free binding energy ΔG is calculated as the energy of the receptor-ligand complex minus the energy of the receptor alone and the ligand alone, as follows:

$$\Delta G(bind) = E\_complex(minimized) - (E\_ligand(minimized) + E\_receptor(minimized))$$

Principal component analysis (PCA)

To study the overall motion of the protein system throughout the entire simulation trajectory, principal component analysis (PCA) for pairwise distance of alpha carbon atoms (Cα) was performed. VMD [170] software was used to prepare the topology and trajectory files. The python package MDTraj [171] was utilized for analysis and visualization. Cα were extracted from 500 snapshots of 50 ns MD trajectories to construct the structure matrix. Every snapshot was aligned to the starting frame before constructing the eigenvectors. The first two principal components (PC1 and PC2) were projected to represent the protein dynamics in the MD trajectories of the AMP-protein and the apo-protein systems.

Results and disscussion

Protein structure modeling and validation

Structure-based design of inhibitors relies on the knowledge of three-dimensional molecular structure of the target proteins. Since, no experimental structures were available for the *C*Las TolC protein, this was done using the homology modeling approach based on the amino acid sequence of *C*Las TolC protein and 3D structure of a related protein. Multiple efflux pump templates were found by BLAST and the selected template had 30.16% sequence identity and 0.35 sequence similarity in Table 8. The trimer homology model was built by the multidrug efflux outer membrane protein (OMP) OprN (PDB: 5iuy) [40, 166]. The predicted homology models were first preprocessed and prepared using Schrödinger's Protein Preparation Wizard. Then the Ramachandran plot model was generated by PROCHECK [167]. According to the PROCHECK report, 89.3% of the residues were located in the most favored regions, 10.6% in the additional allowed regions, 0.1% in the generously allowed regions, and 0% in the disallowed regions in Figure 17. The residue level in the Ramachandran favored score was 99.9% (1091/1092), which met ideal conditions (>98%) [172]. This indicated that the protein backbones were possible, and this conformation was stable. The Z score was predicted by comparison with the non-redundant PDB structure set of the proposed model (S. Figure 9). The Z score showed that the model was less than 1 standard deviation from the mean models. Based on overall structural assessment, the predicted model was considered reliable for further *in silico* studies.

Table 8 Homology models generated using SWISS-MODEL

Template	Chain	GMQE	QSQE	Seq Identity	Seq Similarity	Membrane protein	Ref
5iuy	A	0.71	0.51	30.16	0.35	OprN	[166]
4mt0	A	0.71	0.47	26.05	0.33	MtrE	[173]
3pik	A	0.71	0.43	23.65	0.32	CusC	[174]
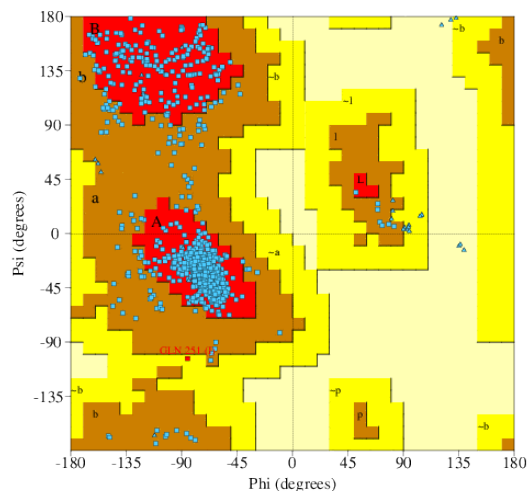
Figure 17 Ramachandran plot of the homology model. The residues are shown as blue dots. The φ values were on the x-axis and the ψ values were on the y-axis. φ was the N(i-1),C(i),Ca(i),N(i) torsion angle and ψ was the C(i),Ca(i),N(i),C(i+1) torsion angle. The red, brown and yellow areas represent the favored, permitted and permissive areas as determined by ProCheck.

Virtual screening of AMPs

Potential AMPs were screened from the APD3 database, including 2619 AMPs which were mostly natural antimicrobial peptides. The AMPs were selected based on previous evidence of inhibitory activity against bacterial efflux pumps. Preliminary screening of 15 antimicrobial peptides associated with bacterial membranes was carried out from the filter search. Eight AMPs with membrane activity against Gram-negative bacteria were further selected based on experimental data listed in Table 9.

The AMPs were screened based on their affinities to the β-barrel structure of the TolC protein. The β-barrel structure consists of residues Ala 94 to Leu 128 and Tyr 300 to Gly338. All the homology models of the peptide candidates had α-helical conformations and positive net charges. These screened peptides came from a diverse range of sources, such as animals, insects, plants, and cultured bacteria. These included temporins, urechistachykinins, colistins, plantaricin, and darobactin. Temporins are a family of antimicrobial peptides from the European red frog Rana *temporaria* [175]. These peptides were originally extracted from frog skin secretion.

69

Tachykinin-related peptides Urechistachykinin I and II were found from the echiuroid worm,

Urechis unicinctus [176]. They have known effects on neurons and G protein-coupled receptors

in animals. Colistin (polymyxin E) has been used for more than 50 years as an antibiotic to treat

infections caused by Gram-positive bacteria such as *Bacillus subtilis* [177, 178]. The bacteriocin

plantaricin JLA-9 was found from Suan-Tsai, a traditional Chinese fermented cabbage [179].

JLA-9 has a broad-spectrum antibacterial activity against Gram-positive and Gram-negative

bacteria under acidic conditions. Trp-rich peptide TetraF2W-RK is a synthetic peptide [180]. The

peptide exhibited good activity in targeting bacterial membranes and was able to inhibit the

antibiotic-resistant bacteria *Staphylococcus aureus* [181]. Darobactin is a newly found antibiotic

that was isolated from *Photorhabdus* spp. in 2019 [182]. Darobactin is effective against

antibiotic-resistant Gram-negative pathogens, has low toxicity, and good pharmacokinetics *in*

vivo.

Table 9 The information of the screened AMPs. All peptides have been shown to be effective
against Gram-negative bacteria.

No	Peptide	Definition[a]	APD3 ID	Length	Hydrophobic residue%	Net Charge
1	LSPNLLKSLL	Temporin H	AP00859	10	50	2
2	LRQSQFVGSR	Urechistachykinin I	AP01480	10	30	3
3	AAGMGFFGAR	Urechistachykinin II	AP01481	10	60	2
4	KTKKKLLKKT	Colistin A	AP02204	10	20	6
5	FLPLIGRVLSGIL	Temporin A	AP00094	13	61	2
6	FWQKMSFA	Plantaricin JLA-9	AP02677	8	62	1
7	WWWLRKIW	TetraF2W-RK	AP02856	8	75	3
8	WNWSKSF	Darobactin	AP03168	7	42	1

a. The peptides modeled in this study were generated using linear amino acid sequences. The
structures of the peptides tested may not be identical to the known AMPs.

Molecular docking analysis (GLIDE)

The conformations with the seven AMP-protein complex docking scores were analyzed for the intermolecular interactions with SP peptide docking. Table 10 shows that the binding affinity of AMPs ranged from -2.787 to -9.605 kcal/mol and -7.678 kcal/mol for the efflux pump inhibitor MRL-494 as the positive control [183]. All the molecules showed multiple hydrogen bonds with the protein complex. Darobactin with the highest binding affinity -9.605 kcal/mol simultaneously formed five hydrogen bonds with residues ASP314, ASN315 on chain G, ASN315, SER316 residues on chain H, and ASN312 residue on chain I. Also, eleven polar residues and four hydrophobic residues were proximal to darobactin at a cut-off distance of 3 Å. Only two negative charged residues ASP314 on chain G and ASP314 on chain H and no positive residues were involved in the interaction. Urechistachykinin II had a binding score of -9.332 kcal/mol and formed six hydrogen bonds. One π-π stack interaction, 13 polar residues, seven hydrophobic bonds, two negatively charged residues, and two positively charged residues were involved in the interaction of urechistachykinin II with the protein. Plantaricin JLA-9 had a binding affinity of -9.002 kcal/mol and formed six hydrogen bonds. One π-π stack interaction, one π-cation interaction, nine polar residues, four hydrophobic bonds, two negatively charged residues were involved in the interaction of plantaricin JLA-9 with the protein. All of the screened AMPs and MRL-494 interacted well with all three sub-chains of TolC, simultaneously. Similar to MRL-494, darabatin, urechistachykin, and planataricin JLA-9 all formed H bonds with one or more of the residues among ASP314 on chain G and ASP314 and ASP315 on chain H, suggesting that these three peptides bind to the same domain as MRL-494.
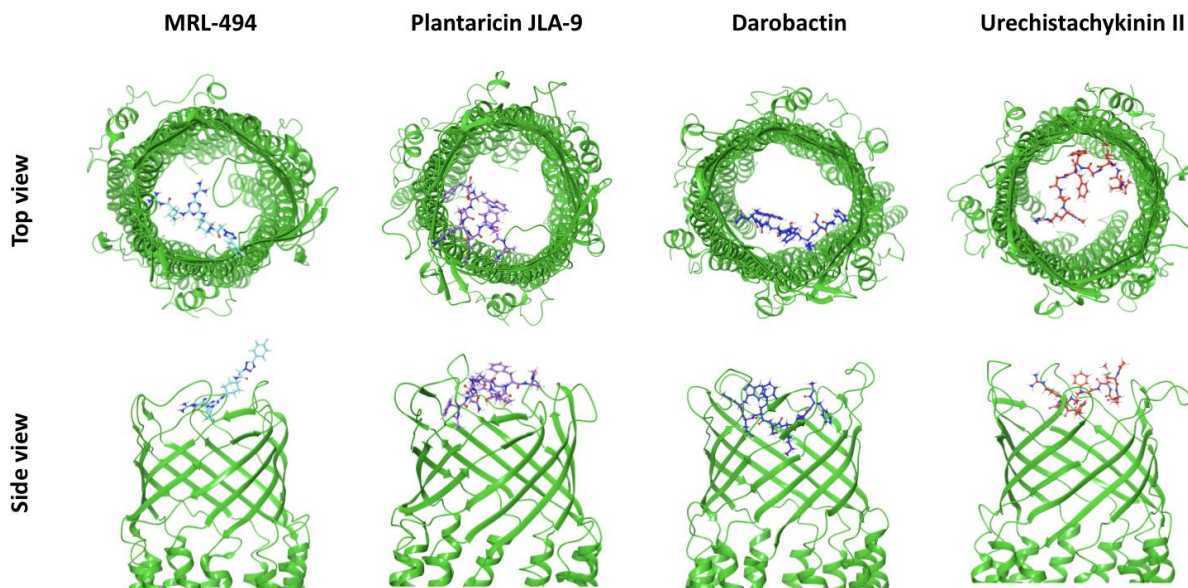
Figure 18 The structural snapshots of the selected AMPs interacting with TolC at the β-barrel sites.

Since the three AMPs darobactin, urechistachykinin II and plantaricin JLA-9 have a good binding affinity and tight interaction with TolC, these three AMPs were selected for further kinetic simulation analysis. In Figure 18 and S. Figure 10, S. Figure 11, and S. Figure 12, the interaction diagram of the three AMPs mentioned above showed that the docking position was at the outlet of the efflux pump. This interaction reduced the surface area at the outlet of the efflux pump to inhibit the efflux function of the protein.

Molecular Dynamics Simulation

To understand the structural and conformational changes during peptide binding, molecular dynamics simulations of 100 ns were performed for five complex systems (apo-protein, known inhibitor, and three selected AMPs). The RMSD of atomic positions and RMSF were monitored to evaluate the stability of the protein and peptide complexes using the simulation trajectory. In Figure 19, the protein Cα RMSDs were compared, and the peptide bound complexes showed lower deviation. All simulations stabilized in the RMSD range of 4 to 6 Å after 50 ns. The mean RMSD of the apoprotein was 4.99 ± 0.487 Å. The average RMSD of

MRL-494 was 4.38 ± 0.338 Å. The mean RMSD of darobactin, plantaricin JLA-9, and

urechistachykinin II were 4.83 ± 0.510 Å, 4.72 ± 0.407 Å, and 4.11 ± 0.258 Å, respectively. The

simulations of urechistachykinin II showed the lowest fluctuation and deviation compared to the

other groups.
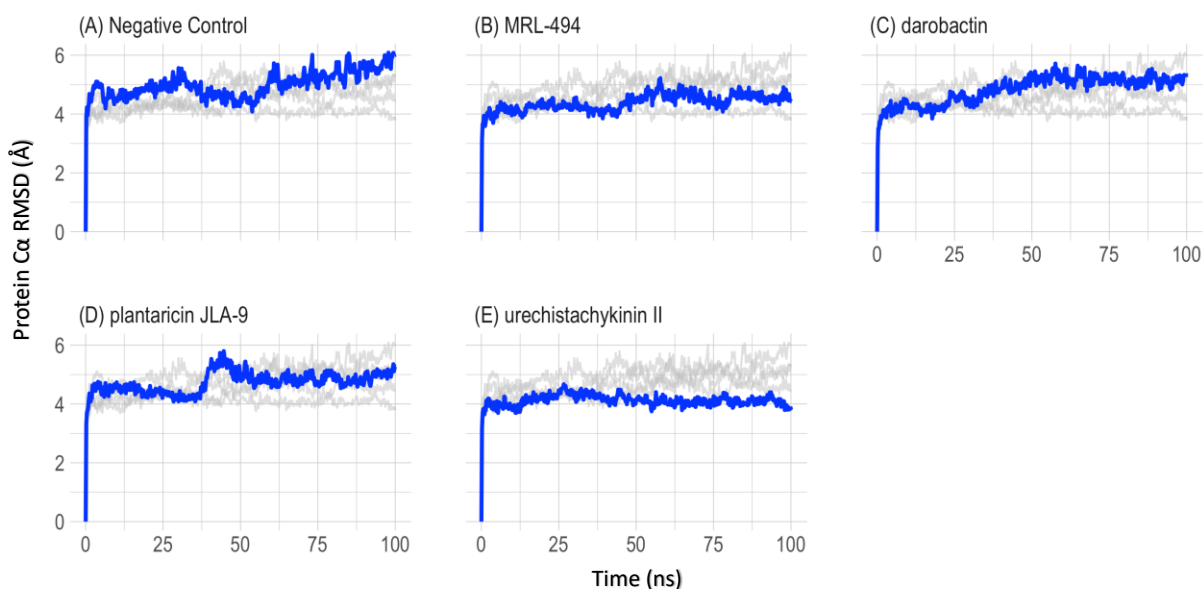
Figure 19 Cα RMSD for the 100ns MD of TolC efflux pump complex. The main data for each
graph are marked in blue, and the other four sets of data used as comparisons are marked in gray.
The RMSD of all simulated complexes of Cα was stable within the interval of 4 to 6 Å.

In Figure 20, Cα RMSF measured the flexibility and stability of protein residues. To

clearly view the contribution of the ligand to the residues, RMSF maps were prepared for each of

the three sub-chains of the trimeric efflux pump protein. The residues with high RMSF values

were always considered as highly flexible regions. The average RMSF at the active binding

region of the apo-protein was 10.4 ± 1.53 Å. And the RMSF of MRL-494 was 3.56 ± 0.513 Å,

the mean RMSF of darobactin, plantaricin JLA-9 and urechistachykinin II were 5.64 ± 0.810Å,

5.51 ± 0.860Å, and 4.62 ± 0.775Å, respectively. All AMP-protein complexes had lower RMSF

values compared to the apo-protein, suggesting that binding could reduce the flexibility of the

efflux protein.

73

Prime MM-GBSA Analysis

MM-GBSA was performed to quantitatively calculate the binding energies between the AMPs and the TolC protein. In Table 11, the overall average binding free energy (ΔG) and the specific binding free energy components are presented. All listed energies were obtained and calculated from the last 50ns of coordinate sampling of the simulated trajectory. The ΔG of MRL-494 was -63.55 ± 8.51 kcal/mol. The ΔG of darobactin, plantaricin JLA-9, and urechistachykinin II were -42.09 ± 11.20 kcal/mol, -51.84 ± 10.90 kcal/mol, and -46.15 ± 12.09 kcal/mol, respectively. The energies were expressed as their mean and standard deviation. The ligand strain energy represents the energy cost of getting the compound into the active pocket. This is the energy difference between the optimized drug-protein complex and the drug outside the binding pocket. The ligand strain energies for the screened peptides were in increasing order, plantaricin JLA-9, darobactin, and urechistachykinin II. Lower ligand strain energies are associated with a ligand more easily accommodating and fitting to its associated active site.

Table 10 AMPs screened by APD3 and the Standard Precision (SP)-peptide mode docking score and interactions with receptor. The selection of residue groups located within 3Å of the AMP ligand.

No	Definition	Docking Score	glide emodel	H-bond	π-π stack	π-cation
1	Temporin H	-8.971	-120.362	G: ASP314; H: ASN315, SER316; I: ASN315, SER316		
2	Urechistachykinin I	-8.610	-88.383	G: ASP314, ASN315; H: ASP314, ASN315, SER316; I: SER316		
3	Urechistachykinin II	-9.332	-146.077	G: LYS113, HIE313; H: ASN315, SER316, PHE317; I: SER316;	I: PHE317	
4	Colistin A	-7.325	-45.045	G: ASP314, SER316, PHE317, TYR320; H: ARG107, ASP314; I: SER316		G: TYR320

5	Temporin A	-2.787	2.406	G: ASP314, ASN315		
6	Plantaricin JLA-9	-9.002	-109.183	G: HIE313, ASP314; H: ASP314, ASN315, SER316; I: SER316	I: TYR320	H: ASP314
7	TetraF2W-RK	-8.700	-122.832	G: ASP314, ASN315, ASN321; I: SER316	G: PHE317	
8	Darobactin	-9.605	-144.059	G: ASP314, ASN315; H: ASN315, SER316; I: ASN312		H: ASP314, PHE317
9	MRL-494	-7.678	-66.449	G: ASP314; H: ASP314, ASN315;		
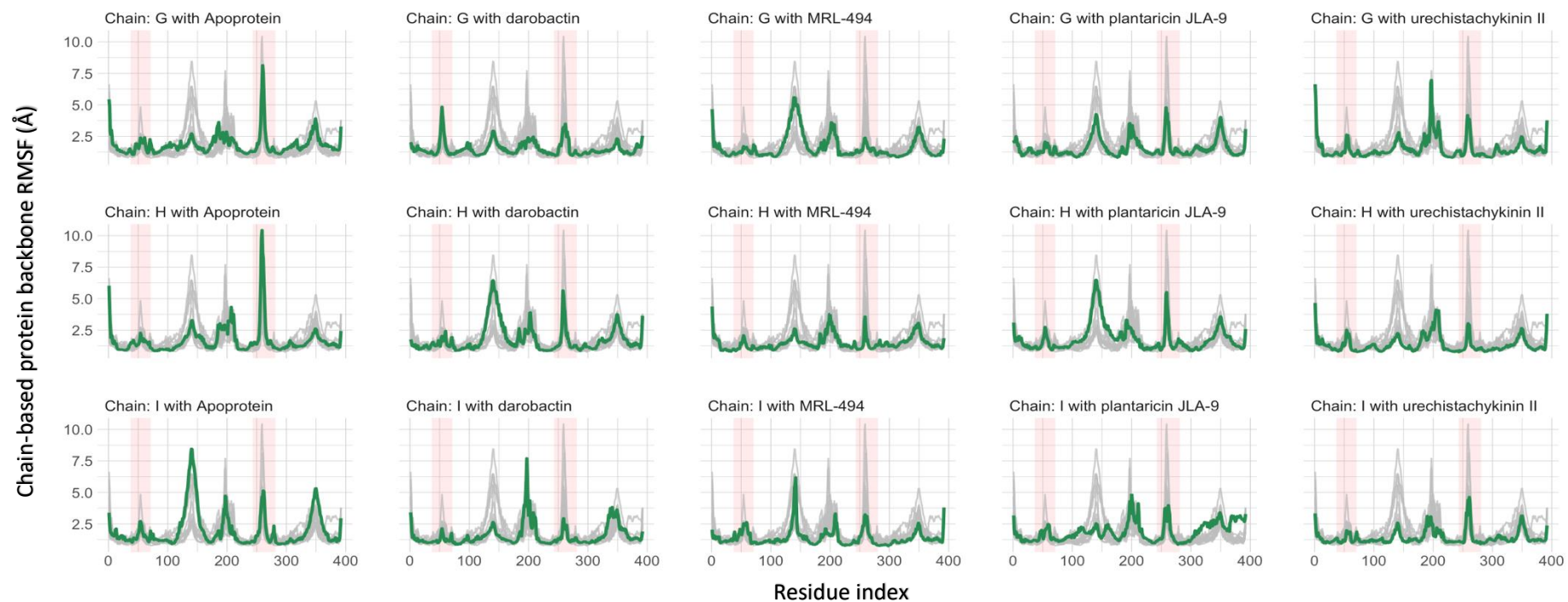
Figure 20 The RMSF for the 100ns MD of TolC efflux pump complex. Backbone RMSFs are plotted separately (green) and compared with all RMSF, including apo-protein (gray) in the other plots. Interested active β-barrel sites are highlighted in red. There are four β-sheets per chain, and each highlighted column represents two β-sheets.

Table 11 AMPs screened by APD3 and the SP-peptide mode docking score and interactions with receptor.

Compound Name	ΔG overall (kcal/mol)	ΔG Coulomb energy (kcal/mol)	ΔG Covalent binding energy (kcal/mol)	ΔG Lipophilic. energy (kcal/mol)	ΔG Van der Waals energy (kcal/mol)	ΔG Generalized Born electrostatic solvation energy (kcal/mol)	ΔG ligand strain (kcal/mol)
Darobactin	-42.09±11.20	-7.78±40.00	1.49±4.29	-8.10±4.60	-48.49±7.58	23.70±35.44	17.69±9.98
Plantaricin JLA-9	-51.84±10.90	-8.24± 14.95	1.34±3.48	-13.09±3.39	-46.83±10.72	18.40± 12.91	10.30±6.76
Urechi-stachykininII	-46.15±12.09	-19.16±24.41	-1.27±4.50	-13.88±2.79	-48.25±8.78	39.78±20.54	29.48±5.98
MRL-494	-63.55±8.51	-39.08±26.15	5.64± 2.20	-12.55±1.97	-43.96±4.64	34.24±25.02	7.35±2.88

Principal component analysis (PCA)

To understand the motion of the simulation, PCA was performed using the pairwise distance method. Five hundred frames were sampled to represent the changes across the simulation (Figure 21A-E). The first two eigenvectors PC1 and PC2 in all five PCA tests contributed approximately 50% of the variance to the motion. During the MD simulation, the conformational changes of the docked peptide ligands and protein receptors are represented by a color gradient from purple to yellow. In Figure 21F, it showed that the first few eigenvectors explained most of the variance, and when the number of eigenvectors reached 100 it explained practically 90% of the total variance. The PC1 vs PC2 scatter plots could explain the structural changes by the AMP interactions. The graphs showed consecutive point changes and no outliers

were found, indicating a smooth simulation. The PC1 of apoprotein exhibits a higher range

(~300) than that of the interacting complexes (~200). This phenomenon also occurred on the

apoprotein's PC2. Based on PCA, clustering of data points suggests structural stabilization while

scattering suggests structural changes. The interacting complexes had low structural variation

according to the first and second principal components. According to the last 50 ns of the

simulation, all AMP groups underwent a point of cluster distribution, which was thought to be a

reduction in protein flexibility due to binding, thus limiting the variability of protein

conformation.
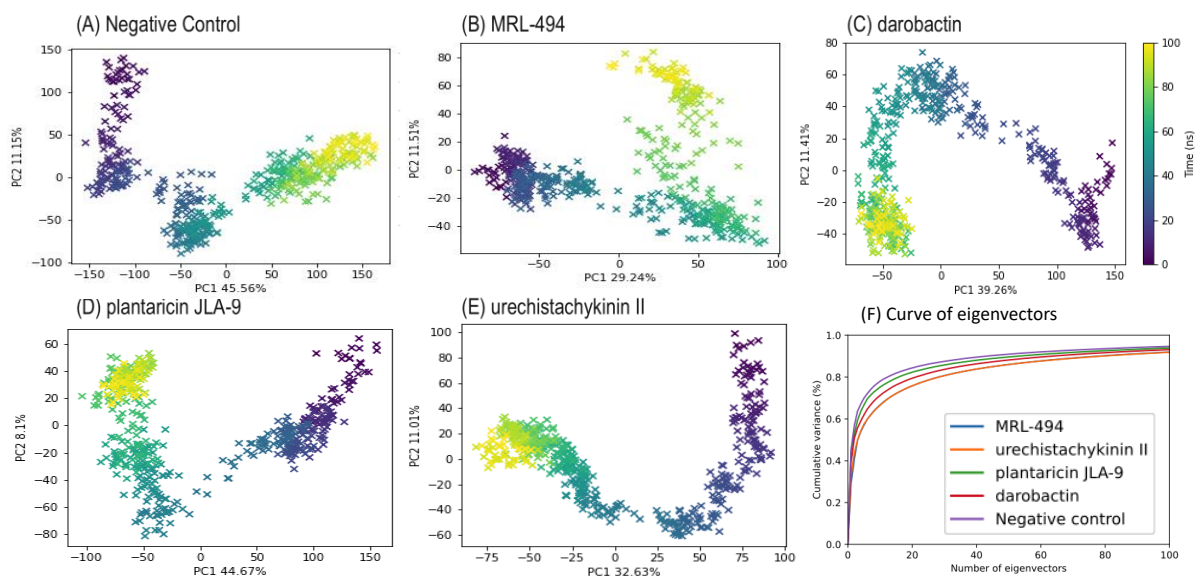
Figure 21 PCA on Cα atoms from MD simulation constructed by first two eigenvectors PC1 and PC2. A) Apoprotein (Negative control) B) MRL-494 C) darobactin D) plantaricin JLA-9 E) urechistachykinin II F) explained variance curves of the first 100 principal components for these five simulations.

Discussion

Multidrug efflux proteins are resistant to many small molecules, including antibiotics. T1SS is the only secretion system currently found in the *C*Las bacteria. Finding an effective AMP to bind and inhibit the outer membrane component may be an effective way to reduce the efflux function of the entire transmembrane protein. Several biological informatic technologies were applied to analyze the potential inhibitors for the *C*Las efflux pump protein.

The number of AMPs in the library of APD3 were narrowed down by adding conditions that were compatible with the membrane-active peptide. Short peptide macromolecules of less than fifteen amino acids in length were selected as one of the screening criteria due to their good bioavailability and stability [184, 185]. As a result, eight AMPs were initially screened for further molecular docking and MD experiments. Molecular docking showed that the binding affinities of six AMPs were higher than that of the positive control. All these peptides bound to the outer membrane position of the efflux pump. Therefore, the three lowest energy macromolecules darobactin, plantaricin JLA-9, and urechistachykinin II were selected for further analyses. To verify the inhibitory binding of the three AMPs, molecular dynamics simulations of the protein-membrane complex in a water box were performed. AMP-interacting systems showed lower fluctuations than that of the apoprotein from the RMSD and RMSF plots. This suggested that the interaction of AMP and receptors limited the flexibility of the outer membrane assembly and thus may provide a partial inhibition of protein activity. The PCA plot also confirmed that simulations with the involvement of AMP had lower fluctuation. These AMP-protein structures also hindered the cylindrical outlet, which can reduce or even block the enforced excretion of intracellular drugs. From the calculations of the various categories of

Gibbs binding free energy, plantaricin JLA-9 exhibited good binding scores and low strain energy suggesting its potential as an efflux pump blocker.

Conclusion

In this study, we screened AMPs as potential inhibitors for *C*Las via blocking TolC, a key protein in the efflux pump of the T1SS system. Multiple *in silico* approaches such as homology modeling, molecular docking, MD simulations, MM-GBSA calculations, and PCA were used to uncover potential AMPs against the outer membrane protein TolC. Three AMPs were found to have good binding affinities from the initial docking study. MD simulations of 100 ns showed the ability of AMPs to reduce protein flexibility and hinder tunneling effects of protein receptors. Based on further PCA and Gibbs free energy calculations, darobactin, plantaricin JLA-9, urechistachykinin II, showed promising ability to block the β barrel entrance of the TolC protein. Further *in vitro* studies (conducted elsewhere) showed that plantaricin JLA-9 has the potential to inhibit *C*Las bacteria by the combined action of antimicrobial activity and blocking effector efflux.

Data statistics and analysis, and software availability

The protein-ligand interaction data and whole Desmond simulation reports were available in the Github repository. (https://github.com/sfernando-BAEN/AMP_project)

The Schrödinger drug discovery platform version 2018-4 was used in this study. On this platform, Desmond used an academic license. The Protein Preparation Kit, Glide, and Prime used a commercial license. MDTraj package version 1.9.4 was performed in the python 3 environment. The studies were conducted as a CRD with three replicates. The RMSD and RMSF plots were prepared by ploty packakge version 4.9.2.1 in R Studio. The statistics of the simulation data were analyzed using R Studio. Data between different ligands were tested by

two-sample unpaired Student's t-test. T test was given by the t.test() function in R-studio. The significance level was $\alpha = 0.05$.

REFERENCES

1. Minasov, G., X. Wang, and B.K. Shoichet, *An ultrahigh resolution structure of TEM-1 beta-lactamase suggests a role for Glu166 as the general base in acylation.* J Am Chem Soc, 2002. **124**(19): p. 5333-40.

2. Rose, A.S., et al., *NGL Viewer: Web-based molecular graphics for large complexes.* Bioinformatics, 2018.

3. Jana, S., A. Ganeshpurkar, and S.K. Singh, *Multiple 3D-QSAR modeling, e-pharmacophore, molecular docking, and in vitro study to explore novel AChE inhibitors.* RSC advances, 2018. **8**(69): p. 39477-39495.

4. Lauria, A., et al., *IKK-β inhibitors: An analysis of drug–receptor interaction by using Molecular Docking and Pharmacophore 3D-QSAR approaches.* Journal of Molecular Graphics and Modelling, 2010. **29**(1): p. 72-81.

5. Liu, G., et al., *Application of 3D-QSAR, pharmacophore, and molecular docking in the molecular design of diarylpyrimidine derivatives as HIV-1 nonnucleoside reverse transcriptase inhibitors.* International journal of molecular sciences, 2018. **19**(5): p. 1436.

6. Ehrlich, P., *Über den jetzigen Stand der Chemotherapie.* Berichte der deutschen chemischen Gesellschaft, 1909. **42**(1): p. 17-47.

7. Wermuth, C.-G., et al., *Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1997).* Annual reports in medicinal chemistry, 1998. **33**: p. 385-395.

8. Gund, P., *Three-dimensional pharmacophoric pattern searching*, in *Progress in molecular and subcellular biology*. 1977, Springer. p. 117-143.

9. Yang, S.-Y., *Pharmacophore modeling and applications in drug discovery: challenges and recent advances.* Drug discovery today, 2010. **15**(11-12): p. 444-450.

10.   Martin, Y.C., *DISCO: what we did right and what we missed.* Pharmacophore perception, development, and use in drug design, 2000. **2**: p. 49.

11.   Jones, G., P. Willet, and R. Glen, *GASP: genetic algorithm superimposition program.* Pharmacophore perception, development, and use in drug design, 2000: p. 85-106.

12.   GALAHAD™ is distributed by Tripos Inc., S.H.R., St. Louis, MO 63144, USA (http://www.tripos.com).

13.   Barnum, D., et al., *Identification of common functional configurations among molecules.* Journal of chemical information and computer sciences, 1996. **36**(3): p. 563-571.

14.   Li, H., J. Sutter, and R. Hoffmann, *HypoGen: an automated system for generating 3D predictive pharmacophore models.* Pharmacophore perception, development, and use in drug design, 2000. **2**: p. 171.

15.   Schneidman-Duhovny, D., et al., *PharmaGist: a webserver for ligand-based pharmacophore detection.* Nucleic acids research, 2008. **36**(suppl_2): p. W223-W228.

16.   Brown, B.P., J. Mendenhall, and J. Meiler, *BCL:: MolAlign: Three-Dimensional Small Molecule Alignment for Pharmacophore Mapping.* Journal of chemical information and modeling, 2019. **59**(2): p. 689-701.

17.   Dixon, S.L., et al., *PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results.* Journal of computer-aided molecular design, 2006. **20**(10): p. 647-671.

18.   Ortuso, F., T. Langer, and S. Alcaro, *GBPM: GRID-based pharmacophore model: concept and application studies to protein–protein recognition.* Bioinformatics, 2006. **22**(12): p. 1449-1455.

19.   Wolber, G. and T. Langer, *LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters.* Journal of chemical information and modeling, 2005. **45**(1): p. 160-169.

20.   Sunseri, J. and D.R. Koes, *Pharmit: interactive exploration of chemical space.* Nucleic Acids Research, 2016. **44**(W1): p. W442-W448.

21.   Chen, J. and L. Lai, *Pocket v. 2: further developments on receptor-based pharmacophore modeling.* Journal of chemical information and modeling, 2006. **46**(6): p. 2684-2691.

22.   Chen, J., et al., *Protein-protein interface analysis and hot spots identification for chemical ligand design.* Current pharmaceutical design, 2014. **20**(8): p. 1192-1200.

23.   Schaller, D., S. Pach, and G. Wolber, *PyRod: Tracing water molecules in molecular dynamics simulations.* Journal of chemical information and modeling, 2019. **59**(6): p. 2818-2829.

24.   Koes, D.R. and C.J. Camacho, *ZINCPharmer: pharmacophore search of the ZINC database.* Nucleic Acids Research, 2012. **40**(W1): p. W409-W414.

25.   Schaller, D., et al., *Next generation 3D pharmacophore modeling.* Wiley Interdisciplinary Reviews: Computational Molecular Science, 2020. **10**(4): p. e1468.

26.   Xu, Y., et al., *CavityPlus: a web server for protein cavity detection with pharmacophore modelling, allosteric site identification and covalent ligand binding ability prediction.* Nucleic acids research, 2018. **46**(W1): p. W374-W379.

27.   Goodford, P.J., *A computational procedure for determining energetically favorable binding sites on biologically important macromolecules.* Journal of medicinal chemistry, 1985. **28**(7): p. 849-857.

28.   Barillari, C., G. Marcou, and D. Rognan, *Hot-spots-guided receptor-based pharmacophores (HS-Pharm): a knowledge-based approach to identify ligand-anchoring atoms in protein cavities and prioritize structure-based pharmacophores.* Journal of chemical information and modeling, 2008. **48**(7): p. 1396-1410.

29.   Tran-Nguyen, V.-K., et al., *All in one: Cavity detection, druggability estimate, cavity-based pharmacophore perception, and virtual screening.* Journal of chemical information and modeling, 2018. **59**(1): p. 573-585.

30.   Verdonk, M.L., J.C. Cole, and R. Taylor, *SuperStar: a knowledge-based approach for identifying interaction sites in proteins.* Journal of molecular biology, 1999. **289**(4): p. 1093-1108.

31.   Yu, W., et al., *Pharmacophore modeling using site-identification by ligand competitive saturation (SILCS) with multiple probe molecules.* Journal of chemical information and modeling, 2015. **55**(2): p. 407-420.

32.   Yu, W., et al., *Site-identification by ligand competitive saturation (SILCS) assisted pharmacophore modeling.* Journal of computer-aided molecular design, 2014. **28**(5): p. 491-507.

33.   Lee, J.Y., et al., *Pharmmaker: Pharmacophore modeling and hit identification based on druggability simulations.* Protein Science, 2020. **29**(1): p. 76-86.

34.   Bakan, A., et al., *Druggability assessment of allosteric proteins by dynamics simulations in the presence of probe molecules.* Journal of chemical theory and computation, 2012. **8**(7): p. 2435-2447.

35.   Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics.* Journal of molecular graphics, 1996. **14**(1): p. 33-38.

36. Zhou, Q.-Y., J. Park, and V. Koltun, *Open3D: A modern library for 3D data processing.* arXiv preprint arXiv:1801.09847, 2018.

37. Zhou, Q.-Y., J. Park, and V. Koltun. *Fast global registration*. in *European conference on computer vision*. 2016. Springer.

38. Park, J., Q.-Y. Zhou, and V. Koltun. *Colored point cloud registration revisited*. in *Proceedings of the IEEE international conference on computer vision*. 2017.

39. Besl, P.J. and N.D. McKay. *Method for registration of 3-D shapes*. in *Sensor fusion IV: control paradigms and data structures*. 1992. International Society for Optics and Photonics.

40. Berman, H.M., et al., *The Protein Data Bank.* Nucleic Acids Research, 2000. **28**(1): p. 235-242.

41. Raschka, S., *BioPandas: Working with molecular structures in pandas DataFrames.* Journal of Open Source Software, 2017. **2**(14): p. 279.

42. Rusu, R.B., N. Blodow, and M. Beetz. *Fast point feature histograms (FPFH) for 3D registration*. in *2009 IEEE international conference on robotics and automation*. 2009. IEEE.

43. Fischler, M.A. and R.C. Bolles, *Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography.* Communications of the ACM, 1981. **24**(6): p. 381-395.

44. Mysinger, M.M., et al., *Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking.* Journal of medicinal chemistry, 2012. **55**(14): p. 6582-6594.

45.   Hunter, J.D., *Matplotlib: A 2D graphics environment.* Computing in science & engineering, 2007. **9**(03): p. 90-95.

46.   Van Der Walt, S., S.C. Colbert, and G. Varoquaux, *The NumPy array: a structure for efficient numerical computation.* Computing in science & engineering, 2011. **13**(2): p. 22-30.

47.   Release, S., *4: Glide.* Schrödinger, LLC, New York, NY, 2018.

48.   Yeung, C.M., et al., *Oximinoarylsulfonamides as potent HIV protease inhibitors.* Bioorganic & medicinal chemistry letters, 2005. **15**(9): p. 2275-2278.

49.   Miller, J.F., et al., *Ultra-potent P1 modified arylsulfonamide HIV protease inhibitors: the discovery of GW0385.* Bioorganic & medicinal chemistry letters, 2006. **16**(7): p. 1788-1794.

50.   Lim, S.P., et al., *Potent Allosteric Dengue Virus NS5 Polymerase Inhibitors: Mechanism of Action and Resistance Profiling.* Plos Pathogens, 2016. **12**(8).

51.   Kwiatkowski, N., et al., *Small-molecule kinase inhibitors provide insight into Mps1 cell cycle function.* Nature Chemical Biology, 2010. **6**(5): p. 359-368.

52.   Braffman, N.R., et al., *Structural mechanism of transcription inhibition by lasso peptides microcin J25 and capistruin.* Proceedings of the National Academy of Sciences of the United States of America, 2019. **116**(4): p. 1273-1278.

53.   Pan, P.W., et al., *Structure and Biochemical Functions of SIRT6.* Journal of Biological Chemistry, 2011. **286**(16): p. 14575-14587.

54.   You, W.J., et al., *Structural basis for the activation and inhibition of Sirtuin 6 by quercetin and its derivatives.* Scientific Reports, 2019. **9**.

55. Wolber, G., et al., *Molecule-pharmacophore superpositioning and pattern matching in computational drug design.* Drug discovery today, 2008. **13**(1-2): p. 23-29.

56. Richmond, N.J., et al., *GALAHAD: 1. Pharmacophore identification by hypermolecular alignment of ligands in 3D.* Journal of computer-aided molecular design, 2006. **20**(9): p. 567-587.

57. Douguet, D. and F. Payan, *SENSAAS (SENsitive Surface As A Shape): utilizing open-source algorithms for 3D point cloud alignment of molecules.* arXiv preprint arXiv:1908.11267, 2019.

58. Eguida, M. and D. Rognan, *A computer vision approach to align and compare protein cavities: application to fragment-based drug design.* Journal of Medicinal Chemistry, 2020. **63**(13): p. 7127-7142.

59. Bertolazzi, P., G. Liuzzi, and C. Guerra. *A global optimization algorithm for protein surface alignment.* in *2009 IEEE International Conference on Bioinformatics and Biomedicine Workshop.* 2009. IEEE.

60. Chen, J. and B. Belaton. *An improved iterative closest point algorithm for rigid point registration.* in *International Conference on Machine Learning and Cybernetics.* 2014. Springer.

61. Kabsch, W., *Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants.* Journal of applied crystallography, 1993. **26**(6): p. 795-800.

62. Wolber, G., A.A. Dornhofer, and T. Langer, *Efficient overlay of small organic molecules using 3D pharmacophores.* Journal of computer-aided molecular design, 2006. **20**(12): p. 773-788.

63.   Agarwal, S., et al. *A hybrid [ICP and GA] image registration algorithm for depth images*. in *2014 International Conference on Smart Structures and Systems (ICSSS)*. 2014. IEEE.

64.   Torres, D. and F. Cuevas. *Three-dimensional Point-cloud Registration using a Genetic Algorithm and the Iterative Closest Point Algorithm*. in *IJCCI (ECTA-FCTA)*. 2011.

65.   Jones, G., P. Willett, and R.C. Glen, *A genetic algorithm for flexible molecular overlay and pharmacophore elucidation.* Journal of computer-aided molecular design, 1995. **9**(6): p. 532-549.

66.   Holland, J.H., *Genetic algorithms.* Scientific american, 1992. **267**(1): p. 66-73.

67.   Trott, O. and A.J. Olson, *AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading.* Journal of computational chemistry, 2010. **31**(2): p. 455-461.

68.   Sengupta, S., M.K. Chattopadhyay, and H.P. Grossart, *The multifaceted roles of antibiotics and antibiotic resistance in nature.* Frontiers in Microbiology, 2013. **4**.

69.   Donowitz, G.R. and G.L. Mandell, *Drug-Therapy - Beta-Lactam Antibiotics .1.* New England Journal of Medicine, 1988. **318**(7): p. 419-426.

70.   Stillerman, M. and S.H. Bernstein, *Streptococcal Pharyngitis Therapy.* American Journal of Diseases of Children, 1964. **107**(1): p. 35-&.

71.   Holten, K.B., *Appropriate prescribing of oral beta-lactam antibiotics.* American family physician, 2000. **62**(3).

72.   Elander, R., *Industrial production of β-lactam antibiotics.* Applied microbiology and biotechnology, 2003. **61**(5-6): p. 385-392.

73.   Rohde, H., et al., *Open-Source Genomic Analysis of Shiga-Toxin-Producing E. coli O104:H4.* New England Journal of Medicine, 2011. **365**(8): p. 718-724.

74.   Grad, Y.H., et al., *Genomic epidemiology of the Escherichia coli O104:H4 outbreaks in Europe, 2011 (vol 109, pg 3065, 2012).* Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(14): p. 5547-5547.

75.   Kong, K.F., L. Schneper, and K. Mathee, *Beta-lactam antibiotics: from antibiosis to resistance and bacteriology.* Apmis, 2010. **118**(1): p. 1-36.

76.   Lyu, J., et al., *Ultra-large library docking for discovering new chemotypes.* Nature, 2019. **566**(7743): p. 224-229.

77.   Bahar, I., C. Chennubhotla, and D. Tobi, *Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation.* Current Opinion in Structural Biology, 2007. **17**(6): p. 633-640.

78.   Mahasenan, K.V., et al., *Conformational dynamics in penicillin-binding protein 2a of methicillin-resistant Staphylococcus aureus, allosteric communication network and enablement of catalysis.* Journal of the American Chemical Society, 2017. **139**(5): p. 2102-2110.

79.   Mangoni, M., D. Roccatano, and A. Di Nola, *Docking of flexible ligands to flexible receptors in solution by molecular dynamics simulation.* Proteins: Structure, Function, and Bioinformatics, 1999. **35**(2): p. 153-162.

80.   Klebe, G., *Virtual ligand screening: strategies, perspectives and limitations.* Drug Discovery Today, 2006. **11**(13): p. 580-594.

81.   Adcock, S.A. and J.A. McCammon, *Molecular dynamics: survey of methods for simulating the activity of proteins.* Chemical reviews, 2006. **106**(5): p. 1589-1615.

82.   Trott, O. and A.J. Olson, *Software News and Update AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading.* Journal of Computational Chemistry, 2010. **31**(2): p. 455-461.

83.   Ahmed, S.A., et al., *Genomic comparison of Escherichia coli O104:H4 isolates from 2009 and 2011 reveals plasmid, and prophage heterogeneity, including shiga toxin encoding phage stx2.* PLoS One, 2012. **7**(11): p. e48228.

84.   Arnold, K., et al., *The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling.* Bioinformatics, 2006. **22**(2): p. 195-201.

85.   Stec, B., et al., *Structure of the wild-type TEM-1 beta-lactamase at 1.55 A and the mutant enzyme Ser70Ala at 2.1 A suggest the mode of noncovalent catalysis for the mutant enzyme.* Acta Crystallogr D Biol Crystallogr, 2005. **61**(Pt 8): p. 1072-9.

86.   Cahill, S.T., et al., *Cyclic Boronates Inhibit All Classes of beta-Lactamases.* Antimicrobial Agents and Chemotherapy, 2017. **61**(4).

87.   Lee, W., et al., *A 1.2-A snapshot of the final step of bacterial cell wall biosynthesis.* Proc Natl Acad Sci U S A, 2001. **98**(4): p. 1427-31.

88.   Isgro, J.T., et al., *Analysis. Specific Heat*, in *NAMD Tutorial.* 2012. p. 50-53.

89.   Vanommeslaeghe, K., et al., *CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields.* Journal of computational chemistry, 2010. **31**(4): p. 671-690.

90.   Yu, W., et al., *Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations.* Journal of computational chemistry, 2012. **33**(31): p. 2451-2468.

91.   Lipinski, C.A., *Rule of five in 2015 and beyond: Target and ligand structural limitations, ligand chemistry structure and drug discovery project decisions.* Advanced Drug Delivery Reviews, 2016. **101**: p. 34-41.

92.   Kahlmeter, G., et al., *European Committee on Antimicrobial Susceptibility Testing (EUCAST) Technical Notes on antimicrobial susceptibility testing.* Clinical Microbiology and Infection, 2006. **12**(6): p. 501-503.

93.   Wiegand, I., K. Hilpert, and R.E.W. Hancock, *Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances.* Nature Protocols, 2008. **3**(2): p. 163-175.

94.   Comitani, F. and F.L. Gervasio, *Exploring Cryptic Pockets Formation in Targets of Pharmaceutical Interest with SWISH.* Journal of Chemical Theory and Computation, 2018. **14**(6): p. 3321-3331.

95.   Galdadas, I., et al., *Defining the architecture of KPC-2 Carbapenemase: identifying allosteric networks to fight antibiotics resistance.* Scientific Reports, 2018. **8**.

96.   Fonze, E., et al., *TEM1 beta-lactamase structure solved by molecular replacement and refined structure of the S235A mutant.* Acta Crystallogr D Biol Crystallogr, 1995. **51**(Pt 5): p. 682-94.

97.   Strynadka, N.C., et al., *Molecular docking programs successfully predict the binding of a beta-lactamase inhibitory protein to TEM-1 beta-lactamase.* Nat Struct Biol, 1996. **3**(3): p. 233-9.

98.   Nikaido, H., *Molecular basis of bacterial outer membrane permeability revisited.* Microbiology and Molecular Biology Reviews, 2003. **67**(4): p. 593-+.

99. Acosta-Gutierrez, S., et al., *Getting Drugs into Gram-Negative Bacteria: Rational Rules for Permeation through General Porins.* Acs Infectious Diseases, 2018. **4**(10): p. 1487-1498.

100. Bajaj, H., et al., *Bacterial Outer Membrane Porins as Electrostatic Nanosieves: Exploring Transport Rules of Small Polar Molecules.* Acs Nano, 2017. **11**(6): p. 5465-5473.

101. Bajaj, H., et al., *Molecular Basis of Filtering Carbapenems by Porins from beta-Lactam-resistant Clinical Strains of Escherichia coli.* Journal of Biological Chemistry, 2016. **291**(6): p. 2837-2847.

102. Richter, M.F., et al., *Predictive compound accumulation rules yield a broad - spectrum antibiotic.* Nature, 2017. **545**(7654): p. 299-+.

103. Weston, G.S., et al., *Structure-based enhancement of boronic acid-based inhibitors of AmpC beta-lactamase.* J Med Chem, 1998. **41**(23): p. 4577-86.

104. Song, W., et al., *Detection of extended-spectrum beta-lactamases by using boronic acid as an AmpC beta-lactamase inhibitor in clinical isolates of Klebsiella spp. and Escherichia coli.* J Clin Microbiol, 2007. **45**(4): p. 1180-4.

105. Rojas, L.J., et al., *Boronic Acid Transition State Inhibitors Active against KPC and Other Class A beta-Lactamases: Structure-Activity Relationships as a Guide to Inhibitor Design.* Antimicrobial Agents and Chemotherapy, 2016. **60**(3): p. 1751-1759.

106. Markham, A., *Tavaborole: First Global Approval.* Drugs, 2014. **74**(13): p. 1555-1558.

107. Lomovskaya, O., et al., *Vaborbactam: Spectrum of Beta-Lactamase Inhibition and Impact of Resistance Mechanisms on Activity in Enterobacteriaceae.* Antimicrobial Agents and Chemotherapy, 2017. **61**(11).

108. Taylor, N.P. *VenatoRx raises $42M to take drug for breaking bacterial resistance to approval*. 2017 [cited 2018; Available from: https://www.fiercebiotech.com/biotech/venatorx-raises-42m-to-take-drug-for-breaking-bacterial-resistance-to-approval.

109. Greenway, G., *Economic impact of zebra chip control costs on grower returns in seven US states.* Am. Potato J., 2014. **91**(6): p. 714-719.

110. Hodges, A.W. and T.H. Spreen *Economic impacts of citrus greening (HLB) in Florida, 2006/07–2010/111. The Institute of Food and Agricultural Sciences*. 2012.

111. Garnier, M., et al., *Genomic characterization of a liberibacter present in an ornamental rutaceous tree, Calodendrum capense, in the Western Cape province of South Africa. Proposal of 'Candidatus Liberibacter africanus subsp capensis'.* Int. J. Syst. Evol., 2000. **50**: p. 2119-2125.

112. Aubert, B. and J. Bové. *Effect of penicillin or tetracycline injections of citrus trees affected by greening disease under field conditions in Réunion Island*. in *International Organization of Citrus Virologists Conference Proceedings (1957-2010)*. 1980.

113. Martinez, A., *Response of the etiologic agent of citrus greening disease in the Philippines to treatment with broad spectrum antibiotics.* Philippine Phytopathology, 1975. **11**: p. 58-61.

114. Martinez, A., D. Nora, and A. Armedilla, *Suppression of symptoms of citrus greening disease in the Philippines by treatment with tetracycline antibiotics.* Plant Dis Rep, 1970. **54**: p. 1007-9.

115. National Academies of Sciences, E. and Medicine, *A review of the citrus greening research and development efforts supported by the Citrus Research and Development Foundation: Fighting a ravaging disease*. 2018: National Academies Press.

116. Kumar, K., et al., *Antibiotic use in agriculture and its impact on the terrestrial environment.* Adv. Agron., 2005. **87**: p. 1-54.

117. Editorial, *Spraying diseased citrus orchards with antibiotics could backfire*, in *Nature*. 2019. p. 283.

118. Serrano, M., E. Kombrink, and C. Meesters, *Considerations for designing chemical screening strategies in plant biology.* Front. Plant. Sci., 2015. **6**(131).

119. Zabotina, O., et al., *Identification and preliminary characterization of a new chemical affecting glucosyltransferase activities involved in plant cell wall biosynthesis.* Mol. Plant, 2008. **1**(6): p. 977-989.

120. Yoshitani, N., et al., *A structure-based strategy for discovery of small ligands binding to functionally unknown proteins: Combination of in silico screening and surface plasmon resonance measurements.* Proteomics, 2005. **5**(6): p. 1472-1480.

121. Prasad, S., et al., *SEC-translocon dependent extracytoplasmic proteins of Candidatus Liberibacter asiaticus.* Front. Microbiol., 2016. **7**.

122. Trott, O. and A.J. Olson, *Software news and update AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading.* J. Comput. Chem., 2010. **31**(2): p. 455-461.

123. Humphrey, W., A. Dalke, and K. Schulten, *VMD: Visual molecular dynamics.* J. Mol. Graph. Model., 1996. **14**(1): p. 33-38.

124. Bakan, A., et al., *Druggability assessment of allosteric proteins by dynamics simulations in the presence of probe molecules.* J. Chem. Theory Comput., 2012. **8**(7): p. 2435-2447.

125. Fernando, S. and T. Fernando, *Antivirals for allosteric inhibition of Zika virus using a homology model and experimentally determined structure of envelope protein.* BMC Res. Notes, 2017. **10**(1): p. 354.

126. Fernando, S., et al., *An approach for Zika virus inhibition using homology structure of the envelope protein.* Mol. Biotechnol., 2016. **58**(12): p. 801-806.

127. Phillips, J.C., et al., *Scalable molecular dynamics with NAMD.* J. Comput. Chem, 2005. **26**(16): p. 1781-802.

128. Koes, D.R. and C.J. Camacho, *ZINCPharmer: pharmacophore search of the ZINC database.* Nucleic Acids Res., 2012. **40**(W1): p. W409-W414.

129. Lipinski, C.A., *Lead-and drug-like compounds: the rule-of-five revolution.* Drug Discov. Today Technol., 2004. **1**(4): p. 337-341.

130. Release, S., *4: Desmond molecular dynamics system.* DE Shaw Research: New York, NY, 2017.

131. Wizard, P.P., *Epik version 2.8.* Schrödinger, LLC, New York, NY, 2014.

132. Release, S., *1: Maestro.* Schrödinger, LLC, New York, NY, 2017. **2017**.

133. Roos, K., et al., *OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules.* Journal of Chemical Theory and Computation, 2019. **15**(3): p. 1863-1874.

134. Ciesielski, G.L., V.P. Hytonen, and L.S. Kaguni, *Biolayer interferometry: A novel method to elucidate protein-protein and protein-DNA interactions in the mitochondrial DNA replisome*, in *Mitochondrial DNA: Methods and Protocols, 3rd Edition*, M. McKenzie, Editor. 2016. p. 223-231.

135.   Raththagala, M., et al., *Structural Mechanism of Laforin Function in Glycogen Dephosphorylation and Lafora Disease.* Molecular Cell, 2015. **57**(2): p. 261-272.

136.   Remmert, M., et al., *HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment.* Nature Methods, 2012. **9**(2): p. 173-175.

137.   Berman, H.M., et al., *The protein data bank.* Nucleic Acids Res., 2000. **28**(1): p. 235-242.

138.   Lipinski, C.A., *Lead- and drug-like compounds: the rule-of-five revolution.* Drug Discov Today Technol, 2004. **1**(4): p. 337-41.

139.   Shah, N.B. and T.M. Duncan, *Bio-layer interferometry for measuring kinetics of protein-protein interactions and allosteric ligand effects.* J Vis. Exp., 2014(84): p. e51383.

140.   Duan, Y., et al., *Complete genome sequence of citrus huanglongbing bacterium,'Candidatus Liberibacter asiaticus' obtained through metagenomics.* Molecular Plant-Microbe Interactions, 2009. **22**(8): p. 1011-1020.

141.   Hodges, A. and T. Spreen, *Economic impacts of citrus greening (HLB) in Florida, 2006/07-2010/11.* Electronic Data Information Source (EDIS) FE903, 2012: p. 32611.

142.   USDA. *Citrus greening threatens America's citrus. Don't risk citrus, don't move citrus.*; Available from: https://www.aphis.usda.gov/aphis/resources/pests-diseases/hungry-pests/the-threat/citrus-greening/citrus-greening-hp.

143.   Blaustein, R.A., G.L. Lorca, and M. Teplitski, *Challenges for managing Candidatus Liberibacter spp.(huanglongbing disease pathogen): Current control measures and future directions.* Phytopathology, 2018. **108**(4): p. 424-435.

144.   Li, S., et al., *Citrus Greening: Management Strategies and Their Economic Impact.* HortScience, 2020. **1**(aop): p. 1-9.

145. (EPA), E.P.A., *Final Registration Decision for the New Use of the Active Ingredient Oxytetracycline Hydrochloride on Citrus Crop Group 10-10*. 2018.

146. (EPA), E.P.A., *Pesticide Emergency Exemptions: Agency Decisions and State and Federal Agency Crisis Declarations*. 2020.

147. Manyi-Loh, C., et al., *Antibiotic Use in Agriculture and Its Consequential Resistance in Environmental Sources: Potential Public Health Implications.* Molecules, 2018. **23**(4).

148. Dall, C. *Lawmakers urge EPA to rethink use of antibiotics on citrus trees*. 2019; Available from: https://www.cidrap.umn.edu/news-perspective/2019/08/lawmakers-urge-epa-rethink-use-antibiotics-citrus-trees.

149. diversity, C.f.b. *Trump Administration Approves Antibiotic Residue on Citrus Fruit*. 2018; Medically Important Antibiotic Oxytetracycline Allowed on Oranges, Tangerines]. Available from: https://www.biologicaldiversity.org/news/press_releases/2018/oxytetracycline-12-06-2018.php.

150. diversity, C.f.b. *EPA Extends 'Emergency' Antibiotic Use on California, Florida Citrus Crops*. 2020; Approval of Streptomycin as Pesticide Sidesteps Full Safety Assessment]. Available from: https://biologicaldiversity.org/w/news/press-releases/epa-extends-emergency-antibiotic-use-on-california-florida-citrus-crops-2020-02-11/.

151. Med, N.A.S.E., *Review of the Citrus Greening Research and Development Efforts Supported by the Citrus Research and Development Foundation: Fighting a Ravaging Disease.* Review of the Citrus Greening Research and Development Efforts Supported by the Citrus Research and Development Foundation: Fighting a Ravaging Disease, 2018: p. 1-269.

152.   Wang, N., et al., *The Candidatus Liberibacter-Host Interface: Insights into Pathogenesis Mechanisms and Disease Control.* Annual Review of Phytopathology, Vol 55, 2017. **55**: p. 451-482.

153.   Costa, T.R., et al., *Secretion systems in Gram-negative bacteria: structural and mechanistic insights.* Nature Reviews Microbiology, 2015. **13**(6): p. 343-359.

154.   Morgan, J.L.W., J.F. Acheson, and J. Zimmer, *Structure of a Type-1 Secretion System ABC Transporter.* Structure, 2017. **25**(3): p. 522-529.

155.   Sugio, A., et al., *Diverse Targets of Phytoplasma Effectors: From Plant Development to Defense Against Insects.* Annual Review of Phytopathology, Vol 49, 2011. **49**: p. 175-195.

156.   Wang, N., *The Citrus Huanglongbing Crisis and Potential Solutions.* Molecular Plant, 2019. **12**(5): p. 607-609.

157.   Clark, K., et al., *An effector from the Huanglongbing-associated pathogen targets citrus proteases.* Nature Communications, 2018. **9**.

158.   Pang, Z.Q., et al., *Citrus CsACD2 Is a Target of Candidatus Liberibacter Asiaticus in Huanglongbing Disease.* Plant Physiology, 2020. **184**(2): p. 792-805.

159.   Lei, J., et al., *The antimicrobial peptides and their potential clinical applications.* American journal of translational research, 2019. **11**(7): p. 3919.

160.   Mannoor, M.S., et al., *Electrical detection of pathogenic bacteria via immobilized antimicrobial peptides.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(45): p. 19207-19212.

161.   Nawrot, R., et al., *Plant antimicrobial peptides.* Folia Microbiologica, 2014. **59**(3): p. 181-196.

162. Guerra-Lupian, M.A., et al., *Localized expression of antimicrobial proteins mitigates huanglongbing symptoms in Mexican lime.* Journal of Biotechnology, 2018. **285**: p. 74-83.

163. Guzman, J.V., et al., *Liposome delivery system of antimicrobial peptides against Huanglongbing.* Phytopathology, 2018. **108**(10).

164. Irigoyen, S., et al., *Plant hairy roots enable high throughput identification of antimicrobials against Candidatus Liberibacter spp.* Nature Communications, 2020. **11**(1).

165. Lin, H., et al., *Complete Genome Sequence of "Candidatus Liberibacter africanus," a Bacterium Associated with Citrus Huanglongbing.* Microbiology Resource Announcements, 2015. **3**(4).

166. Enguene, Y.V.N., et al., *Xenon for tunnelling analysis of the efflux pump component OprN.* Plos One, 2017. **12**(9).

167. Laskowski, R.A., et al., *Procheck - a Program to Check the Stereochemical Quality of Protein Structures.* Journal of Applied Crystallography, 1993. **26**: p. 283-291.

168. Wang, G.S., X. Li, and Z. Wang, *APD3: the antimicrobial peptide database as a tool for research and education.* Nucleic Acids Research, 2016. **44**(D1): p. D1087-D1093.

169. Srinivasan, J., et al., *Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate - DNA helices.* Journal of the American Chemical Society, 1998. **120**(37): p. 9401-9409.

170. Humphrey, W., A. Dalke, and K. Schulten, *VMD: Visual molecular dynamics.* Journal of Molecular Graphics & Modelling, 1996. **14**(1): p. 33-38.

171.   McGibbon, R.T., et al., *MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories.* Biophysical Journal, 2015. **109**(8): p. 1528-1532.

172.   Williams, C.J., et al., *MolProbity: More and better reference data for improved all-atom structure validation.* Protein Science, 2018. **27**(1): p. 293-315.

173.   Lei, H.T., et al., *Crystal Structure of the Open State of the Neisseria gonorrhoeae MtrE Outer Membrane Channel.* Plos One, 2014. **9**(6).

174.   Kulathila, R., et al., *Crystal Structure of Escherichia coli CusC, the Outer Membrane Component of a Heavy Metal Efflux Pump.* Plos One, 2011. **6**(1).

175.   Simmaco, M., et al., *Temporins, antimicrobial peptides from the European red frog Rana temporaria.* European Journal of Biochemistry, 1996. **242**(3): p. 788-792.

176.   Ikeda, T., et al., *2 Novel Tachykinin-Related Neuropeptides in the Echiuroid Worm, Urechis-Unicinctus.* Biochemical and Biophysical Research Communications, 1993. **192**(1): p. 1-6.

177.   Koyama, Y., *A new antibiotic'colistin'produced by spore-forming soil bacteria.* J. Antibiot., 1950. **3**: p. 457-458.

178.   Nation, R.L. and J. Li, *Colistin in the 21st century.* Current Opinion in Infectious Diseases, 2009. **22**(6): p. 535-543.

179.   Zhao, S.M., et al., *Purification and Characterization of Plantaricin JLA-9: A Novel Bacteriocin against Bacillus spp. Produced by Lactobacillus plantarum JLA-9 from Suan-Tsai, a Traditional Chinese Fermented Cabbage.* Journal of Agricultural and Food Chemistry, 2016. **64**(13): p. 2754-2764.

180.   Mishra, B., et al., *Design and surface immobilization of short anti-biofilm peptides.* Acta Biomaterialia, 2017. **49**: p. 316-328.

181. Zarena, D., et al., *The pi Configuration of the WWW Motif of a Short Trp-Rich Peptide Is Critical for Targeting Bacterial Membranes, Disrupting Preformed Biofilms, and Killing Methicillin-Resistant Staphylococcus aureus.* Biochemistry, 2017. **56**(31): p. 4039-4043.

182. Imai, Y., et al., *A new antibiotic selectively kills Gram-negative pathogens.* Nature, 2019. **576**(7787): p. 459-+.

183. Hart, E.M., et al., *A small-molecule inhibitor of BamA impervious to efflux and the outer membrane permeability barrier.* Proceedings of the National Academy of Sciences of the United States of America, 2019. **116**(43): p. 21748-21757.

184. Craik, D.J., et al., *The Future of Peptide-based Drugs.* Chemical Biology & Drug Design, 2013. **81**(1): p. 136-147.

185. Bruno, B.J., G.D. Miller, and C.S. Lim, *Basics and recent advances in peptide and protein drug delivery.* Therapeutic Delivery, 2013. **4**(11): p. 1443-1467.

Supplementary tables

S. Table 1 Definitions of pharmacophores in ELIXIR-A

Pharmacophore types	Abbreviation	Presented colors	Default radius (Å)
Hydrogen Donor	HDR	Silver	0.5
Hydrogen Acceptor	HAC	Orange	0.5
Hydrophobic	HPB	Green	1.0
Aromatic	ARO	Purple	1.1
Positive Ion	PIO	Blue	0.75
Negative Ion	NIO	Red	0.75

S. Table 2 Record of the ELIXIR-A input and output pdb files. Most parts of this table used the standard ATOM record type from PDB database. (http://www.wwpdb.org/documentation/file-format-content/format33/sect9.html#ATOM)

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 – 6	Record name	"ATOM "	
7 – 11	Integer	Serial	Atom serial number
13 – 16	Atom	Name	Atom name
17	Character	altLoc	
18 – 20	Pharmacophore name	resName	Pharmacophore name
22	Character	chainID	
23 - 26	Integer	resSeq	
27	Achar	iCode	
31 – 38	Real (8. 3)	x	Orthogonal coordinates for x in Angstroms.
39 – 46	Real (8. 3)	y	Orthogonal coordinates for y in Angstroms.
47 – 54	Real (8. 3)	z	Orthogonal coordinates for z in Angstroms.
55 – 60	Real (6. 2)	radius	Pharmacophore sphere radius
61 – 66	Real (6. 2)	Tempfactor	
77 – 78	LString (2)	Element	
79 – 80	LString (2)	charge	

S. Table 3 Minimum Inhibitory Concentration of antibiotics and inhibitors

Rownum	1	2	3	4	5	6	7	8	9	10	11	12	
C(mg/L)	128	64	32	16	8	4	2	1		0		NC[a]	Comments
A	-	+	+	+	+	+	+	+		+		-	
B	-	+	+	+	+	+	+	+		+		-	Amoxicillin
C	-	+	+	+	+	+	+	+		+		-	
D	-	+	+	+	+	+	+	+		+		-	

Rownum	1	2	3	4	5	6	7	8	9	10	11	12	
C(mg/L)	128	64	32	16	8	4	2	1		0		NC[a]	Comments
A	-	-	-	-	-	-	+	+		+		-	
B	-	-	-	-	-	-	+	+		+		-	Amoxicillin-Boronic acid (2:1)
C	-	-	-	-	-	-	+	+		+		-	
D	-	-	-	-	-	-	+	+		+		-	
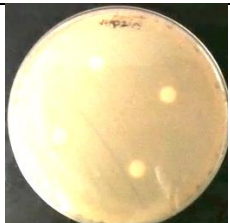
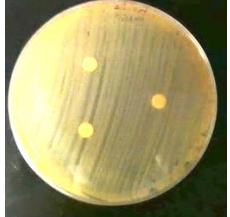| Rownum	1	2	3	4	5	6	7	8	9	10	11	12	
C(mg/L)	128	64	32	16	8	4	2	1		0		NC[a]	Comments
A	-	-	-	-	-	-	-	+		+		-	
B	-	-	-	-	-	-	-	+		+		-	Amoxicillin-Clavulanic acid (2:1)
C	-	-	-	-	-	-	-	+		+		-	
D	-	-	-	-	-	-	-	+		+		-	

a. NC: negative control with sterile LB media. PC: positive control with bacteria and LB media only.

S. Table 4 Minimum Inhibitory Concentration of screened ligands

Rownum	1	2	3	4	5	6	7	8	9	10	11	12	
C(mg/L)	128	64	32	16	8	4	2	1		0		NC[a]	Comments
A	-	-	-	-	-	-	-	+		+		-	
B	-	-	-	-	-	-	-	+		+		-	BACM
C	-	-	-	-	-	-	-	+		+		-	
D	-	-	-	-	-	-	-	+		+		-	
A	-	-	-	-	-	-	-	+		+		-	
B	-	-	-	-	-	-	-	+		+		-	Clavulanic acid
C	-	-	-	-	-	-	-	+		+		-	
D	-	-	-	-	-	-	-	+		+		-	

S. Table 5 Disk diffusion values of selected β-lactamase inhibitor - β-lactam combination with plate images. The disks with mark N were negative control.

Inhibitor	β-lactam/ non β-lactam	Disk content (µg)	Zone Diameter (mm)	Resisted/ susceptible	Image
--	Amoxicillin	20	0.0±0.0	R	
--	Kanamycin[a]	30	26.3±1.2	S	
--	Kanamycin	10	16.0±1.0	S	The disks with mark 3 were 30 µg dose. The blank disks were 10 µg dose.
Clavulanic acid	Amoxicillin	20-10	22.3±4.7	S	Four replicates were the blank disks.
ZINC13777605	Amoxicillin	20-10	0.0±0.0	R	Three replicates were the blank disks.
ZINC75280048	Amoxicillin	20-10	0.0±0.0	R	

ZINC337270	*Amoxicillin*	*20-10*	*0.0±0.0*	*R*		
ZINC3978503	*Amoxicillin*	*20-10*	*0.0±0.0*	*R*		
BAPE	*Amoxicillin*	*20-10*	*0.0±0.0*	*R*	*Three replicates were the disks with B marks.*	
BACM	*Amoxicillin*	*20-10*	*18.8±0.5*	*S*		

S. Table 6 High-affinity probe binding spots from MD simulation. The probes at active binding sites were bolded. Specific information about the probe molecule is described on the plugin website. (http://prody.csb.pitt.edu/tutorials/drugui_tutorial/probes.html)

Pharmacophore hotspots	Drug-like probes	x	y	z	Binding free energy (kcal/mol)
1	**ACETA**	**-1.5**	**-7.5**	**-14**	**-2.62**
2	IPAMA	3.5	0	10	-2.34
3	**ACETA**	**-6.5**	**-3.5**	**-13**	**-2.29**
4	ACAMA	-13.5	-10	2	-2.28
5	**ACETA**	**-3.5**	**-1.5**	**-14**	**-2.21**
6	IPROA	14	-1.5	-5.5	-2.21
7	IBUTA	-14	-3.5	10	-1.99
8	IPROA	-14	-5	-2	-1.98
9	IBUTA	-8.5	-11.5	7.5	-1.85
10	IPAMA	9	-1.5	-11	-1.8
11	IPROA	-4.5	11.5	5	-1.71
12	**ACETA**	**-4**	**-8**	**-13**	**-1.68**
13	**IPROA**	**0**	**-13**	**-12.5**	**-1.67**
14	IPROA	-4	7.5	-4	-1.6
15	IBUTA	-8.5	-6.5	11	-1.58
16	IPROA	-4.5	10.5	7.5	-1.58
17	**IBUTA**	**6**	**-8**	**5.5**	**-1.57**
18	IPROA	5.5	13.5	2.5	-1.56
19	**IPROA**	**-5.5**	**-11**	**-10**	**-1.55**
20	**IPROA**	**2**	**-14**	**-11**	**-1.47**
21	IPAMA	10	-3.5	-14	-1.47
22	IPROA	11.5	-1.5	-6.5	-1.45
23	IPAMA	22	0	-4	-1.44
24	IPAMA	-26	-3.5	2	-1.43
25	IBUTA	1.5	-12	-18.5	-1.41
26	IPROA	11	14	-2	-1.38
27	IPROA	9.5	-16	-12.5	-1.37
28	IPROA	-13	-5.5	11.5	-1.35
29	IPROA	-15	-0.5	10	-1.34
30	**IPROA**	**0**	**-14**	**2.5**	**-1.32**
31	**IPROA**	**11**	**-6**	**8**	**-1.31**
32	**ACETA**	**-0.5**	**-5.5**	**-15**	**-1.31**
33	**IPROA**	**-4.5**	**-11**	**10.5**	**-1.29**
34	ACETA	-6	-5.5	-14.5	-1.29

35	IPROA	22.5	-10.5	-10.5	-1.27
36	ACETA	-8.5	-6.5	-18.5	-1.27
37	ACETA	16.5	0.5	8	-1.25
38	IBUTA	-10	-8.5	10	-1.25
39	IPAMA	-10	5	-13.5	-1.24
40	**IPROA**	**2.5**	**-13**	**-1**	**-1.24**
41	IPROA	16.5	-1	-6	-1.23
42	IPROA	16	-9	3.5	-1.22
43	**IPROA**	**0**	**-13**	**-16.5**	**-1.19**
44	IPROA	-16.5	-5	-7.5	-1.19
45	IPROA	8	-17.5	-11	-1.19
46	IPROA	-1.5	13.5	4	-1.19
47	IPROA	17.5	-1.5	-10.5	-1.18
48	IPROA	-4	9.5	0.5	-1.18
49	**IPROA**	**-5.5**	**-14**	**-9**	**-1.17**
50	IPAMA	12	-4.5	-15.5	-1.17
51	IPROA	-12.5	17	1	-1.17
52	IPROA	18	11.5	-5	-1.17
53	IPROA	9.5	-14	-19	-1.17
54	IPROA	7	-11	-19.5	-1.14
55	**ACAMA**	**4**	**-14**	**-7**	**-1.14**
56	ACETA	-8.5	-10	-18	-1.13
57	**IPROA**	**4**	**-12**	**4**	**-1.1**
58	ACAMA	-12.5	-9	4	-1.1
59	IPROA	14.5	-0.5	-9.5	-1.1
60	IPROA	1.5	14	1.5	-1.08
61	IPROA	22	9	-1	-1.06
62	ACAMA	7.5	-4.5	7.5	-1.05
63	IPROA	18.5	-13.5	-5	-1.05
64	IPROA	10.5	-16	-17.5	-1.04
65	**IPROA**	**-8**	**-11.5**	**-12.5**	**-1.02**
66	IBUTA	-10.5	-9.5	7.5	-1.01
67	ACETA	5.5	14	5.5	-1.01

Supplementary figures


```
5I3P.json_source_out.pdb

ATOM      1  CA  HDR A    1      20.843  64.257  11.837  0.50  0.00
ATOM      2  CA  HDR A    2      20.434  57.245   2.717  0.50  0.00
ATOM      3  CA  HAC A    3      18.799  64.712  12.745  0.50  0.00
ATOM      5  CA  HAC A    5      18.116  58.963   9.883  0.50  0.00
ATOM      6  CA  HAC A    6      20.434  57.245   2.717  0.50  0.00
ATOM      7  CA  HPB A    7      18.684  61.044   6.924  1.00  0.00
ATOM      8  CA  HPB A    8      17.841  57.824  10.699  1.00  0.00
ATOM      9  CA  HPB A    9      19.041  58.772   3.728  1.00  0.00
```
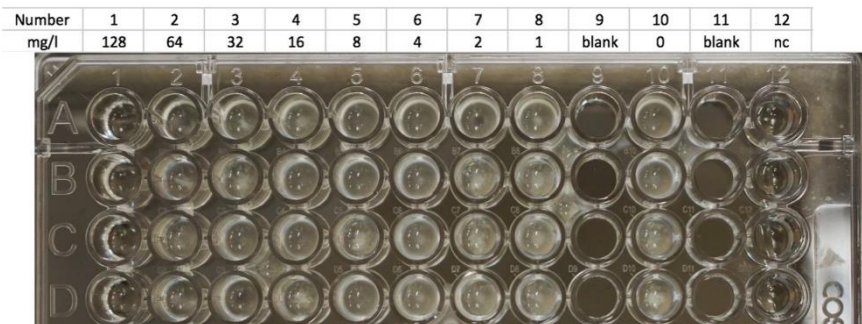
S. Figure 1 Example ELIXIR-A output file.
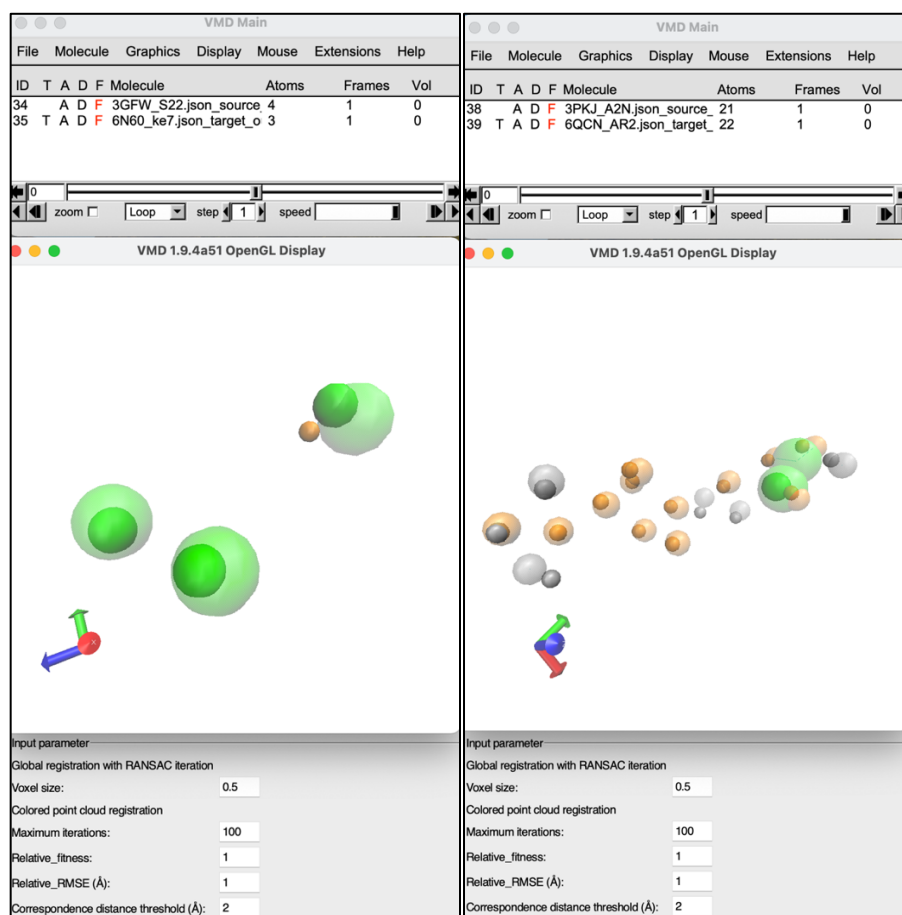
S. Figure 2 The pharmacophore alignment parameters and results for Dual specificity protein kinase TTK (left) and NAD-dependent deacetylase sirtuins (right).

a)

| Number	1	2	3	4	5	6	7	8	9	10	11	12
mg/l	128	64	32	16	8	4	2	1	blank	0	blank	nc

Amoxicillin

b)

Number	1	2	3	4	5	6	7	8	9	10	11	12
mg/l	128	64	32	16	8	4	2	1	blank	0	blank	nc

Amoxicillin : BACM
(1:0.5)

c)
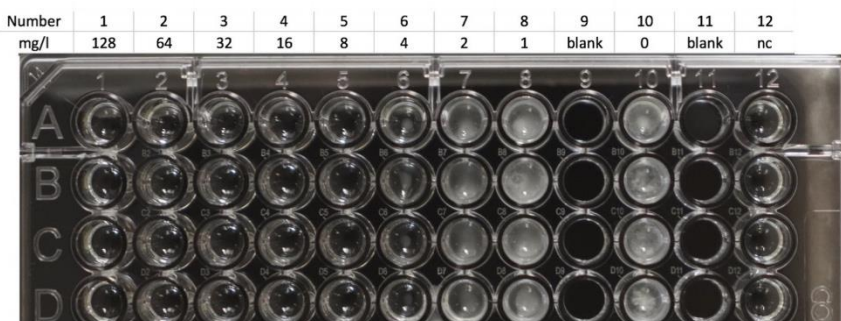
Number	1	2	3	4	5	6	7	8	9	10	11	12
mg/l	128	64	32	16	8	4	2	1	blank	0	blank	nc

Amoxicillin:Clavulanic acid
(1:0.5)

S. Figure 3 MIC results of screened ligands and antibiotics. Noted that the concentration in the figure was related to antibiotic Amoxicillin.
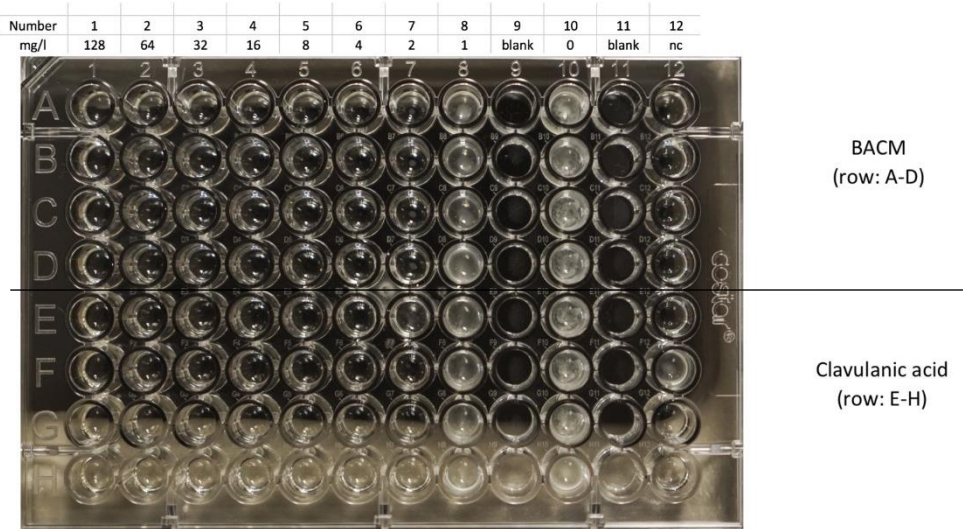
S. Figure 4 MIC results of screened ligands. Noted that the concentration in the figure was related to antibiotic Amoxicillin.
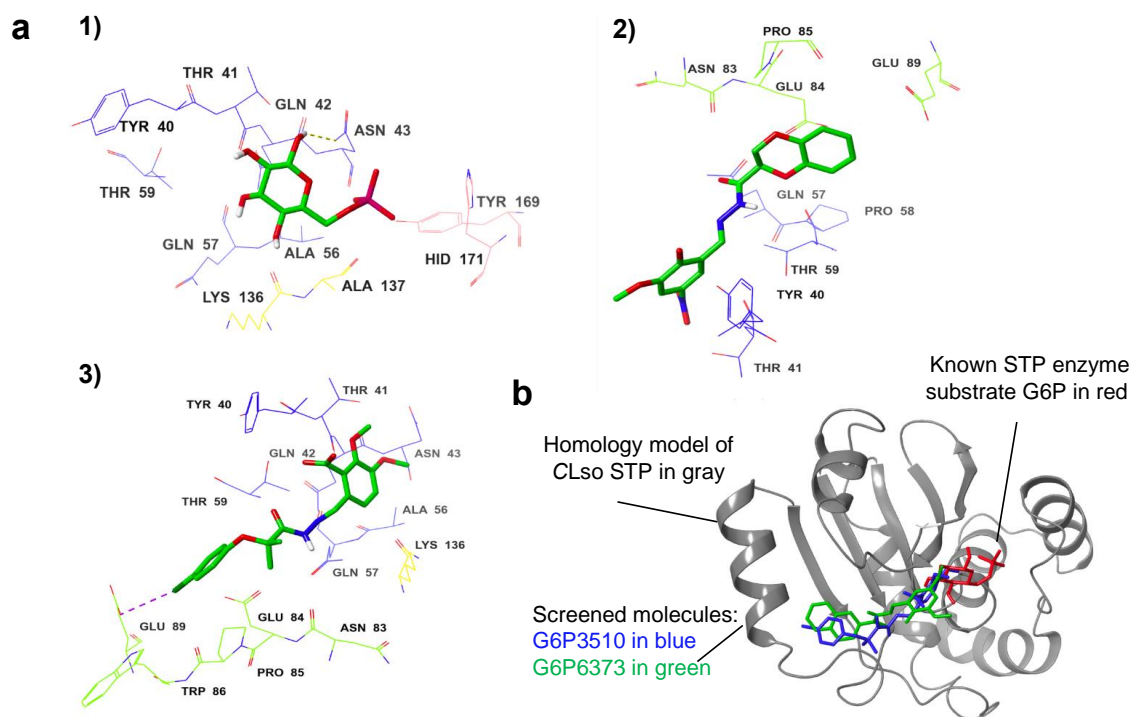
S. Figure 5 Identification of competitive inhibitors (G6P3510 and G6P6373) that bind to *CL*so STP. (a) The interactions of STP with G6P, G6P6373 and G6P3510, respectively. (b) G6P (red) occupying the active site and an isolated ligand (blue) occupying the same site as G6P on *CL*so STP (gray).
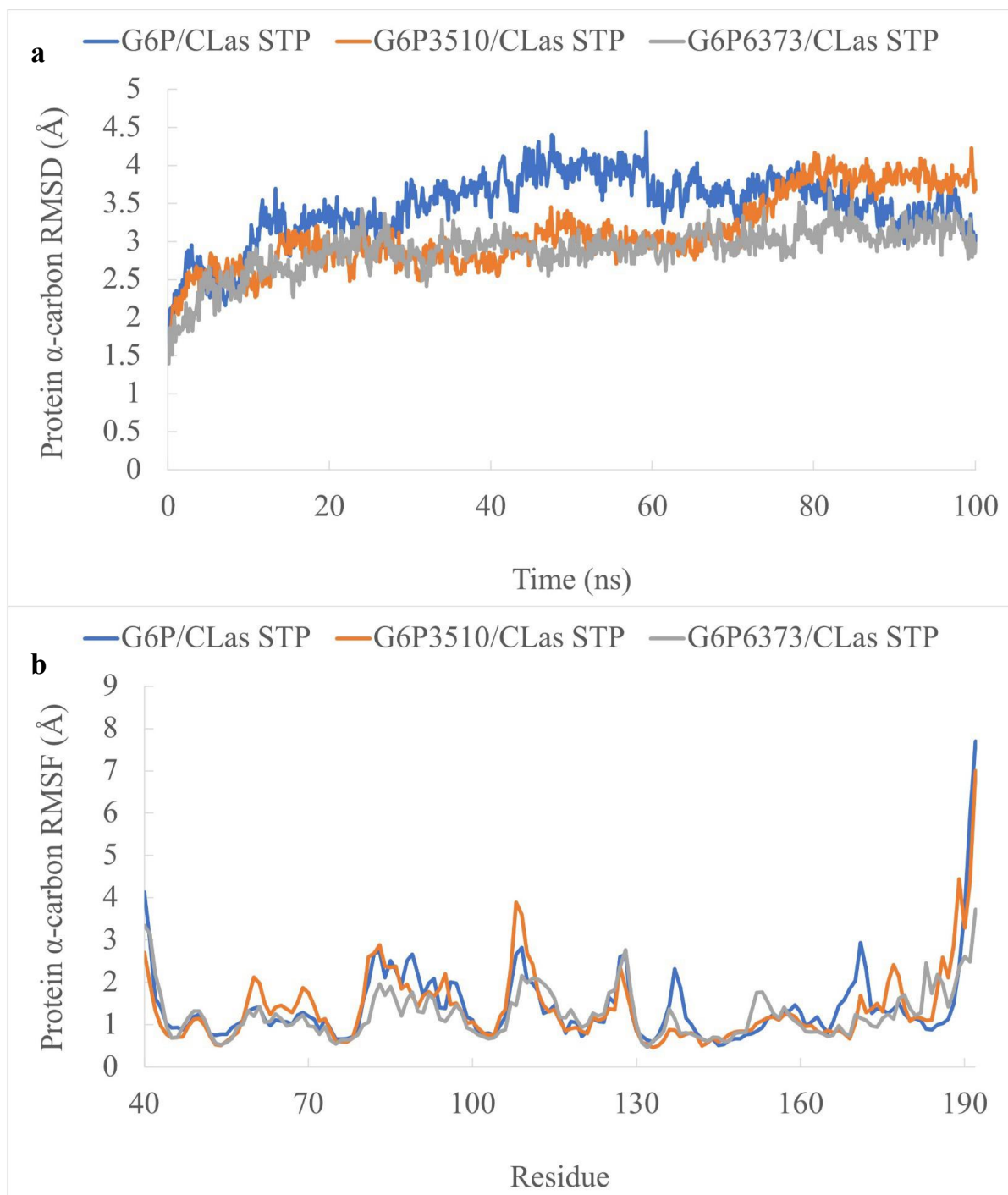
S. Figure 6 a) RMSD of protein α-carbon throughout the 100 ns MD simulations. b) RMSF of protein α-carbon residues from the average of 100 ns MD simulations of the complexes.
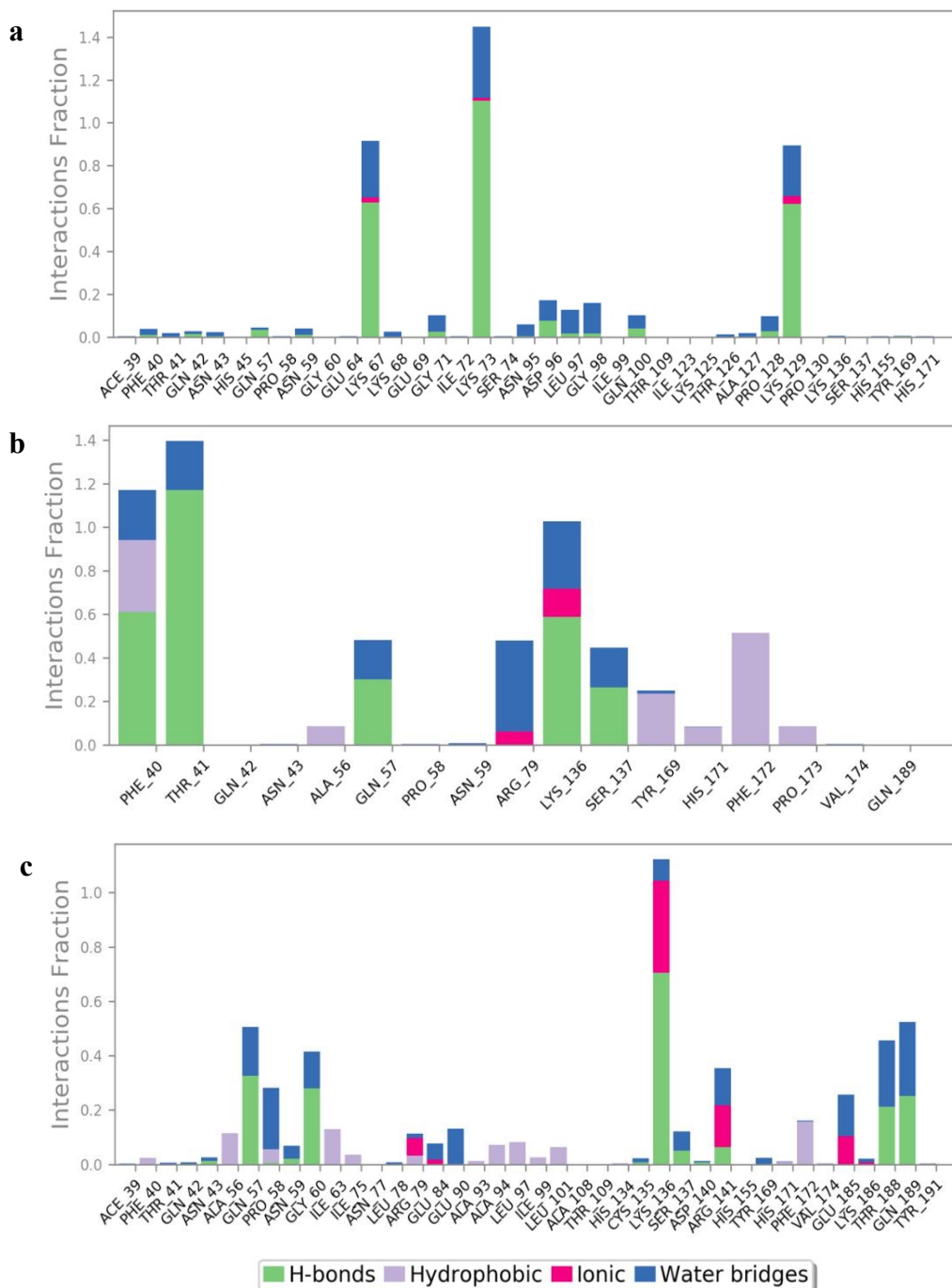
S. Figure 7 Protein-ligand interaction diagram extracted from the complex throughout the 100 ns simulation. a) G6P- *C*Las STP b) G6P3510- *C*Las STP c) G6P6373- *C*Las STP
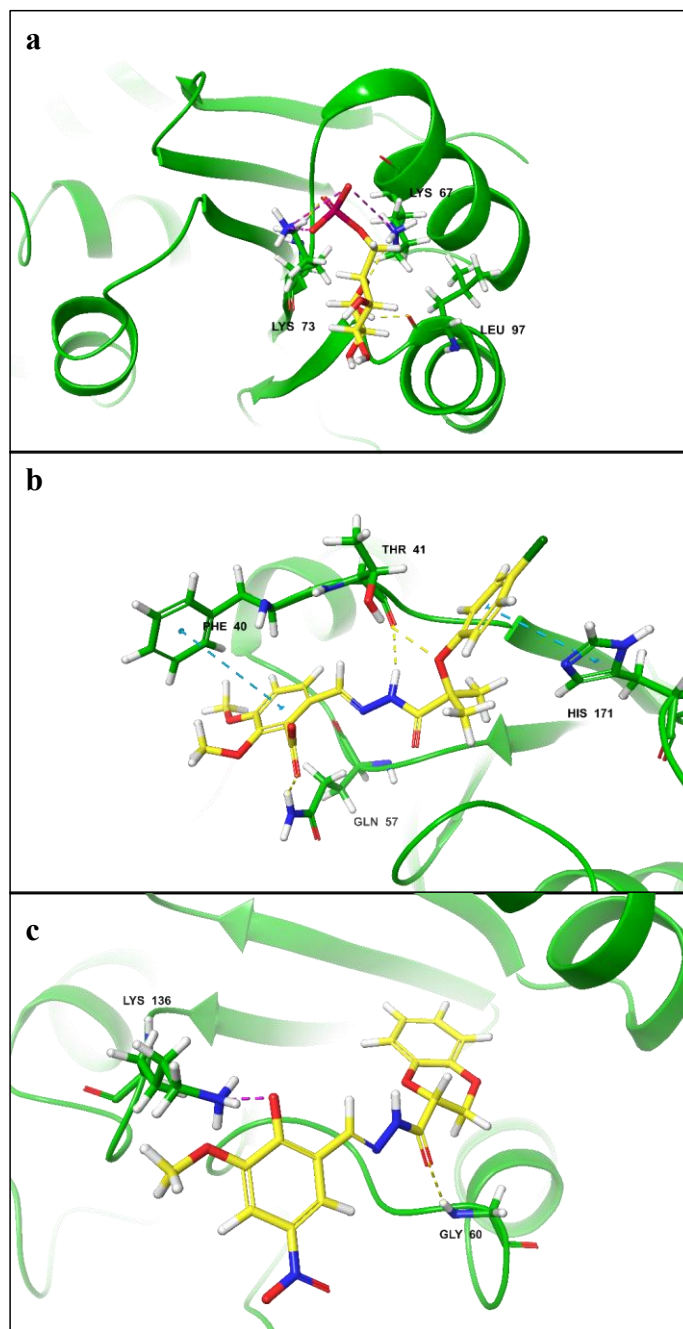
S. Figure 8 Snapshot of the protein-ligand interaction extracted from the complex at the 100ns of the 100 ns simulation. a) G6P- *C*Las STP b) G6P3510- *C*Las STP c) G6P6373- *C*Las STP. *C*Las STP protein is represented in green ribbons and the interacted residues are presented in green based stick representation. Small molecules are presented in yellow based stick representation. Important non-covalent bonding interactions such as hydrogen bonding (yellow), pi-pi stacking (blue), salt bridges (pink) and halogen bonding (purple) are indicated by dashed lines.
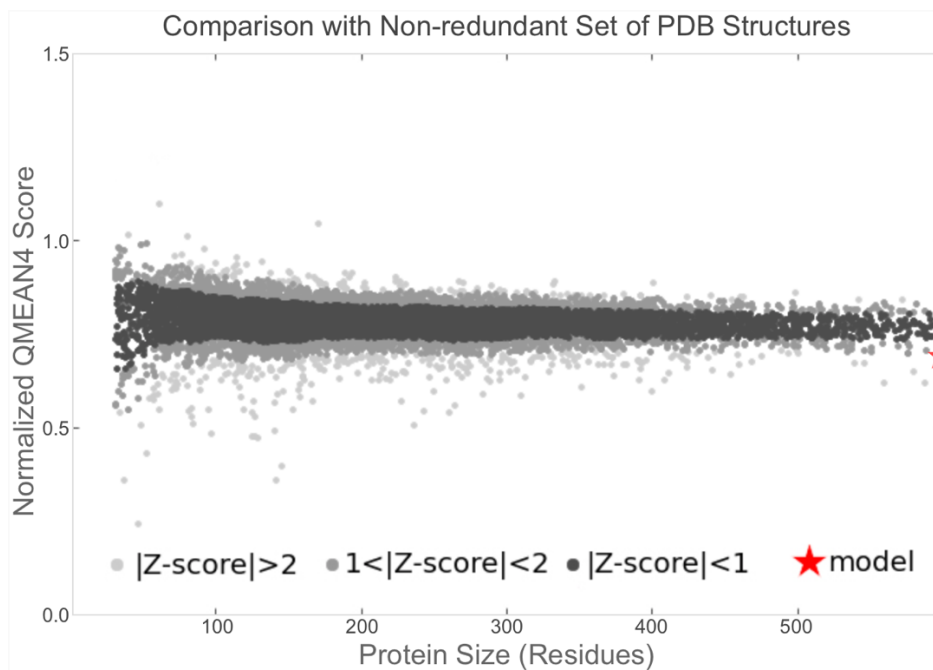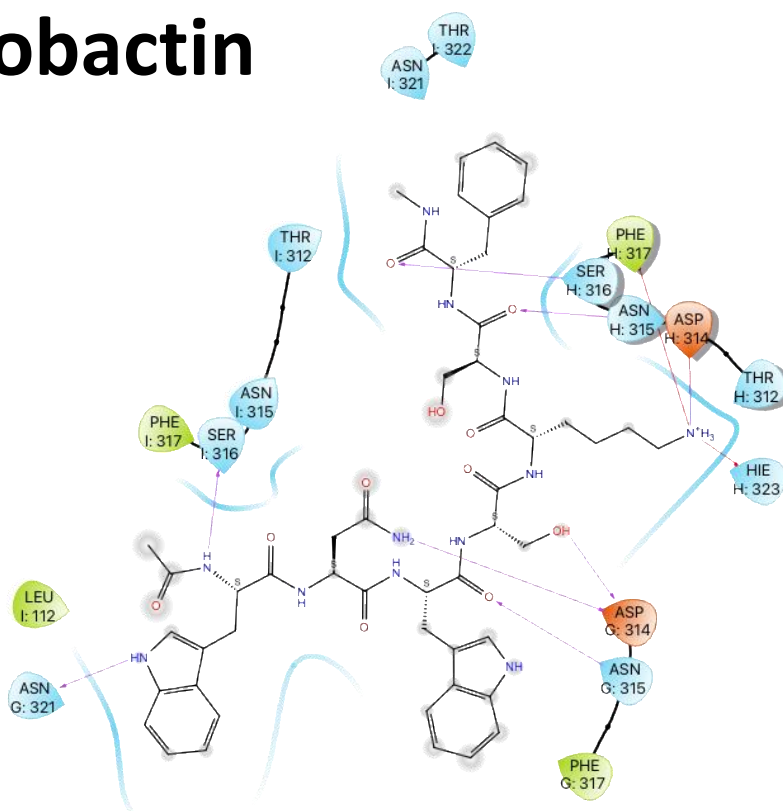
S. Figure 9 Z scores were predicted by comparison with the non-redundant PDB structure set of the proposed model.
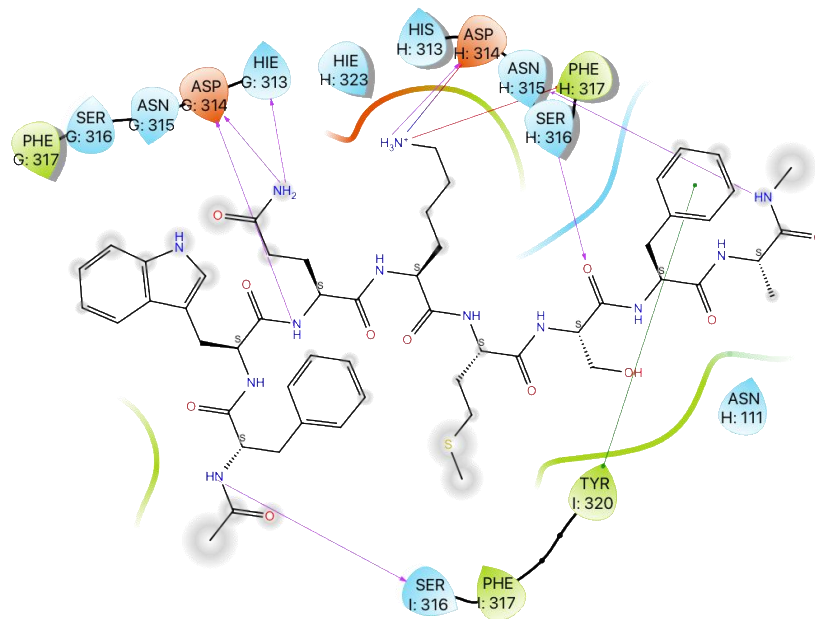
MRL-494
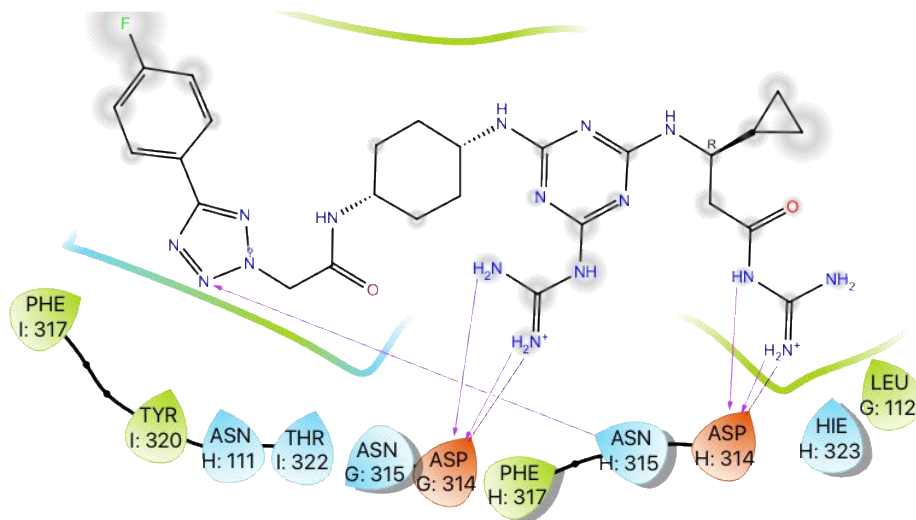
S. Figure 10 Interactions between MRL-494 (positive control) and receptors under SP-peptide docking mode.

Darobactin

S. Figure 11 Interactions between darobactin and receptors under SP-peptide docking mode.

Plantaricin JLA-9

S. Figure 12 Interactions between plantaricin JLA-9 and receptors under SP-peptide docking mode.