5-6-2019

BIOINFORMATICS TECHNIQUES FOR STUDYING DRUG RESISTANCE IN HIV AND STAPHYLOCOCCUS AUREUS

Shrikant Pawar

BIOINFORMATICS TECHNIQUES FOR STUDYING DRUG RESISTANCE IN HIV AND *STAPHYLOCOCCUS AUREUS*

by

Shrikant Pawar

Under the Direction of Dr. Irene. Weber, Ph.D.

ABSTRACT

The worldwide HIV/AIDS pandemic has been partly controlled and treated by antivirals targeting HIV protease, integrase and reverse transcriptase, however, drug resistance has become a serious problem. HIV-1 drug resistance to protease inhibitors evolves by mutations in the PR gene. The resistance mutations can alter protease catalytic activity, inhibitor binding, and stability.

Different machine learning algorithms (restricted boltzmann machines, clustering, etc.) have been shown to be effective machine learning tools for classification of genomic and resistance data. Application of restricted boltzmann machine produced highly accurate and robust classification of HIV protease resistance. They can also be used to compare resistance profiles of different protease inhibitors.

HIV drug resistance has also been studied by enzyme kinetics and X-ray crystallography. Triple mutant HIV-1 protease with resistance mutations V32I, I47V and V82I has been used as a model for the active site of HIV-2 protease. The effects of four investigational antiviral inhibitors was measured for Triple mutant. The tested compounds had significantly worse inhibition of triple mutant with Ki values of 17-40 nM compared to 2-10 pM for wild type protease. The crystal structure of triple mutant in complex with GRL01111 was solved and showed few changes in protease interactions with inhibitor. These new inhibitors are not expected to be effective for HIV-2 protease or HIV-1 protease with changes V32I, I47V and V82I.

Methicillin-resistant *Staphylococcus aureus* (MRSA) is an opportunistic pathogen that causes hospital and community-acquired infections. Antibiotic resistance occurs because of newly acquired low-affinity penicillin-binding protein (PBP2a). Transcriptome analysis was performed to determine how MuM (mutated PBP2 gene) responds to spermine and how Mu50 (wild type) responds to spermine and spermine–β-lactam synergy. Exogenous spermine and oxacillin were found to alter some significant gene expression patterns with major biochemical pathways (iron, sigB regulon) in MRSA with mutant PBP2 protein.

INDEX WORDS: HIV protease, Drug resistance, Machine learning, Restricted Boltzmann Machine, Structure-based drug design, *Staphylococcus aureus*.

BIOINFORMATICS TECHNIQUES FOR STUDYING DRUG RESISTANCE IN HIV AND *STAPHYLOCOCCUS AUREUS*

by

SHRIKANT PAWAR

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy, Ph.D. in the College of Arts and Sciences

Georgia State University

2019

BIOINFORMATICS TECHNIQUES FOR STUDYING DRUG RESISTANCE IN HIV AND *STAPHYLOCOCCUS AUREUS*

by

SHRIKANT PAWAR

Committee Chair: Dr. Irene. Weber

Committee: Dr. Irene. Weber

Dr. Robert. Harrison

Dr. John. Houghton

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

May 2019

DEDICATION

To my parents Mr. Dilip Pawar, Mrs. Dipali Pawar and my loving grandmother Mrs. Sawant.

ACKNOWLEDGEMENTS

Though only my name appears on the cover of this dissertation, a great many people have contributed to its production. I owe my gratitude to all those people who have made this dissertation possible and because of whom my graduate experience has been one that I will cherish forever.

My deepest gratitude is to my advisors, Dr. Irene. Weber, Dr. Robert. Harrison, Dr. Chung-Dar. Lu and Dr. John. Houghton. I have been amazingly fortunate to have advisors who gave me the freedom to explore on my own and at the same time the guidance to the right path. I am extremely grateful to Dr. Weber for holding me to a high research standard and enforcing strict validations for each research result, and thus teaching me how to do research. I hope that one day I would become as good an advisor to my students as my advisers have been to me.

I would give a heartfelt, special thanks to all the people in my department especially Mrs. Tameka. Hudson, Mrs. Ping. Jiang and Mrs. Sonya. Young. I am always inspired and encouraged by their passion on science, full of energy and hard work. Most of all their continuous encouragement, guidance and genuine caring enabled me to attend to life while also earning my Ph.D.

My committee members, Dr. Robert. Harrison and Dr. John. Houghton, their insightful comments and constructive criticisms at different stages of my research were thought-provoking and they helped me focus my ideas. My gratitude extends to Mrs. LaTesha. Warren. She was one of the first friendly faces to greet me when I began this doctoral program and has always been a tremendous help no matter the task or circumstance.

I would like to acknowledge all my lab mates especially Dr. Johnson. Agniswamy, Mrs. Yuan-Fang. Wang, Mr. Andres. Sam. Wong, Mr. Daniel Kneller (Danny), Mrs. Shelly. Burnaman for their valuable discussions, support and friendship. A very special thank you to my roommates and friends Mr. Vishakh. Walia, Mr. Mihiran. Vikram, Dr. Shukkor. Kondengaden and Mr. Rajnish. Sharma. They have always been staunchly supportive and caring. I am very glad we share so many precious memories along the way. For all these, I am eternally grateful. Most importantly, none of this would have been possible without the love and patience of my family. I would like to express my heart-felt gratitude to them, for their complete faith in me, their tolerance of my willfulness, and their great love beyond comparison in the world. This dissertation is dedicated to my parents Mr. Dilip Pawar, Mrs. Dipali Pawar and my loving grandmother Mrs. Sawant.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ATV: Atazanavir;
DRV: Darunavir;
FPV: Fos-Amprenavir;
HIV: Human immunodeficiency virus;
HIV PR: HIV protease;
IDV: Indinavir;
LPV: Lopinavir;
NFV: Nelfinavir;
PCA: Principle components analysis;
PPV: Positive Predictive Value;
RBM: Restricted boltzmann machine;
SQV: Saquinavir;
TPV: Tipranavir;
HAART: Highly Active Antiretroviral Therapy;
PR: HIV Protease;
PI: Protease Inhibitor;
MRSA: Methicillin-resistant Staphylococcus aureus.

1 GENERAL INTRODUCTION

Approximately 37 million people are currently infected with HIV worldwide [1]. HIV/AIDS epidemic has been partly controlled and treated by antivirals targeting different viral proteins, including HIV protease and transcriptase [1]. HIV is a retrovirus with two types, HIV-1 and HIV-2, and HIV-1 is subdivided into groups M, N, and O. Group M is further divided into subtypes with different geographical prevalence. The two HIV types have distinct origins. HIV-1 strains are closer to SIV that naturally infect chimpanzees (SIVcpz) [2], while HIV-2 strains are related to SIV from sooty mangabeys (SIVsm) [3]. HIV-1 group O is the most divergent group, and it has recently been suggested it could have originated from SIV that infected wild gorillas (SIVgor) [4]. SIVgor is related to SIVcpz and thus it is possible that gorillas are an intermediate reservoir of HIV-1 group O [5]. The HIV-1O epidemic pattern is restricted to the West and Central Africa [6], Group N was only identified in 1998 [7], and its origin was traced to a recombination event between the ancestor of group M and SIVcpz [8]. In the early 1990s the sequencing and alignment of viral genes env and gag from different strains of HIV-1 allowed for first time to establish the presence of well-defined HIV-1M genetic clades [9]. All HIV strains previously classified as subtype E based on env gene phylogeny has divergent subtype A classification in gag and pol [10]. According to the current classification, HIV-1 group M is divided into nine different "pure" subtypes or non-recombinant forms (A-D, F-H, J and K) [11]. Genetic differences between sub-subtypes are around 7% in pol, and to date only two subtypes are divided in sub-subtypes, subtype A (A1 through A5) and F (F1 and F2) [12]. The remaining six HIV-1 subtypes (D, F, G, H, J and K) represented together only 10% of new infections in 2004 [13]. HIV-1 non-B subtypes represent a challenge to HAART treatment as there are scarce studies in the literature on the

efficacy of treatment in the context of infection by these subtypes. Different HIV groups and subtypes carry in their genomes genetic signatures and polymorphisms that could alter the structure of viral proteins which are targeted by drugs, thus impairing ARV drug binding and efficacy. HIV-2 is less susceptible to some PIs, such as amprenavir, ritonavir and indinavir [13]. Various HIV-1 antivirals have been developed to block several essential proteins within the viral life cycle. HIV-1 protease, integrase and reverse transcriptase are successful enzyme targets in treatment of HIV-1 antivirals. HIV-1 PR is crucial for the viral maturation step; it processes cleavage sites in Gag-Pol region allowing for assembly of individual structural proteins. Some drugs block viral fusion by blocking gp120/gp41 receptors [14], or by blocking CCR5 co-receptor [15]. Nucleoside inhibitors (NRTI) perform chain termination [16] and non-nucleoside inhibitors (NNRTI) hinder polymerase activity [17] targeting reverse transcriptase. HIV-1 protease inhibitors prevent processing of Gag and Gag-Pol viral polyproteins during the virus maturation [18].

1.1. Resistance to HIV-1 PR inhibitors:

Evolution of HIV to gain resistance to anti-retroviral drugs emerges from a combination of several factors. The factors include the high genetic diversity of the virus, lack of proof-reading and rapid replication. The viral reverse transcriptase lacks a proofreading function and a low fidelity transcription drives an increased rate of spontaneous mutagenesis [14]. The long term usage of HAART (highly active antiretroviral therapy) has caused increase in the selection of resistant strains and the virus containing resistant protease mutants is still capable of reproducing in vivo by successful processing of the Gag and Gag-Pol precursors in the maturation stage. HIV PR mutations are classified as either major or minor resistance mutations by their effect in HAART

treatment [19]. Major mutations are distributed between the active site, flap and distal regions of the protease dimer. Several mechanisms have been described for resistance towards protease inhibitors. Single substitutions in active site cavity residues can result in altered drug interactions, mutations can lower dimer stability by interrupting the interactions at the interface between the two subunits in the dimer, and distal mutations can cause low dimer stability and defects in autoprocessing [30]. Some of the well-known single residue mutations causing resistance are D30N, V32I, I47V, G48V, I50V, Val82 and I84V [21]. Distal major mutation of L76V is associated with resistance to DRV, APV, IDV and LPV, while conferring susceptibility to the other PIs. Protease with this mutation shows low dimer stability and defects in autoprocessing [22].

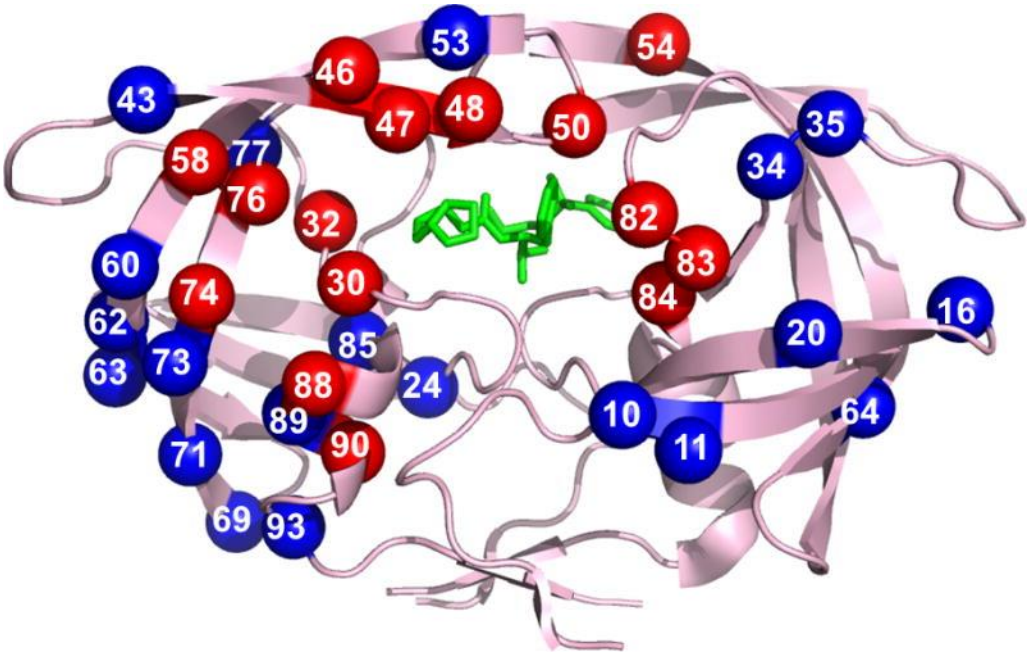
Individual mutations combine to produce highly resistant virus [32]. Several multidrug and highly resistant HIV protease mutants have been characterized to understand how the mutations affect the structure and function of the enzyme. A D25N variant of a multidrug resistant clinical isolate (MDR769) carries 10 mutations L10I, M36V, M46L, I54V, I62V, L63P, A71V, V82A, I84V, and L90M and a variant of MDR769 with V82T mutation shows 11–2600-fold resistance to all the nine available PIs [20]. A clinical isolate from a pediatric patient harboring 22 mutations in the protease (T4S, L10V, I13A, K14R, K20I, A22V, L33I, E35D, M36I, N37D, R41K, K43S, G48A, I54V, I66F, H69K, T74S, V82A, I84V, L89I, L90M and T91S) was studied by kinetic and x-ray analysis in complex with DRV [23]. Some of the mutation combinations (V82A/I84V/G48A) contribute to the collapse of the 80’s loop by directly affecting inhibitor binding to the S1/S1′ subsites inducing long-range structural changes altering inhibitor and substrate binding kinetics [24]. PR20, protease with 20 mutations indicates the higher dissociation constant stemming from loss of intersubunit contacts between the more flexible flaps. Mutations

M36I/I15V/133F in PR20 cause structural rearrangements causing flaps flexibility. Unlike other highly resistant mutants, PR20 retains the majority of hydrogen bond interactions with clinical inhibitors DRV, SQV and APV [25]. Different mutations in the precursor cleavage sites have been associated with resistant protease [15]. Mutations in the gag cleavage sites of NC-p2 and p2-p6 are commonly associated with drug resistance [15].

1.2. Structure of HIV-1 PR:

It is an aspartic protease which forms dimer from identical 99-residue monomers [26]. The dimer interface residues 1-4 from the N-terminus and 96-99 from the C-terminus glue the two monomers [27]. Flaps are flexible and they open and close upon substrate binding, residues I50 and G51 create a transient dimer interface [28]. There are several intermolecular interactions which occur with the movement of flaps. Water molecules play important role in the hydrolysis of substrate; it can form a tetrahedral intermediate causing substrate hydrolysis [28]. Figure 1 shows protease dimer showing sites of resistance mutations. Locations of resistance associated mutations mapped on the HIV protease dimer bound with DRV (PDB: 2IEN). Major and minor mutations are distributed among both monomers and shown in red and blue spheres, respectively [30].

Figure 1: Protease dimer showing sites of resistance mutations. Major and minor mutations are distributed among both monomers and shown in red and blue spheres, respectively [18].



1.3. Action of HIV-1 PR inhibitors:

Highly active anti-retroviral (HAART) therapy consists of different inhibitors targeting multiple life cycle of HIV virus [31]. HIV-1 protease inhibitors act by competitive inhibition. They occupy the PR active site and inhibit proteolysis in a competitive, reversible manner. P2-P2' positions on inhibitors fill in the S2-S2' pockets in PR [32]. A transition state is mimicked by an hydroxyl group between P1 and P1' sites which displaces the active-site water nucleophile [33]. The clinical protease inhibitors used in the current setting are atazanavir, indinavir, nelfinavir, amprenavir, darunavir, lopinavir, tipranavir and saquinavir.

1.4. Triple mutant protease with mutations V32I, I47V and V82I:

HIV-1 PR (PR1) and HIV-2 PR (PR2) share 39–48% amino acid sequence identity and the two enzymes differ in their cleavage site sequences in the viral precursors and in their specificity

for peptide substrates and inhibitors at the P2 positions of peptide substrates [33]. The binding site for clinical inhibitors differs only in the conservative substitution of hydrophobic residues Val32, Ile47, and Val82 in PR1 by Ile32, Val47, and Ile82 in PR2 [34]. Earlier studies showed that PR1 bearing the substitutions, V32I, I47V, and V82I, altered the inhibition but not the binding mode of a tripeptide inhibitor [35]. Tie *et al.* have shown that the PR with three resistance mutations V32I, I47V and V82I resembles PR2. It is poorly inhibited by amprenavir, although saquinavir and darunavir inhibition are similar to values for wild type PR1 and PR2 [36]. Kovalevsky *et al.* also proved that these three mutations in HIV-2 PR contribute fewer interactions with inhibitors relative to HIV-1 PR [37]. The mutation I47V is located at the flaps while the mutations V32I and V82I are away from the flaps. I47V mutation shortens the length of side chain for residue 47 while V32I and V82I mutations increase the size of mutated side chains. The mutated side chains alter hydrophobic interactions with inhibitors. A single substitution V82I in HIV PR-1 is associated with resistance to ATV [19]. Therefore, it is of importance to elucidate potential mechanism of drug resistance of three mutations V32I, I47V, and V82I toward inhibitors and testing if some newer potent inhibitors can effectively target such mutant protease.

1.5. Investigational inhibitors designed to target resistant HIV-1:

GRL01111, GRL0519, GRL0739 and GRL0249 are very potent HIV-1 protease inhibitors developed by structure-guided designs. GRL01111 has an improved backbone ligand-binding site interactions with incorporation of basic-amines at the C4 position of the bis-tetrahydrofuran (bis-THF) ring. This incorporation causes a unique hydrogen bonding interactions with the backbone carbonyl group of Gly48 as well as with the backbone NH of Gly48 in the flap region of the enzyme active site [39]. These ligand-binding site interactions are possibly responsible for their

potent activity. Furthermore, this inhibitor has displayed very potent enzyme inhibitory and antiviral activity [39]. GRL0519 is a novel nonpeptidic human immunodeficiency virus type 1 (HIV-1) protease inhibitor (PI) containing tris-tetrahydrofuranylurethane (tris-THF) and a sulfonamide isostere, is highly potent against laboratory HIV-1 strains and primary clinical isolates. GRL-0519 blocked the infectivity and replication of HIV-1NL4-3 variants selected by up to a 5 μM concentration of ritonavir, lopinavir, or atazanavir [40]. GRL-0519 is also potent against multi-PI-resistant clinical HIV-1 variants isolated from patients who no longer responded to existing antiviral regimens after long-term antiretroviral therapy, highly darunavir (DRV)-resistant variants [40]. Furthermore, the development of resistance against GRL-0519 is substantially delayed compared to other PIs like amprenavir (APV) and DRV. This PI has a tris-THF moiety, compared to the bis-THF moiety present in DRV, has greater water-mediated polar interactions with key active-site residues of protease and the tris-THF moiety and paramethoxy group effectively fill the S2 and S2′ binding pockets [40]. GRL-0739, a novel nonpeptidic HIV-1 protease inhibitor containing a tricycle (cyclohexyl-bis-tetrahydrofuranylurethane [THF]) and a sulfonamide isostere, is highly active against laboratory HIV-1 strains and primary clinical isolates with minimal cytotoxicity [41]. GRL-0739 is highly active against multidrug-resistant clinical HIV-1 variants isolated from patients who no longer responded to existing antiviral regimens after long-term antiretroviral therapy [41]. GRL-0739 has also been showed to have a desirable central nervous system (CNS) penetration property, as assessed using a novel in vitro blood-brain barrier model [41]. The tricyclic ring and methoxybenzene of GRL-0739 have a larger surface and make greater van der Waals contacts with protease than in the case of darunavir. GRL0249 is a novel C3-substituted cyclopentyltetrahydrofuranyl (Cp-THF)-derived HIV-1 protease inhibitor [42]. The C3-functional groups on the Cp-THF ligand maximize the ligand-binding site interactions in

the flap region of the protease [42]. It has displayed the most potent enzyme inhibitory and antiviral activity against a panel of multidrug resistant HIV-1 variants [42].

Specific aim 3 focuses on drug resistance in triple mutant protease to understand if it can be attenuated by formation of new interactions with inhibitors GRL01111 [39], GRL0519 [40], GRL0739 [41] and GRL0249 [42]. We hypothesize that these newer inhibitors will have increased interactions and improved inhibition towards the mutant HIV-1 protease relative to established clinical inhibitors like Lopinavir (LPV) or Darunavir (DRV). Enzyme kinetics were performed for each of the inhibitors with triple mutant to assess the effect relative to Ki values of wild type PR. Furthermore, crystal structure of triple mutant in complex with GRL01111 was solved for comparison with PRWT complex with the same inhibitor.

1.6. Machine learning for prediction of resistance:

Presence of specific mutations in the viral genome can be used to predict drug resistance. Several genotype interpretation algorithms have been utilized for predicting drug resistance. These algorithms either use a Genetics/Bayer Diagnostics genotype interpretation rules [43] to generate the susceptibility of the infecting virus for each drug or apply a score or 'penalty' for each drug such as the algorithm used in Stanford HIV database [44, 45]. Although these methods are fast, they suffer from the major disadvantage of relying on specific known mutations strongly associated with resistance and cannot identify newly appearing resistance mutations. Linear protein sequence based methods omitting potentially valuable information from the three-dimensional protein structure have also been applied in resistance classification. Statistical methods such as cluster analysis and linear discriminant analysis have been evaluated [47], and non-parametric methods have been proposed for high dimensionality data [48]. Several machine

learning algorithms like linear regression, decision trees, neural networks, support vector regression, and bayesian networks have been utilized to predict resistance. The KNN algorithm is a non-parametric method that uses the full training data set finding the K nearest neighbors to a query point and reports their class by majority vote. KNN in training stage is faster and it uses the complete training data in the prediction stage. Another important algorithm is random forest (RF) [47]. The RF algorithm is an ensemble based classifier that works with multiple decision trees to improve the accuracy. Previous studies from our lab have predicted phenotypic results successfully from protease sequences by applying a unified encoding of sequence and protein structure as a feature vector [46]. Yu, et. al. (2014) have utilized support vector machines for classification of resistance mutations [46]. This classification shows high accuracy, sensitivity and specificity for all inhibitors. For PIs the accuracy values range from a low of 0.93 to a high of 0.96, while sensitivity and specificity range from 0.92-0.96 and 0.94-0.98, respectively. The same technique was applied with Artificial Neural Networks (ANN) to classify genotype-phenotype data for resistance and the values calculated for accuracy, sensitivity and specificity for resistance to PIs have a low of 0.91 and reach 0.97 [46]. Significantly higher accuracies have been achieved with this technique. This approach utilizes sequences from publicly available high quality filtered datasets obtained from the Stanford HIV drug resistance database, which includes the results of drug susceptibility tests using the PhenoSense assay (http://hivdb.stanford.edu/pages/genopheno.dataset.html). Eight protease inhibitors, atazanavir (ATV), indinavir (IDV), nelfinavir (NFV), amprenavir (APV), darunavir (DRV), lopinavir(LPV), tipranavir (TPV) and saquinavir (SQV), are included in the datasets.

One approach developed in our laboratory to select a small number of meaningful mutants uses the Mean shift clustering seeking the mode of a density function in the given sample set [65].

By clarifying the relationship between mean shift and the optimization, the algorithm could potentially be applied on clustering and global optimization problems by declaring each mode as representative of one cluster and assigning each data point to the mode it converges to. A non-parametric iterative mean shift with protein encoding method to extract the most representative drug resistant mutants from the Stanford HIV database has been proven successful in identifying representative drug resistant mutants in HIV-1 protease [67, 68]. This selection algorithm works well on selecting drug resistant mutants from both HIV PR and RT inhibitor genotype/phenotype data. Among all the mutants, around 250 most representative mutants were selected for different drugs. Such selection was based on the kernel bandwidth, and the goal correlation (R2) value. After selection, the multiple linear regression was applied on the selected mutants drug resistance values [65]. Identifying a small number of representative mutants will enable laboratory studies of the molecular mechanisms of resistance. In another study, RF algorithm prediction accuracies were examined by five-fold cross-validation on the genotype-phenotype datasets [69]. A supervised machine learning approach for automatic prediction of drug resistance was developed to handle genotype-phenotype datasets of HIV protease (PR) and reverse transcriptase (RT). It predicts the drug resistance phenotype and its relative severity from a query sequence. The accuracy of the classification was higher than 0.973 for eight PR inhibitors and 0.986 for ten RT inhibitors, respectively. The overall cross-validated regression R2-values for the severity of drug resistance were 0.772–0.953 for 8 PR inhibitors and 0.773–0.995 for 10 RT inhibitors [69].

1.7. Delaunay triangulation:

As mentioned in section 1.6, each of the machine learning approaches like mean shift clustering or RF utilizes our efficient in-house sequence-structure encoding technique with

Delaunay triangulation (DT) [46, 65, 67, 69]. DT is a triangulation such that the circumsphere of every d-simplex is empty. It does not contain any of the given points in its interior. DT minimizes the maximum radius of circle and it maximizes the minimum angle. Distance along edges from DT between two vertices is constant time the Euclidean distance between them. There are several algorithms for calculating DT: local improvement, starting with an arbitrary triangulation, algorithm locally flips the faces of pairs of adjacent simplices according to the circumsphere criteria. Incremental construction, DT constructed by successively building simplices whose circumspheres contain no points in P (first point selection is usually a middle point). Incremental insertion, recursively adding and testing by putting points from P at one time in a triangulation (order of adding is not important). Higher dimensional embedding by applying convex hull (smallest convex set that contains X). Divide and conquer, recursive partitioning and local triangulation of point sets. Euclidean distance or Euclidean metric is the straight-line distance between two points in Euclidean space. In three-dimensional Euclidean space, the distance is:

$$d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + (p_3 - q_3)^2}.$$

In MATHLAB '*delaunayn'* function can make DT graphs. T = Delaunayn(X) computes a set of simplices such that no data points of X are contained in any circumspheres of the simplices. The set of simplices forms the Delaunay triangulation. X is an m-by-n array representing m points in n-dimensional space. T is an array where each row contains the indices into X of the vertices of the corresponding simplex.

1.8. Clustering algorithms:

As stated in section 1.6, the mean shift algorithm (Fukunaga and Hostetler 1975; Cheng 1995) is a recursive algorithm that allows us to perform non-parametric mode-based clustering, i.e. clustering data on the basis of a kernel density and has been successfully used to predict sequences with high level resistance [65]. Agniswamy et al 2016 have successfully reported structural analysis of HIV protease variant PRS17 selected by this technique. Other different techniques of cluster analysis which can be applied to improve the existent analysis are KNN: an algorithm which determines the classification of a point, combines the classification of the K nearest points, Agglomerative hierarchical clustering: Hierarchical clustering by starting with each point as a singleton cluster and repeatedly merging two clusters until a single, DBSCAN: Density based clustering algorithm that produces a portioning cluster in which number of clusters is determined by a density algorithm. It becomes a challenging task to identify certain important, representative HIV-1 PR sequences from a pool of several hundred sequences. This paper attempts to answer this question by implying some supervised and unsupervised machine learning techniques. Support vector machine (SVM) and random decision forest (RF) are supervised machine learning techniques, while K means, hierarchical agglomerative and divisive clustering are categorized as unsupervised machine learning techniques, and we have implemented them to identify important PR sequences from a large pool of mutants. While some groups have tried to provide valuable insights into characterization of novel HIV drug resistance mutants using clustering, multidimensional scaling and SVM based feature ranking from relational clinical databases, ours attempts to select resistant sequences from existing pool.

1.9. Transcriptomic analysis of Methicillin-resistant *Staphylococcus aureus* (MRSA):

MRSA is a rapidly emerging *Staphylococcal* infection which has developed resistance to most of the β-lactam antibiotics causing it difficult to treat [49]. Staphylococcus aureus is a gram-positive, round-shaped bacterium that is a member of the Firmicutes, and it is a usual member of the microbiota of the body, frequently found in the upper respiratory tract and on the skin [50]. Staphylococcus aureus is both a commensal bacterium and a human pathogen. Approximately 30% of the human population is colonized with S. aureus [51]. It is a leading cause of bacteremia and infective endocarditis (IE) as well as osteoarticular, skin and soft tissue, pleuropulmonary, and device-related infections. While overall rates of Staphylococcus aureus may have stabilized over the past 20 years, the contribution of methicillin-resistant S. aureus (MRSA) has fluctuated. In Quebec, Canada, the incidence of MRSA bacteremia increased from 0 per 100,000 person-years to 7.4 per 100,000 person-years from 1991 to 2005 [52]. Similar trends of increasing MRSA bacteremia incidence over this time period were seen in Minnesota from 1998 to 2005 [53], Calgary, Canada, from 2000 to 2006 and Oxfordshire, United Kingdom, from 1997 to 2003 [53]. In North America, epidemic community-associated clones of MRSA (e.g., USA300) have been largely responsible for the increase in the incidence of MRSA bacteremia while in the United Kingdom, epidemic health care-associated clones of MRSA [53]. Methicillin-resistant Staphylococcus aureus (MRSA) is an opportunistic pathogen that causes hospital and community-acquired infections varying from mild skin lesions to severe, life-threatening conditions like pneumonia, sepsis, and endocarditis [54]. The resistance occurs because of newly acquired low-affinity penicillin-binding protein (PBP2a), which can build the cell wall when other PBPs are blocked by beta-lactams [55]. Beta-lactam antibiotics irreversibly occupy the serine residue at the active site of penicillin-binding proteins (PBPs) forming a stable ester-linked acyl-enzyme which inhibits the transpeptidation process necessary for cell wall cross-linking [56]. Thus, MRSA seems

to regulate a complex gene expression profiling when challenged with beta-lactams. Many studies have shown that exogenous spermine may become toxic to Staphylococcus aureus [57]. Exogenous spermine exerts a dose-dependent inhibition effect on the growth of E. coli, Salmonella enterica serovar, and Staphylococcus aureus. In *Staphylococcus aureus*, there are four native PBPs (PBP1-4) and one acquired PBP2a is attributive to MRSA [58]. Recently, it was reported that exogenous spermine may affect cell wall synthesis through its interactions with PBP2 and/or PBP2-associated multienzyme machineries to enhance the killing effects of β-lactam antibiotics [59]. Here an unusual finding of down regulation in sigB regulon and most ATP-producing pathways to high dose spermine stress was reported. Exogenous spermine exerts a dose-dependent inhibition effect on the growth of *E. coli, Salmonella enterica serovar* and *Staphylococcus aureus* [60]. PBP2 itself or enzymatic activities associated with the PBP2-dependent complex in cell wall synthesis could be potential target of spermine so we sought to identify a spontaneous mutation derived MRSA strain (MuM) conferring spermine resistance and studying how this mutation can contribute to resistance. Vancomycin works by binding to the D-Ala-D-Ala moiety, and subsequently blocking cross-linkage of the murein monomers during the transpeptidation reactions by PBPs [61]. We also made an intriguing finding where MuM lost spermine–β-lactam synergy and instead gained a spermine-vancomycin synergy. Transcriptome of how MuM responds to spermine and how Mu50 responds to spermine and spermine–β-lactam synergy is still unknown, therefore in this aim we have systematically compared Mu50 (wild type) and MuM response to spermine alone (high dose) or in combination with b-lactam (oxacillin) (both at low dosages) using bioinformatics tools. We hypothesize that exogenous spermine and oxacillin can alter some significant gene expression patterns with major biochemical pathways in *Staphylococcus aureus* with mutant PBP2 protein.

Specific Aim 1: Analysis of drug resistance in HIV protease with Restricted Boltzmann machine.

Specific Aim 2: Computational optimization of previously defined graph based sequence structure resistance prediction technique.

Specific Aim 3: Investigating effectiveness of newer protease inhibitors against HIV-1 triple mutant.

Specific Aim 4: To analyze biological pathways in *Staphylococcus aureus* with mutant penicillin binding protein 2 on spermine and oxacillin stress.

2 ANALYSIS OF DRUG RESISTANCE IN HIV PROTEASE

Pawar S, Freas C, Weber I, Harrison R. (2018) Analysis of drug resistance in HIV protease. *BMC Bioinformatics*, 19(l11):362.

Abstract

Background: Drug resistance in HIV is the major problem limiting effective antiviral therapy. Computational techniques for predicting drug resistance profiles from genomic data can accelerate the appropriate choice of therapy. These techniques can also be used to select protease mutants for experimental studies of resistance and there by assist in the development to next-generation therapies.

Results: The machine learning produced highly accurate and robust classification of HIV protease resistance. Genotype data were mapped to the enzyme structure and encoded using Delaunay triangulation. Generative machine learning models trained on one inhibitor could classify resistance from other inhibitors with varying levels of accuracy. Generally, the accuracy was best when the inhibitors were chemically similar.

Conclusions: Restricted Boltzmann Machines are an effective machine learning tool for classification of genomic and structural data. They can also be used to compare resistance profiles of different protease inhibitors.

2.1 INTRODUCTION

Human Immunodeficiency Virus (HIV) is a major pandemic disease [62]. More than 36 million people have been infected and about half of these people receive anti-retroviral therapy [63]. However, retroviruses like HIV mutate rapidly since the conversion from the RNA genome to DNA is error-prone [64]. They readily form quasi-species and distinct viral strains. Therefore, retroviruses can respond effectively to selective pressures such as drug treatment by mutating to evade the antiviral drug. The development of drug resistance in HIV is an ongoing threat to effective long-term therapy.

Machine learning can predict drug resistance from sequence data with high accuracy as shown by tests on genotype-resistance data for HIV protease and reverse transcriptase [65–72]. The critical improvement in the application of machine learning to drug resistance is the inclusion of structural data in the features. We found that using Delaunay triangulation to encode the protein structure [73] is highly effective. The encoding compresses a protein sequence and its corresponding structure into a feature set consisting of 210 components. The set contains the relative frequencies of each kind of amino acid pair from the structure. Yu's use of compressed encoding in [65] suggested that even fewer features were necessary to encapsulate drug resistance. Therefore, we used Principle Components Analysis (PCA) to explore the remaining redundancy in the data. The availability of a large amount of sequence and resistance data for HIV protease (PR) has proved valuable for method development.

The validity of this approach was verified by experimental studies [74, 75]. Machine learning was used to rigorously select representative highly resistant PR sequences for biochemical and structural characterization. The computationally selected mutant demonstrated several orders of magnitude worse affinity for inhibitors compared to wild type enzyme. The selected mutant had

only one mutation in the inhibitor binding site. Therefore, a high level of resistance was achieved almost exclusively by mutations distal from the active site.

Restricted Boltzmann Machines (RBMs) are a generative machine learning algorithm [76, 77]. RBMs only require positive, or in-class, training data, and often generalize more accurately than other approaches. Training the standard algorithm on large datasets is often computationally infeasible. We have developed a highly efficient version of this algorithm [78, 79]. Using a simplified representation of the hidden and visible spins and replacing a numerical estimate of the gradient with an analytic form results in an algorithm that is at least 14 times faster than the conventional algorithm without compromising the accuracy.

Generative machine learning has not been applied to drug resistance in HIV. Therefore, application of this approach to the analysis of drug resistance is of interest. This paper shows that RBMs are as accurate as other machine learning approaches for these data. Additionally, we studied how well RBMs trained on one drug were able to predict resistance for a different drug.

2.2 METHODS

2.2.1. Datasets and data preparation

Datasets used for the study: The genotype-phenotype datasets were downloaded from the Stanford HIV drug resistance database [80]. Data were used for the HIV protease inhibitors: atazanavir (ATV), nelfinavir (NFV), ritonavir (RTV), indinavir (IDV), lopinavir (LPV), tipranvir (TPV), saquinavir (SQV), fosamprenavir (FPV) and darunavir (DRV). All the datasets were pre-processed using the methods and the cutoff values described previously in [65]. The threshold for resistance recommended by the database curators was used in this work [80]. The results of the

expansion of data for each of the HIV-1 PR inhibitors and proportion of resistant mutants are shown in Table 1.

Table 1The results of the expansion for each of the HIV-1 PR inhibitors

Inhibitor	No. isolates	No. sequences	No. resistant	No. sensitive	Fraction resistant
SQV	1722	10258	4206	6052	41.0
DRV	607	5973	1889	4084	31.6
LPV	1444	10239	5095	5144	49.8
NFV	1771	10911	6170	4741	56.5
IDV	1730	10537	5122	5415	48.6
ATV	1141	8430	4237	4193	50.3
FPV	1681	10521	4405	6116	41.9
TPV	847	7363	2062	5301	28.0

2.2.2. Pre-processing/expansion of the datasets

Wild type HIV PR has a protein sequence of 99 amino acids. Sequences with insertions, deletions, or stop codons were removed. Genomic datasets often include multiple mutations at the same site. In these cases, the data were expanded to multiple sequences with single amino acids at each location to represent a single amino acid sequence for each mutant protein. For example, if one 99-amino acid mutant sequence has two different types of amino acids at one position and another site has three, this one sequence needs to be represented by six unique sequences each differing in only one amino acid substitution. The pre-processing method has been explained in

detail in [65]. Each sequence was accompanied by its inhibitor resistance fold values. The relative resistant fold values for each of the inhibitors ranged from 0 to 800-fold resistance. Finally, the expanded datasets with sequences were allotted a unique identifier number to help recover the original sequences and their respective resistance fold change after analysis.

2.2.3. Encoding structure and sequence with Delaunay triangulation

A graph-based encoding system was utilized to represent the sequence and structural information of the protein [67]. The X-ray crystal structure for HIV-1 PR (3OXC) [58] was used as a template for creating the Delaunay triangulation. The structurally adjacent pairs of amino acids were represented as a vector of the 210 unique pairs of 20 standard amino acids. This graph-based encoding of sequence and structure has been proven to be a promising technique for fast and accurate predictions of resistance from sequence in HIV infections [46, 66].

2.3.4. Principal component analysis

Principal Component Analysis (PCA) using Singular Value Decomposition (SVD) was run on all the HIV-1 PR datasets using the Scikit-Learn machine learning library [82]. The datasets for each inhibitor were analyzed using the Pandas data analysis library [83]. The resistance fold values were not included in the PCA calculations since predicting these values is the goal of this work. The results of this analysis are detailed in the “Results” section.

2.3.5. Training the RBM

The mutants with relative resistant fold less than 3.0 were classified as non-resistant (susceptible) and denoted as 0; while those with relative resistant fold of greater than 3.0 were

classified as resistant and denoted as 1, as used in [65] and consistent with other analyses of the Stanford HIV resistance database [80]. RBMs work best with bit patterns. These bit patterns were generated by scaling and dividing the range of individual features into equal intervals. Each feature of the data was scaled to the range 0 to 1 based on the maximum and minimum values of that feature. The scaled data were divided into eight intervals encoded with three bits per feature. The testing and training sets were scaled independently.

The RBM was trained using gradient descent with the derivative as shown in Eq. 1. The analytic expression for the expected value of the derivative, shown in Eq. 2 and derived in [78], was used.

$$dU/dW_{i,j} = H_j V_i - < dU/dW_{i,j} > \quad (1)$$

$$< dU/dW_{i,j} >= H_j \, e\beta U - e{-}\beta U / e\beta U + e{-}\beta U \quad (2)$$

In these equations, H and V are hidden and visible (or input) layers respectively, β is the inverse temperature, U is the potential energy, and W are the weights used to define the potential.

During training, the layer that gave the best fit for each new data point was updated with a descent step and the other layers were "anti-trained" with a small ascent step. "Anti-training" improves the convergence and training efficiency of the RBM. Anti-training is only feasible when using an analytic expression for the training gradient. An RBM with 150 units in the hidden layer was trained for each category with a constant step-size of 0.1. A step-size of 0.01 was used for anti-training. An RBM was trained for both resistant and non-resistant classes. Class membership was assigned by the fractional reconstruction error, shown in Eq. 3 as defined in [78].

$$R = H_i \, j \, W_{i,j} V_j / |H_i| \, j \, W_{i,j} C_j \text{ where C is the perfect reconstruction.} \quad (3)$$

Five-fold cross validation was used to ensure that the results reflect the error in the models. The models for each fold were trained to convergence with ten iterations and the values for accuracy, positive predictive value, recall, and F from the last iteration were reported.

2.3 RESULTS

2.3.1 Classification with an RBM

The classification results are detailed in Table 2 and show a high degree of accuracy.

Table 2 The accuracy of the machine learning model

Inhibitor	Accuracy	PPV	Recall	F
Idv	0.979	0.974	0.985	0.979
Lpv	0.984	0.977	0.992	0.984
Sqv	0.969	0.963	0.986	0.974
Tpv	0.987	0.984	0.998	0.991
Drv	0.988	0.985	0.998	0.992
Atv	0.983	0.976	0.989	0.983
Nfv	0.978	0.974	0.975	0.975
Fpv	0.988	0.984	0.998	0.991

The nearly uniform values of close to 1.0 for accuracy, PPV, recall, and F-score, show that the models reliably predict both resistant and nonresistant classes. These results compare favorably with our earlier results using non-generative machine learning algorithms [65–67, 69].

2.3.2 Cross-classification with an RBM

RBMs differ from non-generative machine learning methods in an interesting way. It is trivial to train an RBM against one dataset and use it to predict the behavior of another. Table 3 shows the results of a cross-training analysis of resistance data.

Table 3Cross training reveals similarity between the inhibitors

Compound	Atv	Drv	Fpv	Idv	Lpv	Nfv	Sqv	Tpv
Atv	0.990	0.868	0.880	0.955	0.946	0.914	0.893	0.819
Drv	0.767	0.996	0.818	0.786	0.785	0.718	0.792	0.925
Fpv	0.929	0.873	0.981	0.889	0.886	0.822	0.822	0.828
Idv	0.945	0.863	0.880	0.989	0.960	0.905	0.878	0.809
Lpv	0.939	0.892	0.877	0.963	0.988	0.891	0.865	0.837
Nfv	0.923	0.853	0.824	0.918	0.901	0.987	0.837	0.758
Sqv	0.898	0.837	0.825	0.890	0.871	0.840	0.983	0.807
Tpv	0.723	0.929	0.765	0.729	0.728	0.655	0.732	0.993

Each row was trained on one inhibitor and the columns show the accuracy with which that model predicts the resistance for the other inhibitors. The inhibitors generally, but not completely, cross-classify with high accuracy. TPV and DRV seem to have more differences from the other inhibitors. The ability of an RBM trained on resistance to one inhibitor to predict the behavior of resistance to another inhibitor shows that the drug resistance of HIV protease does not fully depend on the type of drug. The existence of cross-resistance is well known and our lab has used similar approaches to identify interesting multi-drug resistant mutants for structural study [69, 74, 75].

2.3.3 Principal component analysis

Figure 2 shows the explained variance for each of the datasets as a function of the number of reduced dimensions. As shown in the figure, there is overlap between some of the datasets (some of the plots depict the same data). This suggests that redundancy exists between datasets and not just within a single dataset. The horizontal line in the figure depicts where at least 95% of the explained variance of the datasets is captured. In most cases, the first principal component explained at least 90% of the observed variance. This was true for the ATV, LPV, NFV, and SQV inhibitors. The remaining four inhibitors had an explained variance ratio between 51% and 87% for the first principal component. For all inhibitors except for DRV, 95% of the explained variance for each dataset was captured within 60 dimensions, suggesting that the data could be further compressed while still minimizing the reconstruction error. For DRV, the explained variance could be reduced to 50 dimensions. These results indicate that a more compact encoding for the resistance data exists, consistent with Yu's results on effectiveness of compressed encoding for machine learning [65].

2.4 DISCUSSION

2.4.1. Classification of resistant mutations of HIV PR

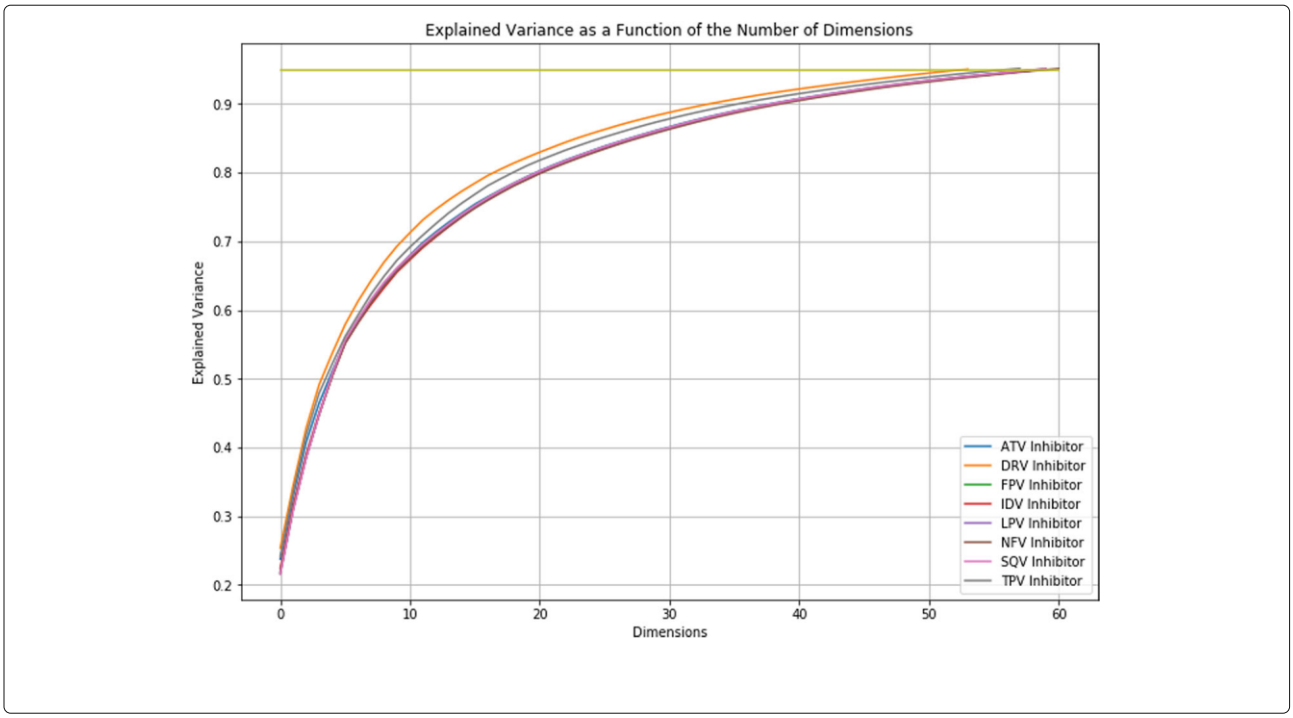
The combination of structure-based encoding and RBMs is an effective technique for the prediction of drug resistance in HIV PR. The five-fold cross validated results in Table 2 clearly demonstrate their success and accuracy. The high values for PPV indicate that the models could be clinically valuable. The use of an RBM is especially interesting because there are essentially no

adjustable parameters in the process. Efficient training algorithms allow the RBM to handle large datasets in reasonable times. While these datasets are not quite big data, they are too big for other machine learning programs [68, 84].

2.4.2. Comparison with other methods

We pioneered the approach of using a unified representation of sequence with 3D structural data expressed as a 210-long feature vector for

Figure 2 Principal Component Analysis on the HIV-1 PR Datasets. The similarity in the curves indicates that the datasets have a similar underlying structure



machine learning [65]. This approach gave improved accuracy for predicting drug resistance for HIV protease and reverse transcriptase compared to using sequence data alone.

Another group reported mean R2 values of >0.95 for regression with ANN using a subset of HIV sequences restricted to subtype B with the data filtered to remove rare variants [71]. Their classification accuracy was less impressive. Structural data can also be represented by molecular mechanics calculations on protein-drug complexes. Molecular interaction components calculated between a drug and 36 single mutants of HIV protease were used for SVM classification of resistance and showed improved accuracy over using sequence alone [70]. These results were comparable to our earlier results, but for a much smaller number of sequences. Feature vectors derived from a four-body statistical potential and n-grams were applied in [72]. This approach also used explicit atomic models for the protease and therefore only a few hundred mutants were included for classification and regression. Their reported accuracy was worse than ours.

Our approach preserves structural information using Delauney Triangulation derived from a single protein structure, and is applicable to any mutant, while eliminating the expensive step of calculating molecular properties for models of every mutant structure.

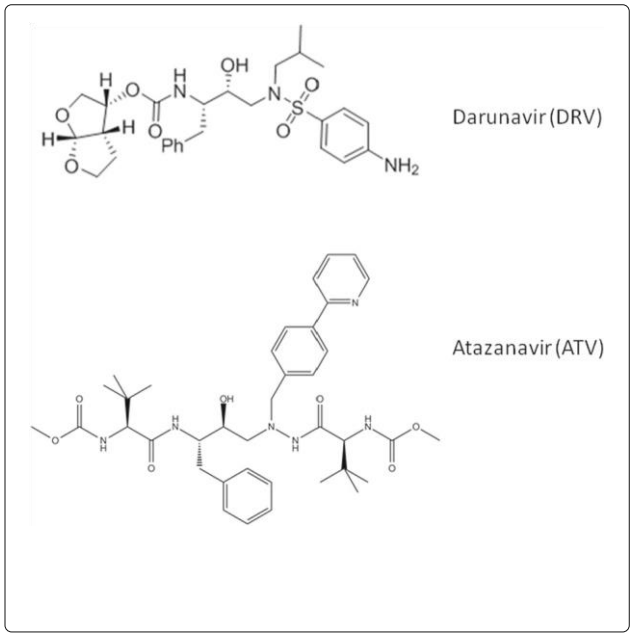
2.4.3 Redundancy in the data

One of the original motivations for exploring graphbased encoding of protein structures was to remove unnecessary data while retaining the critical features for machine-learning based analysis of structure and function [61]. Earlier work [65], which used compressed encoding, hinted that the redundancy was not completely removed from the data. Our use of PCA on the data demonstrated that further compression is possible because the majority of the variance in the data could be captured with 50–60 dimensions instead of the 210 used in the original representation. This strongly suggests that we may be able to extract patterns of mutations associated with drug resistance from the structural data itself.

2.4.4 Inhibitor specific patterns of drug resistance

Another important difference between generative machine learning and more conventional algorithms is that it is logically consistent to apply generative machine learning across categories. Since the RBM is essentially measuring how well it can reconstruct a given data point, it makes sense to ask whether an RBM trained on one inhibitor such as ATV could reconstruct data for a different inhibitor such as DRV. Examples of two inhibitors, Darunavir and Atazanavir are shown in Fig. 3 which demonstrates the diversity of drug chemistry used to inhibit HIV PR.

Figure 3 The chemical structures for a sulfonamide-containing (DRV) and non-sulfonamide-containing (ATV) inhibitor are shown here. These demonstrate the variety of chemistry used in inhibitors



The inhibitors segregate into two main classes in the cross-training analysis. Cross-training results in high accuracy for most inhibitors, with the exception of DRV and TPV, which both incorporate sulphonamides. DRV and TPV, predict each other with reasonable accuracy (92.5%), however they show worse prediction for other inhibitors. While this could be due to chemical similarity, it

could also be due to these being second generation or salvage inhibitors where the full spectrum of resistance mutations has not had time to evolve.

2.5 CONCLUSION

Generative machine learning algorithms such as the RBM are well-suited to the prediction of drug resistance in HIV PR, and likely will work on other systems as well. The graph-based structure/sequence encoding used in this and related work removes much of the redundancy in the data, but does not remove it all. This result suggests that even more efficient encoding schemes are possible. The RBM was used to analyze similarities in resistance profiles for different clinical inhibitors. The analysis suggests that there are at least two main classes of inhibitors for HIV PR.

3 MACHINE LEARNING APPROACH FOR SELECTING DELINEATIVE DRUG RESISTANCE MUTANTS OF HIV-1 PROTEASE*

*In the process of submission.

Abstract

Background: Highly active antiretroviral therapy (HAART) is a combination of different antiretroviral drugs, currently prescribed as an anti-AIDS treatment. Upcoming drug resistance to HAART have caused an alarming concern in its treatment. Several groups have tried to predict drug resistance from genotype data by a variety of algorithms. It becomes a challenging task to identify certain important, representative HIV-1 PR mutants from a pool of several hundred mutants.

Results: This paper attempts to answer this question by implying some supervised and unsupervised machine learning techniques. Support vector machine (SVM) and random forest (RF) are supervised machine learning techniques, while K means, hierarchical agglomerative and divisive clustering are categorized as unsupervised machine learning techniques, and we have implemented them to identify important PR mutations from a pool of mutants. By intersecting sequences with the 2 clustering techniques, 32, 257, 287, 76, 53, 104, 331, and 304 sequences felt common with hierarchical and divisive clustering for each of the ATV, DRV, FPV, IDV, LPV, NFV, SQV and TPV inhibitors. Further, intersecting these with the K means clustering selected only 20, 37, 2 and 35 sequences for FPV, LPV, NFV and SQV inhibitors.

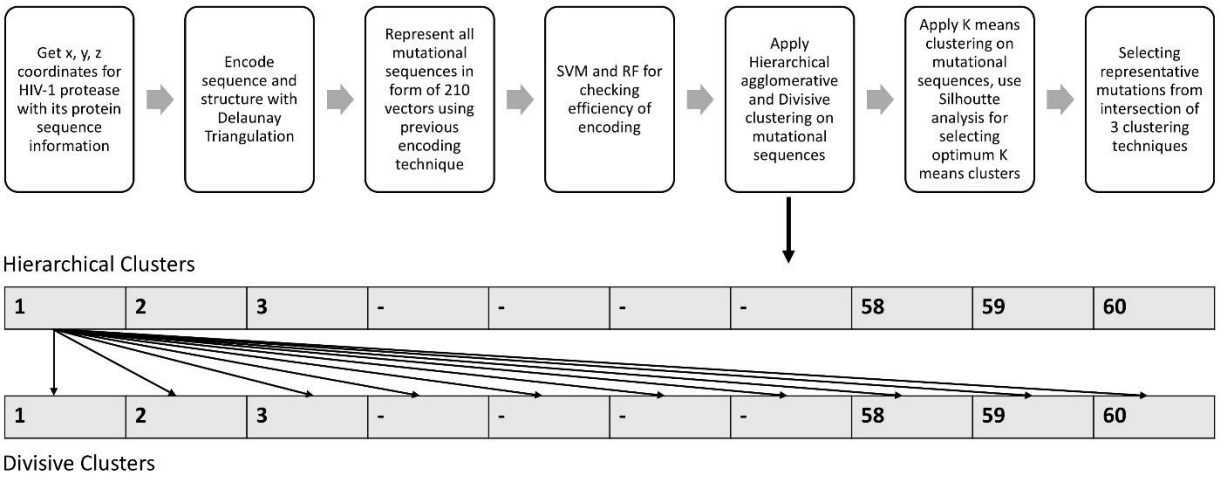
Conclusion: In this study, we have effectively encoded HIV-1 protease sequence and structure information in a vector by using Delaunay triangulation. The effectiveness is tested by SVM and RF learning algorithms. We then have utilized 3 un-supervised machine learning algorithms for identifying most representative mutants. These resistant sequences with mutants will enable laboratory studies of the molecular mechanisms of resistance.

3.1 INTRODUCTION

Acquired Immunodeficiency Syndrome (AIDS) is one of the deadliest pandemic disease with approximately 1.1 million people living with HIV in US [86]. A total of 26 licensed drugs are used in anti-AIDS therapy targeting viral entry, reverse transcription, integration and maturation [87]. An important enzyme, HIV protease (PR) causes processing of viral precursor proteins after budding of virus from the host cell during the maturation stage [88]. Inhibitors to block this PR and subsequently its proteolytic activity have been effective in controlling infections [89]. Highly active antiretroviral therapy (HAART) is a combination of different antiretroviral drugs, currently prescribed as a anti-AIDS treatment [90]. Upcoming drug resistance to HAART have caused an alarming concern in its treatment [90]. Several groups have tried to predict drug resistance from genotype data by a variety of algorithms [91] [92] [93]. It becomes a challenging task to identify certain important, representative HIV-1 PR sequences from a pool of several hundred sequences. This paper attempts to answer this question by implying some supervised and unsupervised machine learning techniques. Support vector machine (SVM) and random decision forest (RF) are supervised machine learning techniques, while K means, hierarchical agglomerative and divisive clustering are categorized as unsupervised machine learning

techniques, and we have implemented them to identify important PR sequences from a large pool of mutants. While some groups have tried to provide valuable insights into characterization of novel HIV drug resistance mutants using clustering, multidimensional scaling and SVM based feature ranking from relational clinical databases, ours attempts to select resistant sequences from existing pool [94]. We have generated a novel pipeline for such analysis as depicted in Figure 4.

Figure 4 Pipeline for analysis: Sequence of steps for the analysis are shown.



3.2 RESULTS

3.2.1 Efficient encapsulation of protein sequence and structure using Delaunay triangulation:

We have efficiently encoded a combined three-dimensional protein structure and sequence information of HIV-1 protease using a graph generated by Delaunay triangulation [95]. Structure template 3OXC [96] was used for HIV-1 PR. All the PR mutant sequences can be efficiently

represented using 210-dimensional vectors with a constant dimensionality. These structure vectors along with their respective resistance fold values were fed in SVM to check the efficiency of encapsulation. SVM is widely used as a supervised learning classifier in the classification area [97]. A clear delineation between resistant and non-resistant sequences through SVM can indicate encapsulation efficiency. In this experiment a three, five and ten fold validations were performed for SVM analysis in a linear kernel mode by implementing Python sklearn toolbox [98]. The results are shown in Table 4.

Table 4Confidence Intervals, P-Value, Sensitivity, Specificity, Positive and Negative Predictions, Balanced Accuracy and Standard Deviation for all the inhibitors from SVM and Random Forest (RF) analysis

Parameters ATV	Three Fold (SVM,RF)	Five Fold (SVM,RF)	Ten Fold (SVM,RF)
Accuracy	0.9742,0.9863	0.9709,0.99	0.9491,0.9964
95 Percent Confidence Interval	(0.9697, 0.9782),(0.983, 0.9891)	(0.9666, 0.9748),(0.9866, 0.9928)	(0.9439, 0.954),(0.9896, 0.9993)
P-Value	<2e-16,<2e-16	<2e-16,<2e-16	<2e-16,<2e-16
Sensitivity	0.9727,0.9878	0.9759,0.9957	0.9472,0.9976
Specificity	0.9757,0.9847	0.9661,0.9844	0.9511,0.9954
Positive Prediction Value	0.9765,0.9848	0.9661,0.9845	0.9512,0.9952
Negative Prediction Value	0.9719,0.9877	0.9758,0.9957	0.9471,0.9977
Balanced Accuracy	0.9742,0.9863	0.971,0.9901	0.9491,0.9965
Stddev FPV	4.10528,4.099989	4.10528,4.101648	4.10528,4.103662
Accuracy	0.9602,0.9998	0.9379,0.9995	0.9386,0.9772
95 Percent Confidence Interval	(0.9554, 0.9647),(0.9991, 1)	(0.9325, 0.9429),(0.9983, 0.9999)	(0.9336, 0.9434),(0.9646, 0.9862)

Parameters	Three Fold (SVM,RF)	Five Fold (SVM,RF)	Ten Fold (SVM,RF)
P-Value	<2e-16,<2e-16	<2e-16,<2e-16	<2e-16,<2e-16
Sensitivity	0.9482,1.0000	0.9278,1.0000	0.9266,1.0000
Specificity	0.969,0.9997	0.9451,0.9990	0.9474,0.9539

Positive Prediction Value	0.9569,0.9997	0.9233,0.9991	0.9273,0.9568
Negative Prediction Value	0.9626,1.0000	0.9483,1.0000	0.9469,1.0000
Balanced Accuracy	0.9586,0.9998	0.9364,0.9995	0.937,0.9769
Stddev IDV	4.104171,4.097711	4.104171,4.104209	4.104171,4.104669
Accuracy	0.9671,0.9821	0.9571,0.9826	0.9505,0.986
95 Percent Confidence Interval	(0.9627, 0.9712),(0.9783, 0.9853)	(0.9525, 0.9613),(0.9782, 0.9863)	(0.946, 0.9548),(0.9757, 0.9928)
P-Value	<2e-16,<2e-16	<2e-16,<2e-16	<2.2e-16,<2e-16
Sensitivity	0.9617,0.9966	0.9547,0.9963	0.9387,1.0000
Specificity	0.9723,0.9674	0.9593,0.9688	0.9617,0.9739
Positive Prediction Value	0.971,0.9685	0.9566,0.9700	0.9585,0.9708
Negative Prediction Value	0.9635,0.9965	0.9575,0.9961	0.9433,1.0000
Balanced Accuracy	0.967,0.9820	0.957,0.9825	0.9502,0.9869
Stddev DRV	4.105339,4.102193	4.105339,4.104663	4.105339,4.102954
Accuracy	0.9895,0.9806	0.9772,0.9827	0.9745,0.9764
95 Percent Confidence Interval	(0.9858, 0.9924),(0.977, 0.9838)	(0.9726, 0.9812),(0.9789, 0.986)	(0.9699, 0.9786),(0.9704, 0.9815)
P-Value	<2.2e-16,<2.2e-16	<2e-16,<2.2e-16	<2e-16,<2.2e-16
Sensitivity	0.996,0.9913	0.9683,0.9915	0.9643,0.9954
Specificity	0.9865,0.9717	0.9813,0.9771	0.9793,0.9733
Positive Prediction Value	0.9709,0.9665	0.9601,0.9652	0.9559,0.8603
Negative Prediction Value	0.9982,0.9927	0.9852,0.9945	0.9833,0.9992
Balanced Accuracy	0.9912,0.9815	0.9748,0.9843	0.9718,0.9843
Stddev NFV	4.108086,4.058676	4.108086,4.055835	4.108086,4.057195
Accuracy	0.9658,0.9921	0.953,0.9915	0.9496,0.9902
95 Percent Confidence Interval	(0.9613, 0.9698),(0.9895, 0.9942)	(0.9484, 0.9574),(0.9883, 0.9941)	(0.9451, 0.9538),(0.9808, 0.9958)
P-Value	<2e-16,<2e-16	<2e-16,<2e-16	<2e-16,<2e-16

Sensitivity	0.9692,0.9993	0.9559,0.9995	0.9538,1.0000
Specificity	0.9612,0.9848	0.9493,0.9831	0.9441,0.9805
Positive Prediction Value	0.9704,0.9851	0.9609,0.9842	0.9571,0.9808
Negative Prediction Value	0.9597,0.9993	0.9428,0.9995	0.9399,1.0000
Balanced Accuracy	0.9652,0.9920	0.9526,0.9913	0.9489,0.9902
Stddev LPV	4.107777,4.104829	4.107777,4.106061	4.107777,4.104499
Accuracy	0.9754,0.9901	0.965,0.9921	0.9639,0.9915
95 Percent Confidence Interval	(0.9714, 0.9789),(0.9872, 0.9925)	(0.9608, 0.9688),(0.989, 0.9946)	(0.9599, 0.9676),(0.9826, 0.9966)
P-Value	<2e-16,<2e-16	<2e-16,<2e-16	<2.2e-16,<2e-16
Sensitivity	0.9791,0.9993	0.9643,0.9995	0.9567,1.0000
Specificity	0.9718,0.9809	0.9656,0.9846	0.971,0.9828
Positive Prediction Value	0.971,0.9812	0.965,0.9851	0.9704,0.9836
Negative Prediction Value	0.9797,0.9993	0.9649,0.9995	0.9575,1.0000
Balanced Accuracy	0.9754,0.9901	0.965,0.9921	0.9639,0.9914
Stddev	4.106898,4.106167	4.106898,4.107035	4.106898,4.10363

Parameters TPV	Three Fold (SVM,RF)	Five Fold (SVM,RF)	Ten Fold (SVM,RF)
Accuracy	0.9701,0.9859	0.9523,0.9901	0.9538,0.9956
95 Percent Confidence Interval	(0.9649, 0.9746),(0.9826, 0.9886)	(0.9465, 0.9576),(0.9869, 0.9927)	(0.9485, 0.9588),(0.9913, 0.9981)
P-Value	<2e-16,<2e-16	<2e-16,<2e-16	<2e-16,<2e-16
Sensitivity	0.9512,0.9918	0.9067,0.9929	0.9136,1.0000
Specificity	0.9774,0.9805	0.9697,0.9879	0.9697,0.9942
Positive Prediction Value	0.9423,0.9785	0.9197,0.9849	0.9224,0.9816
Negative Prediction Value	0.981,0.9926	0.9645,0.9943	0.966,1.0000
Balanced Accuracy	0.9643,0.9862	0.9382,0.9904	0.9416,0.9971
Stddev SQV	4.108681,4.089722	4.108681,4.088172	4.108681,4.089285
Accuracy	0.9681,0.9988	0.9559,0.9986	0.9357,1.0000

95 Percent Confidence Interval	(0.9637, 0.9722),(0.9976, 0.9995)	(0.9512, 0.9602),(0.9969, 0.9995)	(0.9305, 0.9406),(0.9957, 1)
P-Value	<2e-16,<2e-16	<2e-16,<2e-16	<2.2e-16,<2e-16
Sensitivity	0.9622,1.0000	0.941,1.0000	0.9123,1.0000
Specificity	0.9722,0.9976	0.9664,0.9971	0.9519,1.0000
Positive Prediction Value	0.9595,0.9976	0.9517,0.9972	0.9295,1.0000
Negative Prediction Value	0.9741,1.0000	0.9588,1.0000	0.9398,1.0000
Balanced Accuracy	0.9672,0.9988	0.9537,0.9985	0.9321,1.0000
Stddev	4.112695,4.101274	4.112695,4.107813	4.112695,4.10336

We found a high accuracy, sensitivity, 95 percent confidence, positive and negative prediction values and specificity for all the tested inhibitors. For all the PIs the accuracy values range from a low of 0.96 to a high of 0.98, while sensitivity and specificity range from 0.95-0.99 and 0.96-0.98, respectively. A 95 percent confidence values ranged from 0.95-0.99, while the positive and negative prediction values ranged from 0.95-0.97 and 0.95-0.99, respectively. These values were consistent for all the tested fold validations. We further validated these SVM results with another supervised machine learning algorithm, random decision forest (RF) analysis. These high SVM and RF parameters conclusively separates the resistant and non-resistant data into two distinct categories thereby demonstrating an efficient encapsulation

3.2.2 *Hierarchical agglomerative clustering for classifying resistance mutations:*

Hierarchical agglomerative clustering is a ’bottom up’ approach, where each observation starts in its own cluster, followed by a continuous merging with upwards hierarchical movement [99]. The choice of an appropriate metric is important in influencing the shape of the clusters. Distance plays an important factor in cluster distribution and various distance techniques such as euclidean, squared euclidean, manhattan, maximum and mahalanobis distances have been implied

in such clustering [100]. In this analysis, an manhattan or euclidean metric seems appropriate as we are interested in an 'absolute magnitude', identifying vectors having similar mean values [101, 102]. Although, the run-time of agglomerative and divisive clustering are exhaustive and slow with O(n 2 log(n)) and O(2n) complexities [103], we didn't had any run-time issues with our data on an IBM system x3850 X5 server. Hierarchical agglomerative clustering starts with an most disparate observation which initiates a splinter group, which subsequently reassigns observations that are closer to the splinter group than to the old party and results in a division of the selected cluster [102]. Euclidean distances were calculated using function 'dist{stats}' from cluster package in R [103]. This function computes and returns the distance matrix computed between the rows of a data matrix [104]. Hierarchical cluster analysis was performed using function 'hclust{stats}' from cluster package in R [105]. Agglomeration method of 'complete linkage' for finding similar clusters was utilized. With complete linkage, for 'n' observations there are 'n-1' merges with '2n−1 ' possible orderings for the leaves in a dendrogram [102]. Figure 5 depicts the sequences clustered with hierarchical clustering, interestingly most of the high resistance fold sequences with class 2 were clustered in first 10 clusters for all the selected inhibitors, delineating a clean separation between non-resistant and resistant sequences.

Figure 6 Hierarchical and divisive clustering on sequences: Most of the high resistance fold sequences with class 2 were clustered in first 10 clusters for all the selected inhibitors through both hierarchical and divisive clustering delineating a clean separation between non-resistant and resistant sequences.

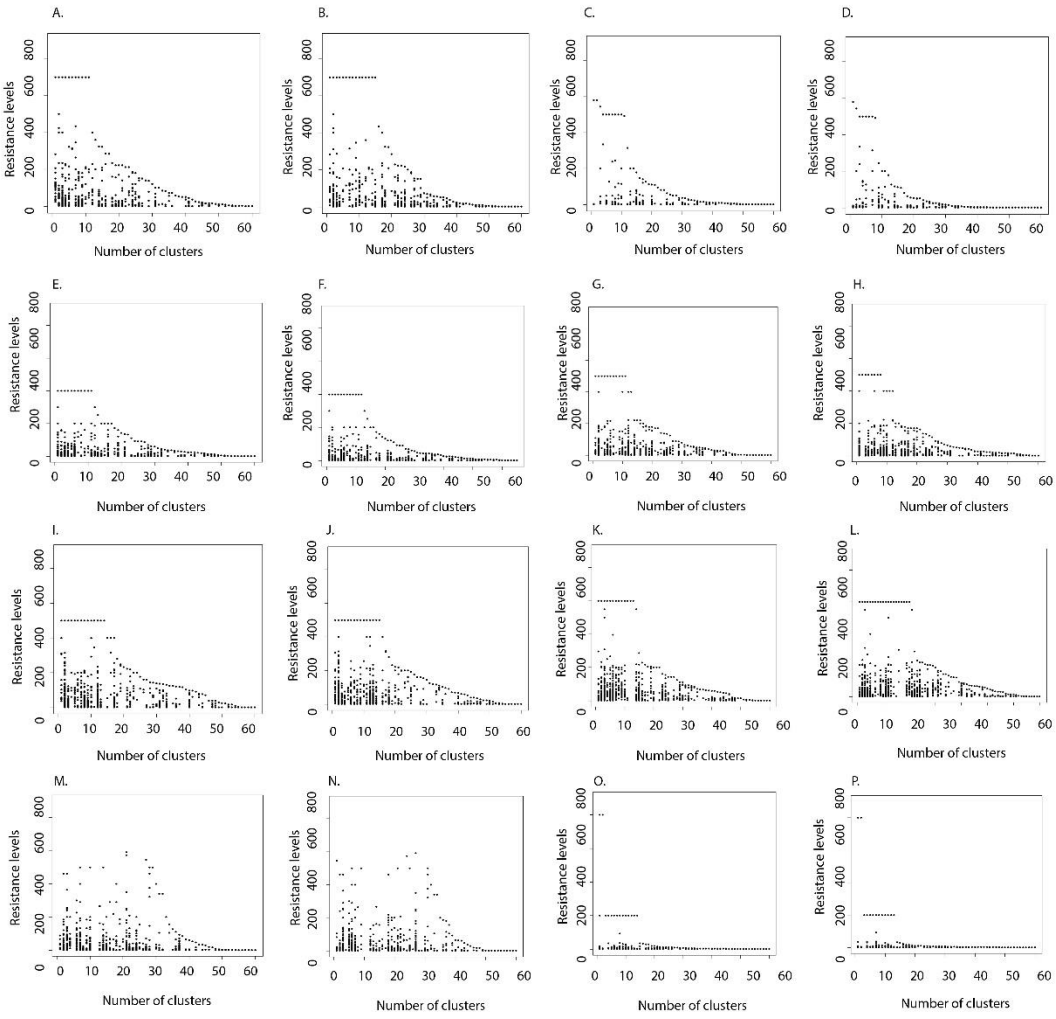


Table 5 shows results from hierarchical agglomerative clustering. We found a total of 1433, 657, 4349, 3018, 7616, 2594, 2092 and 1070 sequences in first 10 clusters for each of the ATV, DRV, FPV, IDV, LPV, NFV, SQV and TPV inhibitors.

Table 5 Number of sequences for each of the 3 classes clustered via Hierarchical (H) clustering for first 10 clusters

Cluster Number	Class	ATV H	DRV H	FPV H	IDV H	LPV H	NFV H	SQV H	TPV H
Total		1433	657	4349	3018	7616	2594	2092	1070
1	0	0	72	50	6	0	56	63	69
	1	88	4	559	50	130	16	24	35
	2	10	8	41	6	0	19	4	1
2	0	11	1	16	0	45	188	24	35
	1	212	29	103	0	231	502	84	1
	2	41	2	73	8	0	213	8	2
3	0	22	0	1877	16	570	4	239	0
	1	229	26	150	4	253	280	181	0
	2	11	19	8	0	0	341	135	16
4	0	19	2	10	422	24	32	0	97
	1	21	15	6	102	44	369	8	87
	2	3	7	8	0	0	11	4	10
5	0	0	0	266	3	0	30	8	8
	1	13	70	465	16	3	205	487	8
	2	7	3	6	0	0	22	71	8
6	0	0	0	82	16	1	1	52	306
	1	0	0	70	18	136	29	227	8
	2	10	32	7	0	0	5	40	1
7	0	2	0	0	0	36	0	125	133
	1	294	24	7	0	9	34	204	120
	2	132	5	3	0	0	4	28	7
8	0	2	0	68	2005	3	0	17	9
	1	17	0	316	141	20	8	4	8
	2	18	128	5	0	0	4	12	3
9	0	16	88	8	26	227	0	0	2
	1	90	91	22	131	157	77	17	4
	2	57	5	24	0	0	49	10	12

10	0	0	15	0	19	0	54	0	2
	1	26	9	66	29	2	38	0	10
	2	82	2	33	0	0	3	16	68

3.2.3 Divisive clustering for classifying resistance mutations:

Divisive clustering is a ’top down’ approach, where all the observations start in one cluster and splitting occurs recursively as one moves down the hierarchy [102]. Divisive cluster analysis was performed using function ’diana{cluster}’ from cluster package in R [104]. This algorithm constructs a hierarchy of clusterings, starting with one large cluster containing all ’n’ observations with divisions until each cluster contains only a single observation [104]. The logical flag was set to true, as the input matrix was considered a dissimilarity matrix obtained from vectors euclidean distances. Table 6 shows results from divisive clustering.

Table 6 Number of sequences for each of the 3 classes clustered via Divisive (D) clustering for first 10 clusters

Cluster Number	Class	ATV D	DRV D	FPV D	IDV D	LPV D	NFV D	SQV D	TPV D
Total		2125	771	4152	2798	5774	4059	3273	1210
1	0	0	6	0	0	0	4	0	29
	1	367	1	162	46	294	26	31	55
	2	136	4	37	8	249	8	1	13
2	0	10	0	18	35	7	236	41	26
	1	104	0	87	685	356	502	295	1
	2	41	4	30	303	643	222	154	2
3	0	33	0	64	0	648	29	243	0
	1	264	17	864	4	301	637	258	0
	2	33	2	7	32	8	356	78	16
4	0	6	0	1967	16	0	4	0	1

	1	61	0	291	150	4	94	454	10
	2	13	16	10	19	53	33	90	4
5	0	25	136	10	0	0	32	0	38
	1	138	363	291	145	80	172	0	34
	2	23	1	8	4	7	10	2	11
6	0	0	0	101	452	36	48	20	134
	1	13	0	345	277	40	606	81	337
	2	8	3	9	11	124	76	42	5
7	0	38	32	0	0	20	8	46	32
	1	76	34	1	277	64	66	209	6
	2	108	3	7	34	76	31	34	1
8	0	4	45	4	0	5	0	5	126
	1	145	13	16	1	90	306	26	82
	2	270	17	7	2	27	231	8	37
9	0	5	2	0	0	2245	0	36	45
	1	44	32	15	46	255	86	81	42
	2	89	17	17	10	7	36	50	1
10	0	0	8	0	0	4	0	481	52
	1	18	10	18	419	85	117	463	34
	2	53	5	8	27	46	83	44	36

We found a total of 2125, 771, 4152, 2798, 5774, 4059, 3273 and 1210 sequences in first 10 clusters for each of the ATV, DRV, FPV, IDV, LPV, NFV, SQV and TPV inhibitors.

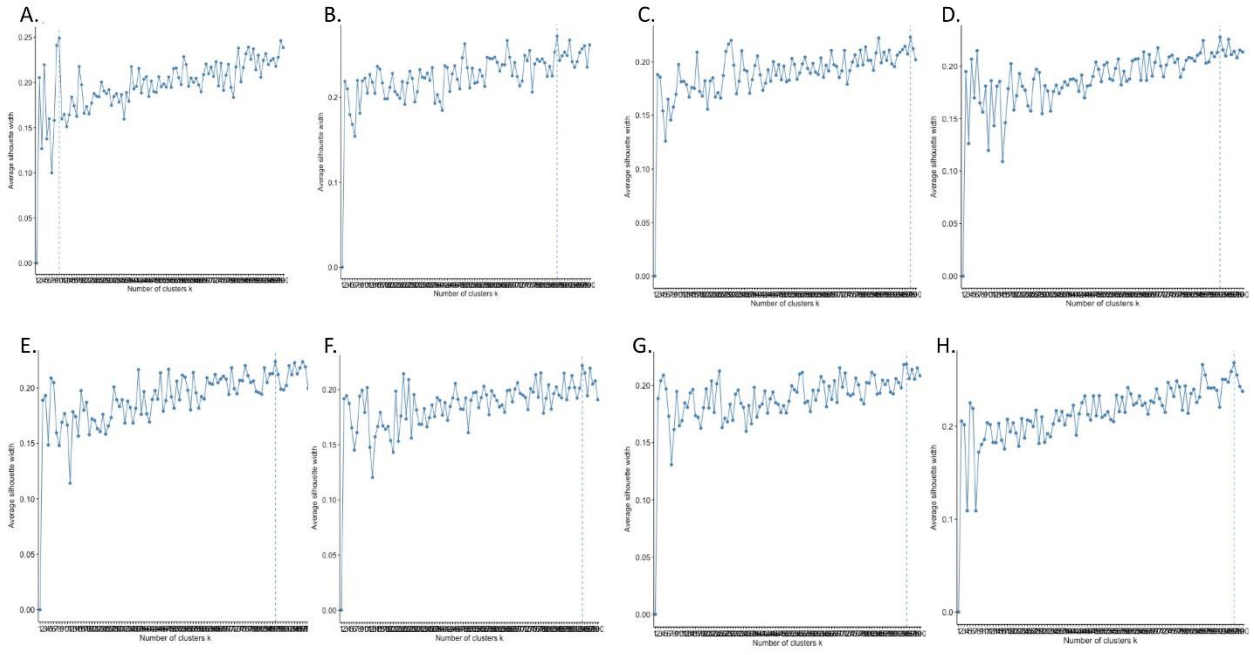
3.2.4 K means clustering for classifying resistance mutations:

K means method aims to partition the points into K groups, such that the sum of squares from points to the assigned cluster centre's is minimized, it at-most tries to put all cluster centre's at the mean of their Voronoi sets [105]. For a set of observations {x1, x2, ..., xn}, K means

clustering aims to partition the ’n’ observations into K sets, S = {S1, S2, ..., Sk}, so as to minimize the within-cluster sum of squares (WCSS), where μi is the mean of points in set Si [103].

K means analysis was performed using function ’kmeans{stats}’ from cluster package in R [104]. The centers or the number of clusters, K were chosen with the silhouette method with K maximum of 100. Figure 5 shows the optimal cluster cutoffs for each of the inhibitors. We also tried other techniques for validating optimal cluster determination by silhouette, some of them were elbow method, clusGap, Mclust, and cascadeKM.

Figure 5 Silhouette method for cluster cutoffs: Optimal cluster cutoffs for each of the inhibitors. A: ATV, B: DRV, C: FPV, D: IDV, E: LPV, F: NFV, G: SQV, and H: TPV



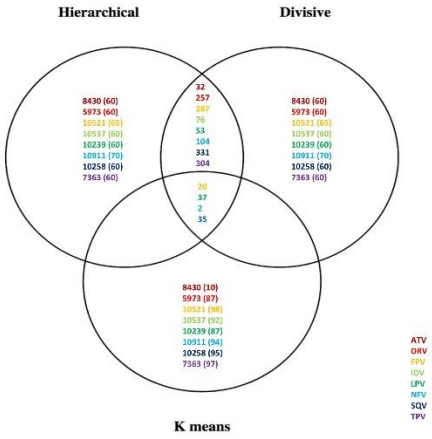
3.2.5 Sequences/mutations selected from intersection between hierarchical agglomerative, divisive and K means clustering:

Table 7 and Figure 6 shows the sequences selected from intersection between hierarchical agglomerative, divisive and K means clustering.

Table 7Number of sequences common via hierarchical (H), divisive (D) and K means (K) clustering. Cluster numbers shown in brackets

Category	ATV	DRV	FPV	IDV	LPV	NFV	SQV	TPV
H and D	32	257	287	76	53	104	331	304
H, D and K	0	0	20 (66)	0	35 (61) , 2 (31)	2 (12)	5 (58), 6 (63), 3 (73), 21 (85)	0

Figure 7 A Venn diagram showing number of mutations found common via hierarchical, divisive and K means clustering: Number of mutations found common ivia hierarchical, divisive and K means clustering. Optimal cluster numbers are shown in brackets



Different cluster combinations between hierarchical agglomerative and divisive were performed as shown in Figure 4. Each cluster of hierarchical was intersected with each cluster of divisive, the cluster combination giving maximum common sequences was selected to intersect with K means clusters. With intersections between hierarchical agglomerative and divisive clustering, for ATV

D and ATV H, cluster combination 1659 gave maximum number of common sequences (94), belonging to class 1 with resistance fold of 55. DRV D and DRV H, cluster combination 1711 gave maximum number of common sequences (38), belonging to classes 0, 1 with resistance folds of 0.4 and 51. FPV D and FPV H, cluster combination 198 gave max number of common sequences (149), belonging to classes and resistance folds of 1, 23; 1, 53; 1, 70; 2, 400; and 0, 0.3. IDV D and IDV H, cluster combination 1113 gave max number of common sequences (146), belonging to classes and resistance folds of 1, 15; 2, 186.9 and 2, 500. LPV D and LPV H, cluster combination 538 gave max number of common sequences (104), belonging to classes and resistance folds of 0, 2.3; 1, 3.2; 1, 21; 1, 76; 1, 82; 2, 100.7; 2, 126; and 2, 280.0. NFV D and NFV H, cluster combination 932 gave max number of common sequences (62), belonging to classes and resistance folds of 1, 48; 1, 68; 1, 72; 1, 78; 1, 97.2; and 2, 286. SQV D and SQV H, cluster combination 1327 gave max number of common sequences (132), belonging to classes and resistance folds of 0, 1; 0, 1.8; 1, 14.0; 1, 14.0; 1, 33; 1, 56; 1, 93; 2, 181; 2, 242; 2, 249.8; 2, 350.0; 2, 591.5; and 2, 1000. TPV D and TPV H, cluster combination 793 gave max number of common sequences (76), belonging to classes and resistance folds of 2, 200; and 0, 0.4. These sequences were then intersected with clusters from K means clustering, and 20 (cluster 66), 37 (Cluster 61, 31), 2 (Cluster 12), and 35 (Cluster 58, 63, 73 and 85) sequences felt in common clusters for each of the FPV, LPV, NFV and SQV inhibitors.

3.3 Materials and methods

3.3.1 Datasets used for the study:

The genotype-phenotype datasets were downloaded from the Stanford HIV drug resistance database [106] (http://hivdb.stanford.edu/pages/genopheno.dataset.html). For HIV-1 PR, the inhibitors atazanavir (ATV), nelfinavir (NFV), ritonavir (RTV), indinavir (IDV), lopinavir (LPV), tipranvir (TPV), saquinavir (SQV), and darunavir (DRV) were tested. All the datasets were pre-processed using the methods and the cutoff values described previously in [93]. The results of the expansion for each of the HIV-1 PR inhibitors were: for SQV 10258 sequences of 1722 isolates, DRV 5973 sequences of 607, LPV 10239 sequences of 1444, NFV 10911 sequences of 1771, IDV 10537 sequences of 1730, ATV 8430 sequences of 1141, FPV 10521 sequences of 1681 and TPV 7363 sequences of 847.

3.3.2 Pre-processing/Expansion of the datasets:

Insertion, deletion, or presence of stop codon in sequences has been removed for data unification. A sequence with 99 amino acids representing protease HIV-1 PR has been used for this data. Datasets usually have multiple mutations at the same sites; these datasets were expanded to multiple sequences with single amino acids at each location to represent a single amino acid sequence for each mutant protein. For example, if one 99-amino acid mutant sequence has two different types of amino-acids, and another site has three, this one sequence needs to be represented in six different sequences each having only one amino-acid substitution. The time complexity to minimize such calculations have been explained in detail [93] and been applied for pre-processing this dataset. Each of these sequences were accompanied by their inhibitor resistance fold values. The relative resistant fold values for sequences in each of the inhibitors ranged from 0-800-fold resistance. Furthermore, the expanded datasets with sequences have been allotted a unique

identifier number which will help to trace back the sequences and its resistance fold change after analysis.

3.3.3 Encoding structure and sequence with Delaunay triangulation:

A graph-based encoding system was utilized to represent the sequence and structural information of the protein [95]. The structurally adjacent pairs of amino acids were represented in form of a vector of the 210 unique pairs of 20 standard amino acids. A amino acid type for each adjacent pair in structure was the parameter utilized for calculating all the 210 vectors. Such graph-based encoding of sequence and structure have been proven to be an promising technique for faster and accurate predictions of resistance from sequence in HIV infections [106]. The X-ray crystal structure for HIV-1 PR (3OXC) was used as a template for creating Delaunay triangulation [96].

3.3.4 Cutoffs for resistance/susceptibility for each drug:

For SVM and RF analysis, for all the HIV-1 PR inhibitors those mutants with the relative resistant fold > 3.0 were classified as resistant, denoted as 1. For K means, hierarchical and divisive clustering, relative resistant 99 denoted as class 1 and anything above relative resistant fold of 100 were classified as class 3.

3.3.5 SVM and RF analysis:

To understand the efficiency of Delaunay triangulation encoding, randomly chosen (k-1)/k sequences (some are drug resistant, while others are non-drug resistant) were selected for training the classifier and the remaining 1/k sequences were used for testing. The SVM and RF parameters

for each of the inhibitors are shown in Table 4. A three, five and ten fold validation was performed for both RF and SVM analysis.

3.3.6 Clustering analysis with K means, Hierarchical and Divisive clustering:

Optimal number of clusters for K means clustering were selected through silhouette method [106]. 10, 87, 98, 92, 87, 94, 95, and 97 clusters were found to be optimal for sequences with ATV, DRV, FPV, IDV, LPV, NFV, SQV and TPV inhibitors. Figure 4 shows the optimal cluster cutoffs for each of these inhibitors. For divisive clustering the height of 60 was found to be optimal for all the inhibitors. For hierarchical clustering, the height of 60 was also found to be optimal for ATV, DRV, IDV, LPV, SQV and TPV inhibitors, while 65 and 70 heights were found to be optimal for FPV and NFV inhibitors. Hierarchical and divisive clustering was performed with resistance levels on 210 vectors for each of the sequences. Table 5 and Table 6 shows number of sequences for each of the 3 classes clustered in first 10 clusters.

3.4 DISCUSSION

Drug resistance in HIV-infected individuals is a serious problem and there are numerous mutations that can occur in any of the 99 amino acids of HIV-1 protease sequence. Thus, making it extremely difficult to select representative mutants for detailed research in the laboratory. In this study, we have effectively encoded HIV-1 protease sequence and structure information in a vector by using Delaunay triangulation. The effectiveness is tested by two supervised machine learning algorithms. We then have utilized 3 un-supervised machine learning algorithms for identifying most representative mutants. By intersecting sequences with the 2 clustering techniques, 32, 257, 287, 76, 53, 104, 331, and 304 sequences felt common with hierarchical and divisive clustering

for each of the ATV, DRV, FPV, IDV, LPV, NFV, SQV and TPV inhibitors. Further, intersecting these with the K means clustering selected only 20, 37, 2 and 35 sequences for FPV, LPV, NFV and SQV inhibitors. This approach selected sequences with L10F, I13V, L33F, E35D, M36I, N37D, I54V, I84V and L90M, L10I, M46I, L63P and A71V mutations. Some important characteristics of these resistant mutations from selected sequences are summarized below. With L10F mutation, a hydrogen bond that is crucial for effective inhibition is lost with NFV [109]. L10I mutation impairs replication complex and confers resistance to SQV [110]. L33F increases non-covalent interactions in the hydrophobic pocket, it also increases rigidity of the 30s and 80s loops compared to wild type protease with expansion of the S1/S1 prime sub site which alters DRV binding [111]. E35D mutation is seen to increase flexibility of the flaps affecting the conformational equilibrium between the closed and semi open conformations of the free protease [112]. Double mutant (V82A and L90M) both confer strong resistance to ritonavir (RTV), and the molecular dynamic simulations reveals a contraction of the ligand binding pocket which is enhanced by I54V mutation [113]. A71V and M36I are also important secondary mutations associated with resistance [114]. These sequences will enable laboratory studies of the molecular mechanisms of resistance.

4 STUDIES OF FOUR INVESTIGATIONAL PROTEASE INHIBITORS WITH HIV-1 PROTEASE BEARING DRUG RESISTANT SUBSTITUTIONS V32I, I47V AND V82I*

*In the process of submission

Abstract

Antiviral inhibitors of HIV-1 protease are effective in HIV/AIDS therapy, although drug resistance is a severe problem. This study examines the effects of four investigational inhibitors against HIV-1 protease with drug resistant mutations of V32I, I47V and V82I (PRTri) that models the inhibitor-binding site of HIV-2 protease. Inhibitors GRL-0249, GRL-0519, GRL-0739 and GRL-1111 introduce diverse chemical modifications on the darunavir scaffold, and form new interactions with wild type protease. The measured inhibition constants for PRTri mutant range from 17-40 nM or three-orders of magnitude worse than for wild type enzyme. The X-ray crystal structure of PRTri mutant in complex with GRL-1111 determined at 1.5 Å resolution shows minor changes in interactions with inhibitor compared with the corresponding wild type PR complex. Instead, mutation to Ile82 induces two alternate conformations for the side chain of Arg8 with new interactions with inhibitor and Leu10.

4.1 Introduction

HIV/AIDS is a pandemic disease with about 37 million people infected worldwide [115]. HIV infection can be controlled by antiviral drugs targeting different stages of viral replication,

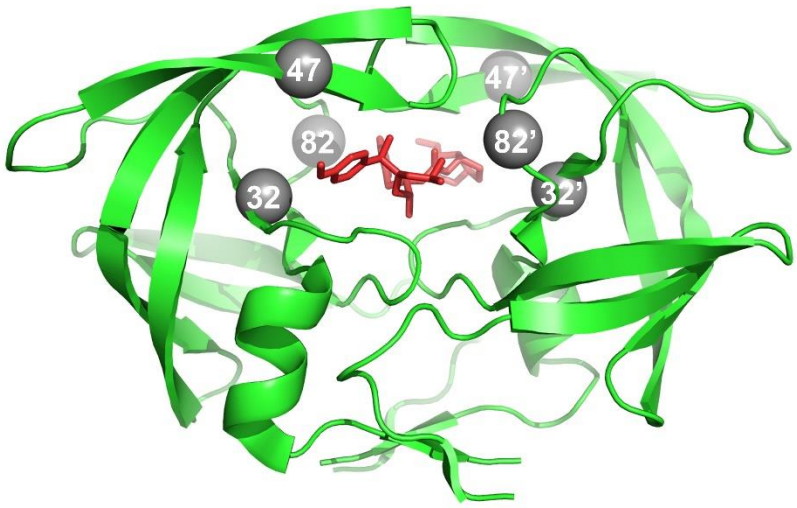
however, the genetic diversity of the virus and rapid selection of drug resistant strains pose a severe challenge [116,117]. The retrovirus HIV includes two types, HIV-1 and HIV-2, and HIV-1 is subdivided into three groups (M, O, and N) and subtypes with different geographical distribution. HIV-2 infections are common in West Africa, however, some drugs designed for HIV-1 are not effective for HIV-2 infections [118].

One important class of antiviral drugs targets the viral protease (PR), which is crucial for production of infectious virus. PR processes cleavage sites in Gag-Pol region during viral maturation to produce individual structural proteins [119]. This aspartic protease is catalytically active as a dimer of 99-residue monomers [120]. Clinical inhibitors bind in the active site cavity of the enzyme and block binding of substrates. Some HIV-1 PR inhibitors, including amprenavir (APV), are significantly less potent against HIV-2 infections probably because the amino acid sequences of HIV-1 and HIV-2 PRs show only about 40% sequence identity [121]. In addition to its natural genetic diversity, the virus has evolved resistance mutations for all clinical protease inhibitors (PIs) [122]. Single, major mutations decrease binding of inhibitor, however, they can be deleterious for viral replication. Resistance mutations can alter the catalytic activity, binding affinity and stability of PR [9,10]. The virus evolves additional mutations that compensate by restoring effective viral replication in the presence of inhibitor [123].

HIV-1 PR with drug resistant mutations V32I, I47V and V82I (PR_{Tri}) has been evaluated as a model for inhibition of HIV-2 PR. The three mutations in PR_{Tri} alter residues in the inhibitor-binding cavity (Figure 8) and represent the changes in the inhibitor binding site of HIV-2 protease. APV showed poor inhibition of both PR_{Tri} and HIV-2 PR at 15- and 19-fold worse than for HIV-

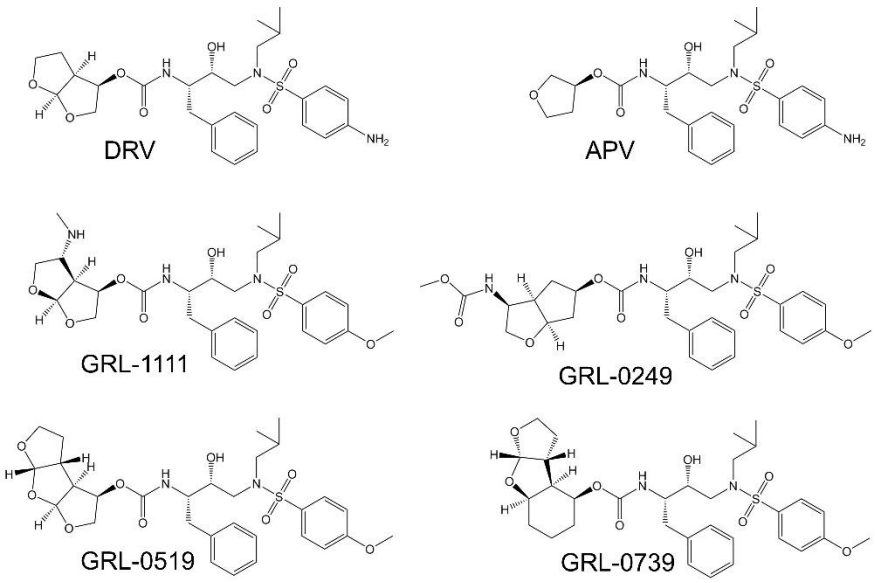
1 PR, while darunavir (DRV) and saquinavir (SQV) were effective inhibitors of all three enzymes [124]. Moreover, the individual mutations of V32I, I47V and V82I are associated with resistance to one or more clinical inhibitors [122]. Hence, PR_{Tri} was chosen to evaluate the efficacy of investigational inhibitors for drug-resistant mutants and HIV-2 PR.

Figure 8: Sites of mutation in HIV-1 PR dimer with inhibitor GRL-1111. PR is shown as green ribbons with sites of the three mutations, V32I, I47V and V82I, indicated as grey spheres, GRL-1111 is shown in red bonds.



The four investigational antiviral inhibitors were designed on the DRV scaffold (Figure 9) with the aim of introducing new interactions of the P2 group with PR. They exhibited potent antiviral activity on wild type and drug resistant strains of HIV-1.

Figure 9: Structures of the clinical and investigational inhibitors used in this study.



Inhibitor GRL-0519 contains *tris*-tetrahydrofuranylurethane (*tris*-THF) as the P2 ligand instead of *bis*-THF in DRV [127-129]. The third THF group introduces new water-mediated hydrogen bond interactions with Gly27 and Arg87 in the PR dimer interface. GRL-1111 has a novel P2 *bis*-THF group with a basic amine that forms direct and water-mediated hydrogen bond interactions with the main chain carbonyl oxygen and amide of Gly48 in the PR flap [130]. GRL-0739 has a novel tricycle cyclohexyl-bis-tetrahydrofuranylurethane at P2, and resembles GRL-0519 in its new water-mediated interactions [131]. This inhibitor showed favorable penetration of the central nervous system (CNS) in an *in vitro* model [18]. GRL-0249 has a 3-(*S*)-*N*-methoxycarbonyl amino substituted cyclopentyltetrahydrofuranyl (Cp-THF) at the P2 group [132].

The carbamate NH forms hydrogen bonds interactions with the main chain of Gly48 similar to those of GRL-1111 and the carbamate carbonyl has a water-mediated interaction with the guanidinium group of Arg8'. Crystal structures confirmed the presence of new interactions of wild type PR with the P2 groups of these inhibitors compared to DRV. Therefore, we hypothesized they would show good inhibition of PR_{Tri}. Inhibition constants were measured for the four inhibitors, and a high resolution X-ray structure was determined for PR_{Tri} in complex with GRL-1111.

4.2 Materials and methods

4.2.1 Inhibitors

Amprenavir (APV) (HPLC purity of 99.7%) was obtained from the AIDS Reagent Program, Division of AIDS, NIAID, NIH. Compounds GRL-0519, GRL-1111, GRL-0739 and GRL-0249 (>95.0% purity by HPLC) were provided by Dr. Arun Ghosh at Purdue University.

4.2.2 Protein purification

The clone for triple mutant PR_{Tri} (V32I, I47V and V82I) includes optimizing mutations of Q7K, L33I, and L63I to decrease autoproteolysis, and C67A, C95A to eliminate cysteine-thiol oxidation [134]. Protein was expressed in E. coli BL21 and purified from inclusion bodies as described previously [135,136] using gel filtration followed by reverse phase chromatography and refolding. Samples were concentrated to 5.0 mg/mL for crystallization or diluted for kinetic assays.

4.2.3. Protein crystallography

PR_{Tri} was mixed with inhibitor (dissolved in dimethylsulfoxide) at a molar ratio of 1:5. Crystals were grown at room temperature by hanging drop vapor diffusion. Each drop contained 1 μL protein and 1 μL reservoir solution. Crystals of PR_{Tri}-GRL-1111 were grown from 0.1 M sodium acetate, pH 4.6, and 2 M NaCl. Crystals were cryo-protected in 30% glycerol and flash frozen in liquid nitrogen. X-ray diffraction data were collected on the SER-CAT 22ID beamline, Advanced Photon Source, Argonne National Laboratory (Argonne, IL), and processed using HKL-2000 [137]. The structure was solved by molecular replacement with PR complex with APV (3NU3) [138] using CCP4i Phaser [139,140]. The structure was refined with SHELX-2014 [141], followed by REFMAC5 [141]. COOT [143] was used for visualization and refitting. Alternate conformations were modeled according to the electron density maps. Anisotropic B factors were applied in the refinement. Structural figures were made using PyMOL [144]. Atomic coordinates and structure factors for PR_{Tri} with GRL-1111 have been deposited in the PDB [145] with ID: XXXX.

4.2.4. Enzyme kinetic assays

Kinetic parameters of PR_{Tri} were determined in 3-5 replicate runs by monitoring hydrolysis of fluorescence substrate derived from the HIV-1 p2/NC cleavage site: Abz-Thr-Ile-Nle-*p*-nitro-Phe-Gln-Arg-NH_2 (BACHEM H-2992) (where Abz is anthranilic acid, Nle is norleucine, and *p*-nitro-Phe is *p*-nitrophenylalanine). Samples were equilibrated at 37 °C for 5 minutes prior to initiating the reactions. Enzyme activity was measured at 37 °C using a PolarStar Optima microplate reader (BMG Labtech) with excitation wavelength at 340 nm and emission wavelength at 420 nm, as described previously [138,146]. To determine catalytic efficiency, 10 μL of PR_{Tri}

(final well concentration of 40-100 nM determined by active site titration with APV) was mixed with 100 µL reaction buffer (100 mM MES pH 5.6, 400 mM NaCl, 1 mM EDTA, and 5% glycerol). Reaction was initiated by adding 100 µL substrate (12-96 µM final concentration). Initial velocities (V_0) were determined using MARS software (BMG Labtech). The K_m and k_{cat} were determined by fitting data to the Michaelis-Menten plot of V_0 vs [substrate].

For inhibition studies, 10 µL PR_{Tri} was mixed with 98 µL reaction buffer and 2 µL inhibitor in DMSO (final well concentration 0-40 µM). Reaction was initiated with 90 µL substrate (final well concentration 60 µM). IC_{50} values were determined using SigmaPlot (Systat Software) by non-linear regression curve fitting to a dose-response plot of V_0 vs [inhibitor]. K_i values were calculated using the equation for tight-binding inhibitor of $K_i = (IC_{50} - 0.5[E]) / (1 + [S]/K_m)$ [147].

For the urea denaturation assay, 10 µL PR_{Tri} was mixed with 100 µL reaction buffer containing 8 different urea concentrations (0-4 M). Reaction was initiated with 90 µL substrate (final well concentration 60-72 µM) in 0-4 M urea. The urea concentration in each well remained the same throughout the experiment. A plot of V_0 vs [urea] was constructed using SigmaPlot (Systat Software) and sigmoidal curve fitting used to determine the urea concentration for 50% maximum velocity (UC_{50}).

4.3 Results

4.3.1 Kinetic Parameters, Inhibition and Stability of PR_{Tri} mutant

Kinetic parameters were determined for PR_{Tri} hydrolysis of a fluorescent substrate analog at 37 °C prior to assaying the effects of inhibitors under the same conditions. PR_{Tri} mutant had k_{cat}

of 299 ± 48 μM/min and K_m of 72 ± 17.7 μM. The catalytic efficiency (k_{cat}/K_m) of 4.2 μM-1 min-1 is similar to that value of 6.5 μM-1 min-1 for wild type enzyme at 26 °C [147].

K_i values were measured for investigational inhibitors (Table 8). Clinical inhibitor APV gave a K_i value of 2.5 nM for PR_{Tri} mutant, which is identical to the value reported previously using a different peptide substrate in an HPLC assay [126]. APV showed 16-fold worse inhibition of the mutant compared to wild type PR, as reported earlier using a different assay [126]. These investigational inhibitors, however, had K_i values ranging from 16.7 to 39.5 nM for the mutant. GRL-0249 with K_i of 16.7 nM was the best of the four inhibitors, and GRL-0519 and GRL-0739 were the worst with K_i values of 38.4 and 39.5 nM, respectively. The tested inhibitors were significantly less effective for the mutant compared to K_i values of 2-10 pM reported for the wild type enzyme [128,130,132,133].

The stability of the mutant and wild type enzymes were assessed by measuring enzyme activity under urea denaturation. A UC_{50} value of 0.97 +/- 0.05 M was obtained for PR_{Tri} mutant. In comparison, the wild type enzyme gave a UC_{50} value of 0.70 +/- 0.07 M under the same conditions. These values suggest the mutant dimer is somewhat more stable than the wild type protease.

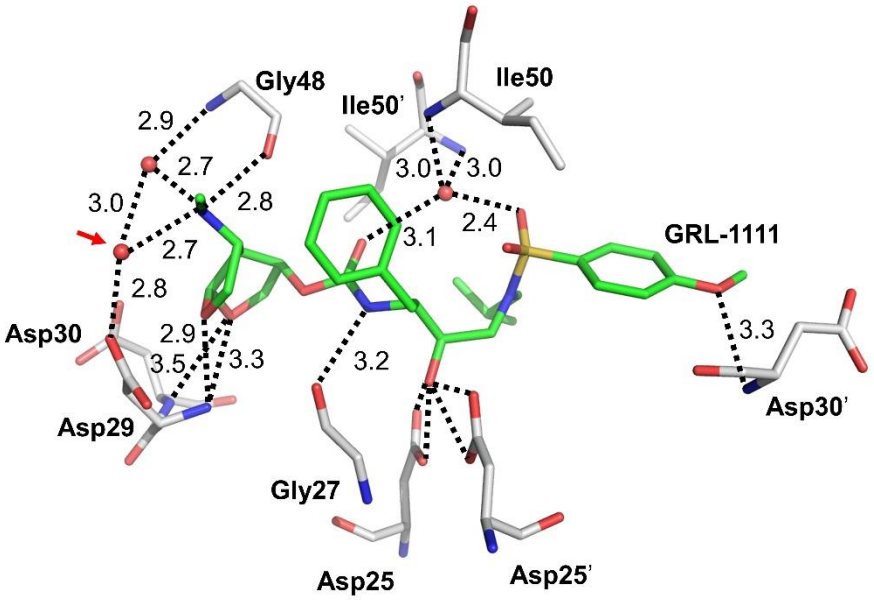
4.3.2 Crystallographic analysis of PR_{Tri} complex with GRL-1111

The crystal structure of PR_{Tri} with GRL-1111 was solved at 1.50 Å resolution in the $P2_12_12$ space group, and refined to R/R_{free} values of 13.4/17.2 %. Crystallographic statistics are listed in

Table 9. The asymmetric unit of the crystal structure contained a dimer of PR_{Tri}, and the inhibitor, GRL-1111, was bound at the active site in two alternate orientations with an occupancy ratio of 0.5/0.5. The solvent was modeled with 133 water, 1 glycerol and 4 formic acid molecules from the crystallization and cryo-protectant solutions. The crystal structure of PR_{Tri}/GRL-1111 was compared with the corresponding complex of wild type PR (PDB ID: 5BRY), which was determined at 1.34 Å resolution in the same space group with isomorphous unit cell dimensions and contained PR dimer with inhibitor bound in two orientations with 0.6/0.4 relative occupancy [130]. The two dimer structures superimposed with a low RMSD value of 0.23 Å on Cα atoms.

The interactions of PR_{Tri} with inhibitor were analyzed in comparison to those in wild type PR complex. GRL-1111 forms six direct hydrogen bonds with PR_{Tri}, excluding the interactions of the central hydroxyl with the catalytic Asp25 and 25' (Figure 10).

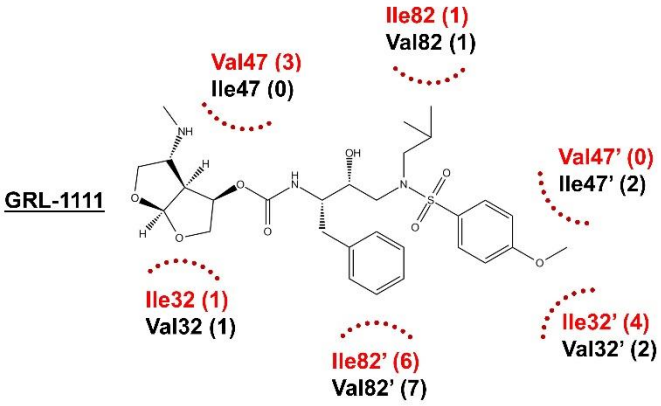
Figure 10: Hydrogen bond interactions of PR_{Tri} with inhibitor GRL-1111. Protein is shown as grey bonds, inhibitor in green bonds. Dotted lines indicate hydrogen bond interactions with interatomic distances in Å.



Additionally, it forms water-mediated interactions with flap residues Gly48, Ile50 and Ile50', while one orientation of inhibitor (designated as the "major" conformation) has a water-mediated interaction with the carboxylate side chain of Asp29. PR-inhibitor hydrogen bond interactions are conserved in the mutant and wild type complexes with differences of no greater than 0.2 Å in length for the major inhibitor conformations. GRL-1111 was designed to incorporate an amine on *bis*-THF at P2 that forms a new direct hydrogen bond interaction with the carbonyl oxygen of Gly48, a water-mediated interaction with the amide of Gly48 in the flap region, and a second water-mediated interactions with the side chain of Asp29. These interactions of GRL-1111 cannot occur with clinical inhibitor DRV, which lacks the basic amine. However, the minor inhibitor conformation of GRL-1111 in PR_{Tri} complex has lost the second water and its interaction with Asp29.

The hydrophobic side chains of residues 32, 47 and 82 are important components of the PR binding site for substrates or inhibitors. In wild type PR/GRL-1111, Val82 interacts with the P1 and P1' groups of inhibitors, while Val32 and Ile47 contribute to the binding site for P2 and P2' groups (Figure 11).

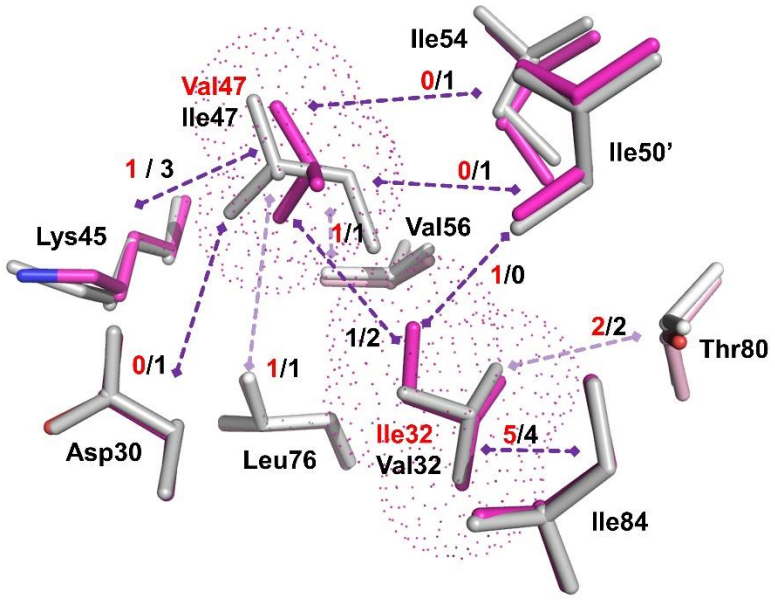
Figure 11: The hydrophobic side chains of residues 32, 47 and 82.



Ile82 in the mutant has similar interactions with inhibitor P1 and P1' as seen for Val82 of wild type enzyme. The P2 group of GRL-1111 has similar interactions with Val32 and Ile32, while Val47 in the mutant has shifted to form new interactions with P2 compared to Ile47 in the flap. The P2' group has gained van der Waals interactions with Ile32' and lost contacts with Val47' relative to those in the wild type structure.

In addition to the contacts with inhibitors, these residues interact with adjacent PR side chains. The side chains of residue 32 and 47 form hydrophobic contacts with each other and with neighboring side chains, as shown in Figure 12.

Figure 12: Figure depicting side chains of residue 32 and 47 forming hydrophobic contacts with each other and with neighboring side chains.



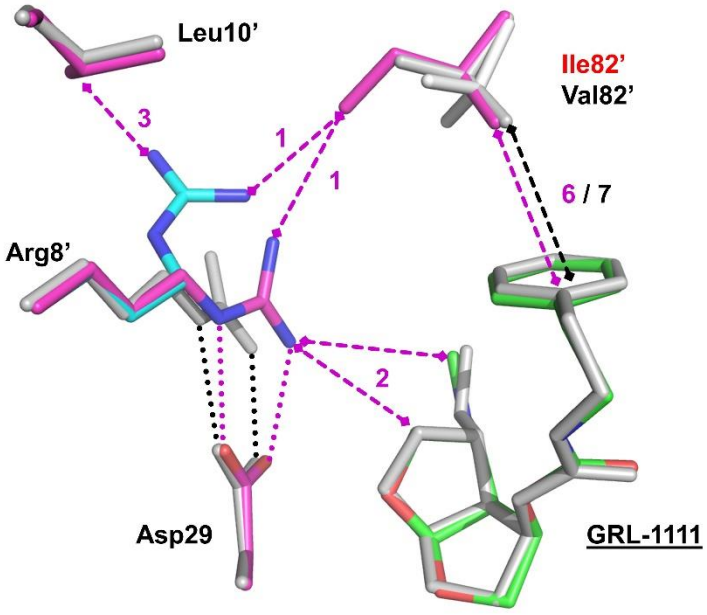
Val32 and Val32' have similar hydrophobic contacts with Thr80/80' and Ile84/84', while Val32' forms an additional van der Waals interaction with the side chain of Ile50 at the tip of the flap. Flap residues Ile47 and 47' show conserved hydrophobic contacts with Lys45/45', Ile54/54' and Ile50'/50, while Ile47' has an additional contact with Val56' in the flap and Leu76' but loses a contact with Asp30'. The majority of these internal hydrophobic contacts are retained in the PR^{Tri} mutant, despite the altered side chains of V32I and I47V. Compared to the wild type complex,

both Ile32 and 32’ lose one hydrophobic contact to Val47 and Val47’, respectively. However, Ile32 has gained a contact with Ile50’ in the mutant relative to the wild type complex. Val47 in the mutant has shifted by about 0.5Å and rotated by 180°, thus forming three extra hydrophobic contacts to the P2 group of the inhibitor but losing contacts with Asp30, Ile54, Ile50’, and Lys45. While Val47’ has the same orientation as the wild type residue, it loses two contacts to the P2’ group of the inhibitor and has fewer contacts with Lys45’ and Leu76’ in the mutant.

The Ca atoms of both Val47 and 47’ shift by about 0.5Å relative to their location in wild type PR. The most significant change illustrated for Val47 in mutant relative to Ile47 in wild type PR. This shift enables Val47 to form 3 new hydrophobic contacts with the P2 group of inhibitor, while losing contacts with neighboring side chains of Asp30, Lys45, Ile54 and Ile50’.

Mutation of Val82 to the longer Ile introduces new van der Waals interactions and significant changes in the side chain conformation of Arg8. In the wild type PR complex, the side chain of Arg8 has a single conformation that forms a key intersubunit ion pair with Asp29’ and no contacts with inhibitor. In the PR_{Tri}/GRL-1111 structure, the longer side chain of Ile82/82’ appears to induce two alternate conformations of Arg8/8’ with equivalent changes in both subunits. The guanidinium group of one conformation of Arg8 retains the ion pair with Asp29’ and also forms a new van der Waals contact with the P2 group of inhibitor. The second conformation of Arg8 is rotated away to form van der Waals contacts with Leu10 instead of the intersubunit ionic interaction with Asp29’. The second conformation of Arg8/8’ is not observed in the corresponding complex of PR_{Tri}/DRV, possibly because DRV lacks the extra amine of GRL-1111. These changes may be due to the local environment, since a multiple mutant PR20 has similar alternative conformations of Arg8/8’ related to the substitution of the larger Phe side chain instead of Leu10 (Figure 13) [149].

Figure 13: Interactions of alternative conformations of Arg8' with neighboring side chains and inhibitor. Mutant is in magenta bonds with alternate conformation of Arg8' in cyan and GRL-1111 in green. Wild type PR and inhibitor are grey. Dotted lines indicate ionic interaction, and dashed lines indicate hydrophobic interactions. CH-pi interactions with the aromatic P1 group of inhibitor are indicated by a single dashed line.



4.4 Discussion

The four tested investigational antiviral compounds were poor inhibitors (K_i values of 16.7 to 39.5 nM) of HIV-1 PR_{Tri} despite their excellent pM inhibition of wild type enzyme [128,130,132,133]. Analysis of protease-inhibitor interactions in the crystal structure of PR_{Tri}/GRL-1111 revealed the loss of one water mediated polar interaction in one orientation of

inhibitor and small changes in hydrophobic contacts. In particular, Val47 loses hydrophobic contacts with flap residues Lys45, Ile50' and Ile54 compared to those of wild type complex. The most substantial change compared to the wild type complex occurred for mutated residue Ile82/82', where the larger side chain produces two alternate conformations for the side chain of Arg8/8'. One conformation of the guanidinium group of Arg8/8' forms new van der Waals contacts with Leu10 and the P2 group of GRL-1111 instead of its typical intersubunit ion pair with Asp29'/29. We previously reported the crystal structures of PR_{Tri} complexes with APV, DRV and SQV [126]. These structures exhibited a single conformation for Arg8/8', except for the APV complex which had two alternate conformations for Arg8' side chain in one subunit and new van der Waals contacts with Ile82' and Leu10' similar to those in the PR_{Tri}/GRL-1111 structure. Moreover, only a single conformation was observed for Arg8/8' in our structures of HIV-2 PR with different inhibitors [126,150]. This change in conformation seen for Arg8/8' in PR_{Tri}/GRL-1111, comprising partial loss of its intersubunit ion pair and new intra-subunit interactions, might be expected to alter the stability of the mutant, however, the UC50 for urea denaturation remains close to the value for wild type PR. In contrast, PR mutant with the single substitution of R8Q, which completely eliminated the ion pair with Asp29', showed decreased stability with UC50 of 0.7 relative to wild type enzyme [151]. Overall, the loss of internal contacts among flap residues and disruption of the Arg8 – Asp29' ion pair are consistent with molecular dynamics simulations suggesting substitutions V32I, I47V and V82I in HIV-2 PR decrease the hydrophobic interactions with APV and DRV [152].

This new structure of PR_{Tri}/GRL-1111 shows how drug resistant mutations of V32I and I47V on opposite sides of the S2/S2' subsites can partially compensate for altered hydrophobic

interactions with inhibitors and other protease residues, while mutation V82I induces alternate conformations of Arg8/8' and introduces new interactions with Leu10 and the P2 group of GRL-1111. These changes appear to be specific for this combination of inhibitor and mutant.

Table 8: K_i values (nM) of tested inhibitors.

	APV	GRL-1111	GRL-0249	GRL-0519	GRL-0739
PR_{Tri}	2.5 ± 0.9	31.4 ± 3.9	16.7 ± 0.8	38.4 ± 1.2	39.5 ± 6.1
PR	0.16 [147]	0.002 [130]	0.002 [132]	0.006 [150]	0.010 [149]

Values for wild type enzyme are from listed citations.

Table 9: Crystallographic Data Collection and Refinement Statistics.

(Values in parentheses are given for the highest resolution shell.)

	PR_{Tri}/GRL-1111
Space group	$P2_12_12$
Unit cell dimensions: (Å)	
A	58.57
B	86.52
C	45.32
Resolution range (Final Shell)(Å)	50-1.50 (1.55-1.50)
Unique reflections	37,188 (2,865)
R_{merge} (%) overall (final shell)	5.9 (49.2)
I/σ(I) overall (final shell)	26.6 (3.1)
Completeness (%) overall (final shell)	96.8 (76.4)
Redundancy (final shell)	6.7 (4.1)
R (%)	13.4
R_{free} (%)	17.2

No. of solvent atoms	151
RMS deviation from ideality	
Bonds (Å)	0.020
Angle distance (degree)	2.29
Average B-factors (Å^2)	
Wilson B-factor	17.7
Main-chain atoms	17.0
Side-chain atoms	23.6
Whole chain atoms	20.2
Inhibitor	15.7
Solvent	28.3

5 SPERMINE AND OXACILLIN STRESS RESPONSE ON THE CELL WALL SYNTHESIS AND THE GLOBAL GENE EXPRESISON ANALYSIS IN METHICILLIN RESISTANCE STAPHYLOCOCCUS AUREUS

Pawar S, Yao X, Lu C. 2018. Spermine and oxacillin stress response on the cell wall synthesis and the global gene expression analysis in Methicillin-resistance *Staphylococcus aureus*. *Genes & Genomics*. 41:43–59. https://doi.org/10.1007/s13258-018-0735-8.

Abstract:

Methicillin-resistant Staphylococcus aureus (MRSA) is a rapidly emerging bacteria causing infection, which has developed resistance to most of the beta-lactam antibiotics because of newly acquired low-affinity penicillin-binding protein (PBP2a), which can continue to build the cell wall when beta-lactams block other PBPs. Exogenous spermine exerts a dose-dependent inhibition effect on the growth of Escherichia coli, Salmonella enterica serovar, and S. aureus. Selection of an MRSA Mu50 derivative which harbors mutation on PBP2 gene (named as MuM) showing spermine resistance and which confers a complete abolishment of spermine-beta-lactam synergy was identified. To further investigate the gene expression changes, a transcriptome profiling of MuM against Mu50 (wild-type) without any treatment, MuM and Mu50 in response to high dose spermine and Mu50 in response to spermine-beta-lactam synergy at 15, 30 and 60 min time points was performed. Functional annotation was further performed to delineate the metabolic pathways associated with the significant genes. A significant down-regulation in the iron regulatory system, potassium channel uptake and polyamine transport system with an up-

regulation in general stress response sigB dependent operon in MuM strain at 15, 30 and 60 min time points with spermine treatment compared to Mu50 strain was observed. Analysis of spermine-dependent synergy with beta-lactams on cell wall synthesis revealed that it significantly reduces the degree of cross-linkage on cell wall with no change in trypsin digestion pattern of purified PBPs and without affecting PBPs expression or PBPs acylation by Bocillin. A strong relation between PBP2 protein and general stress sigB response, iron, potassium and polyamine transport systems was observed. SigB regulon should be activated on stress, which was not seen in some of our previous studies where it was down-regulated in wild-type Mu50 strain with spermine stress. Here, an intriguing finding is made where there seems to be a correction of this abnormal response of no SigB induction to a significant induction by PBP2 mutation. In MuM strain, a significant downregulation of KdpABC operon genes at 15, 30 and 60 min time points on spermine stress is seen, which seems to be absent without spermine treatment. Since KCL has been found to protect the cell against spermine stress in wild-type strain by induction of KdpABC operon, it fails to do so in MuM strain underlying the importance of PBP2 protein in spermine stress. Analysis of spermine-dependent synergy with beta-lactams on cell wall synthesis revealed that it significantly reduces the degree of cross-linkage on cell wall with no change in trypsin digestion patterns of purified PBPs and without affecting PBPs expression or PBPs acylation by Bocillin. Furthermore, spermine does not help in enhancing the binding of beta-lactams to PBPs and binding of spermine to PBPs does not cause conformational changes to PBPs, as tested with trypsin digestion patterns. Future studies on the molecular mechanism of spermine interactions with these systems hold great potential for the development of new therapeutics for MRSA infections.

5.1 INTRODUCTION

Methicillin-resistant *Staphylococcus aureus* (MRSA) is an opportunistic pathogen that causes hospital and community-acquired infections varying from mild skin lesions to severe, life-threatening conditions like pneumonia, sepsis, and endocarditis [154, 155]. The resistance occurs because of newly acquired low-affinity penicillin-binding protein (PBP2a), which can build the cell wall when other PBPs are blocked by beta-lactams [156]. In *Staphylococcus aureus*, there are four native PBPs (PBP1-4), and one acquired PBP2a is attributive to MRSA [157]. Beta-lactam antibiotics irreversibly occupy the serine residue at the active site of penicillin-binding proteins (PBPs) forming a stable ester-linked acyl-enzyme which inhibits the transpeptidation process necessary for cell wall cross-linking [158-173]. Thus, MRSA seems to regulate a complex gene expression profiling when challenged with beta-lactams. Many studies have shown that exogenous spermine may become toxic to *Staphylococcus aureus* [174-177]. Exogenous spermine exerts a dose-dependent inhibition effect on the growth of *E. coli, Salmonella enterica serovar,* and *Staphylococcus aureus*. Recently, we reported that exogenous spermine might affect cell wall synthesis through its interactions with PBP2 and PBP2-associated multienzyme machinery to enhance the killing effects of beta-lactam antibiotics. The two possible ways of making MRSA sensitive by spermine-beta-lactam synergy can be by reducing the cross-linkage of peptidoglycan layer or by inhibiting cell growth. The PBP2 itself or enzymatic activities associated with the PBP2-dependent complex in cell wall synthesis could be a potential target of spermine, so we sought to identify a spontaneous mutation derived MRSA strain (*MuM*) conferring spermine resistance and studying how this mutation can contribute to resistance. The identified *MuM* strain showed a 32-fold increase in tolerance of growth inhibition by spermine and has completely lost

the spermine-beta-lactam synergy. Genome re-sequencing revealed a deprivation of transpeptidase activity with a 7-bp deletion within the *pbpB* gene, which may result in perturbation of cell wall synthesis and change the peptidoglycan structure. We also found that the degradation of *MuM* cell wall materials was significantly faster than that for the *Mu50* control which suggested that cell walls of *MuM* may be less cross-linked and more susceptible to hydrolysis due to the defects in PBP2 protein [177, 178]. Thus, *MuM* is seemed to adjust its metabolic activities to compensate the defective cell wall synthesis. A detailed transcriptomic analysis of how *MuM* and *Mu50* responds to spermine and spermine-beta-lactam synergy is still unknown. Therefore, in the current study, we have systematically compared *Mu50* (wild-type), and *MuM* strain response to spermine alone (high dose) or in combination with b-lactam (oxacillin) (both at low dosages) using microarrays. To study the exogenous spermine effects on cell wall synthesis through its interactions with PBP2 and PBP2-associated multienzyme machinery to enhance the killing effects of beta-lactam antibiotics, comparisons between *Mu50* and *MuM* strains are performed using cell wall synthesis and transcriptomic assays. For transcriptomic experimental design, a three time-point (0, 30, and 60 minutes) microarray data with replicates was generated for each of the two strains treated with spermine, oxacillin or both, and comparisons were made to study within and between treatment (*MuM*) changes when compared to the control strain (*Mu50*). Complementation of pbpB protein was also performed to validate the changes in treatment strain. While some groups have reported genome-wide transcriptional profiling of *Staphylococcus aureus* against the cell wall active antibiotics [177], ours is one of the first attempts in understanding this species response to spermine and beta-lactams with mutated PBP2 protein.

5.2 METHODS

5.2.1 Bacterial strains, plasmids, growth conditions and isolation of spermine-resistant mutants:

Staphylococcus aureus Mu50 (ATCC 700699), RN4220 (kindly provided by R. P. Novick, Newyork, USA) and *Escherichia coli DH5 alpha* (Bethesda Research Laboratories, Maryland, USA) were used for this study. Spontaneous mutants of MRSA *Mu50* (oxacillin MIC of 512 μg/ml and spermine MIC of 1 mM at pH 8.0) were isolated by spreading 1×108 CFU of log-phase cells onto spermine-containing plates (2 to 8 mM, pH 8.0) maintained in the Luria-Bertani (LB) medium and incubated overnight at 37°C. One independent colony was found resistant to spermine (up to 32 mM) in broth and on plates designated as *MuM*. For cell wall composition analysis, *Staphylococcus aureus Mu50* (ATCC700699), *E. coli DH5 alpha* (Bethesda Research Laboratories, Maryland, USA) and *Top10* (Invitrogen, Massachusetts, USA) strains were employed. The plasmids and primers used in this study are listed in Table 10.

Table 8 Plasmids used in this study

Plasmids	Relevant characteristics	Source or reference
pBAD/HisA	Expression vector, Amp	Invitrogen
pBAD/HisD	Expression vector for producing N terminal His tag fusion, Amp	This study
pBAD/HisE	Expression vector for producing C terminal His fusion, Amp	This study
pH6N-PBP1	*pBAD/HisD* expressing N-HisPBP1	This study

pH6C-PBP2	*pBAD*/*HisE* expressing C-HisPBP2	This study
pH6N-PBP3	*pBAD*/*HisD* expressing N-HisPBP3	This study
pH6N-PBP4	*pBAD*/*HisD* expressing N-HisPBP4	This study
pBP2a	*pBAD*/*HisA* expressing no tag PBP2a	This study

Both *E. coli* and *Staphylococcus aureus* strains were grown and maintained in the LB medium cultured in 20mM Tris-HCl at the indicated pH when needed. For *E. coli*, ampicillin was added to the medium as necessary at 100 μg/ml. Antibiotics were added to the medium as necessary at the following concentrations for maintenance of plasmids: ampicillin, 100μg/ml for *E. coli*; kanamycin, 25μg/ml for *E. coli*, 50μg/ml for *S. aureus*; erythromycin, 10μg/ml for *S. aureus*; chloramphenicol, 10μg/ml for *S. aureus*.

5.2.2 Complementation of pbpB:

The *pbpB* gene was reported to be transcribed independently or as a polycistronic RNA from its upstream *prfA* promotor [185]. To ensure optimal expression, a 3.2 kb fragment covering both *prfA* and *pbpB* were amplified by PCR from *Mu50* genomic DNA with the primers *5'-CGC GGA TCC ACA CAT ACT TGT ACT TGC CTC-3'* (forward) and *5'-CGC GGC GCC GAG TGG ATT AGT TGA ATA TAC CTGTTA ATC CAC CGC TG-3'* (reverse). The resulting PCR product was cloned into the *BamHI* and *NarI* sites of the shuttle vector *pCN38*. The recombinant plasmid *pYX9* was first cloned into and extracted from *E. Coli* and then electroporated into *Staphylococcus aureus RN4220* [167, 194]. Plasmid DNA isolated from recombinant strains of *RN4220* was subsequently introduced into *Mu50* and *MuM* strains by electroporation. Meanwhile, empty vector

(*pCN38*) did not rescue the phenotypes in *MuM* (data not shown).

5.2.3 Transcriptional profiling conditions:

Staphylococcus aureus Mu50 and *MuM* were grown in Tris-buffered LB (pH 7.5), and cultures in the exponential phase (optical density at 600 nm [OD600] of around 1.0) were immediately treated with the RNA protection reagent (Qiagen, Maryland, USA) before harvesting. In the first condition, *Mu50* and *MuM* strains were treated with 1mM spermine, and RNA isolated at 0, 15, 30 and 60-minute time-points with spermine and single 0 min time-point without spermine. In the second condition, three treatments of *Mu50* strain with 1mM spermine, 2ng/µl oxacillin and a combination of 1mM spermine, 2ng/µl oxacillin were grown for one hour subsequently followed by RNA isolation. Spermine at 0.5mM was able to potentiate oxacillin MIC in MRSA strain *Mu50*, from 512µg/mL to less than 1µg/ml. Spermine MIC alone is 2-4mM. Therefore, spermine at 0.5 or 1 mM, oxacillin at 1µg/ml, 2µg/ml or 16µg/ml, are the concentrations that do not affect bacterial growth when used alone, but they can kill MRSA when used in combination [167]. So we chose to use 1/4 MIC instead of 1mM for spermine, and 1/32 MIC instead of 16µg/mL for oxacillin. RNA samples were extracted from cells with phenol (Fisher, Massachusetts, USA), digested with RNase-free DNase I (Roche, Indiana, USA) to remove genomic DNA, and purified with RNeasy mini columns (Qiagen, Maryland, USA). The cDNA synthesis, fragmentation, and terminal labeling were carried out as per the protocols of the manufacturer (Affymetrix, Massachusetts, USA). Labeled cDNA was hybridized to the GeneChip *Staphylococcal aureus* genome array. After scanning, the images were processed with GCOS 1.4 software (Affymetrix, Massachusetts, USA). The data was generated for two independent biological replicates.

5.2.4 Microarray analysis:

Mas 5.0 normalization was performed for all the files at 0-minute time point for *Mu50* and *MuM* strains and strains with *MuM-PBP2* and *Mu50-PBP2* complementation plasmid [195]. For calculating upregulated genes, in control of all the P (present) call intensity values were considered for analysis and all the M (marginal) and A (absent) calls were regarded as 100. For treatment, all the genes with intensity values above 500 were considered in the analysis. This rigorous approach gave us significant differences amongst various comparisons avoiding any false positive and false negative results. The exact opposite criteria were applied to find down-regulated genes. Fold changes (greater than 1.5 and less than 1.5) with *MuM-Mu50* and (*MuM* with *PBP2*)-(*Mu50* with *PBP2*) were taken to find up and down-regulated genes in *MuM* strain. A similar method was used for comparison of *MuM* and *Mu50* at 15, 30 and 60-minute time points with spermine treatment. Up and down-regulated genes were calculated and compared with 0-minute time point with no spermine treatment. All the microarray data were analyzed using library 'Affy' package [196] on R platform [197]. Heat maps were generated using library 'gplots' [198]. Heat maps were developed on Z scores, which were calculated by heatmap.2 function of gplots [Z score = (raw intensity - average)/standard deviation]. MA plots [199] for showing differentially expressed genes were calculated as follows: M = Logarithm to base 2 (Treatment/Control), A= 1/2×Logarithm to base2 (Treatment×Control). MA plots were made on R platform with 'plotMA' limma Bioconductor package [200].

5.2.5 Cell wall composition analysis:

MRSA strain *Mu50* was grown in LB broth (pH 8.0, w/ 20mM Tris cl) at 37°C with aeration. Cultures were supplemented with 0.5 mM spermine (Spm) and/or 1 ng/µl oxacillin (Ox) at OD600 of 0.5 and cell growth was continued for 4 hours before harvesting. Isolation of cell wall peptidoglycans and the analysis of the muropeptides by reverse-phase HPLC were carried out by CeCo Labs from Tubingen University, Tubingen, Germany (www.cecolabs.de).

5.2.6 Protein cloning, expression, and purification:

Phusion polymerase (NEB) was used to amplify *pbp1* (or *pbpA*), *pbp3* (or *pbpC*), *pbp4* (or *pbpD*) genes with no N-terminal signal peptide and the putative transmembrane (TM) domain (at N-terminus of *PBP1* and *PBP3* and C-terminus of *PBP4*) from gDNA of MRSA *Mu50* strain. Three plasmids were generated (*pH6N- PBP1, pH6N-PBP3, and pH6N-PBP4*) from PCR product cloned into *pBAD/HisD* expression vector digested at *PstI/EcoRI* restriction enzyme sites. The constructs encoded hexahistidine tag at the N-terminus of *PBP1, PBP3,* and *PBP4* plasmids. Recombinant proteins covering residues M37-D744 for PBP1, Q44-K691 for PBP3, and T25-H400 for PBP4 were expressed from these plasmids. A similar approach was utilized for expressing periplasmic polyamine-binding protein PotD and PotR of the *potABCD* operon. The PCR products were digested with HindIII and EcoRI and cloned into vector *pBAD/HisD* (5’ blunt end of PCR product ligating to SmaI site on the vector) at SmaI/HindIII and SmaI/EcoRI sites. We labeled these plasmids as *pH6NnSP-PotD* and *pH6N-PotR* (hexa-histidine tags were fused to N-terminus of PotD and PotR). *Top10 E. coli* strains harboring plasmid of interest were grown at 30°C in LB medium containing 100 µg/ml ampicillin. At 0.6 to 0.8 OD600 cultures were supplemented with 0.2 percent arabinose and grown for additional 4 hours. After harvestation cells were suspended in buffer A containing 20 mM sodium phosphate [NaPi] buffer, pH 7.4, and

500mM NaCl. The buffer was also supplemented with EDTA-free protease inhibitors (Roche), which after centrifugation at 20, 000×g for 30min at 4°C a HisTrap HP column (GE) pre-equilibrated with Buffer A was applied on soluble fraction. Bound proteins were eluted by increasing concentrations of imidazole (final conc. 500 mM) in a gradient-dependent manner. Protein concentration was performed on fractions using Aminco Ultra-4 centrifugal filter unit (Millipore, Massachusetts, USA) with buffer to 20 mM Tris-HCl (pH 7.6). The region encoding amino acids W59 to S716 for *pbp2* (or *pbpB*) gene was PCR amplified (Phusion, NEB, New England, USA) from *Mu50* genomic DNA with specific primers. Digestion was done with BspHI/SmaI restriction enzymes and the product was cloned in vector *pBAD-HisE* at NcoI/SmaI sites (pH6C-PBP2 having C-terminal hexahistidine tag) which were subsequently introduced into *E. coli Top10* for expression its protein. Induction with 0.2 percent arabinose for 4 hours at 30°C was performed at a log-phase culture. This was followed by French press at 17,000 psi in Buffer A. After centrifugation (20, 000g × for 30 min) pellet was suspended in Buffer A (0mM CHAPS, 10 percent glycerol, 0.1 percent Sarkosyl (detergent) at pH to 9.0. Stepwise elution was performed with detergent-containing Buffer A with 500mM imidazole using a HisTrap HP column (GE). Concentration was performed using Aminco Ultra-4 centrifugal filter unit (Millipore, Massachusetts, USA) at buffer 20 mM Tris-Cl (pH7.6). *mecA* gene (without first 69 bases for its N-terminal anchoring region of 23 amino acids) was amplified from *Mu50* gDNA and cloned into NcoI/HindIII restriction enzyme sites of *pBAD/HisA* for expressing PBP2a protein. The resulting plasmid *pPBP2a* was transformed into *E. coli Top10* which was induced at a log-phase culture with 0.2 percent arabinose for 4 hours at 30°C. Cells were disrupted by French press at 17,000 psi in buffer Q (20mM Tris-Cl, pH7.6). The supernatant after centrifugation at 20, 000g × for 30 min was subjected to ammonium sulfate fractionation. At 30 percent saturation fraction was

precipitated and concentrated by Aminco Ultra-4 centrifugal filter unit (Millipore, Massachusetts, USA). The obtained protein sample was further diluted with Q buffer and applied to HiTrap Q HP column (GE, California, USA). Gradient elution with buffer Q containing 1M KCl was also used, and samples were concentrated again with Aminco Ultra-4 centrifugal filter unit to change the buffer to 20 mM Tris-HCl (pH 7.6).

5.2.7 Construction of expression vector pBAD-HisE:

Expression vector *pBAD-HisE* was constructed with primer pair's *hisF-5'-GGGCAT CATCATCATCATCATTGAATTCTGC-3'* and *hisR-5'-GCAGAATTCAATGAT GATGATGATGATGCCC-3'.* Digestion was performed with SmaI/EcoRI to incorporate this region in the vector *pQF50*. Further digestion was performed with NcoI/EcoRI enzymes to release the partial MCS carrying the His6 region.

5.2.8 Bocillin labeling of PBPs:

Purified PBPs (0.04 µg in 20 µl of 20mM Tris-Cl pH8.0) were incubated with Bocillin of indicated concentrations for 30 min at 37°C with gentle shaking. *Staphylococcus. aureus* (10 ml) was harvested at log-phase and the cell pellet was suspended in 200 µl the lysis buffer (50mM Tris pH7.5, 50mM NaCl, 5-10 µl of lysostaphin (10mg/ml), 5ul of DNaseI (10 Units/µl), 5 µl of RNase A (200 µg/ml) for 15-20 minutes at 37°C. The lysates obtained were labelled by adding Bocillin at a final concentration of 5uM and incubated at 37°C for 30 min. Ranges of Bocillin concentration varied from 0.25-16 uM for concentration-dependent experiments. Samples were pre-incubated with spermine for 30 min at 37°C prior to Bocillin labeling when needed. Samples were then mixed

with Laemmli sample buffer and boiled for 5 minutes prior to SDS-PAGE. The gel was rinsed and scanned with Typhoon scanner with excitation wavelength at 488 nm and emission wavelength at 530 nm. Band intensity was quantified by Image Quant software.

5.2.9 DARTS (Drug Affinity Responsive Target Stability):

The purified protein in buffer containing 1 μg in 20 μl of 50mM Tris-Cl, pH 8.0, 50mM NaCl, 10 mM CaCl2 was incubated at 37°C with or without 1 mM spermine for 25 min followed by trypsin digestion (0.075 μg) at 25°C. The reaction was terminated with Laemmli sample buffer followed by boiling samples for 10min and subsequent SDS-PAGE.

5.3 RESULTS

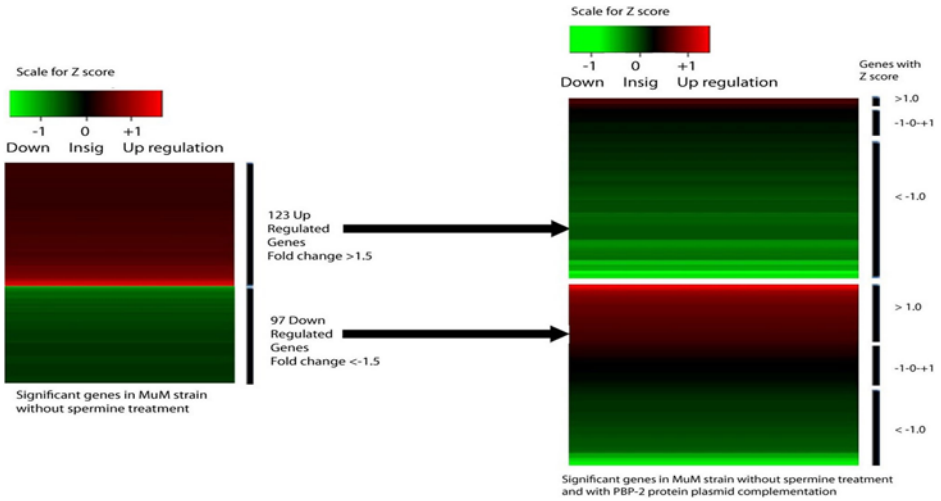
5.3.1 Transcriptome analysis of MuM strain in response to high dose spermine:

5.3.1.1 Differences in Mu50 and MuM gene expression at 0-minute time point without spermine and reversion by MuM/PCN38 complementation:

A total of 220 significantly up and down-regulated genes were found to be expressed in *MuM* strain with no spermine treatment at 0-minute time-point. These genes were then compared with *MuM-PBP2* complementation plasmid to see if the changes reverted to wild-type expression profile. Reversion of expression by *MuM-PBP2* was set to genes whose expression levels felt below -1-fold or went above +1-fold [170]. Three groups were made with genes which reverted to down-regulation, genes which returned to up-regulation, and genes with no change after complementation. Doing so 59 genes which completely reverted to up or down-regulation profiles

and rest 161 genes had no effect by complementation. A heat map in Figure 14 compares these 220 significant genes in *MuM* and *MuM* complemented with the PBP2 plasmid.

Figure 14 Heatmap comparing 220 significant genes in MuM and MuM complemented with PBP2 plasmid strains.



The complete reversion of the up and down regulation status of genes in *MuM* strain with a PBP2 complemented plasmid is because of mutant PBP2 protein. These genes are selected to validate the changes occurring with and without spermine treatment in subsequent analysis.

5.3.1.2 Significant down-regulation in the iron transport system and up-regulation of sigB regulon genes was seen at 15, 30 and 60-minute time points in MuM strain when compared to Mu50 strain with spermine treatment:

A total of 420, 335, and 332 genes were found significantly expressed in *MuM* strain at 15, 30 and 60-minute time points when compared with *Mu50* strain. These genes were then fed in DAVID database [170] to perform functional annotation. Genes with the iron transport system were found to be significantly downregulated with spermine treatment at 15, 30 and 60-minute

time points in *MuM* strain compared to *Mu50* strain. The list of these genes and their fold change expression levels are given in Table 11 and Figure 28.

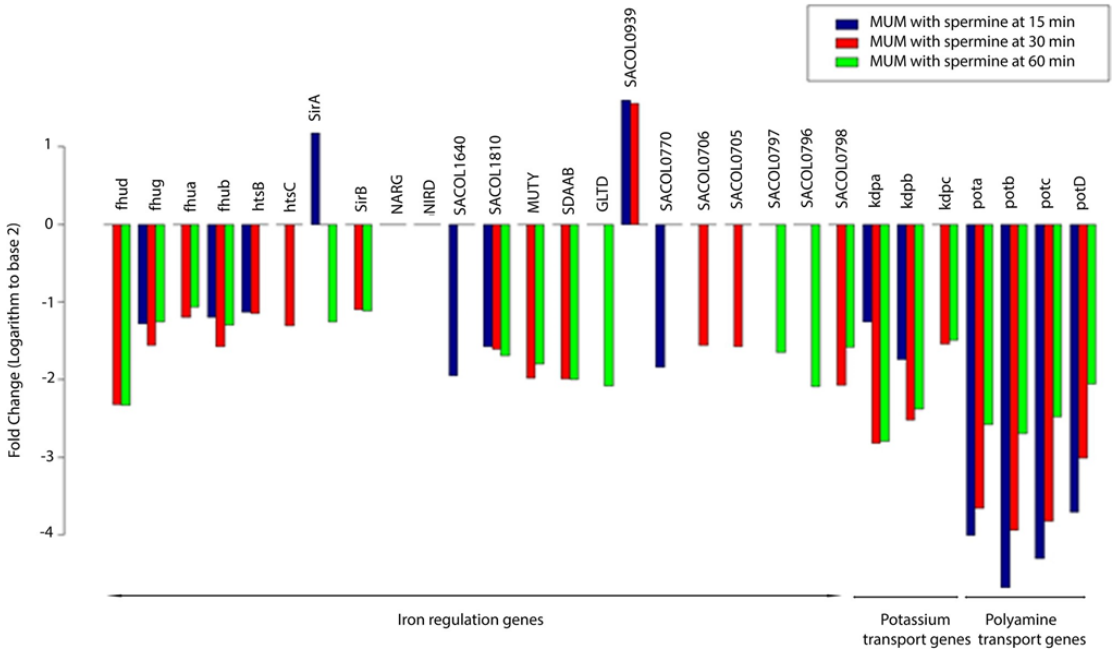
Table 9List of iron regulation, polyamine and potassium transport genes with their significant fold change expression levels (logarithm to base 2)

Gene symbol	Affymetrix ID	*MUM.NOSPM.0*	*MUM.SPM.15*	*MUM.SPM.30*	*MUM.SPM.60*
fhud	sa c914s711 at	1.39	0	− 2.32	− 2.33
fhug	sa c7993s6980 at	0	− 1.28	− 1.56	− 1.26
fhua	sa c5423s4693 a at	0	0	− 1.2	− 1.07
fhub	sa c7989s6976 at	0	− 1.2	− 1.57	− 1.3
htsB	sa c4643s3963 a at	0	− 1.13	− 1.15	0
htsC	sa c4639s3961 a at	0	0	− 1.31	0
sirA	sa c1230s1008 at	0	1.18	0	− 1.26
sirB	sa c1172s953 a at	0	0	− 1.1	− 1.12
NARG	sa c5574s4827 a at	− 2.99	0	0	0
NIRD	sa c5580s4836 a at	− 1.94	0	0	0
SACOL1640	sa c2711s2285 a at	0	− 1.95	0	0

SACOL1810	sa c3357s2894 a at	0	− 1.57	− 1.61	− 1.69
MUTY	sa c3689s3168 a at	0	0	− 1.98	− 1.8
SDAAB	sa c6092s5283 a at	0	0	− 1.99	− 2
GLTD	sa c7412s6438 a at	0	0	0	− 2.08
SACOL0939	sa c8086s7067 a at	0	1.6	1.56	0
SACOL0770	sa c8202s7182 a at	0	− 1.84	0	0
SACOL0706	sa c7993s6980 at	0	0	− 1.56	0
SACOL0705	sa c7989s6976 at	0	0	− 1.57	0
SACOL0797	sa c8283s7260 a at	0	0	0	− 1.65
SACOL0796	sa c8276s7256 a at	0	0	0	− 2.09
SACOL0798	sa c5353s4626 a at	0	0	− 2.07	− 1.58
kdpa	sa c4298s3650 a at	1.88	− 1.26	− 2.82	− 2.8
kdpb	sa c4292s3644 a at	0	− 1.74	− 2.52	− 2.38
kdpc	sa c236s9562 at	1.32	0	− 1.54	− 1.49

pota	sa c5349s4625 a at	− 1.01	− 4.01	− 3.66	− 2.58
potb	sa c9028s7925 a at	− 1.97	− 4.68	− 3.94	− 2.7
potc	sa c795s596 a at	− 1.36	− 4.31	− 3.83	− 2.48
potD	sa c803s604 a at	− 1.69	− 3.71	− 3.01	− 2.06

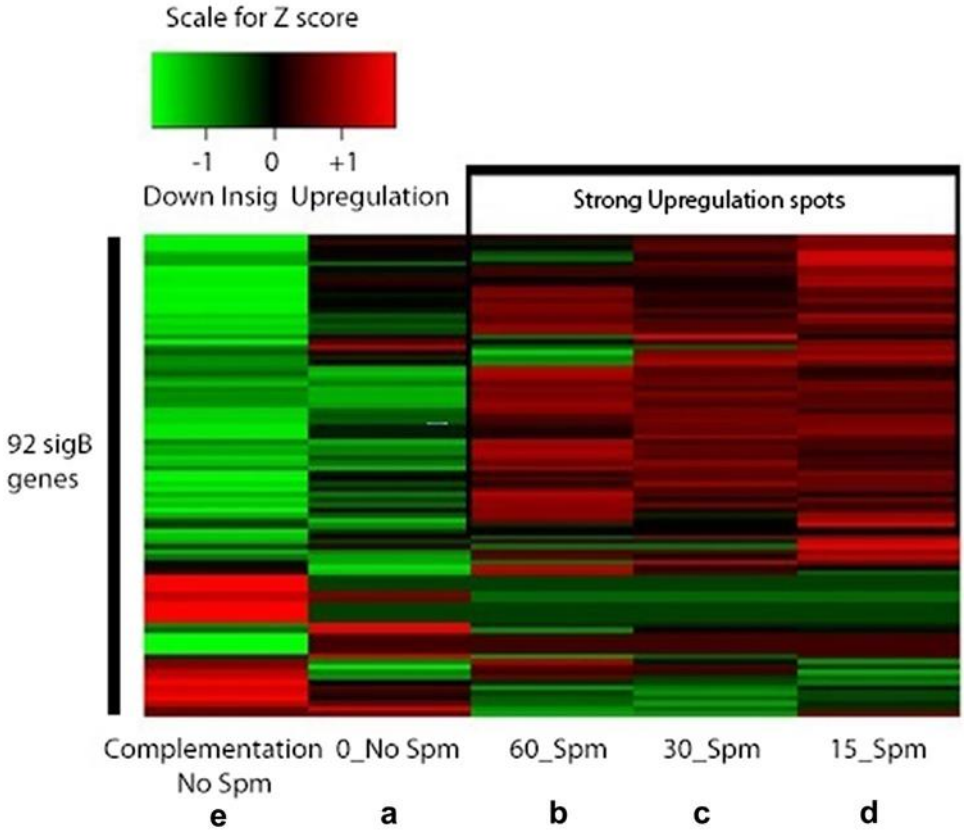
Figure 15 Bar graph with fold changes (logarithm to base 2) for iron regulation, potassium and polyamine transport genes in MuM strain at 15, 30 and 60 min time points with spermine treatment.



A pattern of *sigB* regulon genes with significant up regulation with spermine treatment was seen in *MuM* strain when compared to *Mu50* strain. A heat map comparing expression levels of sigB

operon genes at 0-minute time point with no spermine, 15, 30 and 60-minute time points with spermine treatment is shown in Figure 15.

Figure 16 Heat map on the fold changes comparing 92 sigB genes in MuM and Mu50 strains at 0 min time-point without spermine treatment, MuM at 0 min time-point without spermine treatment and with PBP2 complementation, MuM with spermine treatment at 15, 30 and 60 min time points.



Iron is an essential element of redox systems in many gram-positive and negative bacteria's. Iron transport systems are needed by bacteria to solubilize external Fe3+. The ABC transporters form a major iron transport system in gram-positive bacteria. *Staphylococcus aureus* grows on transferrin as a sole iron source [173]. There has been a recent demonstration of the production of siderophores by *Staphylococcal* strains. *Staphylococcus hyicus* has been shown to produce at least two siderophores, staphyloferrins A and B. Siderophores mobilize transferrin iron and donate it to

siderophore specific ABC transporters in the cytoplasmic membrane [175]. Staphyloferrin A and B also host catecholamine hormones that support the growth of Staphylococcus aureus. Staphyloferrin iron is taken up via the *HtsABC* and *SirABC* systems, whereas catecholate iron is taken up by the newly identified *SstABCD* transport system. Approximately, 42 known transport proteins of Fe3+ siderophore transport systems have been identified [175]. Expression of iron repressible outer membrane proteins (IROMPs) binding to specific iron-containing compounds transport free iron or the iron bound ligand into the cell. The binding of these receptor proteins is mediated by the iron binding repressor protein and ferric uptake regulator (fur) requiring ferrous iron as a cofactor [176]. In *Escherichia coli fhu* operon is composed of the *fhuA, B, C* and *D* genes which are essential for the utilization of ferric siderophores of the hydroxamate-type and mutations in either one of the fhu genes negatively influence the expression of other fhu genes in the operon. Transport across the outer membrane is regulated by *fhuA*, while the cytoplasmic membrane transport is regulated via *fhuC, B,* and *D* genes [177]. By comparing *MuM* with *Mu50* strains at 15, 30 and 60-minute time points with spermine treatment, a significant down-regulation in *HtsABC, SirABC* and *fhu* operons in *MuM* strain was observed. These significant genes included iron ion binding proteins, iron-sulfur cluster binding proteins, and iron ABC transporters. There have been reports for chelation of iron inside the bacterial cell from spermine by forming ternary complexes with Fe2 and phosphate groups [178] [179]. We have shown that spermine can induce stress through iron starvation in *Staphylococcus aureus* which can be the result of altered intracellular iron status [176]. These signals can be reciprocated by intracellular iron regulators channels like *HtsABC, SirABC,* and *fhu* operons. Significant induction of all these iron acquisition operons is seen in *Staphylococcus aureus* on spermine stress indicating a correlation between iron and spermine. We see a significant down-regulation in iron acquisition operons with mutant PBP2

protein pitching the importance of PBP2 to iron acquisition and spermine stress. SigB factors have been shown in various bacteria to be important for survival under extreme conditions influencing virulence and pathogenicity [180]. *SigB* regulon is one of the first regulons to response in stress for bacteria [181]. *SigB* influences the expression of a variety of virulence genes like alpha-hemolysin, lipases, and proteases [181-184]. DNA microarray bases study in *Staphylococcus aureus* has shown that the *sigB* regulon influences the expression of at least 251 genes (198 and 53) genes as positively and negatively controlled respectively) [180]. A sixth sigma factor is necessary to initiate transcription from promoters by binding to the core of RNA polymerase consisting of five subunits [182] [183]. Most bacteria have additional sigma factors called alternative sigma factors which can target specific recognition sequences and control specialized regulons that are activated under stress [179]. Regulators *RsbV* and *RsbW* are conserved in all gram-positive bacteria that have *sigB* regulon [166]. The kinase activity of RsbW determines phosphorylation state of the anti-sigma factor antagonist RsbV is subsequently controlling sigma B activation [182] [184]. The sigma B regulated genes are involved in many cellular processes including cell envelope turnover, biosynthesis, intermediary metabolism, virulence and membrane transport processes where the *SigB* regulon is highly interconnected with other regulatory networks. Moreover, global regulators of stress response like *SarA* are also partly under SigB control [183]. About 92 *sigB* regulated genes were identified to respond at 0, 15, 30 and 60-minute time points. At 0-minute time point with no spermine treatment, 49 genes were found to be significantly up-regulated in *MuM* strain. When compared with 15, 30 and 60-minute time points with spermine treatment, 45, 35 and 30 genes of 49 significantly expressed at 0-time point showed a fold change of greater than 1.5. A pattern of *sigB* regulon genes with significant up-regulation with spermine treatment was seen in *MuM* strain when compared to *Mu50*. *MuM* strain deprives

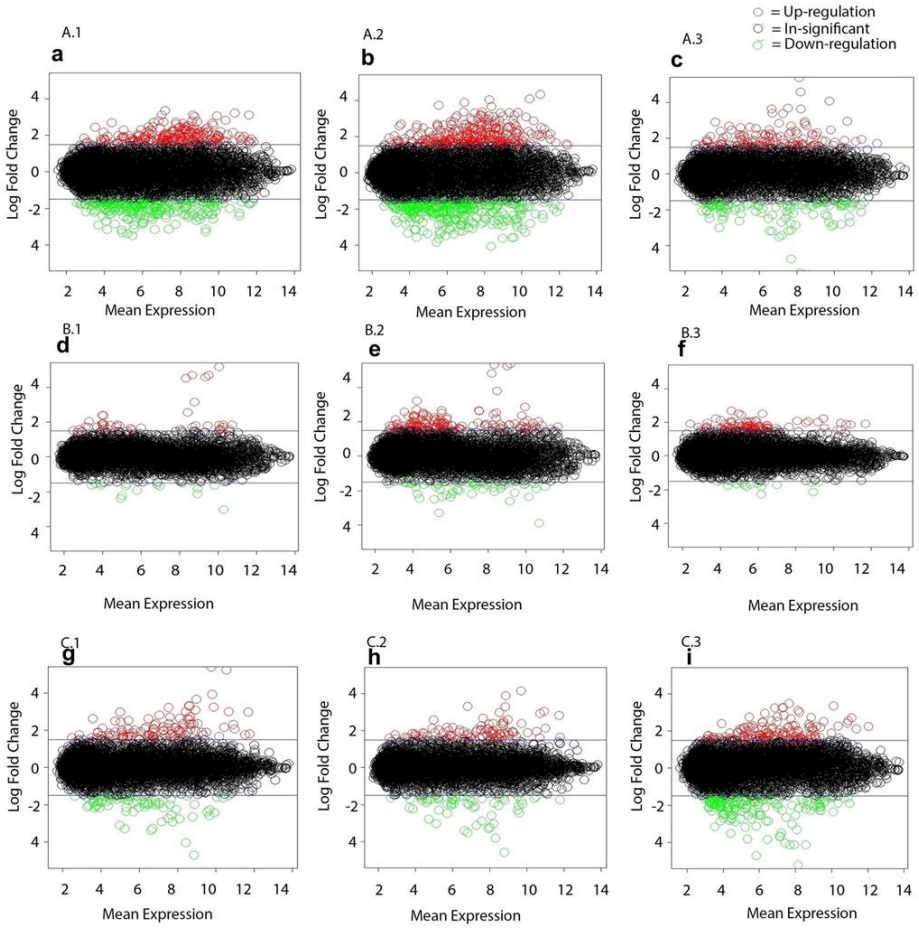
transpeptidase activity of the PBP2 protein, being abnormal physiology, the bacterium may try to compensate it by activating *sigB* regulon in *MuM* strain without spermine stress. Spermine stress may exacerbate this condition and can cause a continued up-regulation of this regulon throughout 15, 30 and 60-minute time points on spermine stress. *SigB* regulon should be activated on stress, which was not seen in some of our previous studies where it was down-regulated in wild-type *Mu50* strain with spermine stress [166]. Here, we have made an intriguing finding where there seems to be a correction of this abnormal response of no *SigB* induction to a significant induction by PBP2 mutation. It seems to be unclear why *SigB* regulon was not induced by spermine stress in wild-type strain which starts responding in *MuM* strain.

5.3.1.3 A total of 488, 146 and 249 significant genes stating differences between and within treatments for Mu50 and MuM strains were selected amongst three categories for functional annotation:

To identify within and between treatment changes, three categories of comparisons were made: 1. Comparison of *Mu50* 15, 30 and 60-minute time-points with spermine treatment and *Mu50* 0 minute time-point without spermine treatment, 2. Comparison of *MuM* 15, 30 and 60-minute time-points with spermine treatment was compared with *MuM* 0 minute time-point without spermine treatment and 3. Comparison of *MuM* 15, 30 and 60-minute time-points with spermine treatment was compared with *Mu50* 0 minute time-point without spermine treatment. A total of 488 significant genes with a fold change of greater and lesser than 1.5 in category 1, 146 significant genes in category 2 and 249 significant genes in category three were selected. MA (log ratio to

mean average) plots for each of these comparisons are plotted in Figure 17 to show the distribution of these differentially expressed genes.

Figure 17 MA plots showing differentially expressed genes in Mu50 and MuM treatments. a.1, a.2, a.3 Mu50 at 15, 30 and 60 min time-points with spermine treatment over Mu50 0 min time-point without spermine treatment. b.1, b.2, b.3 MuM at 15, 30 and 60 min time-points with spermine treatment over MuM 0 min time-point without spermine treatment. c.1, c.2, c.3 MuM at 15, 30 and 60 min time-points with spermine treatment over Mu50 0 min time-point without spermine treatment.



These significant genes were further used for functional annotation.

5.3.1.4 Addition of spermine to Staphylococcus aureus Mu50 strain caused decrease expression of genes in major central metabolic pathways of carbohydrate biosynthesis, glycolysis and TCA

cycle, cell structure, fatty acid and lipid biosynthesis and electron transfer chain:

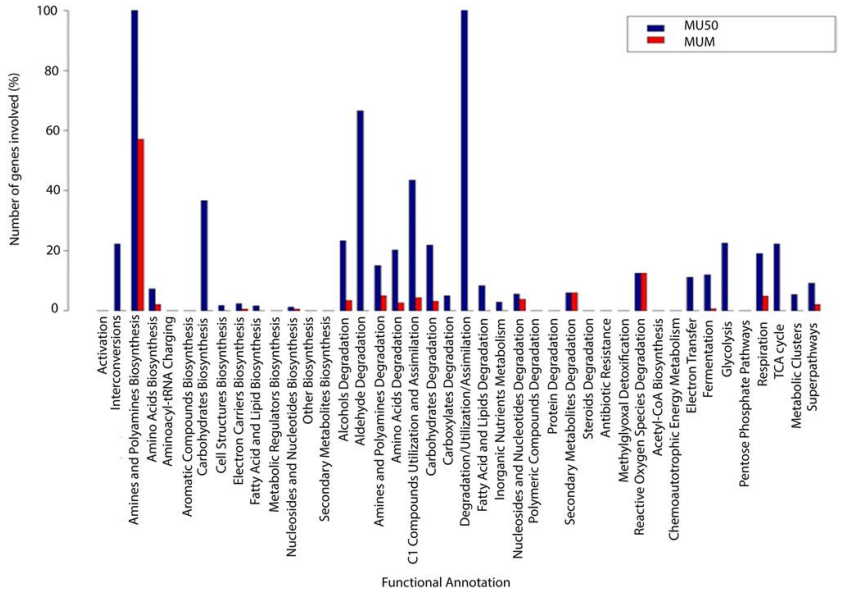
The metabolic pathways and genes involved in *Staphylococcus* were downloaded from Biocyc pathway/genome database collection. Functional annotation was performed to delineate the metabolic pathways associated with the significant genes. Addition of spermine to *Staphylococcus aureus Mu50* strain caused decrease expression of genes in major central metabolic pathways of carbohydrate biosynthesis, glycolysis and TCA cycle, cell structure, fatty acid and lipid biosynthesis and electron transfer chain. Percent number of genes involved in each of the processes affected in *MuM* and *Mu50* strains is shown in Figure 30 with their descriptions in Table 12.

Table 10 List of metabolic processes affected in each of the MuM and Mu50 strains on spermine stress [+: affected, −: not affected]

Process number	Metabolic processes	*MU50*	*MU50 (percent genes)*	*MuM*	*MUM (percent genes)*
1.2	Interconversions	+	22	−	0
2.1	Amines and polyamines biosynthesis	+	100	+	57
2.2	Amino acids biosynthesis	+	7	+	2
2.5	Carbohydrates biosynthesis	+	36	−	0
2.6	Cell structures biosynthesis	+	1	−	0
2.7	Cofactors, prosthetic groups, electron carriers biosynthesis	+	2	+	0.5
2.8	Fatty acid and lipid biosynthesis	+	2	−	0
2.1	Nucleosides and nucleotides biosynthesis	+	1.5	+	0.5
2.11	Other biosynthesis	−	0	+	0.5
3.1	Alcohols degradation	+	23	+	3
3.2	Aldehyde degradation	+	66	+	0
3.3	Amines and polyamines degradation	+	15	+	5
3.4	Amino acids degradation	+	20	+	2
3.5	C1 compounds utilization and assimilation	+	43	+	4

3.6	Carbohydrates degradation	+	21	+	3
3.7	Carboxylates degradation	+	5	−	0
3.8	Degradation/utilization/assimilation—other	+	100	−	0
3.9	Fatty acid and lipids degradation	+	8	−	0
3.1	Inorganic nutrients metabolism	+	2	−	0
3.11	Nucleosides and nucleotides degradation	+	5	+	3
3.14	Secondary metabolites degradation	+	5	+	5
3.15	Steroids degradation	−	0	−	0
4.2	Methylglyoxal detoxification	−	0	+	1
4.3	Reactive oxygen species degradation	+	12	+	12
5.3	Electron transfer	+	11	−	0
5.4	Fermentation	+	11	+	0.5
5.5	Glycolysis	+	22	−	0
5.7	Respiration	+	19	+	4
5.8	TCA cycle	+	22	−	0
6.1	Metabolic clusters	+	5	−	0
7.1	Superpathways	+	9	+	2

Figure 18 Functional annotation was performed to delineate the metabolic pathways associated with the significant genes using Biocyc pathway/genome database collection.



This transcriptome analyses demonstrated a specific pattern of response in *MRSA Mu50* upon high

dose spermine exposure, suggesting spermine alone may alter the intracellular iron status and suppress the *SigB* regulon to exert its toxicity. Combination of two drugs may inhibit bacterial growth in more complicated ways than the single drug does. As for our case, the synergy of spermine and beta-lactams may function by spermine facilitating beta-lactams on perturbation of cell wall synthesis or beta-lactam enhancing spermine attack on unknown targets, or the formation of a new structural configuration that makes the common target more vulnerable to be attacked by spermine or beta-lactams. To obtain a better insight of how two agent's, spermine and beta-lactam might affect each other to enhance the antibacterial effectiveness, we monitored the global changes in gene expression patterns over spermine only or oxacillin only and spermine/oxacillin combination conditions. Considering our isolated *Mu50* derivative *MuM* became spermine-resistant and lost spermine/beta-lactam synergy by a specific mutation on PBP2. We therefore further tested the hypothesis that spermine may interfere with PBP(s) or the coordinated complex in cell wall synthesis. Three possible ways were selected, by looking at cell wall composition, by understanding the expression of PBPs on transcriptional level and translational levels upon challenge by spermine or beta-lactam alone or in combination and looking at the possible interactions between spermine and PBPs with the recombinant PBPs from *Mu50* analyzed via various biochemical approaches. Down-regulation of various carbohydrate metabolism and transport genes was observed. Specifically, the expression of genes like *phosphoglycerate mutase, glyceraldehyde-3-phosphate dehydrogenase, phosphoglycerate kinase, phosphoglyceromutase, glucose-6-phosphate isomerase, phosphoglycerate mutase, fructose-bisphosphatase, fructose-1, 6-bisphosphate aldolase* and *hexulose-6-phosphate synthase* were reduced in *Mu50* strain with spermine treatment at 15, 30 and 60-minute time points respectively. These genes and processes were unaffected in *MuM* strain with spermine treatment. Polycationic molecules like polyamines

play important role in maintaining conformation of negatively charged nucleic acids. Bacteria's have evolved transport systems that can allow uptake of extracellular polyamines. Polyamine spermidine can bind to PotD protein causing conformational change triggering a subsequent conformational change in PotBC leading to ATP hydrolysis and spermidine uptake [186]. Mutations in any of the genes in the *potABCD* operon abolish polyamine transport [186]. After comparing *MuM* strain with spermine treatment at 15, 30 and 60-minute time points, a general trend in down-regulation of *potABCD* operon was observed in all the time-points. The similar pattern was seen in *MuM* without spermine treatment at 0-minute time point but with low fold expression which was amplified by an at least one-fold decrease on spermine treatment. Our data show that there may exist some interaction between potABCD proteins of spermine transport system and PBP2 protein causing down-regulation of *potABCD* operon on spermine stress in PBP2 mutant strain compared to wild-type *Mu50* strain. Co-immunoprecipitation or Bio-molecular fluorescence (BiFC) techniques would be needed to confirm these interactions. Potassium uptake and regulation in bacteria is carried out with kdp operon [188]. *KdpABC* genes constitute *Kdp* operon whose expression is regulated by the KdpDE two-component system in response to potassium limitation or turgor pressure [189]. The spermine can enter the cell and occlude the transmembrane pore domain of inwardly rectifying potassium channel (Kir) [200]. When supplemented with high concentration potassium, we have seen an increase in spermine resistance by counteracting spermine obstruction on the uptake channel in wild-type *Mu50* strain [166]. Addition of potassium (300 mM KCl) has been seen to increase Spermine MIC up to 32-fold [166]. In *MuM* strain, we see a significant down-regulation of *KdpABC* operon genes at 15, 30 and 60-minute time points on spermine stress which seems to be absent without spermine treatment. Since KCL has been found to protect the cell against spermine stress in wild-type strain by induction of

KdpABC operon, it fails to do so in *MuM* strain underlying the importance of PBP2 protein in spermine stress.

5.3.1.5 No significant stimulation of cell-wall-stress stimulon and peptidoglycan synthesis genes in Mu50 strain with only spermine, or only oxacillin or with combination treatment:

Mu50 strain at two hour-time point with only spermine treatment, with only oxacillin and with spermine and oxacillin in combination were compared with *Mu50* two hour-time point without spermine and oxacillin treatment. A significant stimulation of cell-wall-stress stimulon and peptidoglycan synthesis genes with only spermine, or only oxacillin or with combination treatment was not seen. Further analysis of *sigB* regulon, iron, potassium and polyamine transporter genes status was performed. With only spermine and only oxacillin treatment, over 90 percent of genes were found to be insignificant, while 45 percent of genes were found to be down regulated in spermine with oxacillin treatment. This finding was overlapping with our previous finding where there was down regulation of *sigB* regulon to high dose spermine stress [166]. We further saw the induction of iron regulation genes with the spermine-oxacillin combination, while both potassium and polyamine transport genes were found to be insignificant in all the conditions. Cell-wall-active antibiotics can stimulate regulons controlled by RNA polymerase sigma factors. A cell-wall-stress stimulon in *Staphylococcus aureus* has been reported by transcriptional profiling in response to cell wall-active antibiotics. 98 genes upregulated by oxacillin treatment are noticed as a cell-wall-stress stimulon in *Staphylococcus aureus* [192]. These genes are involved in several processes like coenzyme metabolism, energy production, and conversion, lipid metabolism, secondary metabolites biosynthesis, transport and catabolism, translation, ribosomal structure, and

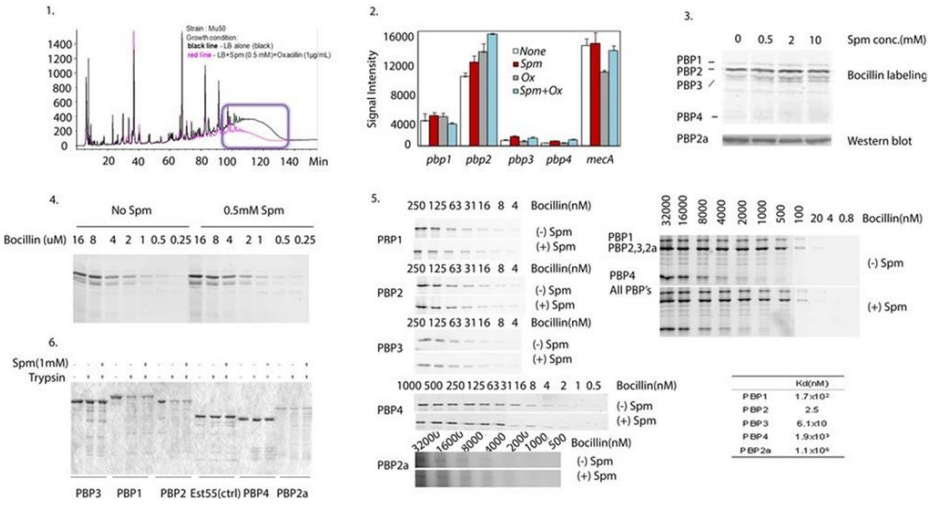
biogenesis. 29 such genes to be up-regulated with a fold change of greater than 1.5 with only spermine treatment were identified. Most of these genes showed similar or higher up-regulation with a spermine-oxacillin synergy at the one-hour time point. Genes potentially related to cell-wall biosynthesis like *murI* (Glutamate racemase), *murZ* (UDP-N-acetylglucosamine 1-carboxylvinyl transferase 2), *PBP2*, *sgtB* (Hypothetical protein like penicillin-binding protein 1A/1B) were not seen to be significantly upregulated with spermine and spermine-oxacillin synergy. Oxacillin binds to PBPs to inhibit the cross-linking step of peptidoglycan biosynthesis [192]. It has been proposed that the cell responds to inhibition of peptidoglycan synthesis by increasing the transcription of the *pbpB* gene to increase the rate of peptidoglycan synthesis [193]. MurZ is the first protein involved in the biosynthesis of UDPN-acetylmuramyl pentapeptide [193]. We didn't find a significant stimulation of cell-wall-stress stimulon in *Mu50* strain by only spermine treatment or with spermine-oxacillin synergy as only 29 percent genes were significantly expressed. Also, we didn't see a significant stimulation of cell-wall biosynthesis genes with only spermine and with spermine-oxacillin treatment, suggesting spermine alone or in combination with oxacillin can weaken cell wall without stimulating genes in cell wall synthesis. When we compared *Mu50* and *MuM* strains with spermine and without oxacillin treatment, a general trend in down-regulation of cell structure biosynthesis genes was seen in the *Mu50* strain which was absent in *MuM*. The peptidoglycan synthesis genes like *bifunctional N-acetylglucosamine-1-phosphate, phosphoglucosamine mutase, UDP-N-acetylglucosamine 2-epimerase, phospho-N-acetylmuramic-pentapeptide-transferase, UDP-N-acetylmuramoylalanine-D-glutamate ligase, UDP-N-acetyl muramate-L-alanine ligase, D-alanine-D-alanine ligase*, and *UDP-N-acetylglucosamine 1-carboxyvinyltransferase* were down-regulated in *Mu50* strain. Some of these genes seemed to be down-regulated in *MuM*, but a general trend in down-regulation was consistent in *Mu50* when

compared to *MuM*. Lipid metabolism genes like *NADPH-dependent glycerol-3-phosphate dehydrogenase, phosphatidylglycerophosphate synthase, enoyl reductase, an acyl carrier protein, 3-oxoacyl-ACP synthase* were seen to be down-regulated in *Mu50* treated with spermine. It has been proposed that spermine can form a complex with PBP2 protein eliciting conformational change to weaken interactions among enzymes in the complex and inhibit the trans-glycosylase activity of PBP2 in cell wall synthesis by increasing the transcription of the *pbpB* gene to increase the rate of peptidoglycan synthesis [166]. Given our observations, it seems to contradict this hypothesis as there is down-regulation of peptidoglycan synthesis genes in Mu50 on spermine stress. Further analysis of these genes with real-time PCR can provide more insights on validating these observations.

5.3.1.6 Analysis of spermine-dependent synergy with beta-lactams on cell wall synthesis: The synergy between spermine and oxacillin reduces the degree of cross-linkage on cell wall:

Peptidoglycan was digested with M1 muramidase, and the fragments were analyzed by HPLC to determine the composition and degree of cell wall cross-linkage in spermine/oxacillin-grown cells. As shown in Figure 19.1, there were no major changes seen in the biochemical composition of PG in the presence of spermine (0.5mM) alone, oxacillin (1ng/µl) alone or their combination.

Figure 19 Chromatograms of peptidoglycan analysis by HPLC. Effect of Spm/beta-lactam on the muropeptide composition and degree of cross-linking of peptidoglycan extracted from S. aureus Mu50.



However, we do saw differences in the degree of cross-linking between glycan chains within the samples. The combination treatment results in a more depressed hump (retention time 110 minutes and after) in this sample, while samples of spermine or oxacillin alone do not cause any significant changes on the chromatograms (data not shown). Depressed humps on chromatogram have also been encountered in Staphylococcus aureus at high doses of beta-lactams that block the biochemical functions of PBPs [160]. Since PG cross-linking is catalyzed by PBPs, this result suggests an enhanced PBP inactivation by oxacillin-spermine synergy. The PBPs contribution in cell wall synthesis can be affected by oxacillin/spermine synergy at multiple levels of transcription and translation; the binding affinity with beta-lactams; the enzymatic activity, which may result from irregular conformation, incorrect localization, masked active sites, etc. To study if PBP2a or the PBPs complex is involved in the synergism, we then tested if spermine suppresses PBP's expression; whether spermine enhances the PBPs-beta-lactam binding and whether spermine can directly interact with PBPs to influence their activities. We did not evaluate whether spermine can delocalize PBP(s) or prevent their recruitment to the septum due to some technical limitations.

5.3.1.7 Transcriptional or translational expression of PBPs is not seen to be affected by spermine treatment:

We compared microarray gene expression levels of PBPs in *MRSA Mu50* treated with spermine (1mM), oxacillin (16ng/µl), or in combination. As seen in Figure 19.2 (signal intensities of different PBP expression), no significant differences were detected between untreated and treated transcripts, suggesting low-dose spermine by itself or in combination with beta-lactams may not affect the *PBP* transcription. To examine whether spermine can influence the PBPs protein levels, lysates from *Mu50* grown in the presence of different concentrations of spermine (0.5mM, 2mM, 10mM) were labeled to Bocillin labeling (for detection of PBP1-4) or Western blot (for detection of PBP2a). Again, we didn't see any significant changes in the quantities of any PBPs as seen in Figure 19.3. Thus, we conclude that spermine may not affect PBPs expression either at the transcriptional or translational level. The PBP2 mutant strain *MuM* became spermine resistant and showed no spermine/beta-lactam synergy which suggested that PBPs could be the target for spermine. In line with this, the reduced PG crosslinking by a combination of low dose spermine and oxacillin further supports the hypothesis of spermine perturbing cell wall integrity in favor of beta-lactam efficacy. However, a series of experiments investigating spermine effects on PBPs, the beta-lactam targets, implies that spermine does not appear to suppress the PBPs expression or alter their interactions with beta-lactams. Besides, no apparent stimulation of cell-wall-stress stimulon in *Mu50* strain by only spermine treatment or with spermine-oxacillin synergy was observed (as only 29 percent of cell wall stimulon genes were significantly expressed). Even so, these results cannot rule out the possibility that the enzymatic activities of PBP-associated multi-enzyme complexes might be modulated directly by binding of spermine to the complexes or indirectly

through delocalization or reconfiguration of the complexes on the membrane. Also, spermine may perturb other cell wall-related factors (instead of PBPs) to achieve the spermine/beta-lactam synergy. Our unpublished data showing other spontaneous mutants with mutation not locating on pbp2 also lose the synergy like *MuM*, suggest that other gene defects may confer similar impacts as that of PBP2 deficiency in *MuM* to block the spermine target. Although we did not see the up-regulation of cell wall stimulon by the spermine/oxacillin combination, the induced expression of iron uptake genes in *Mu50* strain drew our attention. Iron and H2O2 or its precursor superoxide (O2-) are required for the Fenton reaction to generate destructive hydroxyl radicals (-OH), which can trigger a common oxidative damage cellular death pathway by increasing the transcription of the *pbpB* gene to increase the rate of peptidoglycan synthesis [193] [194]. Bactericidal antibiotics including beta-lactams can trigger superoxide production through destabilization of the electron transport chain and in our transcriptome data, iron/heme acquisition was highly stimulated by spermine. Therefore, one possibility was that the enhanced ferrous iron uptake by spermine, together with superoxide generated by beta-lactams significantly intensifies the Fenton reaction and subsequently causes cell death. There is also a possibility of beta-lactams augmenting spermine effects during the synergy. The extensively affected energy production pathways (glycolysis, fermentation, TCA cycle) by spermine may also perturb the electron transport chain. If so, one presumption is: beta-lactam inhibited cell wall synthesis will disturb membrane potential which cooperates with the possible deficiency on the electron transport chain endowed by spermine effect to kill the bacteria. In conclusion, a significant down-regulation in an iron regulatory system, potassium channel uptake and polyamine transport system with an up-regulation in general stress response *sigB* dependent operon in *MuM* strain at 15, 30 and 60-minutes time points with spermine treatment when compared to *Mu50* is observed. The electron

transfer chain, carbohydrate biosynthesis, cell structure fatty acid and lipid biosynthesis systems were seen to be down-regulated in *Mu50* and unaffected in *MuM* strain on spermine stress. Furthermore, analysis of spermine-dependent synergy with beta-lactams on cell wall synthesis revealed that it significantly reduces the degree of cross-linkage on cell wall with no change in trypsin digestion patterns of purified PBPs and without affecting PBPs expression or PBPs acylation by Bocillin.

5.3.1.8 Spermine does not help in enhancing the binding of beta-lactams to PBPs:

To evaluate if spermine can enhance binding of beta-lactams to PBPs binding during synergy, PBPs in *Staphylococcus aureus* lysate or purified PBPs were incubated with Bocillin (fluorescent beta-lactam derivative) with or without spermine treatment. Lysates from *Mu50* grown in the medium free of spermine or oxacillin were pre-incubated with 0.5mM spermine and subjected to labeling by Bocillin of indicated concentrations. The binding affinities were estimated by measuring the fluorescence intensities of acylated PBPs by Bocillin. As illustrated in figure 19.4, the binding kinetics of Bocillin to PBPs were unaffected by the presence of spermine. We then studied beta-lactam-PBP binding affinity individually or in combination with purified PBPs. As shown in Figure 19.5, we calculated affinity of Bocillin for different PBP's through a monophasic saturation curve. A notably higher dissociation constant (Kd) for PBP2a than those of other PBPs and was consistent with its low affinity to beta-lactams and the resultant high beta-lactam MIC for MRSA was seen. We also compared the individual Kd values between spermine treated and untreated samples where no significant variations were detected. Coomassie staining of the same gel ensured no variations in protein amount between samples (data not shown). Our

results suggest that spermine may not affect PBP acylation by beta-lactams.

5.3.1.9 Binding of spermine to PBPs does not cause conformational changes to PBPs, as tested with trypsin digestion patterns:

Binding of spermine to PBPs may cause conformational changes causing sensitivity of ligand-protein complexes against protease digestion. So, we utilized DARTS (drug affinity responsive target stability) method to look for possible conformation changes of PBPs due to spermine interaction. Purified PBP was pre-incubated with or without spermine and subjected to trypsin digestion following a time course with subsequent SDS-PAGE separation. The unrelated carboxylesterase Est55 served as control. As shown in Figure 32.6, no significant difference on the digestion patterns were observed regardless of spermine treatment.

REFERENCES

1. World Health Organization., 2016. Retrieved from https://www.who.int/hiv/en/.

2. Gao, F.; Bailes, E.; Robertson, D.L.; Chen, Y.; Rodenburg, C.M.; Michael, S.F.; Cummins, L.B.; Arthur, L.O.; Peeters, M.; Shaw, G.M.; Sharp, P.M.; Hahn, B.H. Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. Nature 1999, 397, 436–441.

3. Gao, F.; Yue, L.; White, A.T.; Pappas, P.G.; Barchue, J.; Hanson, A.P.; Greene, B.M.; Sharp, P.M.; Shaw, G.M.; Hahn, B.H. Human infection by genetically diverse SIVSM-related HIV-2 in west Africa. Nature 1992, 358, 495–499.

4. Van Heuverswyn, F.; Li, Y.; Neel, C.; Bailes, E.; Keele, B.F.; Liu, W.; Loul, S.; Butel, C.; Liegeois, F.; Bienvenue, Y.; Ngolle, E.M.; Sharp, P.M.; Shaw, G.M.; Delaporte, E.; Hahn, B.H.; Peeters, M. Human immunodeficiency viruses: SIV infection in wild gorillas. Nature 2006, 444, 164.

5. Takehisa, J.; Kraus, M.H.; Ayouba, A.; Bailes, E.; Van Heuverswyn, F.; Decker, J.M.; Li, Y.; Rudicell, R.S.; Learn, G.H.; Neel, C.; Ngole, E.M.; Shaw, G.M.; Peeters, M.; Sharp, P.M.; Hahn, B.H. Origin and biology of simian immunodeficiency virus in wild-living western gorillas. J. Virol. 2009, 83, 1635–1648.

6. Ayouba, A.; Mauclère, P.; Martin, P.M.; Cunin, P.; Mfoupouendoun, J.; Njinku, B.; Souquières, S.; Simon, F. HIV-1 group O infection in Cameroon, 1986 to 1998. Emerg. Infect. Dis. 2001, 7, 466–467.

7. Simon, F.; Mauclère, P.; Roques, P.; Loussert-Ajaka, I.; Müller-Trutwin, M.C.; Saragosti, S.; Georges-Courbot, M.C.; Barré-Sinoussi, F.; Brun-Vénizet, F. Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. Nat. Med. 1998, 4, 1032–1037.

8. Roques, P.; Robertson, D.L.; Souquière, S.; Apetrei, C.; Nerrienet, E.; Barré-Sinoussi, F.; Müller-Trutwin, M.; Simon, F. Phylogenetic characteristics of three new HIV-1 N strains and implications for the origin of group N. AIDS 2004, 18, 1371–1381.

9. HIV sequence database. The Human Retroviruses and AIDS 1994 Compendium. III23-III33. Available online: Http://www.hiv.lanl.gov/content/sequence/HIV/COMPENDIUM/94compendium.html (accessed on 18 July 2009).

10. Robertson, D.L.; Anderson, J.P.; Bradac, J.A.; Carr, J.K.; Foley, B.; Funkhouser, R.K.; Gao, F.; Hahn, B.H.; Kalish, M.L.; Kuiken, C.; Learn, G.H.; Leitner, T.; McCutchan, F.; Osmanov, S.; Peeters, M.; Pieniazek, D.; Salminen, M.; Sharp, P.M.; Wolinsky, S.; Korber, B. HIV-1 nomenclature proposal. Science 2000, 288, 55–56.

11. Gao, F.; Vidal, N.; Li, Y.; Trask, S.A.; Chen, Y.; Kostrikis, L.G.; Ho, D.D.; Kim, J.; Oh, M.D.; Choe, K.; Salminen, M.; Robertson, D.L.; Shaw, G.M.; Hahn, B.H.; Peeters, M. Evidence of two distinct subsubtypes within the HIV-1 subtype A radiation. AIDS Res. Hum. Retroviruses 2001, 17, 675–688.

12. Hemelaar, J.; Gouws, E.; Ghys, P.D.; Osmanov, S. Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. AIDS 2006, 20, W13–W23.

13. Ntemgwa, M.L.; Toni, T.D.; Brenner, B.G.; Camacho, R.J.; Wainberg, M.A. Antiretroviral Drug Resistance in Human Immunodeficiency Virus Type 2 (HIV-2). Antimicrob. Agents Chemother. 2009, 53, 3611–3619.

14. Kilby, J.M., et al., Potent suppression of HIV-1 replication in humans by T-20, a peptide inhibitor of gp41-mediated virus entry. Nat Med, 1998. 4(11): p. 1302-7.

15. Tan Q, Z.Y., Li J, Chen Z, Han GW, Kufareva I, Li T, Ma L, Fenalti G, Li J, Zhang W, Xie X, Yang H, Jiang H, and L.H. Cherezov V, Stevens RC, Zhao Q, Wu B, Structure of the CCR5 chemokine receptor-HIV entry inhibitor maraviroc complex. Science, 2013: p. 1387-1390.

16. Michailidis, E., et al., 4'-Ethynyl-2-fluoro-2'-deoxyadenosine (EFdA) inhibits HIV-1 reverse transcriptase with multiple mechanisms. J Biol Chem, 2014. 289(35): p. 24533-48.

17. Das, K., et al., Roles of conformational and positional adaptability in structure-based design of TMC125-R165335 (etravirine) and related non-nucleoside reverse transcriptase inhibitors that are highly potent and effective against wild-type and drug-resistant HIV-1 variants. J Med Chem, 2004. 47(10): p. 2550-60.

18. Young, T.P., et al., Prevalence, mutation patterns, and effects on protease inhibitor susceptibility of the L76V mutation in HIV-1 protease. Antimicrob Agents Chemother, 2010. 54(11): p. 4903-6.

19. Wensing, A. M., Calvez, V., Günthard, H. F., Johnson, V. A., Paredes, R., Pillay, D., Shafer, R. W., … Richman, D. D. (2016). 2017 Update of the Drug Resistance Mutations in HIV-1. Topics in antiviral medicine, 24(4), 132-133.

20. Wang Y, Liu Z, Brunzelle JS, et al. The higher barrier of darunavir and tipranavir resistance for HIV-1 protease. Biochem Biophys Res Commun. 2011;412(4):737–742.

21. Goldfarb NE, Ohanessian M, Biswas S, et al. Defective hydrophobic sliding mechanism and active Site expansion in HIV-1 protease drug resistant variant Gly48Thr/Leu89Met: mechanisms for the loss of saquinavir binding potency. Biochemistry. 2014;54(2):422–433.

22. Louis JM, Zhang Y, Sayer JM, et al. The L76V drug resistance mutation decreases the dimer stability and rate of autoprocessing of HIV-1 protease by reducing internal hydrophobic contacts. Biochemistry. 2011;50(21):4786–4795.

23. Kozísek M, Henke S, Sasková KG, et al. Mutations in HIV-1 Gag and Pol compensate for the loss of viral fitness caused by a highly mutated protease. Antimicrob Agents Chemother. 2012;56(8):4320–4330.

24. Agniswamy J, Shen C-H, Wang Y-F, et al. Extreme multidrug resistant HIV-1 protease with 20 mutations is resistant to novel protease inhibitors with P1'-pyrrolidinone or P2-tris-tetrahydrofuran. J Med Chem. 2013;56(10):4017–4027. Highly drug resistant mutant PR20 is assessed with new antiviral inhibitors designed for resistant HIV.

25. Zhang Y, Chang Y-CE, Louis JM, et al. Structures of DRV-resistant HIV-1 protease mutant reveal atypical binding of DRV to wide open flaps. ACS Chem Biol. 2014;9(6):1351–1358.

26. Tang, C., et al., Visualizing transient events in amino-terminal autoprocessing of HIV-1 protease. Nature, 2008. 455(7213): p. 693-6.

27. Wlodawer, A., et al., Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. Science, 1989. 245(4918): p. 616-21.

28. Kovalevsky, A.Y., et al., Caught in the Act: the 1.5 A resolution crystal structures of the HIV-1 protease and the I54V mutant reveal a tetrahedral reaction intermediate. Biochemistry, 2007. 46(51): p. 14854-64.

29. Perryman, A. L., Lin, J. H., & McCammon, J. A. (2004). HIV-1 protease molecular dynamics of a wild-type and of the V82F/I84V mutant: possible contributions to drug resistance and a potential new target site for drugs. Protein science : a publication of the Protein Society, 13(4), 1108-23.

30. Weber IT, Agniswamy J. HIV-1 protease: structural perspectives on drug resistance. Viruses. 2009;1(3):1110–1136. Review of the molecular mechanisms suggested by structural and inhibition studies of HIV proteases with single amino acid substitutions.

31. IT, W., Can we design drugs for HIV/AIDS that are less susceptible to resistance? Future Med Chem, 2015.

32. Weber, I.T., D.W. Kneller, and A. Wong-Sam, Highly resistant HIV-1 proteases and strategies for their inhibition. Future Med Chem, 2015. 7(8): p. 1023-38.

33. Menendez-Arias L, Tozser J. HIV-1 protease inhibitors: effects on HIV-2 replication and resistance. Trends Pharmacol Sci. 2008;29:42–49.

34. Hoog SS, Towler EM, Zhao B, Doyle ML, Debouck C, Abdel-Meguid SS. Human immunodeficiency virus protease ligand specificity conferred by residues outside of the active site cavity. Biochemistry. 1996;35:10279–10286.

35. Swairjo MA, Towler EM, Debouck C, Abdel-Meguid SS. Structural role of the 30′s loop in determining the ligand specificity of the human immunodeficiency virus protease. Biochemistry. 1998;37:10928–10936.

36. Tie, Y., et al., Critical differences in HIV-1 and HIV-2 protease specificity for clinical inhibitors. Protein Sci, 2012. 21(3): p. 339-50.

37. Kovalevsky, A.Y., et al., Structural evidence for effectiveness of darunavir and two related antiviral inhibitors against HIV-2 protease. J Mol Biol, 2008. 384(1): p. 178-92.

38. Chen, J., Drug resistance mechanisms of three mutations V32I, I47V and V82I in HIV-1 protease toward inhibitors probed by molecular dynamics simulations and binding free energy predictions. RSC Advances, 2016.

39. Annemarie M. Wensing, M., PhD; Vincent Calvez, MD, PhD; Huldrych F. Günthard, MD; Victoria, M.R.P. A. Johnson, MD, PhD; Deenan Pillay, MD, PhD; Robert W. Shafer, MD;, and M. Douglas D. Richman, 2017 Update of the Drug Resistance Mutations in HIV-1. 2017 Resistance Mutations Update, 2017. 24(4).

40. Ghosh, A.K., et al., Design of HIV-1 Protease Inhibitors with Amino-bis-tetrahydrofuran Derivatives as P2-Ligands to Enhance Backbone-Binding Interactions: Synthesis, Biological Evaluation, and Protein-Ligand X-ray Studies. J Med Chem, 2015. 58(17): p. 6994-7006.

41. Amano, M., et al., GRL-0519, a novel oxatricyclic ligand-containing nonpeptidic HIV-1 protease inhibitor (PI), potently suppresses replication of a wide spectrum of multi-PI-resistant HIV-1 variants in vitro. Antimicrob Agents Chemother, 2013. 57(5): p. 2036-46.

42. Amano M, T.Y., Salcedo-Gómez PM, Parham GL, Nyalapatla PR, Das D3, Ghosh AK, Mitsuya H, A novel tricyclic ligand-containing nonpeptidic HIV-1 protease inhibitor, GRL-0739, effectively inhibits the replication of multidrug-resistant HIV-1 variants and has a desirable central nervous system penetration property in vitro. Antimicrob Agents Chemother, 2015: p. 2625-35.

43. Larder B. Quantitative prediction of HIV-1 phenotypic drug resistance from genotypes: the virtual phenotype (VirtualPhenotype) Antiviral Ther. 2000;15:63–63.

44. Puchhammer‡Stöckl E, Steininger C, Geringer E, Heinz F. Comparison of virtual phenotype and HIV-SEQ program (Stanford) interpretation for predicting drug resistance of HIV strains. HIV Med. 2002;15:200–206. doi: 10.1046/j.1468-1293.2002.00116.x.

45. Schmidt B, Walter H, Moschik B, Paatz C, Van Vaerenbergh K, Vandamme A-M. et al. Simple algorithm derived from a geno-/phenotypic database to predict HIV-1 protease inhibitor resistance. AIDS. 2000;15:1731–1738. doi: 10.1097/00002030-200008180-00007.

46. Yu, X., Weber, I. T., & Harrison, R. W. (2014). Prediction of HIV drug resistance from genotype with encoded three-dimensional protein structure. BMC genomics, 15 Suppl 5(Suppl 5), S1.

47. Sevin AD, DeGruttola V, Nijhuis M, Schapiro JM, Foulkes AS, Para MF. et al. Methods for investigation of the relationship between drug-susceptibility phenotype and human immunodeficiency virus type 1 genotype with applications to AIDS clinical trials group 333. J Infect Dis. 2000;15:59–67. doi: 10.1086/315673

48. DiRienzo G, DeGruttola V, Larder B, Hertogs K. Nonparametric methods to predict HIV drug susceptibility phenotype from genotype. Stat Med. 2003;15:2785–2798. doi: 10.1002/sim.1516.

49. Cimolai, N., MRSA and the environment: implications for comprehensive control measures. Eur J Clin Microbiol Infect Dis, 2008. 27(7): p. 481-93.

50. Masalha M, Borovok I, Schreiber R, Aharonowitz Y, Cohen G (December 2001). "Analysis of transcription of the Staphylococcus aureus aerobic class Ib and anaerobic class III ribonucleotide reductase genes in response to oxygen". Journal of Bacteriology. 183 (24): 7260–72. doi:10.1128/JB.183.24.7260-7272.2001. PMC 95576. PMID 11717286.

51. Allard C , Carignan A , Bergevin M , Boulais I , Tremblay V , Robichaud P , Duperval R , Pepin J . 2008. Secular changes in incidence and mortality associated with Staphylococcus

aureus bacteraemia in Quebec, Canada, 1991-2005. Clin Microbiol Infect 14:421–428. doi:10.1111/j.1469-0691.2008.01965.x.

52. Laupland KB , Ross T , Gregson DB . 2008. Staphylococcus aureus bloodstream infections: risk factors, outcomes, and the influence of methicillin resistance in Calgary, Canada, 2000-2006. J Infect Dis 198:336–343. doi:10.1086/589717.

53. Wyllie DH , Peto TE , Crook D . 2005. MRSA bacteraemia in patients on arrival in hospital: a cohort study in Oxfordshire 1997-2003. BMJ 331:992. doi:10.1136/bmj.38558.453310.8F.

54. Cimolai N. MRSA and the environment: implications for comprehensive control measures. Eur J Clin Microbiol Infect Dis 2008;27:481–93.

55. Enright MC, Robinson DA, Randle G, Feil EJ. The evolutionary history of methicillin-resistant Staphylococcus aureus. Proc Natl Acad Sci 2002;99:7687–92.

56. Hartman BJ, Tomasz A. Low-affinity penicillin-binding protein associated with beta-lactam resistance in Staphylococcus aureus. J Bacteriol 1984;158:513–6.

57. Marzabadi MR. Spermine prevent iron accumulation and depress lipofuscin accumulation in cultured myocardial cells. Free Radic Biol Med 1996;21:375–81.

58. Enright, M.C., et al., The evolutionary history of methicillin-resistant Staphylococcus aureus (MRSA). Proc Natl Acad Sci U S A, 2002. 99(11): p. 7687-92.

59. Yao, X. and C.D. Lu, A PBP 2 mutant devoid of the transpeptidase domain abolishes spermine-beta-lactam synergy in Staphylococcus aureus Mu50. Antimicrob Agents Chemother, 2012. 56(1): p. 83-91.

60. Polyamine Effects on Antibiotic Susceptibility in Bacteria. Antimicrob Agents Chemotherapy, 2007.

61. Pootoolal, J., J. Neu, and G.D. Wright, Glycopeptide antibiotic resistance. Annu Rev Pharmacol Toxicol, 2002. 42: p. 381-408.

62. Estimates of global, regional, and national incidence, prevalence, and mortality of hiv, 1980–2015: the global burden of disease study 2015. The Lancet HIV. 2016;3(8):361–87. https://doi.org/10.1016/S2352-3018(16)30087-X.

63. World Health Organization HIV Page. http://www.who.int/hiv/data/en/. Accessed 6 June 2018.

64. Smyth RP, Davenport MP, Mak J. The origin of genetic diversity in hiv-1. Virus Res. 2012;169(2):415–29. https://doi.org/10.1016/j.virusres.2012.06.015. Retroviral RNA, protein co-factors and chaperones.

65. Yu X, Weber I, Harrison R. Sparse representation for hiv-1 protease drug resistance prediction. In: Proceedings of the 2013 SIAM International Conference on Data Mining. SIAM; 2013. p. 342–9. https://doi.org/10.1137/1.9781611972832.38.

66. Khalid Z, Sezerman O. Prediction of HIV Drug Resistance by Combining Sequence and Structural Properties. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 2018; 966-973.

67. Yu X, Weber IT, Harrison RW. Identifying representative drug resistant mutants of hiv. BMC Bioinforma. 2015;16(17):1.

68. Durham EEA, Yu X, Harrison RW. FDT 2.0: Improving scalability of the fuzzy decision tree induction tool - integrating database storage. In: 2014 IEEE Symposium on Computational Intelligence in Healthcare and e-health (CICARE); 2014. p. 187–90. https://doi.org/10.1109/CICARE.2014.7007853.

69. Shen C, Yu X, Harrison R, Weber IT. Automated prediction of hiv drug resistance from genotype data. BMC Bioinforma. 2016;17(8):278.

70. Tingjun H, Wei Z, Jian W, Wei W. Predicting drug resistance of the hiv-1 protease using molecular interaction energy components. Proteins Struct Funct Bioinforma. 2014;82(6):1099.

71. Sheik Amamuddy O, Bishop NT, Tastan Bishop Ö. Improving fold resistance prediction of hiv-1 against protease and reverse transcriptase inhibitors using artificial neural networks. BMC Bioinforma. 2017;18(1):369. https://doi.org/10.1186/s12859-017-1782-x.

72. Masso M, Vaisman II. Sequence and structure based models of hiv-1 protease and reverse transcriptase drug resistance. BMC Genomics. 2013;14(4):3.

73. Bose P, Yu X, Harrison RW. Encoding protein structure with functions on graphs. In: 2011 IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW). IEEE; 2011. p. 338–44. https://doi.org/10.1109/BIBMW.2011.6112396.

74. Park JH, Sayer JM, Aniana A, Yu X, Weber IT, Harrison R, Louis JM. Binding of clinical inhibitors to a model precursor of a rationally selected multidrug resistant hiv-1 protease is significantly weaker than that to the released mature enzyme. Biochemistry. 2016;55(16):2390–400.

75. Agniswamy J, Louis JM, Roche J, Harrison R, Weber IT. Structural studies of a rationally selected multi-drug resistant hiv-1 protease reveal synergistic effect of distal mutations on flap dynamics. PLoS ONE. 2016;11(12):0168616.

76. Hinton G. UTML TR 2010–003 A Practical Guide to Training Restricted Boltzmann Machines. Toronto: University of Toronto; 2010. http://www.cs.toronto.edu/~hinton/absps/guideTR.pdf. Accessed 2017.

77. Salakhutdinov R, Mnih A, Hinton G. Restricted boltzmann machines for collaborative filtering. In: Proceedings of the 24th International Conference on Machine Learning. ICML '07. New York: ACM; 2007. p.791–8. https://doi.org/10.1145/1273496.1273596.

78. Harrison RW, Freas C. Fuzzy restricted boltzmann machines. In: Melin P, Castillo O, Kacprzyk J, Reformat M, Melek W, editors. Fuzzy Logic in Intelligent System Design. Cham: Springer International Publishing; 2018. p. 392–8.

79. Harrison R, McDermott M, Umoja C. Recognizing protein secondary structures with neural networks. In: 2017 28th International Workshop on Database and Expert Systems Applications (DEXA). DEXA; 2017. p. 62–8.https://doi.org/10.1109/DEXA.2017.29.

80. Rhee S-Y, Gonzales MJ, Kantor R, Betts BJ, Ravela J, Shafer RW. Human immunodeficiency virus reverse transcriptase and protease sequence database. Nucleic Acids Res. 2003;31(1):298–303.

81. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. Nucleic Acids Res. 2000;28(1):235–42. https://doi.org/10.1093/nar/28.1.235.

82. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, et al. Scikit-learn: Machine learning in python. J Mach Learn Res. 2011;12(Oct):2825–30.

83. McKinney W. pandas: a foundational python library for data analysis and statistics. In: Python for High Performance and Scientific Computing. Sebastopol: O'Reilly Media; 2011. p. 1–9.

84. Abu-halaweh NM, Harrison RW. Practical fuzzy decision trees. In: 2009 IEEE Symposium on Computational Intelligence and Data Mining; 2009. p. 211–6. https://doi.org/10.1109/CIDM.2009.4938651.

85. Wensing AMJ, van Maarseveen NM, Nijhuis M. Fifteen years of hiv protease inhibitors: raising the barrier to resistance. Antivir Res. 2010;85(1): 59–74. https://doi.org/10.1016/j.antiviral.2009.10.003. Twenty-five Years of Antiretroviral Drug Development: Progress and Prospects.

86. Center for Disease Control., Prevention: HIV in the United States: At A Glance

87. E, D.C.: Anti-hiv drugs: 25 compounds approved within 25 years after the discovery of hiv. Int J Antimicrob Agents 33(4), 307–20 (2009)

88. Lendeckel U, H.N.: Viral proteases and antiviral protease inhibitor therapy. Springer (2009)

89. Kohl NE, Emini EA: Active human immunodeficiency virus protease is required for viral infectivity. Proc Natl Acad Sci U S A 85(13), 4686–90 (1998)

90. Ahuja TS, G.J. Borucki M: Highly active antiretroviral therapy improves survival of hiv-infected hemodialysis patients. Am J Kidney Dis 36(3), 574–80 (2010)

91. Wang K, S.R.M.J. Jenwitheesuk E: Simple linear model provides highly accurate genotypic predictions of hiv-1 drug resistance. Antivir Ther 9(3), 343–52 (2004)

92. Beerenwinkel N, Schmidt B: Diversity and complexity of hiv-1 drug resistance: a bioinformatics approach to predicting phenotype from genotype. Proc Natl Acad Sci U S A 99(12), 8271–6 (2002)

93. Yu X, H.R. Weber I: Sparse representation for hiv-1 protease drug resistance prediction. SIAM International Conference on Data Mining (SDM13), 342–349 (2013)

94. Lengauer. Characterization of novel hiv drug resistance mutations using clustering, multidimensional scaling and svm-based feature ranking. European Conference on Principles of Data Mining and Knowledge Discovery, 285–296 (2005) 34. Xiaxia Yu, I.T.W., Harrison, R.W.: Identifying representative drug resistant mutants of hiv. BMC Bioinformatics 16(Suppl 17)(S1) (2015)

95. H M Berman, J Westbrook: The protein data bank. Nucleic Acids Research 28, 235–242 (2000)

96. Cortes C, V.V.: Support-vector networks. Machine Learning 20(3), 273–297 (1995)

97. Fabian Pedregosa. Ga¨el Varoquaux: Scikit-learn: ´ Machine learning in python. Journal of Machine Learning Research 12, 2825–2830

98. Rokach Lior, O.M.: Clustering methods. Data mining and knowledge discovery handbook, 321–352 (2005)

99. Institute, S.: The distance procedure: Proximity measures. SAS/STAT 9.2 Users Guide (2009)

100. Peter Rousseeuw, M.H. Anja Struyf: Clustering in an object-oriented environment. Journal of Statistical Software (1997)

101. Kriegel, S.E.Z.A. Hans Peter: The (black) art of runtime evaluation: Are we comparing algorithms or implementations? Knowledge and Information Systems 52, 341–378 (2016)

102. Martin Maechler, A.S.M.H.K.H. Peter Rousseeuw: cluster: Cluster analysis basics and extensions. R package version 2.0.6 (2017)

103. Winter, B.: The F Distribution and the Basic Principle Behind ANOVAs. http://www.bodowinter.com Accessed 2015

104. Hartigan J A, W.M.A.: A k-means clustering algorithm. Applied Statistics 28, 100–108 (1979)

105. Rhee SY, Gonzales MJ: Human immunodeficiency virus reverse transcriptase and protease sequence database. Nucleic Acids Res 31(1), 298–303 (2003)

106. Xiaxia Yu, R.W.H. Irene T Weber: Prediction of hiv drug resistance from genotype with encoded three-dimensional protein structure. BMC Genomics 15(Suppl 5)(S1) (2014)

107. David J Ketchen, J.C.L.S.: The application of cluster analysis in strategic management research: An analysis and critique. Strategic Management Journal 17(6), 441–458 (1996)

108. Vasavia Ramasamy, P.M. Tamizh selvia: Drug resistance mechanism of l10f, l10f/n88s and l90m mutations in hiv-1 protease: Molecular dynamics simulations and binding free energy calculations. Journal of Molecular Graphics and Modelling 75, 390–402 (2017)

109. Flor Parra, P.J.P.-R.P. P´erez Pulido AJ: The hiv type 1 protease l10i minor mutation decreases replication capacity and confers resistance to protease inhibitors. AIDS Res Hum Retroviruses 27(11)(65-70) (2011)

110. Kuipera Bradley, J Keuscha: The l33f darunavir resistance mutation acts as a molecular anchor reducing the flexibility of the hiv-1 protease 30s and 80s loops. Biochemistry and Biophysics Reports 2, 160–165 (2015)

111. Meiselbach H, Horn AH: Insights into amprenavir resistance in e35d hiv-1 protease mutation from molecular dynamics and binding free-energy calculations. J Mol Model 13(2), 297–304 (2007)

112. Dirauf P, S.H. Meiselbach H: Effects of the v82a and i54v mutations on the dynamics and ligand binding properties of hiv-1 protease. J Mol Model 16(10), 1577–83 (2010)

113. Clemente JC, Hemrajani R: Secondary mutations m36i and a71v in the human immunodeficiency virus type 1 protease can provide an advantage for the emergence of the primary mutation d30n. Biochemistry 42(51), 15029–35 (2003)

114. World Health Organization. 2016. Retrieved from: http://www.who.int/hiv/en/.

115. World Health Organization, http://www.who.int/mediacentre/factsheets/fs360/en/, 2016 (accessed Apr. 2019).

116. J. Hemelaar, Implications of HIV diversity for the HIV-1 pandemic, J Infect 66 (2013) 391-400. https://doi.org/10.1016/j.jinf.2012.10.026.

117. D.S. Clutter, M.R. Jordan, S. Bertagnolio, R.W. Shafer, HIV-1 drug resistance and resistance testing, Infect Genet Evol 46 (2016) 292-307. https://doi.org/10.1016/j.meegid.2016.08.031.

118. L. Menendez-Arias, M. Alvarez, Antiretroviral therapy and drug resistance in human immunodeficiency virus type 2 infection, Antiviral Res 102 (2014) 70-86. https://doi.org/10.1016/j.antiviral.2013.12.001.

119. J. Konvalinka, H.G. Krausslich, B. Muller, Retroviral proteases and their roles in virion maturation, Virology 479-480 (2015) 403-417. https://doi.org/10.1016/j.virol.2015.03.021.

120. C. Tang, J.M. Louis, A. Aniana, J.Y. Suh, G.M. Clore, Visualizing transient events in amino-terminal autoprocessing of HIV-1 protease, Nature 455 (2008) 693-696. https://doi.org/10.1038/nature07342.

121. L. Menendez-Arias, J. Tozser, HIV-1 protease inhibitors: effects on HIV-2 replication and resistance, Trends Pharmacol Sci 29 (2008) 42-49. https://doi.org/10.1016/j.tips.2007.10.013.

122. A.M. Wensing, V. Calvez, H.F. Gunthard, V.A. Johnson, R. Paredes, D. Pillay, R.W. Shafer, D.D. Richman, 2017 Update of the Drug Resistance Mutations in HIV-1, Top Antivir Med 24 (2017) 132-133. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5677049/pdf/tam-24-132.pdf.

123. L. Menendez-Arias, Molecular basis of human immunodeficiency virus type 1 drug resistance: overview and recent developments, Antiviral Res 98 (2013) 93-120. https://doi.org/10.1016/j.antiviral.2013.01.007.

124. I.T. Weber, J. Agniswamy, HIV-1 Protease: Structural Perspectives on Drug Resistance, Viruses 1 (2009) 1110-1136. https://doi.org/10.3390/v1031110.

125. M. Nijhuis, R. Schuurman, D. de Jong, J. Erickson, E. Gustchina, J. Albert, P. Schipper, S. Gulnik, C.A. Boucher, Increased fitness of drug resistant HIV-1 protease as a result of acquisition of compensatory mutations during suboptimal therapy, AIDS 13 (1999) 2349-2359. https://pdfs.journals.lww.com/aidsonline/1999/12030/Increased_fitness_of_drug_resistant_HIV_1_protease.6.pdf.

126. Y. Tie, Y.F. Wang, P.I. Boross, T.Y. Chiu, A.K. Ghosh, J. Tozser, J.M. Louis, R.W. Harrison, I.T. Weber, Critical differences in HIV-1 and HIV-2 protease specificity for clinical inhibitors, Protein Sci 21 (2012) 339-350. https://doi.org/10.1002/pro.2019.

127. H. Zhang, Y.F. Wang, C.H. Shen, J. Agniswamy, K.V. Rao, C.X. Xu, A.K. Ghosh, R.W. Harrison, I.T. Weber, Novel P2 tris-tetrahydrofuran group in antiviral compound 1 (GRL-0519) fills the S2 binding pocket of selected mutants of HIV-1 protease, Journal of medicinal chemistry 56 (2013) 1074-1083. https://doi.org/10.1021/jm301519z.

128. A.K. Ghosh, C.X. Xu, K.V. Rao, A. Baldridge, J. Agniswamy, Y.F. Wang, I.T. Weber, M. Aoki, S.G. Miguel, M. Amano, H. Mitsuya, Probing multidrug-resistance and protein-ligand interactions with oxatricyclic designed ligands in HIV-1 protease inhibitors, ChemMedChem 5 (2010) 1850-1854. https://doi.org/10.1002/cmdc.201000318.

129. M. Amano, Y. Tojo, P.M. Salcedo-Gomez, J.R. Campbell, D. Das, M. Aoki, C.X. Xu, K.V. Rao, A.K. Ghosh, H. Mitsuya, GRL-0519, a novel oxatricyclic ligand-containing nonpeptidic HIV-1 protease inhibitor (PI), potently suppresses replication of a wide spectrum of multi-PI-resistant HIV-1 variants in vitro, Antimicrobial agents and chemotherapy 57 (2013) 2036-2046. https://doi.org/10.1128/AAC.02189-12.

130. A.K. Ghosh, C.D. Martyr, H.L. Osswald, V.R. Sheri, L.A. Kassekert, S. Chen, J. Agniswamy, Y.F. Wang, H. Hayashi, M. Aoki, I.T. Weber, H. Mitsuya, Design of HIV-1 Protease Inhibitors with Amino-bis-tetrahydrofuran Derivatives as P2-Ligands to Enhance Backbone-Binding Interactions: Synthesis, Biological Evaluation, and Protein-Ligand X-ray Studies, Journal of medicinal chemistry 58 (2015) 6994-7006. https://doi.org/10.1021/acs.jmedchem.5b00900.

131. M. Amano, Y. Tojo, P.M. Salcedo-Gomez, G.L. Parham, P.R. Nyalapatla, D. Das, A.K. Ghosh, H. Mitsuya, A novel tricyclic ligand-containing nonpeptidic HIV-1 protease inhibitor, GRL-0739, effectively inhibits the replication of multidrug-resistant HIV-1 variants and has a desirable central nervous system penetration

property in vitro, Antimicrobial agents and chemotherapy 59 (2015) 2625-2635. https://doi.org/10.1128/AAC.04757-14.

132. A.K. Ghosh, G.L. Parham, C.D. Martyr, P.R. Nyalapatla, H.L. Osswald, J. Agniswamy, Y.F. Wang, M. Amano, I.T. Weber, H. Mitsuya, Highly potent HIV-1 protease inhibitors with novel tricyclic P2 ligands: design, synthesis, and protein-ligand X-ray studies, Journal of medicinal chemistry 56 (2013) 6792-6802. https://doi.org/10.1021/jm400768f.

133. A.K. Ghosh, B.D. Chapsal, M. Steffey, J. Agniswamy, Y.F. Wang, M. Amano, I.T. Weber, H. Mitsuya, Substituent effects on P2-cyclopentyltetrahydrofuranyl urethanes: design, synthesis, and X-ray studies of potent HIV-1 protease inhibitors, Bioorganic & medicinal chemistry letters 22 (2012) 2308-2311. https://doi.org/10.1016/j.bmcl.2012.01.061.

134. E.M. Wondrak, J.M. Louis, Influence of flanking sequences on the dimer stability of human immunodeficiency virus type 1 protease, Biochemistry 35 (1996) 12957-12962. https://doi.org/10.1021/bi960984y.

135. J.M. Sayer, J. Agniswamy, I.T. Weber, J.M. Louis, Autocatalytic maturation, physical/chemical properties, and crystal structure of group N HIV-1 protease: relevance to drug resistance, Protein Sci 19 (2010) 2055-2072. https://doi.org/10.1002/pro.486.

136. J.M. Louis, R. Ishima, A. Aniana, J.M. Sayer, Revealing the dimer dissociation and existence of a folded monomer of the mature HIV-2 protease, Protein Sci 18 (2009) 2442-2453. https://doi.org/10.1002/pro.261.

137. Z. Otwinowski, W. Minor, Processing of X-ray diffraction data collected in oscillation mode, Methods Enzymol 276 (1997) 307-326.

138. C.H. Shen, Y.F. Wang, A.Y. Kovalevsky, R.W. Harrison, I.T. Weber, Amprenavir complexes with HIV-1 protease and its drug-resistant mutants altering hydrophobic clusters, FEBS J 277 (2010) 3699-3714. https://doi.org/10.1111/j.1742-4658.2010.07771.x.

139. A.J. McCoy, R.W. Grosse-Kunstleve, P.D. Adams, M.D. Winn, L.C. Storoni, R.J. Read, Phaser crystallographic software, J Appl Crystallogr 40 (2007) 658-674. https://doi.org/10.1107/S0021889807021206.

140. M.D. Winn, C.C. Ballard, K.D. Cowtan, E.J. Dodson, P. Emsley, P.R. Evans, R.M. Keegan, E.B. Krissinel, A.G. Leslie, A. McCoy, S.J. McNicholas, G.N. Murshudov, N.S. Pannu, E.A. Potterton, H.R. Powell, R.J. Read, A. Vagin, K.S. Wilson, Overview of the CCP4 suite and current developments, Acta Crystallogr D Biol Crystallogr 67 (2011) 235-242. https://doi.org/10.1107/S0907444910045749.

141. G.M. Sheldrick, Crystal structure refinement with SHELXL, Acta Crystallogr C Struct Chem 71 (2015) 3-8. https://doi.org/10.1107/S2053229614024218.

142. G.N. Murshudov, A.A. Vagin, E.J. Dodson, Refinement of macromolecular structures by the maximum-likelihood method, Acta Crystallogr D Biol Crystallogr 53 (1997) 240-255. https://doi.org/10.1107/S0907444996012255.

143. P. Emsley, B. Lohkamp, W.G. Scott, K. Cowtan, Features and development of Coot, Acta Crystallogr D Biol Crystallogr 66 (2010) 486-501. https://doi.org/10.1107/S0907444910007493.

144. W.L. DeLano, Pymol: An open-source molecular graphics tool, CCP4 Newsletter On Protein Crystallography (2002) 82-92. https://www.ccp4.ac.uk/newsletters/newsletter40/11_pymol.pdf.

145. H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The Protein Data Bank, Nucleic Acids Res 28 (2000) 235-242. https://doi.org/10.1107/S0907444902003451.

146. A. Wong-Sam, Y.F. Wang, Y. Zhang, A.K. Ghosh, R.W. Harrison, I.T. Weber, Drug Resistance Mutation L76V Alters Nonpolar Interactions at the Flap-Core Interface of HIV-1 Protease, ACS Omega 3 (2018) 12132-12140. https://doi.org/10.1021/acsomega.8b01683.

147. R.A. Copeland, D. Lombardo, J. Giannaras, C.P. Decicco, Estimating K-I Values for Tight-Binding Inhibitors from Dose-Response Plots, Bioorganic & medicinal chemistry letters 5 (1995) 1947-1952. https://doi.org/10.1016/0960-894x(95)00330-V.

148. F. Liu, A.Y. Kovalevsky, Y. Tie, A.K. Ghosh, R.W. Harrison, I.T. Weber, Effect of flap mutations on structure of HIV-1 protease and inhibition by saquinavir and darunavir, J Mol Biol 381 (2008) 102-115. https://doi.org/10.1016/j.jmb.2008.05.062.

149. J. Agniswamy, C.H. Shen, A. Aniana, J.M. Sayer, J.M. Louis, I.T. Weber, HIV-1 protease with 20 mutations exhibits extreme resistance to clinical inhibitors through coordinated structural rearrangements, Biochemistry 51 (2012) 2819-2828. https://doi.org/10.1021/bi2018317.

150. A.Y. Kovalevsky, J.M. Louis, A. Aniana, A.K. Ghosh, I.T. Weber, Structural evidence for effectiveness of darunavir and two related antiviral inhibitors against HIV-2 protease, J Mol Biol 384 (2008) 178-192. https://doi.org/10.1016/j.jmb.2008.09.031.

151. B. Mahalingam, J.M. Louis, C.C. Reed, J.M. Adomat, J. Krouse, Y.F. Wang, R.W. Harrison, I.T. Weber, Structural and kinetic analysis of drug resistant mutants of HIV-1 protease, Eur J Biochem 263 (1999) 238-245. https://doi.org/10.1046/j.1432-1327.1999.00514.x.

152. J. Chen, Drug resistance mechanisms of three mutations V32I, I47V and V82I in HIV-1 protease toward inhibitors probed by molecular dynamics simulations and binding free energy predictions, RSC Advances (2016) 58573-58585. https://doi.org/10.1039/C6RA09201B.

153. Foster, T. Medical microbiology. University of Texas Medical Branch at Galveston 1996.

154. Cimolai N. MRSA and the environment: implications for comprehensive control measures. Eur J Clin Microbiol Infect Dis 2008;27:481–93.

155. Pinho MG, de Lencastre H, Tomasz A. An acquired and a native penicillin-binding protein cooperate in building the cell wall of drug-resistant *Staphylococci*. Proc Natl Acad Sci 2001;98:10886–91.

156. Enright MC, Robinson DA, Randle G, Feil EJ. The evolutionary history of methicillin-resistant *Staphylococcus aureus*. Proc Natl Acad Sci 2002;99:7687–92.

157. Ghuysen JM. Serine beta-lactamases and penicillin-binding proteins. Annu Rev Microbiol 1991;45:37–67.

158. Hartman BJ, Tomasz A. Low-affinity penicillin-binding protein associated with beta-lactam resistance in *Staphylococcus aureus*. J Bacteriol 1984;158:513–6.

159. De Jonge BL, Chang YS, Xu N, Gage D. Effect of exogenous glycine on peptidoglycan composition and resistance in a methicillin-resistant *Staphylococcus aureus* strain. Antimicrobial Agents and Chemotherapy 1996;1498–1503.

160. Berger-Bächi B, Tschierske M. Role of fem factors in methicillin resistance. Drug Resist Update 1998;1:325–35.

161. Antignac A, Sieradzki K, Tomasz A. Perturbation of cell wall synthesis suppresses autolysis in *Staphylococcus aureus*: evidence for coregulation of cell wall synthetic and hydrolytic enzymes. J Bacteriol 2007;189:7573–80.

162. Vollmer W, Joris B, Charlier P, Foster S. Bacterial peptidoglycan (murein) hydrolases. FEMS Microbiol Rev 2008;32:259–86.

163. Grossowicz R, Razin NS. Factors influencing the antibacterial action of spermine and spermidine on *Staphylococcus aureus*. J Gen Microbiol 1955;13:436–41.

164. Yao X, Lu CD. Characterization of *Staphylococcus aureus* responses to spermine stress. Curr Microbiol 2014;69:394–403.

165. Yao X, Lu CD. A pbp 2 mutant devoid of the transpeptidase domain abolishes spermine-beta-lactam synergy in *Staphylococcus aureus* mu50. Antimicrob Agents Chemother 2012;56:83–91.

166. Pootoolal J, Neu WJ. Glycopeptide antibiotic resistance. Annu Rev Pharmacol Toxicol 2002;42:381–408.

167. Utaida S, Dunman PM, Macapagal D, Murphy E. Genome-wide transcriptional profiling of the response of *Staphylococcus aureus* to cell-wall-active antibiotics reveals a cell-wall-stress stimulon. Microbiology 2003;149:2719–32.

168. Dalman M, Anthony Deeter GN, Duan ZH. Fold change and p-value cutoffs significantly alter microarray interpretations. BMC Bioinformatics 2012;13:11.

169. Huang da W, Sherman BT. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 2009;37:1–13.

170. Boudewijn LM, de Jonge, Tomasz A. Abnormal peptidoglycan produced in a methicillin-resistant strain of *Staphylococcus aureus* grown in the presence of methicillin: functional role for penicillin-binding protein 2a in cell wall synthesis. Antimicrob Agents Chemother 1993;37:342–6.

171. Haag H, Fiedler HP, Meiwes J, Drechsel H. Isolation and biological characterization of staphyloferrin b, a compound with siderophore activity from *Staphylococci*. FEMS Microbiol Lett 1994;115:125–30.

172. Meiwes J, Fiedler HP, Haag H, Zähner H. Isolation and characterization of staphyloferrin a, a compound with siderophore activity from *Staphylococcus hyicus* dsm 20459. FEMS Microbiol Lett 1990;55:201–5.

173. Chakraborty R, Braun V, Hantke K, Cornelis P. Iron uptake in bacteria with emphasis on *E. coli* and *Pseudomonas*. Springer Briefs in Bio metals 2013;XI:89.

174. Rogers MB, Sexton J, DeCastro J, Calderwood S. Identification of an operon required for ferrichrome iron utilization in *Vibrio cholerae*. J Bacteriol 2000;182:2350–3.

175. Mademidis A. Transport activity of fhua, fhuc, fhud, and fhub derivatives in a system free of polar effects, and stoichiometry of components involved in ferrichrome uptake. Mol Gen Genet 1998;258:156–65.

176. Marzabadi MR. Spermine prevent iron accumulation and depress lipofuscin accumulation in cultured myocardial cells. Free Radic Biol Med 1996;21:375–81.

177. Lomozik L, Jastrzab R. Copper(ii) complexes with uridine, uridine 5'-monophosphate, spermidine, or spermine in aqueous solution. J Inorg Biochem 2003;93:132–40.

178. Bischoff M, Dunman P, Kormanec J, Macapagal D. Microarray-based analysis of the *Staphylococcus aureus* sigmab regulon. J Bacteriol 2004;186:4085–99.

179. Chan PF, Foster SJ, Ingham E, Clements MO. The *Staphylococcus aureus* alternative sigma factor sigmab controls the environmental stress response but not starvation survival or pathogenicity in a mouse abscess model. J Bacteriol 1998;180:6082–9.

180. Hecker M. General stress response of *Bacillus subtilis* and other bacteria. Adv Microb Physiol 2001;44:35–91.

181. Hecker M, Pane-Farre UV. Sigb-dependent general stress response in *Bacillus subtilis* and related gram-positive bacteria. Annu Rev Microbiol 2007;61:215–36.

182. Petersohn A, Brigulla M, Haas S, Hoheisel JD. Global analysis of the general stress response of *Bacillus subtilis*. J Bacteriol 2001;183:5617–31.

183. Pinho MG, Lencastre H. Transcriptional analysis of the *Staphylococcus aureus* penicillin binding protein 2 gene. J Bacteriol 1998;180:6077–81.

184. Kuwahara-Arai K, Kondo N, Hori S, Tateda-Suzuki E. Suppression of methicillin resistance in a meca-containing pre-methicillin-resistant *Staphylococcus aureus* strain is caused by the meci-mediated repression of pbp 2' production. Antimicrob Agents Chemother 1996;2680–5.

185. Shah P. A multifaceted role for polyamines in bacterial pathogens. Mol Microbiol 2008;68:4–16.

186. Epstein W. The roles and regulation of potassium in bacteria. Prog Nucleic Acid Res Mol Biol 2003;75:293–320.

187. Heermann R. The complexity of the 'simple' two-component system kdpd/kdpe in *Escherichia coli*. FEMS Microbiol Lett 2010;304:97–106.

188. Hyeon-Gyu S, Zhe L. Mechanism of the voltage sensitivity of irk1 inward-rectifier k+ channel block by the polyamine spermine. J Gen Physiol 2005;125:413–26.

189. Marquardt JL, Siegele DA, Kolter R, Walsh CT. Cloning and sequencing of *Escherichia coli* murz and purification of its product, a udp-n-acetylglucosamine enolpyruvyl transferase. J Bacteriol 1992;174:5748–52.

190. Qamar A. Dual roles of fmta in *Staphylococcus aureus* cell wall biosynthesis and autolysis. Antimicrob Agents Chemother 2012;56:3797–805.

191. Kohanski M, Dwyer D, Collins J. How antibiotics kill bacteria: from targets to networks. Nat Rev Microbiol 2010;8:423–35.

192. Kohanski MA, Dwyer DJ, Hayete B, Lawrence CA. A common mechanism of cellular death induced by bactericidal antibiotics. Cell 2007;130:797–810.

193. Pepper S, Saunders E, Edwards E, Wilson C. The utility of mas5 expression summary and detection call algorithms. BMC Bioinformatics 2007;273.

194. Gautier L, Cope L, Bolstad BM, Irizarry RA. Affy–analysis of affymetrix genechip data at the probe level. Bioinformatics 2004;20:307–15.

195. R Development Team. R: A language and environment for statistical computing. Vienna, Austria : the R Foundation for Statistical Computing 2011;ISBN: 3-900051-07-0.

196. Gregory R, Warnes G. gplots: Various R programming tools for plotting data. R package version 2009.

197. Yang H. Normalization for cdna microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res 2002;30:15.

198. Ritchie ME, Phipson B, Wu D, Hu Y. Limma powers differential expression analyses for rna-sequencing and microarray studies. Nucleic Acids Res 2015;43:47