AN IN-SILICO DRUG FORAGE AGAINST *TRMD* GENE OF *SALMONELLA TYPHIMURIUM* (LT2)

M.Sc. Thesis
2022

Submitted to
Central Department of Biotechnology
Tribhuvan University
Kirtipur, Kathmandu, Nepal

For Partial Fulfillment of M.Sc. Degree in Biotechnology

By
Bisheshta Nepal
Roll No. 502/074
Registration No: 5-2-0282-0163-2013

Supervisors
Senior Scientist Dr. Pramod Aryal
Prof. Dr. Rajani Malla

ACKNOWLEDGEMENT

First and foremost, I want to specifically thank and appreciate my mentor and supervisor, senior scientist, **Dr. Pramod Aryal**, for his unfailing support and insightful advice. Throughout the duration of my thesis period, his constant focus guided me in the proper direction whenever I encountered a challenge.

I also want to express my gratitude to **Prof. Dr. Rajani Malla** for letting us work under her **Streptomyces project** (**UGC institutional grant**) and her constant guidance and support.

I would also like to express my gratitude to **RECAST** and **Prof. Dr. Rameshwor Adhikari** for their assistance in managing the accommodations throughout the research period. I also appreciate **Nepal Army's** handling of the logistics.

I would like to show my heartfelt gratitude to **University Grant Commission (UGC)** for providing me grant to complete the thesis.

For allowing me to finish my thesis work at Tribhuvan University, **Prof. Dr. Krishna Das Manandhar**, HOD of the Central Department of Biotechnology, has my gratitude as well.

I would like to express my gratitude to my colleagues, Mr. Samiran Subedi, Mr. Siddhartha Gautam Ms. Kabita Kandel, Ms. Suja Maharjan, Ms. Guheshwori Chataut and Mr. Devraj Mainali for their thoughtful contributions and support during the completion of my thesis.

The assistance I received from my seniors Ms. Manju Pun, Ms. Sabina Thapa Magar, Ms. Pooja Pathak, and Ms. Rita Kumari Oli during a time of need is something I would also like to acknowledge. I would like to give special acknowledgement to my senior Ms. Sita Ghimire for providing me with stocks of *Streptomyces*.

Last but not least, I must express my sincere gratitude to my parents and my husband Mr. Darshan Bhattrai for their unwavering inspiration during the period of my thesis.

Bisheshta Nepal

Registration No: 5-2-0282-0163-2013

LIST OF ABBREVIATIONS

AA	Amino Acid
SAM	S- adenosyl methionine
ADMET	Absorption, Distribution, Metabolism, Excretion and Toxicity
TrmD	tRNA methyl transferase D
BE	Binding Energy
CDC	Centers for Disease Control and Prevention
cFID	Consensus Induced-Fit Docking
CLEVER	Chemical Library Editing, Visualizing and Enumerating Resource
COBRA	Constraint-based Reconstruction and Analysis
AMR	Antimicrobial resistece
CSDD	Center for the Study of Drug Development
CT	Computed Tomography
MDR	Multi Drug Resistant
DNA	Deoxyribonucleic Acid
DPP4	Dipeptidyl Peptidase 4
PMQR	plasmid-mediated quinolone resistance
ED	Ensemble Docking
FDA	Food and Drug Administration
Gbk	Gene Bank
HTS	High Throughput Sequencing
LBVS	Ligand-Based Virtual Screening
kD	Kilodaltons
MD	Molecular Dynamics
MERS	Middle East Respiratory Symptom
MM	Molecular Mechanics
NCBI	National Center for Biotechnology Information
nm	Nanometer

NMR	Nuclear magnetic resonance
nsp	Non-structural proteins
ORF	Open reading frame
PAINS	Pan Assay Interference Compounds
PDB	Protein Data Bank
PSA	Polar Surface Area
PTFE	Polytetrafluoroethylene
PVC	Polyvinyl chloride
QM	Quantum mechanics
RBD	Receptor Binding Domain
RNA	Ribonucleic acid
RT-PCR	Real time Polymerase chain reaction
SAM	S-Adenosyl Methionine
SARS	Severe acute respiratory syndrome
SBVS	Structure Based Virtual Screening
Tox	Toxicity
TPP	Thiamine pyrophosphate
tPSA	Topological Polar Surface Area
UTR	Untranslated region
VS	Virtual Screening
WHO	World Health Organization
IM	Inner Membrane
OM	Outer Membrane

LIST OF TABLES

Table 4.1: Gram stain and biochemical tests for identification of bacterial strain 42

Table 4.2 p-BLAST results for *trmD* of *Salmonella Typhimurium* in reference to *trmD Escherichia coli* K-12 and *trmD* Haemophilus influenzae RD-KW20 46

Table 4.3 Z score using various homology modelling tools 51

Table 4.4 Table showing the Ramachandran Analysis of various *trmD* proteins that were created using homology modelling tools 53

Table 4.5 Table showing protein validation using various scores for the models of *trmD* created using MODELLER 54

Table 4.6: Predicted binding sites and heterogens present as predicted by 3DligandSite in *TrmD* of *Salmonella Typhimurium* predicted via Phyre2 webserver 59

Table: 4.7 Table showing the number of ligands (Natural Products) that were reduced after ADMET screening using OSIRIS 63

Table 4.8: Number of various categories of ligands obtained after docking with *trmD* Salmonella (LT2) 64

Table 4.9: Ligands and their interaction with *trmD Salmonella Typhimurium* and human MAT 1A protein 65

Table 4.10: Nature of chemical interaction between SAM and *trmD Salmonella Typhimurium* 67

Table 4.11: Nature of chemical interaction between NP and *trmD Salmonella Typhimurium* 69

Table 4.12: Nature of chemical interaction between KI 1 and *trmD Salmonella Typhimurium* 69

Table 4.13: Nature of chemical interaction between KI 2 and *trmD Salmonella Typhimurium* 71

Table 4.14: Nature of chemical interaction between KI 3 and *trmD Salmonella Typhimurium* 72

Table 4.15: Nature of chemical interaction between KI 4 and *trmD Salmonella Typhimurium* 73

Table 4.16: Nature of chemical interaction between KI 5 and *trmD Salmonella Typhimurium* 74

Table 4.17: Nature of chemical interaction between NM 1 and *trmD Salmonella Typhimurium* 75

Table 4.18: Nature of chemical interaction between NM 2 and *trmD Salmonella Typhimurium* 76

Table 4.19: Nature of chemical interaction between ID 1 and *trmD Salmonella Typhimurium* 77

Table 4.20: Nature of chemical interaction between ID 2 and *trmD Salmonella Typhimurium* 78

Table 4.21: Nature of chemical interaction between ID 3 and *trmD Salmonella Typhimurium* 79

LIST OF FIGURES

Fig 2.1: The characteristics of transfer RNA (tRNA) methyltransferase 5 (TRM5) and tRNA methyltransferase D(*TrmD*)..... 10

Fig 2.2: *TrmD* catalyzes methylation to the N^1 position of G37 in tRNA, using AdoMet as the methyl donor and producing m^1G37 on the 3'-side of the tRNA anticodon (Hou *et. al.,* 2017)..... 11

Fig 2.3: *trmD* of *Escherichia coli* (1P9P) (RCSB.pdb) 12

Figure 2.4: Mechanism of methyl-transfer reaction..... 13

Fig 2.5 Diagramatic illustration of suppression of ribosomal +1-frameshifting by m1G37-tRNA (Hou *et. al.,* 2017) 16

Fig 2.6: Steps in comparative protein structure modeling (Sali and Web, 2020)..... 17

Fig 2.7: CADD in drug design pipeline(SIwoski *et. al.,* 2014) 19

Fig 2.8: Schematic diagram of molecular docking protocols (Brás *et. al.,*. 2014) 26

Fig 4.1 Isolation and revival of Salmonella on XLD Agar (left) and ATCC *Escherichia coli* on LB Agar (right)..... 42

Fig 4.2 Antibiotic Susceptibility Testing of Salmonella (LT2);1- Ertapenem,2-meropenem, 3-Colistin, 4- Gentamicin, 5- Ciprofloxacin..... 43

Fig 4.3 Subculture and revival of Ma and Pc strain of *Streptomyces* in SCA medium..... 44

Fig 4.4 : Resazurin Assay of *Streptomyces* against Salmonella (LT2)..... 45

Fig 4.5 : Multiple sequence allignment of *TrmD* Protein of WHO Prioritised pathogen . 47

Fig 4.6: Phylogenetic tree of *trmD* protein of WHO prioritized pathogens..... 48

Fig 4.7: Structure of *trmD Salmonella Typhimurium* LT2(model *trmD*.B905) using modellar 9.2 taken from PyMol..... 49

Fig 4.8: The Z score plots and energy plot for predicting 3D structure of a protein based on amino acids sequences..... 51

Fig 4.9: Ramachandran plot analysis using SAVES V 5.0 for amino acid residues of *trmD*.B905 (developed using MODELLER)..... 52

Figure 4.10: Structural superposition between MODELLER (*trmD*. B905) (tan) and that of *Escherichia coli* (K12) (PDB:1P9P) (cyan) using UCSF chimera. 55

Fig 4.11: Amino acid alignment of *trmD* of *Salmonella Typhimurium* (LT2) and Escherichia 60

Fig 4.12: Interaction of SAM with active sites of *trmD* Salmonella visualized using PyMol 68

Fig 4.13: Interaction of NP with active sites of *trmD* Salmonella visualized using PyMol (left) and bond length (right) visualized using BIOVA- Discovery Studio..... 69

Fig 4.14: Interaction of KI 1 with active sites of *trmD* Salmonella visualized using PyMol (left) and bond length (right) visualized using BIOVA- Discovery Studio..... 71

Fig 4.15: Interaction of KI2 with active sites of *trmD* Salmonella visualized using PyMol (left) and bond length (right) visualized using BIOVA- Discovery Studio..... 72

Fig 4.16: Interaction of KI 3 with active sites of *trmD* Salmonella visualized using PyMol (left) and bond length (right) visualized using BIOVA- Discovery Studio..... 73

Fig 4.17: Interaction of KI 4 with active sites of *trmD* Salmonella visualized using PyMol (left) and bond length (right) visualized using BIOVA- Discovery Studio..... 74

Fig 4.18: Interaction of KI 5 with active sites of *trmD* Salmonella visualized using PyMol (left) and bond length (right) visualized using BIOVA- Discovery Studio..... 75

Fig 4.19: Interaction of NM 1 with active sites of *trmD* Salmonella visualized using PyMol (left) and bond length (right) visualized using BIOVA- Discovery Studio..... 76

Fig 4.20: Interaction of NM 2 with active sites of *trmD* Salmonella visualized using PyMol (left) and bond length (right) visualized using BIOVA- Discovery Studio..... 77

Fig 4.21: Interaction of ID 1 with active sites of *trmD* Salmonella visualized using PyMol (left) and bond length (right) visualized using BIOVA- Discovery Studio..... 78

Fig 4.22: Interaction of ID 3 with active sites of *trmD* Salmonella visualized using PyMol (left) and bond length (right) visualized using BIOVA- Discovery Studio..... 79

Fig 4.23: Interaction of ID 4 with active sites of *trmD* Salmonella visualized using PyMol (left) and bond length (right) visualized using BIOVA- Discovery Studio..... 80

Fig 4.24: Bond length interaction between the atoms of Antineoplaston..... 82

Fig 4.25: Molecular Orbital Properties (a) HOMO and (b) LUMO..... 83

Fig 4.26: Electon density in Antineoplaston visualized in Gaussview..... 83

TABLE OF CONTENTS

ACKNOWLEDGEMENT ii
LIST OF ABBREVIATIONS iii
LIST OF TABLES v
LIST OF FIGURES vii
TABLE OF CONTENTS x
ABSTRACT xiv
1: INTRODUCTION 1
1.1 Background 1
1.2 Current Studies 2
1.3 Hypothesis 3
1.3.1 Null hypothesis: 3
1.3.2 Alternative hypothesis: 3
1.4 Objectives 3
1.4.1 General Objective 3
1.4.2 Specific Objectives 3
1.5 Rationale 4
1.6 Scope of study 4
2: LITERATURE REVIEW 5
2.1 Review of Literature in Antibiotics Resistance and the underlying mechanism 5
2.1.1 Antimicrobial resistance- a global concern 5
2.1.2 *Salmonella Typhimurium* and Antimicrobial Resistance 6
2.1.3 Antibiotic resistance due to tRNA methylation in Gram Negative Bacteria 6
2.2 *Streptomyces* as an Antibiotic Producer 8
2.3 Literature Review on *trmD* as a Probable Drug Target 9
2.3.1 Structure of *TrmD* 10
2.3.2 tRNA methylation and its role 12
2.3.3 Mechanism of synthesis of m1G37-tRNA by *TrmD*: 12
2.3.4 Mechanism of frameshifting in absence of *TrmD*: 13
2.4 Review of literature on homology modelling 16
2.4.1 MODELLER as a homology modelling tool 18

2.5 Review of Literature related to in-vitro Drug Design 18
2.5.1. Computer Aided Drug Design 18
2.6 Review of literature related to Mutation 29
2.7 Review of literature related to Density Function Theory 31
2.7.1 Density function theory 31
2.8 Review of literature related to Molecular dynamics (MD) simulations 31
2.8.1 Molecular dynamics simulations 31
2.9 Review of literature related to human Cytochrome P450 enzymes 32
3: MATERIALS AND METHODOLOGY 33
3.1 Materials Required for Confirmation and Antimicrobial Sensitivity Test (AST) of test organisms 33
3.1.1 Collection and selection of the test organisms 33
3.1.2 Identification and reconfirmation of the test organisms 33
3.1.3 Revival of *Streptomyces* 35
3.4.1 Secondary metabolite production 35
3.4.2 Resazurin Antimicrobial Assay 36
3.2 Search for bacterial protein target 36
3.2.1 Selection of bacterial protein and obtaining their genomic sequences 36
3.2.2 Alignment search using BLAST tool for *trmD* gene of *Salmonella Typhimurium* 36
3.2.3 Multiple Sequence Alignment search CLUSTAL OMEGA for *trmD* gene of WHO prioritized Pathogens 37
3.3 Homology Modeling of target protein and its validation 37
3.4 Molecular docking simulation 37
3.3.1 Obtaining the dockable crystal structures of the target protein 37
3.3.2 Preparation of ligand database 37
3.3.3 Protein and ligand preparation 38
3.3.4 Setting reference values for docking 38
3.3.5 Structure based Virtual Screening 38
3.3.6 Preparation of ligand database from top hits obtained from docking studies for final docking 39
3.3.7 ADME/TOX filter 39
3.3.8 Final Structure based Virtual screening 39
3.4 Analysis of docking results 40
3.4.1 Binding interaction of ligands with protein target 40

3.4.2 Bond length and bond types responsible for drug ligand interaction 40
3.4.3 Hydrophobic bond interactions 40
3.6 DFT calculations 40
4: RESULTS AND DISCUSSION 41
4.1 Pathogen 41
4.1.1 Gram Staining and Biochemical Test of Pathogen 42
4.1.2 Antibiotic Susceptibility Testing of Pathogen 43
4.2 *Streptomyces* 43
4.2.1 Secondary metabolite Production 44
4.2.2 Resazurin test Assay 44
In-silico 45
4.3 Search for Bacterial Protein Targets 45
4.4 Drug Target Protein Selection 46
4.5 Protein Preparation, Validation and Processing 46
4.5.1 Sequence Retrieval and Alignment 46
4.5.2 Allignment of *trm*D gene of WHO prioritized Pathogens 47
4.5.3 Protein Structure Preparation using various Homology Modeling Tools 48
4.5.4 Protein processing and validation 49
4.6 Molecular docking 55
4.6.1 Protein Preparation 55
4.6.2 Ligand database preparation 57
4.6.3 Identification of Active Binding Site 59
4.6.4 In-silico ADME/Tox tests for possible hits 61
4.6.5 Virtual Screening 63
4.7 Cross reactivity with human proteins: MAT1A as a reference 64
4.8 Analysis of docking 67
4.8.1 Protein ligand interaction 67
4.9 Prospects of Drugs against *trm*D 80
4.10 Mutational Analysis 80
4.11 DFT Analysis 81
4.11.1 Dipole moment and total energy of the molecule 81
4.11.2 Molecular orbital properties 82
4.12 *trm*D as a safe bet with human trm5 protein 83

5. SUMMARY 85
6. CONCLUSION 87
7. RECOMMENDATIONS 87
8. BIBLIOGRAPHY 88
9.APPENDICES 106
9.1 Media composition of ISP2 and ISP4 106
9.2 Data sources for metabolic model reconstruction and refinement 107
9.2.1 DNA sequence and genome annotation databases 107
9.2.2 Protein and enzyme databases 107
9.2.3 Metabolic databases 108
9.2.4 Experimental data repositories 108
9.2.5 Metabolic Model Repositories 109
9.3 Results from OSIRIS property explorer for those passing ADME/Tox filters 109
9.3.1 Natural Products 110
9.3.2 Kinase Inhibitors 111
9.3.3 Nucleoside mimetics 112
9.3.4 Indole Derivatives 113
9.4 Protein Information 114

ABSTRACT

Salmonella, that is multidrug resistant (MDR) poses a serious threat to people. An essential enzyme needed for cell growth and survival, TrmD is an S-adenosyl methionine (AdoMet or SAM)-dependent methyl transferase that creates the methylation m^1G37 in $tRNA^{pro.}$ It is distinct from its eukaryotic and archeal cousin Trm5. In-vitro tests of the effectiveness of Actinomycetes against *Salmonella Typhimurium*, homology modeling of the target protein, molecular docking, predictions of the ADME/T properties of particular antibacterial molecules, research into the druglikeness of particular ligands, and computational quantum mechanical modeling of particular hits are all objectives of this study. In order to comprehend the inhibitory mechanism and the spatial orientation of the ligands, as well as to identify important amino acid residues within the substrate-binding pocket that can be used for structure-based drug design, the study offers an insight into biomolecular interactions. The target protein (modeled using MODELLER) of *Salmonella Typhimurium* (LT2) was docked with a variety of natural products, kinase inhibitors, nucleoside mimics, and indole derivatives. After the compounds were narrowed, the compounds that were able to be effective against target protein and bind with the amino acid L138 in the active site were found to be one Natural Products; Antineoplaston, five kinase inhibitors 2-(4-Fluorophenyl)-N-[2-(1H-indol-3-yl) ethyl] acetamide; N-benzyl-2-(2-methyl-1H-indol-3-yl) acetamide; 1-(1,3-Dimethylpyrazol-4-yl)-3-[(2R)-1-hydroxy-3-phenylpropan-2-yl]urea; ; 1-[(2S)-1-Hydroxy-3-phenylpropan-2-yl]-3-(1-pyridin-4-ylethyl)urea; 1-(1,5-Dimethylpyrazol-3-yl)-3-[(2S)-1-hydroxy-3-phenylpropan-2-yl]urea three indole derivative (N-[(3,4-dimethoxyphenyl)methyl]-3-(1H-indol-6-yl)propenamide; N-[2-(6-methoxy-1H-indol-3-yl)ethyl]-3-phenylpropanamide and 2-(4-fluorophenyl)-N-[2-(5-methoxy-1H-indol-3-yl)ethyl]acetamide. Similarly, two nucleoside mimetics 7-[(3R)-1-Benzoylpyrrolidin-3-yl]-N-methylthieno[2,3-b] pyrazine-6-carboxamide and [3-(2-Methylpropyl)-1,2-oxazol-5-yl]-[(3S)-3-(1H-pyrazolo[3,4-b] pyridin-6-yl) piperidin-1-yl] methanone were found to be effective. Natural Product Antineoplaston's DFT calculation revealed that the substance was a stable and reactive molecule with drug-like properties.

Keywords: ADMET, trmD, homology modelling, drug discovery, molecular docking, therapeutic alternative.

1: INTRODUCTION

1.1 Background

The emergence of antibiotic resistant strains among the pathogens appear to be inevitable due to selective pressure for survival (Tello *et. al.,* 2012). This is why discovery of novel families of antibiotics at a regular interval is a must if modern medicine is to continue in its present form. In an ecologically diverse country like Nepal - 27th in the whole world, there are many yet to be explored places which can be excellent habitats for *Streptomyces* strains that could have the ability to produce novel Antibiotics, eventually helping in coping the present diseases and antibiotic resistance problems. (Butler, 2016)

Salmonella enterica is a leading cause of community-acquired bloodstream infections in many low- and middle-income countries (Reddy *et. al.,* 2010; Deen *et. al.,* 2012). *Salmonella enterica* serovars Typhi, Paratyphi A, Paratyphi B, and Paratyphi C may be referred to collectively as typhoidal *Salmonella*, whereas other serovars are grouped as nontyphoidal *Salmonella* (NTS) (Feasey *et. al.,* 2012).

TrmD is a global determinant of membrane biosynthesis in *E. coli* and *Salmonella enterica* serovar Typhimurium, two major Gram-negative pathogens. m^{1}G37 deficiency caused by *TrmD* depletion disrupts the outer membrane structure and rigidity, sensitizes *E. coli* and *Salmonella* to various classes of antibiotics, and suppresses their development of resistance or persistence upon antibiotic exposure (Masuda *et. al.,* 2020).

In this time of need, repurposing the existing drug library used in the humans can be an avenue forward when the production of antibiotics with higher efficiency may take loger period of time for its development (Wouters *et. al.,* 2020). Drug repurposing also called as, drug repositioning, has high probability of success rates because it explores the existing therapies for new therapeutic purposes and therefore, information related to drug's synthetic methodology, safety and toxicology is available beforehand. Thus, drug repurposing potentially reduces the cost and shortens the time for drug discovery in

comparison to the conventional *de novo* drug discovery. In addition, use of computational biology approach could help in identifying the crucial lead target proteins against which library of thousands of ligands and other small drug like molecules can be virtually screened to narrow down starting molecules which could be then incrementally modified for development of probable drug candidates (March-Vila *et. al.,* 2017).

1.2 Current Studies

In one research, involving 158 pets affected with Typhoid, 36% of the *Salmonella* strains were shown to be resistant to at least one antibiotic, while 13% were found to be resistant to four or more antimicrobials (White DG *et. al.,* 2003). 24 serotypes of *Salmonella*, including *S. Anatum, S. Typhimurium*, and *S. Infantis*, were discovered in another investigation, 36% of them were resistant to one or more antibiotics, while 13% were resistant to four or more. S. Newport strains that produce the CMY-2 AmpC -lactamase were responsible for a human infection in Canada in 2002 (Pitout DD *et. al.,* 2003). Human infections brought on by Salmonella with an enlarged range of cephalosporin resistance have sharply increased in many nations (Su LH *et. al.,* 2004).

In Japan, human samples have been found to contain extended-spectrum -lactamase (ESBL) or AmpC-carrying extended-spectrum cephalosporin-resistant Salmonella (Morita M *et. al.,* 2010)

The most popular drug discovery method, known as "hit to lead identification," entails taking small molecule hits from a high throughput screen (HTS) and turning them into potential lead compounds (Vemuri and Makriyannis, 2015). Modern drug development, which entails the design and chemical synthesis of pharmaceutical molecules, has also benefited from medicinal chemistry.

With a research expense of around 33 billion dollars, only 35 novel substances were registered with the Food and Drug Administration (FDA) in 2003 (Spedding, 2006). A reliable and quicker technique of drug discovery is essential due to the pathogens' rapid rise in treatment resistance and the excessive time and cost requirements for medication development. When combined with the existing in vitro approaches, computational tools

can effectively aid drug discovery and development (Sliwoski *et. al.,* 2014). Due to the fast and steady need for new antibiotics, library compounds found in various databases could be screened using computational methods. To create a medication that operates against any pathogen, identification of the lead protein which is crucial for the pathogen's survival that can be used as a target is necessary for such screenings. Additionally, the molecule should be able to function on a variety of targets or the essential genes to prevent the infection from developing resistance to it quickly. The major focus of this research is the application of various computational tools to speed up the drug development process and perhaps avoid the failure of the created lead compounds in later stages of drug development, such as various pre-clinical trial phases.

1.3 Hypothesis

1.3.1 Null hypothesis:

Potential lead molecules will not be developed against the *trmD* gene of *Salmonella Typhimurium* (LT2).

1.3.2 Alternative hypothesis:

Potential lead molecules will be developed against the *trmD* gene of *Salmonella Typhimurium* (LT2).

1.4 Objectives

1.4.1 General Objective

- Screening of available strains of *Streptomyces* that can produce potent antimicrobial agent against colistin resistant *Salmonella Typhimurium* (LT2) and to identify novel lead molecules against the *trmD* gene of *Salmonella Typhimurium* (LT2) that can act as a drug.

1.4.2 Specific Objectives

- To identify probable drug targets.
- To perform homology modelling to develop target protein.
- To perform molecular docking against the drug targets.
- To look for lead molecules that could potentially be developed as drugs against *trmD* gene of *Salmonella Typhimurium* (LT2) with minimal toxicity to humans.
- To study the thermodynamic and molecular orbital properties of lead compounds.

1.5 Rationale

Since the middle of the 19th century, there has been a sharp decline in the creation of new antibiotics. Development of newer antibiotics is of concern due to the developing resistance in pathogens. Commercial pharmaceutical companies' interest in such investments has decreased due to high research costs and numerous failures at various stages of medication development. The goal of the current research is to provide a reliable approach for target identification, lead identification through virtual screening, and the development of a plausible mechanism of inhibition study method for prospective lead molecules against a therapeutic target.

1.6 Scope of study

The goal of the current investigation is to find possible lead compounds that will work against the *trmD* gene of *Salmonella Typhimurium* (LT2). Also, sought after are reliable screening and antibacterial testing.

2: LITERATURE REVIEW

2.1 Review of Literature in Antibiotics Resistance and the underlying mechanism

2.1.1 Antimicrobial resistance- a global concern

Antimicrobial resistance (AMR) is the inability of microorganisms to be treated with drugs that are intended to kill bacteria, fungi, viruses, and parasites. When antimicrobial agents are used against resistant organisms like bacteria, fungi, viruses, and some parasites, the pathogens are not eradicated and instead continue to spread (Jamal *et. al.*, 2015).

Antimicrobial resistance is the third-most significant public health hazard of the twenty-first century, according to the WHO. Approximately 23,000 Americans and 25,000 Europeans every year pass away from diseases brought on by resistant germs. According to estimates, a child in Asia dies from an infection brought on by resistant bacteria every five minutes (Howard *et. al.,*2014). Antimicrobial resistance has so come to be known as one of the most important public health problems the world currently faces. Our ability to treat infectious diseases is compromised by the widespread, fast emergence of resistant microorganisms, which eventually leads to extended sickness and death (Prestinaci *et. al.,* 2015).

The urgency of the current worldwide crisis is increased by antimicrobial resistance to last-line therapies like carbapenems and colistin as well as a dearth of novel medicines. Antibiotic resistance is predicted to result in around 300 million premature deaths by 2050, costing the global economy $100 trillion. Lack of a strong pipeline for the development of antibiotics has made the issue worse, leading to the creation of illnesses that are nearly incurable and depriving clinicians of a supply of powerful medicines to treat sick patients (Munita & Arias, 2016).

2.1.2 *Salmonella Typhimurium* and Antimicrobial Resistance

The most prevalent serotype to have reported in the contamination of poultry related food is *Salmonella Typhimurium* (Foley *et. al.,* 2008) which is one of the serovar (one of more than 2,600 serotypes) of the genus *Salmonella* of the family Enterobacteriaceae (Wajid *et. al.,* 2019). The threat of multiple drug-resistant (MDR) *Salmonella* to humans is alarming since the haphazard use of antibiotics in poultry leads a direct access to human via food of animal origin (Threlfall E. J., 2002).

An irrational use of antimicrobials has aided the *Salmonella* strains that are resistant to quinolones to circumvent susceptibility to fluoroquinolones as well (Hein *et. al.,* 2006) further leading to treatment failure at times (Wajid *et. al.,* 2019). The major concern is on the production of efflux pumps via selective gene activation mechanism like qnr genes (encoding a group of pentapeptide proteins which bind to DNA gyrase and circumventing the action of quinolones), qepA gene (an quinolone efflux pump,) aac(6')-Ib-cr gene (aminoglycoside acetyltranferase that renders ciprofloxacin ineffective) and oqxAB genes (multidrug resistance efflux pump) (Jacoby *et. al.,* 2014) and extended spectrum beta lactamase (ESBL) production among the drug resistance mechanisms (Hoffmann *et.al.,* 2014; McDermott *et. al.,* 2016). Various resistance of *Salmonella Typhimurium* has been seen against colistin (mcr-9 based phenotypic resistance) (Carroll *et. al.,* 2019), against quinolone (acquisition of plasmid-mediated quinolone resistance (PMQR) genes) (Hopkins *et. al.,* 2008; Casas *et. al.,* 2016; Cummings *et.al.,* 2017).

2.1.3 Antibiotic resistance due to tRNA methylation in Gram Negative Bacteria

Gram-negative bacteria's multi-drug resistance is a serious and growing medical issue. Antibiotics frequently encounter entry and expulsion barriers, which prevent them from reaching high enough intracellular concentrations to have a therapeutic impact. The double-membrane structure of Gram-negative bacteria's cell envelope, which serves as both a permeability barrier and a platform for efflux machines that export drugs, is largely to blame for this issue (Payne *et. al.,* 2007; Silver, 2011).

Resistance mutations quickly appeared in earlier attempts that targeted one membrane protein or one efflux pump at a time (Murakami *et. al.,* 2006). Upon the challenge of the antibiotic during therapy, such mutations are selected, resulting in a population of resistant individuals (Silver, 2011, 2012).

A more potent approach to improving antibiotic efficacy would be to inhibit a mechanism that simultaneously regulates the expression of numerous membrane-associated genes. Such a worldwide mechanism, which has not yet been discovered, could offer a novel anti-bacterial tactic to allow several medications to work, reduce the likelihood of resistance, and hasten bactericidal action. A cell wall, an outer membrane, and a plasma inner membrane (IM) make up the Gram-negative bacteria's cell envelope (OM). The cell wall is a hard, cross-linked matrix of peptidoglycan that gives the cell its mechanical strength, whereas the IM is a fluid lipid bilayer (Holtje, 1998). The OM is composed of lipopolysaccharides in the outer leaflet and phospholipids in the inner leaflet, forming an asymmetric bilayer that prevents substances from diffusing into the cytosol or periplasm while also expelling substances to the external media via membrane-bound efflux transporters (Nikaido, 1998).

In addition to serving as a barrier, *Escherichia coli*'s OM gives the cell a mechanical stiffness comparable to that of the cell wall (Rojas *et. al.,* 2018), demonstrating the significance of robust OM biogenesis for cellular mechanical integrity. Both IM and OM's biogenesis need substantial integration with protein components, which also control the creation of cell walls (Typas *et. al.,* 2011)

Thus, the quality of the entire Gram-negative cell envelope is determined by the generation of membrane proteins; this quality is crucial for creating a permeability barrier and efflux activity against medications as well as for defining cell shape and stability throughout cell growth. (Masuda et.al., 2019)

Codon-specific translation, which has an effect on the rate and caliber of translation at particular codons and has the power to alter gene expression for the development of disease and medication resistance, is one method for the global coordination of protein

synthesis (Rapino *et. al.,* 2018). This type of regulation is different from the way promoters control transcription and the way ribosome-binding sites control translation. (Masuda et.al., 2019)

Codon-specific translation is mediated, mechanistically speaking, by post-transcriptional alterations to the tRNA anticodon or nearby nucleotides. The translation of proline (Pro) codons (CCN) is crucial for membrane-associated genes because Pro is the specific amino acid needed to form polypeptide kinks and to maintain the structure and function of trans-membrane domains (Schmidt *et. al.,* 2016).

The conserved N1 - methylation of G37 on the 3'-side of the tRNA anticodon is necessary for the translation of Pro codons, especially CC[C/U] codons. Without m1G37, tRNA is extremely vulnerable to mistakes called stalling and +1 frameshifting (Gamper *et. al.,* 2015), which disturb the reading frame and cause protein synthesis to end prematurely. (Masuda et.al., 2019)

The conserved tRNA methyl transferase *TrmD* uses S-adenosylmethionine as the methyl donor to produce m1G37 in bacteria (Hou *et. al.,* 2017) Ribosomal frameshifts build up as a result of *TrmD* depletion, and this causes m1G37-tRNA to accumulate, which causes cell death (Gamper *et. al.,* 2015a). We discovered that CC[C/U] codons are frequently present in genes linked with Gram-negative membranes, which suggests that *TrmD*'s m1G37 tRNA methylation may offer a broad way to regulate the biosynthesis of membrane proteins. (Masuda et.al., 2019)

2.2 *Streptomyces* as an Antibiotic Producer

Streptomyces is the largest antibiotic-producing genus that have produced approximately two-thirds of all known antibiotics of microbial origin (Lucas *et. al.,* 2013). Geosmin, an active volatile metabolite with a particular earthy odour, makes the members of genus *Streptomyces* unique (Scholler *et. al.,* 2002). They comprise of 40% of soil bacteria, are chemoorganotrophic gram positive bacteria producing a wide variety of pigments responsible for the color of the vegetative and aerial mycelia (Flärdh *et.al.,* 2009)

,resembles filamentous fungi, have genomes with high GC content and found predominantly in soil and decaying vegetation (Hasani *et. al.,* 2014). They grow as a mycelium of branching hyphal filaments forming septa at regular intervals, creating a chain of uninucleated spores (Ohnishi *et. al.,* 2008).

However, proper nutrients like Carbon and nitrogen sources, oxygen, pH, temperature, divalent ions like $Mn2^{+}$, $Cu2^{+}$, $Fe2^{+}$is of utmost importance for the secondary metabolite production by *Streptomyces* (Gesheva *et. al.,* 2005).

2.3 Literature Review on *trmD* as a Probable Drug Target

TrmD is an S-adenosyl methionine (AdoMet or SAM)-dependent methyl transferase that creates the methylation $m^{1}G37$ in tRNA pro, which makes it easier to generate peptide bonds when lengthening peptides with proline. *TrmD* is essentially different from its eukaryotic and archeal cousin Trm5, being unique to and necessary for bacterial development (Ito *et. al.,* 2015a) The ribosome moves backward with a frameshift of one when one nucleotide is moved in the 5′ direction, while it advances with a frameshift of one when one nucleotide is moved in the 3′ direction (Goto-Ito *et.al.,* 2017). On the 3' side of the anticodon, the methylation $m^{1}G37$ is required for preventing the tRNA+1frameshifting mistake during protein synthesis at the ribosome.

Contrary to missense faults, frameshifting errors change the translational reading frame and introduce premature termination codons, making them invariably fatal. While *TrmD* is a highly conserved sequence in both Gram-positive and Gram-negative bacteria, eukaryotes and archea do not have it. Instead, eukaryotes like humans and archeal animals make m1G37-tRNA via Trm5. *TrmD* and Trm5 are thought of as an analogous pair of enzymes despite using the same chemical substrates and producing the same chemical products, even though they are fundamentally different from one another in nearly every aspect of the methyl transfer reaction, including the global architecture of the enzyme, the binding of AdoMet, the interaction with tRNA, the discrimination of the G37 base, and the control of the rate-limiting step.*TrmD* is therefore considered a high priority antibacterial target for these reasons, and it can offer a means of isolating *TrmD*-specific

medications that wouldn't interact with Trm5, thereby lowering potential side effects (Hou *et. al.*, 2017).

	Trm5	TrmD
Reaction catalyzed	N^1-methylation of tRNA G37	N^1-methylation of tRNA G37
Organisms	Eukaryotes, Archaea	Bacteria
MTase class	Class-I	Class-IV
Protein fold	Rossmann fold, monomer	Deep-trefoil knot, dimer
Cofactor	AdoMet	AdoMet
Substrate requirement	L-shaped tRNA with G37	tRNA anticodon stem loop with G36G37 and D stem
Stoichiometry	1 tRNA/1 Trm5	1 tRNA/2 TrmD

Fig 2.1: The characteristics of transfer RNA (tRNA) methyltransferase 5 (TRM5) and tRNA methyltransferase D(TrmD)

2.3.1 Structure of *TrmD*

TrmD, an essential enzyme required for cell growth and survival, is involved in post translational modification of tRNA (Gerdes *et. al.,* 2003; Hou *et. al.,* 2017) is strongly conserved in sequence among evolutionarily diverse bacterial species (Ahn *et. al.,*2003). *TrmD* is responsible for methyl transfer from S- adenosyl methionine (SAM) to the N^1 position of the G37 base to synthesize m^1G37 on tRNA (Byström *et. al.,* 1982).

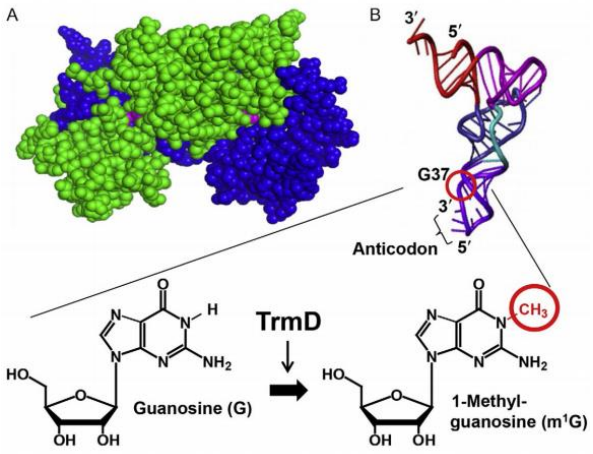


Fig 2.2: TrmD catalyzes methylation to the N^1 position of G37 in tRNA, using AdoMet as the methyl donor and producing m^1G37 on the 3'-side of the tRNA anticodon (Hou et. al., 2017).

The N^1-atom of guanosine at position 37 in transfer RNA (tRNA) is methylated by tRNA methyltransferase 5 (Trm5) in eukaryotes and archaea, and by tRNA methyltransferase D (*TrmD*) in bacteria. The resultant modified nucleotide m^1G37 positively regulates the aminoacylation of the tRNA, and simultaneously functions to prevent the +1 frameshift on the ribosome. Interestingly, Trm5 and *TrmD* have completely distinct origins, and therefore bear different tertiary folds. The importance of the m^1G37 modification is common in the three domains of life. However, depending on the domain, the m^1G37 modification is performed by two different enzymes: tRNA methyltransferase 5 (Trm5) in eukaryotes and archaea, and tRNA methyltransferase D (*TrmD*) in bacteria. Although both enzymes use *S*-adenosyl-L-methionine (AdoMet) as a methyl donor, they have distinct origins and bear different tertiary folds. Trm5 belongs to the most common class of methyltransferases (MTases): class I, which utilizes a Rossmann-fold for the AdoMet binding. *TrmD* belongs to class IV or the SPOUT class, named after SpoU and *TrmD*. The MTases in the SPOUT class possess a deep trefoil knot—which is quite exceptional for proteins—at the C-terminal region of the protein. *TrmD* requires not only G37 but also G36 as a substrate tRNA sequence, and the nine-base pair RNA duplex consisting of the anticodon and D stems with the anticodon loop can serve as its minimum substrate. In contrast, Trm5 requires G37 together with the entire tRNA structure (Christian and Hou, 2007, 2008) (Goto-ito *et.al.,* 2017)

A human TRM5 cDNA has been cloned and recombinant tRNA-N (1)G37 methyltransferase was produced. The recombinant enzyme methylates the N1 position of guanosine 37 (G37) in selected tRNA transcripts utilizing S-adenosyl methionine. G37-methylation by TRM5 occurs regardless of the nature of the nucleotide at position 36. TRM5 also methylates inosine at position 37 unlike *TrmD*, which recognizes the G36pG37 motif preferentially and does not methylate inosine. The TRM5 enzyme is sensitive to subtle changes in the tRNA-protein tertiary interaction leading to loss of activity. The

TrmD enzyme is more tolerant of alterations in tRNA-protein tertiary interactions as long as the core tRNA structure and the G36pG37 are present. The TRM5 enzyme does not have an absolute requirement for magnesium ions, whereas *TrmD* requires magnesium to express activity (Brule *et.al.,* 2004)

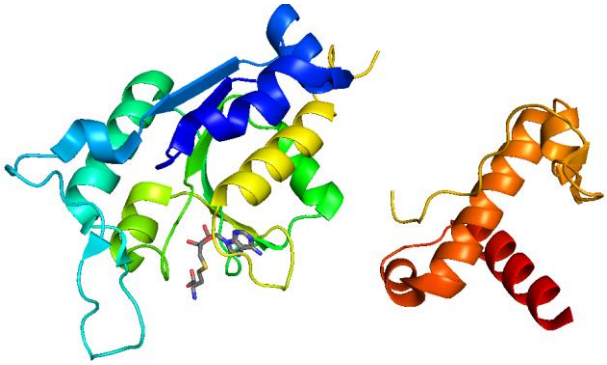


Fig 2.3: trmD of Escherichia coli (1P9P) (RCSB.pdb)

2.3.2 tRNA methylation and its role

The *TrmD* product m^1G37-tRNA helps to maintain the translational reading frame accuracy in bacteria (Björk *et. al.,* 1989) and is a key suppressor of +1 frameshift errors. Various results have shown that in the absence of m^1G37-tRNAPro there is higher chance (raised by almost 10-fold) of inducing +1 frameshifting when in the presence of m^1G37-tRNA the chance is only ~1%. However, at any of the effect of m^1G37-tRNA is smaller (three- to four-fold) on downstream positions. Thus, m^1G37-tRNA is important for minimizing +1-frameshifting throughout all sequence contexts (Hou *et. al.,* 2017)

2.3.3 Mechanism of synthesis of m1G37-tRNA by *TrmD*:

TrmD produces m1G37-tRNA via a post-transcriptional process. The enzyme's effective methylation provides evidence that it catalyzes a primary reaction on tRNA without the need for any previous modifications (Christian *et. al.,* 2010). Unlike Trm5, which demands the entire tRNA structure, *TrmD* simply needs a stem loop structure. Both G37 and a preceding G36, the 3'-nucleotide of the anticodon, are present in methylated tRNA in bacteria. According to research on the *TrmD*-tRNA complex, G36 gets stacked in the empty space between positions 35 and 38 after G37 is flipped out of the anticodon loop

and identified by chain B's flexible linker. G36 therefore offers additional interactions to improve the location of G37 while G37 serves as the substrate for methylation (Kim *et. al.,* 2013).

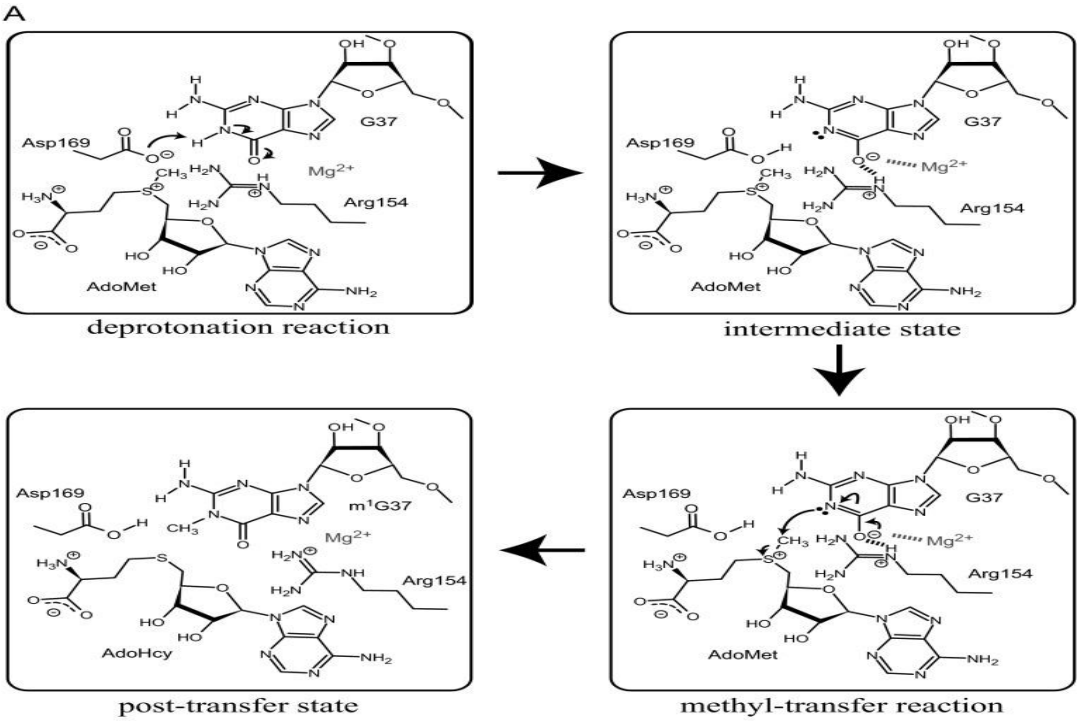


Figure 2.4: Mechanism of methyl-transfer reaction.

The catalytic pocket is invaded by the base moiety of G37, which is flipped out from the anticodon loop. The side-chain of Asp169 in subunit B forms a hydrogen bond with the 1-NH and 2-NH2 groups of G37, making the identification of guanosine unique (Elkins *et. al.,* 2003). *TrmD*'s methyl transfer activities were significantly reduced by the alanine replacements of Arg154 and Asp169, two residues that play fundamental roles in the catalysis of methyl transfer. The 1-NH of G37 and Arg154 are anticipated to donate a proton to Asp169, stabilizing the resulting deprotonated intermediate state until the methyl transfer event (Ito *et. al.,* 2015b).

2.3.4 Mechanism of frameshifting in absence of *TrmD*:

According to Björk, Wikström, and Byström (1989), the bacterial ribosome typically converts three mRNAs into a single amino acid at a rate of about 20 residues per second

by moving the associated tRNA from the A-site (the aminoacyl-tRNA site) to the P-site (the peptidyl-tRNA site) and through the E-site (the exit site). It is discovered that *E. coli* ribosomes only sometimes frameshift in either the 5' or 3'direction, despite these intricate and dynamic movements. The ribosome moves backward with a -1 frameshift when one nucleotide is moved in the 5'direction, and moves forward with a +1 frameshift when one nucleotide is moved in the 3'direction. Unexpectedly, frameshifts are used by bacteria as a "planned" strategy to control gene expression.

For instance, the prfB gene's planned +1 frameshift (for the expression of release factor RF2) and the dnaX gene's planned 1 frameshift (for the expression of DNA polymerase III) (Kim *et. al.,* 2013).

Non-programmed frameshifts, on the other hand, are classified as translational mistakes and typically result from the shifting of a tRNA-ribosome complex on the slick mRNA sequences CC[C/U]-[C/U]. The reading frame is altered and premature termination codons are introduced by the damaging non-programmed frameshift mistakes. Over all sequence contexts, the frequency of such frameshifts is less than 0.003% (Jrgensen & Kurland, 1990). One of the most important suppressors of +1 frameshift mistakes is methylation guanosine37-tRNA. As a result, m1 G37tRNAPRO, a byproduct of the enzyme *TrmD*, is crucial for the precision of the translational reading frame in bacteria (Gamper, *et. al.*, 2015)

The proline-coding mRNA sequence CC[C/U]-[C/U] is the one that is most prone to frameshifts. in every bacterial species. All three of the tRNAPRO isoacceptors are said to possess m1G37 in its native state. The UGG isoacceptor of tRNAPRO reads all four Proline codons either in the 0-frame or the +1-frame with a comparable free energy of stabilization. This makes it the most versatile of the three isoacceptors of tRNAPRO. The GGG isoacceptor, which can read the mRNA slippery sequence with identical stability in the 0- and +1frame, is the following shift-prone tRNA (Gamper, Masuda, Frenkel-Morgenstern, & Hou, 2015).CC[C/U]-N occurs 17,000 times in the K12 genome of *E. coli* and 20,000 times in the LT2 genome of *Salmonella* out of all the sense codons in protein-

coding genes. The high frequency of occurrence of slippery sequences makes it extremely difficult for ribosomes to prevent +1-frameshifting during translation. In *E. coli*, the less common codon sequence CC[C/U]-[C/U] is found 2300 times. (Jrgensen & Kurland, 1990).

These sluggish sequences can appear right near to the codon AUG in the second position or very close to the start codon. Several investigations showed that when cells lack m1G37- tRNAPRO, the sequence is more likely to cause frameshifts at the 2nd codon position. The frequency of +1frameshifting is less than 1% when cells produce m1G37, but it increases about 10-fold when cells lose the methylation. Even though m1G37-tRNA is crucial for preventing +1-frameshifting in all sequence situations, its impact is greatest at the second codon position (Gamper *et. al.,*.2015)

EF-P, the translation factor that promotes the formation of the first peptide bond, maintains the initiator tRNA's stable positioning at the ribosomal-P site, which is necessary for the translation of the second codon in bacterial mRNA because it is unique (Doerfel *et. al.,* 2013). The cell-based reporter test demonstrates that +1frameshifting is reduced by both EF-P and *TrmD*. *TrmD* is necessary for cell growth, but EF-P is necessary for the majority of bacteria to maintain robust cell viability. EF-P and *TrmD* are strictly conserved in bacteria. Therefore, the conserved process is demonstrated to be occuring to take action as early as possible to perhaps retain the precision of the protein synthesis reading frame by the well-regulated actions of both *TrmD* and EF-P. (Ude *et. al.,* 2013).

The conclusion drawn from the foregoing description is that *TrmD* is a significant and essential bacterial enzyme that has the potential to be a high-priority target for antimicrobial drugs due to its necessity for bacterial growth, strict conservation across a wide range of bacterial species, distinct structure and mechanism from that of its human counterpart Trm5, and the presence of an AdoMet binding site, which also allows for the binding of drug-like small molecules. (Goto-ito et.al.,2017)

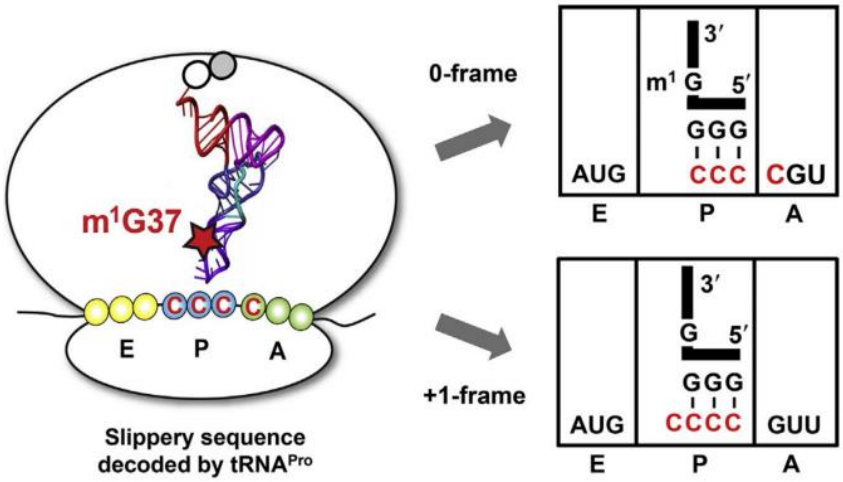


Fig 2.5 Diagramatic illustration of suppression of ribosomal +1-frameshifting by m1G37-tRNA (Hou et. al., 2017)

2.4 Review of literature on homology modelling

A protein's three-dimensional structure can frequently reveal important information. Additionally, structural information is helpful in the process of developing new drugs. The structure might offer suggestions regarding the functional and evolutionary characteristics of the protein. Theoretically, there are three ways to determine the structure of a protein: using experimental data, typically from X-ray crystallography or NMR spectroscopy; using only theoretical methods; or using homology modeling. It is still irrational to think that the structure of more than a tiny portion of the billions of proteins in the world will be investigated by experimental methods in the near future, despite significant advancements in structural genomics research.

The bulk of proteins do not yet appear to be amenable to high-resolution information being provided by purely theoretical methods. Therefore, using homology modeling techniques is the sole approach to obtain structural information for the vast majority of proteins. The notion that proteins with comparable evolutionary histories have similar structures is used in homology modeling techniques. (Wallner and Ellofson, 2004)

A common issue in biology is the functional characterization of a protein sequence. An accurate three-dimensional (3-D) structure of the protein under study usually makes this

work easier. Comparative or homology modeling frequently offers a helpful 3-D model for a protein that is related to at least one known protein structure in the absence of an experimentally established structure. (Marti-Renom, *et. al.,* 2000; Fiser, 2004; Misura and Baker, 2005; Petrey and Honig, 2005; Misura *et. al.,* 2006)

Based mostly on its alignment to one or more proteins with known structures, comparative modeling predicts the 3-D structure of a particular protein sequence (target) (templates) (Wallner and Ellofson, 2004).

There are four key steps in comparative modeling. (Marti-Renom, *et. al.,* 2000)

i) fold assignment, which determines whether the target structure and at least one well-known template structure are similar
ii) alignment between the template with the target sequence (s)
iii) constructing a model based on alignment with the selected template(s)
iv) identifying model flaws.

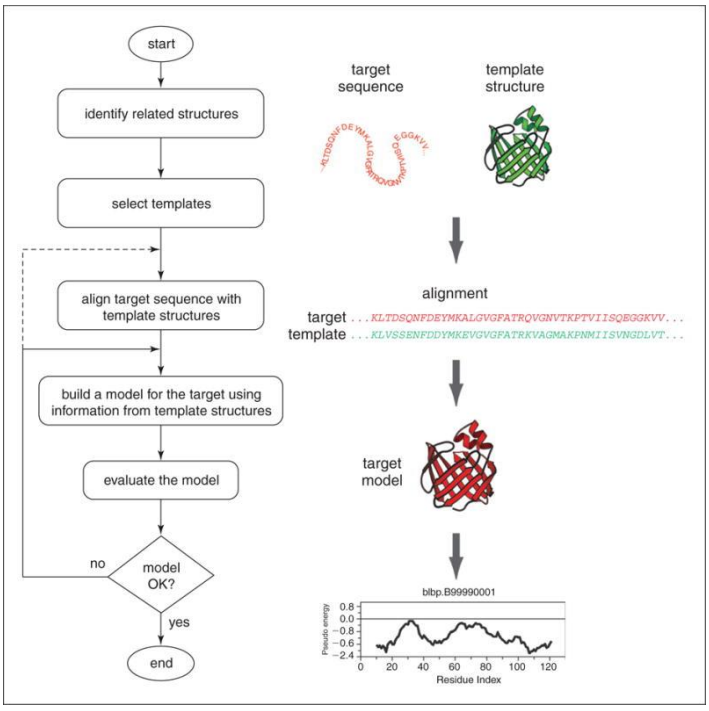


Fig 2.6: Steps in comparative protein structure modeling (Sali and Web, 2020)

2.4.1 MODELLER as a homology modelling tool

A program for modeling comparative protein structures is called MODELLER (Sali and Blundell, 1993; Fiser *et. al.,* 2000). In the most basic scenario, the input consists of a simple script file, the atomic coordinates of the templates, and an alignment of the sequence to be modeled with the template structures. Then, without requiring human input and in only a few minutes on a contemporary PC, MODELLER automatically calculates a model containing every non-hydrogen atom. In addition to model building, MODELLER is capable of other auxiliary tasks such as fold assignment, alignment of two protein sequences or their profiles, multiple alignment of protein sequences and/or structures, calculation of phylogenetic trees, and de novo modeling of loops in protein structures (Marti-Renom *et. al.,* 2004; Madhusudhan *et. al.,* 2006; Madhusudhan *et. al.,* 2009). (Fiser *et. al.,* 2000).

2.5 Review of Literature related to in-vitro Drug Design

2.5.1. Computer Aided Drug Design

CADD is capable of increasing the hit rate of novel drug compounds because it uses a much more targeted search than traditional HTS and combinatorial chemistry. It not only aims to explain the molecular basis of therapeutic activity but also to predict possible derivatives that would improve activity. In a drug discovery campaign, CADD is usually used for three major purposes:

i) filter large compound libraries into smaller sets of predicted active compounds that can be tested experimentally
ii) guide the optimization of lead compounds, whether to increase its affinity or optimize drug metabolism and pharmacokinetics (DMPK) properties including absorption, distribution, metabolism, excretion, and the potential for toxicity (ADMET)

iii) design novel compounds, either by "growing" starting molecules one functional group at a time or by piecing together fragments into novel chemotypes.

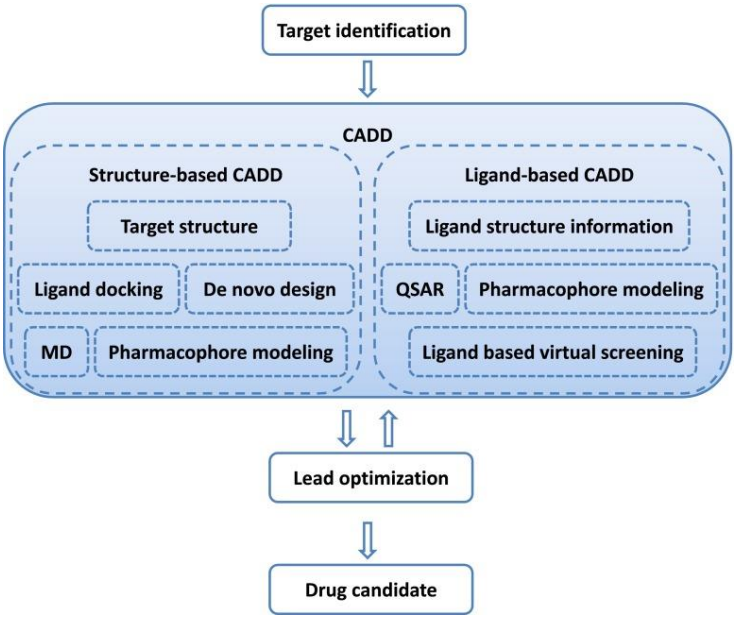


Fig 2.7: CADD in drug design pipeline(SIwoski et. al., 2014)

2.5.1.1 CADD Types

Recently, the value of virtual screening in the drug development process has increased, particularly for lead identification, lead optimization, and scaffold hopping. When compared to high throughput screening, virtual screening is quicker and less expensive when looking for new medications (John *et. al.,* 2010). In general, ligand-based virtual screening and structure-based virtual screening are the two categories of computational approaches for virtual screening. The main focus of the current study is structure-based virtual screening, where we have taken advantage of the corona virus main protease's structural insights for drug discovery purposes. While structure-based drug design, also known as ligand docking methods, makes use of available protein three-dimensional (3D) structures and preferred ligands to study the binding poses and affinities through docking studies, ligand based virtual screening, also known as ligand similarity methods, is based on the idea that ligands with similar physio-chemical structures to active ligands may have

enhanced drug like properties against target compared to random ligands (Hamza *et. al.,* 2012).

2.5.1.1.1 Ligand-Based Virtual Screening (LBVS)

Instead of focusing on the three-dimensional structure of a target protein and the binding modalities and affinities with particular ligands, ligand-based virtual screening methods in particular take advantage of the information extracted from known active ligands for both lead identification and optimization. When there isn't a 3D structure of the target protein, ligand-based approaches are preferred (Hamza *et. al.,* 2012). In drug development, one frequently knows that a known set of ligands are active against the target of interest even while the protein structure for the target of interest is difficult to create. In these situations, the sole option may be ligand-based virtual techniques, which involve discovering new ligands by comparing potential ligands to known active chemicals that are already acting potently against abnormalities or pathological states. (Tresadern *et. al.,* 2009). Instead of focusing on the three-dimensional structure of a target protein and the binding modalities and affinities with particular ligands, ligand-based virtual screening methods in particular take advantage of the information extracted from known active ligands for both lead identification and optimization. When there isn't a 3D structure of the target protein, ligand-based approaches are preferred (Hamza *et. al.,* 2012). In drug development, one frequently knows that a known set of ligands are active against the target of interest even while the protein structure for the target of interest is difficult to create. In these situations, the sole option may be ligand-based virtual techniques, which involve discovering new ligands by comparing potential ligands to known active chemicals that are already acting potently against abnormalities or pathological states. For this technique to be effective, there must be at least one known active substance against the biological target. By using these techniques, the known active compounds can be gathered to choose a query for virtual screening or alignment in ligand-based design.

The factors like an effective similarities measure and a trustworthy scoring function must be addressed by the ligand-based virtual screening approaches (Hamza *et. al.,* 2010).

Despite its numerous drawbacks, this computational strategy is particularly effective against targets of G-protein coupled receptors or protein structures resolved in the apo-form (Rai *et. al.,* 2010). (Jegerschold *et. al.,* 2008).

2.5.1.1.2 Structure-Based Virtual Screening (SBVS)

Target-based virtual screening (TBVS), also known as structure-based virtual screening (SBVS), is focused with determining which ligands will match a molecular target the best. The ligands are then ordered in accordance with their affinity to the target, with the most promising substances typically at the top of the list. The target protein's 3D structure is a requirement for the SBVS in order to anticipate in-silico the interactions between the target and each chemical component (Liu *et. al.,* 2018). According to their interaction with the binding site, the compounds are chosen from several databases and sorted in this approach. Due to its low computational cost and superior results, molecular docking is one of the more dependable SBVS approaches (Meng *et. al.,* 2011).

Despite being created in the 1980s (Kuntz *et. al.,* 1982), this technique gained popularity in the 1990s due to improvements in processing power and accessibility of the target proteins' structural data. The key element of a docking approach is a trustworthy scoring function that can adequately explain the affinity of ligands and molecular targets (Leelanada *et. al.,* 2016).

Many benefits of SBVS include saving time and money, evaluating a synthetic molecule in a lab even before it is synthesized, and more (Maia *et. al.,* 2020). The shortcomings of SBVS include the difficulty in parameterizing the complicated ligand-receptor binding interactions, the inability to accurately anticipate the correct binding poses, and the inability to classify molecules (Lionta *et. al.*, 2014). Nevertheless, numerous studies using SBVS have been used in recent years' drug discovery research (Carpenter *et. al.,* 2018; Dutkiewicz and Mickstacka, 2018; Nunes *et. al.,* 2019).

2.5.1.2 Steps of Structure Based Virtual Screening

In the initial step of SBVS, the desired 3D target structural information is acquired. Homology modeling, Molecular Dynamics (MD) simulations, experimental data (X-ray,

NMR, or neutron scattering spectroscopy), and other sources may all provide the necessary three-dimensional structure of the target protein. When choosing a biological target for SBVS, there are a number of fundamental factors that should come first. These include the receptor's druggability, the binding site chosen, the most pertinent protein structure from the range of options, the incorporation of the receptor's flexibility, the proper assignment of protonation states, and the consideration of water molecules in the binding site, to name a few (Reddy et.al., 2007). The precise prediction of ligand binding locations on biological targets is another crucial element. The target protein in question must be taken into consideration when selecting the set of compounds from a library to be tested in the VS, and the preprocessing steps used to create those libraries must satisfy the requirements for assigning the correct stereochemistry, tautomeric, and protonation states.

The individual compound in the library is then subjected to the docking process utilizing a specific docking program after the successful production of the ligand library and receptor preparation. The structure of the ligand-protein complex is revealed by docking. Now, the scoring function can be used to determine the free energy of binding between the ligand and protein under investigation. The probable compounds are rated after docking and scoring, and post processing can then be used to enable for additional experimental testing (Koppen, 2009).

2.5.1.2.1 Protein preparation

The result of the SBVS method depends on the plausible starting structures for both the protein and the ligand. A PDB file contains a number of heavy atoms, water molecules, cofactors, activators, ligands, and metal ions in addition to a number of protein subunits. In addition, the structure might not contain accurate data on the bond ordering, topologies, or formal atomic charges. Because the X-ray structures cannot clearly discriminate between O and NH2 groups, there can be an improper arrangement of terminal amide groups and asparagine residues. Ionization and tautomeric states

frequently lack correct assignment, too. Due to the limited resolution of a specific protein location, there may be a chance of missing residue side chains or bigger loops, as well as a chance of steric conflicts (Lionta *et. al.,* 2014).

Several protein preparation strategies have been put up to overcome the structural issues discussed above. PROPKA (Li *et. al.,* 2005), H++ (Anandakrishnan *et. al.,* 2012), and SPORES (ten Brink and Exner, 2010) are examples of programs that can be used to determine the protonation states of proteins in certain structures. Similarly, for assigning hydrogen atoms and optimizing protein hydrogen bonds according to an optimal hydrogen bond network, PDB2PQR (Dolinsky *et. al.,* 2007) software is used. Next, partial charges are assigned, residues are capped, metals are treated, missing loops and side chains are filled in, and the protein structure is minimized to prevent steric conflicts (Young *et. al.,* 2007).

Additionally, it must be decided whether water molecules will stay in the binding site or be eliminated. Numerous strategies have been created to address this difficult issue (Michel *et. al.,* 2009). It is better to use best preparation strategies for SBVS rather than employing minimally prepared PDB structures given the impact of protein preparation on docking performance (Sastry *et. al.,* 2013).

2.5.1.2.2 Binding site Identification

The proper implementation of the SBVS system requires accurate binding site prediction. Different methods could be used to predict binding locations. To map the binding hotspots on the 3D protein target using the static techniques, computational solvent mapping with chemical probes is used (from X-ray, MD). FTMap (Ngan *et. al.,* 2012), Fpocket (Le Guilloux *et. al.,* 2009), and others are some typical instances. Another method involves the dynamic evolution over time of the protein and probe, which may entail more than one probe. The right binding sites can subsequently be identified using the generated interaction free energy. This methodology is used in the program MDMix (Seco *et. al.,* 2009). Utilizing water molecules as probes to anticipate potential binding sites in

proteins, as done in JAWS, is the third way for identifying binding sites (Michel *et. al.,* 2009).

2.5.1.2.3 Compound Library Preparation

The compound database is made up of small, drug-like chemicals that are frequently publicly available or may be purchased or synthesized; creating this library is an essential step in the SBVS process. The existence of adequate functional groups to interact with biological protein targets through various chemical interactions, stability and solubility in aqueous conditions, and absence of toxicity and undesired moieties are prerequisites for the tiny drug-like compounds. The "Lipinski Rule of Five" (Lipinski *et. al.,* 2001) is the most well-known rule that defines "drug-likeness" as having a molecular weight less than 500, a lipophilicity (logP) less than 5, fewer than five hydrogen bond donors, and fewer than ten hydrogen bond acceptors. The "Rule of Five" is not followed by 50% of marketed drugs and many drugs made from natural products (Zhang *et. al.,* 2007). The Lipinski Rule of Five has since undergone numerous adaptations and adjustments in an effort to improve forecasts of druglikeness.

A drug-like molecule may have one or more of the following properties: a polar surface area less than 140 2, less than ten rotatable bonds, a logP range of 0.4 to +5.6, a molar refractivity range of 40 to 130, a molecular weight range of 180 to 500, and an atom count range of 20 to 70. (Veber *et. al.,* 2002). According to a novel method known as "Pfizer's Rule of 3/75," substances with estimated partition coefficients (ClogP) less than 3 and topological polar surface areas (TPSA) greater than 75 are around 2.5 times more likely to pass safety tests in vivo (Hughes *et. al.,* 2008). The "Jorgensen Rule-of-Three," another parameter for evaluating the drug-like properties of small molecules, states that the apparent Caco-2 cell permeability should be faster than 22 nm/s, the number of primary metabolites should be less than 7, and the aqueous solubility measured as logS should be greater than -5.7. (Kerns and Di, 2008). Compound selection must include choosing compounds with lead-like properties, excluding known toxicophores or metabolically

liable moieties, and avoiding blind docking of large compound collections like PubChem (30M compounds) and ChemSpider (26M compounds), which is computationally and time-consuming. The efficient filtration of chemicals from a sizable library according to parameters like adsorption, distribution, metabolism, excretion, and toxicity (ADMET) properties is possible with the help of a number of online programs.

By utilizing its many features, such as basic search, filtration (based on steric clashes and toxicity), advanced filtering based on custom chosen physicochemical properties clustering (according to structure and compound physicochemical properties providing representative compounds for each cluster), customized pipeline, and vi, Chembioserver (Athanasiadis *et. al.,* 2012), a publicly available online application, facilitates compound preparation prior to (or after) VS computations.

For SBVS, it should be a top focus to create accurate bond lengths and angles in 3D representations. Along with assigning the correct bond order and filled valences, partial charges (if present), an appropriate protonation state at physiological pH or at the pH of interest, and proper designation of tautomeric states, a probable drug-like molecule must also be free of accompanying fragments such as counter-ions, metals, and solvents (Kalliokosky *et. al.,* 2009). The majority of docking solutions incorporate ligand preparation software like Autodock Tools (Morris *et. al.,* 2009).

2.5.1.2.4 Library Design

For a successful SBVS approach, a tailored library may be more beneficial than openly accessible internet libraries (Lavecchia and Di Giovanni, 2013). The online tool CLEVER (Song *et. al.,* 2008) may be helpful in the library design process because it supports the manipulation of chemical libraries, enumeration of combinatorial chemical libraries by user-specified chemical components, chemical format conversion, and chemical compound analysis and filtration with respect to drug-likeness, lead-likeness, and fragment-likeness based on the physicochemical properties computed from the derived molecules. e-LEA3D is an additional tool for designing libraries; it performs design work

using a user-generated scaffold and reactants obtained, for example, from a chemical vendor (Douguet, 2010).

2.5.1.2.5 Docking and scoring functions

There are several docking software applications in use right now, including AutoDock (Morris *et. al.,* 2009), Dock (Ewing *et. al.,* 2001), FlexX (Rarey *et. al.,* 1996), Glide (Friesner *et. al.,* 2004), and others (Cheng *et. al.,* 2012). In order to rate the compounds following molecular docking, the prediction analysis regarding the interaction between protein and ligand is combined with SBVS scoring. Systematic techniques, stochastic or random torsional searches about rotatable bonds, molecular dynamics simulation techniques, and energy minimization for determining a molecule's energy profiles are all used by docking systems to identify ligand conformational space (Guido et.al., 2008).

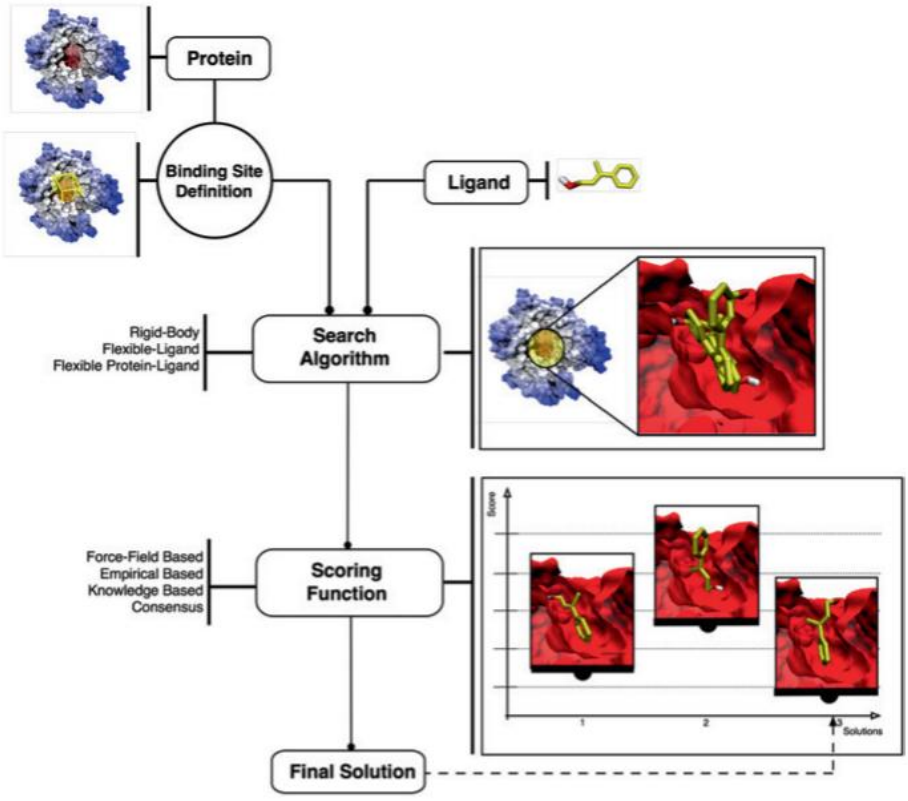


Fig 2.8: Schematic diagram of molecular docking protocols (Brás et. al.,. 2014)

In order to rank compounds, scoring functions are used to evaluate the free energy of association of a ligand to a particular biological target based on a generated docked pose following docking various ligands from a database (Cheng *et. al.,* 2012). In force-field based functions, the binding energy can be calculated by adding the strengths of various chemical interactions such as intermolecular van der Waals, electrostatic interactions, and hydrogen bonds between all of the atoms of the two binding partners in the ligand-protein complex.

In addition, ChemScore and Fresno are two instances of empirical scoring functions that count the number of different kinds of interactions between the two binding partners, including the number of hydrophobic contacts, hydrogen bonds, and rotatable bonds immobilized in complex formation (Lionta *et. al.,* 2014). Another function that addresses hydrogen bonds, hydrophobic effects, and desolvation parameters is HYDE (Schieder *et. al.,* 2012). The effectiveness of the SBVS scheme is based on the construction of an accurate scoring function. Including other characteristics, such as protein flexibility, as done by Spitzer *et. al.*, could help the docking and scoring algorithms.

2.5.1.2.6 Post Processing

A computational chemist frequently finds it difficult to select the molecules that will produce more satisfactory results in the experimental test phase after performing the VS/docking exercise. This is because this step is the rate-limiting step in SBVS. Simplified scoring functions and improper sampling of the conformational space for the ligand result in concerns including unrealistic poses, steric conflicts, twisted amides, an imperfect hydrogen-bonding network, and poses based on shape complementarity (Waszkowycz, 2008). Therefore, an excessively high score brought on by a stance causes the SBVS to perform less effectively. In order to choose the suitable compound set for additional downstream operations, the medicinal chemist typically needs to perform a correct visual evaluation of all docking postures.

Considerable attempts have been made to improve the effectiveness and standard of compound selection (Athanasiadis *et. al.,* 2012).

The binding affinities of ligand-protein complexes can be precisely predicted using a directional approach known as MIECs (molecular interaction energy components), which also distinguishes between ligands that bind and those that do not. (2013) (Ding *et. al.,*.). Malmstrom *et. al.,* demonstrated how the idea of free energy of binding (FEB) rescoring and other empirical docking scores may help with correct lead identification resulting from an SBVS exercise by using a collection of projected binders and non-binders.

As an alternative, Vdw filtering using online resources like ChemBioServer (Athanasiadis *et. al.,* 2012) may assist eliminate compounds with steric conflicts and poses that are outside of the energy minimum, followed by heirarchial clustering to find compounds with related physicochemical attributes (Lionta *et. al.,* 2014).

2.5.1.3 Advancements in SBVS

The effectiveness of SBVS is hampered because the choice of the ideal crystal structure for the receptor target will be under close scrutiny due to the limited information on protein dynamics provided by a single conformation of the protein, which is again dependent upon the crystallization conditions. The efficiency of SBVS is also impacted by conformational changes that happen after ligand binding in both the protein and the ligand. As a result, even though some crystal structures are better as a starting point for SBDD, they may give inaccurate information following the induced fit effects of proteins following ligand binding. Because it gives a more accurate representation of the modeled system, it is vital for docking programs to include receptor flexibility (Korb *et. al.,* 2012).

Despite concentrating on a single protein conformation and taking into account the idea of protein flexibility, potential ligands can be docked utilizing an ensemble or collection of sample structures. Contrary to the conventional single rigid receptor docking approaches, ensemble docking is based on the docking of a single ligand library against

numerous rigid receptor conformations and aids in identifying the realistic structural variation of the protein-target (Craig *et. al.,* 2010).

The downside of this approach, despite the fact that it takes into account any type of protein motion, may be the higher computational power requirement caused by the generation of a larger number of protein conformations in the ensemble and inaccurate outputs from less reliable scoring functions that result in inaccurate predictions of ligand-protein interactions (Osguthorpe *et. al.,* 2012).

Consensus induced fit docking (cFID), used by Kadid *et. al.,*. on COX-2, is another alternative docking procedure that requires the protein binding site to respond to many ligands during SBVS. Due to the method's emphasis on predicted binding modes rather than predicted binding energies, it can also be used to assess advancements in the accurate detection of docked positions (Houston and Walkinshaw, 2013). According to Houston *et. al.,*. 2013, the success rate for pose prediction in consensus induced fit docking is higher than it is for other individual docking programs alone.

2.6 Review of literature related to Mutation

The frequent and severe mutations that have been observed in viruses can be largely attributed to three different phenomena: their vast populations, their rapid generation durations, and their high mutation rates. Every mutation is helped by positive selection, which is handed down to the following generation for widespread dissemination, to evade the host's immune system. High mutation rates are also a result of viral RNA-lack polymerase's proofreading function. (Shao *et. al.,*2017). Errors in replication are accumulated as a result of successive cell divisions. The fact that nature has a mutation rate larger than zero means that some base pairs are always more likely to change than the original genome of each cell. (Strippoli *et. al.,* 2005)

Point mutation is the name given to a change in a single base pair of DNA. Because of individual point mutations, changes in the antigenic sites determine how viruses evolve and ultimately produce new viral subtypes (Carrat and Flahault, 2007).

Such point mutations, or spontaneous mutations, can lead to DNA changes even in the absence of outside influences. A mechanism known as quantum tunneling may be responsible for these point mutations. Quantum biologists believe that electrons act like wave-like particles at the subatomic level because they tunnel through obstacles rather than climbing over them, in accordance with the particle's dual nature as both a wave and a particle (https://the-gist.org/2017/02/schrodingers-mutations/).

By using multiple modeling techniques and quantum methodologies, it has been determined that the spontaneous mutation inside the genome is caused by proton tunneling, a quantum process. The proton spontaneously vanished from one location and spontaneously reappeared nearby. Under specific circumstances, the hydrogen bonds between the two DNA strands behave like waves and spread quickly throughout a large area due to the proton tunneling phenomena. Due to the infrequent discovery of these atoms on the incorrect strand of DNA, this fact largely explains the spontaneous mutation (https://www.surrey.ac.uk/news/new-study-reveals-quantum-physics-can-cause-mutations-our-dna).

Mutations that occur randomly are a result of quantum phenomena. The future replication mistakes can be predicted using a basic understanding of the different quantum leap events.

However, given the Heisenberg uncertainty principle, which is comparable to the genetic information uncertainty principle, it is impossible to forecast errors with absolute certainty in order to predict mutations that are more likely to emerge spontaneously in the future (Mcfadden & Al-Khalili, 1999; Strippoli *et. al.,* 2005).

Transition substitution rates are higher than would be predicted by chance when compared to transversion rates. The mutational hypothesis and the selection hypothesis both support the aforementioned assertion. According to the mutational theory, polymerases' transition mutation rates are larger than their transversion rates. According to a different theory called the selective hypothesis, natural selection favors transitions over transversions (Stoltzfus & Norris, 2016).

2.7 Review of literature related to Density Function Theory

2.7.1 Density function theory

The computational technique used to supplement virtual screening that uses density functional theory appears to be the least common. This may suggest that this method is extremely specialized for the rare circumstances in which unique complex intermolecular interactions and reaction processes must be produced (Gross and Dreizler, 2013). The fundamental problem with this method is that enzyme systems are currently too large to calculate them all in a reasonable amount of time (Gross et.al., 2013). Because of this, it is preferable to study smaller model systems or use hybrid techniques like QM/MM (Senn *et. al.,* 2009) or ONIOM (Verven *et. al.,* 2006). The fact that such calculations are used less frequently does not lessen their significance in CADD because the findings are typically more equivalent to the experimental values.

When compared to other CADD techniques, this technique's emphasis is on precision at the expense of computational cost and efficiency, which could account for its lack of popularity. This might alter as QM/MM (Senn *et. al.,* 2009) and ONIOM (Verven *et. al.,* 2006) approaches gain popularity and as hardware and software advancements take place.

2.8 Review of literature related to Molecular dynamics (MD) simulations

2.8.1 Molecular dynamics simulations

Some of the most crucial computations that come after virtual screening simulations are molecular dynamics simulations. They ought to be viewed as an advanced approach that complements docking. In order to account for the conformational dynamics related to ligand binding, they can also be used prior to docking for conformational sampling and clustering on a protein molecule (Menchon *et. al.,* 2018). By determining the stability of the complex, interatomic interactions, the rate of system fluctuation, and binding free energies, MD simulations are most frequently used to filter and validate the results acquired from protein-ligand complex docking experiments (Menchon *et. al.,* 2018). However, not all virtual screening investigations use

molecular dynamics simulations; as a result, their use is dependent on the structure of the study and the subjects being examined.

For doing molecular dynamics simulations, GROMACS (http://www.gromacs.org/), Amber (http://ambermd.org/), and other programs are widely utilized.

2.9 Review of literature related to human Cytochrome P450 enzymes

Many of the processes that are catalyzed by cytochrome P450 enzymes are vitally crucial for medication response (Guengerich, 2021). Human cytochrome P450 enzymes catalyze numerous metabolic processes that have a significant impact on the biological activities (physiologic, therapeutic, and/or toxic) of xenobiotics like drugs, natural products, general chemicals (e.g., environmental chemicals like pesticides, pro-carcinogens), and physiological compounds. There are several cytochrome P450 isoforms, but CYP2C9, CYP2C19, CYP2D6, and CYP3A4 carry out the majority of responses.

Approximately 60% of oxidized medicines are bio-transformed, at least in part, by CYP3A, making it one of the most significant human enzymes. On chromosome 7q21–q22, the human CYP3A gene cluster is found. It has four genes and two pseudogenes (CYP3AP1 and CYP3AP2) (CYP3A4, CYP3A5, CYP3A7, and CYP3A43). There are 13 exons in each of the four functional genes in the CYP3A locus. The CYP3A4 gene is situated upstream of CYP3A43, which is the sole gene in the cluster to have the reverse orientation and is placed farthest from the centromere (Finta and Zaphiropoulos, 2000). CYP3A4 and CYP3A7 are divided by a pseudogene (CYP3AP2), and CYP3A7 and CYP3A5 are divided by a pseudogene (CYP3AP1).

Every element of the gene structure, including the 5′ upstream sections, introns, exons, and the 3′ untranslated regions (UTRs), as well as the great majority of SNPs that affect each gene, particularly CYP3A4 and CYP3A5, may be detected. These SNPs are not restricted to specific locations (Lee *et. al.,* 2004).

The most researched isoform in the CYP3A subfamily, CYP3A4, is mostly present in the small intestine and adult liver and is thought to be responsible for the majority of metabolic functions attributed to the CYP3A subfamily.

3: MATERIALS AND METHODOLOGY

In vitro- test

3.1 Materials Required for Confirmation and Antimicrobial Sensitivity Test (AST) of test organisms

Test organisms, Nutrient Agar plates, Xylose-Lysine Deoxycholate (XLD) agar plates, Mac-Conkey agar plates, Luria Bertani Agar (LBA), Luria Bertani broth, Gram staining reagents, different biochemical media, Mueller Hinton Agar (MHA) plates, antibiotic discs

3.1.1 Collection and selection of the test organisms

Different bacteria were obtained from Central Department of Biotechnology (CDBT) laboratory and Animal Health Research Division (AHRD). They were strains of *Escherichia coli* and *Salmonella* Typhi.

3.1.2 Identification and reconfirmation of the test organisms

For the identification of the test organisms, the cryopreserved strains were inoculated on LB broth and then subcultured by streaking on the LBA plates. Then, single colony was transferred on respective selective agar plates. *S*. Typhi strain was streaked on Xylose-Lysine Deoxycholate (XLD) agar plates, *Escherichia coli* was streaked on Luria Bertani Agar (LBA), agar plates for further confirmation.

3.1.2.1 Microscopic Observation of the test organisms

Microscopic observation of test organisms was done through Gram staining. A thin and uniform smear of sample was prepared on the microscopic slide by mixing the sample with distilled water and allowed for air dried. Heat fixation was done to fix the smear to the slide. Then, three drops of crystal violet were flooded over the heat fixed smear and allowed to stand for one minute. It was then rinsed with tap water and three drops of Gram's Iodine was flooded and allowed to stand for one minute. Then, rinsing was done by tap water and decolourizing agent (acid alcohol) was flooded and washed by tap water immediately. After that, counter stain was flooded and allowed to stand for four to five seconds. It was rinsed with tap water again and allowed to air dry. The prepared slide was observed under microscope at 40X and 100X (oil immersion) to observe colour, shape and arrangement of the colonies.

3.1.2.2 Biochemical Tests of the test organisms

After Gram staining, different biochemical tests were performed to identify the test organisms.

Indole Test: A sterilized culture tubes containing 5 ml of tryptophan broth were taken. The tubes were aseptically sub-cultured by taking the growth from 18 to 24 hrs tests culture from Luria Bertani broth. The tubes were then incubated at 37°C for 24 hours. Three drops of Kovac's reagent were added to observe the presence or absence of ring to compare with negative control.

Methyl Red –Voges- Proskauer (MR-VP) test: Using test organisms taken from an 18–24-hour pure culture, the MR broth was inoculated. They were then incubated aerobically at 37 °C for 24 hours. Then, 1 ml of the broth was aliquoted to a sterilized culture tube. The remaining broth were incubated for an additional 24 hours. Then 3 drops of methyl red indicator were added to aliquot and the red color was observed immediately. Using test organisms taken from an 18–24-hour pure culture, the VP broth was inoculated. They were then incubated aerobically at 37 °C for 24 hours. Then, 1 ml of the broth was aliquoted to a sterilized culture tube. The remaining broth were incubated for an additional 24 hours. 5% alpha-naphthol and 40% potassium hydroxide were added in the ratio of 3:1 and mixed well to aerate. A pink-red color at the surface within 30 min was observed. The tube was shaken vigorously during the 30-minutes period.

Citrate utilization Test: The Simmon's citrate agar slant was prepared and cooled in slanted position (long slant, shallow butt). The uninoculated medium would be a deep forest green due to the pH of the bromothymol blue. A light inoculum picked from the center of a well-isolated colony was stabbed in butt and streaked the slant back and forth. It was then incubated aerobically at 37°C for up to 24 hours. A color change from green to blue along the slant was observed in case of positive result.

Triple Sugar Iron Agar (TSIA) Test: The TSI agar was cooled in slanted position (long slant, shallow butt). With a sterilized straight inoculation needle the top of a wellisolated colony was touched. The TSI agar was inoculated by first stabbing through the center of the medium to the bottom of the tube and then streaking on the surface of the agar slant. The cap was left on loosely and incubated the tube at 35°C in ambient air for 18 to 24 hours.

3.1.2.3 Antibiotic Sensitivity Test of the test organisms

Preparation of inoculums After the preliminary confirmation of the test organisms, all bacteria were tested in vitro to determine susceptibility towards different antibacterial drugs by agar disc diffusion method as described by Kirby-Bauer and WHO (Biemer, 1973). The 18 hours overnight cultures of test organisms were inoculated in the LB broth and turbidity was observed. Then, required turbidity of the bacterial suspension were adjusted by comparing with 0.5 McFarland standard. Antibiotic sensitivity testing (AST) of the susceptibility test of above strains of bacteria was carried out using MHA plates and was tested in vitro for susceptibility to different antibacterial drugs produced commercially by HiMedia company; Colistin (10 mcg and 25 mcg), Ertapenem (10 mcg), Meropenem (10 mcg), Ciprofloxacin (5 mcg), Gentamicin (10 mcg). For antibiotic sensitivity test, agar disc-diffusion testing method was followed as published by the Clinical and Laboratory Standards Institute (CLSI) for bacteria with slight modification. First, test organism suspension equivalent to that of a 0.5 McFarland standard was evenly spread onto dried MHA plates by dipping a sterile cotton swab twice into the suspension and the agar rim was made finally by the sweep of the swab. MHA plate was incubated at 37°C for half an hour to activate the test organism. Then, antibiotics discs (about 6 mm in diameter) (as mentioned above), were aseptically placed on the agar surface. The plates were incubated at 37°C for eighteen hours. Diameters of inhibition of growth zones were measured and compared with CLSI chart as either sensitive or intermediate or resistant. Colistin resistant bacteria were carried out for further analysis.

3.1.3 Revival of *Streptomyces*

The revival of isolated and stored putative *Streptomyces* were performed by serial ISP2 and ISP4 agar plates were mainly used among all the ISP medias (Shirling and Gottleib, 1996). An inoculum from the stored agar plates was inoculated in ISP2 medium. The plates were then incubated at 30^0 C for 7 to 15 days with regular monitoring until the pure culture was obtained. Colony morphology and different cultural characteristics with pigmentation were recorded.

3.4.1 Secondary metabolite production

Isolated strain Pc were inoculated on first set of autoclaved 50 ml ISP4 broths. The media were then incubated at at 300rpm/28°C for 7 days for secondary metabolite production and kept at 4

°C until use. During that period, after eight days, growth of those two strains were measured in the interval of two days upto twelve days at 600 nm.

3.4.2 Resazurin Antimicrobial Assay

The test organism suspension equivalent to that of a 0.5 McFarland standard was prepared by dilution with LB broth. For resazurin assay, a 96 well microtiter plate was taken. Then, 48 µl of bacterial culture along with 50 µl extract and 2 µl 1% resazurin solution was added in each well to maintain the final volume as 100 µl in each well and final resazurin concentration as 0.02%. The wells having only LB broth and resazurin solution without test organism were taken as positive control and the wells containing only resazurin and test organism without antimicrobial extracts were taken as negative control. 50 mg/ml ampicillin and kanamycin solution were used as positive control of antibiotics. It was then incubated at 37˚C until the color change in negative control was observed as pinkish color. Finally, absorbance reading was taken at 551 nm.

In- silico test

3.2 Search for bacterial protein target

3.2.1 Selection of bacterial protein and obtaining their genomic sequences

Among the highly prioritized pathogen that have been published by WHO in 2017 as Critical and High Priority (Lawe-Davies and Bennett, 2017), *trmD* gene of *Salmonella Typhimurium* (LT2) against whom leads are sought was considered as a target. The whole genome sequences of this protein of *Salmonella Typhimurium* (LT2), which was published in NCBI were taken for alignment and structural analysis with closely related species.

3.2.2 Alignment search using BLAST tool for *trmD* gene of *Salmonella Typhimurium*

The sequence for *trmD* gene of *Salmonella Typhimurium* (LT2) was searched for in NCBI database and BLAST tool in the same database was employed for alignment studies with the intention to find the matches in sequences to other WHO enlisted pathogen.

3.2.3 Multiple Sequence Alignment search CLUSTAL OMEGA for *trmD* gene of WHO prioritized Pathogens

The protein sequence of the *trmD* gene of eight WHO prioritized pathogens *Escherichia coli* (UNIPROT: P0A873), *Salmonella Typhimurium* (UNIPROT: P36245), *Klebsiella pneumoniae* (UNIPROT: A6TCL5), *Serratia marcescens* (UNIPROT: P36244), *Proteus mirabilis* (UNIPROT: B4EUW6), *Neisseria gonorrhea* (UNIPROT: B4RIW3), *Acinetobacter baumanni* (UNIPROT: B7IAT0), *Pseudomonas aeruginosa* (UNIPROT: V6AAA4) was downloaded in FASTA format and the Clustal Omega tool was used for performing the multiple sequence analysis and construction of phylogenetic tree amongst the eight strains in consideration.

3.3 Homology Modeling of target protein and its validation

The lead protein that was selected did not have available crystal structures in the databases and thus various homology modelling tools (MODELLER) were employed to generate the required gene the Z-score analysis and Ramachandran plot analysis was done in proSA web server and SAVES V 5.0 web server respectively to validate the target protein.

3.4 Molecular docking simulation

3.3.1 Obtaining the dockable crystal structures of the target protein

The protein that was developed using modeler (*trmD* B905) was selected as target protein. The protein was downloaded in the pdb format.

3.3.2 Preparation of ligand database

In the present work, structure based virtual screening was performed. Structure based Virtual Screening was done with a primary aim to avoid partiality among the class of molecules and a broad chemical database with structural diversity may include diverse class of molecules that may be sufficient to develop potential leads against the chosen viral target. In the context of global health crisis, opting to a library of compounds which is already suitable for use in human with limited side effect profiles would save both the time and efforts to address this urgency. In this study, a ligand database containing ligand molecules was prepared that included FDA approved

compounds from ZINC database (Irwin and Shoichet, 2005). In one of the works (Wu *et. al.,* 2005) indirubins have been found to be kinase inhibitors. In earlier works, assuming kinase inhibitors could be a potential compound that could compete with SAM due to adenosine moiety, NCI diversity database II was screened and 4-methyl-N-(3-phenylphenyl) piperazin-4-ium-1-carboxamide, an indole derivative had binding energy higher than SAM in *trmD protein*. This showed that the indole containing compounds could be potential leads against pathogens.

3.3.3 Protein and ligand preparation

The dockable protein structures and ligands need certain preparation before docking simulations. Prior to molecular docking, the proteins and ligands need to be prepared for efficient and more accurate docking results using different tools. Protein preparation is done by adding hydrogen atoms, merging non-polar bonds, adding Gasteiger charges and finally converted to pdbqt format that is required for performing docking studies in mgltools (http://mgltools.scripps.edu/). Similarly, ligand preparation was done in Openbabel GUI (O'Boyle *et. al.,* 2011) available in PyRx 0.9.8 setup by energy minimization using forcefield and converted to pdbqt file format, a useable file format for docking afterwards.

3.3.4 Setting reference values for docking

The target protein was docked with the native ligand s-adenosyl methionine (SAM). The highest binding energy thus calculated was taken as a reference value for identifying potential leads from the ligand database, docked against the respective binding pockets of the target proteins.

3.3.5 Structure based Virtual Screening

AutoDock Vina was used to perform virtual screening in a platform, PyRx (Dallakyan and Olson, 2015) against the target protein (*trmD*) with the selected ligand database. The conformation with the lowest docked energy was chosen after the docking interactions and visualized and analyzed using Discovery Studio.

3.3.6 Preparation of ligand database from top hits obtained from docking studies for final docking

Following the docking operations of the chosen ligand dataset to the target protein in consideration, the top hits were identified above the reference value i.e., the value of reference inhibitor SAM in this case

3.3.6.1 Obtaining the 3D SDF of individual hits and merging them to prepare final ligand dataset

The obtained lead compounds after docking studies were searched in the PubChem database for 3D sdf structures and these compounds were downloaded and renamed and merged using the Osiris datawarrior tool to prepare a final dataset of ligands.

3.3.7 ADME/TOX filter

The process of drug discovery and development are very expensive processes and take a lengthy span of time. Many promising lead compounds initially regarded better, fail to reach the clinical stage of drug development mainly in relation to unacceptable pharmacokinetic properties as well as toxicity hazards. The toxic profiles and drug likeness of potential lead compounds obtained from docking studies were predicted using OSIRIS program (Nisha *et. al.,* 2016). OSIRIS calculates various drug relevant properties like molecular weight, cLogP, cLogS, Druglikeness, total polar surface , toxicities like mutagenicity, tumorigenicity, reproductive effects and irritant effects and rotatable bond counts in the lead molecules on the basis of functional groups and chemical bonds present in their structures (Sander *et. al.,* 2015).

3.3.8 Final Structure based Virtual screening

AutoDock Vina was used to perform virtual screening in a platform, PyRx (Dallakyan and Olson, 2015) against the target protein (main protease)with the selected ligand database which contains 3D structures from PubChem database. The hits obtained yielding the higher binding energy than the reference inhibitors were further considered and analyzed for different properties.

3.4 Analysis of docking results

3.4.1 Binding interaction of ligands with protein target

The docking of a lead compound to the target protein involves certain chemical interactions which impart to the stability of the lead molecule in that association. The amino acids residues involved in making possible interactions with the protein structure were looked for using the pymol software version 4.6.0.

3.4.2 Bond length and bond types responsible for drug ligand interaction

The interaction of chemical lead with its target protein always depends upon the different kinds of feasible chemical interactions. A ligand is bound to its target protein with different bond length and different types of interactions which were analyzed by the using Discovery Studio Visualizer.

3.4.3 Hydrophobic bond interactions

Different types of bonding interaction play a crucial role in establishing the stable association of a ligand to the binding pocket of the enzyme. Among them, one of the essential bond types is hydrophobic interaction which can be studied by using Ligplot software version.

3.6 DFT calculations

The selected structures of probable drugs were transferred to the Gaussian 03W (Vittinhoff *et. al.*, 2007) software for optimization with the B3LYP/6–31 G method based on density-functional theory (DFT), to extract a set of a quantum chemistry descriptors, namely the energies of the highest occupied molecular orbit HOMO and the lowest vacant LUMO, the total energy of the molecule (Et) and the dipole moment.

4: RESULTS AND DISCUSSION

In-vitro ASSESSMENT

4.1 Pathogen

The identification and characterization of the pathogenic test organism was done to make sure that the pathogen did not adapt within the storage period. Various literature suggests that the regular and accurate scrutinization measures are of need to maintain the resistance pattern (Bhattacharya *et.al.,* 2013).

The bacterial strains (ATCC and pathogens) that have been stored as cryostocks were retrieved in LB broth. NA plates and respective selective agar plates for respective organisms. After streaking of a loop full of slightly thawed cryopreserved stock of the respective organisms in respective media containing plates, the single colony isolated (Figure 4.1) was taken as the starting material for further works. The respective media in some cases act as selective media, thus giving purity of the strain. Once the single colonies of respective bacteria were isolated, the subsequent works if otherwise not mentioned specific media, the works were carried out in LB or LBA media.

XLD Agar is both a selective and differential medium. It contains yeast extract as a source of nutrients and vitamins. It utilizes sodium deoxycholate as the selective agent and therefore, is inhibitory to gram-positive microorganisms. Xylose is incorporated into the medium since it is fermented by practically all enterics. Lysine is included to enable the *Salmonella* group to be differentiated from the non-pathogens since without lysine, salmonellae rapidly would ferment the xylose and be indistinguishable from non-pathogenic species. After the salmonellae exhaust the supply of xylose, the lysine is attacked via the enzyme lysine decarboxylase, with reversion to an alkaline pH. Degradation of xylose, lactose and sucrose to acid causes phenol red indicator to change its colour to yellow. To add to the differentiating ability of the formulation, an H_2S indicator system, consisting of sodium thiosulfate and ferric ammonium citrate is included for the visualization of the hydrogen sulfide produced, resulting in the formation of

colonies with black centers. The nonpathogenic H_2S producers do not decarboxylate lysine; therefore, the acid reaction produced by them prevents the blackening of the colonies which takes place only at neutral or alkaline pH. (Rasool *et.al.,* 2020)

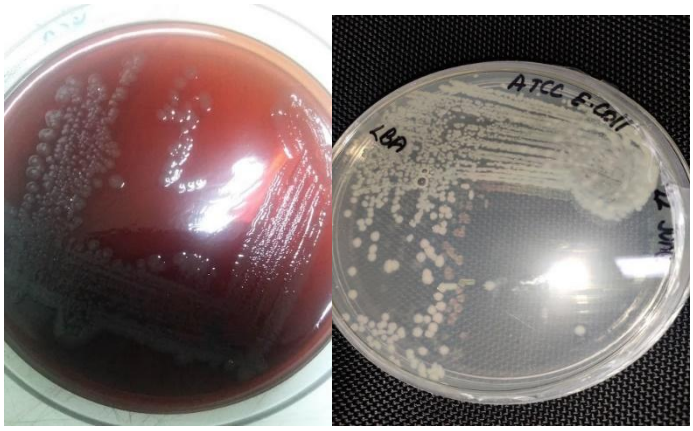


Fig 4.1 Isolation and revival of Salmonella on XLD Agar (left) and ATCC Escherichia coli on LB Agar (right)

4.1.1 Gram Staining and Biochemical Test of Pathogen

Gram staining and biochemical tests are considered to be preliminary tests for the identification and confirmation of bacteria. The results of these tests (Table 4.1) demonstrate Gram stain and various biochemical tests for the respective pathogens and in accordance to be reported for the respective pathogens.

Table 4.1: Gram stain and biochemical tests for identification of bacterial strain

S.N	Test	Organism	
		Salmonella	ATCC *E. coli*
	Gram Staining	Gram negative rod	Gram negative rod
1	Indole	Negative	Positive
2	MR	Positive	Positive
3	VP	Negative	Negative
4	Citrate Utilization	Positive	Negative
5	TSIA	Alkaline/acid	Acid/acid
6	O/F	Fermentative	Fermentative

4.1.2 Antibiotic Susceptibility Testing of Pathogen

The preliminarily confirmed bacterial strains were tested for their antibiotics sensitivity against different classes of antibiotics. The isolated *Salmonella* (LT2) was found to be resistant against Etrapenem, Meropenem, Ciprofloxacin, Colistin and Gentamicin with 7mm, 2mm, 3mm, 0 mm and 4mm zone of inhibition respectively.

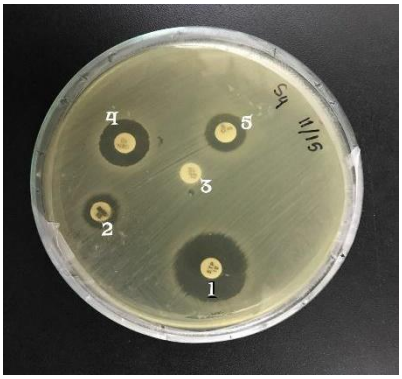


Fig 4.2 Antibiotic Susceptibility Testing of Salmonella (LT2);1- Ertapenem,2-meropenem, 3-Colistin, 4- Gentamicin, 5- Ciprofloxacin

Since the approach was to use the MDR pathogen against the extract the pathogen was taken for further study.

4.2 *Streptomyces*

The *Streptomyces* isolated and screened by the senior (Sita Ghimire) in her thesis for was revived in ISP2. ISP4 and SCA medium by subculturing. The ISP2 media with the soluble starch as sole source of reduced carbon aided the growth of *Streptomyces* (Zainal *et. al.,* 2016) and condemned the growth of bacteria that rely on simple sugars (Wawrik *et. al.,* 2005). This might be the reason that ISP2 encompassed an array of contaminants as the media contained glucose as a carbon source.

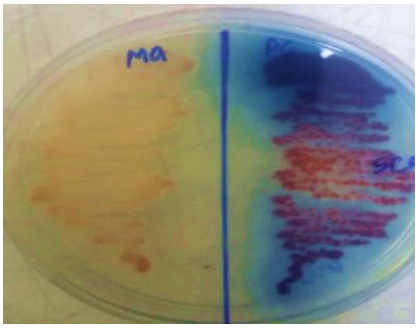


Fig 4.3 Subculture and revival of Ma and Pc strain of Streptomyces in SCA medium

4.2.1 Secondary metabolite Production

For the extraction of secondary metabolites, the revived strains of *Streptomyces* were incubated in ISP4 broths at 300rpm/28°C for 7 days. Apart from the media used, various manipulation of the media, especially carbon source can be done to nullify the negative carbon catabolite effects hindering the production of bioactive secondary metabolites (Ruiz *et. al.,* 2010). The crude extracts were labelled and stored at 4°C for further use. Bioassay was performed against test organisms by the resazurin assay method.

4.2.2 Resazurin test Assay

The Kirby Bauer diffusion method of susceptibility testing against pathogens takes a significant amount of time and labour, and thus the Resazurin test acts as a potent alternative. Resazurin (7-Hydroxy-3*H*-phenoxazin-3-one 10-oxide) is a blue non-fluorescent dye that becomes pink and fluorescent when reduced to resorufin by oxidoreductases present within viable cells. Resorufin is further reduced to hydro resorufin (colorless and nonfluorescent). It is used as an oxidation-reduction indicator in cell viability assays for both aerobic and anaerobic respiration (Chen *et. al.,* 2015).

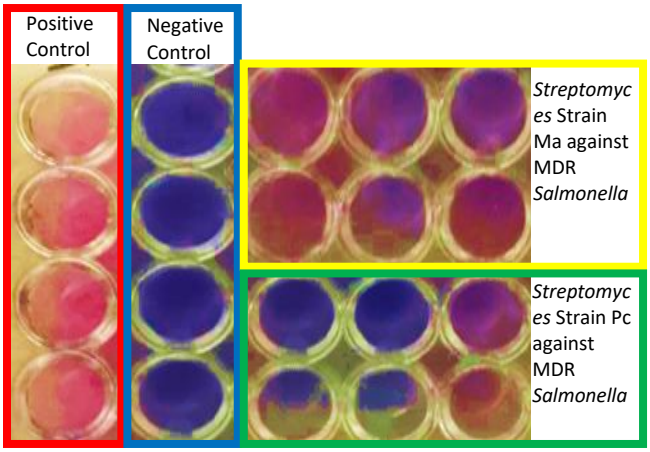


Fig 4.4 : Resazurin Assay of Streptomyces against Salmonella (LT2).

Ma strain of *Streptomyces* was found to have higher efficacy than the Pc strain.

In-silico Assessment

4.3 Search for Bacterial Protein Targets

A number of pathogens are becoming more MDR-resistant, and MDR bugs are emerging that are both carbapenem- and colistin-resistant (Ah *et. al.,* 2014; Potter *et. al.,* 2016) one sample was found in a Nepalese poultry sample, ongoing research. This has rendered all antibiotics previously used to treat these pathogens ineffective. Faster identification of new therapeutic targets and lead compounds directed against them is necessary. In order to produce a broad-spectrum therapy, lead compounds with such protein targets that are difficult for the pathogen to modify are extensively sought after (Bolognesi, 2013).

In addition, the plan must be created to stop the rapid emergence of resistance to these novel medications. It may be difficult for infections to evolve resistance to numerous targets at once, and it is likely impossible for bacteria to live against such developed medications. As a result, genome level sequence alignment of important pathogens may reveal new, shared lead targets for this enzyme that can be used to screen for lead inhibitor compounds. Based on databases of protein-protein interactions, the relevant metabolic pathway or alternative target proteins might then be found.

4.4 Drug Target Protein Selection

TrmD, an essential protein unique to its eukaryotic and archaeal counterpart Trm5, is an S-adenosylmethionine (AdoMet)-dependent methyltransferase synthesizing the methylated m1G37 in tRNA. The efflux activity expels the antibiotics giving rise to bacterial multi drug resistance. TrmD being a major protein responsible for the biosynthesis of membrane protein can be a drug target as the depletion of *trm*D and consequently m1G37-tRNA accumulates ribosomal frameshifts and leads to cell death by resulting the termination of protein synthesis and sensitization of the bacteria to antibiotics (Gamper *et. al.,* 2015a; Hou *et. al.,* 2017; Masuda *et. al.,* 2019).

4.5 Protein Preparation, Validation and Processing

Since the crystal structure of the *trm*D of *Salmonella Typhimurium* (LT2) was not available in RCSB PDB and SCOP databases, the protein structure was prepared using various homology modelling tools and then validated before processing.

4.5.1 Sequence Retrieval and Alignment

The FASTA sequence of *trm*D of *Salmonella Typhimurium* LT2 (NP_461604.1) was retrieved from NCBI. To find out the similar protein with highest percent similarity the sequence was subjected to BLAST (Basic Local Alignment Search Tool). The highest similarity (96.47%) was found with the *Escherichia coli* (K- 12).

*Table 4.2 p-BLAST results for trm*D *of Salmonella Typhimurium in reference to trm*D *Escherichia coli K-12 an*D *trm*D *Haemophilus influenzae R*D*-KW20*

Subject sequence	**Max score**	**Total score**	**Query coverage**	**E-value**	**Percentage identity**
*trm*D *Escherichia coli* K-12	510	510	100	0.0	96.47
*trm*D *Haemophilus influenzae* RD-KW20	407	407	95	2e-145	82.36

4.5.2 Allignment of *trm*D gene of WHO prioritized Pathogens

Allignment of the *trm*D gene of eight WHO prioritized pathogens *Escherichia coli* (UNIPROT: P0A873), *Salmonella Typhimurium* (UNIPROT: P36245), *Klebsiella pneumoniae* (UNIPROT: A6TCL5), *Serratia marcescens* (UNIPROT: P36244), *Proteus mirabilis* (UNIPROT: B4EUW6), *Neisseria gonorrhea* (UNIPROT: B4RIW3), *Acinetobacter baumanni* (UNIPROT: B7IAT0), *Pseu*D*omonas aeruginosa* (UNIPROT: V6AAA4) was done by clustal omega to identify the similar amino acid sequences. This is critical since the similarity in active binding site of a ligand suggests an idea about a same drug being viable for all of the *trm*D gene. All of them were found to have similarity in the active site region 132–141, which is usually conserved in case if *trm*D (Elkins *et. al.,* 2003)

```
CLUSTAL O(1.2.4) multiple sequence alignment


sp|B7IAT0|TRMD_ACIB5        ----MFFAVITLFPEMFDAITAYGISGRAAKRDIVQVTCINPRDFAEGNYRRVDERPFGG        56
sp|B4RIW3|TRMD_NEIG2        ----MLIQAVTIFPEMFDSITRYGVTGRANRQGIWQFEAVNPRKFADNRLGYIDDRPFGG        56
tr|V6AAA4|V6AAA4_PSEAI      MDKRLWVGVVSIFPEMFRAISEYGITSRAVKQGLLTLTCWNPRDYTEDRHQTVDDRPFGG        60
sp|B4EUW6|TRMD_PROMH        ----MWIGVISLFPEMFRAITEYGVTGRAVKHGLLNVECWDPRDFTYDRHHTVDDRPYGG        56
sp|P36244|TRMD_SERMA        ----MFIGIVSLFPEMFRAITDYGVTGRAVKNGLLSVQCWSPRDFTYDRHRTVDDRPYGG        56
sp|P36245|TRMD_SALTY        ----MFIGIVSLFPEMFRAITDYGVTGRAVKKGLLNIQSWSPRDFAHDRHRTVDDRPYGG        56
sp|P0A873|TRMD_ECOLI        ----MWIGIISLFPEMFRAITDYGVTGRAVKNGLLSIQSWSPRDFTHDRHRTVDDRPYGG        56
sp|A6TCL5|TRMD_KLEP7        ----MWIGIISLFPEMFRAITDYGVTGRAVKNGLLSIESWSPRDFTHDRHRTVDDRPYGG        56
                               : .  :::***** :*: **::.** :..:  . . .**.:: ..   :*:**:**

sp|B7IAT0|TRMD_ACIB5        GPGMVMMAEPLAKAINHAKQLASRAGCVHVPVVYMSPQGKTLNEQAVQQFVDYDGLIVLC        116
sp|B4RIW3|TRMD_NEIG2        GPGMIMMAPPLHAAIEHAKA-QSSQ---TAKVIYLSPQGKPLTHQKAAELAELTHLILLC        112
tr|V6AAA4|V6AAA4_PSEAI      GPGMVMKIKPLEGALADARQAAGGR---KAKVIYLSPQGRQLTQAGVRELAEEEALILIA        117
sp|B4EUW6|TRMD_PROMH        GPGMLMMVQPLREAIHQAKA-AAGD---GAKVIYLSPQGRKLDQQGVCELATNEKLILVC        112
sp|P36244|TRMD_SERMA        GPGMLMMVQPLREAIHAAKA-AAGE---GAKVIYLSPQGRKLDHTGVCELAANQKMILVC        112
sp|P36245|TRMD_SALTY        GPGMLMMVQPLRDAIHAAKA-AAGE---GAKVIYLSPQGRKLDQAGVSELATNQKLILVC        112
sp|P0A873|TRMD_ECOLI        GPGMLMMVQPLRDAIHAAKA-AAGE---GAKVIYLSPQGRKLDQAGVSELATNQKLILVC        112
sp|A6TCL5|TRMD_KLEP7        GPGMLMMVQPLRDAIHAAKA-AAGE---GAKVIYLSPQGRKLDQAGVSELATNQKLILVC        112
                            ****:*   **  *:  *:   .     . *:*:****: * .  . ::.    :*::.

sp|B7IAT0|TRMD_ACIB5        GRYEGVDERLIQHYVDQEWSIGDYVLSGGELPAMVLLDSIIRRLPNVMSDEQSAIQDSFV        176
sp|B4RIW3|TRMD_NEIG2        GRYEGIDERLLQSSVDEEISIGDFVVSGGELPAMMLMDAVLRLVPGILGDIQSAEQDSFS        172
tr|V6AAA4|V6AAA4_PSEAI      GRYEGIDERFIEEHVDEEWSIGDYVLSGGELPAMVLVDAVTRLLPGALGHADSAEEDSFT        177
sp|B4EUW6|TRMD_PROMH        GRYEGIDERVIQTEIDEEWSVGDYVLSGGELPAMIMIDAVARFVPGVLGHAASAKEDSFA        172
sp|P36244|TRMD_SERMA        GRYEGIDERVIQTEIDEEWSIGDYVLSGGELPAMTLIDSVARFIPGVLGHQASAEEDSFA        172
sp|P36245|TRMD_SALTY        GRYEGVDERVIQTEIDEEWSIGDYVLSGGELPAMTLIDSVARFIPGVLGHEASAIEDSFA        172
sp|P0A873|TRMD_ECOLI        GRYEGIDERVIQTEIDEEWSIGDYVLSGGELPAMTLIDSVSRFIPGVLGHEASATEDSFA        172
sp|A6TCL5|TRMD_KLEP7        GRYEGIDERVIQTEIDEEWSIGDYVLSGGELPAMTLIDSVSRFIPGVLGHEASATEDSFA        172
                            *****:***.::  :*:* *:**:*:******** ::*:: * :*. :..  ** :***

sp|B7IAT0|TRMD_ACIB5        DGLLDCPQYTKPDQFEGLDVPEILKSGHHANIEKWRFLQRYQRTLERRPELIEQVTLTKQ        236
sp|B4RIW3|TRMD_NEIG2        SGILDCPHYTKPLEFQGMAVPEVLRSGNHGLIAEWRLEQSLRRTLERRPDLLEKRVLIPK        232
tr|V6AAA4|V6AAA4_PSEAI      DGLLDCPHYTRPEVYADKRVPEVLLSGNHEHIRRWRLQQALGRTWERRADLLDSRSLSGE        237
sp|B4EUW6|TRMD_PROMH        EGLLDHPHYTRPEVLDGMAVPAVLLSGNHAHINRWRMKQSLGRTWLRRPELLESLALTDE        232
sp|P36244|TRMD_SERMA        DGLLDCPHYTRPEVLEGMEVPPVLLSGNHAEIRRWRLKQSLGRTWLRRPELLESLALTDE        232
sp|P36245|TRMD_SALTY        DGLLDCPHYTRPEVLEGMEVPPVLLSGNHAEIRRWRLKQSLGRTWLRRPELLENLALTEE        232
sp|P0A873|TRMD_ECOLI        EGLLDCPHYTRPEVLEGMEVPPVLLSGNHAEIRRWRLKQSLGRTWLRRPELLENLALTEE        232
sp|A6TCL5|TRMD_KLEP7        DGLLDCPHYTRPEVLEEMEVPPVLLSGNHAEIRRWRLKQSLGRTWLRRPELLENLALTEE        232
                            .*:** *:**:*       ** :* **:*  * .**: *   **  ** :*::.  *  :

sp|B7IAT0|TRMD_ACIB5        QKKWLSDEQG-------------     246
sp|B4RIW3|TRMD_NEIG2        ESRLLNKILQEQREIQS------     249
tr|V6AAA4|V6AAA4_PSEAI      EQKLLAEYIRQRDDS--------     252
sp|B4EUW6|TRMD_PROMH        QRVLLAEFQQEHLQRTAE-----     250
sp|P36244|TRMD_SERMA        QAVLLAEFQREHQARQQDYEGNV     255
sp|P36245|TRMD_SALTY        QARLLAEFKTEHAQQQHKHDGMA     255
sp|P0A873|TRMD_ECOLI        QARLLAEFKTEHAQQQHKHDGMA     255
sp|A6TCL5|TRMD_KLEP7        QAKLLSEFKTEHAQQQHKHDGQA     255
                            :   * .
```

*Fig 4.5 : Multiple sequence allignment of Trm*D *Protein of WHO Prioritise*D *pathogen*

A phylogenetic tree cladogram was also created using clustal omega tool. It depicted the close relationship between the pathogens, it seems from the tree below that *trm*D of *Klebsiella* is closer to *E.coli* than *Salmonella* .

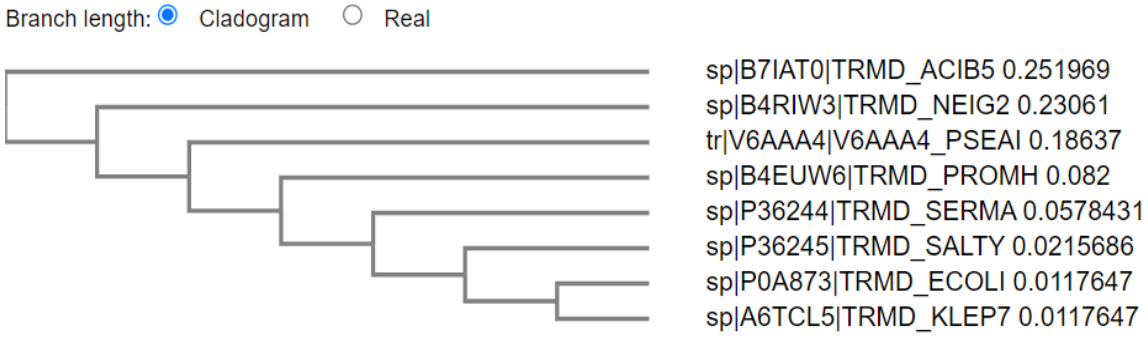


*Fig 4.6: Phylogenetic tree of trm*D *protein of WHO prioritize*d *pathogens*

4.5.3 Protein Structure Preparation using various Homology Modeling Tools

The crystal structure of target protein *trm*D *Salmonella Typhimurium* LT2 was not available in RCSB PDB and SCOP databases. Thus, its primary AA sequence was retrieved from NCBI and were used to predict their tertiary structures using various web-based tools including MODELLER (https://salilab.org/modeller/) CPH model(https://services.healthtech.dtu.dk/service.php?CPHmodels-3.2), Phyre2(http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index , (PS)2v2 (http://ps2.life.nctu.edu.tw/) .

4.5.3.1 MODELLER

A computer application called Modeller, which is frequently abbreviated as MODELLER, is used for homology modeling to create models of the tertiary and quaternary structures of proteins that are rare. It uses a technique known as fulfilment of spatial restrictions, which was inspired by protein nuclear magnetic resonance spectroscopy (protein NMR), in which a set of geometrical requirements are utilized to generate a probability density

function for each atom's placement in the protein. The approach is dependent on an input sequence alignment between the target amino acid sequence to be modeled and a template protein whose structure has been determined. (Fiser and Sali, 2003)

Additionally, the program includes restricted functions for the ab initio structure prediction of protein loop sections, which are notoriously difficult to predict using homology modeling since they frequently exhibit significant levels of variability even across similar proteins. (Kuntal *et.al.,* 2010). In our study, MODELLER gave five different structures of *trm*D (*trm*D.B901, *trm*D.B902, *trm*D.B903, *trm*D.B904, *trm*D.B905). The best one out of 5 was selected after the protein processing and validation.

It should be emphasized that a model's needed quality primarily depends on the application for which it is intended. For instance, developing mutagenesis experiments can entirely rely on low-accuracy models, whereas structure-based virtual screening (SBVS) applications demand more accuracy. (Berman *et. al.,* 2006).

*Fig 4.7: Structure of trm*D *Salmonella Typhimurium LT2(model trm*D*.B905) using modellar 9.2 taken from PyMol*

4.5.4 Protein processing and validation

4.5.4.1 Z- score and energy plot analysis

Homology models, which are computationally produced approximations of a protein structure, may be seriously flawed. Before beginning any type of computational

experiments involving drug development, it is essential to evaluate the precision and dependability of theoretical and experimental models of protein structures (Berman *et. al.,* 2006).

The efficiency of the findings from subsequent studies is determined by the accuracy of the predicted structures, which is a crucial component of successful molecular docking procedures for which Z-score was thought to be a standard parameter of quality assessment (Wiederstein and Sippl, 2007). The quality of the predicted 3D-protein structures was assessed using ProSA-server (https://prosa.services.came.sbg.ac.at/prosa.php) . The Z-score of the input structure must fall within the range of scores typically attained by natural proteins of comparable size, as shown by ProSA. ProSA-web is an interactive web service for finding flaws in 3D theoretical and experimental models of proteins (Wiederstein and Sippl, 2007). The first plot in the picture (figure a), which plots energies as a function of amino acid sequence location, illustrates the local model quality. Positive values typically refer to aspects of the input structure that are incorrect or faulty. Similar to this, the second plot (Figure b) distinguishes structures from other sources (X-ray, NMR) using various color codes. In order to determine whether the Z-score of the query structure falls within the average scoring range for native proteins of a similar size.

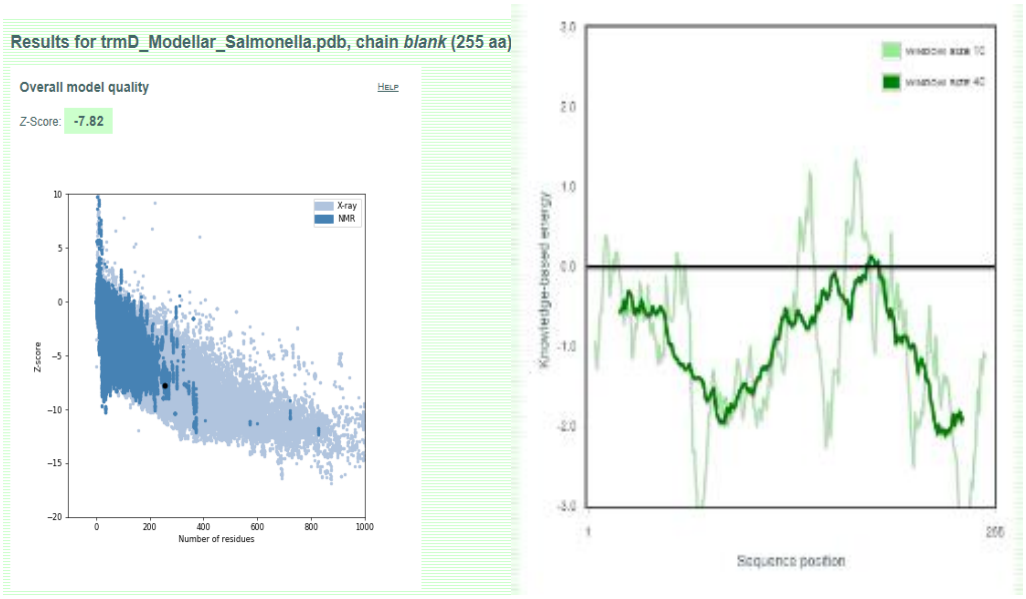


Fig 4.8: The Z score plots and energy plot for predicting 3D structure of a protein based on amino acids sequences.

a) Z-Score plot of model predicted (*trm*D.B905) by MODELLER with structures available in database.
b) Energy plot for amino acid residues for the 3D-structure of *trm*D.B905 of *Salmonella Typhimurium* LT2 predicted from MODELLER.

The structure predicted by MODELLER (*trm*D.B905) had the lowest Z-score (-7.82) and it also modelled the most complete structure of *trm*D.

Table 4.3 Z score using various homology modelling tools

Name of server	**Z score**	**AA no.**
MODELLAR	-7.82	255
CPH model	-7.79	246
Ps2v2	-7.03	255
Phyre2	-6.49	239

4.5.4.2 Ramachandran Plot Analysis

Correct stereochemistry is a protein model's primary prerequisite. Programs like PROCHECK (Laskowski, 1993) and WHATCHECK are used to examine anomalies such phi/psi angle combinations that are positioned in forbidden zones, steric collisions, and unfavorable bond lengths and angles (Hooft *et. al.,* 1996). These programs also examine the stereochemical characteristics of the model's residues and provide an assessment of the model or structure as a whole (Bhattacharya *et. al.,* 2007). It's crucial to draw attention to the model's implausible conformations by analyzing bond geometry through the use of Ramachandran plots. In protein structures, some phi and psi angle conformations are prohibited because they produce steric hindrance, or collisions between atoms. Typically, a good model will have 90% of its residues in the permitted areas of a Ramachandran plot (Laskowski, 1993).

More than 90% of the residues in the protein crystal structure under evaluation were in acceptable regions, meaning that only a small number of phi/psi angle combinations were present in the prohibited zone, as seen in the figure below. Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favored regions.

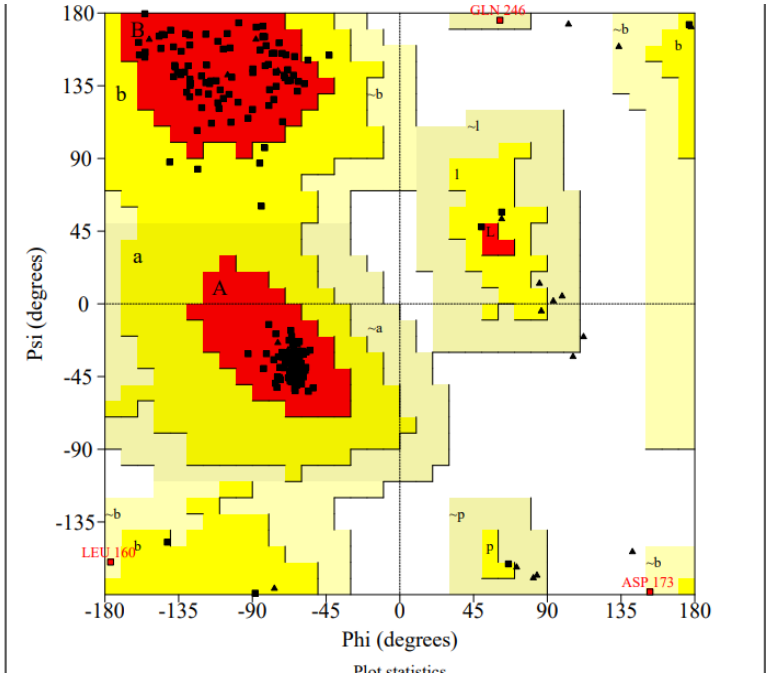


Fig 4.9: Ramachandran plot analysis using SAVES V 5.0 for amino acid residues of trmD.B905 (Developed using MODELLER)

4.5.4.3 Selection of protein based on validation tools

All of the models that were developed using homology modelling tools were analysed for Z- score, Energy plot, Ramachandran plot. The most favoured structure was found to be the one developed from MODELLER (*trm*D.B905). In the table below (table 4.4), the illustration of protein validation done by using Ramachandran plot is shown. Out of the 4 tools, MODELLER and Ps2v2 are the ones with none amino acids lying in the disallowed region. Even though, the model generated from Ps2v2 has higher percentage (93.10%) of amino acids in the most favoured region in comparison to MODELLER (92.60%), it (Ps2v2) still could not be considered to be generating a good model as it had the Z-score (-7.03) and energy plot more than 0, in general, positive values correspond to problematic or erroneous parts of a model (Wiederstein and Sippl, 2007), in comparison to the model of MODELLER with the Z-score 0f -7.82 and the energy score equal to 0.

Table 4.4 Table showing the Ramachandran Analysis of various trmD proteins that were created using homology modelling tools

Tool	Ramachandran Analysis (SAVES v5.0)				Z-Score (ProSA)	Energy Plot	AA no
	Most Favoured Region	Allowed Region	Generously allowed region	Disallowed region			
MODELLER	92.60%	6.00%	1.40%	0.00%	-7.82	=0	255
CPH Model	86.50%	11.50%	1.40%	0.50%	-7.79	=0	246
Ps2v2	93.10%	6.50%	0.50%	0.00%	-7.03	>0	255
Phyre2	92.50%	7.00%	0.50%	0.00%	-6.49	>0	239

MODELLER modelled 5 models, which were also analysed for the best one to be selected. One additional validation tool DOPE score (given by MODELLER) was also added to analyse the 5 models. DOPE (Discrete optimized protein energy) is a statistical potential used to assess homology models in protein structure prediction. DOPE is based

on an improved reference state that corresponds to non-interacting atoms in a homogeneous sphere with the radius dependent on a sample native structure, it thus accounts for the finite and spherical shape of the native structures. (Fiser and Sali, 2003) In the given table (table 4.5), the model with the highest DOPE score (-24986.94922), most negative Z-score (-7.82), energy plot that equaled to 0 and none amino acids in the disallowed region was found to be model *trm*D.B905. Thus, this model was taken forward for molecular docking with ligands.

Table 4.5 Table showing protein validation using various scores for the models of trmD created using MODELLER

Model name	DOPE Score (MODELLER)	Z-Score (ProSA)	Energy Plot (ProSA)	Ramachandran Analysis (SAVES v5.0) Disallowed residues
*trm*D.B901	-24954.08203	-7.24	>0	0.50%
*trm*D.B902	-24513.67773	-7.10	>0	0.90%
*trm*D.B903	-24885.43164	-7.16	>0	0.00%
*trm*D.B904	-24847.59961	-7.14	>0	0.90%
*trm*D.B905	-24986.94922	-7.82	=0	0.00%

4.5.4.4 Superimpose of the model protein with reference protein

The model of protein *trm*D *Salmonella Typhimurium* (LT2) that was developed using MODELLER (*trm*D. B905) and that of *Escherichia coli* (K12) (PDB:1P9P) were structurally aligned using Chimera software version 1.14. employing sequence and align function inbuilt in the software was superimposed to found out the RMSD value to be 0.631 and Q-score 0.456. This suggests that the protein developed using MODELLER is highly similar to the

reference protein 1P9P. Similarly, the sequence similarity was found to be 96% between both the proteins.

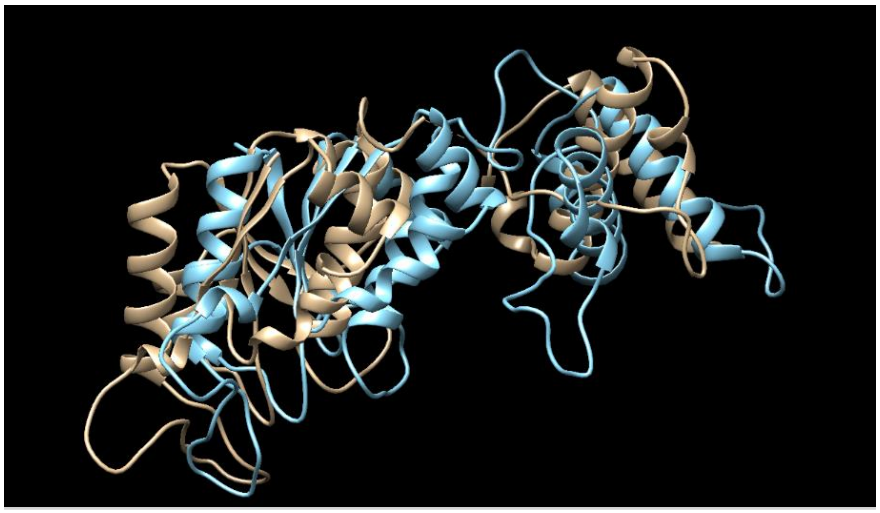


Figure 4.10: Structural superposition between MODELLER (trmD. B905) (tan) and that of Escherichia coli (K12) (PDB:1P9P) (cyan) using UCSF chimera.

4.6 Molecular docking

4.6.1 Protein Preparation

The target protein structure (*trm*D.B905) that was obtained from our study using homology modeling tool (MODELLER) was used for docking simulation. It is necessary to convert the PDB format into the Pdbqt format before to running docking simulations in the PyRx platform using the Autodock Vina tool. The hydrogen atoms must be added using the same method that was used to remove the water molecules from the crystal structure that was deposited at the RCSB facility. Deleting the water molecules further simplifies computing and helps empty the potential binding pocket of the water molecules that could have affected the pose search due to their loosely bound state and frequent absence from binding. Having stated that, it should be conserved that any water molecules revealed to be engaged in potential binding interactions with proteins should be preserved (Wong *et. al.,* 2010). The balance between the enthalpy of the water molecule's interactions and the entropy change brought on by transferring the water to

the bulk solvent determines whether displacing a certain structural water would increase binding (Dunitz *et. al.,* 1994).

Since no such water molecule with crucial binding interactions was discovered in our investigation, all of the water molecules were eliminated during the protein-making process. The ability of a functional group to generate hydrogen bonds relies on the position of its hydrogen atoms and is crucial for the stability of protein-ligand complexes (Lippert *et. al.,* 2009). Therefore, it is crucial to recognize hydrogen bonds and their characteristics, which can be accomplished by having a precise understanding of the locations of hydrogen atoms in proteins. The high mobility of hydrogen atoms introduces a number of degrees of freedom, including torsional changes where the position of the hydrogen atom is rotated around the final heavy-atom bond in a residue, protonation states where the number of hydrogen atoms at a functional group may change, and tautomeric states where a hydrogen atom is capable of changing its binding partner (Lippert *et. al.,* 2010). Before doing structure-based computations, side-chain flips in glutamine, asparagine, and histidine residues—common crystallographic ambiguities—must also be discovered. Additionally, before conducting any pharmaceutical research, the hydrogen atoms must be added because the experimental method used to determine the majority of the publicly available structures in the Protein Data Bank (PDB) may have left ambiguities in the protein structures. This is particularly true if the experimental method used was X-ray crystallography (Berman et. al, 2000).

The merging of non-polar bonds if any present were also conducted. The docking conformation and the energy score of the resulting complex depends on the calculation of accurate partial charges on the protein and ligand both, possibly leading to more accurate estimation of complex geometry and binding energy (Tsai *et. al.,* 2008). There are several charge calculation methods which lead to significant differences in the partial charges assigned to the different atoms (Bikadi et.al., 2009). The Gasteiger charges are determined on the basis of electronegativity equilibration and is the measure of partial charges present in an protein tertiary structure (https://www.researchgate.net/post/What_is_the_differences_between_Kollman_and_

Gasteiger charges) . Gaisteiger charges computed for our particular protein structure which was found to be -7.9933. Any docking simulation experiments begin with the conversion of the particular protein's tertiary structure into the dockable pdqt format.

4.6.2 Ligand database preparation

Early-stage drug discovery focuses on improving the pharmacological characteristics of ligands (possible leads against biological targets) and the potency of these compounds. This can be done by identifying lead compounds that have pharmacological action against a biological target (Lionta el. al., 2014). There are various classes of compounds, both natural and synthetic, with various chemical and structural properties that can be used as drugs against any disease or viral infection. The choice of ligand database is always a crucial and determining factor for any kind of appropriate drug discovery against a chosen target (Glaab *et. al.,* 2016).

The ligand library may contain already known drug substances for repositioning, synthetic compounds that are similar to lead or drug compounds for subsequent structural optimization, or other natural or xenobiotic compounds, depending on the goal and type of the study (e.g., drug development, toxin identification, pesticide development) (Schames *et. al.,* 2004).

4.6.2.1 Natural Products

Since people are known to eat large amounts of natural goods derived from plants, natural products from the ZINC database were chosen as the ligands because the chemicals present in natural products would have a higher possibility of passing the human trial. According to "http://zinc.docking.org/browse/catalogs/natural-products," a third of all pharmaceuticals are natural goods or ones that are comparable to them. Natural products' potential use against newly emerging pathogens, which are less likely to fail clinical trials if relevant against the target pathogens, could be investigated. From the database of natural compounds containing zinc, 3888 molecules were chosen to serve as ligands. From the ZINC database (http://zinc15.docking.org), the ligands were downloaded.

4.6.2.2 Kinase Inhibitors and Indole Derivatives

As SAM has an adenosyl moiety and kinases have an ATP binding domain, these proteins that biosynthesize or use SAM may be susceptible to inhibition by kinase inhibitors (Villamor *et. al.,* 2013). In particular as anticancer medications, protein kinase inhibitors are a significant class of targeted therapeutic medicines (Grant, 2009). According to Gillam *et. al.,*2000, human cytochrome P450 2A6 is known to metabolize indole, and the source of indole may be the gut microflora's metabolism of tryptophan. This suggests that indoles are digested in humans, demonstrating that they cannot be harmful, and that they can be efficiently transported in humans through the gut (Banoglu *et. al.,* 2001). Additionally, indole has been proposed as a pharmacological framework for drug development (Suenkel *et. al.,* 2013). A number of indole derivatives have also been shown to be kinase inhibitors, including its metabolite Indirubin (Wu *et. al.,* 2005) and certain indole derivatives (Kiliç *et. al.,* 2009; lgen *et. al.,* 2011). Indole derivatives are likely candidates to block the SAM binding site of the *trm*D protein because of the near structural proximity of the indole ring to the adenosyl moiety of SAM. Indole derivatives were therefore thought to be potential sources of ligands for virtual screening.

4.6.2.3 Nucleoside Mimetics

Both Gram-positive and Gram-negative bacterial pathogens have highly conserved tRNA-(N1 G37) methyltransferase (*Trm*D), which is necessary for proliferation. *Trm*D is also considerably different from its human orthologue TRM5, making it a good target for the development of brand-new antibiotics. *Haemophilus influenzae Trm*D was utilized to screen a collection of chemical fragments, and it revealed inhibitory, fused thieno-pyrimidones that were competitive with the physiological methyl donor substrate S-adenosylmethionine (SAM). (Hill et.al., 2013) in their study fused thieno-pyrimidones were identified as competitive inhibitors of SAM binding, with nanomolar binding affinity and a lack of activity against human homologs of *Trm*D. Therefore, nucleoside mimetics were also thought to be potential source of ligands.

4.6.3 Identification of Active Binding Site

3DLigandSite (http://www.sbg.bio.ic.ac.uk/3dligandsite/advanced.cgi) , a web server that superimposes the ligands bound to the structures similar to the query and thus predicts the binding site (Wass *et. al.,* 2010), was used to predict the ligand-binding sites for the predicted 3D-structures of *trm*D of *Salmonella Typhimurium* for which crystal structures have not yet been elucidated.

To design a location for molecular docking, all of the AA residues close to the binding site are highlighted as indicated.

Table 4.6: Predicted binding sites and heterogens present as predicteD by 3DligandSite in TrmD of Salmonella Typhimurium predicted via Phyre2 webserver

Heterogens present in Predicted Binding Site

Heterogen	**Count**	**Source structures**
SAM	5	1uak_A,2yy8_B,2v3k_A,1v2x_A,2qmm_B
SAH	5	1uam_A,1ual_A,1p9p_A,1mxi_A,2ha8_B

Residue	Aminoacid	Contact	Average Distance
86	TYR	9	0.26
87	LEU	10	0.00
88	SER	10	0.03
89	PRO	6	0.18
90	GLN	4	0.00
111	VAL	3	0.07
112	CYS	6	0.32
113	GLY	10	0.00
114	ARG	10	0.24
115	TYR	9	0.02
116	GLU	9	0.07
117	GLY	8	0.27

131	TRP	10	0.21
132	SER	10	0.11
133	ILE	10	0.01
134	GLY	10	0.01
136	TYR	9	0.07
137	VAL	7	0.36
138	LEU	10	0.00
139	SER	10	0.08
140	GLY	10	0.00
141	GLY	10	0.00
142	GLU	5	0.30
144	PRO	10	0.02

The protein sequence of *trm*D *Escherichia coli* (UNIPROT: POA873) was aligned with the protein sequence of *trm*D of *Salmonella Typhimurium* (UNIPROT: P36245). The alignment showed that both the sequence shared maximum similarity in the amino acid residues. This illucidated that the active site of *trm*D *Escherichia coli* (RCSB:1P9P) can also be used as an active site for *trm*D of *Salmonella Typhimurium*.

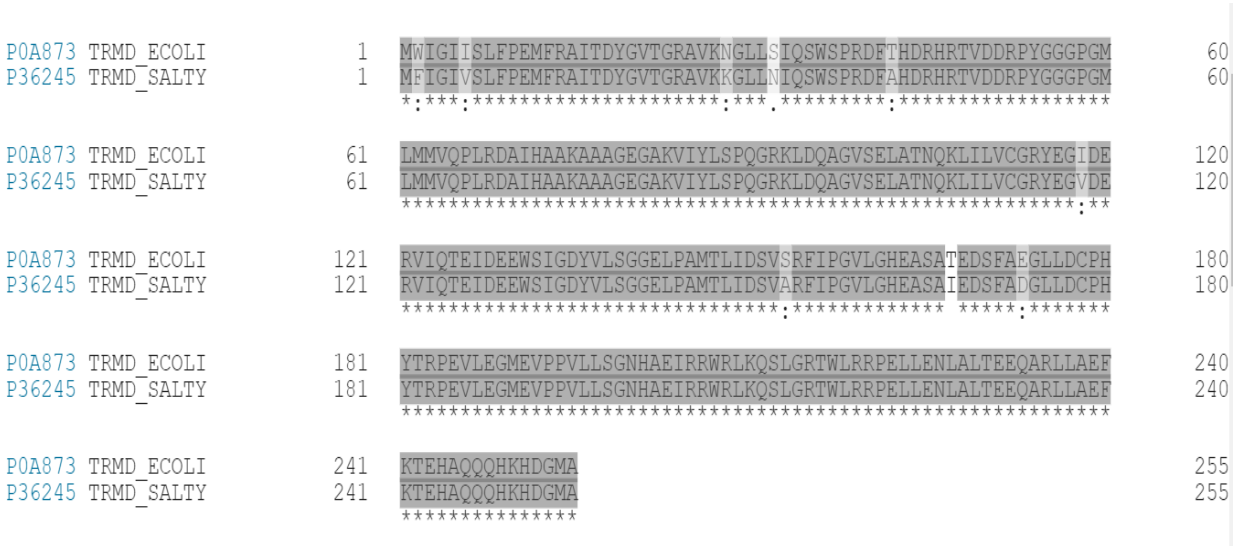


Fig 4.11: Amino acid alignment of trmD of Salmonella Typhimurium (LT2) and Escherichia coli(K12) (Uniprot/kb)

4.6.4 In-silico ADME/Tox tests for possible hits

It was now possible to use all of the ligands with negative binding energies that were larger than those of the reference inhibitor in additional molecular docking simulations. However, ADME/tox characteristics, which are often tested at the end of the drug discovery pipeline, are significant because they are responsible for the failure of 60% of therapeutic compounds during the drug development process(https://www.sciencedirect.com/topics/pharmacology-toxicologyandpharmaceuticalscience/adme#:~:text=6.7%20ADME%2DTox%20Prediction&text=ADMET%20properties%20play%20an%20important,during%20the%20drug%20development%20process.&text=Such%20drugs%20may%20show%20poor,might%20be%20toxic%20in%20nature). Therefore, it could be a good idea to execute ADMET in the first step. Leads obtained from research are typically unsuccessful in clinical trials due to their inability to bind to the target protein and carry out the expected function as well as the toxicity those chemicals impart (Hughes *et. al.,* 2011). As a result, it would appear more prudent to evaluate ADMET and physiochemical properties early in the drug discovery process. Using OSIRIS, the toxic profiles and druglikeness of the pre-screened ligands were investigated.

The selection criteria include the following values for molecular descriptors:

- Mol weight – 200 to 500 Daltons
- cLogP – -3 to 6
- cLogS – greater than -4 i.e. -4 to -2
- Hydrogen bond donors – 0 to 5
- Hydrogen bond acceptors – 0 to 10 (Lipinski, 2004)
- Topological Polar Surface Area (tPSA) – 0 to 120
- Rotatable bonds – 10 or less (Veber *et. al.,* 2002)
- Druglikeness – Positive value

Another vital consideration for drug discovery is the molecule's molecular weight. Drug-like substances, with a few exceptions, typically have a molecular weight of less than 550 daltons and are capable of modulating various biochemical processes to detect, prevent, or treat diseases (Ngo *et. al.,* 2018).

A well-known indicator of a compound's hydrophilicity is its logP value, which is equal to the logarithm of its partition coefficient between n-octanol and waterlog($c_{octanol}/c_{water}$), Poor absorption or penetration is a result of high logP values and low hydrophilicity. It has been established that a compound's clogP similarity should not be more than 5.0 in order for it to have a decent chance of being well absorbed. A drug's lipophilicity, a crucial physicochemical factor that affects its absorption, distribution, metabolism, excretion, and toxicity, can also have negative effects on the body's ability to process and eliminate certain substances (Arnot *et. al.,* 2012). It has been demonstrated that this molecular characteristic is crucial for receptor binding, blood-brain barrier penetration, oral availability, and solubility (Hansch *et. al.,* 1962 & 1987).

The features of a compound's absorption and distribution are greatly influenced by its water solubility. Any medicine that is to be absorbed must be present in solution at the absorption site (Savjani *et. al.,* 2012). The basic goal is to avoid substances with low solubility because they typically have poor absorption. Additionally, one of the key factors in reaching the desired drug concentration in the systemic circulation and the necessary pharmacological response is solubility (Sharma *et. al.,* 2009) Even after oral administration, those medications that are poorly water soluble frequently need high doses to attain therapeutic plasma concentrations (Vemula *et. al.*, 2010).

In order for ligand binding to be specific, hydrogen bonds are crucial (Wade *et. al.*, 1989). Too many hydrogen bond donor / acceptor groups make the process of releasing the hydrogen bond from the water molecules necessary to pass the membrane energetically highly expensive. The number of rotable bond counts measures the molecular flexibility and is in turn associated with the oral bioavailabilty of drugs (Veber *et. al.*, 2002) and most of the drug like compounds have 1-10 rotatable bonds (Khanna *et. al.*, 2010).

The best way to define a molecule's topological polar surface area (TPSA) is to think of it as the surface sum over all polar atoms or molecules, notably oxygen and nitrogen, as well as any hydrogen atoms that are linked to those atoms. It has to do with how well the drug can penetrate cells, and molecules with TPSA greater than 140 angstroms are

regarded as having poor cell penetration and being inferior therapeutics (Pajouhesh *et. al.*, 2005).

On the basis of functional groups found in their structures, the lead molecules' various drug-relevant molecular descriptors, such as molecular weight, cLogP, cLogS, druglikeness, and toxicities like mutagenicity, tumorigenicity, reproductive effects, and irritant effects, can be calculated using OSIRIS (Sander *et. al.,* 2015)

Table: 4.7 Table showing the number of ligands (Natural Products) that were reduced after ADMET screening using OSIRIS

Category of Ligands	**Initial Number**	**Number after ADME/tox**
Biogenic FDA	321	24
Natural Products FDA	136	18
Natural Products in Trial	452	87
Natural Products in World	333	43
Natural Products in Man	1072	58
Natural Products in vivo	1574	131
Total	**3888**	**361**

4.6.5 Virtual Screening

Using AutoDock Vina, docking was done on the Pyrx 0.9.8 platform (https://pyrx.sourceforge.io), which has a user interface that is simple to use. Protein *trm*D of *Salmonella Typhimurium* (LT2) was docked to screen ligands with the number of modes set to 16 and the exhaustiveness set to 32. The S-adenosyl methionine (SAM) showed a binding affinity of -7.9 Kcal/mol after the docking simulations. *Trm*D is in charge of synthesizing m1G37 on tRNA by methylating S-adenosylmethionine (SAM) and the N1 position of the G37 base (Byström *et. al.,* 1982). According to reports, the stronger the binding of the ligand to the target, the higher the negative value of binding affinity

(Dallakyan and Olson, 2015). Thus, for further study, molecules with binding energies bigger in negative value than SAM were taken into consideration. Out of 361 Natural Products that passed ADME/T test, 11 ligands exhibited higher binding energy (BE) than the native SAM towards *trm*D of *Salmonella Typhimurium* (LT2).

Table 4.8: Number of various categories of ligands obtained after docking with trmD Salmonella (LT2)

Types of Ligands	**Binding Energy with SAM**	**Ligand with BE more that with SAM**
Natural Products	-7.9	11
Kinase Inhibitor	-7.9	31
Indole Derivatives	-7.9	111
Nucleoside mimetics	-8.7	99

4.7 Cross reactivity with human proteins: MAT1A as a reference

Studying these top hit compounds' cross reactivity with key human proteins can be a deft tactic to help the drug's effectiveness when used in humans because they were identified by screening from a library of FDA-approved chemicals. S-adenosylmethionine synthetase, also known as MAT (methionine adenosyltransferase), is a crucial human enzyme that produces SAM, the main methyl group donor and precursor for polyamine and glutathione synthesis (Lu *et. al.,* 2008; Mato *et. al.,* 1997). By cleaving the ATP triphosphate chain at both ends, MAT catalyzes the transfer of the adenosyl group from ATP to the sulfur atom of Met (L-methionine) (Markham *et. al.,* 1987). The majority of transmethylation processes occur in adult livers, where the MAT1A gene product is produced, demonstrating the critical role MAT1A plays in preserving healthy human metabolism (Shafqat *et. al.,* 2013)

After filtration with MAT1A, the number of top hits from Natural products were found to be 2 in number. Kinase Inhibitors were filtered down to 5 whereas Nucleoside mimetics and Indole Derivatives were filtered to 2 and 3 in number

Table 4.9: Ligands and their interaction with trmD Salmonella Typhimurium and human MAT 1A protein

Category of Ligand	**Library Obtained from**	**Name of the compound**	**Abbreviations**	**Database ID**	**BE with *Trm*D**	**BE with hMAT 1A**
Native Ligand	Zinc 15 (https://zinc15.docking.org/)	SAM (S-adenosyl methionine)	SAM	ZINC000004214738	-7.9	-7.4
Natural Products	Zinc 15 (https://zinc15.docking.org/)	Scopolamine	NP	ZINC000100229736	-8	-7.3
Kinase Inhibitors	Asinex (https://www.asinex.com/) and Uorsy (https://uorsy.com/)	2-(4-Fluorophenyl)-N-[2-(1H-indol-3-yl)ethyl]acetamide	KI 1	PB26473052	-8.7	-6.5
		N-benzyl-2-(2-methyl-1H-indol-3-yl)acetamide	KI 2	PB339495502	-8.7	-7.3
		1-(1,3-Dimethylpyrazol-4-yl)-3-[(2R)-1-hydroxy-3-phenylpropan-2-yl]urea	KI 3	PB1756362042	-8.9	-7
		1-[(2S)-1-Hydroxy-3-phenylpropan-2-yl]-3-(1-pyridin-4-ylethyl)urea	KI 4	PB1779507632	-8.9	-7.2
		1-(1,5-Dimethylpyrazol-3-yl)-3-[(2S)-1-hydroxy-3-phenylpropan-2-yl]urea	KI 5	PB1780727081	-8.7	-7.2
Nucleoside Mimetics	Asinex (https://www.asinex.com/	7-[(3R)-1-Benzoylpyrrolidin-3-yl]-N-methylthieno[2,3-b]pyrazine-6-carboxamide	NM 1	BDE 32008964	-10.3	-7.8

		[3-(2-Methylpropyl)-1,2-oxazol-5-yl]-[(3S)-3-(1H-pyrazolo[3,4-b]pyridin-6-yl)piperidin-1-yl]methanone	NM 2	BDE 33185743	-9	-7.8
		N-[(3,4-dimethoxyphenyl)methyl]-3-(1H-indol-6-yl)propanamide	ID 1	P095-0542	-8.1	-6.5
Indole Derivatives	Chemdiv (https://www.chemdiv.com/)	N-[2-(6-methoxy-1H-indol-3-yl)ethyl]-3-phenylpropanamide	ID 2	Y020-1432	-8.7	-6
		2-(4-fluorophenyl)-N-[2-(5-methoxy-1H-indol-3-yl)ethyl]acetamide	ID 3	Y031-2293	-8.5	-7.3

4.8 Analysis of docking

4.8.1 Protein ligand interaction

Following the ADMET filter and the human MAT 1A filter, the final leads Natural Products (NP), Kinase Inhibitors (KI), Nucleoside Mimetics (NM), and Indole Derivates (ID) identified from the library were further chosen for examining their interaction with the protein under consideration. The ligand and macromolecule were examined in PyMol after the docking findings to identify the critical amino acid residues involved in ligand binding in the active binding cavity of the crystal structure. For this, the protein-ligand interaction within 5 angstroms of the protein structure's active binding region was taken into consideration.

By examining the type of contact and bond length necessary to stabilize the protein ligand interaction, it is possible to further show how the protein ligand interaction works. The interactions between ligands and the protein's amino acid residues can take many different forms. The bond length and kind of interaction between the protein's amino acid residues and the lead compounds, as determined by molecular docking simulations, are shown in the tables below.

4.8.1.1 Protein Interaction with SAM

Table 4.10: Nature of chemical interaction between SAM and trmd Salmonella Typhimurium

Amino acid residues (within 5 A°)	**Bond length (A°)**	**Interaction type**	**Chemical group of ligand involved**
Tyr 115	2.79	H-bond	Hydrogen of furan ring
Val 118	2.26	H- bond	Oxygen of furan ring
Ser- 139	1.85	H- bond	Hydrogen of furan ring
Leu 138	2.28	H- bond	Hydrogen

Tyr 86	2.86	H-bond	oxygen
Ile 133	1.73		Hydrogen

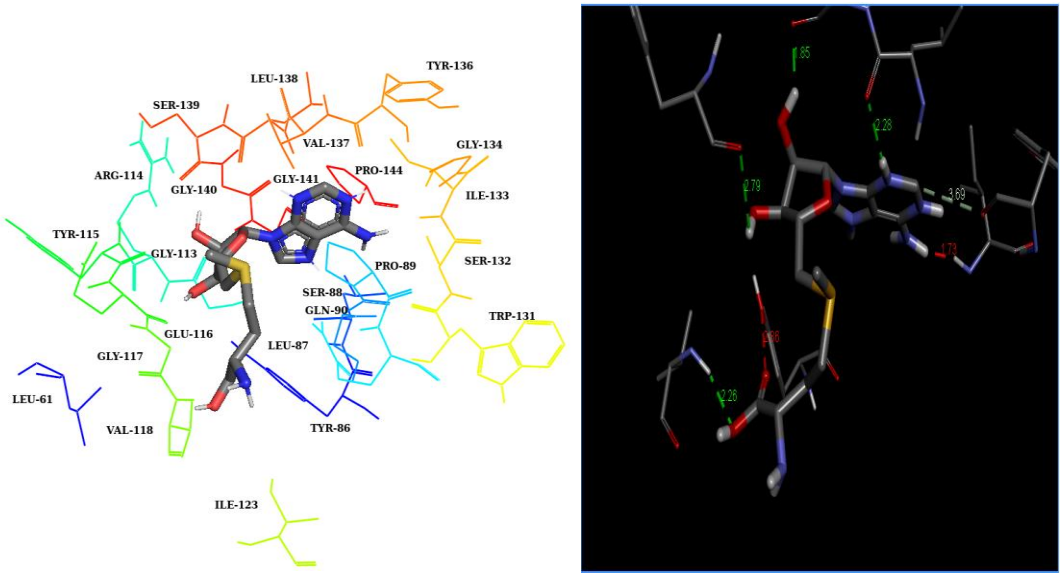


Fig 4.12: Interaction of SAM with active sites of trmD Salmonella visualized using PyMol (left) and bond length (right) visualized using BIOVA- discovery Studio

The junction between the N and C-terminal domains of *trm*D contains the cofactor-binding site, which is hidden at one end of the N-terminal catalytic domain. The pocket contains the adenine moiety of SAM in particular. Residues 132-141 make up the adenine-binding loop, which encircles the adenine ring. (Elkins et.al.,2003). In their investigation of *trm*D in *E. coli* It was discovered that the hydrogen bonds between the backbone nitrogen atoms of L138 and SAM were 3.1 A, which was higher than the value discovered in our investigation (2.28 A). Y86 in our research formed a hydrogen bond with 2.86 A whereas Y86 OH linked with 2.5 A with SAM (Elkins *et. al.,* 2003). The 4% difference in structure between *Salmonella Typhimurium* (LT2) and *Escherichia coli* (k12) *trm*D could be the cause for this. This implies that Y86 and L138 are essential aminoacids for interacting with ligands.

4.8.1.2 Protein Interaction with Natural Product (NP)

Table 4.11: Nature of chemical interaction between NP and trmD Salmonella Typhimurium

Amino acid residues (Within 5 A°)	Bond length (A°)	Interaction type	Chemical group of ligands involved
Pro 89	3.85	Pi-alkyl	Benzene ring
Tyr 115	2.73	H- bond	Ring (x)
Ile 133	4.70	Pi- alkyl	Benzene
Pro 144	4.20	Pi- alkyl	Benzene
Leu 138	5.22	Pi- alkyl	Benzene

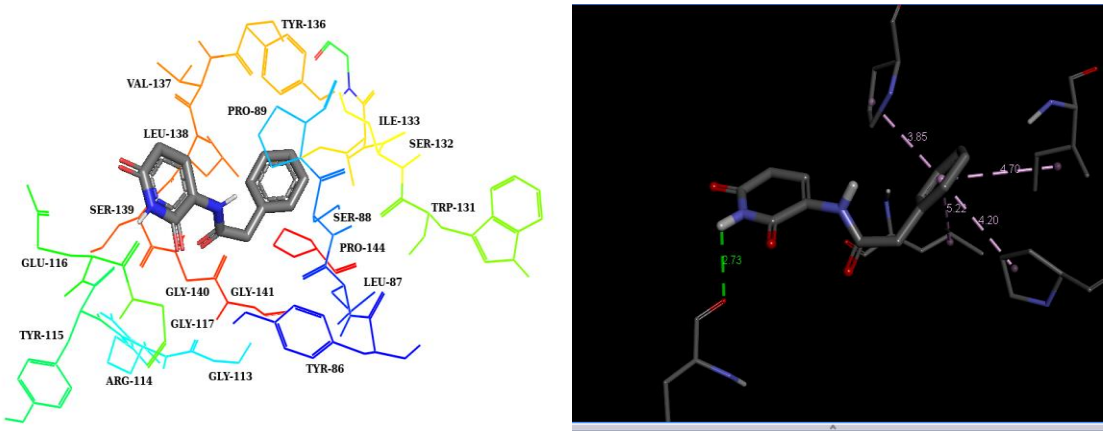


Fig 4.13: Interaction of NP with active sites of trmD Salmonella visualized using PyMol (left) and bond length (right) visualized using BIOVA- discovery Studio

4.8.1.3 Protein Interaction with Kinase Inhibitors (KI)

Table 4.12: Nature of chemical interaction between KI 1 and trmD Salmonella Typhimurium

Amino acid residues (Within 5 A°)	**Bond length (A°)**	**Interaction type**	**Chemical group of ligands involved**
Leu 87	2.78	H- bond	Hydrogen
Gly 134	2.84	H- bond	Fluorine
Tyr 136	2.91	H- bond	Fluorine of Benzene ring
Glu 116	4.83	Amide π- stacked	x- ring
	5.73	Amide π- stacked	x- ring
Val 137	4.30	Pi- alkyl	x-ring
Pro 89	3.85	Pi- alkyl	Fluoro- benzene ring
Ile 133	4.76	Pi- alkyl	Fluoro- benzene ring
Pro 144	4.19	Pi- alkyl	Fluoro- benzene ring
Leu 138	5.30	Pi- alkyl	Fluoro- benzene ring
Gly 134	3.08	H- bond	Fluorine

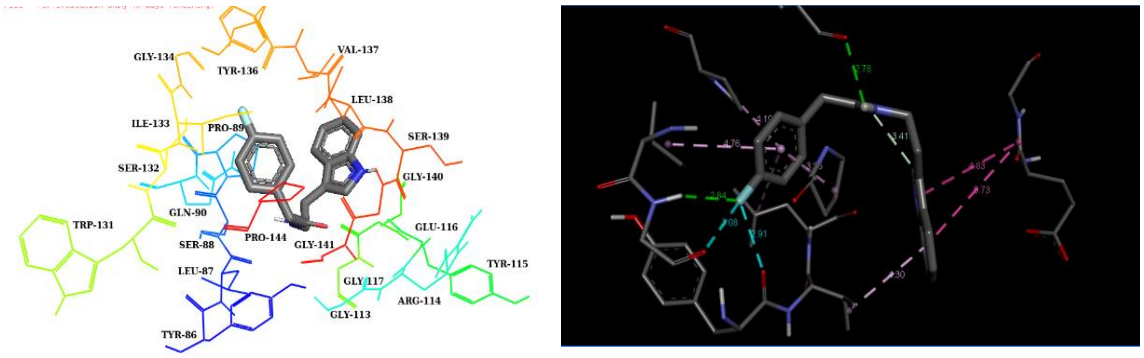


*Fig 4.14: Interaction of KI 1 with active sites of trm*D *Salmonella visualized using PyMol (left) and bond length (right) visualized using BIOVA- discovery Studio*

*Table 4.13: Nature of chemical interaction between KI 2 and trm*D *Salmonella Typhimurium*

Amino acid residues (Within 5 A°)	**Bond length (A°)**	**Interaction type**	**Chemical group of ligands involved**
Leu 87	2.64	H- bond	Hydrogen
Pro 89	3.93	Pi- alkyl	Benzene-ring
	4.40	Pi- alkyl	x- ring
	4.69	Pi- alkyl	x- ring
Val 137	4.91	Pi- alkyl	x-ring
Ile 133	4.81	Pi- alkyl	benzene ring
Pro 144	4.11	Pi- alkyl	benzene ring
Leu 138	5.25	Pi- alkyl	benzene ring

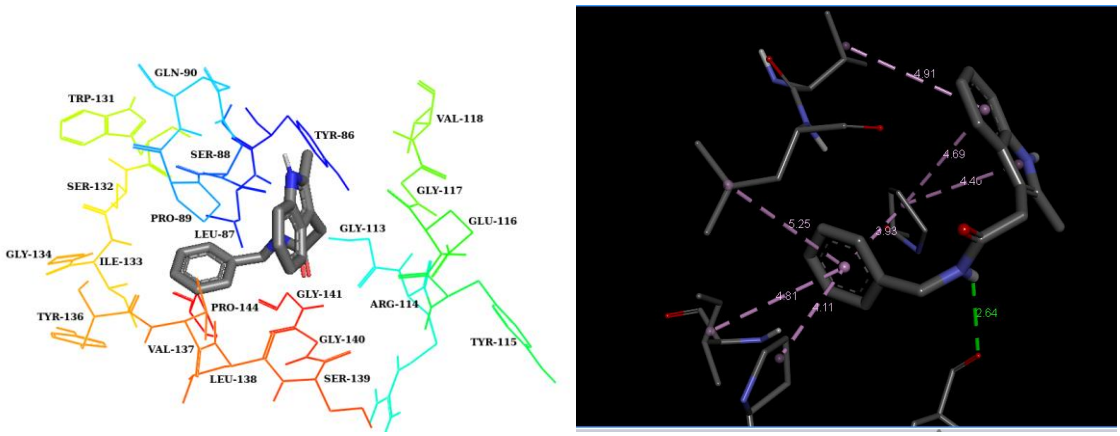


*Fig 4.15: Interaction of KI2 with active sites of trm*D *Salmonella visualized using PyMol (left) and bond length (right) visualized using BIOVA- discovery Studio*

*Table 4.14: Nature of chemical interaction between KI 3 and trm*D *Salmonella Typhimurium*

Amino acid residues (Within 5 A°)	**Bond length (A°)**	**Interaction type**	**Chemical group of ligands involved**
Leu 138	2.50	Hydrogen	Terminal oxygen
Leu 87	2.09	H- bond	Terminal Nitrogen
Tyr 86	2.63	H- Bond	Terminal Nitrogen
Leu 138	5.22	Pi- alkyl	Benzene
Pro 89	3.87	Pi- alkyl	Benzene
Ile 133	4.80	Pi- alkyl	Benzene
Pro 144	4.17	Pi- alkyl	Benzene
Gly 113	3.51	Carbon-Hydrogen	Central carbon

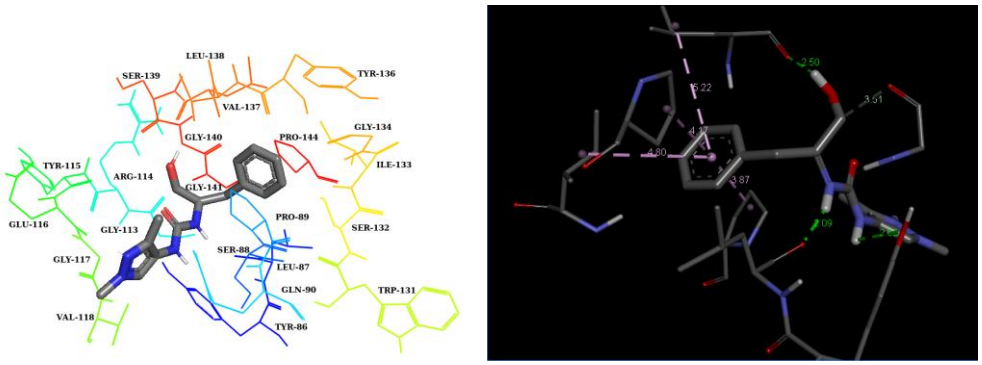


*Fig 4.16: Interaction of KI 3 with active sites of trm*D *Salmonella visualized using PyMol (left) and bond length (right) visualized using BIOVA- discovery Studio*

*Table 4.15: Nature of chemical interaction between KI 4 and trm*D *Salmonella Typhimurium*

Amino acid residues (Within 5 A°)	**Bond length (A°)**	**Interaction type**	**Chemical group of ligands involved**
Val 118	2.81	Hydrogen	Hydrogen of Pyridine ring
Tyr 86	2.51	H- bond	Hydrogen
	2.72	H- Bond	Terminal Nitrogen
Leu 87	2.20	H-bond	Hydrogen
Pro 89	3.86	Pi- alkyl	Benzene
Ile 133	4.80	Pi- alkyl	Benzene
Pro 144	4.18	Pi- alkyl	Benzene
Leu 138	5.24	Pi-alkyl	Benzene
Gly 113	3.53	Carbon-hydrogen	Methyl
Val 118	3.21	Pi donor-hydrogen	Pyridine ring

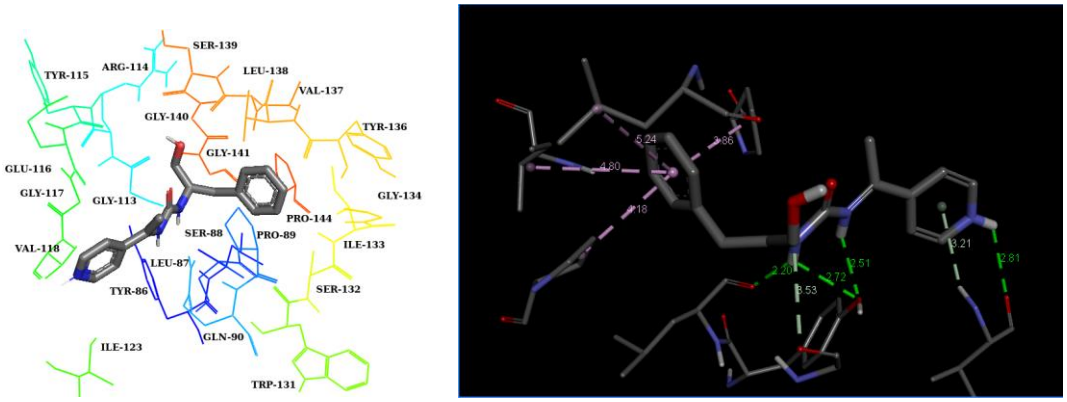


*Fig 4.17: Interaction of KI 4 with active sites of trm*D *Salmonella visualized using PyMol (left) and bond length (right) visualized using BIOVA- discovery Studio*

*Table 4.16: Nature of chemical interaction between KI 5 and trm*D *Salmonella Typhimurium*

Amino acid residues (Within 5 A°)	**Bond length (A°)**	**Interaction type**	**Chemical group of ligands involved**
Tyr 86	2.07	Hydrogen	Hydrogen
Leu 138	1.97	H- bond	Hydrogen
	5.23	Pi-alkyl bond	Benzene ring
Leu 87	2.51	H-bond	Hydrogen
Pro 89	3.86	Pi- alkyl	Benzene
	5.42	Pi-alkyl	x- ring
Glu 116	2.99	H-bond	Hydrogen
Pro 144	4.19	Pi- alkyl	Benzene
Ile 133	4.89	Pi-alkyl	Benzene
Gly 113	3.63	Carbon-hydrogen	Methyl

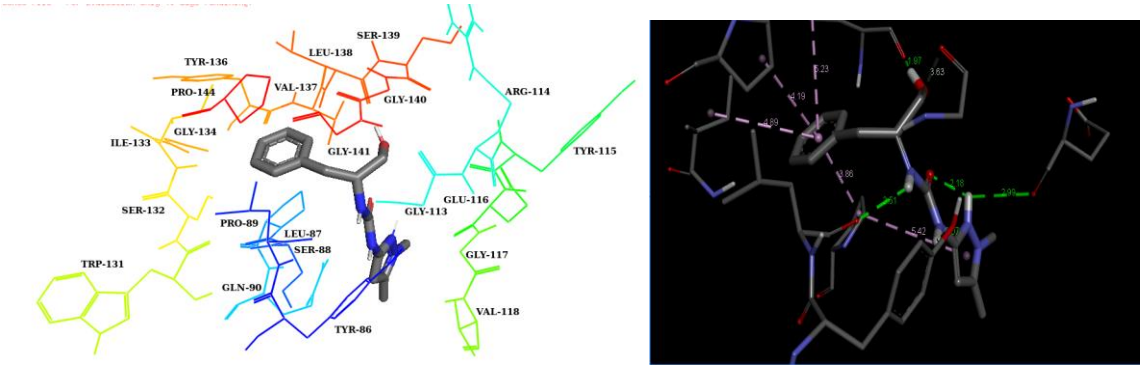


*Fig 4.18: Interaction of KI 5 with active sites of trm*D *Salmonella visualized using PyMol (left) and bond length (right) visualized using BIOVA- discovery Studio*

4.8.1.4 Protein Interaction with Nucleoside Mimetics (NM)

*Table 4.17: Nature of chemical interaction between NM 1 and trm*D *Salmonella Typhimurium*

Amino acid residues (Within 5 A°)	**Bond length (A°)**	**Interaction type**	**Chemical group of ligands involved**
Glu 116	2.51	Hydrogen	Hydrogen
Leu 138	2.73	H- bond	x- ring
	3.63	Carbon-hydrogen	x-ring
	5.14	Pi- alkyl	Benzene ring
Gln 90	2.46	Hydrogen	Oxygen
Gly 141	2.77	Carbon-hydrogen	oxygen
Pro 89	5.28	Pi-alkyl	Benzene
	3.80	Pi- alkyl	Benzene
Ile 133	4.71	Pi-alkyl	Benzene

Pro 144	4.29	Pi- alkyl	Benzene

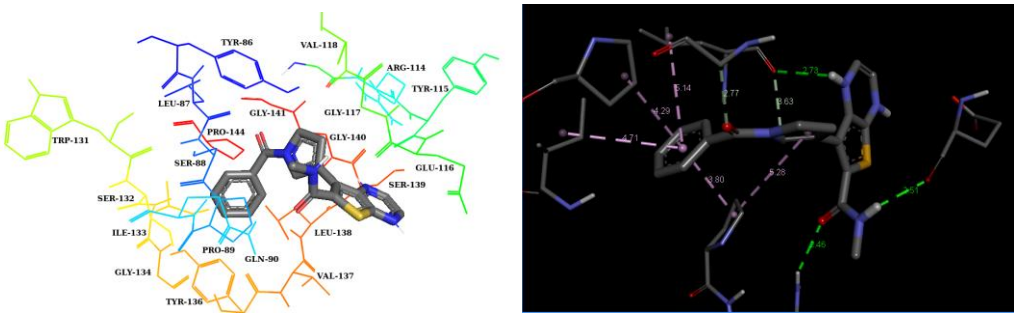


*Fig 4.19: Interaction of NM 1 with active sites of trm*D *Salmonella visualized using PyMol (left) and bond length (right) visualized using BIOVA- discovery Studio*

*Table 4.18: Nature of chemical interaction between NM 2 and trm*D *Salmonella Typhimurium*

Amino acid residues (Within 5 A°)	**Bond length (A°)**	**Interaction type**	**Chemical group of ligands involved**
Tyr 115	2.15	Hydrogen	Hydrogen
	2.97	H- bond	x- ring
Ser 139	2.50	Hydrogen	hydrogen
	2.36	Hydrogen	Hydrogen
Leu 138	4.59	Alkyl	Terminal carbon
Pro 144	4.71	Alkyl	Terminal carbon
Pro 89	4.28	Alkyl	Terminal carbon
Gly 117	3.64	Carbon-hydrogen	x-ring
Pro 89	4.92	Pi- alkyl	x-ring

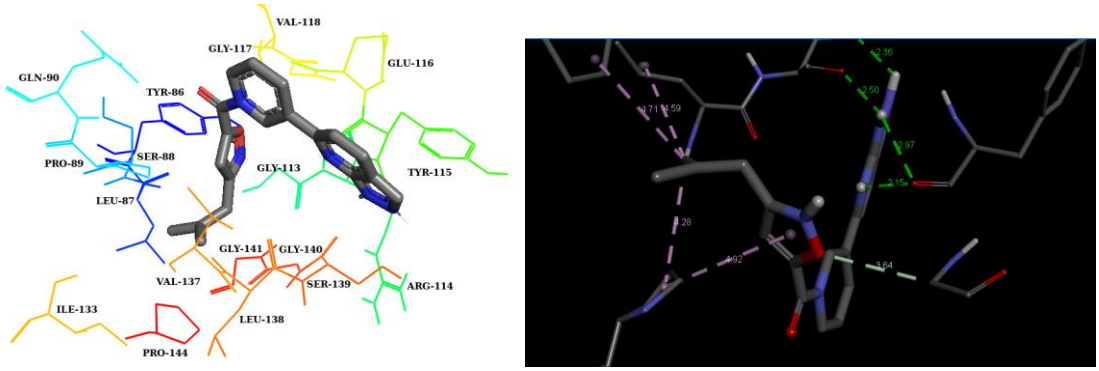


*Fig 4.20: Interaction of NM 2 with active sites of trm*D *Salmonella visualized using PyMol (left) and bond length (right) visualized using BIOVA- discovery Studio*

4.8.1.5 Protein Interaction with Indole Derivatives (ID)

*Table 4.19: Nature of chemical interaction between I*D *1 an*D *trm*D *Salmonella Typhimurium*

Amino acid residues (Within 5 A°)	**Bond length (A°)**	**Interaction type**	**Chemical group of ligands involved**
Tyr 115	1.96	Hydrogen	Hydrogen
Gln 90	2.17	Hydrogen	Oxygen
	2.56	Hydrogen	Oxygen
Pro 144	4.02	Pi- alkyl	Indole ring
	4.52	Pi- alkyl	Indole ring
Pro 89	4.14	Pi- alkyl	Indole ring
	4.06	Pi- alkyl	Indole ring
	4.13	Pi- alkyl	Benzene ring
Val 137	5.40	Pi- alkyl	Benzene ring
Leu 138	3.53	Carbon-hydrogen	Carbon chain

Gly 140	4.15	Amide pi-stacked	Indole ring
Ile 133	4.47	Pi-alkyl	Indole ring
Ser 132	3.61	Carbon-hydrogen	Indole ring

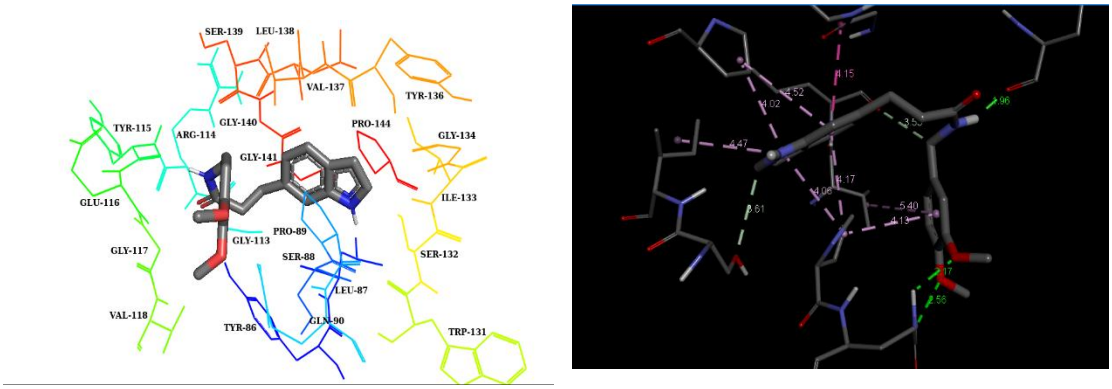


Fig 4.21: Interaction of ID 1 with active sites of trmD Salmonella visualized using PyMol (left) and bond length (right) visualized using BIOVA- discovery Studio

Table 4.20: Nature of chemical interaction between ID 2 and trmD Salmonella Typhimurium

Amino acid residues (Within 5 A°)	**Bond length (A°)**	**Interaction type**	**Chemical group of ligands involved**
Gly 113	2.30	Hydrogen bond	Oxygen
Tyr 86	2.66	Hydrogen bond	Oxygen
Pro 89	3.81	Pi- alkyl	Benzene ring
	5.11	Pi- alkyl	Indole ring
Pro 144	4.25	Pi- alkyl	Benzene ring
Leu 138	5.20	Pi- alkyl	Benzene ring
Ile 133	4.77	Pi- alkyl	Benzene ring
Tyr 115	3.78	Carbon-hydrogen	Carbon

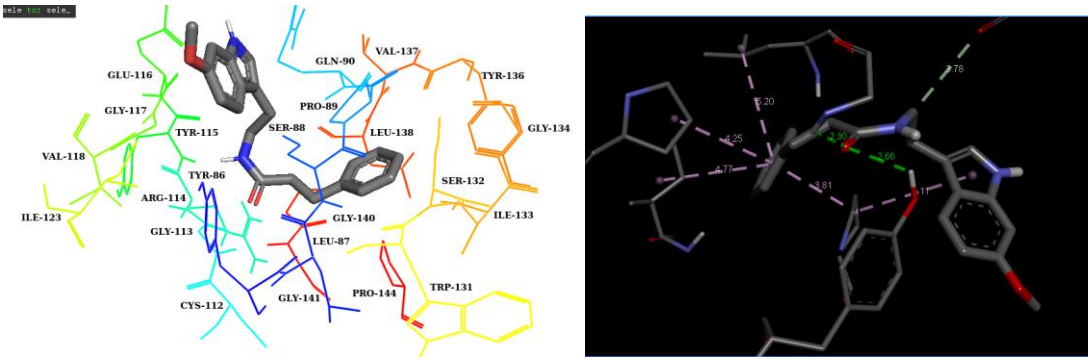


*Fig 4.22: Interaction of ID 3 with active sites of trm*D *Salmonella visualized using PyMol (left) and bond length (right) visualized using BIOVA- discovery Studio*

*Table 4.21: Nature of chemical interaction between ID 3 and trm*D *Salmonella Typhimurium*

Amino acid residues (Within 5 A°)	**Bond length (A°)**	**Interaction type**	**Chemical group of ligands involved**
Tyr 115	3.10	Hydrogen bond	Indole ring
Leu 138	2.94	Hydrogen bond	Indole ring
	5.25	Pi- alkyl	Benzene ring
Gly 134	2.69	Hydrogen bond	Fluorine of benzene ring
Leu 87	2.87	Hydrogen bond	Hydrogen
Pro 89	5.50	Pi- alkyl	Indole ring
	3.83	Pi- alkyl	Benzene ring
Ile 133	4.75	Pi- alkyl	Benzene ring
Pro 144	4.22	Pi- alkyl	Benzene ring
Tyr 136	3.02	Hydrogen bond	Fluorine
Val 137	4.34	Pi- alkyl	Indole ring

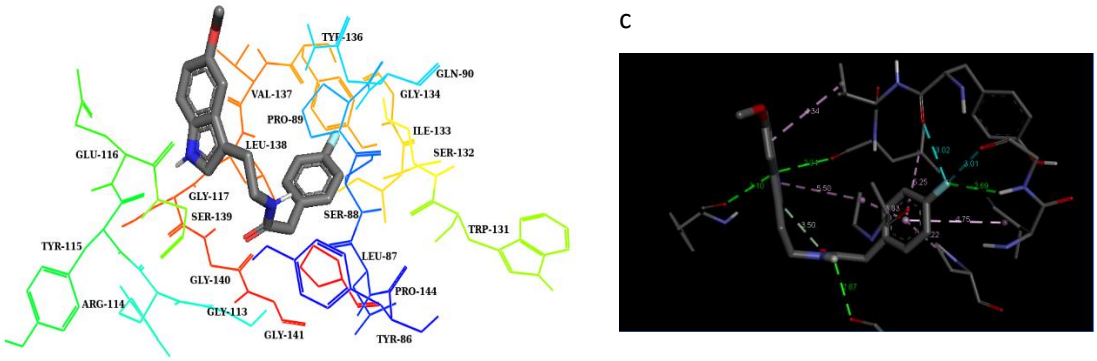


*Fig 4.23: Interaction of ID 4 with active sites of trm*D *Salmonella visualized using PyMol (left) and bond length (right) visualized using BIOVA- discovery Studio*

4.9 Prospects of Drugs against *trm*D

The adenine loop, extending approximately from residues 132 to 141, is very strongly conserved in all bacterial *Trm*D proteins and the active site to bind the adenine of SAM lies in the residues 132–141 (Elkins et.al., 2003). In a study about Thienopyrimidine derivatives against bacterial *trm*D done by Zhong *et. al.,* 2019, they found out that the thienopyrimidinone ring was tightly bound in the adenine pocket and was hydrogen-bonded to residues Leu138, Tyr141, and Leu143. In another study of Tha, *et. al.,* 2020, they found an indole derivative 1-Methyl-3,4-bis(3-indolyl) maleimide found to inhibit *Trm*D (Tha et.al., 2020). In our study as well (shown in table 4.10 – 4.21), all of the top hits ligand were found to be bonded in with the protein at Leu 138, where, ID 3, KI 5, NM 1 were found to be in H-bond interaction as well suggesting that these serve as a prospective candidates against *trm*D.

4.10 Mutational Analysis

Mutations can be particularly harmful to a cell or organism because they frequently result in changes to the structure of an encoded protein, as well as a reduction or loss of its expression. Since the DNA sequence is altered, all copies of the encoded protein are also

affected. (https://www.ncbi.nlm.nih.gov/books/NBK21578/) . In a study conducted by Elkins *et. al.,* in 2003, alanine was substituted for half of the residues in the active site. In all bacterial *Trm*D proteins, the adenine loop, which roughly spans residues 132 to 141, is highly conserved. Site-directed mutations that were relevant were created and their effects on catalytic activity were evaluated. The substantial impacts of mutation on catalytic activity were discovered to be primarily caused by two evident structural considerations. First, residues in the adenine-binding loop and throughout the active site engage directly with the cofactor or with side chains of other residues via their peptide nitrogen or carbonyl oxygen atoms. The cofactor-binding pocket's tertiary structure is kept stable by these interactions. Second, several residues in the active site include side chains that interact with monomers in hydrophobic or ionic ways across the interface, which could impair catalytic activity by destabilizing dimers. Extensive mutagenesis of residues in or close to this region was instructive given the adenine-binding loop's apparent important function in SAM binding. (Elkins *et. al.,* 2003)

4.11 DFT Analysis

4.11.1 Dipole moment and total energy of the molecule

The dipole moment being widely used parameter, which has been shown to explain observable chemical and physical properties of molecules in many different contexts has application in drug discovery attracts high interest (Wojciechowski *et. al.,* 2014 and Ioakimidis *et. al.,* 2008). As compounds with large dipole moments are generally more soluble in water and less likely to be absorbed through lipophilic membranes, the dipole moment has been useful in the assessment of cell permeability and oral bioavailability of drugs (Ioakimdis *et. al.,* 2008 and Matuszek *et. al.,* 2016). The dipole moment for the selected compound Natural Product (Antineoplaston) showed a value of 7.4870 Debye and this is in conjunction with the fact that drug like molecules have a dipole moment less than 10 Debye (Periera *et. al.,* 2018). Since dipole moment directly relates with cell permeability and oral biovailabilty of drugs, proper analysis of dipole moment can also act as proper descriptor for the estimation of drug loading capacities as well (Wu *et. al.,* 2016). Similarly, the molecule showed a total energy of -836.95199857 a.u. (in Hartree). The higher total energy of the molecule suggests the less stable and more reactive nature of the molecule.

Further, free energy is a pivotal criterion to represent the interactions of binding partners where both the sign and magnitude are important to express the likelihood of bimolecular events occurring and greater negative values indicate improved thermodynamic properties (Moniruzzaman *et. al.*, 2018). In this study, it is found that the values are negative meaning the binding will occur spontaneously without any extra energy expenditure.

4.11.2 Molecular orbital properties

To evaluate their chemical reactivity, the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) were computed using DFT analysis. While LUMO energy denotes the molecule's ability to take electrons from the partner protein, HOMO energy suggests the region of the tiny molecules that can give electrons during the complex formation (Banavath *et. al.,* 2014). Natural Product Antineoplaston HOMO and LUMO were calculated to be -0.18311 and -0.0899 eV and the HOMO-LUMO gap energy to be 0.09321 electron volts, respectively.

Because adding electrons to a high-lying LUMO and/or removing electrons from a low-lying HOMO is energetically advantageous in any potential reaction, large HOMO-LUMO gaps are associated with high kinetic stability and low chemical reactivity, and small HOMO-LUMO gaps are important for low chemical stability (Aihara *et. al.,* 1999). The smaller HOMO-LUMO gap in the case of our ligand denotes increased chemical reactivity. Additionally, the HOMO-LUMO gap energy, or the difference between HOMO and LUMO energy, denotes the electronic excitation energy required to calculate the compounds' molecular reactivity and stability (Zheng *et. al.,* 2013).

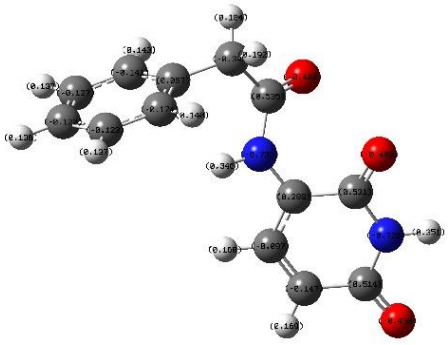
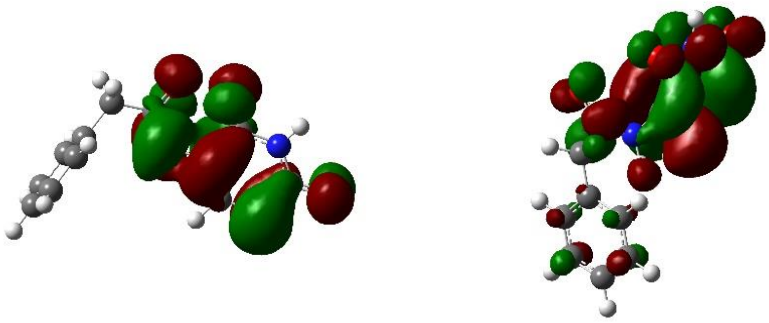


Fig 4.24: Bond length interaction between the atoms of Antineoplaston



(a) LUMO (b) HOMO

Fig 4.25: Molecular Orbital Properties (a) HOMO and (b) LUMO

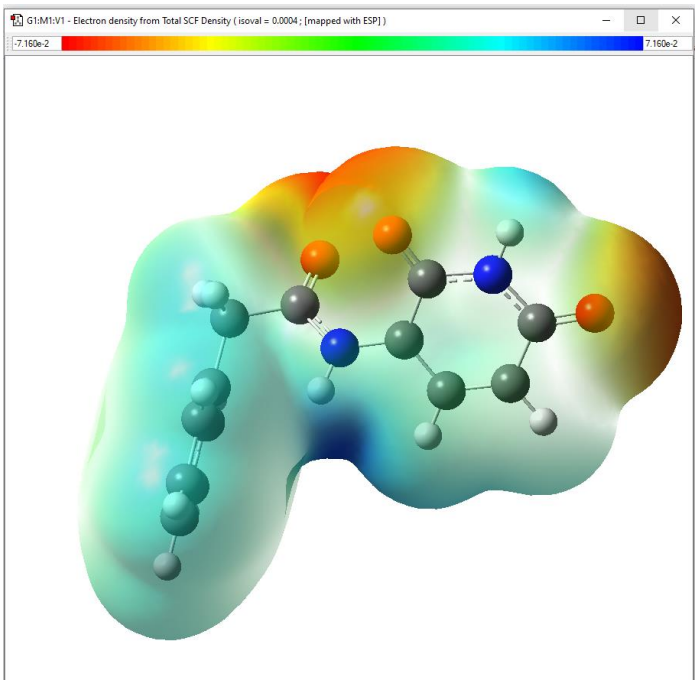


Fig 4.26: Electon density in Antineoplaston visualized in Gaussview

4.12 *trm*D as a safe bet with human trm5 protein

Regardless of the type of nucleotide at position 36, Trm5 still methylates G37. In contrast to *Trm*D, which preferentially identifies the G36pG37 motif and does not methylate inosine, TRM5 also methylates inosine at position 37. *Trm*D is the subject of fresh

evidence demonstrating that some tRNA species also recognize A at position 36. The tRNA-protein tertiary connection can vary subtly, and this can cause the TRM5 enzyme to lose function. If the G36pG37 and the core tRNA structure are present, the *Trm*D enzyme is more tolerant of changes in tRNA-protein tertiary interactions. Unlike *Trm*D, which needs magnesium to express activity, the TRM5 enzyme does not absolutely require magnesium ions (Brule *et.al.,* 2004). This elucidates that the drug development against *trm*D gene is efficient as it donot share any analogy with the eukaryotic trm5 protein.

5. SUMMARY

With the increase in resistant bacteria, it is important to develop a drug which acts in such metabolic pathways that are not quickly mutated. *Streptomyces* is a ubiquitous class of bacteria with potential of antibiotics production. In this study, a colistin resistant *Salmonella Typhimurium* was found to be impotent against Ma strain rather that Pc strain of *Streptomyces* (grown in a tweaked media for indole production*)* as evident by the Resazurin assay performed. When the number of MDR *Salmonella* are increasing exponentially, it is important to develop a quick and fast way to screen the available drugs against them and in-vitro tests takes years to complete. Another approach to overcome that, would be virtual screening of the available drug candidates against the desired protein of interest. The role of *Trm*D in a S-adenosyl methionine (AdoMet or SAM)-dependent methyl transfer to creates the methylation m1G37 in tRNA pro and it being essentially different from its eukaryotic and archeal cousin Trm5 makes it as a probable protein target. Due to the lack of availability of the protein in the database, homology modelling using MODELLER was employed to generate the protein. Based on the hypothesis that some available ligands from database (Natural Products, Kinase Inhibitors, Nucleoside Mimetics and Indole Derivatives) may be effective against the *trm*D (tRNA methylation D) of *Salmonella Typhimurium* (LT2) strain, the molecular docking was performed against Natural Products, Kinase Inhibitors, Nucleoside mimetics and Indole Derivatives. The compounds were narrowed based on their ADME/T and virtual binding energy that exhibited higher binding efficiencies than native ligand SAM. From the virtual screening, as in the study of Elkins *et. al.,* 2003, the compounds that were able to be effective against target protein and bind with the amino acid L138 in the active site were found to be one Natural Products Antineoplaston, five kinase inhibitors 2-(4-Fluorophenyl)-N-[2-(1H-indol-3-yl) ethyl] acetamide; N-benzyl-2-(2-methyl-1H-indol-3-yl) acetamide; 1-(1,3-Dimethylpyrazol-4-yl)-3-[(2R)-1-hydroxy-3-phenylpropan-2-yl]urea; ; 1-[(2S)-1-Hydroxy-3-phenylpropan-2-yl]-3-(1-pyridin-4-ylethyl)urea; 1-(1,5-Dimethylpyrazol-3-yl)-3-[(2S)-1-hydroxy-3-phenylpropan-2-yl]urea three indole derivative (N-[(3,4-dimethoxyphenyl)methyl]-3-(1H-indol-6-yl)propenamide; N-[2-(6-

methoxy-1H-indol-3-yl)ethyl]-3-phenylpropanamide and 2-(4-fluorophenyl)-N-[2-(5-methoxy-1H-indol-3-yl)ethyl]acetamide were screened out as drug candidates against tRNA methyl transferase of *Salmonella* sps. Similarly, two nucleoside mimetics 7-[(3R)-1-Benzoylpyrrolidin-3-yl]-N-methylthieno[2,3-b] pyrazine-6-carboxamide and [3-(2-Methylpropyl)-1,2-oxazol-5-yl]-[(3S)-3-(1H-pyrazolo[3,4-b]pyridin-6-yl)piperidin-1-yl]methanone were found to be effective . Indole derivatives and their metabolites have been reported to function as kinase inhibitors. Thus, screening of indole derivatives was done as competitive SAM inhibitors because of adenosine moiety of SAM that is derived from ATP on the hypothesis that the kinase inhibition by indole could also be competing with the adenosine binding pocket of kinases. These leads suggested that the compounds with pyridine and indole ring with similarity in adenosyl moeity of SAM has some impact. Moreover, from Natural Products Antineoplaston which has been used in trials studying the treatment of Sarcoma, Lymphoma, Lung Cancer, Liver Cancer, and Kidney Cancer, among others was also found to be effective against *trm*D of Salmonella, interestingly, has piperidine ring as well. For the Natural Product Antineoplaston, the thermodynamic properties and molecular orbital properties supported the drug like properties of the compound. With the use of structure based virtual screening, density function theory analysis, mutation prediction and efficacy of developed hits and drug-drug interaction studies, this study provides us the framework to develop the potential therapeutics for main protease of COVID-19 encompassing the different genres of computational chemistry.

6. CONCLUSION

The increased antimicrobial resistance problems in the poultry and clinically isolated pathogens have rendered all the developed antibiotics ineffective. From the present study, with computational approaches to find out new drug candidates revealed potent inhibitor against *Salmonella* sps. It was hypothesized few of the kinase Inhibitors, Nucleoside mimetics, Indole Derivatives, and Natural Products screened from the database could be putative competitive inhibitors for SAM binding pocket in *trm*D gene of *Salmonella Typhimurium* (LT2). One of the revived strains of *Streptomyces* (Ma) was found to give good results in the resazurin assay of its secondary metabolite. The computational drug designing screened several compounds with good drugability and potential against the target protein. All of them were found to be bound with the target protein in the active site amino acid residue L138 similar to the studies done before. In addition, the hypothesis set to study the thermodynamic properties and molecular orbital properties of developed hit, assessing the drug like properties of the hit, along with the drug-drug interaction study rendered the strong support for its prospects to start the lead optimization in drug discovery process.

7. RECOMMENDATIONS

Thus, it is recommended that the interactions between the top hit Antineoplaston and *trm*D of Salmonella be verified by carrying out the molecular dynamics (MD) simulations. This molecule can be further pursued as probable drug candidate for drug development. Respective enzymes inhibition kinetics could be studied and riboswitch structure could also be investigated. Animal testing and toxicity testing can be the way forward to strengthen the candidacy of this compound as potential suitor.

8. BIBLIOGRAPHY

Ahn, H. J., Kim, H. W., Yoon, H. J., Lee, B. I., Suh, S. W., & Yang, J. K. (2003). Crystal structure of tRNA(m1G37) methyltransferase: insights into tRNA recognition. *The EMBO journal*, *22*(11), 2593–2603. https://doi.org/10.1093/emboj/cdg269

Anandakrishnan, R., Aguilar, B., & Onufriev, A. V. (2012). H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. Nucleic acids research, 40(Web Server issue), W537–W541. https://doi.org/10.1093/nar/gks375

Athanasiadis, E., Cournia, Z., & Spyrou, G. (2012). ChemBioServer: a web-based pipeline for filtering, clustering and visualization of chemical compounds used in drug discovery. *Bioinformatics (OxforD, EnglanD), 28(22)*, 3002–3003. https://doi.org/10.1093/bioinformatics/bts551

Berg E. L. (2014). Systems biology in drug discovery and development. D*rug* D*iscovery to*D*ay*, *19*(2), 113–125. https://doi.org/10.1016/j.drudis.2013.10.003

Bhattacharya S. (2013). Early diagnosis of resistant pathogens: how can it improve antimicrobial treatment?. *Virulence*, *4*(2), 172–184. https://doi.org/10.4161/viru.23326

Björk, G. R., Wikström, P. M., & Byström, A. S. (1989). Prevention of translational frameshifting by the modified nucleoside 1-methylguanosine. *Science (New York, N.Y.)*, *244*(4907), 986–989. https://doi.org/10.1126/science.2471265

Brulé, H., Elliott, M., Redlak, M., Zehner, Z. E., & Holmes, W. M. (2004). Isolation and characterization of the human tRNA-(N1G37) methyltransferase (TRM5) and comparison to the *Escherichia coli Trm*D protein. *Biochemistry*, *43*(28), 9243–9255. https://doi.org/10.1021/bi049671q

Butcher, E. C., Berg, E. L., & Kunkel, E. J. (2004). Systems biology in drug discovery. *Nature biotechnology*, *22*(10), 1253–1259. https://doi.org/10.1038/nbt1017

Byström, A. S., & Björk, G. R. (1982). The structural gene (*trm*D) for the tRNA(m1G)methyltransferase is part of a four polypeptide operon in *Escherichia coli* K-12. *Molecular & general genetics : MGG*, *188*(3), 447–454. https://doi.org/10.1007/BF00330047

Carpenter, K. A., Cohen, D. S., Jarrell, J. T., & Huang, X. (2018). Deep learning and virtual drug screening. *Future meDicinal chemistry, 10(21),* 2557–2567. https://doi.org/10.4155/fmc-2018-0314

Carrat, F., & Flahault, A. (2007). Influenza vaccine: The challenge of antigenic drift. *Vaccine, 25(39–40).* 6852–6862.

Carroll, L. M., Gaballa, A., Guldimann, C., Sullivan, G., Henderson, L. O., & Wiedmann, M. (2019). Identification of Novel Mobilized Colistin Resistance Gene *mcr-9* in a Multidrug-Resistant, Colistin-Susceptible Salmonella enterica Serotype Typhimurium Isolate. *mBio*, *10*(3), e00853-19. https://doi.org/10.1128/mBio.00853-19

Casas, M. R., Camargo, C. H., Soares, F. B., da Silveira, W. D., & Fernandes, S. A. (2016). Presence of plasmid-mediated quinolone resistance determinants and mutations in gyrase and topoisomerase in Salmonella enterica isolates with resistance and reduced susceptibility to ciprofloxacin. D*iagnostic microbiology an*D *infectious* D*isease*, *85*(1), 85–89. https://doi.org/10.1016/j.diagmicrobio.2016.01.016

Chandra N. (2009). Computational systems approach for drug target discovery. *Expert opinion on* D*rug* D*iscovery*, *4*(12), 1221–1236. https://doi.org/10.1517/17460440903380422

Chen, J. L., Steele, T. W. J., & Stuckey, D. C. (2015). Modeling and Application of a Rapid Fluorescence-Based Assay for Biotoxicity in Anaerobic Digestion. *Environmental Science an*D *Technology*, *49*(22), 13463–13471. https://doi.org/10.1021/acs.est.5b03050

Christian, T., & Hou, Y. M. (2007). Distinct determinants of tRNA recognition by the *Trm*D and Trm5 methyl transferases. *Journal of molecular biology*, *373*(3), 623–632. https://doi.org/10.1016/j.jmb.2007.08.010

Christian, T., Evilia, C., Williams, S., & Hou, Y. M. (2004). Distinct origins of tRNA(m1G37) methyltransferase. *Journal of molecular biology*, *339*(4), 707–719. https://doi.org/10.1016/j.jmb.2004.04.025

Christian, T., Lahoud, G., Liu, C., & Hou, Y.-M. (2010). Control of catalytic cycle by a pair of analogous tRNA modification enzymes. *Journal of Molecular Biology*, *400*(2), 204–217. https://doi.org/10.1016/j.jmb.2010.05.003

Craig, I. R., Essex, J. W., & Spiegel, K. (2010). Ensemble docking into multiple crystallographically derived protein structures: an evaluation based on the statistical analysis of enrichments. *Journal of chemical information an*D *mo*D*eling, 50(4),* 511–524. https://doi.org/10.1021/ci900407c

Csermely, P., Agoston, V., & Pongor, S. (2005). The efficiency of multi-target drugs: the network approach might help drug design. *Tren*D*s in pharmacological sciences*, *26*(4), 178–182. https://doi.org/10.1016/j.tips.2005.02.007

Cummings, K. J., Rodriguez-Rivera, L. D., Norman, K. N., Ohta, N., & Scott, H. M. (2017). Identification of a Plasmid-Mediated Quinolone Resistance Gene in Salmonella Isolates from Texas Dairy Farm Environmental Samples. *Zoonoses an*D *public health*, *64*(4), 305–307. https://doi.org/10.1111/zph.12318

Detection of *gyrA* Gene in *Salmonella enterica* serovar Typhi Isolated from Typhoid Patients in Baghdad. *Pakistan journal of biological sciences : PJBS*, *23*(10), 1303–1309. https://doi.org/10.3923/pjbs.2020.1303.1309

Ding, B., Wang, J., Li, N., & Wang, W. (2013). Characterization of small molecule binding. I. Accurate identification of strong inhibitors in virtual screening. *Journal of chemical information an*D *mo*D*eling, 53(1),* 114–122. https://doi.org/10.1021/ci300508m

Dolinsky, T. J., Czodrowski, P., Li, H., Nielsen, J. E., Jensen, J. H., Klebe, G., & Baker, N. A. (2007). PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic aci*D*s research, 35(Web Server issue),* W522–W525. https://doi.org/10.1093/nar/gkm276

Douguet D. (2010). e-LEA3D: a computational-aided drug design web server. *Nucleic aci*D*s research*, *38*(Web Server issue), W615–W621. https://doi.org/10.1093/nar/gkq322

Dutkiewicz, Z., & Mikstacka, R. (2018). Structure-Based Drug Design for Cytochrome P450 Family 1 Inhibitors. *Bioinorganic chemistry an*D *applications*, *2018*, 3924608. https://doi.org/10.1155/2018/3924608

Elkins, P. A., Watts, J. M., Zalacain, M., van Thiel, A., Vitazka, P. R., Redlak, M., Andraos-Selim, C., Rastinejad, F., & Holmes, W. M. (2003). Insights into catalysis by a knotted *Trm*D tRNA methyltransferase. *Journal of molecular biology*, *333*(5), 931–949. https://doi.org/10.1016/j.jmb.2003.09.011

Elkins, P. A., Watts, J. M., Zalacain, M., van Thiel, A., Vitazka, P. R., Redlak, M., … Holmes, W. M. (2003). Insights into catalysis by a knotted *Trm*D tRNA methyltransferase. *Journal of Molecular Biology*, *333*(5), 931–949.

Ewing, T. J., Makino, S., Skillman, A. G., & Kuntz, I. D. (2001). DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *Journal of computer-ai*D*e*D *molecular* D*esign, 15(5),* 411–428. https://doi.org/10.1023/a:1011115820450

Finta, C., & Zaphiropoulos, P. G. (2000). The human cytochrome P450 3A locus. Gene evolution by capture of downstream exons. *Gene, 260(1-2),* 13–23. https://doi.org/10.1016/s0378-1119(00)00470-4

Fiser A. (2004). Protein structure modeling in the proteomics era. *Expert review of proteomics*, *1*(1), 97–110. https://doi.org/10.1586/14789450.1.1.97

Flärdh, K., & Buttner, M. J. (2009). *Streptomyces* morphogenetics: Dissecting differentiation in a filamentous bacterium. *Nature Reviews Microbiology*.

https://doi.org/10.1038/nrmicro1968

Flärdh, K., & Buttner, M. J. (2009). *Streptomyces* morphogenetics: Dissecting differentiation in a filamentous bacterium. *Nature Reviews Microbiology*. https://doi.org/10.1038/nrmicro1968

Foley, S. L., Lynne, A. M., & Nayak, R. (2008). Salmonella challenges: prevalence in swine and poultry and potential pathogenicity of such isolates. *Journal of animal science*, *86*(14 Suppl), E149–E162. https://doi.org/10.2527/jas.2007-0464

Foley, S. L., Nayak, R., Hanning, I. B., Johnson, T. J., Han, J., & Ricke, S. C. (2011). Population dynamics of Salmonella enterica serotypes in commercial egg and poultry production. *Applie*D *an*D *environmental microbiology*, *77*(13), 4273–4279. https://doi.org/10.1128/AEM.00598-11

Forst C. V. (2002). Network genomics--a novel approach for the analysis of biological systems in the post-genomic era. *Molecular biology reports*, *29*(3), 265–280. https://doi.org/10.1023/a:1020437311167

Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., & Shenkin, P. S. (2004). Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of me*D*icinal chemistry, 47(7),* 1739–1749. https://doi.org/10.1021/jm0306430

Galperin, M. Y., Walker, D. R., & Koonin, E. V. (1998). Analogous enzymes: independent inventions in enzyme evolution. *Genome research*, *8*(8), 779–790. https://doi.org/10.1101/gr.8.8.779

Gamper HB, Masuda I, Frenkel-Morgenstern M, and Hou YM (2015b). The UGG Isoacceptor of tRNAPro Is Naturally Prone to Frameshifts. Int J Mol Sci 16, 14866–14883. [PubMed: 26140378]

Gamper, H. B., Masuda, I., Frenkel-Morgenstern, M., & Hou, Y. M. (2015). The UGG Isoacceptor of tRNAPro Is Naturally Prone to Frameshifts. *International journal of molecular sciences*, *16*(7), 14866–14883. https://doi.org/10.3390/ijms160714866

Gamper, H. B., Masuda, I., Frenkel-Morgenstern, M., & Hou, Y. M. (2015). Maintenance of protein synthesis reading frame by EF-P and m(1)G37-tRNA. *Nature communications*, *6*, 7226. https://doi.org/10.1038/ncomms8226

Gerdes, S. Y., Scholle, M. D., Campbell, J. W., Balázsi, G., Ravasz, E., Daugherty, M. D., Somera, A. L., Kyrpides, N. C., Anderson, I., Gelfand, M. S., Bhattacharya, A., Kapatral, V., D'Souza, M., Baev, M. V., Grechkin, Y., Mseeh, F., Fonstein, M. Y., Overbeek, R., Barabási, A. L., Oltvai, Z. N., … Osterman, A. L. (2003). Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. *Journal of bacteriology*, *185*(19), 5673–5684. https://doi.org/10.1128/jb.185.19.5673-5684.2003

Gesheva, V., Ivanova, V., & Gesheva, R. (2005). Effects of nutrients on the production of AK-111-81 macrolide antibiotic by *Streptomyces* hygroscopicus. *Microbiological Research*, *160*(3), 243–248. https://doi.org/10.1016/j.micres.2004.06.005

Goto-Ito, S., Ito, T., Ishii, R., Muto, Y., Bessho, Y., & Yokoyama, S. (2008). Crystal structure of archaeal tRNA(m(1)G37)methyltransferase aTrm5. *Proteins*, *72*(4), 1274–1289. https://doi.org/10.1002/prot.22019

Guengerich F. P. (2021). Inhibition of Cytochrome P450 Enzymes by Drugs-Molecular Basis and Practical Applications. Biomolecules & therapeutics, *10.4062/biomolther.2021.102. ADvance online publication.* https://doi.org/10.4062/biomolther.2021.102

Habeeb Rasool, K., Hammood Hussein, N., & Mahamed Taha, B. (2020). Molecular Zainal Abidin, Z. A., Abdul Malek, N., Zainuddin, Z., & Chowdhury, A. J. K. (2016). Selective isolation and antagonistic activity of actinomycetes from mangrove forest of Pahang, Malaysia. *Frontiers in Life Science*, *9*(1), 24–31. https://doi.org/10.1080/21553769.2015.1051244

Hamza, A., Wei, N. N., & Zhan, C. G. (2012). Ligand-based virtual screening approach using a new scoring function. *Journal of chemical information an*D *mo*D*eling, 52(4),* 963–974. https://doi.org/10.1021/ci200617d

Hasani, A., Kariminik, A., & Isaazadeh, K. (2014). Streptomycetes : Characteristics and Their Antimicrobial Activities. *International Journal of A*D*vance*D *Biological an*D *Biome*D*ical Research*, *2*(1), 63–75.

Hein, I., Flekna, G., Krassnig, M., & Wagner, M. (2006). Real-time PCR for the detection of Salmonella spp. in food: An alternative approach to a conventional PCR system suggested by the FOOD-PCR project. *Journal of microbiological metho*D*s*, *66*(3), 538–547. https://doi.org/10.1016/j.mimet.2006.02.008

Hoffmann, M., Zhao, S., Pettengill, J., Luo, Y., Monday, S. R., Abbott, J., Ayers, S. L., Cinar, H. N., Muruvanda, T., Li, C., Allard, M. W., Whichard, J., Meng, J., Brown, E. W., & McDermott, P. F. (2014). Comparative genomic analysis and virulence differences in closely related salmonella enterica serotype heidelberg isolates from humans, retail meats, and animals. *Genome biology an*D *evolution*, *6*(5), 1046–1068. https://doi.org/10.1093/gbe/evu079

Höltje J. V. (1998). Growth of the stress-bearing and shape-maintaining murein sacculus of *Escherichia coli*. *Microbiology an*D *molecular biology reviews : MMBR*, *62*(1), 181–203. https://doi.org/10.1128/MMBR.62.1.181-203.1998

Hopkins, K. L., Day, M., & Threlfall, E. J. (2008). Plasmid-mediated quinolone resistance in Salmonella enterica, United Kingdom. *Emerging infectious* D*iseases*, *14*(2), 340–342. https://doi.org/10.3201/eid1402.070573

Hou, Y. M., Matsubara, R., Takase, R., Masuda, I., & Sulkowska, J. I. (2017). *Trm*D: A Methyl Transferase for tRNA Methylation With m^1G37. *The Enzymes*, *41*, 89–115. https://doi.org/10.1016/bs.enz.2017.03.003

Hou, Y. M., Matsubara, R., Takase, R., Masuda, I., & Sulkowska, J. I. (2017). *Trm*D: A Methyl Transferase for tRNA Methylation With m^1G37. *The Enzymes*, *41*, 89–115. https://doi.org/10.1016/bs.enz.2017.03.003

Houston, D. R., & Walkinshaw, M. D. (2013). Consensus docking: improving the reliability of docking in a virtual screening context. *Journal of chemical information an*D *mo*D*eling, 53(2),* 384–390. https://doi.org/10.1021/ci300399w https://doi.org/10.1002/anie.200802019

Hughes, J. D., Blagg, J., Price, D. A., Bailey, S., Decrescenzo, G. A., Devraj, R. V., Ellsworth, E., Fobian, Y. M., Gibbs, M. E., Gilles, R. W., Greene, N., Huang, E., Krieger-Burke, T., Loesel, J., Wager, T., Whiteley, L., & Zhang, Y. (2008). Physiochemical drug properties associated with in vivo toxicological outcomes. *Bioorganic & me*D*icinal chemistry letters, 18(17),* 4872–4875. https://doi.org/10.1016/j.bmcl.2008.07.071

Ito, T., Masuda, I., Yoshida, K., Goto-Ito, S., Sekine, S., Suh, S. W., Hou, Y. M., & Yokoyama, S. (2015). Structural basis for methyl-donor-dependent and sequence-specific binding to tRNA substrates by knotted methyltransferase *Trm*D. *Proceed*D*ings of the National Aca*D*emy of Sciences of the Unite*D *States of America*, *112*(31), E4197–E4205. https://doi.org/10.1073/pnas.1422981112

Jacoby, G. A., Strahilevitz, J., & Hooper, D. C. (2014). Plasmid-mediated quinolone resistance. *Microbiology spectrum*, *2*(5), 10.1128/microbiolspec.PLAS-0006-2013. https://doi.org/10.1128/microbiolspec.PLAS-0006-2013 Schöller, C. E. G., Gürtler, H., Pedersen, R., Molin, S., & Wilkins, K. (2002). Volatile metabolites from actinomycetes. *Journal of Agricultural an*D *Foo*D *Chemistry*, *50*(9), 2615–2621. https://doi.org/10.1021/jf0116754

Jegerschöld, C., Pawelzik, S. C., Purhonen, P., Bhakat, P., Gheorghe, K. R., Gyobu, N., Mitsuoka, K., Morgenstern, R., Jakobsson, P. J., & Hebert, H. (2008). Structural basis for induced formation of the inflammatory mediator prostaglandin E2. *Proceed*D*ings of the National Aca*D*emy of Sciences of the Unite*D *States of America, 105(32),* 11110–11115. https://doi.org/10.1073/pnas.0802894105

John, S., Thangapandian, S., Sakkiah, S., & Lee, K. W. (2011). Discovery of potential pancreatic cholesterol esterase inhibitors using pharmacophore modelling, virtual screening, and optimization studies. *Journal of enzyme inhibition an*D *me*D*icinal chemistry, 26(4),* 535–545. https://doi.org/10.3109/14756366.2010.535795

Jørgensen, F., & Kurland, C. G. (1990). Processivity errors of gene expression in *Escherichia coli*. *Journal of molecular biology*, *215*(4), 511–521. https://doi.org/10.1016/S0022-2836(05)80164-0

Kalid, O., Toledo Warshaviak, D., Shechter, S., Sherman, W., & Shacham, S. (2012). Consensus Induced Fit Docking (cIFD): methodology, validation, and application to the discovery of novel Crm1 inhibitors. *Journal of computer-ai*D*e*D *molecular* D*esign, 26(11),* 1217–1228. https://doi.org/10.1007/s10822-012-9611-9

Kalliokoski, T., Salo, H. S., Lahtela-Kakkonen, M., & Poso, A. (2009). The effect of ligand-based tautomer and protomer prediction on structure-based virtual screening. *Journal of chemical information an*D *mo*D*eling, 49(12),* 2742–2748. https://doi.org/10.1021/ci900364w

Kim, H. S., Nagore, D., & Nikaido, H. (2010). Multidrug efflux pump MdtBC of *Escherichia coli* is active only as a B2C heterotrimer. *Journal of bacteriology*, *192*(5), 1377–1386. https://doi.org/10.1128/JB.01448-09

Kim, J., Xiao, H., Bonanno, J. B., Kalyanaraman, C., Brown, S., Tang, X., … Almo, S. C. (2013). Structure-guided discovery of the metabolite carboxy-SAM that modulates tRNA function. *Nature*, *498*(7452), 123–126. https://doi.org/10.1038/nature12180

Köppen H. (2009). Virtual screening - what does it give us?. Current opinion in drug discovery & development, 12(3), 397–407.

Korb, O., Olsson, T. S., Bowden, S. J., Hall, R. J., Verdonk, M. L., Liebeschuetz, J. W., & Cole, J. C. (2012). Potential and limitations of ensemble docking. *Journal of chemical information an*D *mo*D*eling, 52(5),* 1262–1274. https://doi.org/10.1021/ci2005934

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

Lahoud, G., Goto-Ito, S., Yoshida, K., Ito, T., Yokoyama, S., & Hou, Y. M. (2011). Differentiating analogous tRNA methyltransferases by fragments of the methyl donor. *RNA (New York, N.Y.)*, *17*(7), 1236–1246. https://doi.org/10.1261/rna.2706011

Lavecchia, A., & Di Giovanni, C. (2013). Virtual screening strategies in drug discovery: a critical review. *Current meDicinal chemistry, 20(23),* 2839–2860. https://doi.org/10.2174/09298673113209990001

Le Guilloux, V., Schmidtke, P., & Tuffery, P. (2009). Fpocket: an open source platform for ligand pocket detection. *BMC bioinformatics, 10,* 168. https://doi.org/10.1186/1471-2105-10-168

Lee, S. J., Usmani, K. A., Chanas, B., Ghanayem, B., Xi, T., Hodgson, E., Mohrenweiser, H. W., & Goldstein, J. A. (2003). Genetic findings and functional studies of human CYP3A5 single nucleotide polymorphisms in different ethnic groups. *Pharmacogenetics, 13(8),* 461–472. https://doi.org/10.1097/00008571-200308000-00004

Leelananda, S. P., & Lindert, S. (2016). Computational methods in drug discovery. *Beilstein journal of organic chemistry, 12,* 2694–2718. https://doi.org/10.3762/bjoc.12.267

Li, H., Liu, S. M., Yu, X. H., Tang, S. L., & Tang, C. K. (2020). Coronavirus disease 2019 (COVID-19): current status and future perspectives. *International journal of antimicrobial agents, 55(5),* 105951. https://doi.org/10.1016/j.ijantimicag.2020.105951

Lionta, E., Spyrou, G., Vassilatis, D. K., & Cournia, Z. (2014). Structure-based virtual screening for drug discovery: principles, applications and recent advances. *Current topics in meDicinal chemistry, 14(16),* 1923–1938. https://doi.org/10.2174/1568026614666140929124445

Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced drug delivery reviews, 46(1-3), 3–26. https://doi.org/10.1016/s0169-409x(00)00129-0

Lucas, X., Senger, C., Erxleben, A., Grüning, B. A., Döring, K., Mosch, J., … Günther, S. (2013). StreptomeDB: A resource for natural compounds isolated from *Streptomyces* species. *Nucleic AciDs Research*, *41*(D1). https://doi.org/10.1093/nar/gks1253

Madhusudhan, M. S., Webb, B. M., Marti-Renom, M. A., Eswar, N., & Sali, A. (2009). Alignment of multiple protein structures based on sequence and structure features. *Protein engineering, Design & selection : PEDS*, *22*(9), 569–574. https://doi.org/10.1093/protein/gzp040

Maia, E., Assis, L. C., de Oliveira, T. A., da Silva, A. M., & Taranto, A. G. (2020). Structure-Based Virtual Screening: From Classical to Artificial Intelligence. *Frontiers in chemistry,* 8, 343. https://doi.org/10.3389/fchem.2020.00343

Marti-Renom, M. A., Madhusudhan, M. S., Fiser, A., Rost, B., & Sali, A. (2002). Reliability of assessment of protein structure prediction methods. *Structure (LonDon, EnglanD : 1993)*, *10*(3), 435–440. https://doi.org/10.1016/s0969-2126(02)00731-1

Martí-Renom, M. A., Stuart, A. C., Fiser, A., Sánchez, R., Melo, F., & Sali, A. (2000). Comparative protein structure modeling of genes and genomes. *Annual review of biophysics anD biomolecular structure*, *29*, 291–325. https://doi.org/10.1146/annurev.biophys.29.1.291

Martí-Renom, M. A., Stuart, A. C., Fiser, A., Sánchez, R., Melo, F., & Sali, A. (2000). Comparative protein structure modeling of genes and genomes. *Annual review of biophysics anD biomolecular structure*, *29*, 291–325. https://doi.org/10.1146/annurev.biophys.29.1.291

Masuda, I., Matsubara, R., Christian, T., Rojas, E. R., Yadavalli, S. S., Zhang, L., Goulian, M., Foster, L. J., Huang, K. C., & Hou, Y. M. (2019). tRNA Methylation Is a Global

Determinant of Bacterial Multi-drug Resistance. *Cell systems*, *8*(4), 302–314.e8. https://doi.org/10.1016/j.cels.2019.03.008

Masuda, I., Takase, R., Matsubara, R., Paulines, M. J., Gamper, H., Limbach, P. A., & Hou, Y. M. (2018). Selective terminal methylation of a tRNA wobble base. *Nucleic aci*D*s research*, *46*(7), e37. https://doi.org/10.1093/nar/gky013

Materi, W., & Wishart, D. S. (2007). Computational systems biology in drug discovery and development: methods and applications. D*rug* D*iscovery to*D*ay*, *12*(7-8), 295–303. https://doi.org/10.1016/j.drudis.2007.02.013

McDermott, P. F., Tyson, G. H., Kabera, C., Chen, Y., Li, C., Folster, J. P., Ayers, S. L., Lam, C., Tate, H. P., & Zhao, S. (2016). Whole-Genome Sequencing for Detecting Antimicrobial Resistance in Nontyphoidal Salmonella. *Antimicrobial agents an*D *chemotherapy*, *60*(9), 5515–5520. https://doi.org/10.1128/AAC.01030-16

Mcfadden, J., & Al-Khalili, J. (1999). A quantum mechanical model of adaptive mutation. *BioSystems, 50,* 203–211.

Menchon, G., Maveyraud, L., & Czaplicki, G. (2018). Molecular Dynamics as a Tool for Virtual Ligand Screening. *Metho*D*s in molecular biology (Clifton, N.J.), 1762,* 145–178. https://doi.org/10.1007/978-1-4939-7756-7_9

Meng, X. Y., Zhang, H. X., Mezei, M., & Cui, M. (2011). Molecular docking: a powerful approach for structure-based drug discovery. *Current computer-ai*D*e*D D*rug* D*esign, 7(2),* 146–157. https://doi.org/10.2174/157340911795677602

Michel, J., Tirado-Rives, J., & Jorgensen, W. L. (2009). Prediction of the water content in protein binding sites. *The journal of physical chemistry. B, 113(40),* 13337–13346. https://doi.org/10.1021/jp9047456

Misura, K. M., Chivian, D., Rohl, C. A., Kim, D. E., & Baker, D. (2006). Physically realistic homology models built with ROSETTA can be more accurate than their templates. *Proceed*D*ings of the National Aca*D*emy of Sciences of the Unite*D *States of America*, *103*(14), 5361–5366. https://doi.org/10.1073/pnas.0509355103

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of computational chemistry, 30(16),* 2785–2791. https://doi.org/10.1002/jcc.21256

Murakami, S., Nakashima, R., Yamashita, E., Matsumoto, T., & Yamaguchi, A. (2006). Crystal structures of a multidrug transporter reveal a functionally rotating mechanism. *Nature*, *443*(7108), 173–179. https://doi.org/10.1038/nature05076

Ngan, C. H., Bohnuud, T., Mottarella, S. E., Beglov, D., Villar, E. A., Hall, D. R., Kozakov, D., & Vajda, S. (2012). FTMAP: extended protein mapping with user-selected probe molecules. *Nucleic aci*D*s research*, *40*(Web Server issue), W271–W275. https://doi.org/10.1093/nar/gks441

Nikaido H. (1998). Antibiotic resistance caused by gram-negative multidrug efflux pumps. *Clinical infectious* D*iseases : an official publication of the Infectious* D*iseases Society of America*, *27 Suppl 1*, S32–S41. https://doi.org/10.1086/514920

Nunes, R. R., Fonseca, A., Pinto, A., Maia, E., Silva, A., Varotti, F. P., & Taranto, A. G. (2019). Brazilian malaria molecular targets (BraMMT): selected receptors for virtual high-throughput screening experiments. *Memorias* D*o Instituto Oswal*D*o Cruz*, *114*, e180465. https://doi.org/10.1590/0074-02760180465

Ohnishi, Y., Ishikawa, J., Hara, H., Suzuki, H., Ikenoya, M., Ikeda, H., … Horinouchi, S. (2008). Genome sequence of the streptomycin-producing microorganism *Streptomyces* griseus IFO 13350. *Journal of Bacteriology*, *190*(11), 4050–4060. https://doi.org/10.1128/JB.00204-08

Osguthorpe, D. J., Sherman, W., & Hagler, A. T. (2012). Generation of receptor structural ensembles for virtual screening using binding site shape analysis and clustering. *Chemical biology & D*r*ug D*e*sign, 80(2),* 182–193. https://doi.org/10.1111/j.1747-0285.2012.01396.x

Petrey, D., & Honig, B. (2005). Protein structure prediction: inroads to biology. *Molecular cell*, *20*(6), 811–819. https://doi.org/10.1016/j.molcel.2005.12.005

Prestinaci, F., Pezzotti, P., & Pantosti, A. (2015). Antimicrobial resistance: a global multifaceted phenomenon. *Pathogens anD global health*, *109*(7), 309–318. https://doi.org/10.1179/2047773215Y.0000000030

Rai, B. K., Tawa, G. J., Katz, A. H., & Humblet, C. (2010). Modeling G protein-coupled receptors for structure-based drug discovery using low-frequency normal modes for refinement of homology models: application to H3 antagonists. *Proteins, 78(2),* 457–473. https://doi.org/10.1002/prot.22571

Rapino, F., Delaunay, S., Rambow, F., Zhou, Z., Tharun, L., De Tullio, P., Sin, O., Shostak, K., Schmitz, S., Piepers, J., Ghesquière, B., Karim, L., Charloteaux, B., Jamart, D., Florin, A., Lambert, C., Rorive, A., Jerusalem, G., Leucci, E., Dewaele, M., … Close, P. (2018). Codon-specific translation reprogramming promotes resistance to targeted therapy. *Nature*, *558*(7711), 605–609. https://doi.org/10.1038/s41586-018-0243-7

Rarey, M., Kramer, B., Lengauer, T., & Klebe, G. (1996). A fast flexible docking method using an incremental construction algorithm. *Journal of molecular biology*, *261*(3), 470–489. https://doi.org/10.1006/jmbi.1996.0477

Reddy, A. S., Pati, S. P., Kumar, P. P., Pradeep, H. N., & Sastry, G. N. (2007). Virtual screening in drug discovery -- a computational perspective. *Current protein & peptiDe science*, *8*(4), 329–351. https://doi.org/10.2174/138920307781369427

Rojas, E. R., Billings, G., Odermatt, P. D., Auer, G. K., Zhu, L., Miguel, A., Chang, F., Weibel, D. B., Theriot, J. A., & Huang, K. C. (2018). The outer membrane is an essential load-bearing element in Gram-negative bacteria. *Nature*, *559*(7715), 617–621. https://doi.org/10.1038/s41586-018-0344-3

Ruiz, B., Chávez, A., Forero, A., García-Huante, Y., Romero, A., Snchez, M., … Langley, E. (2010). Production of microbial secondary metabolites: Regulation by the carbon source. *Critical Reviews in Microbiology*. https://doi.org/10.3109/10408410903489576

Sams-Dodd F. (2005). Target-based drug discovery: is something wrong?. D*rug* D*iscovery to*D*ay*, *10*(2), 139–147. https://doi.org/10.1016/S1359-6446(04)03316-1

Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R., & Sherman, W. (2013). Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of computer-ai*D*e*D *molecular* D*esign, 27(3),* 221–234. https://doi.org/10.1007/s10822-013-9644-8

Schmidt, B. J., Papin, J. A., & Musante, C. J. (2013). Mechanistic systems modeling to guide drug discovery and development. D*rug* D*iscovery to*D*ay*, *18*(3-4), 116–127. https://doi.org/10.1016/j.drudis.2012.09.003

Schmidt, T., Situ, A. J., & Ulmer, T. S. (2016). Structural and thermodynamic basis of proline-induced transmembrane complex stabilization. *Scientific reports*, *6*, 29809. https://doi.org/10.1038/srep29809

Schneider, N., Hindle, S., Lange, G., Klein, R., Albrecht, J., Briem, H., Beyer, K., Claußen, H., Gastreich, M., Lemmen, C., & Rarey, M. (2012). Substantial improvements in large-scale redocking and screening using the novel HYDE scoring function. *Journal of computer-ai*D*e*D *molecular* D*esign, 26(6),* 701–723. https://doi.org/10.1007/s10822-011-9531-0

Schöller, C. E. G., Gürtler, H., Pedersen, R., Molin, S., & Wilkins, K. (2002). Volatile metabolites from actinomycetes. *Journal of Agricultural an*D *Foo*D *Chemistry*, *50*(9), 2615–2621. https://doi.org/10.1021/jf0116754

Seco, J., Luque, F. J., & Barril, X. (2009). Binding site detection and druggability index from first principles. *Journal of me*D*icinal chemistry*, *52*(8), 2363–2371. https://doi.org/10.1021/jm801385d

Senn, H. M., & Thiel, W. (2009). QM/MM methods for biomolecular systems. *Angewan*D*te Chemie (International e*D*. in English), 48(7),* 1198–1229. https://doi.org/10.1002/anie.200802019

Shao, W., Li, X., Goraya, M. U., Wang, S., & Chen, J.-L. (2017). Evolution of Influenza A Virus by Mutation and Re-Assortment. *International Journal of Molecular Sciences, 18(8),* 1650. https://doi.org/10.3390/ijms18081650

Silver L. L. (2011). Challenges of antibacterial discovery. *Clinical microbiology reviews*, *24*(1), 71–109. https://doi.org/10.1128/CMR.00030-10

Sliwoski, G., Kothiwale, S., Meiler, J., & Lowe, E. W., Jr (2013). Computational methods in drug discovery. *Pharmacological reviews*, *66*(1), 334–395. https://doi.org/10.1124/pr.112.007336

Song, C. M., Lim, S. J., & Tong, J. C. (2009). Recent advances in computer-aided drug design. *Briefings in bioinformatics, 10(5),* 579–591. https://doi.org/10.1093/bib/bbp023

Stoltzfus, A., & Norris, R. W. (2016). On the Causes of Evolutionary Transition:Transversion Bias. *Molecular Biology an*D *Evolution, 33(3),* 595. https://doi.org/10.1093/MOLBEV/MSV274

Strippoli, P., Canaider, S., Noferini, F., D'Addabbo, P., Vitale, L., Facchin, F., Lenzi, L., Casadei, R., Carinci, P., Zannotti, M., & Frabetti, F. (2005). Uncertainty principle of genetic information in a living cell. *Theoretical Biology & Me*D*ical Mo*D*elling, 2,* 40. https://doi.org/10.1186/1742-4682-2-40

ten Brink, T., & Exner, T. E. (2010). pK(a) based protonation states and microspecies for protein-ligand docking. *Journal of computer-ai*D*e*D *molecular* D*esign, 24(11),* 935–942. https://doi.org/10.1007/s10822-010-9385-x

Threlfall E. J. (2002). Antimicrobial drug resistance in Salmonella: problems and perspectives in food- and water-borne infections. *FEMS microbiology reviews*, *26*(2), 141–148. https://doi.org/10.1111/j.1574-6976.2002.tb00606.x

Tresadern, G., Bemporad, D., & Howe, T. (2009). A comparison of ligand based virtual screening methods and application to corticotropin releasing factor 1 receptor. *Journal of molecular graphics & mo*D*elling, 27(8),* 860–870. https://doi.org/10.1016/j.jmgm.2009.01.003

Tseng, C. Y., & Tuszynski, J. (2015). A unified approach to computational drug discovery. D*rug* D*iscovery* *to*D*ay*, *20*(11), 1328–1336. https://doi.org/10.1016/j.drudis.2015.07.004

Typas, A., Banzhaf, M., Gross, C. A., & Vollmer, W. (2011). From the regulation of peptidoglycan synthesis to bacterial growth and morphology. *Nature reviews. Microbiology*, *10*(2), 123–136. https://doi.org/10.1038/nrmicro2677

Veber, D. F., Johnson, S. R., Cheng, H. Y., Smith, B. R., Ward, K. W., & Kopple, K. D. (2002). Molecular properties that influence the oral bioavailability of drug candidates. *Journal of me*D*icinal chemistry*, *45*(12), 2615–2623. https://doi.org/10.1021/jm020017n

Wajid, M., Awan, A. B., Saleemi, M. K., Weinreich, J., Schierack, P., Sarwar, Y., & Ali, A. (2019). Multiple Drug Resistance and Virulence Profiling of Salmonella enterica Serovars Typhimurium and Enteritidis from Poultry Farms of Faisalabad, Pakistan. *Microbial* D*rug resistance (Larchmont, N.Y.)*, *25*(1), 133–142. https://doi.org/10.1089/mdr.2018.0121

Wallner, B., & Elofsson, A. (2007). Prediction of global and local model quality in CASP7 using Pcons and ProQ. *Proteins*, *69 Suppl 8*, 184–193. https://doi.org/10.1002/prot.21774

Waszkowycz B. (2008). Towards improving compound selection in structure-based virtual screening. D*rug* D*iscovery* *to*D*ay, 13(5-6),* 219–226. https://doi.org/10.1016/j.drudis.2007.12.002

Wawrik, B., Kerkhof, L., Kukor, J., & Zylstra, G. (2005). Effect of different carbon sources on community composition of bacterial enrichments from soil. *Applie*D *an*D *Environmental Microbiology*, *71*(11), 6776–6783. https://doi.org/10.1128/AEM.71.11.6776-6783.2005

Webb, B., & Sali, A. (2016). Comparative Protein Structure Modeling Using MODELLER. *Current protocols in bioinformatics*, *54*, 5.6.1–5.6.37. https://doi.org/10.1002/cpbi.3

Webb, B., & Sali, A. (2016). Comparative Protein Structure Modeling Using MODELLER. *Current protocols in bioinformatics*, *54*, 5.6.1–5.6.37. https://doi.org/10.1002/cpbi.3

White, T. A., & Kell, D. B. (2004). Comparative genomic assessment of novel broad-spectrum targets for antibacterial drugs. *Comparative anD functional genomics*, *5*(4), 304–327. https://doi.org/10.1002/cfg.411

Yang, K., Bai, H., Ouyang, Q., Lai, L., & Tang, C. (2008). Finding multiple target optimal intervention in disease-related molecular network. *Molecular systems biology*, *4*, 228. https://doi.org/10.1038/msb.2008.60

Young, T., Abel, R., Kim, B., Berne, B. J., & Friesner, R. A. (2007). Motifs for molecular recognition exploiting hydrophobic enclosure in protein-ligand binding. Proceedings of the *National AcaDemy of Sciences of the UniteD States of America, 104(3),* 808–813. https://doi.org/10.1073/pnas.0610202104

9.APPENDICES

9.1 Media composition of ISP2 and ISP4

ISP2 growth media

Ingredients	Gms / Litre
Peptic digest of animal tissue	5.000
Yeast extract	3.000
Malt extract	3.000
Dextrose	10.000
Agar	20.000
Final pH (at 25°C)	6.2±0.2

ISP4 growth media

Ingredients	Gms / Litre
Starch, soluble	10.000
Dipotassium phosphate	1.000
Magnesium sulphate. heptahydrate	1.000
Sodium chloride	1.000
Ammonium sulphate	2.000
Calcium carbonate	2.000
Ferrous sulphate, heptahydrate	0.001
Manganous chloride, 7H2O	0.001

9.2 Data sources for metabolic model reconstruction and refinement

9.2.1 DNA sequence and genome annotation databases

EMBL	http://www.ebi.ac.uk/embl/	General nucleotide sequence database
GenBank	http://www.ncbi.nlm.nih.gov/Genbank/	General nucleotide sequence database
Integr8	http://www.ebi.ac.uk/integr8/	Integrated information on complete genomes
CMR	http://cmr.jcvi.org/	Integrated information on complete prokaryotic genomes
IMG	http://img.jgi.doe.gov/	Integrated system for analysis and annotation of microbial genomes
SEED	http://seed-viewer.theseed.org/	Integrated system for analysis and annotation of genomes using functional subsystems

9.2.2 Protein and enzyme databases

ENZYME	http://www.expasy.ch/enzyme/	Enzyme nomenclature database providing extensive information on all enzymes with an associated EC number
UniProt	http://www.ebi.ac.uk/uniprot/	Universal Protein Resource gathering protein sequences and annotations from SwissProt (manually reviewed), trEMBL (computer annotated), and PIR
TransportDB	http://www.membranetransport.org/	Predictions of membrane transport proteins for fully sequenced genomes
PSORTdb	http://db.psort.org/	Repository of experimentally determined and predicted protein localizations
Prolinks	http://prolinks.mbi.ucla.edu/	Database of predicted functional links between proteins

STRING	http://string.embl.de/	Database of known and predicted protein–protein interactions

9.2.3 Metabolic databases

CheBI	http://www.ebi.ac.uk/chebi/	Database on small molecules of biological interest
Pubchem	http://pubchem.ncbi.nlm.nih.gov/	Database on small molecules
LipidMaps	http://www.lipidmaps.org/	Database on lipid metabolites
Reactome	http://www.reactome.org/	Curated database of biological pathways
KEGG	http://www.genome.jp/kegg/	Suite of databases comprising information on compounds, reactions, pathways, genes/proteins
BioCyc	http://www.biocyc.org/	Collection of organism-specific pathway/genome databases, including a curated multiorganism pathway database: MetaCyc
UniPathway	http://www.grenoble.prabi.fr/obiwarehouse/unipathway/	Curated resource of metabolic pathways linked to UniProt enzyme database
UM-BBD	http://umbbd.msi.umn.edu/	Database on microbial biocatalytic reactions and biodegradation pathways

9.2.4 Experimental data repositories

IntAct	http://www.ebi.ac.uk/intact/	Repository of reported protein interactions
DIP	http://dip.doe-mbi.ucla.edu/	Database of experimentally determined interactions between proteins

Array Express	http://www.ebi.ac.uk/aerep/	Public repository of microarray data
GEO	http://www.ncbi.nlm.nih.gov/geo/	Public repository of microarray data
ASAP	http://asap.ahabs.wisc.edu/	Repository of results of functional genomics experiments for selected bacterial species
E. coli multi-omics DB	http://ecoli.iab.keio.ac.jp/	Comprehensive dataset of transcriptomic, proteomic, metabolomic, and fluxomic experiments for *E. coli* K12
Systomonas	http://www.systomonas.de/	Repository of 'omics' datasets and molecular networks for Pseudomonads species
PubMed	http://www.pubmed.org/	Database on biomedical literature

9.2.5 Metabolic Model Repositories

BiGG	http://bigg.ucsd.edu/	Repository of reconstructed genome-scale metabolic models
BioModels	http://www.ebi.ac.uk/biomodels/	Database of mathematical models of biological systems

9.3 Results from OSIRIS property explorer for those passing ADME/Tox filters

9.3.1 Natural Products

Preference Index	zinc_id	Molecule Name	Total Mol weight	cLogP	cLogS	H-Acceptors	H-Donors	Polar Surface Area	Druglikeness	Rotatable Bonds	Binding Energy with *trm*D	Binding Energy with hMAT 1A
2	ZINC000001590366	Antineoplaston A10	246	0.92	-2.09	5	2	75.27	3.5451	3	-8.3	-7.2
2.2	ZINC000100229736	Scopolamine	303	0.85	-2.92	5	1	62.3	2.875	5	-8	-7.3
4.6	SAM	SAM	399	-4.1	-3.8	11	5	182.63	-10.574	7	-7.9	-7.4

9.3.2 Kinase Inhibitors

Chemical ID	Molecule Name	Preference Index	Total Molweight	cLogP	cLogS	H-Acceptors	H-Donors	Polar Surface Area	Drug likeness	Rotatable Bonds	Binding Energy with *trm*D	Binding Energy with hMAT1A
SAM		**4.6**	**399.451**	**-4.06**	**-3.819**	**11**	**5**	**266.47**	**-10.574**	**7**	**-7.9**	**-7.4**
PB177950763 2	1-[(2S)-1-Hydroxy-3-phenylpropan-2-yl]-3-(1-pyridin-4-ylethyl)urea	2.8	299.373	1.36	-2.583	5	3	74.25	0.61	6	-8.9	-7.2
PB178072708 1	1-(1,5-Dimethylpyrazol-3-yl)-3-[(2S)-1-hydroxy-3-phenylpropan-2-yl]urea	2.8	288.35	1.81	-3.104	6	3	79.18	2.254	5	-8.7	-7.2
PB175636204 2	1-(1,3-Dimethylpyrazol-4-yl)-3-[(2R)-1-hydroxy-3-phenylpropan-2-yl]urea	2.8	288.35	0.83	-2.573	6	3	79.18	2.245	5	-8.9	-7

	2-(4-			3.	-							
PB	Fluorophenyl)-N-			0	3							
264	[2-(1H-indol-3-		296	8	.				1.0			
730	yl)ethyl]acetamid		.34	4	8			44.8	03			
52	e	2	4	9	6	3	2	9	3	5	-8.7	-6.5
					-							
				2.	3							
PB				9	.							
339	N-benzyl-2-(2-		278	5	8				2.8			
495	methyl-1H-indol-		.35	1	0			44.8	55			
502	3-yl)acetamide	1.8	4	9	2	3	2	9	6	4	-8.7	-7.3

9.3.3 Nucleoside mimetics

Chemical ID	Preference Index	Molecule Name	Total Molweight	cLogP	cLogS	H-Acceptors	H-Donors	Polar Surface Area	Drug likeness	Rotatable Bonds	Binding Energy with *trm*D	Binding Energy with hMAT1A
ZINC000421473	**4.68**		**399.47**	**-4.0.6**	**-3.18**	**11**	**5**	**266.47**	**-10.574**	**7**	**-7.9**	**-7.9**

		7-[(3R)-1-		2	-							
		Benzoylpyrrolidin-3-		.	3							
BDE		yl]-N-		5	.							
320		methylthieno[2,3-	36	7	5				6.			
089		b]pyrazine-6-	6.4	7	6			103	55			
64	2	carboxamide	44	4	4	6	1	.43	2	3	-10.3	-7.8
		[3-(2-Methylpropyl)-		2	-							
		1,2-oxazol-5-yl]-[(3S)-		.	3							
BDE		3-(1H-pyrazolo[3,4-		7	.							
331		b]pyridin-6-	35	3	6				6.			
857		yl)piperidin-1-	3.4	9	7			87.	19			
43	2.4	yl]methanone	25	3	4	7	1	91	92	4	-9	-7.8

9.3.4 Indole Derivatives

ID							H-					
N	Pref		Tot	c	c	H-	D	Pola	Dr	Rot	Bindin	Binding
U	ere		al	L	L	Ac	o	r	ugl	ata	g	energy
M	nce		Mol	o	o	ce	n	Surf	ike	ble	Energ	with
BE	Inde		wei	g	g	pt	or	ace	ne	Bon	y with	hMAT
R	x	Molecule Name	ght	P	S	ors	s	Area	ss	ds	*trm*D	1A
				2								
P0		N-[(3,4-		.	-							
95		dimethoxypheny		8	3							
-		l)methyl]-3-(1H-	338	6	.				0.5			
05		indol-6-	.40	8	7			63.3	61			
42	2.8	yl)propanamide	6	4	4	5	2	5	75	7	-8.1	-6.5
Y0		N-[2-(6-methoxy-										
20		1H-indol-3-		3	-							
-		yl)ethyl]-3-	322	.	3				1.2			
14		phenylpropanam	.40	3	.			54.1	99			
32	2.6	ide	7	6	8	4	2	2	1	7	-8.7	-6

				8	3								
				5	4								
		2-(4-		3	-								
Y0		fluorophenyl)-N-		.	3								
31		[2-(5-methoxy-		0	.								
-		1H-indol-3-		1	8					1.0			
22		yl)ethyl]acetami	326	4	7			54.1	66				
93	2.4	de	.37	9	8	4	2	2	2	6	-8.5	-7.3	

9.4 Protein Information

Protein ID	Protein name	Modellar Structure	Active site(chain B/D)	PyRx Coordinates (Center)	Dimension
POA873	*Trm*D (tRNA methyl transferase)		85 I 86 Y 87 L 88 S 89 P 90 Q 112 C 113 G 114 R 115 Y 116 E 117 G 118 V 131 W 132 S 133 I	X: 31.50 Y: 43.127 Z: 36.268	X: 28.490 Y: 25.613 Z: 21.514

134 G
135 D
136 Y
137 V
138 L
139 S
140 G
141 G
143 L
144P