

**T.R.**  
**GEBZE TECHNICAL UNIVERSITY**  
**GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES**

**COMPUTATIONAL CHARACTERIZATION OF LIGAND  
SPECIFICITY AND PROMISCUITY OF STAPHYLOCOCCUS  
AUREUS NORA EFFLUX PUMP**

**ESRA BÜŞRA IŞIK**  
**A THESIS SUBMITTED FOR**  
**MASTER OF SCIENCE**  
**DEPARTMENT OF BIOENGINEERING**  
**BIOINFORMATICS AND SYSTEMS BIOLOGY PROGRAM**

**GEBZE**  
**2023**

**T.R.**  
**GEBZE TECHNICAL UNIVERSITY**  
**GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES**

**COMPUTATIONAL CHARACTERIZATION  
OF LIGAND SPECIFICITY AND  
PROMISCUITY OF STAPHYLOCOCCUS  
AUREUS NORA EFFLUX PUMP**

**ESRA BÜŞRA IŞIK**  
**A THESIS SUBMITTED FOR  
MASTER OF SCIENCE  
DEPARTMENT OF BIOENGINEERING  
BIOINFORMATICS AND SYSTEMS BIOLOGY  
PROGRAM**

**THESIS SUPERVISOR**  
**ASSIST. PROF. DR. ONUR SERÇİNOĞLU**

**GEBZE**  
**2023**

GTÜ Fen Bilimleri Enstitüsü Yönetim Kurulu'nun 07/02/2023 tarih ve 2023/11 sayılı kararıyla oluşturulan jüri tarafından 16/02/2023 tarihinde tez savunma sınavı yapılan Esra Büşra Işık'ın tez çalışması Biyomühendislik Anabilim Dalı Biyoinformatik ve Sistem Biyolojisi Programında YÜKSEK LİSANS tezi olarak kabul edilmiştir.

**JÜRİ**

ÜYE

(TEZ DANIŞMANI) : Dr. Öğr. Üy. Onur SERÇİNOĞLU

ÜYE

: Dr. Öğr. Üy. Saliha Ece ACUNER ZORLUUYSAL

ÜYE

: Dr. Öğr. Üy. Pınar PİR

**ONAY**

Gebze Teknik Üniversitesi ..... Enstitüsü Yönetim Kurulu'nun  
...../...../..... tarih ve ...../..... sayılı kararı.

## SUMMARY

*Staphylococcus aureus* is a gram-positive bacterial pathogen and highly adaptive to environmental conditions and causes various disorders. Excessive usage of antibiotics may result in development of antibiotic resistance in *S. aureus*. A major antibiotic resistance mechanism is increase in the activity of transmembrane multi-drug efflux pump proteins. NorA is the most studied efflux pump in *S. aureus*, which belongs to the Major Facilitator Superfamily (MFS). NorA has been shown to contribute to resistance against a variety of small molecules, such as hydrophilic fluoroquinolone antibiotics, antiseptics, biocides, and dyes such as ethidium bromide. Moreover, a variety of small molecule inhibitors of NorA have been found. A remarkable feature of NorA inhibitors is their structural dissimilarity, which suggests promiscuous ligand recognition. Here, we aimed to characterize the structural aspects of ligand specificity and promiscuity of NorA by analyzing the binding behavior of NorA inhibitors into potential ligand binding sites within the internal cavity of the NorA protein structure. For this purpose, a list of known NorA inhibitors were docked into detected ligand binding pockets on an ensemble of modelled NorA structures generated via molecular dynamics simulations. A clustering analysis based on molecular similarity and binding fingerprints indicated that binding sites of chemically different ligands were consistent, in line with the promiscuous nature of NorA. On the other hand, residues 47F, 19I, 98R, 16F, 140F, 78F, 23I, 22V, 105A, and 44V were found to play a prominent role in ligand binding. Our findings may guide further research on discovery of novel NorA inhibitors.

**Key Words:** *Staphylococcus aureus*, Bacteria, Efflux pump, NorA, Antibiotic Resistance.

## ÖZET

*Staphylococcus aureus*, çevresel koşullara uyum kapasitesi yüksek ve çeşitli enfeksiyon hastalıklarına neden olan gram pozitif patojen bir bakteri türüdür. Bilinçsiz ve aşırı miktarlarda antibiyotik kullanımı *S. aureus* bakterisinde antibiyotik direncinin gelişmesine neden olabilmektedir. Antibiyotik direncinin kazanılmasında önemli rol oynayan mekanizmalardan biri, transmembran çoklu ilaç dışa atım pompası proteinlerinin aktivitesindeki artıştır. NorA, Majör Fasilitatör Süperfamilyası'na (MFS) mensup bir *S. Aureus* proteindir ve bu tür pompa proteinleri arasında en çok çalışılan dışa atım pompasıdır. NorA'nın hidrofilik florokinolon antibiyotikler, antiseptikler, biyositler ve etidyum bromür gibi boyaları da içeren çeşitli küçük moleküllere karşı *S. aureus*'un direnç kazanmasına katkıda bulunduğu gösterilmiştir. Ayrıca, NorA'nın çeşitli küçük molekül inhibitörleri de bulunmuştur. Bu NorA inhibitörlerinin dikkate değer bir özelliği, oldukça farklı yapılara sahip olmalarına rağmen NorA tarafından seçimsiz olarak tanınmalarıdır. Bu tez çalışmasında, NorA inhibitörlerinin NorA protein yapısının iç kavitesindeki potansiyel ligand bağlanma bölgelerine bağlanma davranışını analiz ederek, NorA'nın özgül/seçimsiz ligand tanımalarının yapısal yönlerini karakterize edilmesi amaçlanmıştır. Bu amaçla, öncelikle bilinen NorA inhibitörleri, moleküler dinamik simülasyonları yoluyla elde edilen NorA konformasyonlarında tespit edilen ligand bağlama ceplerine yerleştirilmiştir. Moleküler benzerliğe ve bağlanma bölgelerine dayanan bir kümeleme analizi, kimyasal olarak farklı ligandların bağlanma bölgelerinin büyük ölçüde benzerlik gösterdiğini ortaya koymuştur. Elde edilen sonuçlar, NorA yapısında bulunan 47F, 19I, 98R, 16F, 140F, 78F, 23I, 22V, 105A ve 44V amino asitlerinin ligand bağlanmasında önemli bir rol oynadığına da işaret etmektedir. Bu tez çalışmasında elde edilen bulgular, yeni NorA inhibitörlerinin keşfi üzerine yapılacak çalışmalara rehberlik edebilecektir.

**Anahtar Kelimeler:** *Staphylococcus aureus*, Bakteri, Dış Atım Pompası, NorA, Antibiyotik Direnci.

## ACKNOWLEDGEMENTS

I have started my MSc on Gebze Technical University in September 2020 and became a Research Assistant at Bezmialem Vakif University in January 2021. Becoming a researcher has been a blessing therefore I would like to offer my sincere thanks to all of my lecturers at Gebze Technical University, Tunahan Çakır, Pınar Pir, Saliha Durmuş and Onur Serçinoğlu. Their courses thought me a lot about the field bioinformatics, which is growing fast all around the world, and in Türkiye. I also would love to offer my special thanks to my PI at Bezmialem Vakif University, Mohammad Asif Khan, who has been a great boss during these two and a half years and offered his support whenever he can. My thesis supervisor, Onur Serçinoğlu has been a great instructor during the planning, execution, and finalization of this thesis, henceforth I would like to thank him for all his help and congratulate him for graduating his first master student.

I am hoping to be a great representative of the field of bioinformatics, with growing benefits to the society. The wishes only can come true with strong mindset and support from the right people. I would like to thank my family members, my friends, my colleagues, and my boyfriend for their pouring support during this time and hoping to reflect the effects of it by becoming a beneficial person to the society.

This project was supported by TUBITAK BİDEB 2210/A Scholarship. I would like to thank TUBITAK for selecting me to for this support. I would also like to acknowledge computing resources made available by ULAKBİM's TRGRID for MD simulations, TEI High Performance Computing Laboratory at GTU Faculty of Engineering for molecular docking simulations, and BVU Big Data and Bioinformatics Laboratory for data analysis.

With the hopes of a healed world.

# TABLE of CONTENTS

|                                                                     | <b><u>Page</u></b> |
|---------------------------------------------------------------------|--------------------|
| SUMMARY                                                             | v                  |
| ÖZET                                                                | vi                 |
| ACKNOWLEDGEMENTS                                                    | vii                |
| TABLE of CONTENTS                                                   | viii               |
| LIST OF ABBREVIATIONS AND ACRONYMS                                  | x                  |
| LIST OF FIGURES                                                     | xi                 |
| LIST OF TABLES                                                      | xii                |
| <br>                                                                |                    |
| 1. INTRODUCTION                                                     | 1                  |
| 1.1. Bacteria                                                       | 1                  |
| 1.2. Antibiotic Resistance                                          | 1                  |
| 1.3. Transport Proteins in Bacteria                                 | 2                  |
| 1.4. Mechanism of Transport of MFS                                  | 7                  |
| 1.5. Ligand Binding Specificity and Promiscuity of MFS Transporters | 10                 |
| 1.6. <i>Staphylococcus aureus</i>                                   | 11                 |
| 1.7. NorA Efflux Pump of <i>S. aureus</i>                           | 12                 |
| 1.8. Aim and Hypothesis                                             | 16                 |
| 2. MATERIALS AND METHODS                                            | 17                 |
| 2.1. Collection of NorA Inhibitors from ChEMBL Database             | 19                 |
| 2.2. Modeling of 3D Structure of NorA                               | 20                 |
| 2.3. Molecular Dynamics (MD) Simulations                            | 22                 |
| 2.4. Binding Pocket Detection from MD Simulation Trajectories       | 25                 |
| 2.5. Molecular Docking Simulations                                  | 26                 |
| 2.6. Data Analysis                                                  | 27                 |
| 2.6.1. Clustering Based on Pairwise Tanimoto Similarity Scores      | 27                 |
| 2.6.2. Clustering Based on Protein-Ligand Interaction Profiles      | 28                 |
| 2.6.3. Confusion Matrix                                             | 29                 |
| 3. RESULTS                                                          | 30                 |
| 3.1. Molecular Similarity Between NorA Inhibitors                   | 30                 |
| 3.2. Modeling of 3D Structure of NorA                               | 33                 |
| 3.3. MD Simulations                                                 | 35                 |

|      |                                                                  |    |
|------|------------------------------------------------------------------|----|
| 3.4. | Putative Binding Pockets in Inward-Facing Conformation of NorA   | 38 |
| 3.5. | Interaction Between Ciprofloxacin and NorA                       | 40 |
| 3.6. | Characterization of Interactions Between NorA and Its Inhibitors | 42 |
| 3.7. | Confusion Matrix                                                 | 48 |
| 4.   | DISCUSSION                                                       | 50 |
| 5.   | CONCLUSION                                                       | 54 |
|      | REFERENCES                                                       | 55 |
|      | BIOGRAPHY                                                        | 67 |
|      | APPENDICES                                                       | 68 |
|      | Appendix A: Publications Made in the Scope of Thesis Study       | 68 |
|      | Other Appendices (CD)                                            | 68 |

## LIST OF ABBREVIATIONS AND ACRONYMS

| <u>Abbreviations<br/>and Acronyms</u> | <u>Explanations</u>                                            |
|---------------------------------------|----------------------------------------------------------------|
| $\mu\text{m}$                         | : Micrometer                                                   |
| ATP                                   | : Adenosine Triphosphate                                       |
| CDS                                   | : Centers for Disease Control and Prevention                   |
| MATE                                  | : Multiple Antimicrobial Extrusion Protein Family              |
| MD                                    | : Molecular Dynamics                                           |
| MDR                                   | : Multiple Drug Resistant                                      |
| MFS                                   | : Major Facilitator Superfamily                                |
| MIC                                   | : Minimum Inhibitory Concentration                             |
| PDB                                   | : Protein Data Bank                                            |
| POPC                                  | : 1-palmitoyl-2-oleyl-glycerol-3-phospho-choline               |
| RMSD                                  | : Root Mean Square Deviation Standard                          |
| RMSF                                  | : Root Mean Square Fluctuation Standard                        |
| RND                                   | : Resistance-Nodulation-Cell Division Transporter Super Family |
| <i>S. aureus</i>                      | : <i>Staphylococcus aureus</i>                                 |
| SDR                                   | : Specific Drug Resistant                                      |
| SMR                                   | : Small Multidrug Resistant Transporter Family                 |
| SuMD                                  | : Supervised Molecular Dynamics                                |
| TM                                    | : Transmembrane                                                |

## LIST OF FIGURES

| <b><u>Figure No:</u></b>                                                                | <b><u>Page</u></b> |
|-----------------------------------------------------------------------------------------|--------------------|
| 1. 1: Superfamilies and families of solute transporters.                                | 4                  |
| 1. 2: Schematic illustration of the MFS-fold transporter topology.                      | 5                  |
| 1. 3: Motif A of MFS efflux pumps and their multiple sequence alignments.               | 6                  |
| 1. 4: Motif C of MFS efflux pump and their multiple sequence alignments.                | 7                  |
| 1. 5: Schematic highlighting the local substrate gating rearrangements.                 | 8                  |
| 1. 6: Three distinct models of SLC transporters for alternating-access.                 | 9                  |
| 1. 7: Outward open Cryo-EM structure of NorA Efflux pump with Fab25.                    | 15                 |
| 2. 1. Workflow of the study.                                                            | 18                 |
| 3. 1. Cluster map of NorA inhibitors based on Tanimoto similarity scores.               | 30                 |
| 3. 2. A model of NorA efflux pump inward-facing conformation.                           | 34                 |
| 3. 3. Ramachandran plot of the selected NorA model for MD simulation.                   | 34                 |
| 3. 4. RMSD scores obtained from both MD simulations.                                    | 36                 |
| 3. 5. RMSF plots obtained from simulations.                                             | 37                 |
| 3. 6 NorA homology model used in the study with fluctuated region.                      | 38                 |
| 3. 7. Putative ligand binding pockets.                                                  | 39                 |
| 3. 8. Ciprofloxacin binding residues highlighted on NorA efflux pump.                   | 41                 |
| 3. 9. Ciprofloxacin binding residues.                                                   | 41                 |
| 3. 10. Inhibitor binding residues highlighted on NorA efflux pump of <i>S. aureus</i> . | 43                 |
| 3. 11 Inhibitor binding residues.                                                       | 44                 |
| 3. 12. Heatmap of binding site clusters obtained from PLIP analysis.                    | 44                 |
| 3. 13. Ciprofloxacin complex with NorA Efflux Pump.                                     | 46                 |
| 3. 14. ChEMBL3774527 of Group 1 complex with NorA Efflux Pump.                          | 46                 |
| 3. 15. ChEMBL1649602 of Group 2 complex with NorA Efflux Pump.                          | 47                 |
| 3. 16. ChEMBL1650999 of Group 3 complex with NorA Efflux Pump.                          | 47                 |
| 3. 17. ChEMBL2158999 of Group 4 complex with NorA Efflux Pump.                          | 48                 |
| 3. 18. Confusion matrix.                                                                | 49                 |

## LIST OF TABLES

| <b><u>Table No:</u></b>                                       | <b><u>Page</u></b> |
|---------------------------------------------------------------|--------------------|
| 3. 1. NorA efflux pump inhibitor groups by Tanimoto clusters. | 32                 |
| 3. 2. NorA efflux pump inhibitor groups by PLIP clusters.     | 45                 |



# 1. INTRODUCTION

## 1.1. Bacteria

Bacterial species are single-celled microorganisms and can be found all over the world. They come in a huge variety of forms, sizes, and metabolic capacities, making them incredibly diverse. In addition to on and in the human body, bacteria can also be found in soil, water, and the air. They come in a variety of shapes, including rod-shaped, spherical, spiral, and even in clusters. They can be categorized according to their structure, staining propensities, nutrient needs, oxygen needs, and a variety of other factors. Bacteria are essential for keeping the ecosystem in balance, and the biotechnology sector relies on their genetic variety. Their effect on continuity of life is also significant considering their role as pharmaceuticals as they have been used for centuries to produce antibiotics, enzymes, hormones, vaccines, and they are also used in the development of new drugs. For example, *Staphylococcus aureus* produces active compounds against Gram-positive bacteria [1]–[4]. While certain bacteria can cause disease and hence, would be pathogenic, others can be helpful, such as those that aid in the breakdown of organic materials. Pathogenic bacterial species cause threat to humans and are major cause of infections and diseases, and eventually death [5]. A total of 10% of the emerging infections are caused by bacterial pathogens [6]. Emergence and re-emergence of the infectious agents, increased transmission with human migration and antibiotic resistance are the major factors contributing to disease formation [7].

## 1.2. Antibiotic Resistance

Antibiotics are the chemical molecules that are used for the treatment of bacterial diseases. They do this by modifying the functioning of proteins, or by interfering with the structure of nucleic acids, cell walls, and the metabolic route that bacteria use [8]. Excessive usage and abuse of antibiotics by humans and animals often lead to antibiotic resistance [9]. Pathogenic bacteria are considered to have developed resistance to all antibiotics discovered to date [10]. In 2019, antibiotic-resistant

bacteria affected 2.8 million individuals, and it caused the death of around 35,000 people in the United States, based on the reports of Centers for Disease Control and Prevention (CDC) [11]. This number is expected to be 10 million by 2050, causing 100 trillion USD [12].

Specific drug resistant (SDR) bacterial species, gain resistance against only one type of drug molecule, while multiple drug resistant (MDR) bacterial species can show resistance to variety of drug molecules. Resistance to antibiotics by the bacteria reduces the efficacy and efficiency of the existing drugs, therefore the treatment of disorders caused by MDR bacterial species has been a global issue [13].

Several mechanisms are utilized by bacteria to resist the action of antibiotics. Bacteria can develop enzymes that can inactivate antibiotics before they can cause any damage, which is known as enzymatic inactivation. Efflux pumps of the bacterial species can also pump antibiotics out of the cell membrane. This process may stop antibiotics from reaching the intended targets inside the cell. Bacteria also can alter the target sites of antibiotics, such as by altering the structure of a particular protein, preventing the antibiotics from interacting with the target site. Some bacteria have the capacity to alter the chemical structure of antibiotics to render them inactive. Bacteria can also have genetic alterations that give them antibiotic resistance. These mutations can occur in the efflux pump coding genes or in the target sites of antibiotics. Horizontal gene transfer can also aid in acquiring bacterial resistance genes from other bacteria. To transfer genetic material between bacteria, methods like plasmid exchange or transduction may be used. It is significant to note that some bacteria may possess numerous resistance mechanisms that are not exclusive of one another. A significant public health concern is the rise and spread of antibiotic resistant bacteria, which emphasizes the requirement for the creation of novel medicines and alternative therapies [13]. Therefore, research for production of new drugs or enhancing the activity of already existing drugs are carried out all over the world.

### **1.3. Transport Proteins in Bacteria**

Antibiotic accumulation inside bacterial cells may be prevented due to the action of transport proteins. These proteins are important for cellular homeostasis, growth, cellular response, defense against xenobiotics and cellular metabolism [14], [15]. Ion

channels and transporters are two types of proteins that take part in the flow of ions and other molecules across cell membranes. Transmembrane proteins named ion channels create a pore in the cell membrane that allows ions to pass through. Transmembrane proteins called transporters, also referred to as pumps, utilize energy and go through conformational changes to carry ions, peptides, and drugs across the cell membrane. Therefore, transporters are also termed drug efflux pumps [16], [17]

Drug efflux pumps are grouped into primary transporters that utilize ATP as energy source to transport drugs across membrane, and secondary active transporters that utilize energy of concentration gradient created by primary active transporters to transport molecules across membrane [18], [19]. Primary transporters obtain the energy required for transport from ATP hydrolysis, whereas in secondary transporters the required energy is obtained indirectly via carriage of substances by coupling with transport of another compound as the driving force [20], [21]. This compound can be an ion such as  $H^+$  or  $Na^+$  and is transported across cell membrane.

Secondary active transporters can be uniporters, which only transport a single substrate; antiporters, which carry the ion and substrate in the opposite directions; or symporters, that carry both in the same direction. Substrates as peptides, vitamins, amino acids, neurotransmitters, sugars are transported by secondary active transporters [22]. Resistance-nodulation-cell division transporter superfamily (RND), small multidrug resistant transporter family (SMR), multiple antimicrobial extrusion protein family (MATE), and major facilitator superfamily (MFS) are the four groups of bacterial secondary active transporters that transport antibiotics [23]. Figure 1.1 lists the families of transporters [24].

MFS is the largest family found in the secondary transporters. MFS proteins comprise approximately 25% of bacterial membrane transmembrane proteins. These transporters are found in the membranes of all the kingdoms of living organisms [25]. Transport Classification Database includes 89 families within this superfamily [26]. MFS transporters share common features as structural organization; as all consist of 12 transmembrane domains, and both N- and C- terminals facing cytoplasm; and mechanism of transport; they are facilitated diffusion transporters that transfer molecules across the membrane along their concentration gradient without energy. These transporters are also similar because they transport a wide variety of small molecules such as sugars, amino acids, and drugs, and are regulated by environmental changes as pH, temperature, and substrate concentration. They are involved in

antibiotic resistance, and they have high sequence diversity due to selective forces. MFS transporters regulate food absorption, osmoregulation, cell membrane ion gradients and many other crucial physiological processes. They are also involved in many diseases, including diabetes, cystic fibrosis, and cancer [24].

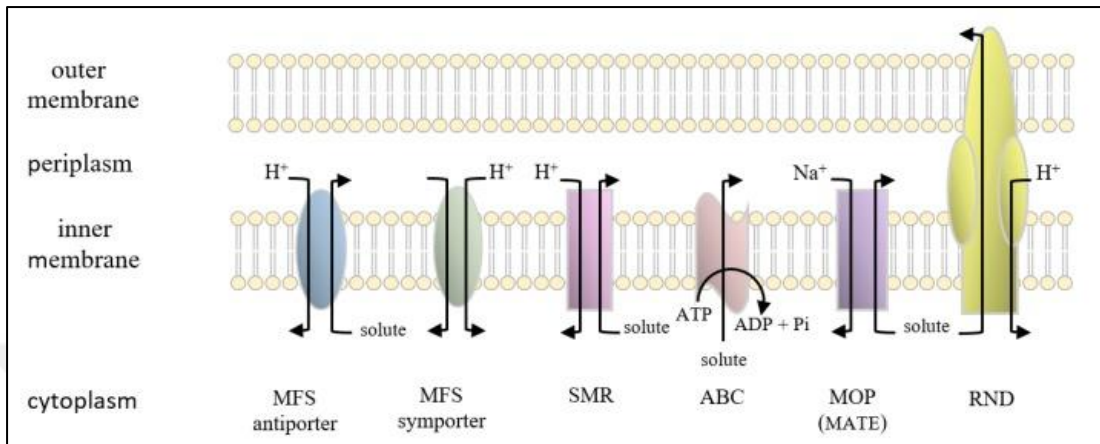


Figure 1. 1: Superfamilies and families of solute transporters. The outer and inner membranes of bacteria are illustrated with superfamilies and families of solute transporters. SMR is the small multi-drug resistant family; ABC is the ATP-binding cassette transporter superfamily; MOP is the multidrug/oligosaccharide-lipid/polysaccharide flippase superfamily; and RND is the resistance-nodulation-cell division superfamily. H<sup>+</sup> and Na<sup>+</sup> respectively represent protons and sodium ions.

These transporter proteins share protein-fold made up from transmembrane helices repeated for four times. The N-terminal domain is composed of transmembrane helices 1-3 and 4-6, while C-terminal domain includes transmembrane helices 7-9 and 10-12. Figure 1.2 shows a schematic illustration of the structure [27]. Transmembrane helices 1,4,7 and 9 are a part of substrate binding pocket while others take role in conformational changes and lipid bilayer interactions. The structure of the most MFS transporters is yet to be elucidated [28].

Short sequences of amino acids known as "amino acid sequence motifs" play a distinct structural or functional role in proteins. They are frequently located in areas such as the receptor binding site or the active site of an enzyme. Transmembrane domains as MFS also can include highly conserved amino acid sequence motifs. Motif A and Motif C are important for MFS transporters. Amino acids "G [X]3 D R/K X G R R/K" are found in Motif A and resides within cytoplasmic loop [29], [30]. These residues were found to be taking active role in forming components of transport pathway, regulating conformational changes and many other important functions [31]–

[36]. Motif C that includes amino acids “G [X]8 G [X]3 G P[X]2 G G” found in helix 5 [37]. The motif was found to be important for conferring antibiotic resistance [38]–[42]. These sequence motifs are also highlighted in Figure 1.3 and Figure 1.4 [24].

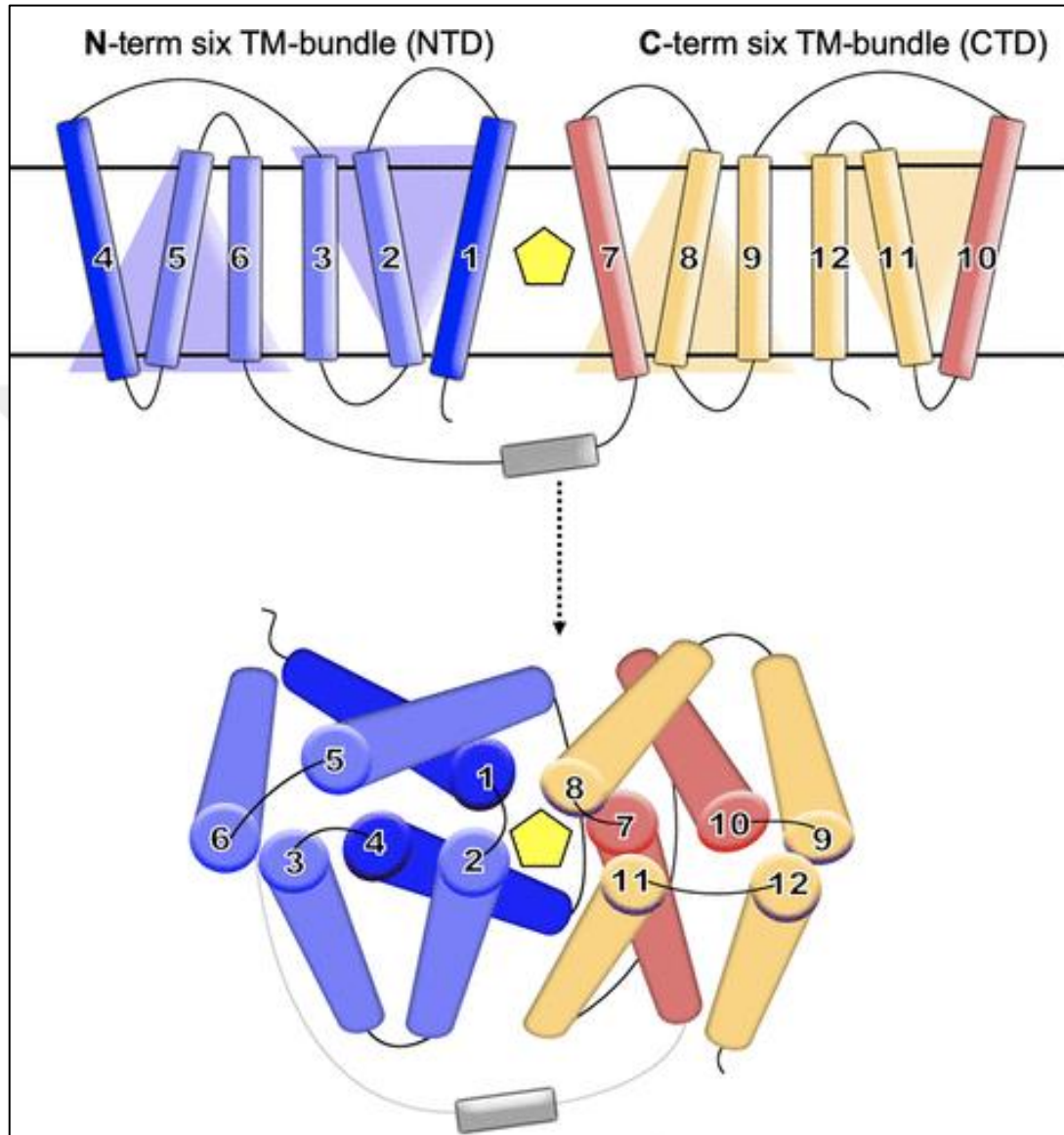


Figure 1. 2: Schematic illustration of the MFS-fold transporter topology. The N-terminal bundle (TM1–6; light blue/blue) and the C-terminal bundle (TMs 7–12; light orange/red) are joined by a cytosolic loop, which may contain structural features at times (gray).

| Efflux pump, Database identifier        | Motif A Sequences   |
|-----------------------------------------|---------------------|
| MdfA, P0AEY8                            | G p l s D r i G R r |
| tr   A8SPX9   A8SPX9_ECOLX              | G k r s D r f G R r |
| TetG, gi   13186195   gb   AAB24796.2   | G q l s D s y G R r |
| Q54806                                  | G n a a D r y G R r |
| Q50392                                  | G s l g D r l G R r |
| O52557                                  | G s f s D r y G R k |
| Bmr1,sp   P33449   BMR1_BACSU           | G r w v D r f G R k |
| ProP, sp   P0C0L7   PROP_ECOLI          | G n l g D k y G R q |
| ProP, sp   P0C0L7   PROP_ECOLI_2        | G l l s D r f G R r |
| YajR, sp   P77726   YAJR_ECOLI          | G l l s D r i G R k |
| TetA, P02982                            | G a l s D r f G R r |
| CmlA1, P32482                           | G p l s D r l G R r |
| LmrA, sp   P46104   LMRA_STRLN          | G r i a D l f G R r |
| LmrP, gi   15674147   ref   NP_268322.1 | G f f a D r n G R k |
| Bcr, P28246                             | G p m a D s f G R k |
| NorB, Q7A5M0                            | G g l a D k y G R i |
| gi   12044692   emb   CAC19830.1        | G r w s D k l G R r |
| Q48658                                  | G f f a D r n G R k |
| Bmr1, P33449                            | G r w v D r f G R k |
| EmrD, P31442                            | G p i s D r v G R r |
| TraI, P46104                            | G r i a D l f G R r |
| Bmr2, P39843                            | G r w v D r f G R k |
| ActVA1, Q53903                          | G v l g D r l G R r |

Figure 1. 3: Motif A of MFS efflux pumps and their multiple sequence alignments. Amino acids in the alignment (right), sequence database identifiers (left). The motif is found between the helix 2 and helix 3.

| Efflux pump, Database identifier         | Motif C (Antiporter Motif) Sequences |
|------------------------------------------|--------------------------------------|
| Tcr, P02983                              | G f i g s i v a l G e g l G P s i G  |
| tr   A8SPX9   A8SPX9_ECOLX               | G w l g a s f g l G l i a G P i i G  |
| TetG, gi   13186195   gb  <br>AAB24796.2 | G y m g a c y g a G n i a G P a l G  |
| gi   12044692   emb   CAC19830.1         | G r l g a a f g a G l i a G P a i G  |
| NorA, P0A0J7                             | G y m s a i i n s G f i l G P g i G  |
| Bmr1, P33449                             | G y m s a a i s t G f i i G P g i G  |
| Bmr1, P33449_2                           | G m n s m f t s i G n v f G P i i G  |
| RifP, O52557                             | G l w s a v s s f G y a s G P l i G  |
| Bmr1, sp   P33449   BMR1_BACSU           | G y m s a a i s t G f i i G P g i G  |
| Bmr1, sp   P33449  <br>BMR1_BACSU_2      | G m n s m f t s i G n v f G P i i G  |
| Bmr2, P39843                             | G y v s a a i s t G f i i G P g a G  |
| Bmr2, P39843_2                           | G m n s t y t s l G n i f G P a l G  |
| NorA, sp   P0A0J7   NORA_STAAU           | G y m s a i i n s G f i l G P g i G  |
| TetA, P02982                             | G f m s a c f g f G n v a G P v l G  |
| Tcrb, P23054                             | G l i g s l v a n G e g v G P a i G  |

Figure 1. 4: Motif C of MFS efflux pump and their multiple sequence alignments. Amino acids in alignment (right), sequence database identifiers (left). The motif is found in helix 5.

## 1.4. Mechanism of Transport of MFS

Uncovering the mechanism of transporter proteins important to understand antibiotic resistance in bacterial species because they take part in the extrusion of drugs and other toxic compounds. The molecules are transferred across the membrane from an area of higher concentration to of lower without the need of energy, via facilitated diffusion. MFS transporters are thought to work by a "substrate-binding protein" mechanism in which this site is exposed to only one side of the membrane at a time.

The conformational change may be caused by the binding of the substrate to the extracellular domain, which then triggers a change in the conformation of the transmembrane domains, allowing the substrate to cross the membrane. During this process, protein can go through various conformations as inward open, inward-

occluded, outward open, outward-occluded conformational states [43]–[50] as given in the Figure 1.5 [27].

The transformation between different states occurs with changes in rigid-body transitions and structural changes in transmembrane helices [51]. The "alternating-access model" can be defined by three models, according to molecular details: “rocker-switch”, “rocking-bundle”, or “elevator” mechanism demonstrated at Figure 1.6 [27]. Core substrate binding sites are rocked over by the C- and N-domains, according to rocker-switch paradigm. Changes in conformation from inner to outward and vice versa are caused by this rigid-body motion. For this transition, it is believed that salt bridges, motions on the A- and B- helices, and binding of substates are crucial [52], [53]. But it was noted that this model is insufficient on its own to account for the development of occluded states. Therefore, a new "clamp-and-switch model" was consequently suggested. Based on this hypothesis, during the binding of the substrate or while the substrate is being released, occluded states are created by pore-lining helices bend to create the occluded state [51].

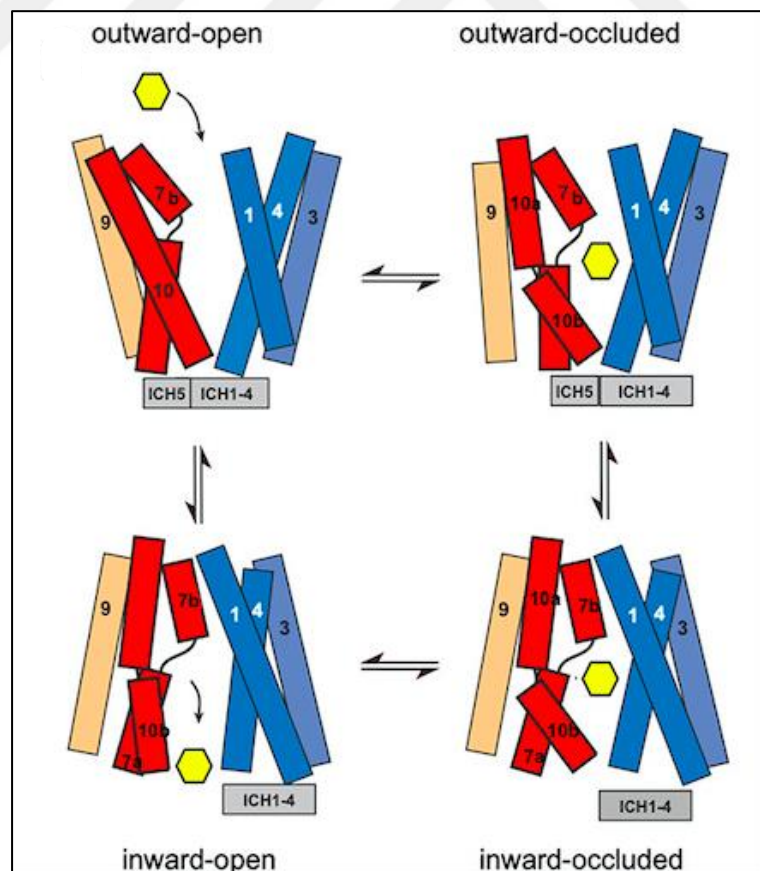


Figure 1. 5: Schematic highlighting the local substrate gating rearrangements.

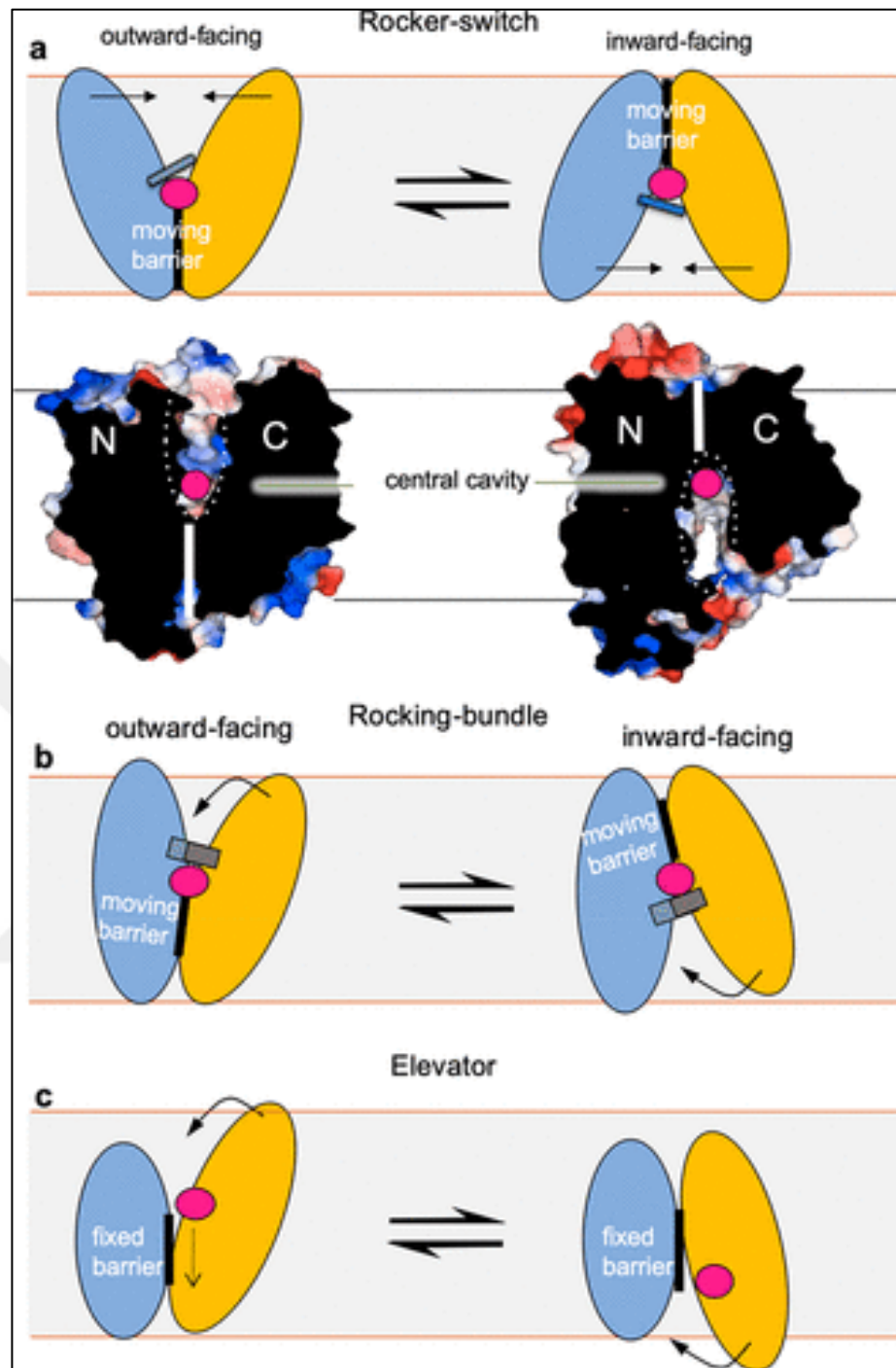


Figure 1. 6: Three distinct models of SLC transporters for alternating-access. Pink sphere is the substrate, N terminal is given as light blue and C terminal is given as light orange. (a) In the rocker-switch mechanism, structurally identical bundles N and C rearrange symmetrically around the core substrate-binding site to alternate access to the opposite side of the membrane. (a) The rocking-bundle mechanism alternates membrane access by reorganizing physically different bundles asymmetrically around the core substrate-binding site. (c) In the elevator mechanism, the two bundles are significantly divergent, and the C-terminal bundle moves the substrate across the membrane while the N-terminal bundle remains fixed, often due to oligomerization.

Using a combination of transporter structures in multiple functional states along with biochemical analysis, biophysical analysis, and in silico molecular dynamics (MD) calculations, researchers have been able to interpret more specific details of the transport mechanisms of different transporters over the past decade. However, more needs to be done to reveal these mechanisms.

## **1.5. Ligand Binding Specificity and Promiscuity of MFS Transporters**

The ability of MFS proteins in bacteria to specifically bind and transport particular ligands or substrates across the cell membrane is referred to as ligand binding specificity. Bacterial MFS transporters have been found to exhibit a high degree of promiscuity, which means they can move a wide range of chemically different substrates [54]. It is believed that this promiscuity serves as an adaptation to the various environments in which bacteria live. This enables bacteria to utilize various nutrition sources and adapt to various circumstances [31], [37], [55].

The highly conserved structural organization of MFS transporters enables them to have a broad range of substrate specificities and it is suggested to be the cause of their high level of promiscuity. It was also stated that there are two basic principles for MFS transporters [56]–[58]. According to the first principle, transporter has large and flexible binding site that recognizes various substrates. This structural flexibility also allows for changes in the protein's conformation, which can lead to new substrate specificities. The "non-consensus" residues that interact with the substrate in the binding pocket can increase the selectivity of substrate binding. The presence of other proteins, such as chaperones or other transporters, can also affect the specificity of MFS transporters [54], [59]–[66]. According to second principle, the interactions between substrate and transporter are very weak, in contrast to normal enzyme-ligand systems. However, the bond between substrate and transporter is quite particular and selective. The selectivity of MFS transporters can change between different bacteria and even between distinct MFS transporters in the same bacteria, which can have implications for the transport of different molecules and for the development of antibiotics targeting these transporters [67].

It is believed that the promiscuity of MFS transporters contributes to antibiotic resistance in bacteria. It has been discovered that certain MFS transporters can pump

out a wide variety of antibiotics, making them MDR pumps. It is stated that bacteria could have developed this resistance mechanism to increase their chances of surviving in an environment where the concentration of antibiotics are high. This MDR feature, also named promiscuity, makes it difficult to understand binding characteristics of the transporters. These statements are used to explain the characteristics for many systems, however more work needs to be done to uncover exact mechanism of recognition, binding and subsequent transport of various substrates and the residues that take part in this process needs explanation [68]–[70].

## 1.6. *Staphylococcus aureus*

*Staphylococcus aureus* (*S. aureus*) is an aerobic to facultative anaerobic bacterial pathogen which has high degree of environmental adaptability. Because of its thick peptidoglycan cell wall, *S. aureus* is classified as a gram-positive bacteria, which also contributes to antibiotic resistance. They have spherical shape and their size change from 0.5 to 1.5  $\mu\text{m}$  and they can be found in nature in various combinations from single cells to clusters. They cannot move or form spores however some *S. aureus* may contain polysaccharide capsule [71].

Normally, they are found in several parts of the human body as gastrointestinal tract, skin, nares, and they share a benign relationship. The bacteria can be spread via contamination or individuals can carry the bacteria without showing any symptoms but can spread the bacteria to others. Various disorders, e.g., skin infections as boils, impetigo, cellulitis; respiratory infections as sinusitis and pneumonia; infection in bloodstream named bacteremia; severe form of bacteremia named septicemia, osteomyelitis, endocarditis, toxic shock syndrome, food poisoning, scalded skin syndrome are caused by *S. aureus* [72]–[75].

*S. aureus* has virulence factors that contributes to pathogenicity. These bacteria can produce toxins and enzymes that damage host tissues and evade the host's immune response. Their cell wall contains teichoic acids, assisting the bacteria in adhering to the host tissues. *S. aureus* can also produce catalase and coagulate enzymes. Catalase helps to protect the bacteria from the toxic effects of hydrogen peroxide while coagulase helps the bacteria to evade the host immune system by forming a protective clot around itself. The bacteria can also form biofilms that are encased in a protective

matrix and are resistant to antibiotics and host defenses [76]. Antibiotics are used for the treatment of *S. aureus* infections and are said to be effective; however, the development of resistance to these drugs were also a problem. They are resistant to several antibiotics as fluoroquinolones,  $\beta$ -lactamics, aminoglycosides, macrolides, lincosamides, chloramphenicol, sulfonamides, streptomycin, and tetracycline [77].

*S. aureus* can develop antibiotic resistance via several mechanisms. These mechanisms include production of enzymes known as beta-lactamases, which can degrade and render inactive specific classes of antibiotics like penicillin and cephalosporins. Another mechanism includes the mutation or acquisition of genes that cause resistance to antibiotics. Efflux pumps are also used as a resistance mechanism because these pumps can force drugs out of the bacteria cell, rendering them useless [78].

Since some strains of the bacteria have developed resistance to various medicines and can be challenging to treat, these mechanisms of *S. aureus* antibiotic resistance have grown to be a significant public health concern. By maintaining personal hygiene, including frequent hand washing, keeping wounds and bruises clean and covered, and avoiding direct contact with infected individuals, spread of *S. aureus* can be prevented. To prevent the emergence of antibiotic resistance, it is also crucial to use antibiotics carefully.

## **1.7. NorA Efflux Pump of *S. aureus***

In *S. aureus*, NorA efflux pump has the highest number of studies [79]. The gene encoding NorA is found to be one of the core genes in the *S. aureus* and it is found in the genome of most of the *S. aureus* strains identified to date [80]. This gene, norA, was identified in 1986 in the chromosome of *S. aureus* [81]. It is stated that three alleles of *S. aureus* include 10% difference in nucleotide sequences and 5% difference in protein sequences of NorA [82]–[84]. These alleles are referred as norAI, norAII and norAIII. Recent studies showed that overexpression or inhibition of norA expression is important in evolution of *S. aureus* strains. Variations of the expression norA gene effects the resistance and evolvability of the bacterial strain [85].

NorA is an antiport transporter that initiates the transport of antimicrobial substances across the membrane via proton motive force. NorA efflux pump

contributes to the resistance against hydrophilic fluoroquinolones (as ciprofloxacin), antiseptics, dyes as ethidium bromide, biocides in case of the overexpression of the protein [86]–[88]. The capability of the NorA to pump structurally and chemically different substrates make NorA a promiscuous transmembrane protein. NorA plays an important role in hastening the evolution of ciprofloxacin resistance. The protein either increases the resistance or enhances the advantages of traditional ciprofloxacin resistance mutations that affect DNA topoisomerase [85].

Several studies been conducted to understand and inhibition mechanism of NorA efflux pump. To understand efflux pump inhibition different molecules been tried on various strains of the *S. aureus* with several known substrates of the pump. These molecules include essential oils, petroleum extracts, dichloromethane extract [89], 1,8-naphthyridines [90], ketoconazole [91], menadione (vitamin K3) [92], riparins [73]. However, these compounds had high toxicity levels to be approved for clinical use. Therefore, in one study, clinically approved drugs as pimozone, maraviroc, nilotinib and naftifine were also tried and found to be potential efflux pump inhibitors (EPI) [93]. These molecules are found to be decreasing the minimal inhibitory concentration (MIC) of antibiotics, which are the substrates of the efflux pumps. MIC can be defined as the lowest concentration for the substrates, in this case the antibiotics, for which no growth of the bacterial cells was observed. Hence, by using these inhibitors, the level of bacterial growth is sought to be decreased.

The novel inhibitors were also generated and tried on NorA Efflux Pump. In one study, researchers used a combination of virtual screening, chemical synthesis, and experimental validation to identify and characterize several compounds that showed promising inhibitory activity against NorA. They also investigated the mechanisms by which the inhibitors worked and found that they appeared to act by blocking the activity of the transporter [94]. In another study, natural flavones nucleus was used as starting point and a series of chemical modifications was used to synthesize a library of compounds, which were then screened for activity against NorA. Several compounds showed promising inhibitory activity, and further modifications were made to improve their potency. The most promising compound was a 2-(4-propoxyphenyl) quinoline derivative, which was found to be a potent inhibitor of NorA and had good antimicrobial activity against *S. aureus* [95]. Chalcone derivatives were also tried against NorA, using both whole cells and enriched everted membrane vesicles. Several compounds showed inhibitory activity against the NorA pump, with

one compound, named CH3, being the most potent. Further experiments revealed that CH3 was able to significantly reduce the efflux of several NorA substrates, including ethidium bromide and ciprofloxacin, in both whole cells and enriched everted membrane vesicles. The authors also investigated the mechanism of action of CH3 and found that it appeared to act by inhibiting the energy required for efflux rather than by directly binding to the NorA protein [96]. In another study, researchers synthesized a series of 2,1-benzothiazine 2,2-dioxide derivatives and tested their biological activity against several bacterial strains, including *Staphylococcus aureus*. Several compounds showed good antibacterial activity, with one compound, named 14g, being the most potent against *S. aureus*. The 14g appeared to act by inhibiting the bacterial DNA gyrase enzyme, which is the same mechanism of action as quinolone antibiotics [97]. Indole derivatives were also synthesized, and their biological activity was tested against several bacterial strains. It was found that several of the compounds showed good antibacterial activity, with some showing even better activity than the antibiotics used as controls. Mechanism of action of the most promising compound was investigated and found that it appeared to act by inhibiting the bacterial DNA gyrase enzyme, which is the same mechanism of action as quinolone antibiotics [98]. In another study, researchers began by screening a library of indole-based compounds and identified a lead compound with moderate inhibitory activity against NorA. Through rational design and optimization, they were able to improve the potency of the lead compound by several orders of magnitude, resulting in a new compound with excellent activity against NorA and good antimicrobial activity against *S. aureus*. This compound appeared to act by blocking the activity of the NorA protein [99].

NorA contains 388 amino acids and belongs to MFS transporters and has 12-transmembrane (TM) domain and part of the drug:H<sup>+</sup> anti-porter subfamily (DHA12) [100], [101]. 12 hydrophobic TM -helices with N- and C-terminal domains are positioned in the cytoplasm, joined by hydrophilic loops, and organized with pseudo-twofold symmetry [102].

Before the structure of the NorA have been discovered experimentally, the studies included NorA protein structure mostly based on homology modeling. For instance, the molecular substrate recognition of NorA efflux pump were investigated by supervised molecular dynamics study by using ciprofloxacin as the substrate and homology models of NorA as protein structure. In the study, the ciprofloxacin interacted with various amino acids on the protein. These amino acids included 127K,

264K, 109M, 133S, 136I, 51Q, 310R, 222E, 303F, and 225Y. Considering the promiscuous property of the NorA, this pathway was suggested for ciprofloxacin and study suggested different pathways could be encountered in case of structurally different compounds [102].

In another study, the elements contributing the substrate binding to NorA efflux pump was highlighted using *in vivo* assays, using techniques as side directed mutagenesis and measuring the conformational changes by FRET and microscale thermophoresis analysis. In the study, researchers confirmed the regions that are conserved in MFS transporters, Motif-A, Motif-B and Motif-C that were in N-terminal, were also conserved in NorA. The Motif-A in NorA was found to be stabilizing the outward confirmation, Motif-B was found to play essential role in substrate efflux and Motif-C mediated the conformational change and mutations of this motif caused fluoroquinolone resistance. It was also reported that two more motifs Motif-1 and Motif-2 were identified in NorA homologues. These motifs were found to be important for stabilizing the outward confirmation of NorA. The study also highlighted the amino acids 16F, 222E, 303F and 307D are important for binding activity of NorA [103].

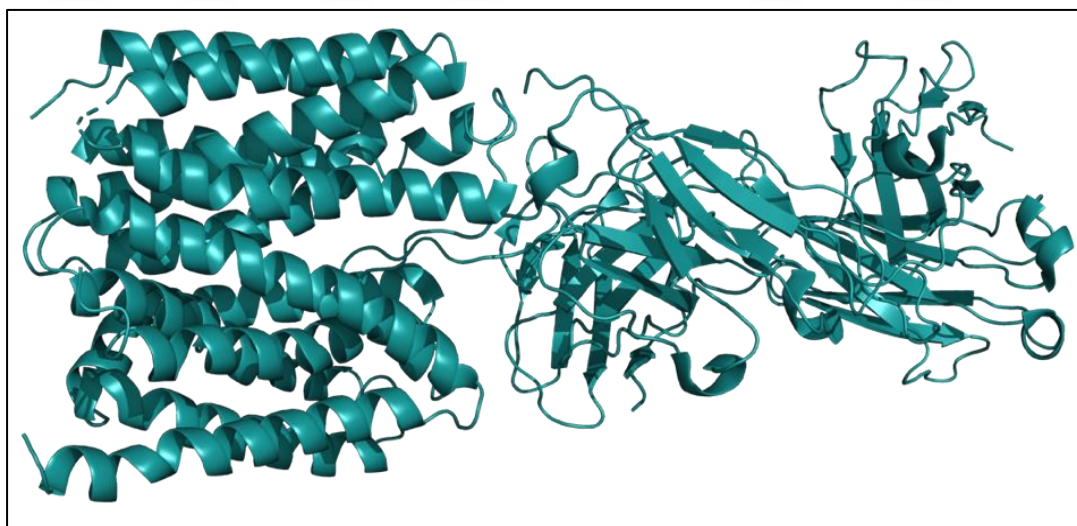


Figure 1. 7: Outward open Cryo-EM structure of NorA Efflux pump with Fab25 (PDB ID: 7LO7).

The cryo-electron microscopy (cryo-EM) mechanism was used in a recent study to establish structure of NorA efflux pump with *in vivo* drug resistance assays and NorA–Fab25 (EMD-23463; PDB ID: 7LO7) and NorA–Fab36 (EMD-23464; PDB

ID: 7LO8) structures were examined in the study. The protein was captured in outward-open confirmation, and it was found that 12 TM -helices make up NorA, which is made up of two 6-TM bundles (the N- and C-terminal domains) that cross the substrate-binding pocket. Through interactions between TM4, TM5, TM10, and TM11, the pocket is closed from the cytoplasmic side of the membrane. Interactions between TM2-TM11 and TM5-TM8 allow for communication between the N- and C-terminal domains. These TM pairs have distinctive "hourglass" forms, which probably aid conformational exchange via the "rocker-switch" mechanism [104].

The substrate binding pockets included amino acids 16F, 140F, 303F, 306F, 98R, 222E, 307D and 310R. The aromatics residues (F) are thought to be facilitating the binding to aromatic drugs as fluoroquinolones and other ionizable residues are important for binding to charged substrates. NorA also has MFS sequence motifs, such as Motif-A that is important for stability of the conformer and Motif-C that takes part in conformational exchange [104]. The Cryo-EM structure interacting with a compound named Fab25 used in the study (PDB ID: 7LO7) is given in the Figure 1.7 [105].

## **1.8. Aim and Hypothesis**

The aim of this thesis work is to characterize the ligand binding specificity/promiscuity of NorA efflux pump. Specifically, we aim to establish an understanding of how NorA efflux pump recognizes ligands that are chemically diverse by detecting the possible binding sites on the NorA efflux pump structure and exploring ligand binding behavior in these pockets via molecular docking simulations. We also intend to explore the relationship between the structure and sequence of NorA by identifying amino acids that play a significant role in ligand recognition.

## 2. MATERIALS AND METHODS

This thesis work aims to characterize ligand promiscuity/selectivity of *S. aureus* NorA efflux pump using computational structural biology methods. For this purpose, protein structure modeling, molecular dynamics, and molecular docking simulations have been performed. First, the 3D structure of NorA was generated using protein structure modeling methods. In particular, the inward-facing conformation of NorA was considered for modeling, as this is the conformation that recognizes and transports the substrates from the inside to the outside of the bacterial cell. The structural stability of the generated homology model was validated by performing classical molecular dynamics simulations. Next, multiple NorA conformations were selected from the generated simulation trajectories. Each of these conformations were used as receptors in subsequent molecular docking simulations in an ensemble docking context. Finally, molecular docking simulations were performed, and results were analyzed with respect to the predicted binding affinities and residue contact patterns of known inhibitors of NorA. The overall workflow used is shown in Figure 2.1. The procedure followed in each of the steps shown below is given in the following sections.

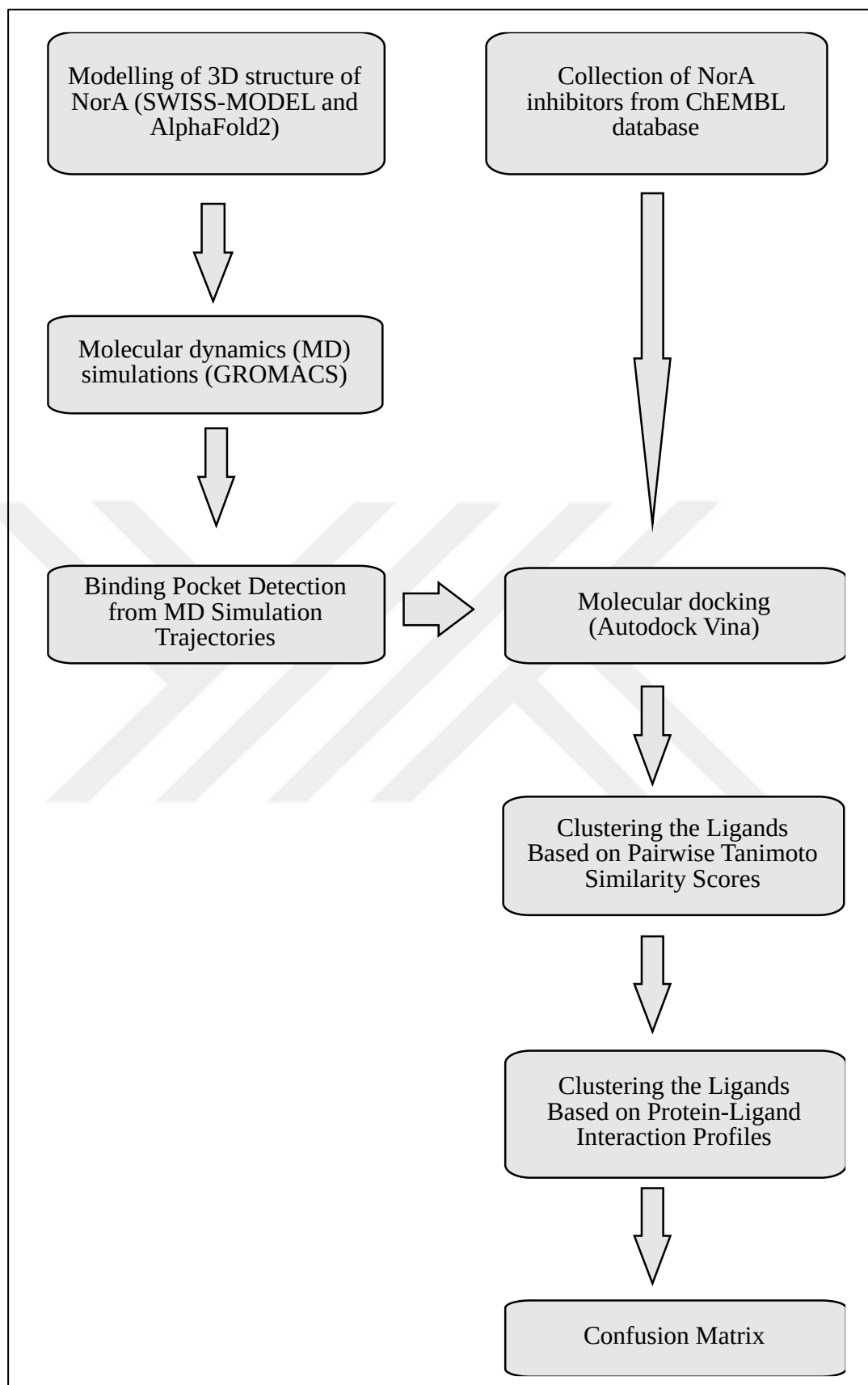


Figure 2. 1: Workflow of the study.

## 2.1. Collection of NorA Inhibitors from ChEMBL Database

The effectiveness of the potential inhibitors is measured with various assays, including the minimum inhibitory concentration (MIC) assay, used to determine the lowest concentration of a drug that can inhibit the growth of a bacterial culture; or the efflux pump assay that measures the amount of the fluorescent substrate (such as ethidium bromide) that is transported out of the bacterial cells by efflux pumps. For NorA, comparing the activity of the pump in the presence and absence of a potential inhibitor may directly indicate a measure of the inhibition level. The effectiveness of the drugs can also be expressed by measures as IC<sub>50</sub> (half maximal inhibitory concentration) that is measured by controlling a specific biological process such as the growth of a bacteria culture or the activity of an enzyme. IC<sub>50</sub> is defined as the concentration of a compound needed to inhibit the biological process by 50%. The lower the IC<sub>50</sub> value, the more effective the compound is at inhibiting the process [106]–[109]. The ChEMBL is a manually curated database which includes bioactive molecules [110]. For these molecules, a standard value is reported by calculating the  $-\log_{10}$  of the inhibition levels of the molecules that are reported in the studies. Basically, the formula is explained as:  $-\log_{10}$  (molar XC<sub>50</sub>, IC<sub>50</sub>, AC<sub>50</sub>, EC<sub>50</sub>, K<sub>i</sub>, K<sub>d</sub> or Potency) [111].

A variety of chemicals has been tested for their inhibitory activity on NorA. Here, a list of NorA inhibitors were obtained from ChEMBL, and stored in CSV file format. These studies surrounding these inhibitors are mentioned in section 1.7. The inhibitors were obtained by searching for term “NorA” in the search bar of ChEMBL web interface and browsing then to the associated compounds section [112]. A total of 600 NorA inhibitors were available in the ChEMBL database by the time of this search (13.08.2021). The list of NorA inhibitor molecules included those whose activities were obtained based on strains of the organism *S. aureus* that included mutated NorA sequences. Moreover, activities of several inhibitors were also tested multiple times in different studies, leading to duplications. Hence, a pre-processing of the dataset was performed using a simple script written in R programming language [113]. Initially, from the downloaded CSV file, assays that included the “mutant” and “mutation” terms in their description were removed. Then, the repeated molecules (duplicates) were removed from the dataset. After that, the number of molecules were reduced to

256. To achieve a more reliable set of direct NorA inhibitors, only molecules with inhibitory activities reported in terms of “% reduction in NorA efflux activity” were retained. From this set of molecules, only molecules that are reported to have standard value (i.e., inhibition level) less than 20% were removed to eliminate molecules with limited activity (and thus no significant inhibitory activity).

## 2.2. Modeling of 3D Structure of NorA

The protein sequence of *S. aureus* NorA includes 388 amino acids and displays a typical MFS fold with 12 transmembrane segments. No experimentally resolved structure of NorA was available at the beginning of this work, although two structures of this MFS pump in complex with an antibody were resolved via CryoEM experiments recently [104]. Moreover, the resolved structures included outward-facing conformations of NorA, which was considered less relevant than the inward-facing conformation as the primary aim was to characterize ligand (inhibitors as well as substrates) recognition by the pump, and the substrates are recognized by the inward-facing conformation. An AlphaFold2 model of NorA in its inward-facing conformation appeared on UniProt database later (AF-Q2G0A2-F1), yet it was also unavailable at the beginning of this work either. Hence, homology modeling and threading tools were used to model the 3D structure of NorA. Homology modeling and threading tools require the sequence of the molecule, and homology modeling tools can either propose experimentally validated structures as templates, or they can allow user to upload their own template structures to create 3D model of the given amino acid sequence. Phyre2 [114], SWISS-MODEL [115], AlphaFold2 [116] can be given as examples for homology modeling tools, while I-TASSER [117] is a known example of threading tools. Phyre2, I-TASSER tools only require the amino acid sequence of the protein for the modeling, whereas SWISS-MODEL also gives the to the users to upload their own template. The difference between homology modeling and threading lies on the fact that homology models are created based on structures of homologous proteins, yet threading is based on structures of homologs in the fold-level, therefore threading algorithms are in general considered more useful for sequences with no close homologs with known structures [118].

In this project, polypeptide sequence of norAI, which is the most common allele of this protein, was selected for structural modeling (GenBank ID: BAA14147.1) [80] [119]. Several homology modeling and threading tools were used as alternatives to generate a model of NorA structure. Initially, SWISS-MODEL web tool was used with default options, which proposed two protein structure templates that had similar amino acid sequence of norAI strain. Secondly, Phyre2 web portal was used with default options, which had multiple results obtained by using multiple protein structures as templates, these structures were proposed by the algorithm. Thirteen homology models from the Phyre2 results were selected for further analysis. After this, the protein structures that were used as a template for modeling of norAI, mentioned in the literature, were used as templates and SWISS-MODEL was used to model the protein. The templates were chosen from the work of Drew, D. [27] and had PDB ID of 6SM4 and 2GFP [120], [121]. Finally, a model of another less common allele of NorA, NorAII, which was available at the AlphaFold2 Protein Structure Database [122], was also used as a template for modeling of NorAI structure using SWISS-MODEL.

The resulting 3D structure models of each tool were then analyzed with Swiss-Model Structure Assessment[123] and QMEANBrane [124] tools to obtain scores that help assess the quality of the generated structures. To this end, several metrics were taken into consideration, such as the MolProbity score, sequence identity and coverage, QMEANDisCo score, and QMEANBrane score. Ramachandran plots were also examined.

MolProbity score is calculated based on bond angles, bond lengths, and Ramachandran angles. A low MolProbity score indicates a higher quality structure [125]. QMEANDisCo (Quality Measure for the Assessment of Non-local Distance Restraints in Crystallography) is a scoring function that measures of how well the non-local distance restraints agree with the electron density map of the structure. The score is a number between 0 and 1. A low QMEANDisCo score indicates better agreement and a higher quality structure [126]. QMEANBrane (Quality Measure for the Assessment of Embedding in Membranes) evaluates the quality of protein structures that are embedded in a lipid bilayer. The shape of the protein within the membrane as well as the agreement of the lipids with their experimental electron density map and the agreement of the protein with its experimental electron density map are all factors that contribute to the final score. Scores range from 0 to 1, with 0 denoting complete agreement and 1 denoting no agreement [127]. The Ramachandran plot is a graphical

representation of the conformation of the peptide backbone in terms of the two dihedral angles, phi and psi, used in protein structure prediction and analysis. Based on a statistical examination of the conformations of recognized protein structures, the plot displays the permitted and prohibited regions for these angles. The plot's "preferred regions" are those where most residues are located, whereas the "disallowed" or "disfavored" parts are those where little to no residue is located [128]. Here, the NorA structure with the highest quality was expected to have low MolProb, QMEANDisco scores, whereas the number of psi-phi angle combinations in the favored regions in Ramachandran plot as high as possible, and QMEANBrane score in expected range. The template structures used in homology modeling were also selected based on their conformational states to use only inward-facing MFS structures as templates. The model that yielded the most favorable scores in terms of the above-mentioned metrics was then selected as the best model, and suitable results used in further analysis including MD and molecular docking simulations.

### **2.3. Molecular Dynamics (MD) Simulations**

Molecular dynamics (MD) simulations predict the movement of atoms in molecular system based on interatomic interactions calculated by Newton's law of motion [129], [130]. These simulations have been used to simulate the dynamic behavior of proteins since 1970s and have been proven to be useful for observing changes in protein structure on an atomistic scale in the face of perturbations such as mutations, ligand addition, and changes in environmental conditions such as pH, temperature, pressure, and chemical composition. MD simulations can also be used for examining the conformational changes, interactions between proteins and small molecules, the properties of materials, lipid bilayers and solutions [131].

A force field is a mathematical model that represents the interactions between atoms and molecules in an MD simulation. The force-field is used to calculate the total energy of a system as a function of the atomic and molecular locations and velocities. This energy is utilized to compute the forces acting on each atom and molecule to model their motion through time. Atomic and molecular interactions are described by a series of equations constituting a force field. These equations account for the different types of possible contacts, including electrostatic, van der Waals, and

covalent interactions. Typically, the force field's parameters are derived via experiments or quantum mechanical calculations, and they are utilized to precisely simulate the interactions between atoms and molecules. There are various forms of force fields, including classical and quantum mechanical force fields. Classical force fields are often utilized in MD simulations and are based on classical mechanics. These force fields are used to simulate the interactions between atoms and molecules using simple mathematical functions and are generally computationally efficient. Quantum mechanical force fields, on the other hand, are based on quantum mechanics and provide a more precise description of the interactions between atoms and molecules, albeit at a greater processing cost [132]. The time steps used in MD simulations are usually on the femtoseconds (fs) range, since most chemical processes occur on a femtosecond time scale.

Several procedures are required to prepare a system for MD simulations. First, the system is prepared by specifying the molecules, initial positions of their atoms, and the boundaries of the simulation box. Second a topology file that specifies the bonded and unbonded interactions between atoms in a system is created. Typically, this file contains information about the types and amounts of atoms, as well as any constraints or restrictions that should be imposed to the system. Third, a force-field suitable for simulation of the system is selected, and the potential energy of the system is minimized to eliminate bad angles in the system that may lead to unstable MD simulations. Fourth step is equilibration, in which the conformation of the protein is allowed to interact with and adjust to its environment under controlled temperature and pressure levels. Fifth, initial velocities are assigned to atoms and molecules in accordance with the desired temperature of the system. Finally, once the system is ready with necessary simulation settings are defined, the simulation is started [133]–[135].

Several tools that simplify these processes and generate necessary input parameters for MD simulations to be conducted with popular MD simulations programs have been developed in the past. One such software is CHARMM-GUI which is a web-based platform, which is developed since 2006 and used for preparation of input files for MD simulations. Simulations packages such as CHARMM, NAMD, GROMACS, AMBER and GENESIS accept inputs generated by CHARMM-GUI. The platform contains various modules for MD simulations of a

broad range of cellular components such as glycolipids, building bilayers, soluble proteins [134], [136], [137].

The Membrane/Bilayer Builder module of CHARMM-GUI web service was used to generate complex of protein-bilayer system or bilayer-only system for NorA MD simulations. Building a membrane using CHARMM-GUI includes several steps. First, system is defined by specifying the kind and number of lipids, the size of the box used for the simulation is decided, and the boundary conditions for the system to be simulated. Then, Lipid Builder module is utilized to specify the type of lipids to be used in the simulation and the number of lipids in the system. Then the Solvation Wizard module used to introduce solvent molecules to the system, such as water. Utilize the Energy Minimization module to reduce the system's energy, which can eliminate any high-energy conformations that may be present. Afterwards, the Equilibration Wizard is utilized to equilibrate the system to the appropriate temperature and pressure. Finally, Parameter Wizard module is used to assign force field parameters based on the type and attributes of the atoms and molecules to the force field [138]–[141]

There is several software designed for the purpose of MD and GROMACS is one of the most utilized package utilized to perform MD, since it is a free and community-driven project. It is designed for biochemical molecules and used for non-biological systems and can run both graphics processing units (GPUs) and central processing units (CPUs) [136], [137], [142]–[145].

In this project, two independent 100 nanoseconds (ns)-long molecular dynamics (MD) simulations were conducted to (1) examine stability of the generated NorA model and (2) obtained potential ligand binding pockets from conformation selected from the generated simulation trajectories. Computational resources kindly provided by TRUBA TR-GRID were used to perform these simulations. The protein was inserted inside a POPC (1-palmitoyl-2-oleyl-glycerol-3-phospho-choline) bilayer prior to simulation by using CHARMM-GUI [102]. The steps involved in building a membrane around the protein included uploading the PDB structure of the protein and choosing the chain to be used. Since the structure had only one chain, it was selected. In the second step, PDB manipulation options were not used and left at default settings. The third step included correcting the orientation of the protein, positioning, and area calculation options. In this step, PDB orientation was selected to be used and other parameters were not used. In the fourth step, system size was determined and POPC

was added to the system. The system included NorA, and 248 POPC, 49 potassium ions, 54 chloride ions and 18425 TIP3 water molecules. Then, these components were assembled. Simulations were conducted by using CHARMM36 as force field and GROMACS (version 2018.2) as the simulation engine [142], [146]. For both simulations, system was minimized for a maximum of 5000 steps until the maximum force reached a level lower than 10 kJ/mol. Cutoff scheme Verlet was used which generates a pair list with buffering and the buffer size was selected as 1.2. Van-der Waals and Coulomb interaction cut off distances were used as 1.2 nm. For treatment of long-range electrostatic interactions, Particle Mesh Ewald (PME) method was used. Following minimization, the systems were subjected to 3 equilibration steps each including 125000 steps using 1 femtoseconds time step at 310K in the NVT ensemble. Then 3 more equilibration steps each including 250000 steps using 2 femtoseconds were conducted with same settings. Positional restrains were applied to NorA and lipid molecules during equilibration steps. Finally, unrestrained production simulations were conducted for 100 ns, in the NPT ensemble by setting pressure at 1 bar and temperature at 310 K and using.

For the analysis of MD simulations Root-Mean Deviation Score (RMSD) and Root-Mean Fluctuation Score (RMSF) were calculated for the structure using ProDy [147]. These scores represent the difference between the atoms while the molecule is moving in time, compared to the starting point and calculation individual residue flexibility.

## **2.4. Binding Pocket Detection from MD Simulation Trajectories**

A major goal of this thesis work is characterization of ligand promiscuity of NorA using computational structural biology tools. To this end, prediction of interactions between ligands/inhibitors and potential binding sites of NorA ligands within the internal cavity of the protein is required. This, in turn, requires definition of binding sites on the protein structure. To consider all possible binding sites on the NorA structure as well as the influence of protein flexibility, we extracted conformations from MD simulations trajectories, and predicted binding pockets on these structures prior to docking simulations. Binding pocket detection may be achieved by existing online tools that are designed to predict binding pockets or

cavities on a given protein structure, such as DoGSiteScorer [148], GRID [149], [150], POCKET [151], and SurfNet [152], [153].

Each MD simulation conducted here yielded 1002 conformations. From each simulation, 10 conformations were selected in a stepwise manner, corresponding to conformations at 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100 ns simulation time. Potential ligand binding sites on each of these conformations were detected using DoGSiteScorer web service, which was available at proteins.plus website. The tool reports pockets, along with their respective volumes and a DrugScore value. DrugScore is a measure of the druggability potential of a binding site. The calculation of DrugScore value is based on a combination of structural, energetic, and pharmacophoric such as the size, shape, and electrostatic properties of the site, as well as the binding energy of the ligand. The score ranges from 0 to 1, with a higher value indicating a higher druggability potential of the binding site. Here, pockets with DrugScore values smaller than 0.65 were eliminated. Finally, a total of 45 binding pockets were obtained. and used to define docking grids in molecular docking simulations in an “ensemble docking” context. These pockets were thus taken to represent the variability of the NorA internal cavity in the context of ligand recognition.

## **2.5. Molecular Docking Simulations**

Molecular docking is a computational procedure that is often used to predict the binding of small molecules to macromolecules. In addition to its widespread use in computational Structure-based Drug Design (SBDD) campaigns, the method can also help characterize binding behaviors and elucidate biochemical processes of the small molecules [154]. During molecular docking, small molecule is oriented towards its potential receptor, and its conformation is modified many times. For each conformation that produces a ligand-macromolecule interaction (i.e., a binding “pose”), a binding affinity prediction is performed, or the obtained pose is scored using a scoring function [155]. More than 60 tools have been developed in the past for molecular docking simulations and subsequent scoring of obtained poses, such as DOCK [156], AutoDock [157], FlexX [158], GOLD [159].

As described in Section 2.1, 96 small molecules that are known to have inhibitory effect on NorA efflux pump, as well as ciprofloxacin, a known substrate of the NorA, were docked into the detected binding pockets. AutoDock Vina (v1.1.2) was used to perform molecular docking simulations. The information required for the formation of grids were extracted from DogSiteScorer results. The X-Y-Z coordinates and the radius of the box is given in the results, which was integrated into the Autodock Vina codes. The files required by AutoDock Vina were prepared by converting the structure files (either in pdb or sdf format) to pdbqt file format using Open Babel [160]. The exhaustiveness level was set to 512 to perform a comprehensive conformational search. A total of  $45 \times 96 = 4320$  docking runs were performed. Since multiple pockets were used for each inhibitor, multiple binding affinities were predicted by AutoDock Vina. Among these, only the pose that yielded the lowest binding free energy (kcal/mol) for each inhibitor was considered in further protein-ligand interaction profiling analysis. Moreover, only the inhibitors with binding affinities smaller than -7 kcal/mol were included in the analysis. Binding site characterization was performed by using a pdb file of protein-ligand complexes and an offline installation of PLIP [161]. The whole process was automated using custom Python scripts. These scripts and the files generated during the analysis are given in Appendix B.

## **2.6. Data Analysis**

### **2.6.1. Clustering Based on Pairwise Tanimoto Similarity Scores**

Similarity of molecular substances can be calculated by computing various molecular similarity metrics. Tanimoto coefficient is a score that is used for quantification of similarity between two small molecules. The score is calculated by calculating the ratio of the number of elements common to both molecules to the total number of properties in both molecules [162]. These elements are also called chemical fingerprints and generation of these fingerprints is achieved by detecting all linear paths of atoms and bonds, branching points and cyclic patterns. Large and scattered data is better understood by clustering the similar sets with each other. Clustering analysis is used to organize and classify the patterns into groups. There are different approaches utilized by researchers for clustering large datasets. Hierarchical and

partitioned approaches are the main two branches for clustering methods [163]. In hierarchical approach, researchers aiming to analyze big data can create clusters that are consisting of subjects with related properties [164]. Here, 96 small molecule inhibitors of NorA obtained from ChEMBL database (as described in section 2.1.) were clustered based on their Tanimoto similarity scores between each other. For this purpose, first Smiles IDs of the ligands were obtained from the ChEMBL during the first download, then these IDs were subtracted from the CSV file and used as the input for the analysis. Tanimoto similarities were calculated by using rdkit module on Python (v.3.8). The molecules were then clustered based on the obtained scores. The clusters were visualized as heatmap using python seaborn module. The code written for this analysis is given in Appendix B.

### **2.6.2. Clustering Based on Protein-Ligand Interaction Profiles**

Small molecules and macromolecule can form various type of chemical interactions, as hydrogen bonds, hydrophobic contacts, salt bridges, and water bridges. It is important to characterize interaction types formed between the protein-ligand complex to achieve a better understanding of the function as well as molecular recognition behavior of the protein at hand. This type of information can also help with drug discovery efforts [165].

The same 96 inhibitors described previously were also clustered based on the results of a binding site analysis of obtained molecular docking poses. Interactions between each molecule and putative NorA binding pockets were obtained by using PLIP, which is a tool that allows detection and visualization of non-covalent interactions in protein-ligand complexes. For each inhibitor, PLIP generated a list of different types of interactions between the ligand and the protein, including hydrogen bonds, hydrophobic contacts, salt bridges, and water bridges. All these interactions were taken to consideration. For each ligand, a list of ligand contacting NorA amino acids was generated. An amino acid was considered ligand-contacting if it made at least one type of interaction with the inhibitor. Once these lists were obtained, the inhibitors were clustered based on the list of NorA residues (amino acids) they contact. These “PLIP clusters” were then visualized as heatmaps using python seaborn module. The code written for this analysis is given in the Appendix B.

### 2.6.3. Confusion Matrix

A “confusion matrix” was constructed to compare overlap between the two sets of clusters obtained from Tanimoto similarity scores and PLIP-based binding site analysis. This is a matrix which included “Tanimoto clusters” and “PLIP clusters” in its rows and columns, respectively. Each cell in the matrix included the number of molecules shared between sets of clusters. To enable the construction of this matrix, the same number of clusters was used in each clustering analysis.



## 3. RESULTS

### 3.1. Molecular Similarity Between NorA Inhibitors

First, NorA inhibitors were gathered from ChEMBL as explained in section 2.1. After removing duplicates and molecules with low effects (lower than 20% in standard value field), 96 molecules were left for analysis. A list of these molecules is given in the Appendix B. To get an overview of similarity/difference between these molecules, Tanimoto similarity scores between each inhibitor were calculated. The molecules were then clustered based on these scores, as described in 2.6.1. The obtained cluster map is given in the Figure 3.1 in the form of a dendrogram.

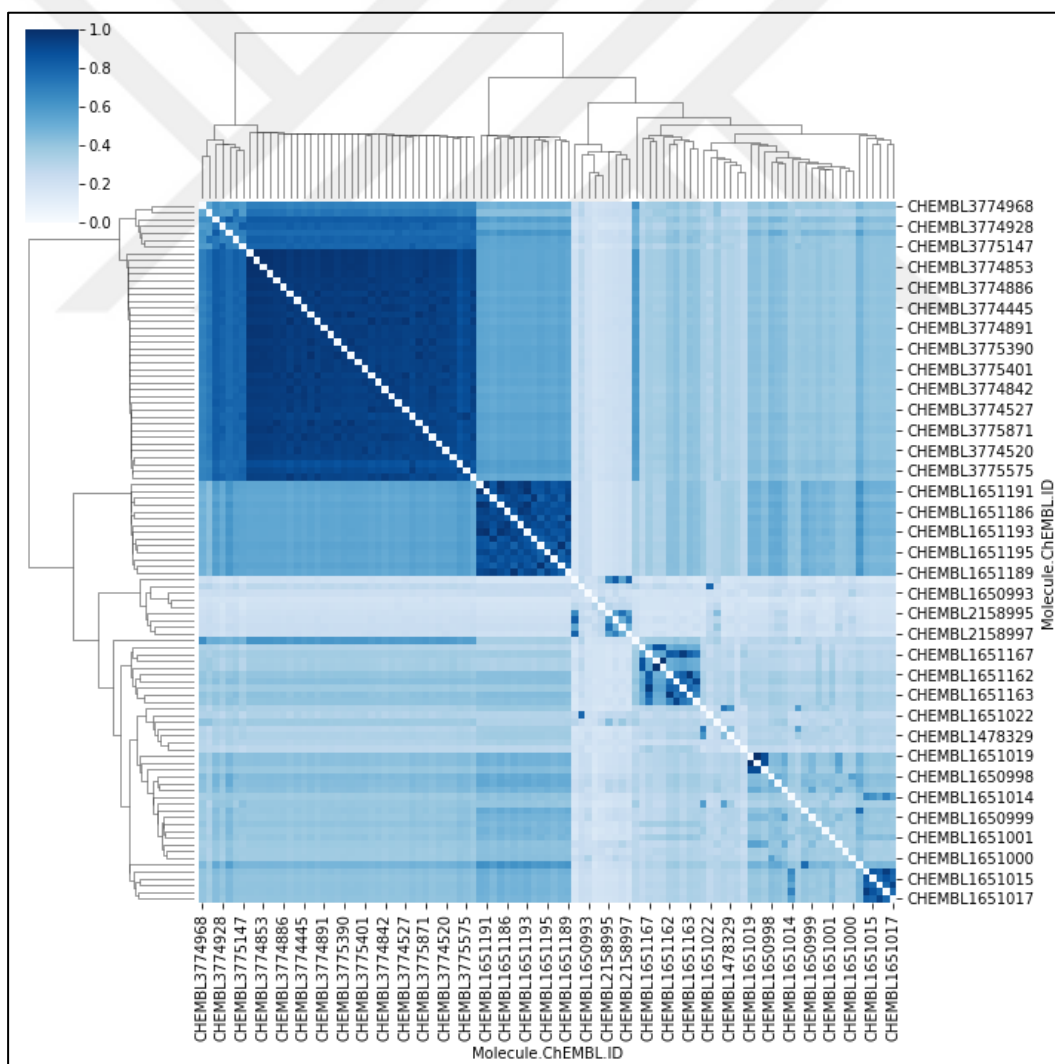
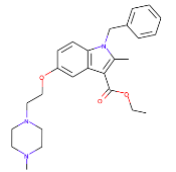
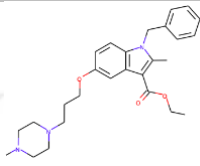
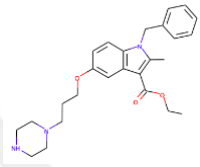
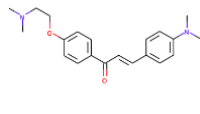


Figure 3. 1: Cluster map of NorA inhibitors based on Tanimoto similarity scores.

Two groups of molecules, in the top-left in Figure 3.1, were observed to include two sets of highly similar small molecules. The rest of the molecules; however, did not yield distinct groups of highly similar molecules, despite a group of molecules in the bottom-right in Figure 3.1, which yielded Tanimoto similarity scores between 0.4 and 0.6. As expected, this finding is in line with the previously indicated promiscuity in NorA ligand recognition. Nevertheless, the molecules were separated to four distinct clusters based on the results of this dendrogram to enable comparison between binding site similarity-based clustering results. The molecules that were found to be in the same cluster were then examined in the respective studies in which they were reported. Accordingly, ligands in Cluster 1 included 9 different molecules. These molecules originated from 3 different publications and included chalcones, indole-based compounds and other novel inhibitors designed to inhibit norA. The molecule with ChEMBL ID “CHEMBL3774527” had the highest standard value (and thus inhibition level), therefore it has been chosen as the representative of this cluster. The molecules in this cluster had median of 89 with 28,04 standard deviation as standard value, and a median of 435,57 and 54,82 standard deviation for molecular weight. The 38 ligands in the Cluster 2 also originated from 3 different publications and included quinolone and hydroxyquinolone derivatives, indole-based compounds and other novel inhibitors designed to inhibit norA. The molecule CHEMBL3774675 had the highest standard value, therefore it has been chosen as the representative of the cluster. The molecules in this cluster had median of 52,8 with 25,41 standard deviation as standard value and median of 391,5 and 55,82 standard deviation for molecular weight. Cluster 3 included 33 molecules from two different publications, and only indole-based compounds to inhibit norA. The molecule CHEMBL3775062 included the highest standard value, therefore it has been chosen as the representative of the cluster. The molecules in this cluster had median of 89,7 with 25,86 standard deviation as standard value and median of 450,58 and 82,78 standard deviation for molecular weight. Finally, Cluster 4 had 17 molecules coming from 7 different publications and there were a variety of chemical groups in this cluster including quinolone and hydroxyquinolone derivatives, chalcones, quinolone-like analogues, indole-based compounds, and other novel inhibitors. These molecules had median of 54,4 with 24,82 standard deviation for standard value and median of 436,53 with 81,27 standard value for molecular weight. These results are also summarized in Table 3.1. A more detailed version of the Table 3.1, which also includes the ChEMBL IDs of inhibitors, is given in Appendix B.

Table 3. 1: NorA efflux pump inhibitor groups by Tanimoto clusters.

| Cluster No. | Publication<br>ChEMBL ID's                                                                                    | No. of Ligands | Chemical<br>Groups/Classes<br>of the Ligands                                                                                                      | Median<br>of<br>Standard<br>Values<br>(+/- STD) | Median of<br>Molecular<br>Weights [-<br>+ STD] | 3D Structure of<br>Representative<br>Molecule                                                          |
|-------------|---------------------------------------------------------------------------------------------------------------|----------------|---------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------|------------------------------------------------|--------------------------------------------------------------------------------------------------------|
| Cluster 1   | CHEMBL2162179,<br>CHEMBL1657423,<br>CHEMBL3777946                                                             | 9              | Chalcones, Indole-<br>Based<br>Compounds,<br>Novel Inhibitors                                                                                     | 89 +/-<br>28,04                                 | 435,57 +/-<br>54,82                            | <br>CHEMBL3774527   |
| Cluster 2   | CHEMBL1657423,<br>CHEMBL1817147,<br>CHEMBL3777946                                                             | 38             | Quinolone and<br>Hydroxyquinolone<br>Derivatives,<br>Indole-Based<br>Compounds,<br>Novel Inhibitors                                               | 52,8 +/-<br>25,41                               | 391,5 +/-<br>55,81                             | <br>CHEMBL3774675   |
| Cluster 3   | CHEMBL3777946,<br>CHEMBL3879153                                                                               | 33             | Indole-Based<br>Compounds                                                                                                                         | 89,7 +/-<br>25,86                               | 450,58 +/-<br>82,78                            | <br>CHEMBL3775062  |
| Cluster 4   | CHEMBL1657423, CHEMBL1817147, CHEMBL3226388,<br>CHEMBL2162040, CHEMBL2162041,<br>CHEMBL3777946, CHEMBL2162042 | 17             | Quinolone and<br>Hydroxyquinolone<br>Derivatives,<br>Chalcones,<br>Quinolone-like<br>Analogues,<br>Indole-Based<br>Compounds,<br>Novel Inhibitors | 54,4 +/-<br>24,82                               | 436,53 +/-<br>81,27                            | <br>CHEMBL2158999 |

### 3.2. Modeling of 3D Structure of NorA

No experimentally determined inward-facing conformation of NorA has been available at the start of this thesis work. Therefore, to obtain a 3D structure of IF NorA conformation, protein structure modeling methods including homology modeling and threading were utilized. For homology modeling, 23 experimental structures of different MFS transporters in their IF conformations and a computationally generated model of another less common allele of NorA (NorAII), available at the AlphaFold Protein Structure Database, were used as templates. Then, the obtained models were examined using the Structure Assessment tool of Swiss-Model. Quality metrics taken into consideration included the MolProb score, Ramachandran favored score, sequence identity and coverage, QMeanDisCo, QMeanBrane. These scores were calculated for each of the molecules and the results are given in the Appendix B.

Based on the structure assessment, the NorAI model generated by using the AF2 NorAII structure yielded MolProb score of 0.77 with 97.93 Ramachandran favorite score and 0.63 QMeadBrane score. When compared to those of other models, these scores were found to be the most favorable. The generated structure is shown in in Figure 3.2. As expected, the modeled structure of the NorA had 12 transmembrane structures as in other MFS transporters.

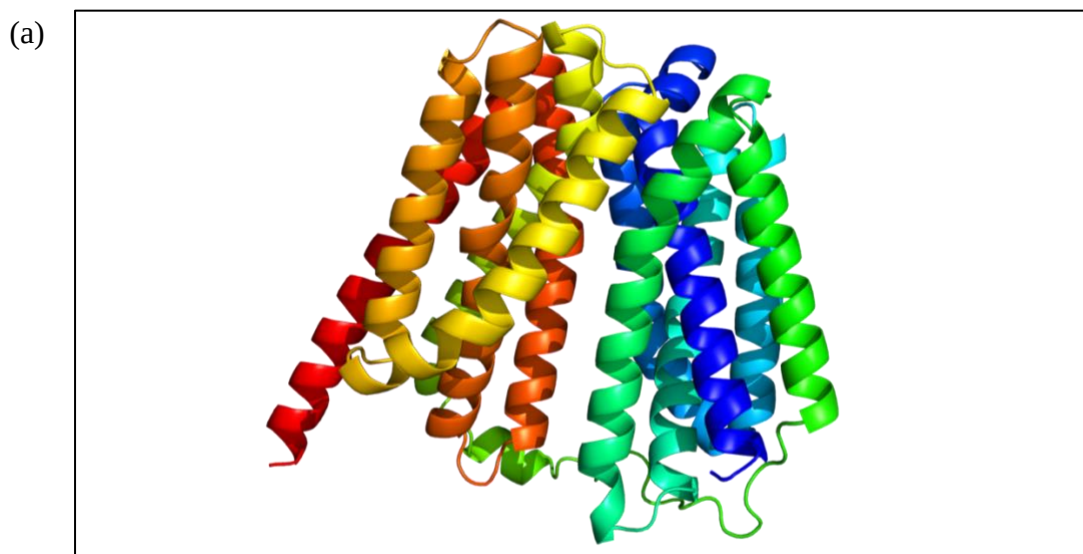


Figure 3. 2: A model of NorA efflux pump inward-facing conformation. The side view of the model. (b) The view of the model from the bottom.

(b)



Figure 3. 3: Continue.

The Ramachandran plot of the model is also given in Figure 3.3. The dihedral angles of protein backbone were clearly found to be in the accepted areas with a favored percentage of 97.93%. Hence, this structure was used as initial structure in MD simulations described in the next section.

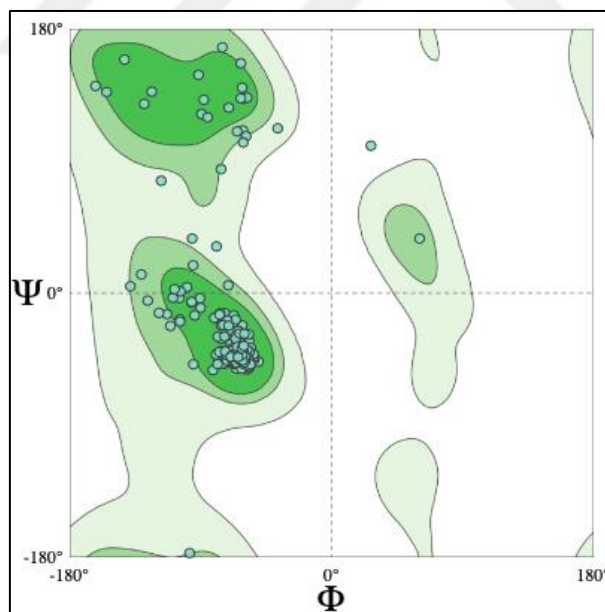


Figure 3. 4: Ramachandran plot of the selected NorA model for MD simulation. The NorA model produced by AF2 was found to show the highest quality (Ramachandran favored percentage 97.93%, QMEANBrane expected range <100000, Molprobit score 0.77).

### 3.3. MD Simulations

Two molecular dynamics (MD) simulations for, each 100 nanoseconds (ns) long, were performed to (1) examine stability of the generated NorA model, and (2) to obtain putative ligand (substrate/inhibitor) binding pockets within the internal cavity of NorA (in case the structures are deemed sufficiently stable). These MD simulations were named based on the date of MD simulation started (08\_12\_2021 and 09\_21\_2021). The RMSD and RMSF time profiles were calculated for both simulations (Figure 3.4 and 3.5).

RMSD plots of both simulations displayed an initial increase (as expected from a typical classical MD simulation). However, the simulation 09\_21\_2021 was found to yield a more stable RMSD when compared with simulation 08\_12\_2021. The latter simulation showed an increasing trend in the RMSD with increasing simulation time.

RMSF plots given in the Figure 3.5 were obtained by computing fluctuations about average positions of only carbon alpha atoms of each amino acids. According to RMSF plot of both simulations, residues 175-200 showed relatively higher level of fluctuations (highlighted in Figure 3.6). This region is not part of any secondary structure element (i.e., helix in the case of NorA): it is a linker element that connects helices N-terminal helix 6 and C-terminal helix 7. These helices are given in Figure 1.2. High levels of RMSF in this region are thus expected.

Our results in this section indicated that one of the MD simulations conducted (simulation 08\_12\_2021) did not appear to have completed equilibration based on the RMSD plot. However, RMSF plot of this simulation did not show clear differences with the other simulation conducted. In contrast, there is a larger fluctuation in the C-terminus of the protein, which apparently affected RMSD plot as well. A visual inspection of this simulation also showed that the backbone of the obtained NorA IF model remained stable throughout the simulation trajectory. The generated NorA model was thus concluded to be a stable IF conformation of NorA.

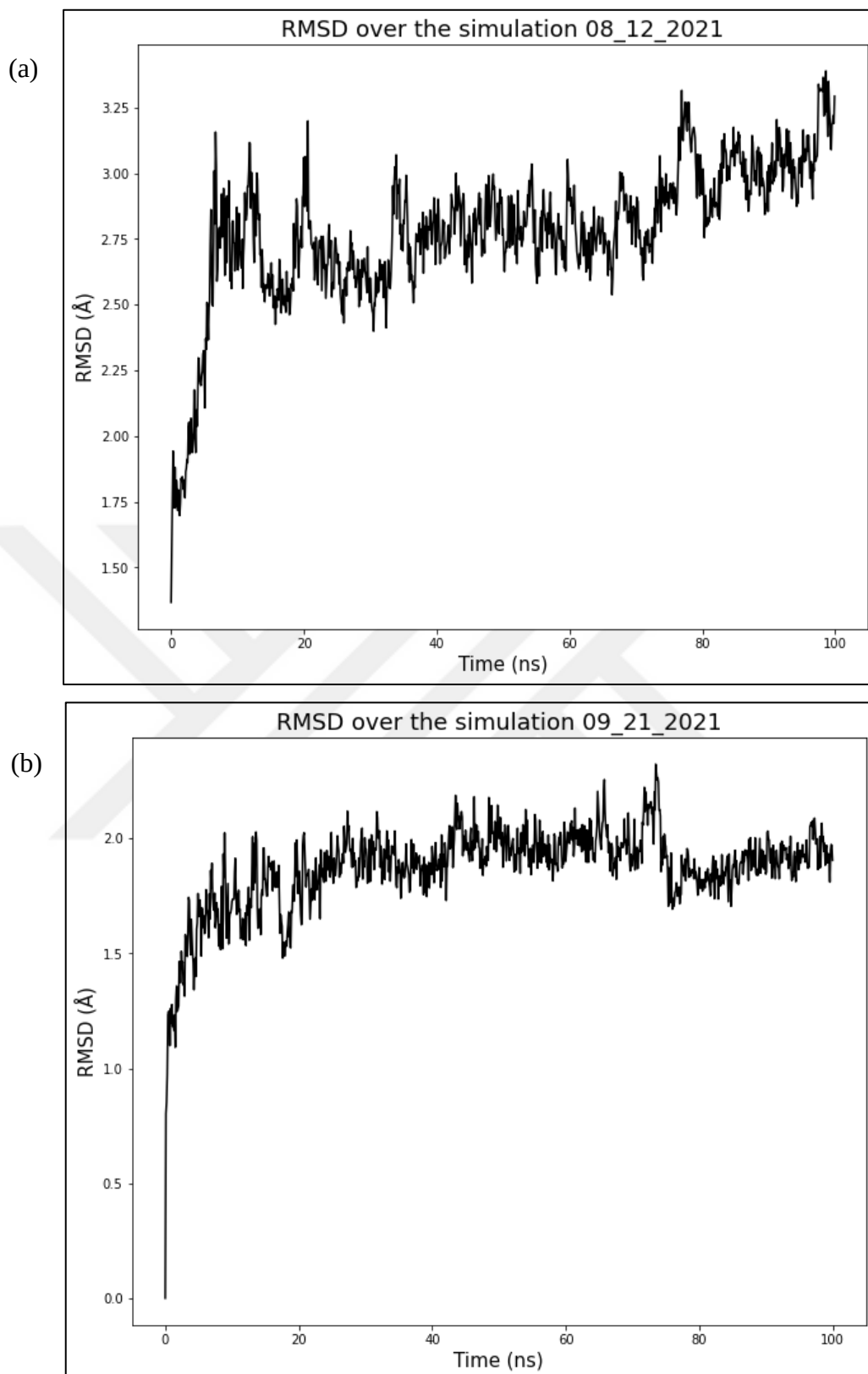


Figure 3. 5: RMSD scores obtained from both MD simulations.(a) RMSD over the simulation 08\_12\_2021. (b) RMSD over the simulation 09\_21\_2021.

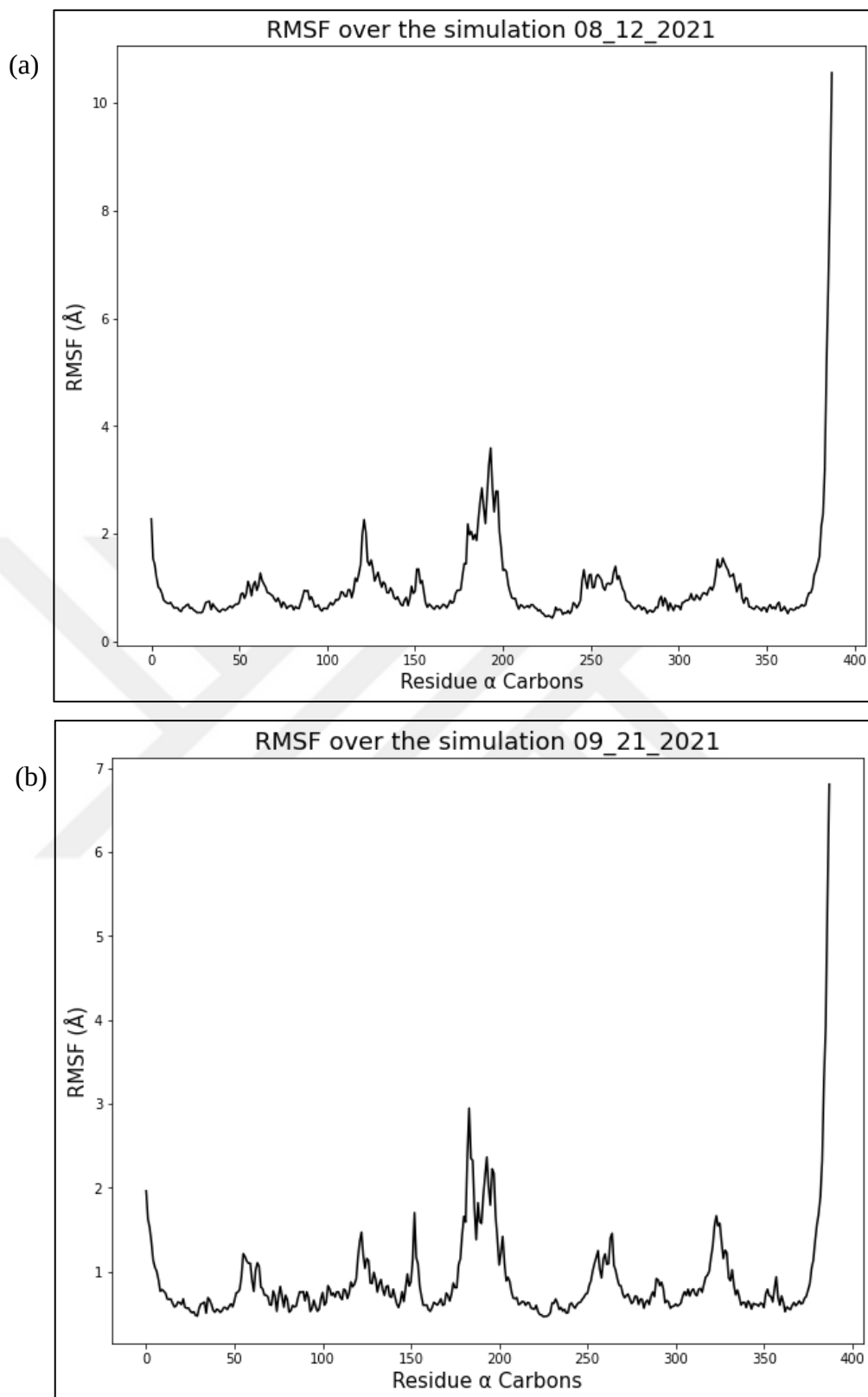


Figure 3. 6: RMSF plots obtained from simulations.(a) RMSD over the simulation 08\_12\_2021. (b) RMSD over the simulation 09\_21\_2021.

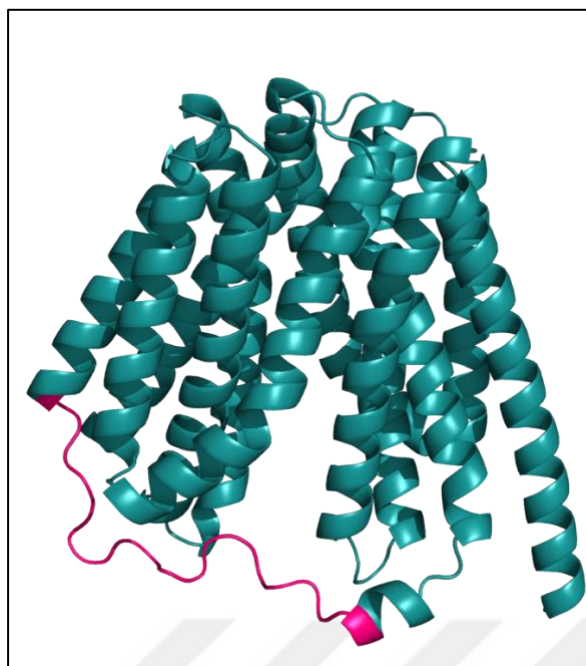


Figure 3. 7: NorA homology model used in the study with fluctuated region according to RMSF scores (residues between 175-200 amino acids) highlighted in pink.

### 3.4. Putative Binding Pockets in Inward-Facing Conformation of NorA

From each MD simulation trajectory, 11 snapshots were taken (or in other words, conformations were selected). These conformations were saved in separate PDB files and named based on their respective frame numbers. Thus, 22 different conformations were obtained. Putative ligand binding pockets in each of these conformations were then detected using DogSiteScorer. Detected pockets in each conformation were filtered based on the drug score calculated by the server (only those having a DrugScore higher than 0.65) and the volume of the pockets (only those with a volume higher than 400 Å<sup>3</sup>). The remaining binding pockets for each conformation are used for further analysis. A visual overview of these pockets is given in the Figure 3.7. Volumes, drug scores, and amino acids found in each pocket in each conformation are given in the Appendix B.

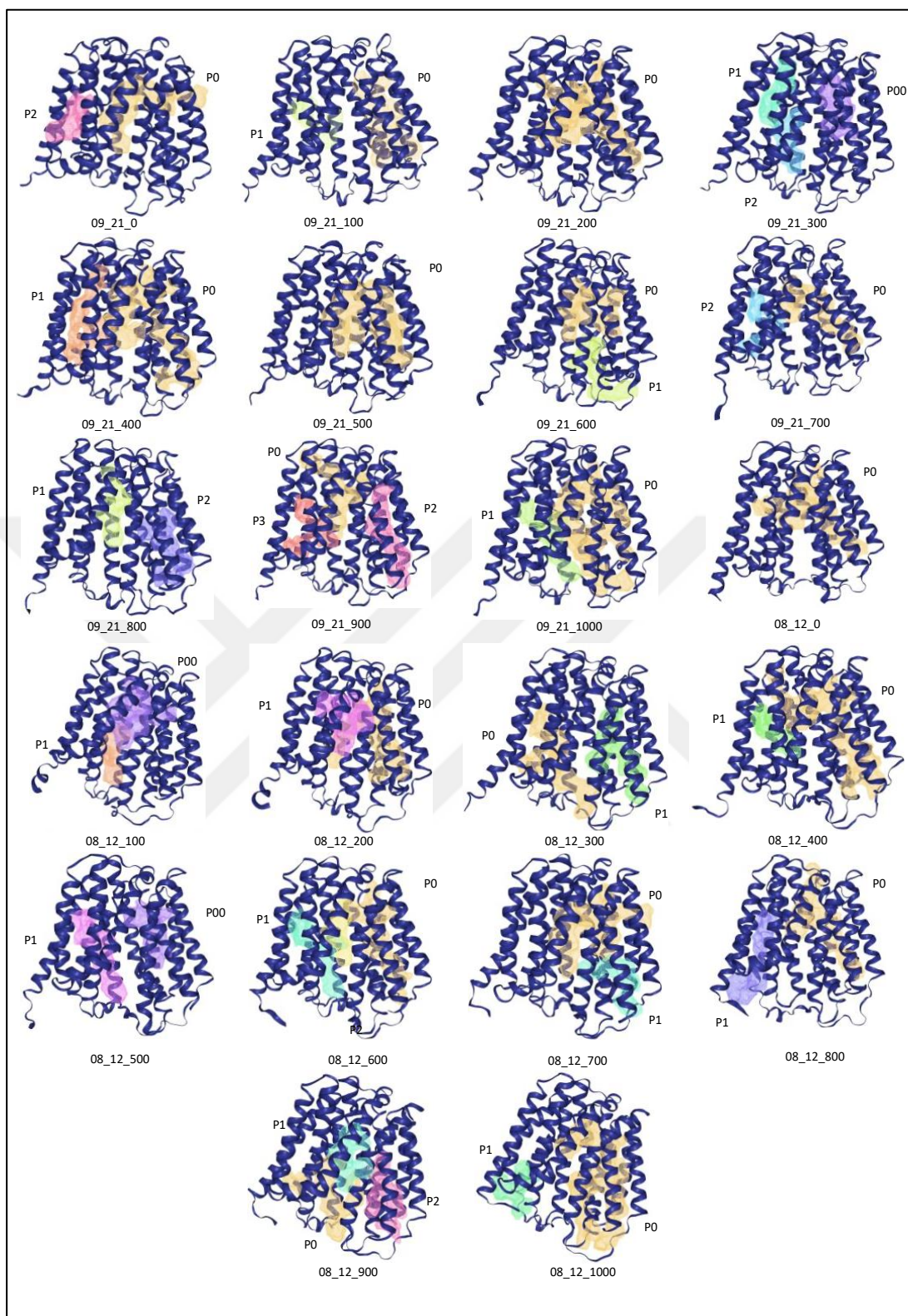


Figure 3. 8: Putative ligand binding pockets in NorA IF conformation as predicted by DogSiteScorer.

### 3.5. Interaction Between Ciprofloxacin and NorA

Ciprofloxacin is a known substrate of NorA. Therefore, we first docked ciprofloxacin into each of the detected binding given in Figure 3.7 using AutoDock Vina to characterize the substrate binding pocket in NorA IF conformation. The pocket 09\_21\_0\_P\_0 had the lowest binding score of -8.9, while other pockets had scores ranging from -8.7 to 18.6. Since pocket 09\_21\_0\_P\_0 had the best binding score, PLIP algorithm was used to obtain amino acids that are found on the binding sites between the molecule and the protein. According to the results of this analysis, 98R, 19I, 22V, 44V, 40L, 43L, 85F were found to be the residues that recognize and bind ciprofloxacin. Hence, these residues were identified to form the substrate-binding site of NorA (Figure 3.8 and 3.9).

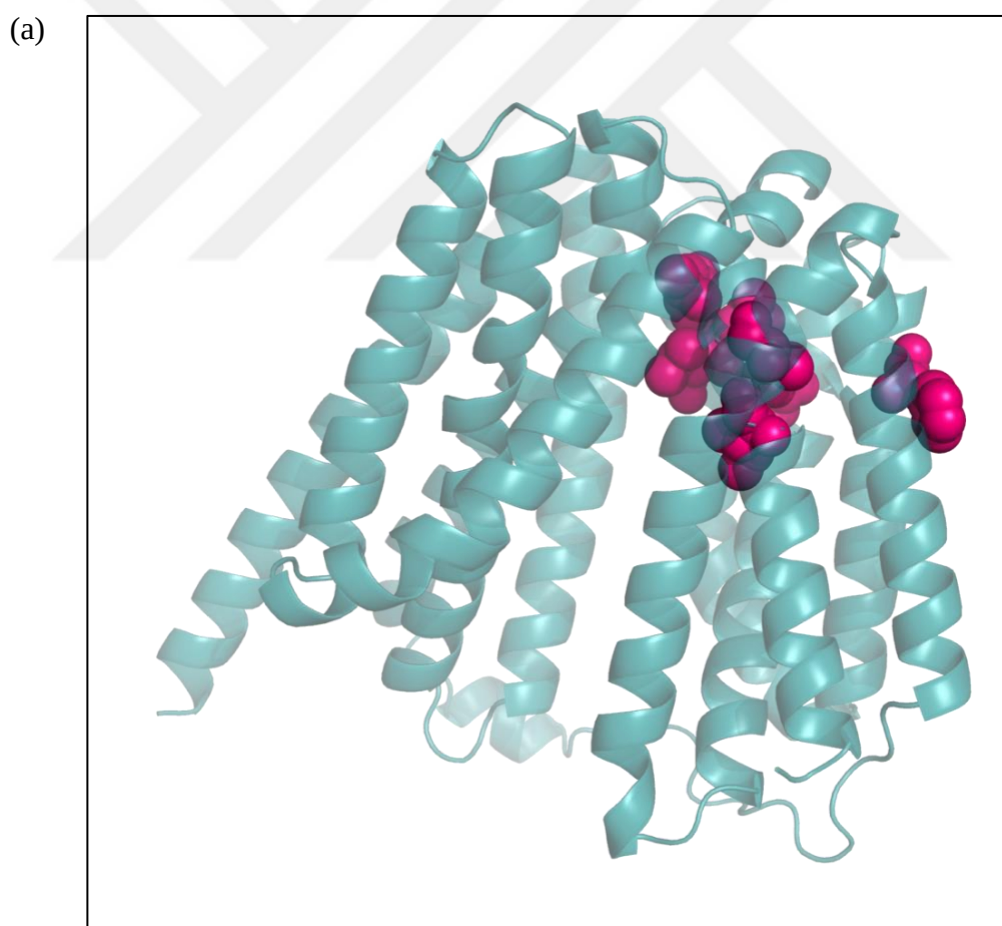


Figure 3. 9: Ciprofloxacin binding residues highlighted on NorA efflux pump of *S. aureus*.(a) The side view of the binding site in homology model. (b) The view of the binding site of the model from the bottom.

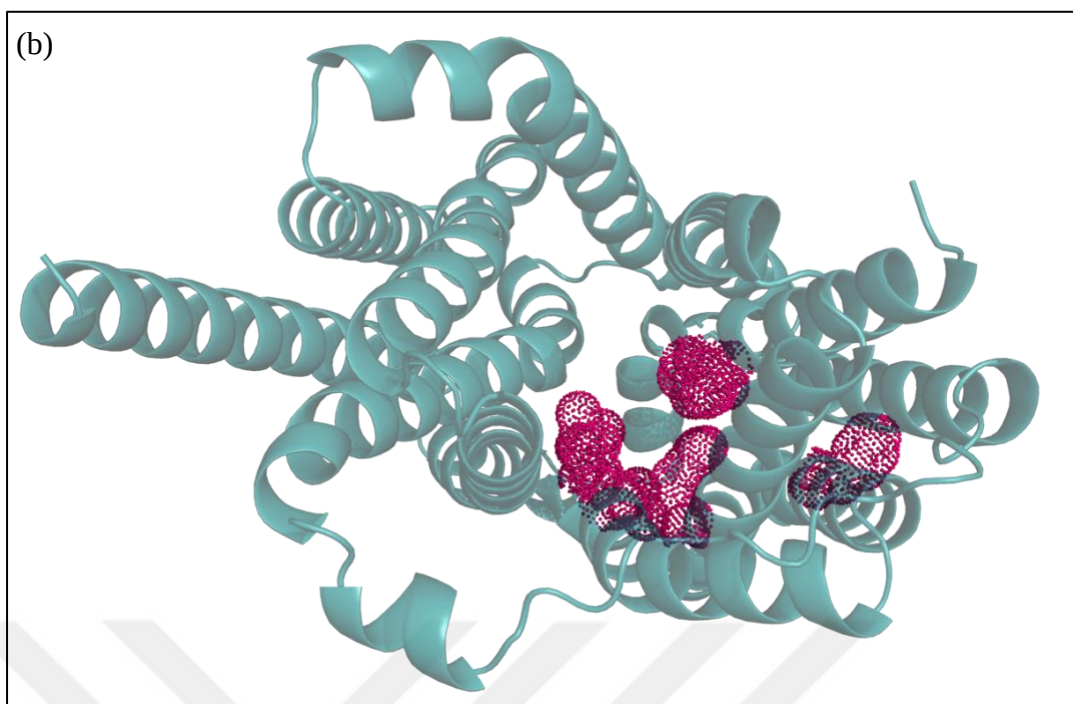


Figure 3. 10: Continue.

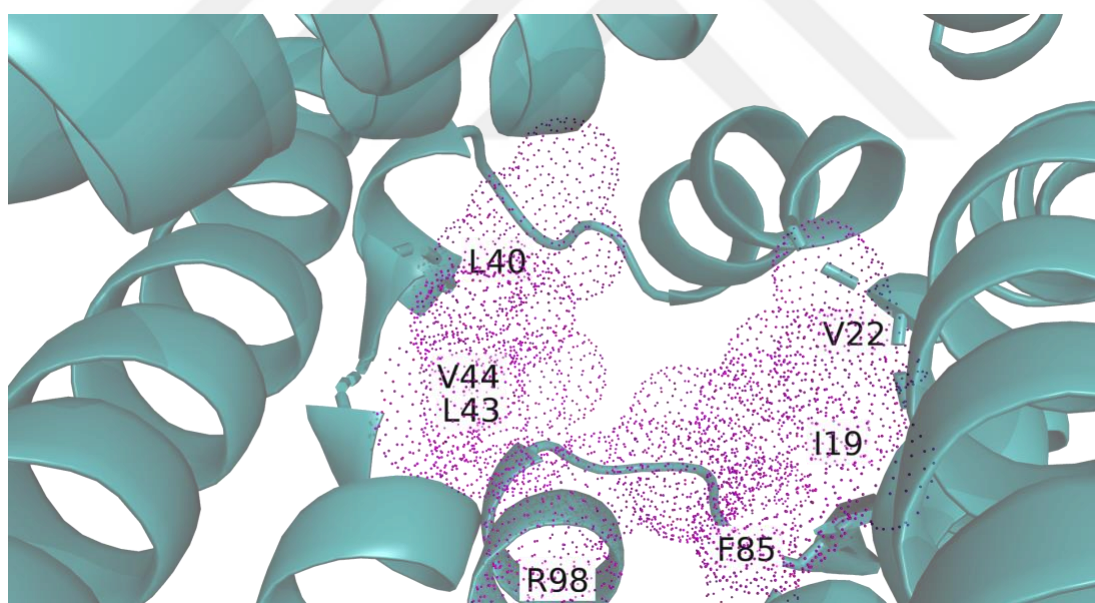


Figure 3. 11: Ciprofloxacin binding residues 98R, 19I, 22V, 44V, 40L, 43L, 85F on NorA efflux pump, highlighted.

### 3.6. Characterization of Interactions Between NorA and Its Inhibitors

The inhibitors of NorA were docked into all the pockets detected in Section 3.4 (and shown in Figure 3.7) using AutoDock Vina. Following docking simulations, only poses that yielded the lowest binding free energy estimates were included in protein-ligand interaction profiling analysis. Then, for each inhibitor, an array of ligand-contacting amino acids (residues) was generated. These arrays were then subjected to a clustering analysis to identify groups of inhibitors that showed similar contact patterns with NorA. As a result of this analysis, 47F, 19I, 98R, 16F, 140F, 78F, 23I, 22V, 105A, and 44V were identified as positions that contacted most of NorA inhibitors. These residues are highlighted in the Figure 3.10 and Figure 11. A full list of ligand-contacting residues is given in Appendix B.

From the cluster map, 4 groups (clusters) of inhibitors with similar binding patterns to each other were identified. These groups are highlighted in Table 3.2 and the binding residues of these molecules can be examined in the heatmap of Figure 3.13. It is important to note that, not all the binding residues are shown in the heatmap. The inhibitor molecules in the Group 1, formed high number of bonds with the amino acids 47F, 19I, 98R, 78F, 16F, 105A, 108V, 22V, 15I, 44V, 223T, 140F, 23I, 26L, 51Q, 85F, 40L, 82E, 75L, 340N. The molecule with ChEMBL ID “CHEMBL3774527” was chosen as representative for Group 1. The amino acids 52M, 333S, 110P, 188F, 109M, 310R, 113T, 334T, 336T, 332N, 313I, 309I, 255Q, 222E, 218L, 214L were found to be interacting with the molecules in the second group, however the number of bonds formed with these amino acids were much lower compared with Group 1. The molecule CHEMBL1649602 were chosen as the representative of this group. Group 3 included amino acids 23I, 22V, 140F, 47F, 19I, 222E, 26L, 85F, 98R, 44V, 223T, 40L, 226S, 16F and molecule CHEMBL1650999 was chosen as representative. Finally, the molecules in Group 4 included amino acids 140F, 303F, 244I, 23I, 222E, 141I, 26L, 19I, 226S, 225Y, 16F, 22V, 307D, 47F, 44V and the molecule CHEMBL2158999 was chosen as the representative of this group. The interaction patterns of these representative molecules are given in Figure 3.13 to Figure 3.17.

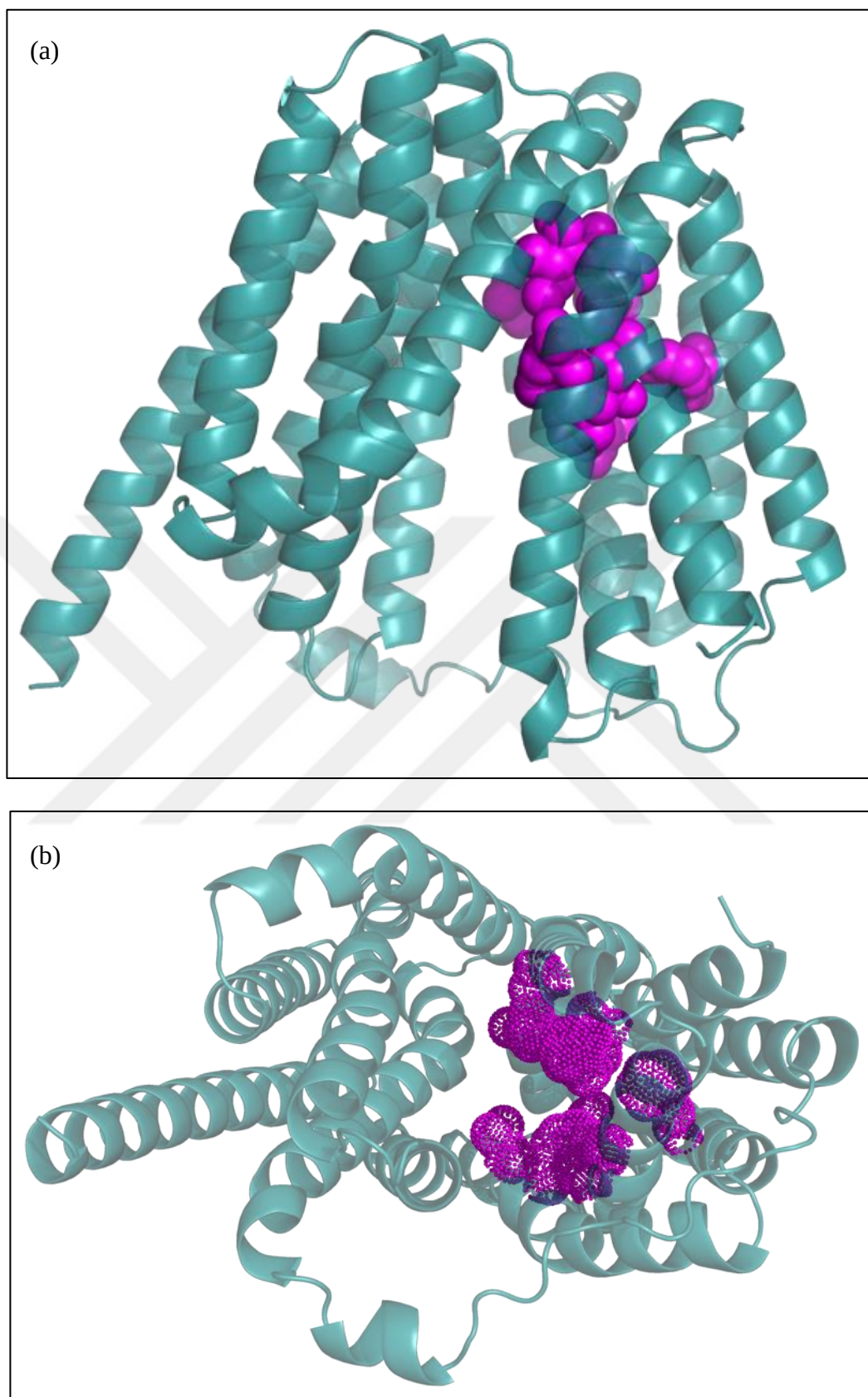


Figure 3. 12: Inhibitor binding residues highlighted on NorA efflux pump of *S. aureus*.(a) The side view of the binding site in homology model. (b) The view of the binding site of the model from the bottom.

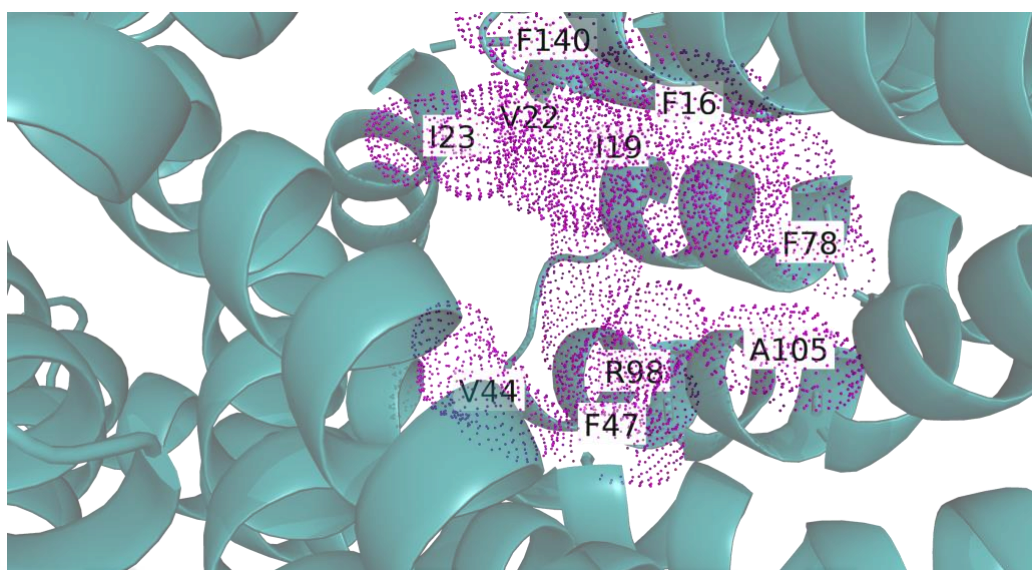


Figure 3. 13: Inhibitor binding residues 47F, 19I, 98R, 16F, 140F, 78F, 23I, 22V, 105A, 44V highlighted on NorA model.

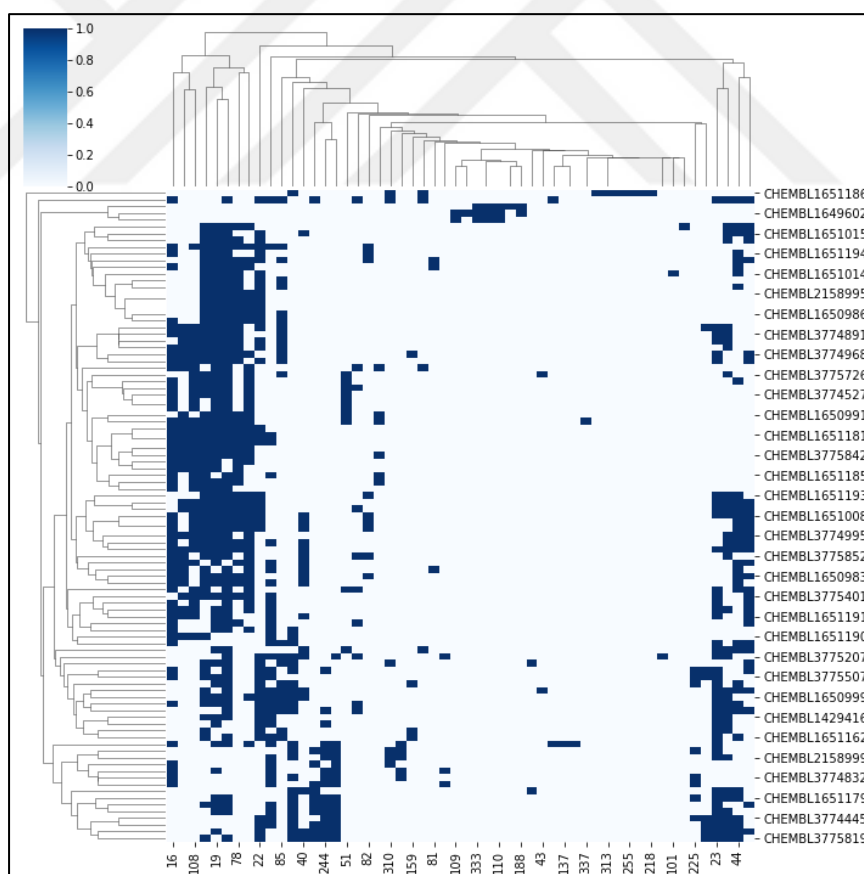


Figure 3. 14: Heatmap of binding site clusters obtained from PLIP analysis. The clusters with similar binding characteristics are highlighted with red boxes, the number of cluster is also indicated.

Table 3. 2: NorA efflux pump inhibitor groups by PLIP clusters.

| Group Name | ID            | Molecular Species | 2D Structure | Complex with NorA from Molecular Docking Simulation |
|------------|---------------|-------------------|--------------|-----------------------------------------------------|
| Group 1    | CHEMBL3774527 | Neutral           |              |                                                     |
| Group 2    | CHEMBL1649602 | Neutral           |              |                                                     |
| Group 3    | CHEMBL1650999 | Acid              |              |                                                     |
| Group 4    | CHEMBL2158999 | Neutral           |              |                                                     |
| Substrate  | Ciprofloxacin | Zwitterion        |              |                                                     |

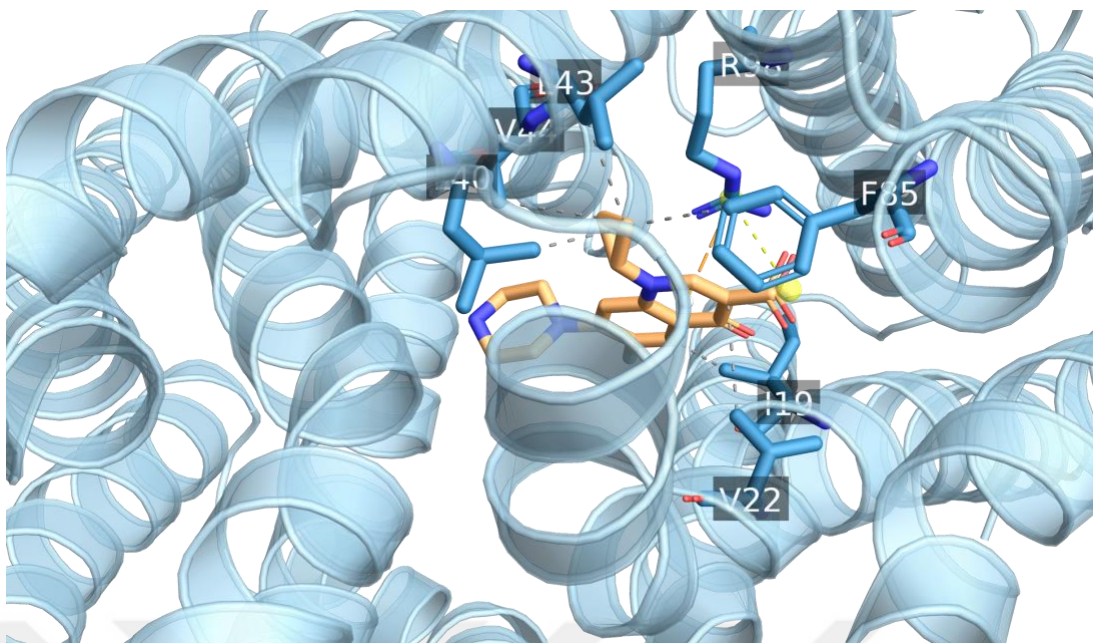


Figure 3. 15: Ciprofloxacin complex with NorA Efflux Pump.

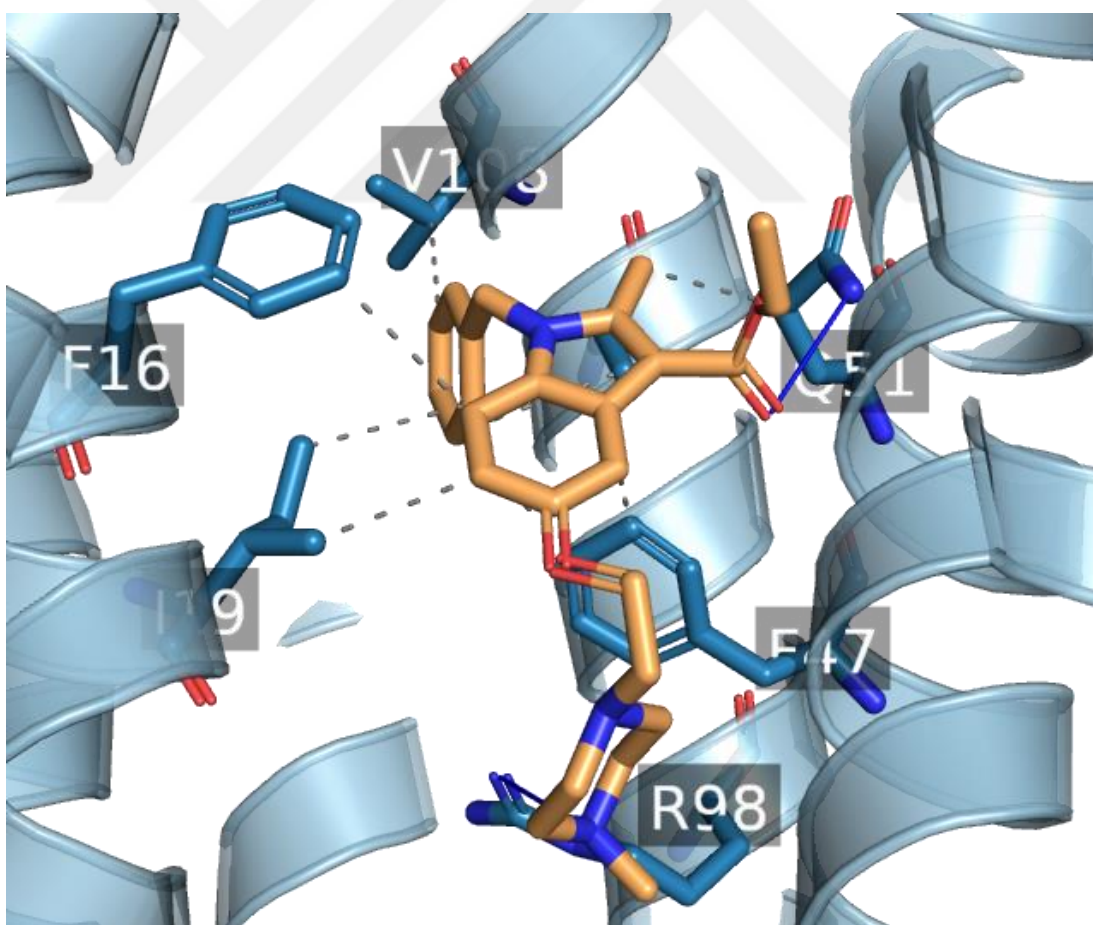


Figure 3. 16: CHEMBL3774527 of Group 1 complex with NorA Efflux Pump.

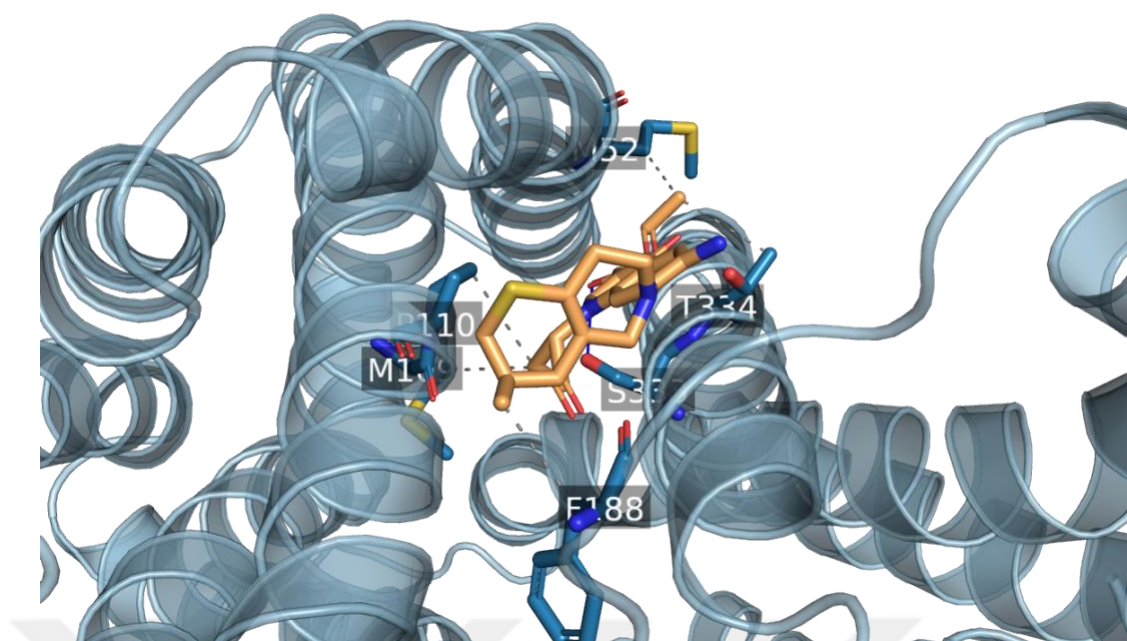


Figure 3. 17. CHEMBL1649602 of Group 2 complex with NorA Efflux Pump.

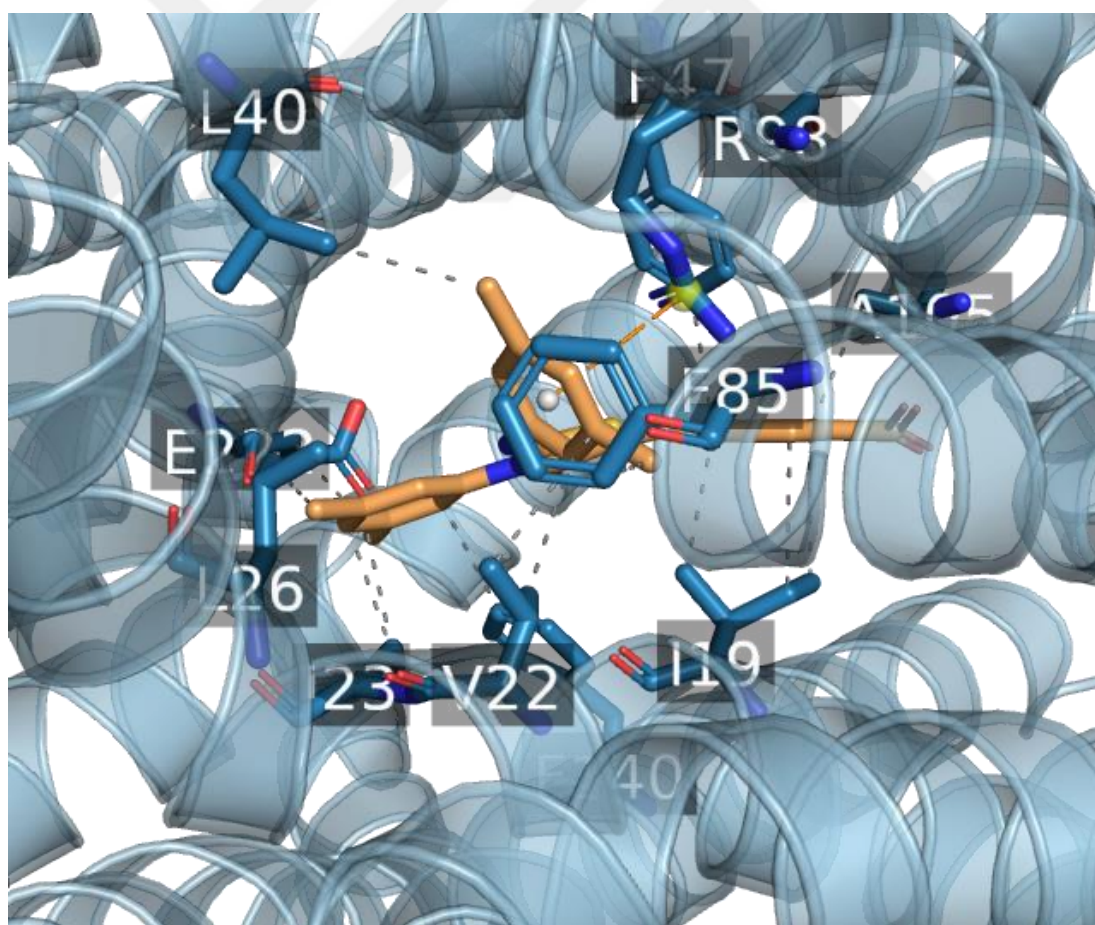


Figure 3. 18. CHEMBL1650999 of Group 3 complex with NorA Efflux Pump.

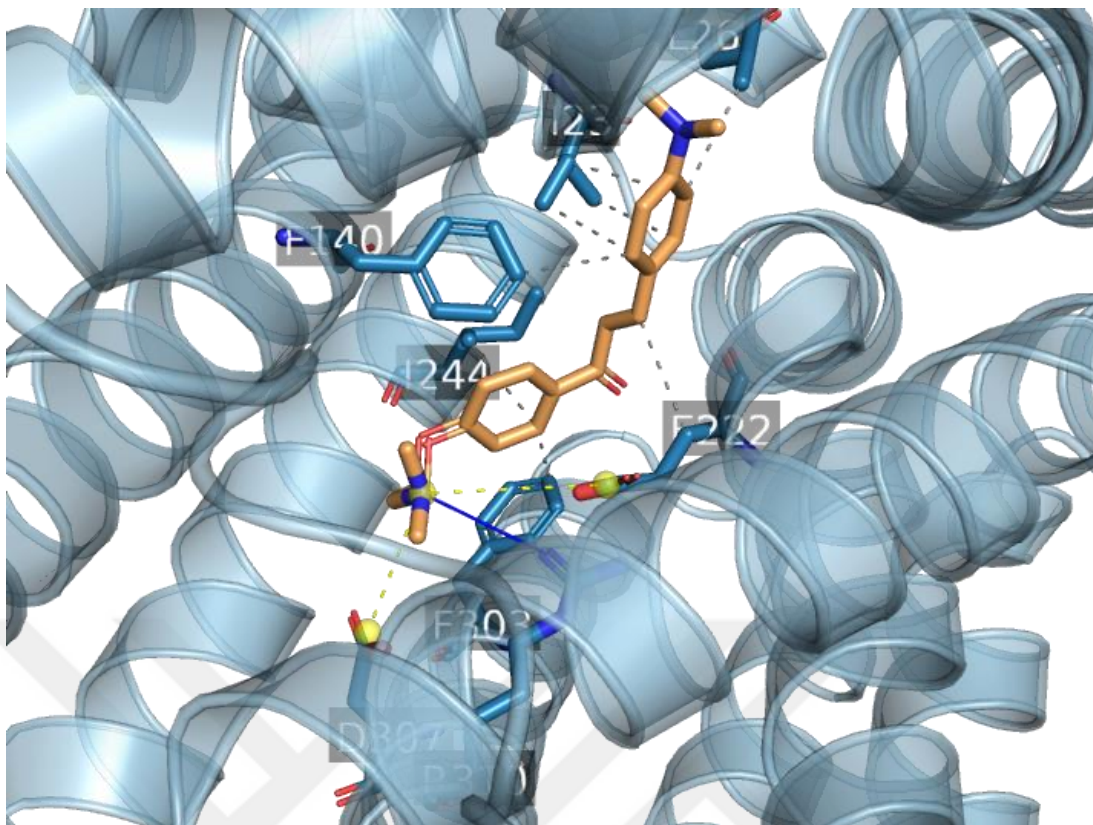


Figure 3. 19: CHEMBL2158999 of Group 4 complex with NorA Efflux Pump.

### 3.7. Confusion Matrix

The 96 inhibitor molecules of norA were clustered based on Tanimoto similarity scores as explained in the section 2.6.1. These molecules were also clustered based on the binding site characteristics on norA residues, as explained in section 2.6.2. A confusion matrix was generated to compare the overall similarity between these clusters. The matrix is given in the Figure 3.18. According to the matrix, even though the inhibitor molecules belonged to different Tanimoto clusters, most of these molecules preferred the same PLIP cluster. Out of 96 inhibitors, 62 of them preferred PLIP cluster 0, 4 inhibitors preferred PLIP cluster 1, 17 inhibitors preferred PLIP cluster 2 and 14 molecules preferred PLIP cluster 3. These results indicate that, despite differences in their molecular structures, most inhibitors preferred the same binding site in the NorA internal cavity.

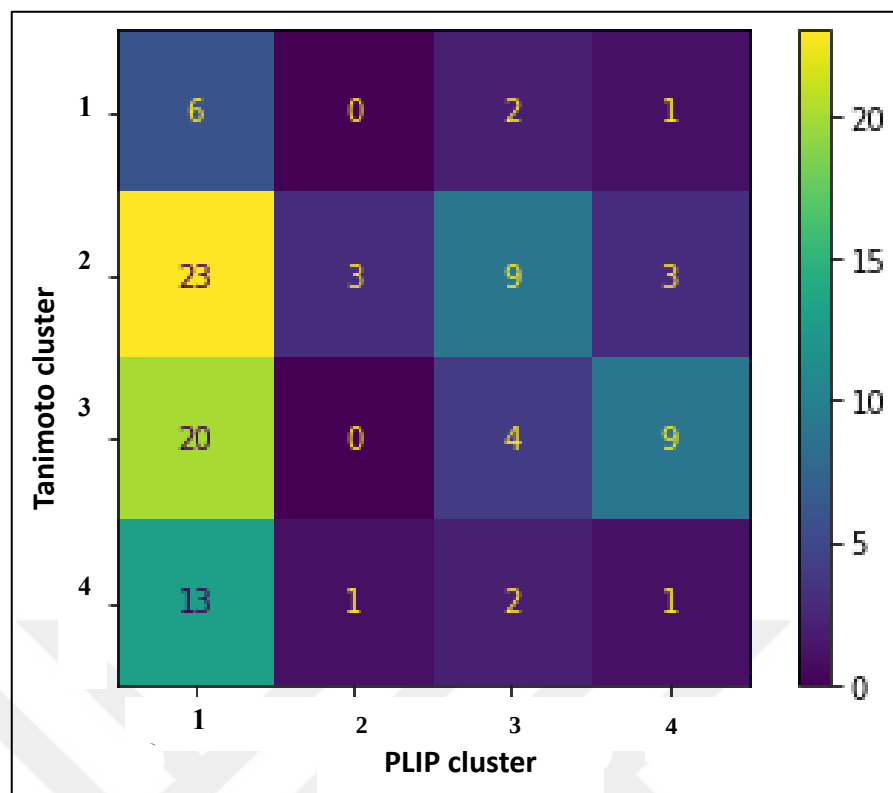


Figure 3. 20: Confusion matrix generated by comparison of Tanimoto scores and binding site clusters of inhibitors.

## 4. DISCUSSION

*S. aureus* is a type of bacteria that can cause a variety of infections, including skin infections such as boils and impetigo, respiratory infections such as sinusitis and pneumonia and food poisoning. Some strains of *S. aureus* are resistant to antibiotics, making them difficult to treat. Multidrug-resistant *S. aureus* (MDRSA) refers to strains of the bacteria that are resistant to multiple classes of antibiotics. MDRSA strains have developed resistance mechanisms to different classes of antibiotics, such as beta-lactams, macrolides, tetracyclines, and others, making it harder to treat infections caused by these bacteria. In this study, inhibitors of the NorA efflux pump that were previously reported at the ChEMBL database were collected, and Tanimoto similarity scores between each of them were calculated. Then, the ligands were clustered based on these Tanimoto similarity scores. As given in the heatmap of Figure 3.1, a major cluster that included a large group of molecules, and a few minor clusters that included smaller groups of molecules, were obtained. In Table 3.1. chemical groups and classes, median of standard values and molecular weight of these ligands were given. A variety of the chemical groups were clearly included in the inhibitors list, since the inhibitors included chalcones, indole-based compounds, quinolones, among others. The standard values of the molecules and molecular weights were different between Cluster 1-3 and Cluster 2-4. Hence, there was no significant difference between these inhibitors, based on standard values and molecular weights.

The major facilitator superfamily (MFS) is a group of transmembrane transport proteins responsible for transporting a wide variety of small molecules across the cell membrane, such as sugars, amino acids, ions, and drugs. NorA is a type of efflux pump that belongs to the major facilitator superfamily (MFS) of transmembrane transport proteins. Structurally, MFS transporters are composed of 12 transmembrane alpha helices, with the N- and C-termini located in the cytoplasm. The transmembrane helices form a central pore through which the substrate is transported. In this project, several tools were utilized to obtain 3D model of the NorA efflux pump, and the amino acid sequence of the most common allele of NorA, NorAI, was used to model the structure. A Cryo-EM structure of the NorA has been discovered recently, yet in an outward-facing conformation, which is not the state in which the substrate is recognized. The model which yielded the best quality metrics was obtained by homology modeling of the norAI structure based on a computational model of NorAII

obtained from the AlphaFold2 database. This model had a typical MFS-fold (12 transmembrane structures), with 97.93% of backbone dihedral angles in the favored region of Ramachandran plot, QMEANBrane score in the expected range ( $<100000$ ), and a Molprobit score of 0.77.

Since the inward-facing structure of NorA was not available prior to this study, putative binding pockets within the NorA structure were also not characterized. Such characterization was aimed here by using molecular docking simulations. To perform molecular docking simulations in an ensemble docking context, we extracted multiple IF conformations of NorA from MD simulations. During simulations, the movement of the amino acids can modify the internal cavity within the protein, and thus putative binding pockets as well. The simulations yielded multiple binding pockets, which were then filtered based on their DrugScore values and volumes. The pockets with drug score smaller than 0.65 and volumes smaller than  $400 \text{ \AA}^3$  were removed since these pockets were considered unlikely to bind small molecules.

The exact mechanism of substrate recognition and binding by the NorA efflux pump is still not fully understood. In a study by Palazzotti et al., supervised molecular dynamics simulations revealed 127K, 264K, 109M, 133S, 136I, 51Q, 310R, 222E, 303F, and 225Y residues on NorA efflux pump as residues that interacted with ciprofloxacin. In a study by Shang et al. substrate binding sites were examined via side directed mutagenesis and FRET analysis and the study highlighted the amino acids 16F, 222E, 303F and 307D are important for substrate binding activity of NorA. However, in the recent study by Brawley et.al., researchers discovered the structure of NorA by Cryo-EM. According to the results of this study, it was found that ~65% of substrate-binding pocket included lipophilic residues and contained 16F, 140F, 303F and 306F, that are aromatic residues. These residues are believed to be facilitating binding to aromatic drugs. Ionizable residues as 98R, 222E, 307D and 310R were also found in the pocket, that forms positively and negatively charged patches. These patches are thought to aid drug binding due to positively charged substrates. Ciprofloxacin is a type of antibiotic that belongs to the fluoroquinolone class of drugs and works by inhibiting the bacterial DNA gyrase enzyme that is used for DNA replication and repair. It is reported that over expression of Nora efflux pump can lead to resistance to ciprofloxacin. According to our analysis here, 98R, 19I, 22V, 44V, 40L, 43L, 85F were the amino acids that were found to be the residues that binds to ciprofloxacin. The residue 98R were also highlighted in the above-mentioned study by

Brawley et.al. in the substrate binding site. According to Brawley et.al, a positively charged patch is formed by 98R in the N-terminal. The findings of this project thus agree with this of these experimental study.

A particularly interesting feature of NorA is its promiscuous ligand binding behavior. Indeed, a variety of NorA inhibitors were discovered previously. A set of small molecule inhibitors of NorA efflux pump were docked into various binding pockets detected in MD simulations of this protein. Then, contact patterns of each of these inhibitors with NorA were analyzed. Based on this analysis, 47F, 19I, 98R, 16F, 140F, 78F, 23I, 22V, 105A and 44V residues were found to interact with most of the inhibitors. The research by Shang et al. also suggested 16F as important for substrate binding activity of NorA. When compared with Brawley et al. residues 98R, 16F and 140F were found to be overlapping. Importantly, 98R were found in both substrate and inhibitor interactions here. However, the regions 16F and 140F were specific to inhibitors, while 222E was specific to ciprofloxacin. These residues can be used to highlight important aspects of binding mechanism of NorA efflux pump. The residues that were found to be important based on our analysis can also be important for the substrate and inhibitor binding mechanisms and can be used for further studies.

The inhibitor molecules were clustered based on Tanimoto similarity scores, as well as binding site profiles because of PLIP analysis. The results of PLIP analysis are given in Figure 3.12 as a heatmap. The heatmap highlights the preferences of inhibitors with the residues they are interacting. These molecules were clustered to four groups and these groups are given in Table 3.2. These groups belonged to various molecular species as acids, bases and neutral compounds. In statistics, confusion matrix is used to compare real values with test values to see if test values show similarity to real values. By this approach, the reliability of the statistical models can be obtained. In this project, the Tanimoto similarity cluster and PLIP clusters were compared with confusion matrix. The matrix is given in Figure 3.18. Based on the matrix, it can clearly be seen that most of the inhibitors in different Tanimoto clusters were accumulated in one PLIP cluster. Out of 96 inhibitors, 62 of them preferred PLIP cluster 1, 4 inhibitors preferred PLIP cluster 2, 17 inhibitors preferred PLIP cluster 3 and 14 molecules preferred PLIP cluster 4.

The inhibitor molecules were found to be interacting with various amino acids on the protein. The residues were compared with the studies mentioned previously. 16F, 140F were found in the Groups 1,3, and 4 and was mentioned in the studies by

Shang et al. and Brawley et al. The residues 303F, 307D and 222E were mentioned in all three studies, and 303F, 307D was only found in Group 4 and 222E Groups 3 and 4. 98R mentioned by Brawley et al. was found in Group 1,3 and ciprofloxacin. 310R mentioned by Palazzotti et. al and Brawley et al. was only found in Group 2.

These findings indicate that, even though the inhibitors are different according to their chemical properties and groups, they prefer to bind to specific residues on NorA efflux pump. This result agrees with the promiscuous character of NorA efflux pump, meaning that it can bind to molecules that are chemically dissimilar. Even though most of the inhibitors preferred one distinct cluster, the clusters that are preferred by the minority, Cluster 3 and 4 can also be important for binding characteristics of NorA efflux pump. All of clusters have common amino acids, and the movement of the inhibitors within the protein can be achieved with interactions of these residues. These results can also be used as a starting point of more comprehensive studies that aim to understand NorA efflux pump of other species, as well as preparation of inhibitory molecules against the fight with antibiotic resistance.

## 5. CONCLUSION

In this study, potential ligand binding pockets of NorA and binding characteristics of known NorA inhibitors were characterized based on a homology model. 47F, 19I, 98R, 16F, 140F, 78F, 23I, 22V, 105A and 44V were found to be predominant residues involved in ligand binding. Comparison of protein-ligand interaction profiles and Tanimoto similarity scores of NorA inhibitors revealed that a single binding pocket is responsible for binding a wide variety of ligands, confirming the promiscuous character of the NorA efflux pump. These findings can potentially serve as a launching point for additional research aimed at elucidating the NorA efflux pump of other species and the development of compounds that suppress antibiotic resistance.

## REFERENCES

- [1] Baldeweg F., Hoffmeister D., Nett M., (2019) “A genomics perspective on natural product biosynthesis in plant pathogenic bacteria”, *Nat Prod Rep*, 36 (2), 307–325.
- [2] Shi Y. M., Bode H.B., (2018) “Chemical language and warfare of bacterial natural products in bacteria–nematode–insect interactions”, *Nat Prod Rep*, 35 (4), 309–335.
- [3] Tobias N. J., Shi Y. M., Bode H. B., (2018), “Refining the Natural Product Repertoire in Entomopathogenic Bacteria” *Trends Microbiol*, 26 (10), 833–840.
- [4] Newstead L. L., Varjonen K., Nuttall T., Paterson G. K., (2020), “Staphylococcal-Produced Bacteriocins and Antimicrobial Peptides: Their Potential as Alternative Treatments for Staphylococcus aureus Infections” *Antibiotics*, 9 (2), 40.
- [5] Beceiro A., Tomás M., Bou G., (2013), “Antimicrobial Resistance and Virulence: a Successful or Deleterious Association in the Bacterial World?”, *Clin Microbiol Rev*, 26 (2), 185.
- [6] Dong J., Olano J. P., McBride J. W., Walker D. H., (2008), “Emerging pathogens: challenges and successes of molecular diagnostics”, *J Mol Diagn*, 10 (3), 185–197.
- [7] Vouga M., Greub G., (2016), “Emerging bacterial pathogens: the past and beyond”, *Clin Microbiol Infect*, 22 (1), 12–21.
- [8] Hutchings M., Truman A., Wilkinson B., (2019), “Antibiotics: past, present and future”, *Curr Opin Microbiol*, 51, 72–80.
- [9] Web 1, (2019), <https://apps.who.int/iris/handle/10665/112642> (Accessed: 17/01/2023).
- [10] Prestinaci F., Pezzotti P., Pantosti A., (2015), “Antimicrobial resistance: a global multifaceted phenomenon”, *Pathog Glob Health*, 109 (7), 309–318.
- [11] Web 2, (2019), <https://www.cdc.gov/drugresistance/pdf/threats-report/2019-ar-threats-report-508.pdf> (Accessed: 17/01/2023).
- [12] Murray C. J., (2022), “Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis”, *The Lancet*, 399 (10325), 629–655.
- [13] Blair, M. A. Webber M. A., Baylay A. J., Ogbolu D. O., Piddock L. J. V., (2015), “Molecular mechanisms of antibiotic resistance”, *Nat Rev Microbiol*, 13 (1), 42–51.

- [14] Diallinas G., 2016, “Dissection of Transporter Function: From Genetics to Structure”, *Trends in Genetics*, 32 (9), 576–590.
- [15] Redhu A. K., (2018), “Molecular Basis of Substrate Polyspecificity of the *Candida albicans* Mdr1p Multidrug/H<sup>+</sup> Antiporter”, *J Mol Biol*, 430, 682–694.
- [16] Shi Y., (2015), “Common Folds and Transport Mechanisms of Secondary Active Transporters”, *Annurev Biophysics*, 42 (1), 51–72.
- [17] Yan N., (2015), “Structural Biology of the Major Facilitator Superfamily Transporters”, *Annurev Biophysics*, 44, 257–283.
- [18] Levy S. B., (1992), “Active efflux mechanisms for antimicrobial resistance.” *Antimicrob Agents Chemother*, 36 (4), 695.
- [19] Higgins C. F., (2007), “Multiple molecular mechanisms for multidrug resistance transporters”, *Nature*, 446 (7137), 749–757.
- [20] Sadée W., Drübbisch V., Amidon G. L., (1995), “Biology of membrane transport proteins”, *Pharm Res*, 12 (12), 1823–1837.
- [21] West I. C., (1980), “Energy coupling in secondary active transport”, *Biochim Biophys Acta*, 604 (1), 91–126.
- [22] Boudker O., Verdon G., (2010), “Structural perspectives on secondary active transporters”, *Trends Pharmacol Sci*, 31 (9), 418–426.
- [23] Kumar S., Varela M. F., (2012), “Biochemistry of Bacterial Multidrug Efflux Pumps”, *Int J Mol Sci*, 13 (4), 4484.
- [24] Kumar S., Lekshmi M., Parvathi A., Ojha M., Wenzel N., Varela M. F., (2020), “Functional and Structural Roles of the Major Facilitator Superfamily Bacterial Multidrug Efflux Pumps”, *Microorganisms*, 8 (2), 2.
- [25] Gregor M., Dang S., Yan N., Ronald Kaback H., (2013), “Evolutionary mix-and-match with MFS transporters”, *Proc Natl Acad Sci*, 110 (15), 5870–5874.
- [26] Web 3, (2023), <https://tcdb.org/search/result.php?tc=2.A.1.9>, (Accessed: 18/01/2023).
- [27] Drew D., North R. A., Nagarathinam K., Tanabe M., (2021), “Structures and General Transport Mechanisms by the Major Facilitator Superfamily (MFS)”, *Chem Rev*, 121 (9), 5289–5335.
- [28] Zhang X. C., Zhao Y., Heng J., Jiang D., (2015), “Energy coupling mechanisms of MFS transporters”, *Protein Science*, 24 (10), 1560–1579.
- [29] Griffith J. K., (1992), “Membrane transport proteins: implications of sequence comparisons”, *Curr Opin Cell Biol*, 4 (4), 684–695.

- [30] Henderson P. J. F., (1991), “Studies of translocation catalysis”, *Biosci Rep*, 11 (6), 477–538.
- [31] Reddy V. S., Shlykov M. A., Castillo R., Sun E. I., Saier M. H., (2012), “The Major Facilitator Superfamily (MFS) Revisited”, *FEBS J*, 279 (11), 2022.
- [32] Yamaguchi A., Ono N., Akasaka T., Noumi T., Sawai T., (1990) “Metal-tetracycline/H<sup>+</sup> antiporter of *Escherichia coli* encoded by a transposon, Tn10. The role of the conserved dipeptide, Ser65-Asp66, in tetracycline transport.”, *Journal of Biological Chemistry*, 265 (26), 15525–15530.
- [33] Kimura T., Shiina Y., Sawai T., Yamaguchi A., (1998), “Cysteine-scanning mutagenesis around transmembrane segment III of Tn10-encoded metal-tetracycline/H<sup>+</sup> antiporter”, *J Biol Chem*, 273 (9), 5243–5247.
- [34] Yamaguchi A., Akasaka T., Kimura T., Sakai T., Adachi Y., Sawai T., (1993), “Role of the conserved quartets of residues located in the N- and C-terminal halves of the transposon Tn10-encoded metal-tetracycline/H<sup>+</sup> antiporter of *Escherichia coli*”, *Biochemistry*, 32 (21), 5698–5704.
- [35] Kimura T., Nakatani M., Kawabe T., Yamaguchi A., (1998), “Roles of conserved arginine residues in the metal-tetracycline/H<sup>+</sup> antiporter of *Escherichia coli*”, *Biochemistry*, 37 (16), 5475–5480.
- [36] Jiang D., (2013), “Structure of the YajR transporter suggests a transport mechanism based on the conserved motif A”, *Proc Natl Acad Sci*, 110 (36), 14664–14669.
- [37] Marger M. D., Saier M. H., (1993), “A major superfamily of transmembrane facilitators that catalyse uniport, symport and antiport”, *Trends Biochem Sci*, 18 (1), 13–20.
- [38] Varela M. F., Sansom C. E., Griffith J. K., (1995), “Mutational analysis and molecular modelling of an amino acid sequence motif conserved in antiporters but not symporters in a transporter superfamily”, *Mol Membr Biol*, 12 (4), 313–319.
- [39] Ginn S. L., Brown M. H., Skurray R. A., (2000), “The TetA(K) tetracycline/H(+) antiporter from *Staphylococcus aureus*: mutagenesis and functional analysis of motif C” *J Bacteriol*, 182 (6), 1492–1498.
- [40] Konishi S., Iwaki S., Kimura-Someya T., Yamaguchi A., (1999), “Cysteine-scanning mutagenesis around transmembrane segment VI of Tn10-encoded metal-tetracycline/H(+) antiporter” *FEBS Lett*, 461 (3), 315–318.
- [41] Jin J., Krulwich T. A., (2002), “Site-directed mutagenesis studies of selected motif and charged residues and of cysteines of the multifunctional tetracycline efflux protein Tet(L)”, *J Bacteriol*, 184 (6), 1796–1800
- [42] Jesus M., Jin J., Guffanti A. A., Krulwich T. A., (2005), “Importance of the GP dipeptide of the antiporter motif and other membrane-embedded proline

and glycine residues in tetracycline efflux protein Tet(L)", *Biochemistry*, 44 (38), 12896–12904.

- [43] Newstead S., (2011), "Crystal structure of a prokaryotic homologue of the mammalian oligopeptide-proton symporters, PepT1 and PepT2", *EMBO J*, 30 (2), 417–426.
- [44] Kumar H., Kasho V., Smirnova I., Finer-Moore J. S., Kaback H. R., Stroud R. M., (2014), "Structure of sugar-bound LacY", *Proc Natl Acad Sci USA*, 111 (5), 1784–1788.
- [45] Sun L., (2012), "Crystal structure of a bacterial homologue of glucose transporters GLUT1-4", *Nature*, 490 (7420), 361–366.
- [46] Quistgaard E. M., Löw C., Moberg P., Trésaugues L., Nordlund P., (2013), "Structural basis for substrate transport in the GLUT-homology family of monosaccharide transporters", *Nat Struct Mol Biol*, 20 (6), 766–768.
- [47] Deng D., (2015), "Molecular basis of ligand recognition and transport by glucose transporters" *Nature*, 526 (7573), 391–396.
- [48] Yan H., (2013), "Structure and mechanism of a nitrate transporter", *Cell Rep*, 3 (3), 716–723.
- [49] Pedersen B. P., (2013), "Crystal structure of a eukaryotic phosphate transporter" *Nature*, 496 (7446), 533–536.
- [50] Ethayathulla A. S., Yousef M. S., Amin A., Leblanc G., Kaback H. R., Guan L., (2014), "Structure-based mechanism for Na<sup>+</sup>/melibiose symport by MelB", *Nature Communications*, 5 (1), 1–11.
- [51] Quistgaard E. M., Löw C., Guettou F., Nordlund P., (2016), "Understanding transport by the major facilitator superfamily (MFS): structures pave the way", *Nat Rev Mol Cell Biol*, 17 (2), 123–132.
- [52] Huang Y., Lemieux M. J., Song J., Auer M., Wang D. N., (2003), "Structure and mechanism of the glycerol-3-phosphate transporter from *Escherichia coli*", *Science*, 301 (5633), 616–620.
- [53] Law C. J., Maloney P. C., Wang D. N., (2008), "Ins and outs of major facilitator superfamily antiporters", *Annu Rev Microbiol*, 62, 289–305.
- [54] Henderson P. J. F., (1991), "Sugar transport proteins", *Curr Opin Struct Biol*, 1 (4), 590–601.
- [55] Pao S. S., Paulsen I. T., Saier M. H., (1998), "Major facilitator superfamily", *Microbiol Mol Biol Rev*, 62 (1), 1–34.
- [56] Neyfakh A. A., (2002), "Mystery of multidrug transporters: the answer can be simple", *Mol Microbiol*, 44 (5), 1123–1130.

- [57] Zheleznova E. E., Markham P. N., Neyfakh A. A., Brennan R. G., (1997), "Preliminary structural studies on the multi-ligand-binding domain of the transcription activator, BmrR, from *Bacillus subtilis*.", *Protein Sci*, 6 (11), 2465.
- [58] Zheleznova E. E., Markham P. N., Neyfakh A. A., Brennan R. G., (1999), "Structural basis of multidrug recognition by BmrR, a transcription activator of a multidrug transporter", *Cell*, 96 (3), 353–362, Feb. 1999, doi: 10.1016/S0092-8674(00)80548-6.
- [59] Augustin R., (2010), "The protein family of glucose transport facilitators: It's not only about glucose after all", *IUBMB Life*, 62 (5), 315–333.
- [60] Cura A. J., Carruthers A., (2012), "Role of monosaccharide transport proteins in carbohydrate assimilation, distribution, metabolism, and homeostasis", *Compr Physiol*, 2 (2), 863–914.
- [61] Smith D. E., Cl  men  on B., Hediger M. A., (2013), "Proton-coupled oligopeptide transporter family SLC15: physiological, pharmacological and pathological implications", *Mol Aspects Med*, 34 (2–3), 323–336.
- [62] Roth M., Obaidat A., Hagenbuch B., (2012), "OATPs, OATs and OCTs: the organic anion and cation transporters of the SLCO and SLC22A gene superfamilies", *Br J Pharmacol*, 165 (5), 1260–1287.
- [63] Hagenbuch B., Stieger B., (2013), "The SLCO (former SLC21) superfamily of transporters", *Mol Aspects Med*, 34 (2–3), 396–412.
- [64] Koepsell H., (2013), "The SLC22 family with transporters of organic cations, anions and zwitterions", *Mol Aspects Med*, 34 (2–3), 413–435.
- [65] Reimer R. J., (2013), "SLC17: A functionally diverse family of organic anion transporters", *Mol Aspects Med*, 34 (0), 350
- [66] Zhao R., Goldman I. D., (2013), "Folate and thiamine transporters mediated by facilitative carriers (SLC19A1-3 and SLC46A1) and folate receptors", *Mol Aspects Med*, 34 (2–3), 373–385.
- [67] Zgurskaya H. I., Nikaido H., (2000), "Multidrug resistance mechanisms: drug efflux across two membranes", *Mol Microbiol*, 37 (2), 219–225.
- [68] Alegre K. O., Paul S., Labarbuta P., Law C. J., (2016), "Insight into determinants of substrate binding and transport in a multidrug efflux protein", *Scientific Reports*, 6 (1), 1–12.
- [69] Lewinson O., Adler J., Sigal N., Bibi E., (2006), "Promiscuity in multidrug recognition and transport: the bacterial MFS Mdr transporters", *Mol Microbiol*, 61 (2), 277–284.
- [70] Wong K., Ma J., Rothnie A., Biggin P. C., Kerr I. D., (2014), "Towards understanding promiscuity in multidrug efflux pumps", *Trends Biochem Sci*, 39 (1), 8–16.

- [71] Reddy P. N., Srirama K., Dirisala V. R., (2017), “An Update on Clinical Burden, Diagnostic Tools, and Therapeutic Options of *Staphylococcus aureus*”, *Infect Dis*, 10.
- [72] Tong S. Y. C., Davis J. S., Eichenberger E., Holland T. L., Fowler V. G., (2015) “*Staphylococcus aureus* Infections: Epidemiology, Pathophysiology, Clinical Manifestations, and Management”, *Clin Microbiol Rev*, 28 (3), 603.
- [73] Costa L. M., (2016), “Inhibition of the NorA efflux pump of *Staphylococcus aureus* by synthetic riparins”, *J Appl Microbiol*, 121 (5), 1312–1322.
- [74] Mah T. F. C., . O’Toole G. A, (2001), “Mechanisms of biofilm resistance to antimicrobial agents”, *Trends Microbiol*, 9 (1), 34–39.
- [75] Tuchscher L., Löffler B., (2016), “*Staphylococcus aureus* dynamically adapts global regulators and virulence factor expression in the course from acute to chronic infection”, *Curr Genet*, 62 (1), 15–17.
- [76] Akanbi O. E., Njom H. A., Fri J., Otigbu A. C., Clarke A. M., (2017), “Antimicrobial Susceptibility of *Staphylococcus aureus* Isolated from Recreational Waters and Beach Sand in Eastern Cape Province of South Africa”, *Int J Environ Res Public Health*, 14, 9.
- [77] Schindler B. D., Jacinto P., Kaatz G. W., (2013), “Inhibition of drug efflux pumps in *Staphylococcus aureus*: current status of potentiating existing antibiotics”, *Future Microbiol*, 8 (4), 491–507.
- [78] Foster T. J., (2017), “Antibiotic resistance in *Staphylococcus aureus*. Current status and future prospects”, *FEMS Microbiol Rev*, 41 (3), 430–449.
- [79] Yu J. L., Grinius L., Hooper D. C., (2002), “NorA Functions as a Multidrug Efflux Protein in both Cytoplasmic Membrane Vesicles and Reconstituted Proteoliposomes”, *J Bacteriol*, 184 (5), 1370.
- [80] Costa S. S., (2019), “Genetic diversity of *norA*, coding for a main efflux pump of *Staphylococcus aureus*”, *Front Genet*, 10, 710.
- [81] Ubukata K., Itoh-Yamashita N., Konno M., (1989), “Cloning and expression of the *norA* gene for fluoroquinolone resistance in *Staphylococcus aureus*”, *Antimicrob Agents Chemother*, 33 (9), 1535–1539.
- [82] Yoshida H., Bogaki M., Nakamura S., Ubukata K., Konno M., (1990), “Nucleotide sequence and characterization of the *Staphylococcus aureus* *norA* gene, which confers resistance to quinolones”, *J Bacteriol*, 172 (12), 6942–6949.
- [83] Kaatz G. W., Seo S. M., Ruble C. A., (1993), “Efflux-mediated fluoroquinolone resistance in *Staphylococcus aureus*”, *Antimicrob Agents Chemother*, 37 (5), 1086–1094.
- [84] Noguchi N., Okada H., Narui K., Sasatsu M., (2004), “Comparison of the nucleotide sequence and expression of *norA* genes and microbial

susceptibility in 21 strains of *Staphylococcus aureus*”, *Microb Drug Resist*, 10 (3), 197–203.

- [85] Papkou A., Hedge J., Kapel N., Young B., MacLean R. C., (2020), “Efflux pump activity potentiates the evolution of antibiotic resistance across *S. aureus* isolates”, *Nat Commun*, 11, 1.
- [86] Bambeke F., Pages J.M., Lee V., (2006), “Inhibitors of bacterial efflux pumps as adjuvants in antibiotic treatments and diagnostic tools for detection of resistance by efflux”, *Recent Pat Antiinfect Drug Discov*, 1 (2), 157–175.
- [87] Zimmermann S., Tuchscher L., Rödel J., Löffler B., Bohnert J. A., (2017), “Optimized efflux assay for the NorA multidrug efflux pump in *Staphylococcus aureus*”, *J Microbiol Methods*, 142, 39–40.
- [88] Kaatz G. W., Thyagarajan R. V., Seo S. M., (2005), “Effect of promoter region mutations and *mgrA* overexpression on transcription of *norA*, which encodes a *Staphylococcus aureus* multidrug efflux transporter”, *Antimicrob Agents Chemother*, 49 (1), 161–169.
- [89] Espinoza J., (2019), “Essential oil, extracts, and sesquiterpenes obtained from the heartwood of *Pilgerodendron uviferum* act as potential inhibitors of the *Staphylococcus aureus* NorA multidrug efflux pump”, *Front Microbiol*, 10, 337.
- [90] Oliveira-Tintino C. D. M., (2021), “The 1,8-naphthyridines sulfonamides are NorA efflux pump inhibitors”, *J Glob Antimicrob Resist*, 24, 233–240.
- [91] El-Baky R. M. A., Sandle T., John J., Abuo-Rahma G. E. D. A., Hetta H. F., “A novel mechanism of action of ketoconazole: inhibition of the NorA efflux pump system and biofilm formation in multidrug-resistant *Staphylococcus aureus*”, *Infect Drug Resist*, 12, 1703–1718.
- [92] Tintino S. R., (2020), “Effect of Vitamin K3 Inhibiting the Function of NorA Efflux Pump and Its Gene Expression on *Staphylococcus aureus*”, *Membranes (Basel)*, 10 (6), 1–18.
- [93] Zimmermann S., (2019), “Clinically Approved Drugs Inhibit the *Staphylococcus aureus* Multidrug NorA Efflux Pump and Reduce Biofilm Formation”, *Front Microbiol*, 10, 2762.
- [94] Brincat J. P., (2011), “Discovery of novel inhibitors of the NorA multidrug transporter of *staphylococcus aureus*”, *J Med Chem*, 54 (1), 354–365.
- [95] Sabatini S., (2011), “Evolution from a natural flavones nucleus to obtain 2-(4-Propoxyphenyl) quinoline derivatives as potent inhibitors of the *S. aureus* NorA efflux pump”, *J Med Chem*, 54 (16), 5722–5736.
- [96] Holler J. G., Slotved H. C., Molgaard P., Olsen C. E., Christensen S. B., (2012), “Chalcone inhibitors of the NorA efflux pump in *Staphylococcus aureus* whole cells and enriched everted membrane vesicles”, *Bioorg Med Chem*, 20 (14), 4514–4521.

- [97] Pieroni M., Sabatini S., Massari S., Kaatz G. W., Cecchetti V., Tabarrini O., (2012), "Searching for innovative quinolone-like scaffolds: synthesis and biological evaluation of 2,1-benzothiazine 2,2-dioxide derivatives", *Medchemcomm*, 3 (9), 1092–1097.
- [98] Lepri S., (2016), "Indole Based Weapons to Fight Antibiotic Resistance: A Structure-Activity Relationship Study", *J Med Chem*, 59 (3), 867–891.
- [99] Buonerba F., (2017), "Improved Potency of Indole-Based NorA Efflux Pump Inhibitors: From Serendipity toward Rational Design and Development", *J Med Chem*, 60 (1), 517–523.
- [100] Paulsen I. T., Brown M. H., Skurray R. A., (1996), "Proton-dependent multidrug efflux systems", *Microbiol Rev*, 60 (4), 575–608.
- [101] Handzlik J., Matys A., Kieć-Kononowicz K., (2013), "Recent Advances in Multi-Drug Resistance (MDR) Efflux Pump Inhibitors of Gram-Positive Bacteria *S. aureus*" *Antibiotics (Basel)*, 2 (1), 28–45.
- [102] Palazzotti D., (2019), "Deciphering the Molecular Recognition Mechanism of Multidrug Resistance *Staphylococcus aureus* NorA Efflux Pump Using a Supervised Molecular Dynamics Approach", *Int J Mol Sci*, 20, 16.
- [103] Shang Y., (2022), "Evolutionary conservative analysis revealed novel functional sites in the efflux pump NorA of *Staphylococcus aureus*", *Journal of Antimicrobial Chemotherapy*, 77 (3), 675–681.
- [104] Brawley D. N., (2022), "Structural basis for inhibition of the drug efflux pump NorA from *Staphylococcus aureus*", *Nat Chem Biol*, 18 (7), 706–712.
- [105] Web 4, (2023), [https://www.wwpdb.org/pdb?id=pdb\\_000071o7](https://www.wwpdb.org/pdb?id=pdb_000071o7), (Accessed: 23/01/2023).
- [106] Lage O. M., Ramos M. C., Calisto R., Almeida E., Vasconcelos V., Vicente F., (2018), "Current Screening Methodologies in Drug Discovery for Selected Human Diseases", *Mar Drugs*, 16, 8.
- [107] Andrews J. M., (2001), "Determination of minimum inhibitory concentrations", *J Antimicrob Chemother*, 48 (1), 5–16.
- [108] He Y., (2016), "The changing 50% inhibitory concentration (IC<sub>50</sub>) of cisplatin: a pilot study on the artifacts of the MTT assay and the precise measurement of density-dependent chemoresistance in ovarian cancer", *Oncotarget*, 7 (43), 70803.
- [109] Sebaugh J. L., (2011), "Guidelines for accurate EC<sub>50</sub>/IC<sub>50</sub> estimation", *Pharm Stat*, 10 (2), 128–134.
- [110] Mendez D., (2019), "ChEMBL: Towards direct deposition of bioassay data", *Nucleic Acids Res*, 47 (1), 930–940.

- [111] Bento A. P., (2014), “The ChEMBL bioactivity database: an update”, *Nucleic Acids Res*, 42, 1083.
- [112] Web 5, (2023), [https://www.ebi.ac.uk/chembl/target\\_report\\_card/CHEMBL5114/](https://www.ebi.ac.uk/chembl/target_report_card/CHEMBL5114/), (Accessed: 20/01/2023).
- [113] Web 5, (2017), [https://www.scirp.org/\(S\(351jmbntvnsjt1aadkposzje\)\)/reference/ReferencesPapers.aspx?ReferenceID=2144573](https://www.scirp.org/(S(351jmbntvnsjt1aadkposzje))/reference/ReferencesPapers.aspx?ReferenceID=2144573), (Accessed: 17/01/2023).
- [114] Kelley L. A., Mezulis S., Yates C. M., Wass M. N., Sternberg M. J. E., (2015), “The Phyre2 web portal for protein modeling, prediction and analysis”, *Nat Protoc*, 10 (6), 845–858.
- [115] Waterhouse A., (2018) “SWISS-MODEL: homology modelling of protein structures and complexes”, *Nucleic Acids Res*, 46, (296).
- [116] Jumper J., (2021), “Highly accurate protein structure prediction with AlphaFold”, *Nature*, 596 (7873), 583–589.
- [117] Roy A., Kucukural A., Zhang Y., (2010), “I-TASSER: a unified platform for automated protein structure and function prediction”, *Nature Protocols*, 5 (4), 725–738.
- [118] Brylinski M., Skolnick J., (2010), “Comparison of structure-based and threading-based approaches to protein functional annotation”, *Proteins*, 78 (1), 18–134.
- [119] Web 6, (1990), <https://www.ncbi.nlm.nih.gov/protein/BAA14147.1>, (Accessed: 17/01/2023).
- [120] Web 7, (2020), <https://www.rcsb.org/structure/6sm4>, (Accessed: 23/01/2023).
- [121] Web 8, (2006), <https://www.rcsb.org/structure/2gfp>, (Accessed: 23/01/2023).
- [122] Web 9, (2021), <https://alphafold.ebi.ac.uk/>, (Accessed: 17/01/2023).
- [123] Bienert S., (2017), “The SWISS-MODEL Repository-new features and functionality”, *Nucleic Acids Res*, 45 (1), 313–319.
- [124] Studer G., Biasini M., Schwede T., (2014), “Assessing the local structural quality of transmembrane protein models using statistical potentials (QMEANBrane)”, *Bioinformatics*, 30 (17), 505.
- [125] Williams C. J., (2018), “MolProbity: More and better reference data for improved all-atom structure validation”, *Protein Sci*, 27 (1), 293–315.

- [126] Studer G., Rempfer C., Waterhouse A. M., Gumienny R., Haas J., Schwede T., (2020), “QMEANDisCo—distance constraints applied on model quality estimation”, *Bioinformatics*, 36 (6), 1765–1771.
- [127] Studer G., Biasini M., Schwede T., (2014), “Assessing the local structural quality of transmembrane protein models using statistical potentials (QMEANBrane)”, *Bioinformatics*, 30, 17.
- [128] Hollingsworth S. A., Karplus P. A., (2010), “A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins”, *Biomol Concepts*, 1 (3–4), 271.
- [129] McCammon J. A., Gelin B. R., Karplus M., (1977), “Dynamics of folded proteins”, *Nature*, 267 (5612), 585–590.
- [130] Karplus M., McCammon J. A., (2002), “Molecular dynamics simulations of biomolecules”, *Nat Struct Biol*, 9 (9), 646–652.
- [131] Hollingsworth S. A., Dror R. O., (2018), “Molecular Dynamics Simulation for All” *Neuron*, 99 (6), 1129–1143.
- [132] Lindorff-Larsen K., Maragakis P., Piana S., Eastwood M. P., Dror R. O., Shaw D. E., (2012), “Systematic validation of protein force fields against experimental data”, *PLoS One*, 7, 2.
- [133] Betz R., (2017), “Dabble”, *Zenodo*.
- [134] Jo S., Kim T., Iyer V. G., Im W., (2008), “CHARMM-GUI: a web-based graphical user interface for CHARMM”, *J Comput Chem*, 29 (11), 1859–1865.
- [135] Madhavi G., Adzhigirey M., Day T., Annabhimoju R., Sherman W., “Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments”, *J Comput Aided Mol Des*, 27 (3), 221–234.
- [136] Brooks B. R., (2009), “CHARMM: The Biomolecular Simulation Program”, *J Comput Chem*, 30 (10), 1545.
- [137] Lee J., (2016), “CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field” *J Chem Theory Comput*, 12 (1), 405–413.
- [138] Jo S., Kim T., Im W., (2007), “Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations”, *PLoS One*, 2 (9), 880.
- [139] Wu E. L., (2014), “CHARMM-GUI Membrane Builder toward realistic biological membrane simulations”, *J Comput Chem*, 35 (27), 1997–2004.

- [140] Jo S., Lim J. B., Klauda J. B., Im W., (2009), “CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes”, *Biophys J*, 97 (1), 50.
- [141] Lee J., (2019), “CHARMM-GUI Membrane Builder for Complex Biological Membrane Simulations with Glycolipids and Lipoglycans”, *J Chem Theory Comput*, 15 (1), 775–786.
- [142] Abraham M. J., (2015), “GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers”, *SoftwareX*, 1 (2), 19–25.
- [143] Bowers K. J. (2006), “Scalable algorithms for molecular dynamics simulations on commodity clusters”, *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing*, 6.
- [144] Eastman P., (2017), “OpenMM 7: Rapid development of high performance algorithms for molecular dynamics”, *PLoS Comput Biol*, 13, 7.
- [145] Phillips J. C., (2005), “Scalable molecular dynamics with NAMD”, *J Comput Chem*, 26 (16), 1781–1802.
- [146] Huang J., Mackerell A. D., (2013), “CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data”, *J Comput Chem*, 34 (25), 2135–2145.
- [147] Bakan A., Meireles L. M., Bahar I., (2011), “Structural bioinformatics ProDy: Protein Dynamics Inferred from Theory and Experiments”, *BIOINFORMATICS APPLICATIONS NOTE*, 27 (11), 1575–1577.
- [148] Volkamer A., Kuhn D., Rippmann F., Rarey M., (2012), “DoGSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment”, *Bioinformatics*, 28 (15), 2074–2075.
- [149] Goodford P. J., (1985), “A computational procedure for determining energetically favorable binding sites on biologically important macromolecules”, *J Med Chem*, 28 (7), 849–857.
- [150] Kastenholtz M. A., Pastor M., Cruciani G., Haaksma E. E. J., Fox T., (2000), “GRID/CPCA: a new computational tool to design selective ligands”, *J Med Chem*, 43 (16), 3033–3044.
- [151] Levitt D. G., Banaszak L. J., (1992), “POCKET: a computer graphics method for identifying and displaying protein cavities and their surrounding amino acids”, *J Mol Graph*, 10 (4), 229–234.
- [152] Laskowski R. A., (1995), “SURFNET: A program for visualizing molecular surfaces, cavities, and intermolecular interactions”, *J Mol Graph*, 13 (5), 323–330.

- [153] Glaser F., Morris R. J., Najmanovich R. J., Laskowski R. A., Thornton J. M., (2006), “A method for localizing ligand binding pockets in protein structures”, *Proteins*, 62 (2), 479–488.
- [154] Nocedal J., Wright S., (2006), “Solutions to the Selected Problems in Numerical Optimization”, 524.
- [155] Pagadala N. S., Syed K., Tuszynski J., (2017), “Software for molecular docking: a review”, *Biophys Rev*, 9 (2), 91–102.
- [156] Venkatachalam C. M., Jiang X., Oldfield T., Waldman M., (2003), “LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites”, *J Mol Graph Model*, 21 (4), 289–307.
- [157] Österberg F., Morris G. M., Sanner M. F., Olson A. J., Goodsell D. S., (2002), “Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock”, *Proteins*, 46 (1), 34–40.
- [158] Rarey M., Kramer B., Lengauer T., Klebe G., (1996), “A fast flexible docking method using an incremental construction algorithm”, *J Mol Biol*, 261 (3), 470–489.
- [159] Jones G., Willett P., Glen R. C., Leach A. R., Taylor R., (1997), “Development and validation of a genetic algorithm for flexible docking”, *J Mol Biol*, 267 (3), 727–748.
- [160] O’Boyle N. M., Banck M., James C. A., Morley C., Vandermeersch T., Hutchison G. R., (2011), “Open Babel: An Open chemical toolbox”, *J Cheminform*, 3 (10), 1–14.
- [161] Salentin S., Schreiber S., Haupt V. J., Adasme M. F., Schroeder M., (2015), “PLIP: fully automated protein-ligand interaction profiler”, *Nucleic Acids Res*, 43 (1), 443–447.
- [162] Bajusz D., Rácz A., Héberger K., (2015), “Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations?”, *J Cheminform*, 7 (1), 1–13.
- [163] Jain A. K., Murty M. N., Flynn P. J., (1999), “Data clustering”, *ACM Computing Surveys (CSUR)*, 31 (3), 264–323.
- [164] Zhang Z., Murtagh F., Poucke S., Lin S., Lan P., (2017), “Hierarchical cluster analysis in clinical research with heterogeneous study population: highlighting its visualization with R”, *Ann Transl Med*, 5 (4), 75–75.
- [165] Liu Q., Kwoh C. K., Li J., (2013), “Binding affinity prediction for protein-ligand complexes based on  $\beta$  contacts and B factor”, *J Chem Inf Model*, 53 (11), 3076–3085.

## **BIOGRAPHY**

My undergraduate degree was in Molecular Biology and Genetics, at Istanbul Medeniyet University, Turkey. During my studies, I have been mostly involved in wet-lab research. In my final year, I was introduced to Bioinformatics, Systems Biology and Structural Biology, which had a profound impact on my future research direction. I am currently pursuing a MSc study at Gebze Technical University under the Bioinformatics and Systems Biology program. Separately, I am a Research Assistant at Big Data and Bioinformatics Lab at Life Sciences and Biotechnology Institute, Bezmialem Vakif University, Turkey. Currently, I am involved in several projects covering a diverse areas of bioinformatics, namely genomics, transcriptomics, proteomics, and structural biology. I also have been granted by TUBITAK BIGG program and opening a company with my colleagues for data analysis.

## APPENDICES

### Appendix A: Publications Made in the Scope of Thesis Study

Işık E. B., Serçinoğlu O. (2022), Computational Characterization of Ligand Specificity and Promiscuity of Staphylococcus aureus NorA Efflux Pump, 6<sup>th</sup> GTU Graduate Studies Symposium, Gebze, Turkey, June 1-2, 2022.

Işık E. B., Serçinoğlu O. (2022), Computational Characterization of Ligand Specificity and Promiscuity of Staphylococcus aureus NorA Efflux Pump, 30<sup>th</sup> Conference on Intelligent Systems for Molecular Biology, Online, July 10-14, 2022.

Işık E. B., Serçinoğlu O. (2022), Investigation of the Ligand Binding Characteristics of Staphylococcus aureus NorA Efflux Pump, 15<sup>th</sup> International Symposium on Health Informatics and Bioinformatics, HIBIT, METU, Erdemli, Mersin, Turkey, October 20-21, 2022.

### Other Appendices (CD)

- **Ligands used in the study, with their metadata and binding site and Tanimoto score clusters obtained in the study.**
- **Comparison of models of NorA obtained by homology modeling and threading, with their structural assessment scores.**
- **Binding pockets on NorA conformations generated from MD simulations, with X-Y-Z coordinates and residues forming the binding pocket.**
- **Cluster of ligands based on Tanimoto similarity scores.**
- **PLIP results including the ligands, NorA conformations and binding site residues.**
- **Code files used for the study.**