

# Computational models of biophysical processes across the membrane

A Thesis

Submitted For the Degree of  
DOCTOR OF PHILOSOPHY  
in the Faculty of Science

by

Malay Ranjan Biswal



THEORETICAL SCIENCES UNIT  
JAWAHARLAL NEHRU CENTRE FOR ADVANCED SCIENTIFIC RESEARCH  
Bangalore – 560 064, India

AUGUST 2021



## DECLARATION

I hereby declare that the matter embodied in the thesis entitled “**Computational models of biophysical processes across the membrane**” is the result of investigations carried out by me at the Theoretical Sciences Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, India under the supervision of Dr. Meher K. Prakash and that it has not been submitted elsewhere for the award of any degree or diploma.

In keeping with the general practice in reporting scientific observations, due acknowledgement has been made whenever the work described is based on the findings of other investigators. Any omission that might have occurred by oversight or error of judgement is regretted.

*Malay Ranjan Biswal*

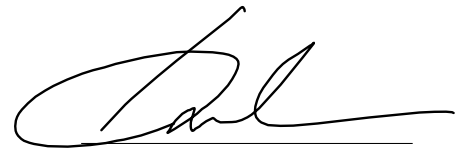
---

Malay Ranjan Biswal



## CERTIFICATE

I hereby certify that the matter embodied in this thesis entitled “**Computational models of biophysical processes across the membrane**” has been carried out by Mr. Malay Ranjan Biswal at the Theoretical Sciences Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, India under my supervision and that it has not been submitted elsewhere for the award of any degree or diploma.



Dr. Meher K. Prakash  
(Research Supervisor)



# Acknowledgements

In these six years, I have been so fortunate to have so many wonderful people who have tremendously helped me, directly and indirectly, to reach this milestone. It was a fantastic experience, and I learned many things. When I first came to JNCASR for the interview to the Theoretical Sciences Unit, I felt very nice to see such welcoming and supporting faculty here. Fortunately, I was selected here for the PhD under Dr Meher K Prakash in the Computational Biology lab and joined on the 28th of July 2015.

I am incredibly grateful to my supervisor, Dr Meher K Prakash, for his guidance and support since I joined him for my PhD. On the first day of my joining, he taught me that we should always thank people if we take help from them because we generally get help but forget to thank them later. It was the first lesson for my communication. He has always given importance to my communication skills, as I was terrible at them. Whatever little I have improved in my communication skill, the credit goes to him as he has been very patient with me and corrected every time I have made a mistake. He has always been very optimistic and encouraging about solving any problems and always there for help. He told me one beautiful thing: my problem is like my patient, and I have to keep it alive somehow; there is no escape. I had never thought solving biology problems using computational tools are so interesting. Thanks to him, he inspired me in these subjects. I have made so many mistakes these years, but he has always been extremely patient and always tried to correct me. He helped me how to think and how to approach problems. I have always enjoyed the interaction and long conversation with him, whether it is scientific or non-scientific. Apart from science, he has also taught me many valuable lessons which will always be helpful throughout life.

In the beginning, I came to know that to get a PhD I need to have a published article. And to have a published article, I should discover something new. That time I was scared if I can find anything new. Thanks to all the collaborators, Prof. Jayanta Halder, Prof. Ravi Manjithaya, Dr Jiaul Hoque, Dr Sreedevi Padmanabhan, Dr Sruthi C K, Dr Sandhya Rai and Ayan Majumder, for the exciting and fruitful discussions which helped to understand the field better. And with their help, it was easy to understand the subject and find some interesting phenomena in the area. I am very thankful to Sreedevi for

teaching me important things in biology.

I enjoyed the courses here, which were very interesting and a great learning experience. Thanks to all the course instructors, Prof. Umesh Waghmare, Prof. Swapan K. Pati, Prof. Kavita Jain, Prof. Srikanth Sastry, Prof. Vidhyadhiraja, Prof. Balasubramanian Sundaram, Prof. Ravi Manjithaya, Prof. Hemalatha Balaram, Prof. Kaustuv Sanyal, and Prof. Namita Surolia, I could get comfortable with programming and learn fascinating topics.

I acknowledge all the faculties from TSU, Prof. Umesh Waghmare, Prof. Swapan K. Pati, Prof. Kavita Jain, Prof. Srikanth Sastry, Prof. Sobhana Narasimhan, Prof. Vidhyadhiraja, Prof. Subir Das, for their support in my PhD time. It was always nice to interact with them. I also thank all the past and present chairman of TSU, secretary and Chandan for their support and help. I also thank all the members of the GSAC committee for helping in my PhD thesis.

During the absence of my supervisor, I never had to worry about my stipend. I thank Prof. Swapan K. Pati, Prof. Srikanth Sastry, Prof. Vidhyadhiraja and Prof. Subir Das for signing my scholarship form in the absence of my supervisor. They were always ready to help.

I am very thankful to JNCASR for financially supporting me in my PhD duration. I also thank everyone in the admin and academic section who processed all my official documents hassle-free. I thank all the Complab members who were very helpful in fixing any computer and internet related problems that made my work smooth. I am grateful to the CCMS administrators, Anoop and Suresh, for their help in using the clusters for my simulation. I thank all the mess staff and hostel staff who took care of the necessities like food and made my stay in the hostel very comfortable. I thank all the doctors and nurses at Dhanvantari for taking care of my health.

This PhD duration would have been tough without the presence of my highly supportive past and present labmates. I was really lucky to have such amazing people who made the lab atmosphere so cool and friendly. The non-scientific and sometimes utterly illogical discussions with them were always significant, and I always enjoyed them. I am deeply indebted to Sruthi for being there always whenever I needed any help. However busy she was, she would always come and help solve my problems. She taught me almost everything I required for my PhD. Thanks to her, I also started learning the Carnatic violin. Many thanks to Sandhya for all the help in my project and for all the festivals, cooking parties, and get-togethers we celebrated in her home with her husband, Antarip, who made those moments more enjoyable. I thank Soumalya for all the creative ideas and many planings of "*ek din hum yeh banate hain*" (let's make it one day). I very much appreciate Himanshu and Brijesh for giving important (and useless) updates about the



world news. I thank all other labmates, Anupam, for teaching smart usage of Linux and vim; Soutick, Ashlin, Ayan and Rashi, for all the fun-filled coffee break discussions.

I have greatly benefitted from many supportive seniors, Priyanka, Tarak, Sudip, and Anshul, who spent their valuable time and energy and helped me solve some troubles in my course work and projects. Special thanks to Priyanka for all the care and cycling trips near ICTS.

I thank all the Jnc buddies, Anusha, Shivaram, Usha, Ravishankar, Pradeep, PK, Divya, Rajendra, Ganesh, Priyanka and Deepak, for all the trips, Mafia, Uno, snacks and dinner parties, birthday celebrations, and fun conversations during lunch and dinner. Special thanks to Anusha for all the love, care, and always supporting me mentally and emotionally. Also, for long lectures on mental health and productivity. Thanks to Yagyik for helping me during coursework. I also acknowledge Abhiroop, Madhulika, Subhajit, Nalina, Manoj, Shreyan, Niloyendu, and Pushkar for the musical jamming sessions; specially Abhiroop for sharing various knowledge about music production.

I thank all my friends from Pondicherry University - Binay, Kiruthiga, Sadhana, Riju, Zeeshan, Nalini, Chinamayee, Pavani, Nikhil, Ganguly, Belormi, Shivshankar, Shakthi with whom I sometimes have long phone calls and messages. Thanks to Pavani for immensely helping me whenever I visited Kanpur and giving me a Lucknow tour.

I also would like to take this opportunity to thank the professors at Pondicherry University - Prof. SVM Sathyanarayana, Prof. Alok Sharan and Prof. KVP Lata, who had always inspired me to pursue research with their thoughtful advice.

And finally, a heartfelt thanks to my Bapa, Maa, Bhai and Bhauja, for all the love, support, sacrifices and helping me in whatever possible way they could do despite all challenges.



# Synopsis

A lot of fascinating phenomena occur across cell membranes. In this thesis, I address two interesting facets of life across cell membranes using computational tools – what makes membrane-active drugs effective, what is the signal that allows certain proteins to cross cell membranes even when they lack a conventional signal peptide. While both these subjects are areas of rapidly growing importance, the former is a well-explored subject with several gaps in knowledge arising from the antibiotic resistance and new classes of drug candidates, and the latter is mostly an uncharted territory.

In **Chapter 1**, I give a general introduction to the membrane active drugs as well as to the transport of proteins across membranes.

## *Part I. Mechanism of action and design of membrane-active drugs.*

Antibiotic resistance is a serious problem not only for global health but also for food security. The reason for this antibiotic resistance is that conventional antibiotics target crucial bacterial enzymes, and even a single point mutation in the enzyme can make antibiotics ineffective. Drugs which mimic antimicrobial peptides (AMPs) in targeting bacterial membranes thus have the potential to replace traditional antibiotics while combating antibiotic resistance. However, the use of AMPs is not an easily scalable solution, as they have high production costs, poor pharmacokinetics, and high susceptibility to proteolysis. A better understanding of the properties and mechanism of membrane active drugs and a basis for rational design may help improve the viability of AMPs or AMP-like drugs against pathogens. Chapters 2-5 discuss some of these ideas.

## *Part II. Unconventional protein secretion across membranes.*

Most of the proteins that are secreted out of the cell for various metabolic activities and functions, follow a conventional pathway through the endoplasmic reticulum and Golgi apparatus. All these proteins have a signal peptide at the N-terminus. Interestingly, several recent studies show secretion of proteins under cellular stress, even in the absence of conventional signal peptides. The way this unconventional protein secretion (UPS)

is triggered remains an open question. Recent experimental data on the evidence of unconventional secretion allows the possibility of bioinformatic analyses for developing hypotheses. Chapter 6 and 7 tried to use these data to develop hypotheses for unconventional protein secretion.

### *Part I*

In **Chapter 2**, I present the detailed mechanism of selective action of a new class of drug-candidate molecules on bacterial membranes using advanced molecular dynamics simulations. Four different systems were studied with two different choices of the drug candidate by changing the charge spacing in the acyl chains and two different model membranes representing the bacterial and human RBC membranes. The study establishes that the basis for the selectivity of the action is dictated by the spacing of the charges that is commensurate with the thickness of the model bacterial membrane. The study also explores the simultaneous insertion and aggregation of these molecules using detailed atomistic simulations, something not explored even in the studies of AMPs to the best of our knowledge. The results show that the molecules are primarily attracted by the charges, and they initially self-aggregate and subsequently disrupt the membrane.

In **Chapter 3**, I present our work on a model for the rational design of AMPs against *A. baumannii*. *A. baumannii* is a pathogen which is mostly associated with hospital-acquired infections, and is an example of bacteria showing resistance to many antibiotics, including the last line of drugs. Despite its importance, very few experimental studies are performed to study the effect of AMPs against these bacteria. We trained an artificial neural network with all the available data on AMP activity against *A. baumannii*. The model was used to screen the library of naturally occurring AMPs for their effectiveness, simultaneously also validating their toxicity. The lead molecules detected in the process may serve as a starting point for performing detailed antibiotic activity studies on *A. baumannii*.

In **Chapter 4**, I present the results of our attempt to capture the activity of AMPs using short molecular dynamics simulations. We re-examine synthetic antimicrobial peptides, CAMEL which were earlier studied for their Quantitative Structure Activity Relationship (QSAR). These QSAR models were developed as meta-models using descriptive parameters that are derived from sequence based property predictors. Since biological parameters are usually associated with their dynamics, we used short molecular dynamics in octanol to obtain descriptive parameters of each peptide and developed an artificial neural network based model.

In **Chapter 5**, I introduce a new concept by asking how long membrane active drugs

remain effective as the bacteria adapt to this perturbation as well. Bacteria respond to membrane active drugs by changing their membrane composition, typically by lysylation. Thus, over time, the drug is faced with newer membrane compositions and its activity against them needs to remain high. Conceptually one drug molecule weighed against bacterial membranes that are continuously adapting is the opposite of a typical scenario of screening multiple drugs against one bacterial target such as the works described in **Chapter 3** and **Chapter 4**. But specificity against the membrane of the targeting bacteria has not been clearly understood. We curated data of activity of daptomycin against *Staphylococcus aureus* from various experimental studies and using an artificial neural network model, we could explore for the first time the effect of the same drug on multiple membrane composition.

## *Part II*

In **Chapter 6**, I begin part II of the thesis which focuses on finding a signature motif for unconventional protein secretion (UPS). The only experimental evidence till date which identified a potential signature motif for UPS is the one which identified that a diacidic 'DE' motif was important in the secretion of SOD1 protein. Since the two amino acid motif may be quite ubiquitous, we ask if it is possible to identify the structural context in which this DE motif appears to better characterize the nature of the potential signature motif. Using charge, hydrophobicity and structural flexibility around the DE motifs in the 57 UPS proteins and 43 non-secreted proteins we curated from different physiological and pathological conditions, we could identify that the DE in an ordered region has higher odds of being an UPS signal.

In **Chapter 7**, we continue on the general theme of finding the UPS signature motif by analyzing the recent data on around 200 RNA binding proteins that are secreted through the autophagy pathway. Using this largest data set of established UPS cargos, we performed several bioinformatic analyses searching for signal motifs that are 2, 3 or 4 amino acids long. Our results in which we compare these RNA binding proteins to the entire large data set of conventionally secreted proteins led us to the finding that presence of one of the four tri-acidic motifs (DDE, DED, DEE, EEE) in the neighborhood of acts as a statistically significant discriminator of UPS proteins. The discovery is incidentally validated using the SOD1 data that is discussed in chapter 6, is one of the early hypotheses in this new and developing field that requires further validation.

Thus in my thesis, I use a range of techniques – artificial neural networks, molecular dynamics, bioinformatic analyses to ask some of the questions that are arising newly in

the fields of drug action on membranes or transport of proteins across membranes.

# List of Publications

1. S. Padmanabhan, M. R. Biswal, R. Manjithaya and M. Prakash, “Exploring the context of diacidic motif DE as a signal for unconventional protein secretion in eukaryotic proteins,” *Wellcome Open Research.*, vol. 13, no. 6, 2018
2. M. R. Biswal\*, S. Rai\*, and M. Prakash, “Molecular dynamics based antimicrobial activity descriptors for synthetic cationic peptides,” *Journal of Chemical Sciences*, vol. 131, no. 2, 2019
3. A. Majumder\*, M. R. Biswal\*, and M. Prakash, “One Drug Multiple Targets: An Approach To Predict Drug Efficacies on Bacterial Strains Differing in Membrane Composition,” *ACS Omega*, vol. 4, no. 3, 2019
4. A. Majumder\*, M. R. Biswal\*, and M. Prakash, “Computational screening of antimicrobial peptides for *Acinetobacter baumannii*,” *Plos one*, vol. 14, no. 10, 2019
5. M. R. Biswal\*, S. Padmanabhan\*, R. Manjithaya and M. Prakash, “Early bioinformatic Implication for a triacidic amino acid motifs in Autophagy-Dependent Unconventional Secretion of Mammalian Proteins,” *Frontiers in cell and developmental biology*, vol. 10, 2022
6. M. R. Biswal, S. Rai, C. K. Sruthi, J. Hoque, J. Haldar and M. Prakash, “Detailing the Molecular Mechanism of Selective Antibiotic Action Against Bacterial Membranes via Self Assembly and Membrane Penetration,” (*manuscript under preparation*)

\*Authors contributed equally





# Contents

Acknowledgements	v
Synopsis	ix
List of Publications	xiii
List of Figures	xxi
List of Tables	xxix
<b>1 Introduction</b>	<b>1</b>
1.1 Life across the membrane . . . . .	1
1.2 Antibiotic Resistance . . . . .	2
1.3 Antimicrobial Peptides and their action mechanism . . . . .	3
1.3.1 Rational strategies to improve or replace AMPs . . . . .	4
1.3.2 Challenges and problems in the AMPs . . . . .	6
1.4 Unconventional Protein Secretion . . . . .	7
1.4.1 Computational methods . . . . .	8
1.4.2 Challenges and recent development . . . . .	9
Bibliography . . . . .	9
<b>PART I: Mechanism of action and design of membrane-active drugs</b>	<b>17</b>
<b>2 Detailing the Molecular Mechanism of Selective Antibiotic Action Against Bacterial Membranes via Self Assembly and Membrane Penetration</b>	<b>19</b>

<b>Abstract</b>	<b>20</b>
2.1 Introduction . . . . .	20
2.2 Methods . . . . .	21
2.3 Results . . . . .	24
2.3.1 Membrane is not disrupted by a single molecule . . . . .	24
2.3.2 Clusters of the molecules disrupts membranes . . . . .	24
2.3.3 Selectivity in membrane disruption . . . . .	25
2.3.4 Mechanism of drug interaction with membrane . . . . .	28
2.4 Discussion . . . . .	30
2.4.1 Self assembly, aggregation, disruption . . . . .	30
2.4.2 Differential selectivity of the molecules . . . . .	30
2.5 Conclusions . . . . .	31
Bibliography . . . . .	32
<b>3 Computational screening of antimicrobial peptides for <i>Acinetobacter baumannii</i></b>	<b>37</b>
<b>Abstract</b>	<b>38</b>
3.1 Introduction . . . . .	38
3.2 Methods . . . . .	39
3.2.1 Curation of data . . . . .	39
3.2.2 Parameter computation . . . . .	39
3.2.3 Artificial neural network . . . . .	40
3.3 Results . . . . .	41
3.3.1 Curated data for AMPs and their effectiveness . . . . .	41
3.3.2 Quantitative Models for AMP activity . . . . .	41
3.3.3 Selecting the best model . . . . .	42
3.3.4 Predicting the results for naturally occurring AMPs . . . . .	43
3.3.5 Parameter importance in model . . . . .	45
3.3.6 Relevance of predictions for MDR strains . . . . .	45
3.3.7 Conclusions . . . . .	46
Bibliography . . . . .	46
<b>4 Molecular dynamics based antimicrobial activity descriptors for synthetic cationic peptides</b>	<b>55</b>
<b>Abstract</b>	<b>56</b>
4.1 Introduction . . . . .	56

4.2	Methods . . . . .	57
4.2.1	Structure and dynamics simulation . . . . .	57
4.2.2	Neural Network Model . . . . .	57
4.3	Results . . . . .	58
4.3.1	Dynamical properties . . . . .	58
4.3.2	Model for CAMEL-s activity prediction . . . . .	59
4.3.3	Importance of the variables used . . . . .	59
4.4	Discussion . . . . .	61
4.5	Summary . . . . .	62
	Bibliography . . . . .	62
<b>5</b>	<b>One drug multiple targets: An approach to predict drug efficacies on bacterial strains differing in membrane composition</b>	<b>65</b>
	<b>Abstract</b>	<b>66</b>
5.1	Introduction . . . . .	66
5.2	Results . . . . .	67
5.2.1	Membrane descriptors . . . . .	67
5.2.2	Membrane bilayer model . . . . .	68
5.2.3	Inner and outer leaflet models . . . . .	69
5.3	Discussion . . . . .	70
5.3.1	One drug multiple membrane targets . . . . .	70
5.3.2	Possible mechanism of action . . . . .	71
5.3.3	Predicting daptomycin activity . . . . .	71
5.4	Conclusion . . . . .	72
5.5	Methods . . . . .	73
	Bibliography . . . . .	75
<b>PART II: Unconventional protein secretion across membranes</b>		
<b>83</b>		
<b>6</b>	<b>Exploring the context of diacidic motif DE as a signal for unconventional protein secretion in eukaryotic proteins</b>	<b>85</b>
	<b>Abstract</b>	<b>86</b>
6.1	Introduction . . . . .	86
6.2	Methods . . . . .	87
6.2.1	Data curation . . . . .	87

6.2.2	Variables describing the physical context . . . . .	88
6.2.3	Most relevant DE motif . . . . .	88
6.2.4	Multivariate binary logistic regression . . . . .	89
6.2.5	LIR motif analysis . . . . .	89
6.2.6	Co-evolution analysis . . . . .	89
6.3	Results . . . . .	89
6.3.1	Physical nature of the DE neighbourhood as a context . . . . .	89
6.3.2	Signalling context via phosphorylation of the motif . . . . .	91
6.3.3	Signalling context via LC3 interacting region (LIR) analysis . . . . .	91
6.3.4	Co-evolutionary context . . . . .	92
6.4	Discussion . . . . .	92
6.5	Conclusions . . . . .	94
	Bibliography . . . . .	94
<b>7</b>	<b>Early bioinformatic evidence for a tri-acidic amino acid motif as a possible signature for autophagy dependent unconventional protein secretion</b>	<b>99</b>
	<b>Abstract</b>	<b>100</b>
7.1	Introduction . . . . .	100
7.2	Methods . . . . .	102
7.3	Results . . . . .	103
7.3.1	Acidic motifs top the differential motif analysis . . . . .	103
7.3.2	Acidic motifs appear in the proximity of LIR motifs . . . . .	104
7.3.3	Phosphorylatable amino acids occur preferentially in the UPS proteins	104
7.4	Discussion . . . . .	105
7.4.1	Triacidic motif is potentially a signal for UPS . . . . .	105
7.4.2	Phosphorylation may be activating the signal . . . . .	106
7.4.3	LIR motif in the proximity of triacidic motif is discriminatory . . . . .	107
7.4.4	Deriving the hypothesis for the signal for UPS . . . . .	107
7.5	Conclusions . . . . .	109
	Bibliography . . . . .	111
<b>8</b>	<b>Conclusions and Future Outlook</b>	<b>113</b>
<b>A</b>	<b>Detailing the Molecular Mechanism of Selective Antibiotic Action Against Bacterial Membranes via Self Assembly and Membrane Penetration</b>	<b>117</b>
	Bibliography . . . . .	121

<b>B</b>	<b>Computational screening of antimicrobial peptides for <i>Acinetobacter baumannii</i></b>	<b>123</b>
	Bibliography . . . . .	133
<b>C</b>	<b>Molecular Dynamics Based Antimicrobial Activity Descriptors for Synthetic Cationic Peptides</b>	<b>137</b>
<b>D</b>	<b>One drug multiple targets: An approach to predict drug efficacies on bacterial strains differing in membrane composition</b>	<b>141</b>
	Bibliography . . . . .	154
<b>E</b>	<b>Exploring the context of diacidic motif DE as a signal for unconventional protein secretion in eukaryotic proteins</b>	<b>157</b>



# List of Figures

2.1	(a) Chemical structure of the MeDiAms ‘A’ and ‘B’. The distance between the methylated ammonium group in molecule ‘A’ is larger than molecule ‘B’. (b) MIC and HC <sub>50</sub> of the two MeDiAms. Both have comparable MIC (~1.5 μg/ml and 2 μg/ml respectively), but very different HC <sub>50</sub> s (805 μg/ml and 60 μg/ml respectively). The structures of MeDiAms were optimized at B3LYP/6-31G* [22–24] level of theory. . . . .	22
2.2	Bacterial membrane with (a) molecule ‘A’ and (b) molecule ‘B’ (shown in red color) after 50 ns of classical molecular dynamics simulation. The phosphorous atom in the bacterial lipid head group is shown in golden color. After 50 ns, the molecule and membrane interaction results in a small membrane distortion, which is however far from the membrane disruption required for action. . . . .	24
2.3	Free energy profile of the system containing six molecules of ‘A’, obtained using umbrella sampling. At 7.4 Å free energy profile shows a metastable state. . . . .	25
2.4	Metastable structures corresponding to the minimum free energy configuration of the system obtained from the umbrella sampling simulations ( <b>Figure 2.3</b> ) of MeDiAm ‘A’ (a, b) and MeDiAm ‘B’ (c, d) are shown. Bacterial membrane that is disrupted the molecules ‘A’ in a representation (a) without showing the ‘A’ molecules (b) and with the ‘A’ molecules suggests the mechanism of action. Similarly, (c) and (d) show the disruption to the bacterial membrane by molecule ‘B’. The phosphate ions of the lipid head group interacts with the positively charged nitrogen of these molecules. Color representations in are the same as in <b>Figure 2.2a</b> . . . .	26

2.5	The dark patches in the membrane thickness map indicate a pinching in the membrane and potential for the antibiotic effectiveness of the molecule. Comparing the four plots explains the selectivity and effectiveness seen in the two molecules. . . . .	27
2.6	2D number density of the lipids along the x-y plane (A) Bacterial Membrane (Averaged over last 20 ns) (B) Molecule ‘A’ in bacteria (averaged over 6 ns) (C) Molecule ‘B’ in bacteria (averaged over 6 ns) . . . . .	27
2.7	(a)The free energy profile obtained from the Metadynamics simulations and the black solid line represents the minimum free energy pathway connecting the fully soluble molecules to the final disrupted membrane state. The radius of gyration of the molecules and the distance from the center of the membrane were used as two independent variables to capture the molecular aggregation and membrane insertion (b) Free energy profile along the minimum energy pathway mentioned above. (c) to (f) are the snapshot of the system at different states (A,B,C,D respectively), which shows the alignment and position of the molecules and the membrane. The membrane is denoted by the phosphorous (yellow balls) in the head group.	29
2.8	A snapshot describing the membrane disruption and the formation of a water channel across it. The disrupted membrane is shown (a) with the molecule and (b) without the MeDiAm molecules. One may notice that the phosphorous of the membrane interacts with the charged groups of the molecules thinning the membrane to the point that water penetrates it.	31
3.1	MIC versus different parameters. The AMPs used in the analysis along with the sequences and biophysical parameters are given in <b>Table B.1</b> of <b>Appendix B</b> . . . . .	42
3.2	Comparison of the experimental and calculated MIC ( $\mu\text{g/ml}$ ) of curated AMPs on <i>A. baumannii</i> obtained from Model-1, calculated by using 8 hidden neurons. Training (purple circles), validation (orange squares) and test (green diamonds) sets are shown. The data used in the analysis is shown in <b>Table B.1</b> of <b>Appendix B</b> . . . . .	43
3.3	The relative importance of the different parameters in Model-1 is shown (blue color). Aliphatic index influences the outcomes of the predictions the most in this model. The green line shows the average and standard deviation value of the variable importance by taking into account all the models of 10 fold cross validation. . . . .	46



4.1	(a)RMSD of the peptide was calculated with reference to the starting structure and (b) Ellipticity at 222 nm value calculated for some of the peptides solvated in octanol, shows that the physical parameters stabilizes after 5 ns. We used data from 5 ns to 10 ns in our analysis. . . . .	58
4.2	The experimental vs the calculated values of biological activity for the training (purple circle), validation (orange square) and the test set (green diamond). Blue color area corresponds to 95% confidence interval. . . . .	60
4.3	The relative importance of the different variables used in our model was calculated. The difference in $R^2$ of predictions when each individual variable is replaced by its average value is considered as a measure of importance. Av. denotes average in the figure above. . . . .	60
4.4	Comparison between the calculations from our work with those from earlier reported works.[18, 19] (a) Calculations from Cherkasov <i>et al.</i> training and validation sets ( $R^2 = 0.86$ ) and (b) test set ( $R^2 = 0.46$ ) (c) Calculations from Torrent <i>et al.</i> training and validation sets ( $R^2 = 0.75$ ) and (d) Test set ( $R^2 = 0.45$ ). Our calculations had $R^2$ 0.86, 0.62 and 0.65 for training, validation and test sets respectively. . . . .	61
5.1	Comparison of the experimental MIC ( $\mu\text{g}/\text{mL}$ ) and MIC ( $\mu\text{g}/\text{mL}$ ) of daptomycin calculated using 8 neurons in the hidden layer on different total membrane compositions. Training (purple circles), validation (orange squares) and test (green diamonds) sets are shown. The details of the source of the data as well the method by which the MICs were obtained are given in <b>Table D.1</b> of <b>Appendix D</b> . . . . .	69
5.2	Comparison of the experimental ( $\mu\text{g}/\text{mL}$ ) and MIC ( $\mu\text{g}/\text{ml}$ ) of daptomycin calculated on different membrane compositions, using the data only from (a) the outer leaflet, with 8 neurons in the hidden layer, or (b) the inner leaflet, with 4 neurons in the hidden layer. Training (purple circles), validation (orange squares) and test (green diamonds) sets are shown. The data used in the analysis is shown in <b>Table D.4</b> of <b>Appendix D</b> . . . . .	70
5.3	Prediction of daptomycin MIC ( $\mu\text{g}/\text{mL}$ ) values for a complete range of variation of lipid composition in the (a) outer leaflet and (b) total membrane. The blue-green color represents the %CL percentage for positive values of MIC and the circle size represents MIC. Since we trained our model on the non-monotonous trends ( <b>Figure D.1</b> of <b>Appendix D</b> ), the prediction from the model gives few negative MIC values ( $\mu\text{g}/\text{mL}$ ), represented in red color, over a small parametric region. . . . .	72

6.1	Illustration of the definition of structural order in NADP Isocitrate dehydrogenase (PDB:2B0T). The protein has multiple DE motifs, in ordered, loop and border regions. . . . .	88
6.2	Dependence of secretory nature on the local properties of DE motif in cancer protein data (Groups 3B, 3C in Dataset 1[20]) on the extreme values of the descriptive parameters. (A) Maximum charge (B) minimum hydrophobicity (C) minimum charge (D) maximum hydrophobicity of the neighbouring amino acids were calculated for all the occurrences of the DE motifs, in each of the proteins appearing in the cancer secretome data.	90
7.1	A reanalysis of the multiple sequence alignments obtained from Cruz et al.[6]. A comparison among the mammalian sequences, highlighted in purple, shows a common triacidic motif DEE (highlighted in yellow), rather than a diacidic motif (shown in red colored text) when comparing sequences across all species. . . . .	107
7.2	An analysis of the sequence-proximity of the triacidic motif with the LIR motif among some of the proteins from the class I of the UPS-ATG data set is shown. . . . .	108
7.3	An analysis of the structural-proximity of the triacidic motif with the LIR motif among the proteins from the class I of the UPS-ATG data set for which structures are known is shown. The blue and red colors indicate the LIR and triacidic motifs. For convenience, only the closest pair is shown and other occurrences of LIR or triacidic motifs are not shown. . . . .	109
7.4	An analysis of the structural-proximity of the triacidic as well as the KXX motifs with all LIR motifs occurring among the proteins from the class I of the UPS-ATG data set for which structures are known is shown. The blue, green and red colors indicate the LIR, KXX and triacidic motifs. One may notice that in some structures LIR is close to the triacidic motif, and in others to the KXX motif. . . . .	110
A.1	Thickness profile of the membrane (A) Bacterial Membrane (Averaged over last 20ns) (B) Molecule ‘A’ in bacteria (C) Molecule ‘B’ in bacteria; All data corresponding to the 7 Å distance between the center of mass of the aggregates for both molecule ‘A’ and ‘B’. The plots are averaged over last 6 ns of the simulation for molecules ‘A’ and ‘B’. . . . .	117

A.2	Thickness profile of the membrane (A) Human Membrane (Averaged over last 20ns) (B) Molecule ‘A’ in human (C) Molecule ‘B’ in human; All data is corresponding to the 7 Å distance between the center of mass of the aggregates for both molecule ‘A’ and ‘B’. The plots are averaged over last 6 ns of the simulation for molecules ‘A’ and ‘B’.	118
A.3	Deuterium order parameter (a) Chain 1 (b) Chain 2; data averaged over 6ns	118
A.4	Comparison of the rotation of the lipid molecules to the splay between the lipid acyl chains (A) Molecule ‘A’ in bacteria (B) Molecule ‘B’ in bacteria; for lipids within 5 Å of the molecules.	119
A.5	Formation of a water channel across the membrane (A) Molecule ‘B’ in bacteria (B) Molecule ‘A’ in bacteria	120
A.6	3D structure of (a) Molecule ‘A’ and (b) Molecule ‘B’.	120
A.7	Partial charges on lipid head group shown for DPPC, DPPE and DPPG lipid molecules.	120
A.8	3D and 2D structure of (a) DPPC, (b) DPPE and (c) DPPG are shown.	121
B.1	Histogram of all the parameters corresponding to the AMPs shown in <b>Table B.1</b> . The raw data of these parameters is in <b>Data File in GitHub</b> ( <a href="https://github.com/malayrb/Thesis/blob/main/Ch3/File_S2.xlsx">https://github.com/malayrb/Thesis/blob/main/Ch3/File_S2.xlsx</a> ).	126
B.2	Flow chart showing the schematic of how the ANN models were developed. The part embedded in the dotted box is the logic of selecting the best models in traditional QSAR method, if the goal had been to be content with the prediction for the test set with 7 peptides. In such methods, typically a person conducting the test or the person developing the model but chooses to stay blind to this additional test data, eventually compares the performance of the predictions on this small test set. However, in our case, the goal is to make predictions for the 2338 AMPs for which no activity measurements are yet available, a secondary validation criteria was used to screen the models further.	127
B.3	Comparison of the experimental MIC ( $\mu\text{g/ml}$ ) and predicted MIC ( $\mu\text{g/ml}$ ) values of AMP obtained from the 10 fold cross validation calculation. The neural network is trained with one hidden layer consisting of 6 neurons.	129
B.4	Comparison of the experimental MIC ( $\mu\text{g/ml}$ ) and predicted MIC ( $\mu\text{g/ml}$ ) values of AMP obtained from the 10 fold cross validation calculation. The neural network is trained with one hidden layer consisting of 8 neurons.	130

B.5	Comparison of the experimental MIC ( $\mu\text{g/ml}$ ) and predicted MIC ( $\mu\text{g/ml}$ ) values of AMP obtained from the 10 fold cross validation calculation. The neural network is trained with one hidden layer consisting of 10 neurons.	131
B.6	Comparison of the experimental and calculated MIC ( $\mu\text{g/ml}$ ) of curated AMPs on <i>A. baumannii</i> obtained from Model-2, calculated by using 6 hidden neurons. Training (purple circles), validation (orange squares) and test (green diamonds) sets are shown. The data used in the analysis for the peptides given in <b>Table B.1</b> . The raw data is in <b>Data File in GitHub</b> ( <a href="https://github.com/malayrb/Thesis/blob/main/Ch3/File_S2.xlsx">https://github.com/malayrb/Thesis/blob/main/Ch3/File_S2.xlsx</a> ).	132
B.7	The relative importance of the different parameters is shown for Model-2 (blue color). Aliphatic index continues to be the most relevant variable, similar to <b>Figure 3.3</b> . The green line shows the average and standard deviation value of the variable importance by taking into account all the models of 10 fold cross validation.	132
C.1	The SASA and volume over 10ns simulation shows the system gets stabilized within 10ns	138
C.2	Neural network predicted using simulation in water	139
D.1	Dependence of the experimentally determined daptomycin MIC values on the phospholipid compositions in a) total membrane, b) inner leaflet and c) outer leaflet. The radius of the circle represents MIC ( $\mu\text{g/ml}$ ) and the change in color represents CL concentration. The data shows ‘non-monotonous’ trends, as the magnitude of MIC goes through variations, rather than a straightforward increase with the parameters.	144
D.2	Comparison of the experimentally determined MIC ( $\mu\text{g/ml}$ ) of daptomycin with the MIC calculated using the 10-fold cross validation for total membrane composition. The neural network is trained with one hidden layer consisting of 8 neurons. The choice of the 8-neurons was made after trying calculations with different number of neurons. The mean square error and standard deviation for these choices of neurons are shown in <b>Table D.3</b> .	146
D.3	Comparison of the experimentally determined MIC of daptomycin ( $\mu\text{g/ml}$ ) and the MIC obtained from the 10-fold cross validation using the outer leaflet composition. The neural network is trained with one hidden layer consisting of 8 neurons.	150

D.4	Comparison of the experimentally determined MIC of daptomycin ( $\mu\text{g}/\text{ml}$ ) with the MIC calculated with 10-fold cross validation for inner leaflet composition. The neural network is trained with one hidden layer consisting of 4 neurons. . . . .	152
D.5	Dependence of daptomycin MIC ( $\mu\text{g}/\text{ml}$ ) values in our predictions across the entire range of %PG and %LPG compositions of the inner leaflet. The blue-green and red color represents the change in CL percentage for positive and negative MIC values respectively and radius of the circle increases with increase in absolute value of MIC. Negative MIC is not physically meaningful and we believe it is an artefact of training a non-monotonous variation of MIC on a limited data. . . . .	153
E.1	Comparison between the secretory and non-secretory proteins with their charges and hydrophobicity for various groups . . . . .	165



# List of Tables

3.1	Using the 2 different models, we predicted the activity of 2338 naturally occurring AMPs documented in the AMP database. The complete list of predictions are given in the <b>Data File in GitHub</b> ( <a href="https://github.com/malayrb/Thesis/blob/main/Ch3/File_S2.xlsx">https://github.com/malayrb/Thesis/blob/main/Ch3/File_S2.xlsx</a> ). However, of these the AMPs which had consistent predictions from both the models ( $\Delta\text{MIC} \leq 5 \mu\text{g/ml}$ ) were selected and presented in this table. All of these were peptides listed below were non-toxic according to the predictions from ToxinPred ( <a href="http://crdd.osdd.net/raghava/toxinpred/">http://crdd.osdd.net/raghava/toxinpred/</a> ).[58] . . . . .	44
5.1	10-fold cross validation analysis was repeated with a hidden layer between the input and output layers, and by varying the number of neurons in the hidden layer. Here we tabulated the number of times, out of 10, $R_{test}^2 > 0$ , as well as the mean squared error (MSE) and standard deviation (SD). MSE and SD were used for choosing the optimal number of neurons for each of the membrane compositions studied. . . . .	69
7.1	. . . . .	103
7.1(a)	Discriminatory motif analysis of the UPS-ATG set relative to the Swiss-Prot database. The E-value denotes the significance level of the hypothesis that the sequence appears selectively in the UPS-ATG dataset, and with these extremely small values of it, it is practically the same as the p-value.	103
7.1(b)	A summary of the motifs that occur at least 30% more in the UPS-ATG dataset than in the CPS dataset. The total number of proteins in both the sets as well as the percentage of occurrence in them are both shown.	103
7.2	Propensity of occurrence of the various motifs in the proximity of LIR motif in the unconventionally secreted proteins as well as in the conventionally secreted proteins. . . . .	105

7.3	The occurrence of proximal phosphorylatable amino acids close to motifs appearing in the differential analysis is shown here. The motifs are sorted in the decreasing order of occurrence in the UPS-ATG group. . . . .	106
7.4	The structural distance between the triacidic motif and the LIR region in some of the class I proteins from the UPS-ATG positive dataset, when there is no sequence proximity, is shown. . . . .	108
B.1	Details of all the 75 antimicrobial peptides and their experimental protocol used in the present work. Qualitative test set is indicated by blue color, and with a lower bound in MIC. . . . .	123
B.2	Experimental and predicted MIC values of the 9 qualitative AMP data that were used for an additional test. Before accepting a model, at least 5 of these 9 results were verified to be more than the lower bound suggested by the experiments (to within a factor of 2). An additional condition $R^2 > 0.6$ was used with the test data set from the quantitative data. . . .	128
B.3	10-fold cross validation analysis was performed with a hidden layer between the input and output layers. The hidden layer architecture with 6, 8, and 10 neurons respectively were independently modelled. Here we tabulated the number of times, out of 10, $R_{test}^2 > 0.6$ , as well as the mean squared error (MSE) and standard deviation (SD). The overall error was optimal with the choice of 8 neurons, although all three architectures performed satisfactorily. . . . .	128
B.4	Comparison of the experimental and predicted MIC values for the MDR <i>A. baumannii</i> strains. The results were satisfactory for most of these cases. Polydim-I has extremely low values in one of the parameter used in the model ( $\mu_H$ ), which is not falling in the training range. . . . .	133
D.1	Summary of the data curated from multiple experimental sources on different <i>S. aureus</i> strains is shown. Total membrane phospholipid composition is described in terms of phosphatidylglycerol (PG), lysyl-PG (LPG), cardiolipin (CL), inner leaflet LPG (iLPG) and outer leaflet LPG (oLPG) and the daptomycin MIC values are given in $\mu\text{g/mL}$ . All the MIC values are evaluated according to standard Etest, except when mentioned. . . . .	143
D.2	$R_{training}^2$ , $R_{validation}^2$ and $R_{test}^2$ values obtained from the 10 fold validation calculation for total membrane composition. The neural network is trained with one hidden layer consisting of 8 neurons. A summary of mean square error and standard deviation when the number of neurons is varied is shown in <b>Table D.3</b> . . . . .	145



D.3	Predictions for the data points that were not measured by Etest. The best model developed based on Etest results was used for making this cross-prediction for the data measured by other methods. All the data used for this calculation is given in <b>Table D.1 &amp; D.4</b> . . . . .	147
D.4	Inner and outer leaflet of membrane phospholipid compositions of the different <i>S. aureus</i> strains in terms of lysl-phosphatidylglycerol (lysl-PG), phosphatidylglycerol (PG), cardiolipin (CL) and the daptomycin MIC values in $\mu\text{g}/\text{ml}$ . These numbers were derived as discussed in the <b>Methods section</b> of the article. All the MIC values are evaluated according to standard Etest, except when noted otherwise. . . . .	148
D.5	$R^2_{\text{training}}$ , $R^2_{\text{validation}}$ and $R^2_{\text{test}}$ values obtained from the 10 fold validation for outer leaflet composition. The neural network is trained with one hidden layer consisting of 8 neurons. . . . .	149
D.6	$R^2_{\text{training}}$ , $R^2_{\text{validation}}$ and $R^2_{\text{test}}$ values obtained from the 10-fold cross validation calculation for inner leaflet composition. The neural network is trained with one hidden layer consisting of 4 neurons. . . . .	151
D.7	P-values of the variables used in the analysis, obtained from our neural network analysis performed using total membrane and outer leaflet composition with 8 neurons, and inner leaflet composition with 4 neurons. We report for two independent variables, since the percentage of the third lipid (CL) is 100-PG-LPG. The p-values for the training set are good. . .	153
E.2	Probability of various cases considering the three factors (Order, charge and hydrophobicity) affecting the secretion process and the p-value for each case with individual p-value for each variable . . . . .	157
E.1	All possible cases considered for analysis. If at least one DE motif is present in disordered region (or ordered region) then disordered (or ordered) is important. There are cases where the DE motif is in the border between ordered and disordered region. We have looked those as both ordered (O) and disordered (D) cases. . . . .	160
E.3	Binary logistic regression – odds ratio of the groups analyzed . . . . .	161
E.4	<b>P propensity of the amino acid insertion at DXE sites among secretory and non-secretory proteins:</b> Among the secretory and non-secretory proteins with a DXE motif, we analyzed to see how often X is one of the three O-phosphorylated amino acids (S, T, Y) or one of the three N-phosphorylated amino acids (H, R, K) . . . . .	162

E.5	<b>Propensity of the amino acid insertion at DEX/EDX sites among secretory and non-secretory proteins:</b> Among the secretory and non-secretory proteins with a DEX/EDX motif, we analyzed to see how often X is one of the three O-phosphorylated amino acids (S, T, Y) or one of the three N-phosphorylated amino acids (H, R, K) . . . . .	163
E.6	Secondary structure of DE motif in secretory and non-secretory proteins. The DE motif is ordered (O) if they are in $\alpha$ -helix or $\beta$ -sheets and disordered (D) if they are in loop or missing residues. Separate cases has been considered if they are in between ordered and disordered . . . . .	164

# Chapter 1

## Introduction

### 1.1 Life across the membrane

The cell membrane is one of the most distinguishing characteristics of life as we find it on earth. It is made up of lipid molecules, and separates the cell's internal from its external environment. The amphipathic nature of the lipids with a hydrophilic head group and a hydrophobic aliphatic chain, facilitates their self-aggregation as membranes and forms a containment region for the cell. The cells derive their nutrients as well as the signals for their growth or death, from the outside environment, and across this protective membrane. The selective transport of signals, molecules or ions is facilitated by the transmembrane proteins that are embedded in the membranes and further exchanges with the environment occur through vesicular formation. So many fascinating aspects of cellular life happen across the membranes and continue to offer challenging problems for investigation.

When one looks at the cellular activities, they can all be traced back to the fundamental and secret code of life, the DNA. DNA is transcribed to RNA, which is then translated to polypeptides or proteins. These proteins are then responsible for the various activities of the cell. Unlike proteins and DNA, where the functional aspects may be interpreted from the primary sequence or the structure, membranes large do not lend themselves to such easy functional interpretations. The membranes being not so specific in their molecular recognitions and being large in size, adds to the complexity of studying them. Not surprisingly, a significant emphasis in biology, biological physics or computational biology is given to the study of DNA, RNA and proteins. However, a lot of fascinating biological phenomena happen across the membrane continues to be unearthed. In response to the global threat of antimicrobial resistance, a new strategy to develop antibiotics which target bacterial membranes rather than the conventional strategy of targeting their proteins is emerging. Rational methods to support this strategy are urgently required.

Similarly, newer and unconventional aspects of protein secretion across the membrane under various conditions including extreme cellular stress are being identified, and they need to be understood. These two unconventional aspects form the topics of investigation in this thesis.

## 1.2 Antibiotic Resistance

The discovery of the first antibiotic penicillin marked the advent of a new era of healthcare. Ever since, antibiotics have been extensively used with a high success for controlling bacterial infections not just in humans but also in animal husbandry as growth stimulants and for prophylactic protection against plant pathogens. Due to these reasons, there is a spread of millions of tons of antibiotics in the biosphere, which has affected the environment's bacterial world.[1–3] In the past decade several multi-drug resistance strains or superbugs, which can not be treated with any known antibiotic have emerged. Less than a century after the discovery of penicillin, we are faced with a large reliance on antibiotics and a simultaneous witnessing of the development of resistance to them. While there are evidences for the antibiotic resistant genes from arctic samples from the pre-antibiotic era, the role of the antibiotics in triggering adaptive defence mechanisms in bacteria is undeniable. For example, the extensive use of vancomycin led to the widespread adoption of vancomycin resistant enterococci into the normal food chain.[4, 5] The heavy use of antibiotics also gave rise to multi-drug resistant (bacteria) or superbugs.[6] As the bacteria have been developing to most antibiotics starting from the first antibiotic Penicillin to the recent ones, it poses a major threat to the treatment of the diseases in the future.[7] Without a viable solution, there is a risk that the quality of healthcare and life may slip to the pre-antibiotic days.

Antibiotics are very selective in their action which gives them both an advantage and disadvantage against bacteria. The advantage being that they mostly target specific pathogen cells and not the host cells. They can understand both physiological and biochemical differences between host cells and pathogens and selectively kill or inhibit the target bacterial growth while causing no or minimal harm to the host. The disadvantage of this specificity is that very often even a single point mutation in the bacterial target region may render an antibiotic ineffective. While the clinical resistance for some of the antibiotics developed within a few years of their introduction, it is easy to track an even rapid adaptation of bacteria within a few days of serial-passage experiments. The intrinsic bacterial adaptation and evolution when challenged by drugs is thus something that may be witnessed almost in realtime, as in the recent Mega-Plate experiments.

Antibiotics are majorly classified into four categories based on the mechanism of their action: antibiotics targeting cell wall, antibiotics inhibiting protein synthesis, antibiotics

inhibiting DNA replications, and those inhibiting folic acid metabolism.[8] The bacteria evolved a strategy to resist each of these classes of antibiotics - (a) Efflux pumps on bacterial membranes export the antibiotics and prevent them from reaching their target. These are present in the cytoplasmic membrane and are specific to antibiotics. Similarly, porins present in the outer membrane are responsible for the modulation of membrane permeability. (b) Since the antibiotics are target specific, alterations in the bacterial target site can make an antibiotic ineffective. These changes are generally caused by modification or mutation in the bacterial chromosomes or genes. (c) There are also other mechanisms where the bacteria release enzymes into the environment, modifying or damaging the antibiotics. Given all these bacterial adaptation strategies, one needs to develop newer drug design strategies, motivated from the natural ecology of bacteria.

### **1.3 Antimicrobial Peptides and their action mechanism**

In order to tackle antibiotic resistance, the attention is now shifting towards a class of drugs which act by disrupting bacterial membranes, rather than block their proteins. Antimicrobial peptides (AMPs) are one such class of drugs which act on bacterial membranes. They were first characterized in the 1980s, and subsequently found in every multicellular organism as a key element of its innate immune system.[9] AMPs are generally short peptides with usually 15 to 50 amino acids long, and are likely to be amphipathic  $\alpha$ -helices.[10] The net positive charge of AMPs, typically between +1 to +9, helps them interact with the membrane and their amphiphilic nature which makes them soluble in aqueous and lipid environments. Further, a clear topographic separation of charges, one face of the helical peptide being mostly hydrophilic and the other face being mostly hydrophobic, is a characteristic of the AMPs.

In general, the membranes surrounding the bacteria are negatively charged, and the membranes of multicellular organisms are zwitterionic in nature. The positively charged AMPs, gets electrostatically attracted to the negatively charged membranes of microbes. Typical AMPs are disordered in an aqueous solution, but they form the secondary structure upon contact with the membrane, either on it, when embedded into it. [11, 12] AMPs act by disrupting the bacterial membranes. Several mechanisms of action of AMPs against bacterial membranes have been proposed. These mechanisms can be roughly classified into two categories – membrane curvature modulation and phase separation. Membrane curvature modulation includes barrel stave model,[13–15] toroidal pore formation,[16–19] and carpet mechanisms.[18, 20–22] The amphipathic structure of these AMPs can cause the modulation in the curvature of the membrane. The interfacial

region of the AMPs expand to form a void which creates a positive curvature strain and thinning of the membrane.

AMPs have been studied extensively, and many clinical trials have been conducted. Some of them are used as approved drugs, such as daptomycin and colistin.[23] However, despite the inspiration from naturally occurring AMPs, and their effectiveness against bacteria, and the reduced possibility of the development of resistance to them, there are several limitations that do not allow AMPs to be viable drugs. Analyzing the hydrophobic moment of the AMPs, which quantifies the degree of amphipathicity, it is clear that peptides with larger hydrophobic moments have higher bacterial membrane disruption potential and hemolytic activity against mammalian cells.[24, 25] Further, poor pharmacokinetics, susceptibility to proteolysis and high production costs limit the adoption of AMPs. Thus, the search for AMP inspired membrane acting drugs continues.

### 1.3.1 Rational strategies to improve or replace AMPs

As there is an urgent need to develop methods and find suitable alternatives to conventional antibiotics as well as AMPs, the molecular details of the mechanism of action AMPs and predictors of their effectiveness need to be understood. In this very complicated and time-consuming search for the mechanisms of action and newer drugs, computational tools will be helpful to guide the rational drug design. With the advancement in computer technology, many methods to better understand the peptide/protein interactions with the membrane.[26] have evolved. Using these tools and techniques, it becomes easier to simulate, screen or process a large number of peptides to gain insights for improving the drug candidates. For example, molecular dynamics simulation can help us understand the atomistic details of the structure and function of a peptide and machine learning can help us understand how the sequence or structure properties of peptides correlate with their biological activity.

#### Molecular dynamics simulation of AMPs

As Richard Feynman once stated, "If we were to name the most powerful assumption of all, which leads one on and on in an attempt to understand life, it is that all things are made of atoms, and that everything that living things do can be understood in terms of the jiggings and wiggings of atoms." Molecular dynamics (MD) simulation has emerged as an extremely powerful method to understand these jiggings and wiggings of biomolecules effectively[27, 28]. Molecular dynamics which is based on numerically integrating classical Newton's laws of motion for each of the atoms in the system, while satisfying the relevant thermodynamic properties. All possible forces of bonded and non-bonded interactions between every pair of atoms are calculated during the simulation, combining the information on the various parameters such as partial charges, mass, bond

length, bond angle, and a knowledge of the force fields which govern the specific types of interactions. MD tracks the detailed movement of every atom of a molecule at a detail that can be illuminating but difficult to probe in experiments.[29] MD simulation is thus like a very powerful microscope that can help in visualizing atom-by-atom interactions between various biomolecules.

With the recent advances in computer technology, the complexity of the systems that can be studied continues to increase. Some of these simulations have been extensively used in the study of the interaction between AMPs and membranes. Various mechanisms are also explored using the MD simulation to understand peptide and lipid interactions.[30] Such simulations show how the AMPs self assemble in cell membrane, as well as binding, folding, insertion and translocation with atomistic details of interaction between membrane and peptides. Many MD simulations have validated the physical experimentation of AMPs acting on membranes. These simulation are also becoming common in the drug discovery, and easier to implement.[31] However, simultaneously the need to study larger biological systems with more intricate details also grows. Or alternative ways to include several unknowns in biological systems implicitly are required.

### **Artificial intelligence models for AMPs**

Machine learning and artificial neural networks help the computer to learn from the existing data and then utilize the learning to build predictive models or to identify hidden patterns in the data. The fundamentals and theory of machine learning such as the Bayes' theorem, principal component analysis, multiple linear regression, least-square fitting, etc., which are extensively used in today's machine learning algorithms have been there far before the invention of the modern computers.[32] In earlier times, the quantitative structure activity relationship (QSAR) models were efficiently able to screen and optimize a small number of datasets. Design and optimization of the peptides was traditionally performed using a manual interpretation of the sequences and a screening of their substitution libraries.[33, 34] But these methods were very labour intensive and now the focus has shifted towards data-driven processes. With the development of machine learning and computer technology, several new methods and models to understand the properties and activity of antimicrobial peptides against bacteria have been developed. Improved QSAR models are being widely used in pharmaceuticals and drug discovery.[35]

In machine learning, the data preparation begins by identifying suitable variables, knowing as features of descriptors, that can describe the various important characteristics of the molecule. The simplest descriptors can be a charge, hydrophobicity, molecular weight, etc. and more complicated can be some structural parameters for which molecular dynamics simulation may be used. These descriptors as independent variables, and the

observed outcome are used as the data set to train the artificial intelligence algorithm. With the advancement in the machine learning technology, better tools were developed to calculate effective descriptors of AMPs combining NMR and *in silico* methods to improve the predictions.[36–38] Following the training, the model validated with data not used for the training, sometimes using methods known as the k-fold cross-validation, where the entire dataset divided into k-parts is used iteratively by training on k-1 parts, and testing the 1 part that was not used in training. Over the years, newer techniques such as the Artificial Neural Network, Decision Tree, XGBoost, Random Forest, etc are becoming available, continuing to offer newer options to explore for specific data sets. Apart from these models, some other complex machine learning tools like Deep-Learning and artificial intelligence are used that can process very large datasets, and with suitable data sets and the right biophysical questions these methods will continue to offer improved predictive capabilities and insights.

### 1.3.2 Challenges and problems in the AMPs

Various molecular dynamics studies were performed to understand the effect of drug candidates on membranes. The studies have shown either the self-assembly or insertion of the molecules inside the membrane. The combined effect of self-assembly and aggregation has not been studied so far. Further, the delicate balance between the charges and atomistic details can not be captured in coarse-grained models that are commonly used in the membrane simulations. Can an atomistic mechanism be identified which can show the behaviour of simultaneous insertion and assembly of drug candidates that will help in the disruption of membrane? Will such a study reveal the differences in specificity and sensitivity across different target membranes? Can a rational guidance for drug development be provided which can help in combating infection from the most dangerous multi-drug resistant bacteria? Despite better computational tools and machine learning, without the proper descriptors the predictions will remain poor. Many antimicrobial activity predictions use properties derived from sequences alone. If structural descriptors using MD simulation are used, will the prediction of the AMPs be improved? The traditional drug designing approach involves screening multiple drugs against a single type of bacteria or target. The methodology is fairly standard. But it is not clearly understood how a particular drug will be effective once the bacteria adapt by lysylating their phospholipids. If the bacteria change their membrane composition, will the same AMP continue to be effective against them?



## 1.4 Unconventional Protein Secretion

Another interesting phenomenon that occurs across the membrane is called membrane trafficking, where cargo like proteins and macromolecules occurs in the cell or transported to the outside of the cell. It is a very fundamental and important process in all types of cells. Proteins are secreted out to the extracellular space for various activities like growth, cell homeostasis, cytokinesis, defence, structural maintenance, hormone release, and neurotransmission. In the liver, 70% of the total proteins produced are secreted out to blood. Similarly, pancreatic cells produce several enzymes that are secreted out of the cell to intestinal regions for various digestive processes.[39] Most of the cargo proteins are secreted to the target sites through well documented ‘conventional’ pathways that consist of the Endoplasmic Reticulum (ER) and Golgi complex. The cargo protein made in the lumen of ER travels to the cis-Golgi network of the Golgi complex via COPII vesicles. The proteins then travel from cis-Golgi to trans-Golgi network via medial-Golgi with the process called cisternal progression. Finally, the proteins are sorted into secretory vesicles, which gets integrated into the plasma membrane releasing the proteins to the extracellular environment.[40]

These conventional secretion mechanisms have received enough attention and there are several ways to identify the part of the protein that serves as a signal that marks it for this molecular transport. SP serves the purpose of a ‘zipcode’ that marks the pathway and target location.[41, 42] The signal peptide (SP) is around 16-30 amino acids long [43] consists of three main regions: amino-terminal positively charged region or n-region (1 to 5 amino acid long); a central hydrophobic region or h-region (7 to 15 amino acid long); and a polar carboxy-terminal region or c-region (3 to 7 amino acid long). Beyond this region, there is a stretch of an amino acid called the cleavage site, which is cleaved by a type of enzyme called signal peptidase.

Apart from the conventional pathway mentioned above, some proteins are also secreted, in hitherto unknown pathways and without a clearly known signal. This phenomenon which occurs in several situations including the secretory autophagy conditions triggered during starvation, is known as the unconventionally protein secretion (UPS or UPS) of ‘leaderless’ (without a well-defined signal) proteins. There are four types of UPS - type I, II, III, and IV. The leaderless cytoplasmic proteins, which are directly transported out of the cell either through plasma membrane pores [44] or through ABC transporters,[45] come under type I and type II UPS, respectively. The type III UPS includes the leaderless cytoplasmic proteins that are transported through membrane-bound organelles such as autophagosomes, lysosomes, endosomes and exosomes. [46–49] Whereas type IV proteins contain either signal peptide or transmembrane domain but bypass the Golgi when they

are delivered to the plasma membrane.[50]

The first time, protein interleukin (IL)-1 revealed its secretion without any signal peptide in response to the cell injury and inflammation.[51] Later again it was reported about the secretion of IL-1 $\beta$  from activated human monocytes.[52] Since then, many other proteins which lack signal peptides were discovered in the extracellular space of eukaryotic cells. But understanding their pathways and secretion has remained a mystery. Some attempts were taken to understand the unconventional secretion of IL-1 $\beta$ . For example, once it was proposed that the shedding of micro-vesicles in response to activation of P2X<sub>7</sub> receptor is a major pathway for rapid IL-1 $\beta$  release.[53] Another work suggested that IL-1 $\beta$  release is mediated by multi-vesicular bodies as a vehicle.[54] However, no such studies could identify the single pathway of IL-1 $\beta$  secretion. Some of the computational models were developed to understand signals for the unconventional protein secretion but had limited success.[55] Even artificial neural networks could not explicitly reveal patterns in unconventional secretory proteins.[56, 57]

UPS is mostly induced by cellular stresses e.g., nutrient starvation,[58] mechanical stress,[59] inflammation[60] and ER stress.[61, 62] During many diseases like Alzheimer's, diabetes, allergies, and cancer, UPS can be triggered. Different trafficking pathways of UPS cargoes related to various human diseases have been discussed.[63] It was found that autophagy also mediates the secretion of proteins of both conventional and unconventional secretion. Autophagy is in general a self degradative process when there is a need for energy during nutrient stress. It also helps in maintaining the cellular homeostasis by removing misfolded proteins and other damaged organelles. Apart from being involved in self-digestion, autophagy-related (ATG) genes play an important role in UPS.[64] This type of unconventional secretion is called secretory autophagy. In the core of autophagy, there is a LC3 domain in the phagophore region where the LC3 interacting region (LIR) of a protein targets. These LIR regions are four amino acid long sequence,  $\Theta$ XX $\Gamma$ , where X is any amino acid,  $\Theta$  is one of W/F/Y and  $\Gamma$  is one of I/L/V. There are also certain special LIR motifs called xLIR which is derived from experimentally verified functional LIRs. These have the consensus sequences - [ADEFGLPRSK][DEGMSTV][**WFY**][DEILQTV][ADEFHIKLMPTV][**ILV**].

### 1.4.1 Computational methods

Several tools based on machine learning are available to identify the N-terminal signal peptide meant for the conventional secretion.[65, 66] SignalP is the most widely used online tool for predicting the conventional secretion by identifying the N-terminal signal peptides. Apart from SignalP, there are other tools that have advantages under certain circumstances.[67] The predictive accuracy for the signal peptide is very high. But UPS

lacks any clear signal. The data on UPS of proteins has been very limited, not leaving ample scope for the development of the predictive models. There are some prediction tools available which generally use the conventional secretory proteins by removing the signal peptides, with a hypothesis that the conventional and unconventional secretion share common properties and features which can be related to protein function outside the cell and not to the secretory pathways.[56, 68]. Some of the recent tools such as OutCyte,[69] and ExoPred,[70] developed models using recent machine learning methods, XGBoost and Random Forest respectively. The former was trained on in-house experimental secretome data, and the latter on the exosome data. The predictions of these models for new data sets has not been satisfactory.

### 1.4.2 Challenges and recent development

As mentioned, using machine learning approaches many advances in the prediction of conventional protein secretion were achieved. But as the data sets are limited, detailed analysis and predictions of the chances of unconventional secretion of proteins has not been unravelled. Recently, one article revealed that the diacidic motif “DE” plays an important role in the secretion of superoxide dismutase 1 (SOD1) and Acb1 proteins by comparing the analogs from human, mouse and yeast.[71] This first and concrete observation raises a hope that DE motif could be a signal for the unconventional protein secretion. But given the potential ubiquity of a two amino acid motif, the question on what additional factors, make that set the context of DE motif specific for UPS needs to be explored. Is there any other context like charge, hydrophobicity, structural order, etc., which adds to the secretory nature? Or are there some other types of acidic motifs which can contribute to the secretion, along with some other factor like LC3 interacting region (LIR) motifs, as they are said to be in connection with the protein secretion? All these need an investigation in order to identify the potential ‘signal’ or the context that increases the odds of the unconventional secretion.

Given the importance and the complexity of the various phenomena that occur across the membrane, we explore two aspects noted above - the possibility of a rational design of membrane active antibiotics, and the factors that define the possibility of unconventional secretion.

## Bibliography

- [1] D. I. Massé, N. M. C. Saady, and Y. Gilbert, “Potential of biological processes to eliminate antibiotics in livestock manure: an overview,” *Animals*, vol. 4, no. 2, pp. 146–163, 2014.

- [2] A. K. Sarmah, M. T. Meyer, and A. B. Boxall, "A global perspective on the use, sales, exposure pathways, occurrence, fate and effects of veterinary antibiotics (vas) in the environment," *Chemosphere*, vol. 65, no. 5, pp. 725–759, 2006.
- [3] K. Kumar, S. C. Gupta, Y. Chander, and A. K. Singh, "Antibiotic use in agriculture and its impact on the terrestrial environment," *Advances in agronomy*, vol. 87, pp. 1–54, 2005.
- [4] B. E. Murray, "Vancomycin-resistant enterococci," *The American journal of medicine*, vol. 102, no. 3, pp. 284–293, 1997.
- [5] Y. Cetinkaya, P. Falk, and C. G. Mayhall, "Vancomycin-resistant enterococci," *Clinical microbiology reviews*, vol. 13, no. 4, pp. 686–707, 2000.
- [6] J. Davies and D. Davies, "Origins and evolution of antibiotic resistance," *Microbiology and molecular biology reviews*, vol. 74, no. 3, pp. 417–433, 2010.
- [7] S. Sengupta, M. K. Chattopadhyay, and H.-P. Grossart, "The multifaceted roles of antibiotics and antibiotic resistance in nature," *Frontiers in microbiology*, vol. 4, p. 47, 2013.
- [8] G. Kapoor, S. Saigal, and A. Elongavan, "Action and resistance mechanisms of antibiotics: A guide for clinicians," *Journal of anaesthesiology, clinical pharmacology*, vol. 33, no. 3, p. 300, 2017.
- [9] T.-H. Lee, K. N. Hall, and M.-I. Aguilar, "Antimicrobial peptide structure and mechanism of action: a focus on the role of membrane structure," *Current topics in medicinal chemistry*, vol. 16, no. 1, pp. 25–39, 2016.
- [10] H.-C. Chen, J. H. Brown, J. L. Morell, and C. Huang, "Synthetic magainin analogues with improved antimicrobial activity," *Febs Letters*, vol. 236, no. 2, pp. 462–466, 1988.
- [11] K. Matsuzaki, M. Harada, T. Handa, S. Funakoshi, N. Fujii, H. Yajima, and K. Miyajima, "Magainin 1-induced leakage of entrapped calcein out of negatively-charged lipid vesicles," *Biochimica et Biophysica Acta (BBA)-Biomembranes*, vol. 981, no. 1, pp. 130–134, 1989.
- [12] K. Matsuzaki, M. Harada, S. Funakoshi, N. Fujii, and K. Miyajima, "Physicochemical determinants for the interactions of magainins 1 and 2 with acidic lipid bilayers," *Biochimica et Biophysica Acta (BBA)-Biomembranes*, vol. 1063, no. 1, pp. 162–170, 1991.

- [13] G. Baumann and P. Mueller, "A molecular model of membrane excitability," *Journal of supramolecular structure*, vol. 2, no. 5-6, pp. 538–557, 1974.
- [14] E. Breukink and B. de Kruijff, "The lantibiotic nisin, a special case or not?," *Biochimica et Biophysica Acta (BBA)-Biomembranes*, vol. 1462, no. 1-2, pp. 223–234, 1999.
- [15] G. Ehrenstein and H. Lecar, "Electrically gated ionic channels in lipid bilayers," *Quarterly reviews of biophysics*, vol. 10, no. 1, pp. 1–34, 1977.
- [16] T. Hara, H. Kodama, M. Kondo, K. Wakamatsu, A. Takeda, T. Tachi, and K. Matsuzaki, "Effects of peptide dimerization on pore formation: Antiparallel disulfide-dimerized magainin 2 analogue," *Biopolymers: Original Research on Biomolecules*, vol. 58, no. 4, pp. 437–446, 2001.
- [17] T. Hara, Y. Mitani, K. Tanaka, N. Uematsu, A. Takakura, T. Tachi, H. Kodama, M. Kondo, H. Mori, A. Otaka, *et al.*, "Heterodimer formation between the antimicrobial peptides magainin 2 and pglA in lipid bilayers: a cross-linking study," *Biochemistry*, vol. 40, no. 41, pp. 12395–12399, 2001.
- [18] K. Matsuzaki, "Magainins as paradigm for the mode of action of pore forming polypeptides," *Biochimica et Biophysica Acta (BBA)-Reviews on Biomembranes*, vol. 1376, no. 3, pp. 391–400, 1998.
- [19] M. R. Yeaman and N. Y. Yount, "Mechanisms of antimicrobial peptide action and resistance," *Pharmacological reviews*, vol. 55, no. 1, pp. 27–55, 2003.
- [20] G. Taubes, "The bacteria fight back," *Science*, vol. 321, no. 5887, pp. 356–361, 2008.
- [21] Y. Shai, "Mechanism of the binding, insertion and destabilization of phospholipid bilayer membranes by  $\alpha$ -helical antimicrobial and cell non-selective membrane-lytic peptides," *Biochimica et Biophysica Acta (BBA)-Biomembranes*, vol. 1462, no. 1-2, pp. 55–70, 1999.
- [22] Y. Shai and Z. Oren, "From "carpet" mechanism to de-novo designed diastereomeric cell-selective antimicrobial peptides," *Peptides*, vol. 22, no. 10, pp. 1629–1641, 2001.
- [23] H. B. Koo and J. Seo, "Antimicrobial peptides under clinical investigation," *Peptide Science*, vol. 111, no. 5, p. e24122, 2019.
- [24] M. Zasloff, "Antimicrobial peptides of multicellular organisms," *nature*, vol. 415, no. 6870, pp. 389–395, 2002.

- [25] N. Pathak, R. Salas-Auvert, G. Ruche, M.-h. Janna, D. McCarthy, and R. G. Harrison, "Comparison of the effects of hydrophobicity, amphiphilicity, and  $\alpha$ -helicity on the activities of antimicrobial peptides," *Proteins: Structure, Function, and Bioinformatics*, vol. 22, no. 2, pp. 182–186, 1995.
- [26] P. G. Aronica, L. M. Reid, N. Desai, J. Li, S. J. Fox, S. Yadahalli, J. W. Essex, and C. S. Verma, "Computational methods and tools in antimicrobial peptide research," *Journal of Chemical Information and Modeling*, 2021.
- [27] D. Tieleman and M. Sansom, "Molecular dynamics simulations of antimicrobial peptides: from membrane binding to trans-membrane channels," *International Journal of Quantum Chemistry*, vol. 83, no. 3-4, pp. 166–179, 2001.
- [28] S. Arasteh, M. Bagheri, and B. Goliaei, "Membrane selectivity of small cyclic antimicrobial hexapeptides studied by molecular dynamics simulation," in *Journal of Peptide Science*, vol. 20, pp. S279–S280, Wiley-Blackwell 111 River St, Hoboken 07030-5774, NJ USA, 2014.
- [29] J. P. Ulmschneider and M. B. Ulmschneider, "Molecular dynamics simulations are redefining our view of peptides interacting with biological membranes," *Accounts of chemical research*, vol. 51, no. 5, pp. 1106–1116, 2018.
- [30] N. Palmer, J. R. Maasch, M. D. Torres, and C. de la Fuente-Nunez, "Molecular dynamics for antimicrobial peptide discovery," *Infection and Immunity*, vol. 89, no. 4, pp. e00703–20, 2021.
- [31] A. Ganesan, M. L. Coote, and K. Barakat, "Molecular dynamics-driven drug discovery: leaping forward with confidence," *Drug discovery today*, vol. 22, no. 2, pp. 249–269, 2017.
- [32] J. Friedman, T. Hastie, R. Tibshirani, *et al.*, *The elements of statistical learning*, vol. 1. Springer series in statistics New York, 2001.
- [33] K. Hilpert, R. Volkmer-Engert, T. Walter, and R. E. Hancock, "High-throughput generation of small antibacterial peptides with improved activity," *Nature biotechnology*, vol. 23, no. 8, pp. 1008–1012, 2005.
- [34] K. Hilpert, M. R. Elliott, R. Volkmer-Engert, P. Henklein, O. Donini, Q. Zhou, D. F. Winkler, and R. E. Hancock, "Sequence requirements and an optimization strategy for short antimicrobial peptides," *Chemistry & biology*, vol. 13, no. 10, pp. 1101–1107, 2006.

- [35] R. Perkins, H. Fang, W. Tong, and W. J. Welsh, "Quantitative structure-activity relationship methods: Perspectives on drug discovery and toxicology," *Environmental Toxicology and Chemistry: An International Journal*, vol. 22, no. 8, pp. 1666–1679, 2003.
- [36] K. Kavousi, M. Bagheri, S. Behrouzi, S. Vafadar, F. F. Atanaki, B. T. Lotfabadi, S. Ariaeenejad, A. Shockravi, and A. A. Moosavi-Movahedi, "Iampe: Nmr-assisted computational prediction of antimicrobial peptides," *Journal of Chemical Information and Modeling*, vol. 60, no. 10, pp. 4691–4701, 2020.
- [37] F. Fallah Atanaki, S. Behrouzi, S. Ariaeenejad, A. Boroomand, and K. Kavousi, "Bipep: Sequence-based prediction of biofilm inhibitory peptides using a combination of nmr and physicochemical descriptors," *ACS Omega*, vol. 5, no. 13, pp. 7290–7297, 2020.
- [38] F. Nasiri, F. F. Atanaki, S. Behrouzi, K. Kavousi, and M. Bagheri, "Cpacpp: *in silico* cell-penetrating anticancer peptide prediction using a novel bioinformatics framework," *ACS Omega*, vol. 6, no. 30, pp. 19846–19859, 2021.
- [39] H. Lodish, A. Berk, S. L. Zipursky, P. Matsudaira, D. Baltimore, and J. Darnell, "Intracellular ion environment and membrane electric potential," in *Molecular Cell Biology. 4th edition*, WH Freeman, 2000.
- [40] M. C. Lee, E. A. Miller, J. Goldberg, L. Orci, and R. Schekman, "Bi-directional protein transport between the er and golgi," *Annual Review of Cell and Developmental Biology*, vol. 20, no. 1, pp. 87–123, 2004. PMID: 15473836.
- [41] G. von Heijne, "The signal peptide," *The Journal of membrane biology*, vol. 115, no. 3, pp. 195–201, 1990.
- [42] N. Zheng and L. M. Gierasch, "Signal sequences: the same yet different," *Cell*, vol. 86, no. 6, pp. 849–852, 1996.
- [43] K. Kapp, S. Schrempf, M. K. Lemberg, and B. Dobberstein, "Post-targeting functions of signal peptides," *Protein Transport into the Endoplasmic Reticulum*, pp. 1–16, 2009.
- [44] J. P. Steringer, S. Bleicken, H. Andreas, S. Zacherl, M. Laussmann, K. Temmerman, F. X. Contreras, T. A. Bharat, J. Lechner, H.-M. Müller, *et al.*, "Phosphatidylinositol 4, 5-bisphosphate (pi (4, 5) p2)-dependent oligomerization of fibroblast growth factor 2 (fgf2) triggers the formation of a lipidic membrane pore implicated in unconventional secretion," *Journal of Biological Chemistry*, vol. 287, no. 33, pp. 27659–27669, 2012.

- [45] J. P. McGrath and A. Varshavsky, "The yeast *ste6* gene encodes a homologue of the mammalian multidrug resistance p-glycoprotein," *Nature*, vol. 340, no. 6232, pp. 400–404, 1989.
- [46] N. Dupont, S. Jiang, M. Pilli, W. Ornatowski, D. Bhattacharya, and V. Deretic, "Autophagy-based unconventional secretory pathway for extracellular delivery of *il-1 $\beta$* ," *The EMBO journal*, vol. 30, no. 23, pp. 4701–4711, 2011.
- [47] P. Ejlerskov, I. Rasmussen, T. T. Nielsen, A.-L. Bergström, Y. Tohyama, P. H. Jensen, and F. Vilhardt, "Tubulin polymerization-promoting protein (tppp/p25 $\alpha$ ) promotes unconventional secretion of  $\alpha$ -synuclein through exophagy by impairing autophagosome-lysosome fusion," *Journal of Biological Chemistry*, vol. 288, no. 24, pp. 17313–17335, 2013.
- [48] M. T. Lotze and K. J. Tracey, "High-mobility group box 1 protein (hmgb1): nuclear weapon in the immune arsenal," *Nature Reviews Immunology*, vol. 5, no. 4, pp. 331–342, 2005.
- [49] R. Manjithaya and S. Subramani, "Role of autophagy in unconventional protein secretion," *Autophagy*, vol. 6, no. 5, pp. 650–651, 2010.
- [50] A. G. Grieve and C. Rabouille, "Golgi bypass: skirting around the heart of classical secretion," *Cold Spring Harbor perspectives in biology*, vol. 3, no. 4, p. a005298, 2011.
- [51] P. E. Auron, A. C. Webb, L. J. Rosenwasser, S. F. Mucci, A. Rich, S. M. Wolff, and C. A. Dinarello, "Nucleotide sequence of human monocyte interleukin 1 precursor cDNA," *Proceedings of the National Academy of Sciences*, vol. 81, no. 24, pp. 7907–7911, 1984.
- [52] A. Rubartelli, F. Cozzolino, M. Talio, and R. Sitia, "A novel secretory pathway for interleukin-1 beta, a protein lacking a signal sequence.," *The EMBO journal*, vol. 9, no. 5, pp. 1503–1510, 1990.
- [53] A. MacKenzie, H. L. Wilson, E. Kiss-Toth, S. K. Dower, R. A. North, and A. Surprenant, "Rapid secretion of interleukin-1 $\beta$  by microvesicle shedding," *Immunity*, vol. 15, no. 5, pp. 825–835, 2001.
- [54] Y. Qu, L. Franchi, G. Nunez, and G. R. Dubyak, "Nonclassical *il-1 $\beta$*  secretion stimulated by p2x7 receptors is dependent on inflammasome activation and correlated with exosome release in murine macrophages," *The Journal of Immunology*, vol. 179, no. 3, pp. 1913–1925, 2007.



- [55] A. Lonsdale, M. J. Davis, M. S. Doblin, and A. Bacic, “Better than nothing? limitations of the prediction tool secretomep in the search for leaderless secretory proteins (lsps) in plants,” *Frontiers in plant science*, vol. 7, p. 1451, 2016.
- [56] J. D. Bendtsen, L. J. Jensen, N. Blom, G. Von Heijne, and S. Brunak, “Feature-based prediction of non-classical and leaderless protein secretion,” *Protein Engineering Design and Selection*, vol. 17, no. 4, pp. 349–356, 2004.
- [57] J. D. Bendtsen, L. Kiemer, A. Fausbøll, and S. Brunak, “Non-classical protein secretion in bacteria,” *BMC microbiology*, vol. 5, no. 1, pp. 1–13, 2005.
- [58] D. Cruz-Garcia, A. J. Curwin, J.-F. Popoff, C. Bruns, J. M. Duran, and V. Malhotra, “Remodeling of secretory compartments creates cups during nutrient starvation,” *Journal of cell biology*, vol. 207, no. 6, pp. 695–703, 2014.
- [59] H. Schotman, L. Karhinen, and C. Rabouille, “dgrasp-mediated noncanonical integrin secretion is required for drosophila epithelial remodeling,” *Developmental cell*, vol. 14, no. 2, pp. 171–182, 2008.
- [60] K. Schroder and J. Tschopp, “The inflammasomes,” *cell*, vol. 140, no. 6, pp. 821–832, 2010.
- [61] H. Y. Gee, S. H. Noh, B. L. Tang, K. H. Kim, and M. G. Lee, “Rescue of  $\delta f508$ -cftr trafficking via a grasp-dependent unconventional secretion pathway,” *Cell*, vol. 146, no. 5, pp. 746–760, 2011.
- [62] J. Jung, J. Kim, S. H. Roh, I. Jun, R. D. Sampson, H. Y. Gee, J. Y. Choi, and M. G. Lee, “The hsp70 co-chaperone dnajc14 targets misfolded pendrin for unconventional protein secretion,” *Nature communications*, vol. 7, no. 1, pp. 1–15, 2016.
- [63] J. Kim, H. Y. Gee, and M. G. Lee, “Unconventional protein secretion—new insights into the pathogenesis and therapeutic targets of human diseases,” *Journal of cell science*, vol. 131, no. 12, p. jcs213686, 2018.
- [64] M. Ponpuak, M. A. Mandell, T. Kimura, S. Chauhan, C. Cleyrat, and V. Deretic, “Secretory autophagy,” *Current opinion in cell biology*, vol. 35, pp. 106–116, 2015.
- [65] C. Guda, “ptarget: a web server for predicting protein subcellular localization,” *Nucleic acids research*, vol. 34, no. suppl\_2, pp. W210–W213, 2006.
- [66] J. J. A. Armenteros, K. D. Tsirigos, C. K. Sønderby, T. N. Petersen, O. Winther, S. Brunak, G. von Heijne, and H. Nielsen, “Signalp 5.0 improves signal peptide

- predictions using deep neural networks,” *Nature biotechnology*, vol. 37, no. 4, pp. 420–423, 2019.
- [67] X. Hui, Z. Chen, J. Zhang, M. Lu, X. Cai, Y. Deng, Y. Hu, and Y. Wang, “Computational prediction of secreted proteins in gram-negative bacteria,” *Computational and Structural Biotechnology Journal*, 2021.
- [68] K. K. Kandaswamy, G. Pugalenthi, E. Hartmann, K.-U. Kalies, S. Möller, P. Suganthan, and T. Martinetz, “Spred: A machine learning approach for the identification of classical and non-classical secretory proteins in mammalian genomes,” *Biochemical and biophysical research communications*, vol. 391, no. 3, pp. 1306–1311, 2010.
- [69] L. Zhao, G. Poschmann, D. Waldera-Lupa, N. Rafiee, M. Kollmann, and K. Stühler, “Outcyte: a novel tool for predicting unconventional protein secretion,” *Scientific reports*, vol. 9, no. 1, pp. 1–9, 2019.
- [70] A. Ras-Carmona, M. Gomez-Perosanz, and P. A. Reche, “Prediction of unconventional protein secretion by exosomes,” *BMC bioinformatics*, vol. 22, no. 1, pp. 1–13, 2021.
- [71] D. Cruz-Garcia, N. Brouwers, J. M. Duran, G. Mora, A. J. Curwin, and V. Malhotra, “A diacidic motif determines unconventional secretion of wild-type and als-linked mutant sod1,” *Journal of Cell Biology*, vol. 216, no. 9, pp. 2691–2700, 2017.

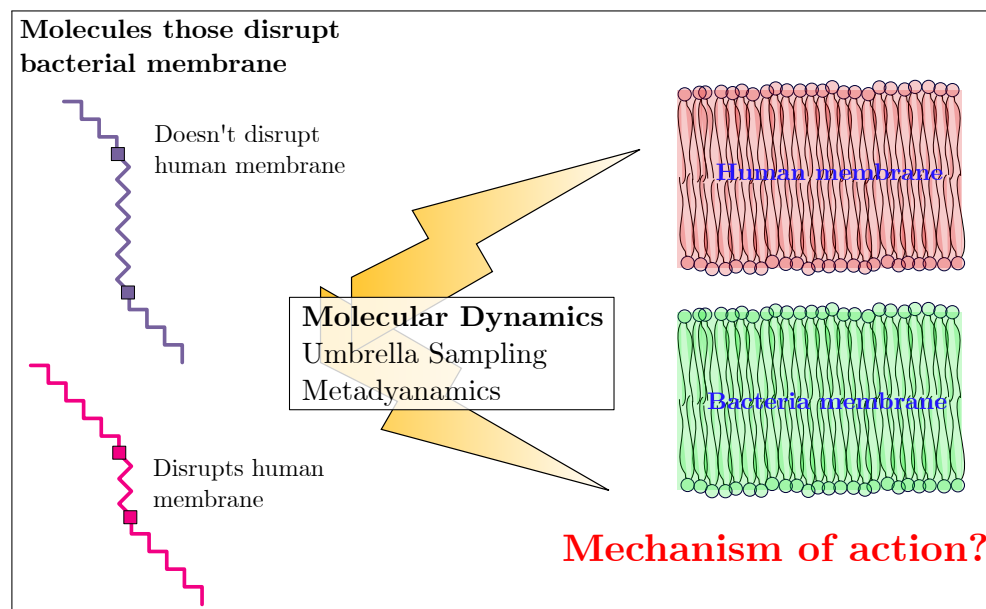
## **PART I**

# **Mechanism of action and design of membrane-active drugs**



## Chapter 2

# Detailing the Molecular Mechanism of Selective Antibiotic Action Against Bacterial Membranes via Self Assembly and Membrane Penetration



## Abstract

Molecules which disrupt bacterial membranes are being seen as a solution to the problem of antibiotic resistance. Recently, a new class of small cationic amphiphilic molecules

have been developed as an alternative to the conventional drugs in avoiding antibiotic resistance and to the antimicrobial peptides in avoiding difficult synthesis routes. Although the experiments suggest that these molecules selectively disrupt the bacterial membrane, the mechanism of action of these molecules is not known. In this work, we use advanced molecular dynamics simulations to investigate the interaction between two such Membrane Disruptive Amphiphilic (MeDiAm) molecules with prototypical bacterial and red blood cell membranes. The cationic groups of these drug candidates interact with the head group of the lipid molecules, thinning the membrane, suggesting the possibility of disruption. Interestingly the disruptive action is by self-assembly of the small molecules. The all atom molecular dynamics simulations also highlight the subtle differences in the action of two different types of drugs candidate on two different membrane models. The differences in charge separations and the length of the alkyl chains lead to a differential selectivity of the drugs between the bacterial and the human membrane, as seen in the differential thinning of the membranes in our calculations.

## 2.1 Introduction

The development of resistance by bacteria to a broad spectrum of antibiotics is currently one of the major challenges in medicine[1]. This points towards an urgent need to develop novel antimicrobial drugs having low propensity for the bacteria to develop resistance against. The reason for the antibiotic resistance is that conventional antibiotics target specific sites of bacterial proteins and even single point mutations in these sites are sufficient to make the drugs ineffective. In this respect, drugs that affect the integrity of bacterial membranes with high selectivity were considered as a solution. Naturally occurring antimicrobial peptides (AMPs)[2, 3] are one such class of molecules which have bacteriocidal action by membrane disruption. But, high costs of production, poor pharmacokinetics and susceptibility to proteolysis, have offset the large scale pharmaceutical application of these peptides.[4]

Given the drawbacks of AMPs, techniques to develop non-natural peptidomimetics that mimic the membrane disruptive action of natural AMPs [5–13] with facile synthesis pathway[14, 15] have been developed. They are aimed to improve the plasma stability and selectivity, thereby, surpassing some of the limitations associated with AMPs.[6, 13] However, designing peptidomimetics involves complex molecular frameworks required to tune the structure-activity relationship, limiting their large scale synthesis and use in medicine.[13] Recently, certain Membrane Disrupting Amphiphilic (MeDiAm) molecules have been developed, that are devoid of any amino acid (consist of non peptidic amide bonds), having a positive charge of two units on the quaternary ammonium groups that are separated by a hydrophobic alkyl chain (**Figure 2.1a**).[16] The basic idea

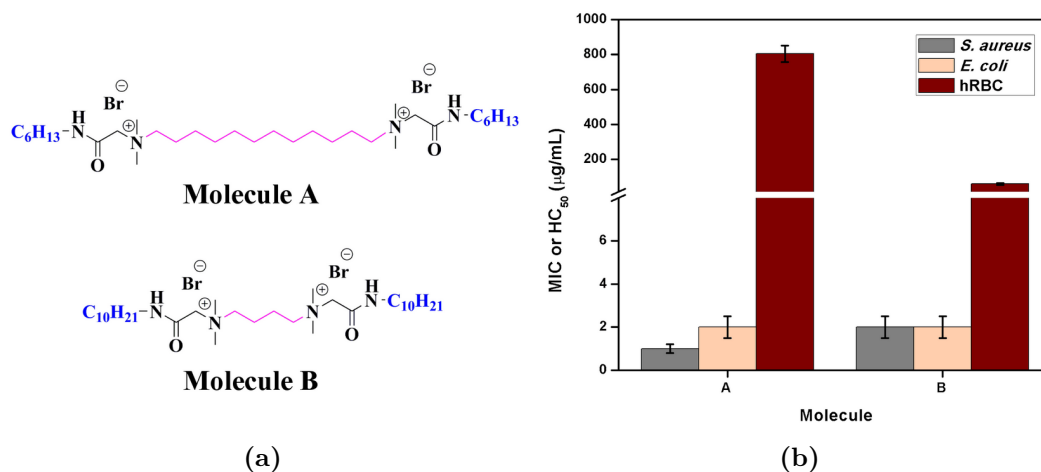
behind the synthesis of these molecules of varying chain length is to systematically assess the amphiphilic balance and to fine-tune the structure activity relationship. The first goal is achieved by varying the lipophilic alkyl chain length pending from the quaternary ammonium groups whereas for the second goal, the lipophilic methylene spacer is varied between the amino groups of the diaminoalkane framework. Several selective antibacterial small molecule peptidomimetics are found to be selectively active against bacterial membranes and non toxic against red blood cells (RBCs)[17, 18]. The lower chance of bacteria developing to these drug candidates and their efficacy against existing drug resistance strains was demonstrated. Their antibacterial efficacy has been evaluated against a wide range of drug sensitive bacteria such as *S. aureus*, *E. coli*, and *P. aeruginosa* and drug-resistant bacteria such as methicillin-resistant *S. aureus* (MRSA), vancomycin-resistant *E. faecium* (VRE),  $\beta$ -lactam-resistant *K. pneumoniae*, norfloxacin-resistant *S. aureus* (NRSA), and colistin-resistant *E. coli* (CREC).[16]

The mechanism of action of these small molecules as suggested by propidium iodide permeation and fluorescence spectroscopy is through the disruption of the bacterial membrane. Since the molecules are of a newer class, lack in ordered structure and amphiphilicity compared to AMPs, and are smaller in size compared to other antimicrobial polymers, [19] the mechanism of action needs to be investigated. Understanding the dynamics of the interaction between the bacterial membrane and the drug molecule would serve as an important tool to rationally design novel drug molecules possessing the essential features that make them antimicrobial in behaviour with a low rate of haemolysis. In this respect a molecular level understanding holds a key role in developing and intuition about the atomic level mechanism.[20, 21]

Molecular dynamics simulation have been employed for understanding the mechanism of interaction of AMPs with membranes. However, these studies are restricted to the interaction of one-type of membrane (bacterial) with one class of drug molecules, and thus do not probe the sensitivity and specificity of potential drug activity. Further the scope is usually limited to exploring the peptides starting in water and depositing on the membrane or of the peptides embedded in membrane causing a membrane distortion. In this work, we explore the pathway of the action of two different drug candidates and their specificity towards model bacterial membranes over the RBCs. Atomistic molecular dynamics combined with advanced sampling methods was used for this purpose.

## 2.2 Methods

Two small amphiphilic MeDiAm molecules, which showed the best and least selectivity were considered in our study (**Figure 2.1a**). These two MeDiAms were selected by considering their minimum inhibition concentration (MIC) and the concentration need



**Figure 2.1:** (a) Chemical structure of the MeDiAms ‘A’ and ‘B’. The distance between the methylated ammonium group in molecule ‘A’ is larger than molecule ‘B’. (b) MIC and HC<sub>50</sub> of the two MeDiAms. Both have comparable MIC ( $\sim 1.5\mu\text{g}/\text{ml}$  and  $2\mu\text{g}/\text{ml}$  respectively), but very different HC<sub>50</sub>s ( $805\mu\text{g}/\text{ml}$  and  $60\mu\text{g}/\text{ml}$  respectively). The structures of MeDiAms were optimized at B3LYP/6-31G\* [22–24] level of theory.

to destroy 50% of human RBCs (HC<sub>50</sub>). Although their molecular formula is same, they show very different HC<sub>50</sub>s ( $740\mu\text{g}/\text{ml}$  and  $55\mu\text{g}/\text{ml}$ ). The sp<sup>3</sup> hybridised nitrogen atoms possess a formal positive charge causing the whole molecule to be positively charged. These nitrogen atoms are separated by an alkyl chain which varies in the length, and hence determines the charge separation in the molecule. The three dimensional geometries were obtained by performing a genetic algorithm based search using the OpenBabel 2.3.2 tool.[25] Initially, ten structures were obtained and the energy was estimated using the universal force field (UFF).

### Generation of the membranes

The models for both bacterial membrane and human membrane were constructed using an online tool, CHARMM-GUI[26], with two types of lipid molecules DPPE and DPPG for bacterial membrane and DPPC for human membrane. In bacterial membrane, a total of 116 lipid molecules were considered. The ratio of DPPE and DPPG in each leaflet of the lipid bilayer was 7:51. The net charge of each DPPG molecule was -1. Where as in human membrane, total 124 DPPC lipid molecules were considered with 62 in each leaflet. DPPC was overall neutral.

### Molecular Dynamics

As a first simulation, one molecule (molecule ‘A’) was inserted into the model bacterial and RBC membranes, and molecular dynamics simulation were performed on both these systems. CHARMM-36[27] and TIP3P[28] force field were used for the lipid molecules



and water molecules respectively. A solvation box, containing around 120 lipid molecules was chosen and the solvated system had nearly 35,000 atoms. For the molecular dynamics simulation we used GROMACS 5.0.5[29]. For these molecules we used a force field that was developed in our group by performing B3LYP/6-31G\* level calculations for all interactions with high penalty. Over 400 ps, the position restraints on the lipid molecules were relaxed slowly in 6 steps. NVT ensemble were used for first two steps and NPT ensemble for the next 4 steps were used at temperature 303 K and pressure 1 atm. SHAKE[30] algorithm was used to hold the hydrogen atom covalent bond. For total Coulombic interaction, PME[31] was employed. van der Waals interaction cut-off was 12 Å with a smooth switching function at 10 Å. Structural coordinates of the system were saved every 10 ps. The procedure was again repeated for molecule 'B'.

### **Umbrella Sampling**

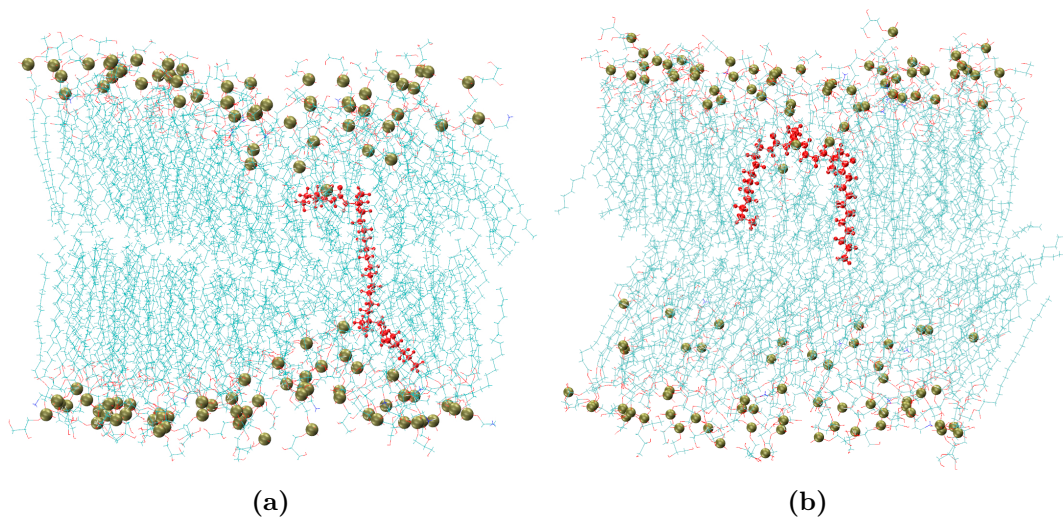
For the advance sampling analysis Umbrella sampling[32] method was used. A total of six molecules were inserted vertically into the membrane. The collective variable employed in this case was the distance of center of mass between two groups of molecules ( $\xi$ ). During the umbrella sampling method the distance between the centers of mass was reduced from 10 Å to 3 Å in 1 Å steps. A force constant of 956 kcal/mol/nm<sup>2</sup> was used on the ( $\xi$ ), and in each umbrella 10 ns of simulation was performed. The free energy profile obtained after the sampling were analysed using WHAM[33].

### **Metadynamics**

In order to understand the aggregation of these molecules and their stability inside the membrane, two dimensional metadynamics[34] simulation was performed on a system of six molecules 'B' in the bacterial membrane. A larger system was used for these calculations and the total number of atoms were around 100,000. The collective variables were radius of gyration ( $R_{gyr}$ ) for studying aggregation between six molecules and distance in the z-direction between the center of mass of the molecules and center of mass of all the phosphorous atoms of the lipids ( $X_{com}$ ). Radius of gyration was computed between the heavy atoms of the molecules. The metadynamics simulation was performed using GROMACS 5.1.4[29] patched with PLUMED 2.3.0[35]. The Gaussian width for the radius of gyration and distance in the z-direction was 2 Å. The biasing hills of height 0.3 kcal/mol was deposited every 2 ps. The simulation was run in NPAT ensemble where temperature was 303 K and pressure was 1 atm. The coordinate files were saved every 10 ps. All other parameters were same as in the previous molecular dynamics run.

## 2.3 Results

### 2.3.1 Membrane is not disrupted by a single molecule



**Figure 2.2:** Bacterial membrane with (a) molecule ‘A’ and (b) molecule ‘B’ (shown in red color) after 50 ns of classical molecular dynamics simulation. The phosphorous atom in the bacterial lipid head group is shown in golden color. After 50 ns, the molecule and membrane interaction results in a small membrane distortion, which is however far from the membrane disruption required for action.

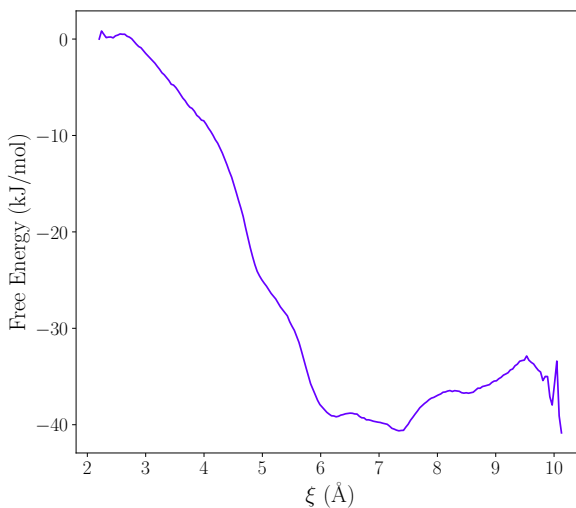
Two separate 50 ns classical molecular dynamics simulations of the bacterial membrane with a single molecule ‘A’ and molecule ‘B’ were performed. It was observed that methylated ammonium group of the molecule interacts with the phosphate group of the membrane. The phosphate groups of the MeDiAm interact either with both the leaves of the membranes (molecule ‘A’) (**Figure 2.2a**) or just the same leaf (molecule ‘B’) (**Figure 2.2b**). The difference between these two can be easily interpreted from the length of the acyl chain between the methylated ammonium group in the two cases. However, with a single MeDiAm molecule, sufficient damages or changes in the human or bacterial membrane were not observed.

### 2.3.2 Clusters of the molecules disrupts membranes

In order to study the possible consequences of these MeDiAm molecules aggregating in the membrane, six molecules ‘A’ were inserted into the membrane approximately 10 to 15 Å apart and their effect on the membrane was studied. The choice of the number of molecules was justified *a posteriori*, after the stable cluster size was found to be of 4 molecules. Since the aggregation of these molecules or the disruption of membrane did not happen in a classical molecular dynamics, we resorted to umbrella sampling

simulation. The distance between the center of mass of one molecule to the center of mass of remaining five molecules combined ( $\xi$ ) was used as a collective variable in umbrella sampling. The free energy profile obtained using umbrella sampling (**Figure 2.3**) has a minimum at  $\xi = 7.4$  Å. The structure corresponding to this stable minimum energy configuration (**Figure 2.4a**) shows that there is an aggregation of 4 molecules and that the integrity of the membrane is compromised, which we use as an *a posteriori* justification for the sufficiency of six molecules in the simulations. We then repeated these umbrella sampling simulation for the other three systems as well. The results showed a variable disruption of the membranes consistent with the experimental observations. We used this as a validation of the force fields in capturing the experimental data. **Figure 2.4d** points that the combined effect of the molecules helps in the disruption of the bacterial membrane.

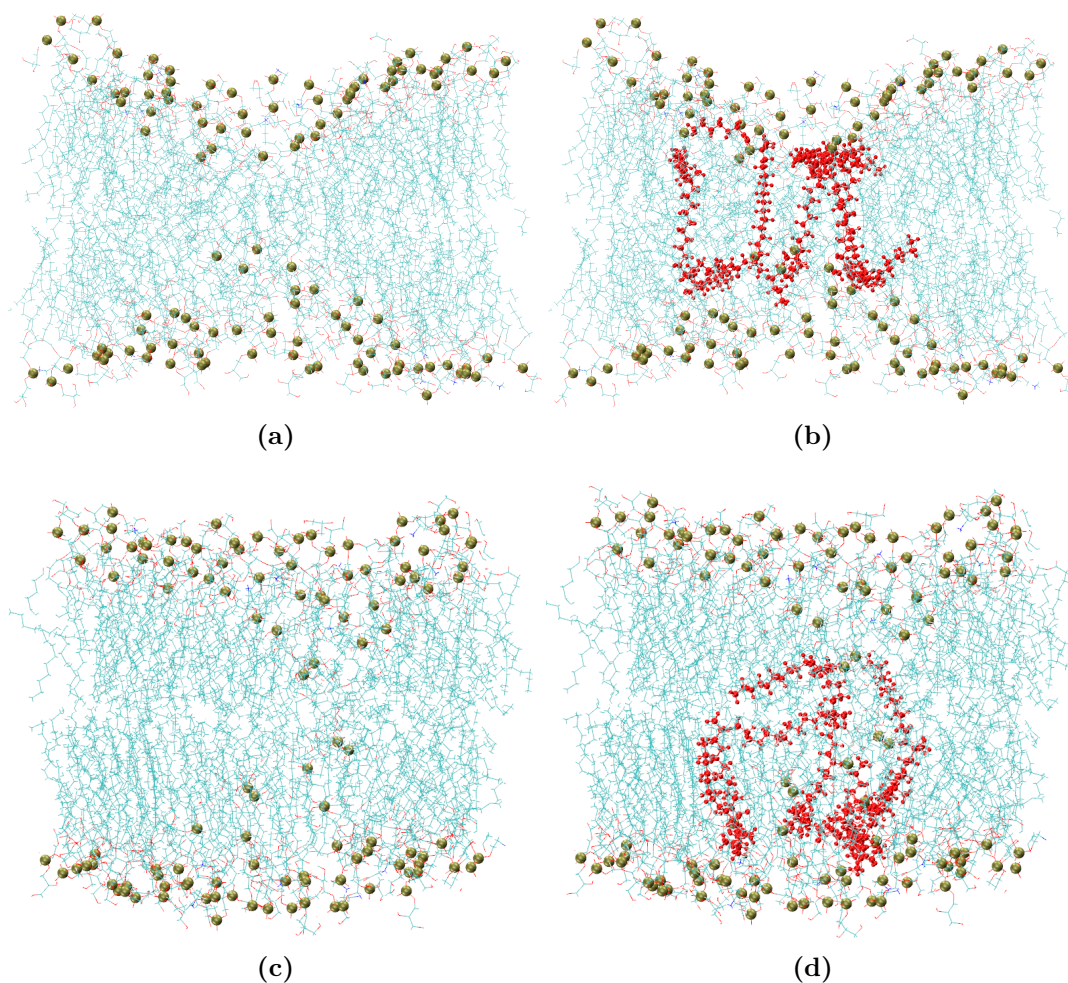
**Figure 2.3:** Free energy profile of the system containing six molecules of ‘A’, obtained using umbrella sampling. At 7.4 Å free energy profile shows a metastable state.



### 2.3.3 Selectivity in membrane disruption

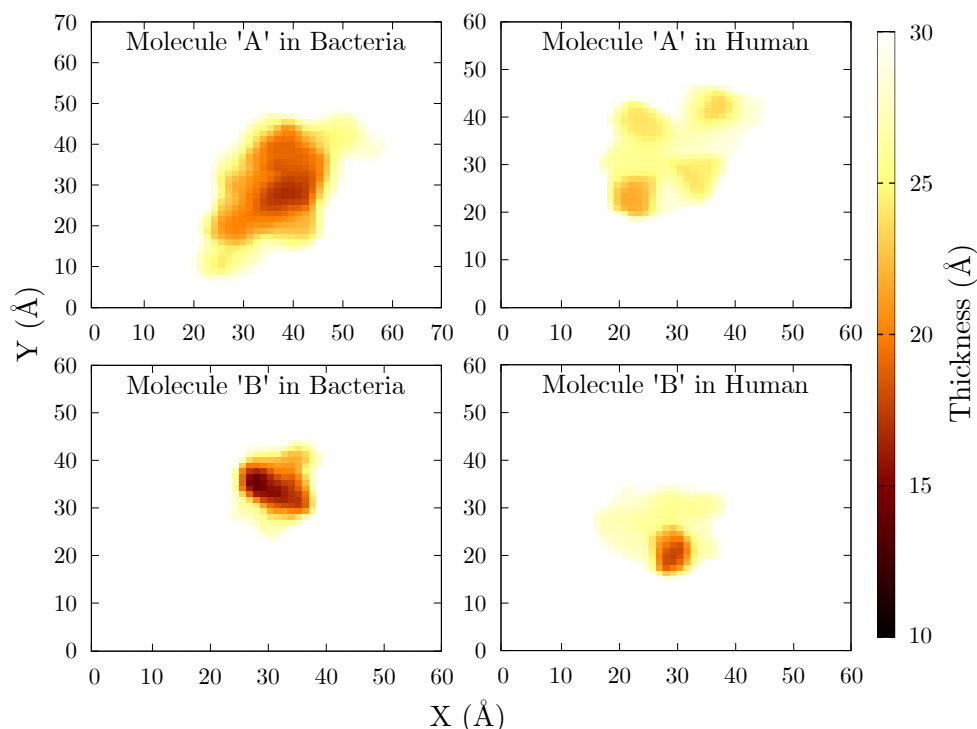
Umbrella sampling simulations on all four systems, molecules ‘A’ and ‘B’ on human and bacterial membrane were analyzed to understand the basis for selectivity. The minimum energy configurations from these four systems are compared in **Figure 2.5**. The distance between the phosphorous atoms of the upper and the lower leaflets was defined as the membrane thickness for that specific structural frame, and the average thickness of the membrane for the last 2 ns of the umbrella sampling corresponding to the minimum energy configurations are shown (**Figure 2.5**). We observed that the molecule ‘B’ causes disruption to both human and bacterial membranes, where as molecule ‘A’ is being more selective and shows effective pinching in bacterial membrane.

As depicted in **Figure 2.5**, lateral inhomogeneity was observed due to the insertion of these molecules. It is evident from the plots that a pinching of the membrane due to



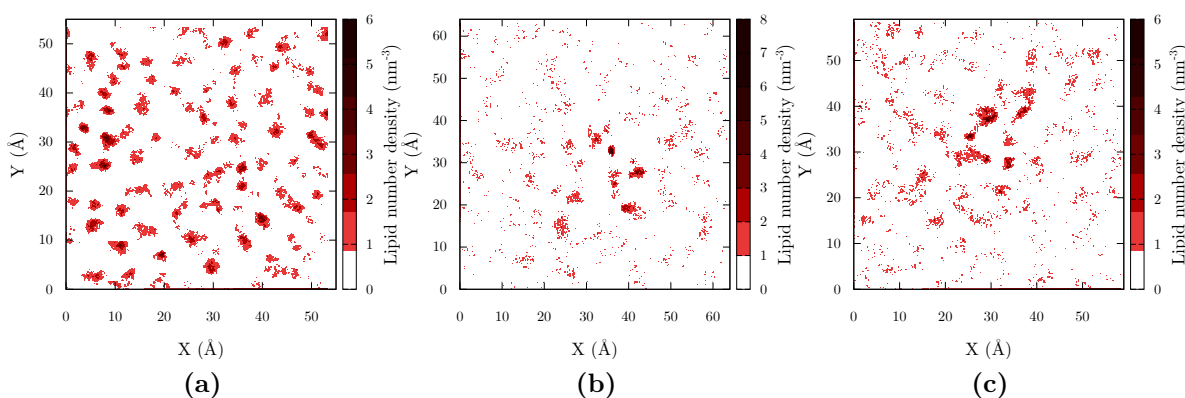
**Figure 2.4:** Metastable structures corresponding to the minimum free energy configuration of the system obtained from the umbrella sampling simulations (**Figure 2.3**) of MeDiAm ‘A’ (a, b) and MeDiAm ‘B’ (c, d) are shown. Bacterial membrane that is disrupted the molecules ‘A’ in a representation (a) without showing the ‘A’ molecules (b) and with the ‘A’ molecules suggests the mechanism of action. Similarly, (c) and (d) show the disruption to the bacterial membrane by molecule ‘B’. The phosphate ions of the lipid head group interacts with the positively charged nitrogen of these molecules. Color representations in are the same as in **Figure 2.2a**.

the insertion of these molecules. However this membrane pinching is much more localized in the case of molecule ‘B’ than in molecule ‘A’. This suggest that the molecule ‘B’ be more lethal than its other counterpart. It is evident that molecule ‘B’ tends to show aggregation, but there is no exclusive separation of the polar and non polar groups within the structure. However, molecule ‘A’ does not show any signs of aggregation and the molecules independently span over the total lipid bilayer. This difference in the behavior can be attributed to the long side chain in molecule ‘B’ that causes an increase in the van der Waals interaction as a result of which these molecules cluster together. In molecule



**Figure 2.5:** The dark patches in the membrane thickness map indicate a pinching in the membrane and potential for the antibiotic effectiveness of the molecule. Comparing the four plots explains the selectivity and effectiveness seen in the two molecules.

'A', though the central chain is long it is bound between two positively charged methyl ammonium groups, and thus their van der Waals interactions are counteracted by the electrostatic repulsion of the positively charged groups. The 2D number density was



**Figure 2.6:** 2D number density of the lipids along the x-y plane (A) Bacterial Membrane (Averaged over last 20 ns) (B) Molecule 'A' in bacteria (averaged over 6 ns) (C) Molecule 'B' in bacteria (averaged over 6 ns)

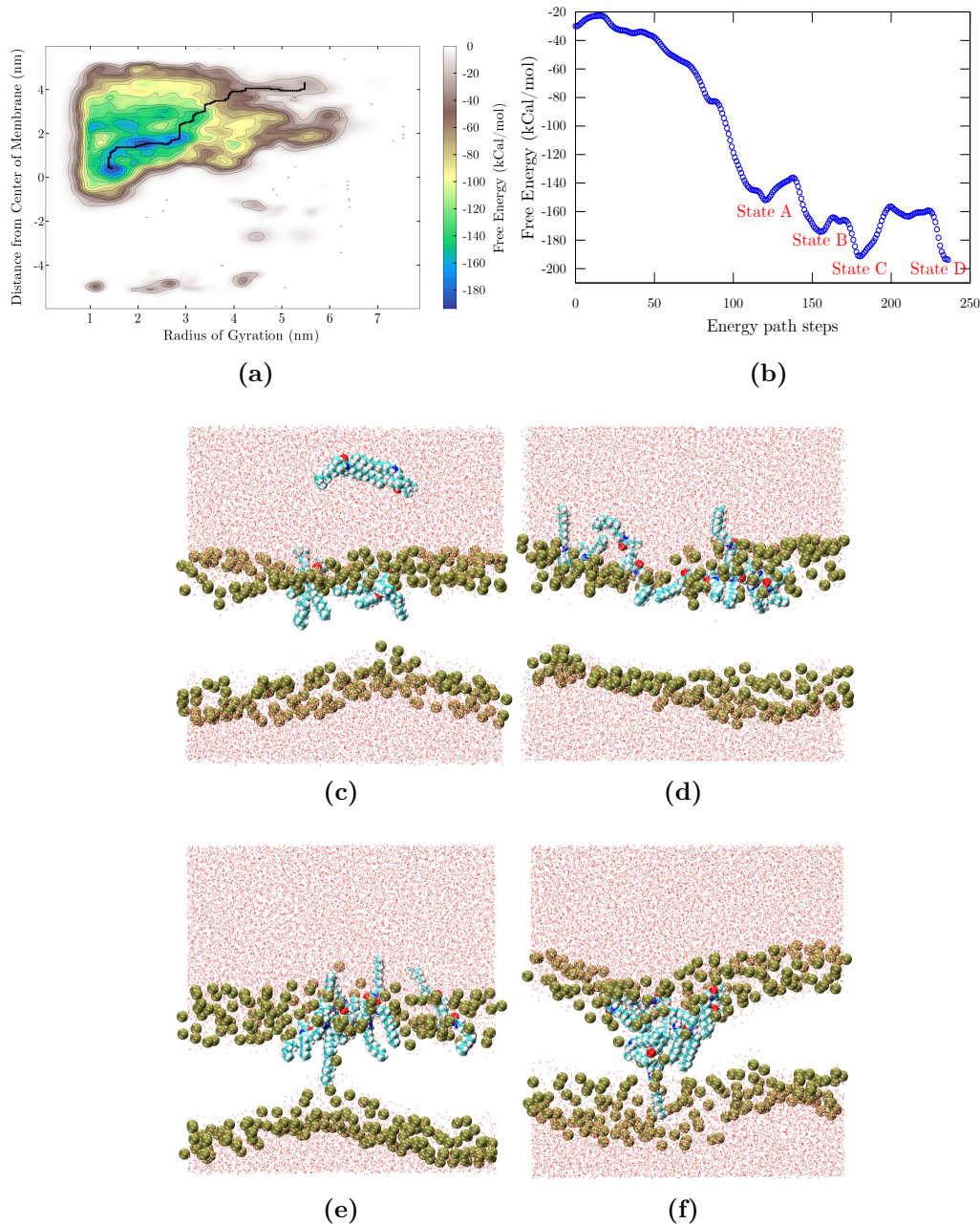
plotted to analyse the effect of inserting these molecules on the membrane (**Figure 2.6**). This indicates the distribution of the lipid molecules along the x-y plane. As is depicted

in **Figure 2.6**, the addition of the molecules caused an aggregation of the lipid molecules, thereby affecting the integrity of the membrane. Interestingly, molecule ‘A’ causes a hole formation suggesting the membrane to be completely disrupted due to the presence of these molecules. Molecule ‘B’ also caused the aggregation of the lipid molecules, but the hole formation was not very distinctly observed in this case.

### 2.3.4 Mechanism of drug interaction with membrane

The umbrella sampling simulations were performed to analyse the aggregation of molecules which were already inserted into the membrane. The selectivity and sensitivity interpreted from the membrane thickness profiles validates that the simulation and the forcefields we used capture the phenomenon being modelled. With this validation, we performed a Metadynamics calculation which simultaneously biased the insertion of these molecules into the membrane and their aggregation, to map the pathway from membrane penetration to the molecule assembly. This is complementary to the extensive aggregation studies in the four systems and was performed just for the case of molecules ‘B’ interacting with bacterial membrane.

The free energy profile for the dynamics along the two CVs was obtained using Metadynamics, the most likely pathway of simultaneous aggregation and insertion is studied. The minimum energy pathway starting from dispersed molecules in water and ultimately aggregating inside the membrane is obtained using MEPSA[36] python script (**Figure 2.7a**). While moving through the minimum free energy pathway the system encounters three intermediate metastable state excluding the final target (marked in **Figure 2.7b**) indicated as A, B, C, and D. All the configurations corresponding to the collective variables are investigated. In the state A, it is observed that most of the time four molecules are resting on or inside the membrane and two are in water. In state B, mostly five molecules are in membrane and one is outside. In state C, we see three kinds of configurations. One where, all the molecules are near the surface of the membrane but all inside. In another instance where five molecules are inside and one is outside the membrane but close to it. And in another case, one molecule is in water far from membrane, but five molecules are close to each other. And finally in state D, all the molecules are inside the membrane, clustered and some structure suggest the pinching of the membrane too.



**Figure 2.7:** (a) The free energy profile obtained from the Metadynamics simulations and the black solid line represents the minimum free energy pathway connecting the fully soluble molecules to the final disrupted membrane state. The radius of gyration of the molecules and the distance from the center of the membrane were used as two independent variables to capture the molecular aggregation and membrane insertion (b) Free energy profile along the minimum energy pathway mentioned above. (c) to (f) are the snapshot of the system at different states (A,B,C,D respectively), which shows the alignment and position of the molecules and the membrane. The membrane is denoted by the phosphorous (yellow balls) in the head group.

## 2.4 Discussion

### 2.4.1 Self assembly, aggregation, disruption

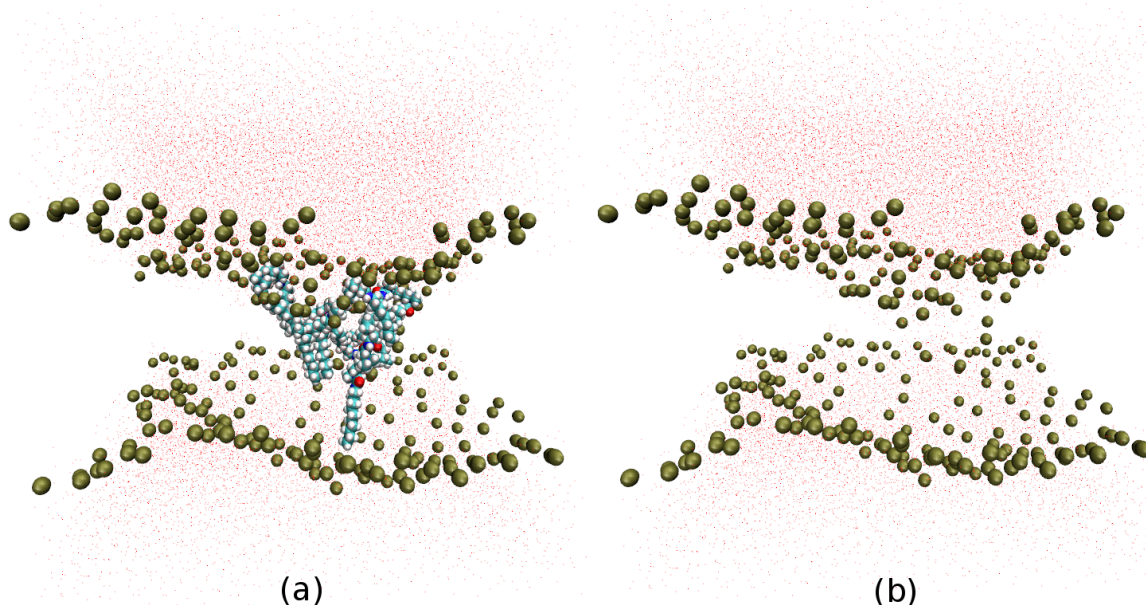
Combining the result of Metadynamics for aggregation and disruption, with those from the umbrella sampling simulations on the subtle differences in activity and selectivity, we infer a mechanism of action of this new class of molecules. As mentioned earlier, the bacterial membrane is negatively charged while the human membrane is neutral. Hence, the positively charged MeDiAms should have a stronger electrostatic interaction with the bacterial membrane thereby causing a selective rupture of the bacterial membrane. The propensity of a charged molecule to interact with a neutral but polar lipid bilayer can not be completely overlooked. However, due to a longer alkyl separation between the two ammonium groups (molecule ‘A’), this molecule spans between the two charged leaflets of the membrane. It is interesting to note that the extent of membrane thinning is less for the human membrane with respect to the bacterial membrane. It may be speculated from here that molecule ‘A’ probably has selectivity towards the human membrane. This selectivity can be attributed to the lack of aggregation in this case that reduces the positive charge density and hence the extent of electrostatic interaction with the membrane. Thus, it may be concluded that, although both molecule ‘A’ and ‘B’ are able to disrupt the bacterial membrane, the selectivity of molecule ‘A’ towards this process makes it a better drug candidate.

The overall mechanism of interaction and action of these MeDiAm molecules may thus be interpreted from our simulations. The MeDiAm molecules are drawn to the membrane and their cationic part first rests on the membrane surface due the partial negative charge of the head-group of the lipids. These molecules form an aggregated structure inside the membrane, as could be seen from minimum free energy structures (**Figure 2.8**). These molecular clusters disturb the integrity of the membrane, by bending the membrane and in some cases resulting in the formation of a water channel in the membrane. The entire mechanism may thus be summarized as one in which the dissolved molecules deposit on the membrane, followed by their self-assembly, aggregation in the membrane, resulting in a final disruption.

### 2.4.2 Differential selectivity of the molecules

Different membrane disruption mechanisms have been proposed for the AMPs. Three major pathways discussed in literature so far are the barrel-stave, carpet and toroidal mechanism.[37–40] Amphiphilicity in the secondary structure is peculiar to majority of the classes of AMPs. However, these MeDiAms do not have the facial amphiphilicity, but still have the potential to perturb the membrane’s integrity. The mechanism of action





**Figure 2.8:** A snapshot describing the membrane disruption and the formation of a water channel across it. The disrupted membrane is shown (a) with the molecule and (b) without the MeDiAm molecules. One may notice that the phosphorous of the membrane interacts with the charged groups of the molecules thinning the membrane to the point that water penetrates it.

is thus expected to be different from that of AMPs. The presence of a cationic group induces the selectivity towards binding to the anionic bacterial membrane and not the zwitterionic human membrane.[38, 41–43] Once the binding takes place, the hydrophobic components interact with the lipid chains, causing the disruption of the membrane.

## 2.5 Conclusions

In this work, the interaction of the small amphiphilic molecules with a prototypical bacterial membrane has been explored by applying enhanced sampling techniques to the all atom molecular dynamics simulation. The small molecules are highly flexible and contain charged ammonium groups along with a long hydrophobic alkyl tail. Their antimicrobial activity has been experimentally tested on a broad-spectrum of gram-positive and gram-negative bacteria.[16] Our results show a difference in the clustering ability of the two molecules and this can be controlled by tuning the distance between the charged groups and the alkyl tail length. The amphiphilic small molecules induce a segregation of the lipid molecules, resulting in the formation of domains. This is shown to induce a bilayer leakage resulting in the entry of water across the membrane. A partial selectivity can be induced in case of molecule ‘A’ which can be attributed to the tuning of the

ammonium group distance in such a way that it can span over the height of the lipid bilayer but not form aggregates.

## Bibliography

- [1] WHO, *Antimicrobial resistance: global report on surveillance*. World Health Organization, 2014.
- [2] X. Zhao, H. Wu, H. Lu, G. Li, and Q. Huang, “Lamp: a database linking antimicrobial peptides,” *PloS One*, vol. 8, no. 6, p. e66557, 2013.
- [3] A. A. Bahar and D. Ren, “Antimicrobial peptides,” *Pharmaceuticals*, vol. 6, no. 12, pp. 1543–1575, 2013.
- [4] M. Zasloff, “Antimicrobial peptides of multicellular organisms,” *Nature*, vol. 415, no. 6870, pp. 389–395, 2002.
- [5] B. P. Mowery, S. E. Lee, D. A. Kissounko, R. F. Epanand, R. M. Epanand, B. Weisblum, S. S. Stahl, and S. H. Gellman, “Mimicry of antimicrobial host-defense peptides by random copolymers,” *J. Am. Chem. Soc.*, vol. 129, no. 50, pp. 15474–15476, 2007.
- [6] C. M. Goodman, S. Choi, S. Shandler, and W. F. DeGrado, “Foldamers as versatile frameworks for the design and evolution of function,” *Nat. Chem. Biol.*, vol. 3, no. 5, pp. 252–262, 2007.
- [7] A. K. Marr, W. J. Gooderham, and R. E. Hancock, “Antibacterial peptides for therapeutic use: obstacles and realistic outlook,” *Curr. Opin. Pharmacol.*, vol. 6, no. 5, pp. 468–472, 2006.
- [8] U. K. Tirlapur and K. König, “Cell biology: targeted transfection by femtosecond laser,” *Nature*, vol. 418, no. 6895, pp. 290–291, 2002.
- [9] A. Ojha and T. Joshi, “A review on the role of pharmacogenomics in drug discovery and development,” *Int. J. Pharm. Sci. Res.*, vol. 7, no. 9, p. 3587, 2016.
- [10] C. Ghosh, M. M. Konai, P. Sarkar, S. Samaddar, and J. Haldar, “Designing simple lipidated lysines: Bifurcation imparts selective antibacterial activity,” *ChemMedChem*, vol. 11, no. 21, pp. 2367–2371, 2016.
- [11] I. S. Radzishhevsky, S. Rotem, D. Bourdetsky, S. Navon-Venezia, Y. Carmeli, and A. Mor, “Improved antimicrobial peptides based on acyl-lysine oligomers,” *Nat. Biotechnol.*, vol. 25, no. 6, p. 657, 2007.

- [12] D. Liu, S. Choi, B. Chen, R. J. Doerksen, D. J. Clements, J. D. Winkler, M. L. Klein, and W. F. DeGrado, "Nontoxic membrane-active antimicrobial arylamide oligomers," *Angew. Chem., Int. Ed.*, vol. 43, no. 9, pp. 1158–1162, 2004.
- [13] F. Nederberg, Y. Zhang, J. P. Tan, K. Xu, H. Wang, C. Yang, S. Gao, X. D. Guo, K. Fukushima, L. Li, J. L. Hedrick, and Y.-Y. Yang, "Biodegradable nanostructures with selective lysis of microbial membranes," *Nat. Chem.*, vol. 3, no. 5, pp. 409–414, 2011.
- [14] C. Ghosh, G. B. Manjunath, P. Akkapeddi, V. Yarlagadda, J. Hoque, D. S. Uppu, M. M. Konai, and J. Haldar, "Small molecular antibacterial peptoid mimics: the simpler the better!," *J. Med. Chem.*, vol. 57, no. 4, pp. 1428–1436, 2014.
- [15] M. M. Konai, C. Ghosh, V. Yarlagadda, S. Samaddar, and J. Haldar, "Membrane active phenylalanine conjugated lipophilic norspermidine derivatives with selective antibacterial activity," *J. Med. Chem.*, vol. 57, no. 22, pp. 9409–9423, 2014.
- [16] J. Hoque, M. M. Konai, S. Gonuguntla, G. B. Manjunath, S. Samaddar, V. Yarlagadda, and J. Haldar, "Membrane active small molecules show selective broad spectrum antibacterial activity with no detectable resistance and eradicate biofilms," *J. Med. Chem.*, vol. 58, no. 14, pp. 5486–5500, 2015.
- [17] R. Dey, K. De, R. Mukherjee, S. Ghosh, and J. Haldar, "Small antibacterial molecules highly active against drug-resistant staphylococcus aureus," *MedChemComm*, 2019.
- [18] J. Hoque, M. M. Konai, S. S. Sequeira, S. Samaddar, and J. Haldar, "Antibacterial and antibiofilm activity of cationic small molecules with spatial positioning of hydrophobicity: an in vitro and in vivo evaluation," *J. Med. Chem.*, vol. 59, no. 23, pp. 10750–10762, 2016.
- [19] U. Baul, K. Kuroda, and S. Vemparala, "Interaction of multiple biomimetic antimicrobial polymers with model bacterial membranes," *J. Chem. Phys.*, vol. 141, no. 8, p. 084902, 2014.
- [20] G. G. Dodson, D. P. Lane, and C. S. Verma, "Molecular simulations of protein dynamics: new windows on mechanisms in biology," *EMBO Rep.*, vol. 9, no. 2, pp. 144–150, 2008.
- [21] J. Mortier, C. Rakers, M. Bermudez, M. S. Murgueitio, S. Riniker, and G. Wolber, "The impact of molecular dynamics on drug design: applications for the characterization of ligand–macromolecule complexes," *Drug discovery today*, vol. 20, no. 6, pp. 686–702, 2015.

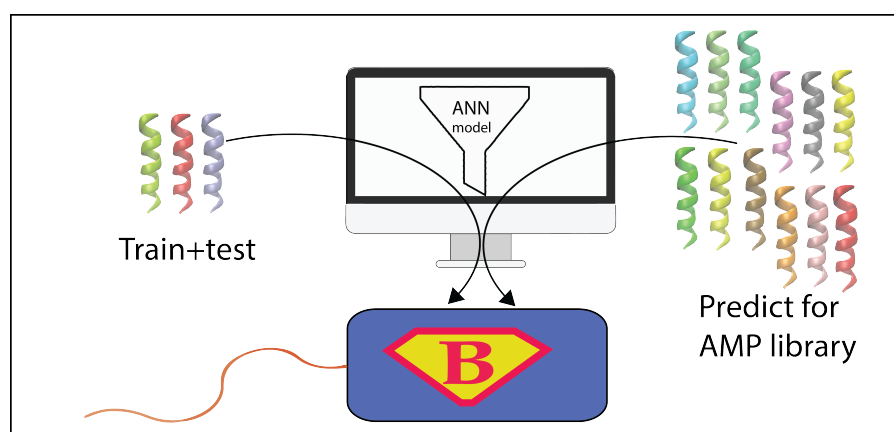
- [22] A. D. Becke, "Density-functional thermochemistry. iii. the role of exact exchange," *J. Chem. Phys.*, vol. 98, pp. 5648–5652, 1993.
- [23] P. J. Stephens, F. Devlin, C. Chabalowski, and M. J. Frisch, "Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields," *J. Phys. Chem.*, vol. 98, no. 45, pp. 11623–11627, 1994.
- [24] W. J. Hehre, R. Ditchfield, and J. A. Pople, "Self-consistent molecular orbital methods. xii. further extensions of gaussian-type basis sets for use in molecular orbital studies of organic molecules," *J. Chem. Phys.*, vol. 56, no. 5, pp. 2257–2261, 1972.
- [25] Y. Ishikawa, "A script for automated 3-dimensional structure generation and conformer search from 2-dimensional chemical drawing," *Bioinformatics*, vol. 9, no. 19, p. 988, 2013.
- [26] J. Lee, X. Cheng, J. M. Swails, M. S. Yeom, P. K. Eastman, J. A. Lemkul, S. Wei, J. Buckner, J. C. Jeong, Y. Qi, S. Jo, V. S. Pande, D. A. Case, C. L. Brooks, III, A. D. MacKerell, Jr., J. B. Klauda, and W. Im, "Charmm-gui input generator for namd, gromacs, amber, openmm, and charmm/openmm simulations using the charmm36 additive force field," *J. Chem. Theory Comput.*, vol. 12, no. 1, pp. 405–413, 2016.
- [27] J. B. Klauda, R. M. Venable, J. A. Freites, J. W. O'Connor, D. J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A. D. MacKerell Jr, and R. W. Pastor, "Update of the charmm all-atom additive force field for lipids: validation on six lipid types," *J. Phys. Chem. B*, vol. 114, no. 23, pp. 7830–7843, 2010.
- [28] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," *J. Chem. Phys.*, vol. 79, no. 2, pp. 926–935, 1983.
- [29] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, "Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," *SoftwareX*, vol. 1, pp. 19–25, 2015.
- [30] J.-P. Ryckaert, G. Ciccotti, and H. J. Berendsen, "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes," *J. Comput. Phys.*, vol. 23, no. 3, pp. 327–341, 1977.
- [31] T. Darden, D. York, and L. Pedersen, "Particle mesh ewald: An  $n \cdot \log(n)$  method for ewald sums in large systems," *J. Chem. Phys.*, vol. 98, no. 12, pp. 10089–10092, 1993.

- [32] G. M. Torrie and J. P. Valleau, “Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling,” *J. Comput. Phys.*, vol. 23, no. 2, pp. 187–199, 1977.
- [33] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, “The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method,” *J. Comput. Chem.*, vol. 13, no. 8, pp. 1011–1021, 1992.
- [34] A. Laio and M. Parrinello, “Escaping free-energy minima,” *Proc. Natl. Acad. Sci.*, vol. 99, no. 20, pp. 12562–12566, 2002.
- [35] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi, “Plumed 2: New feathers for an old bird,” *Comput. Phys. Commun.*, vol. 185, no. 2, pp. 604–613, 2014.
- [36] I. Marcos-Alcalde, J. Setoain, J. I. Mendieta-Moreno, J. Mendieta, and P. Gómez-Puertas, “Mepsa: minimum energy pathway analysis for energy landscapes,” *Bioinformatics*, vol. 31, no. 23, pp. 3853–3855, 2015.
- [37] K. Matsuzaki, “Control of cell selectivity of antimicrobial peptides,” *Biochim. Biophys. Acta, Biomembr.*, vol. 1788, no. 8, pp. 1687–1692, 2009.
- [38] R. M. Epanand, S. Rotem, A. Mor, B. Berno, and R. F. Epanand, “Bacterial membranes as predictors of antimicrobial potency,” *J. Am. Chem. Soc.*, vol. 130, no. 43, pp. 14346–14352, 2008.
- [39] K. A. Brogden, “Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria?,” *Nat. Rev. Microbiol.*, vol. 3, no. 3, pp. 238–250, 2005.
- [40] M. R. Yeaman and N. Y. Yount, “Mechanisms of antimicrobial peptide action and resistance,” *Pharmacol. Rev.*, vol. 55, no. 1, pp. 27–55, 2003.
- [41] U. Schlattner, M. Tokarska-Schlattner, S. Ramirez, A. Brückner, L. Kay, C. Polge, R. F. Epanand, R. M. Lee, M.-L. Lacombe, and R. M. Epanand, “Mitochondrial kinases and their molecular interaction with cardiolipin,” *Biochim. Biophys. Acta, Biomembr.*, vol. 1788, no. 10, pp. 2032–2047, 2009.
- [42] R. F. Epanand, M. A. Schmitt, S. H. Gellman, and R. M. Epanand, “Role of membrane lipids in the mechanism of bacterial species selective toxicity by two  $\alpha/\beta$ -antimicrobial peptides,” *Biochim. Biophys. Acta, Biomembr.*, vol. 1758, no. 9, pp. 1343–1350, 2006.

- 
- [43] P. Joanne, C. Galanth, N. Goasdoué, P. Nicolas, S. Sagan, S. Lavielle, G. Chassaing, C. El Amri, and I. D. Alves, “Lipid reorganization induced by membrane-active peptides probed using differential scanning calorimetry,” *Biochim. Biophys. Acta, Biomembr.*, vol. 1788, no. 9, pp. 1772–1781, 2009.

## Chapter 3

# Computational screening of antimicrobial peptides for *Acinetobacter baumannii*



## Abstract

*Acinetobacter baumannii*, has been developing resistance to even the last line of drugs. Antimicrobial peptides (AMPs) to which bacteria do not develop resistance easily may be the last hope. A few independent experimental studies have designed and studied the activity of AMPs on *A. baumannii*, however the number of such studies are still limited. With the goal of developing a rational approach to the screening of AMPs against *A. baumannii*, we carefully curated the drug activity data from 75 cationic AMPs, all measured with a similar protocol, and on the same ATCC 19606 strain. A quantitative model developed and validated with a part of the data. While the model may be used

This chapter is reprinted from "<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0219693>" under the Creative Commons Attribution License.

for predicting the activity of any designed AMPs, in this work, we perform an *in silico* screening for the entire database of naturally occurring AMPs, to provide a rational guidance in this urgently needed drug development.

## 3.1 Introduction

*Acinetobacter baumannii* (*A. baumannii*) [1] is mainly implicated in hospital infections and is responsible for 80% of the *Acinetobacter* infections. *A. baumannii* can also be found on normal human skin, but it generally does not pose a threat to a healthy person [2–5], besides the not-so-frequent skin and soft tissue infections, infections in the surgical site, urinary tract infection, etc [6, 7]. In the past 30 years, *A. baumannii* has evolved into a multidrug resistant (MDR) [8–10] opportunistic pathogen that selectively infects seriously ill patients in intensive care unit (ICU), trauma or burn patients [2–4, 11]. The presence of intrinsic efflux pump and high rates of genetic adaptation, contributes to adaptation against the antibiotics [12–14]. Besides, it also possesses several beta-lactamase genes which offer resistance against beta-lactam antibiotics [15, 16]. *A. baumannii* has also been developing resistance against carbapenem [17] which had been one of the last line of drugs against it. Combination therapies such as of colistin, polymixin B, and tigecycline are used to treat MDR strains, but these are complex compared to a single drug when it comes to quantification of the effect and the validation of their safety [18–20]. Due to the growing concern about MDR, new types of antimicrobial agents are needed.

Antimicrobial peptides (AMP) are a fundamental part of the innate defense system and are reportedly present in organisms from bacteria and fungus to humans [21, 22]. Although several modes of AMP activity, including DNA damage [23], RNA damage [24] and targeting ribosomes [25–27] regulatory enzymes [28] or other proteins [29] have been proposed, it is generally believed that the positively charged AMPs act by disrupting the bacterial membrane [30–32] and the membrane disruption is one of the key factor for the AMP activity [29, 33, 34]. Because of this fundamental difference in the mechanism compared to the traditional drugs, it is believed that the bacteria do not develop resistance easily against AMPs [21]. The low toxicity of AMPs towards human cells and their tendency not to result in resistant strains makes them an ideal rational choice as the next generation antimicrobial agents [35–37], possibly eventually becoming effective drugs for *A. baumannii*.

Quantitative Structure and Activity Relationship (QSAR) [38] is an approach in computer aided rational drug design, which uses biophysical or biochemical parameters of the molecules to develop a quantitative relation with the measured activities. Once validated, the computational model can be used for predicting the activities of the possible drug candidates and for pre-screening them. Recent studies have developed a QSAR



relation using 29 small molecule drug candidates which act on the oxphos metabolic path of *A. baumannii* [39]. As noted above, since bacteria are less likely to develop resistance against AMP based drugs, we focus on QSAR for AMPs against *A. baumannii*.

The present work has three major objectives. Several experimental groups have independently evaluated the activity of AMP against *A. baumannii*. We curated these experimental results against a single, well studied target, ATCC 19606 strain, whose activity is quantified using Clinical and Laboratory Standards Institute (CLSI) or related protocols.[40] We developed a computational model using neural networks to rationally predict the activity from the biochemical attributes of the AMP. Since *A. baumannii* is a growing threat, while realizing the potential limitations of training on 75 peptides, we also predict the activity of all the naturally occurring AMPs in the AMP database to enable a rational screening of AMPs against *A. baumannii*.

## 3.2 Methods

### 3.2.1 Curation of data

Training QSAR models with data from multiple sources, obtained with different protocols and on different strains can lead to poor predictive capabilities.[41] In order to standardize the data used in the analysis, we used three criteria for inclusion – the tests should be on ATCC 19606 strain, with cationic antimicrobial peptides and studied according to the CLSI or equivalent guidelines. With these inclusion criteria, we believed that the mechanism of antibiotic action will be similar and the data curated from different sources can be compared. Since data availability was limited, we had to include data from different groups. AMP sequence and activity data against *A. Baumannii* was curated from different sources [42–52] and is presented in **Table B.1** of **Appendix B**. The curated AMP data set had the activity of 75 AMPs with their length ranging from 10 to 43 amino acids and charges in the range +1 to +12. Of these, for 63 AMPs the MIC was available (referred to as quantitative data ), and for the remaining 12, only the lower bound of minimum inhibitory concentration (MIC) (referred to as the qualitative data).

### 3.2.2 Parameter computation

For each peptide eight parameters were calculated. *in vivo* aggregation propensity is calculated by using a web-based software AGGRESCAN [53]. Where the aggregation propensity is calculated on the basis of aggregation- propensity scale of amino acids. *in vitro* aggregation propensity is calculated by using TANGO software (with ionic strength 0.0 M, pH 7.0 and T=298 K)[54], where we only consider the  $\beta$ -sheet aggregation term. Aliphatic index of the peptides is calculated as described by Ikai [55]. Grand

average hydrophathy is calculated on the scale given by Kyte-Doolittle [56] and the hydrophobic moment is calculated by using HELIQUEST software [57]. Apart from the above mentioned parameters, charge, length and molecular weights were calculated using basic python script. The toxicity of the AMPs was predicted using *ToxinPred* (<http://crdd.osdd.net/raghava/toxinpred/>).[58] The method allows for the prediction of toxicity of peptides shorter than 50 amino acids. However, this was not a limitation as peptides longer than that are anyways complicated to synthesize and may not be ideal drug candidates.

### 3.2.3 Artificial neural network

Since the available data is limited, we used both the quantitative and the qualitative data, albeit with different proportions, to train and test the models. We used 63 of the MIC values from the quantitative data and 3 from the qualitative data for which the cited lower bound was treated as the MIC for the purpose of this analysis. We performed a 10-fold cross validation to check the robustness of our models. To do the 10-fold cross validation, we divide the data set into 10 different test sets, each contains 7 data points. We performed the artificial neural network (ANN) calculation for each test set by taking remaining 53 data points for training and 6 data points for validation. Rest of the 9 points from the qualitative data are used for an independent qualitative test. The activity of the AMPs was predicted by ANN model with an open module for machine learning called Scikit-learn [59] in Python using Multi-layer Perceptron Regressor function. For the activation function, logistic function was used and low memory BFGS optimization algorithm was used as a solver. Constant learning rate of 0.001 was used with maximum iteration of 5000. Three independent neural network calculations have been performed to do the 10-fold cross validation, by using one hidden layer of 6 neurons, 8 neurons and 10 neurons, respectively. 2500 trial runs in each case were made by taking 50 different random initializations for the input biases and 50 random choices for the training and validation sets. We screened the results of these 2500 trials with  $R_{training}^2 > 0.7$  and  $R_{validation}^2 > 0.6$ . Two best models were selected based on the result obtained from the 10-fold cross validation. The models were expected to perform with  $R_{test}^2 > 0.8$  for the quantitative data and at least 5 predictions for the qualitative data set. These models were then used to predict the MIC values of a complete AMP database [60–62] (<https://aps.unmc.edu>).

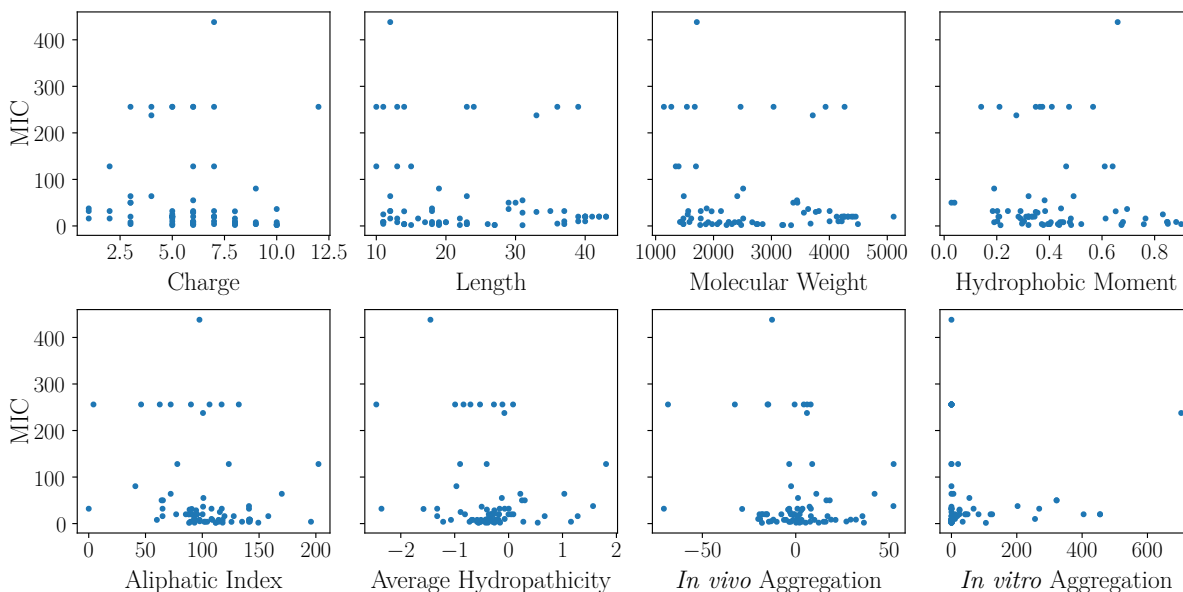
## 3.3 Results

### 3.3.1 Curated data for AMPs and their effectiveness

The data on the activity of AMPs on *A. baumannii* is scattered in literature. We curated the data mainly with the goal of developing a quantitative model, and hence restricted the focus to the most commonly studied ATCC 19606 strain. To maintain uniformity of standards, we included studies which were performed according to CLSI or equivalent guidelines. The sequence data and the antimicrobial activity of these peptides measured as the MIC was gathered (**Table B.1** of **Appendix B**). Overall, the comprehensive collection of the data on AMP activity allowed a classification based on the various biophysical parameters which are commonly used for developing a quantitative relation with activity: (1) charge, which draws the AMPs selectively to anionic membrane, (2) length, reflecting how it has to be commensurate with the membrane thickness for an improved activity [63, 64] (3) molecular weight, which gives an idea of the bulkiness and membrane penetration efficiency (4) hydrophobic moment ( $\mu_H$ ), [57] which quantifies the amphipathic characters required to form pores in the membrane, (5) aliphatic index [55], which indicates the volume of aliphatic content (A, V, I and L) of the peptide, (6) grand average of hydropathy (GRAVY) based on Kyte-Doolittle hydropathicity scale [56], (7) *in vivo* aggregation propensity, calculated by using a web-based software AGGRESCAN [53] and (8) *in vitro* ( $\beta$ -sheet) aggregation propensity, calculated by using TANGO software (with ionic strength 0.02 M, pH 7.0 and temperature 298 K) [54]. The *in vitro* aggregation, before interaction with the membrane can at times stop proteolytic degradation [65] by the bacteria but in many other cases reduce the drug potency [66, 67]. Further, the aggregation propensity affects the barrel-stave [68] and carpet mechanisms [63] of action differently. Toxicity of peptides obtained from ToxinPred[58] was categorical, and it was used only to classify the AMPs from the database as potential drug candidates, and not for the activity prediction. The distribution of the eight parameters for all the curated AMPs are given in **Figure B.1** of **Appendix B** and their individual relation with MIC in **Fig 3.1**, which shows that each of the parameters individually is not sufficient to describe the activity.

### 3.3.2 Quantitative Models for AMP activity

ANN model was used to obtain the relationship between the various above-mentioned parameters and MIC values (**Methods**). A schematic of how we developed the model is shown in **Figure B.2** of **Appendix B**. The first step was to create a model with the activity data from 75 AMPs, of which some were used for an internal assessment of the quality of predictions. The second step was to use the test set in the 75 AMP



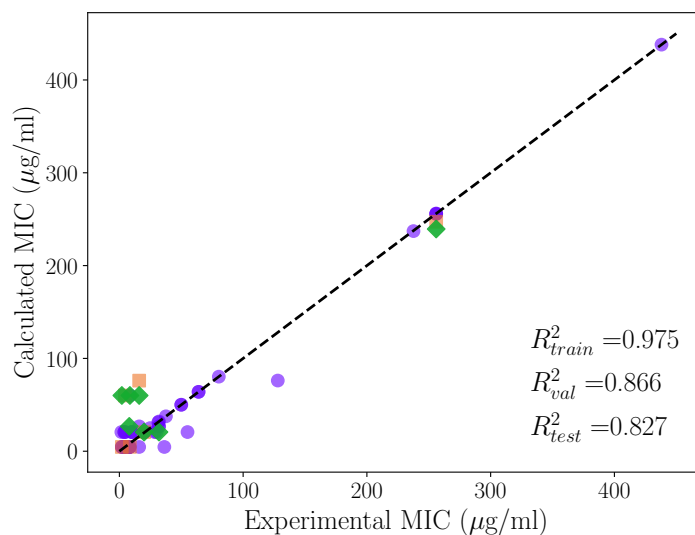
**Figure 3.1:** MIC versus different parameters. The AMPs used in the analysis along with the sequences and biophysical parameters are given in **Table B.1** of **Appendix B**.

data analysis as a secondary validation for refining the choice of model that can be used for making the predictions for the AMP database. The details are as follows. Out of the 75 AMPs curated, for 12 of them a lower bound of MIC, as being greater than a certain value (**Table B.2** of **Appendix B**), rather than a precise number was cited. To include them in the analysis, and not to reduce the data size which is already small (75 AMPs), we created two independent test sets, one in which a quantitative MIC comparison was made (referred to as quantitative data) and another qualitative one in which the calculated MIC was checked if it was more than the experimental lower bound (referred to as qualitative data). The combined data set with quantitative and qualitative data was used to construct training, validation and test sets (**Methods**). We performed a 10-fold cross validation with three different architectures with 6, 8 and 10 hidden neurons respectively. The overall error in the architecture with 8 neurons was optimal, thus justifying a small sampling around it with 6 and 10 neurons (**Table B.3** of **Appendix B**). However, all three architectures were satisfactory in their predictions (**Figure B.3**, **B.4** and **B.5** respectively of **Appendix B**), resulting in many models, which qualify for the criteria ( $R_{training}^2 > 0.7$  &  $R_{validation}^2 > 0.6$ ). Several of these models also had good predictions for the test sets, which are about 10% of the data.

### 3.3.3 Selecting the best model

In a traditional QSAR analysis, the choice of the best model would be guided by the combination of the best  $R_{training}^2$  and  $R_{validation}^2$ , following which  $R_{test}^2$  on a small fraction

of the data, in our case 7 data points, comes as a consequence. Since the goal of screening through the large set of potential AMPs whose activities against an extremely important pathogen are not yet available is more ambitious than performing well on these 7 points, we performed a secondary validation check to select the best models. We used two additional criteria:  $R_{test}^2 > 0.6$  for the quantitative and that at least 5 predictions in qualitative data set were correct to within a factor of 2 (**Table B.2 of Appendix B**). Two models satisfied these conditions, with  $R_{test}^2 > 0.8$  and they were selected. The best among these models (referred to as Model-1) obtained from the calculation with 8 hidden neurons, had good predictions ( $R_{training}^2 = 0.975$ ,  $R_{validation}^2 = 0.866$  and  $R_{test}^2 = 0.827$ ). The experimental MIC for the quantitative data set versus MIC values predicted from Model-1 is shown in **Fig 3.2**. Results obtained from another model (Model-2) are given in **Figure B.6 of Appendix B**.



**Figure 3.2:** Comparison of the experimental and calculated MIC ( $\mu\text{g/ml}$ ) of curated AMPs on *A. baumannii* obtained from Model-1, calculated by using 8 hidden neurons. Training (purple circles), validation (orange squares) and test (green diamonds) sets are shown. The data used in the analysis is shown in **Table B.1 of Appendix B**.

### 3.3.4 Predicting the results for naturally occurring AMPs

Considering the health threat *A. baumannii* is posing, and the potential of AMPs for antibiotic-resistance-free activity, we propose a rational basis for an *in silico* screening of AMPs active against *A. baumannii*. Our models were used to predict the MIC values of the 2338 AMPs obtained from database[60–62] (<https://aps.unmc.edu>) of naturally occurring AMPs. We made the predictions from Model-1 and Model-2 (**Data File in GitHub** [https://github.com/malayrb/Thesis/blob/main/Ch3/File\\_S2.xlsx](https://github.com/malayrb/Thesis/blob/main/Ch3/File_S2.xlsx)). In order to reduce the risk of a poorly trained ANN model with limited data, we filtered these

results for a consistent prediction that is within  $\Delta\text{MIC} \leq 5\mu\text{g/ml}$  for both the models (**Table 3.1**). Despite the potential statistical limitations of training and validating on 75 AMPs, a pre-screening to rationally sort multiple AMPs with their predicted activity, *in vitro* and *in vivo* aggregation potential, toxicity and length (a surrogate for synthetic complexity), all are provided in **Table 3.1** and in the **Data File in GitHub** ([https://github.com/malayrb/Thesis/blob/main/Ch3/File\\_S2.xlsx](https://github.com/malayrb/Thesis/blob/main/Ch3/File_S2.xlsx)). The computational scripts and the predictions are made accessible (**GitHub Folder:** <https://github.com/malayrb/Thesis/tree/main/Ch3/Models>), to provide an immediate access to a pool of rational choices that can help progress towards large scale experimental testing, considering the extreme urgency of developing effective strategies to combat the superbug, *A. baumannii*

**Table 3.1:** Using the 2 different models, we predicted the activity of 2338 naturally occurring AMPs documented in the AMP database. The complete list of predictions are given in the **Data File in GitHub** ([https://github.com/malayrb/Thesis/blob/main/Ch3/File\\_S2.xlsx](https://github.com/malayrb/Thesis/blob/main/Ch3/File_S2.xlsx)). However, of these the AMPs which had consistent predictions from both the models ( $\Delta\text{MIC} \leq 5\mu\text{g/ml}$ ) were selected and presented in this table. All of these were peptides listed below were non-toxic according to the predictions from ToxinPred (<http://crdd.osdd.net/raghava/toxinpred/>).[58]

Peptide	Sequence	Length	Model-1	Model-2
			MIC( $\mu\text{g/ml}$ )	MIC( $\mu\text{g/ml}$ )
AP01466	VNWKKILGKIIKVAK	15	2.84	6.20
AP00143	KKLLKWLKLL	11	9.08	4.59
AP01456	VGKTWIKVIRGIGKSKIKWQ	20	9.34	4.60
AP00708	GFKRIVQRIKDFLRNLV	17	9.38	4.59
AP00161	GLWSKIKTAGKSVAKAAAKAAVKAVTNAV	29	14.24	10.44
AP00577	GLFTLIKGAAKLIGKTVAKEAGKTGLELMACKITNQ	37	14.24	15.64
AP00608	KRIVQRIKDFLR	12	14.40	14.25
AP01525	SWLSKTYKKLENSAKKRRISEGIAIAIQGGPR	31	16.38	20.78
AP00869	ILPLVGNLLNDLL	13	17.60	20.67
AP00425	GCWSTVLGGLKKFAKGGLEAIVNPK	25	18.23	20.86
AP01388	GLLSGILNSAGLLGNLIGSLSN	23	21.02	20.67
AP00733	LLGDFFRKAREKIGEEFKRIVQRIKDFLRNLVPRTES	37	21.70	19.43
AP01387	GLLSGILNTAGLLGNLIGSLSN	23	22.83	20.67
AP00061	GIGGVLLSAGKAALKGLAKVLAEKYAN	27	23.57	20.66
AP00210	GMASKAGAIAGKIAKVALKAL	21	25.07	20.26
AP00006	GNNRPVYIPQRPHPRI	18	27.10	26.77
AP00007	GNNRPVYIPQRPHPRL	18	27.10	26.77

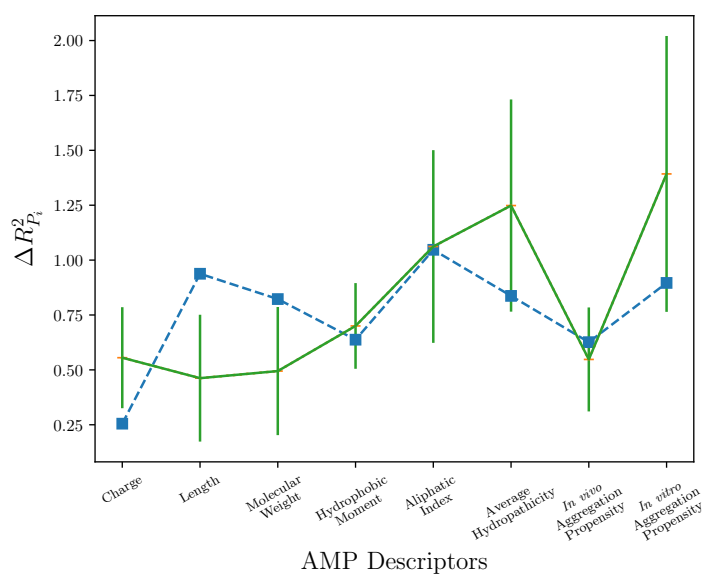
Peptide	Sequence	Length	Model-1	Model-2
			MIC( $\mu\text{g/ml}$ )	MIC( $\mu\text{g/ml}$ )
AP00024	GVS GHGQHGVHG	12	27.10	27.98
AP00025	HGVS GHGQHGVHG	13	27.10	26.77
AP00141	RKKWFW	6	27.10	26.77
AP00150	ILPWKWPWPWRR	13	27.10	26.77
AP00152	VRRFPWWPFLRR	13	27.10	26.77
AP00169	GRPNPVNTKPTPYRL	16	27.10	26.77
AP00170	VDKGSYLPRPTPPRPIYNRN	20	27.10	26.77
AP00172	GKPRPYSPRPTSHPRPIRV	19	27.10	26.79
AP00190	HPLKQYWWRPSI	12	27.10	26.77
AP00191	ECRRLCYKQRCVTYCRGR	18	27.10	26.77
AP00211	RRWCFRVCYRGFCYRKCR	18	27.10	26.77
AP00212	RRWCFRVCYKGFYRKCR	18	27.10	26.77
AP00213	KWCFRVCYRGICYRKCR	17	27.10	26.77

### 3.3.5 Parameter importance in model

It is important to know which are the parameters ( $P_i$ ) that are most responsible for the activity on *A. baumannii*. In the combined training and validation set used for accepting the models, we replaced ( $P_i$ ) with its average  $\langle P_i \rangle$  and measure the difference  $\Delta R_{P_i}^2 = R_{training+validation}^2 - R_{training+validation, \langle P_i \rangle}^2$ .  $\Delta R_{P_i}^2$  is treated as reflecting the importance of the parameter. The results obtained from Model-1 are given in **Fig 3.3** and the result obtained from another model is given in **Figure B.7** of **Appendix B**. From our calculations, we found out that the aliphatic index is the most important parameter in both the models.

### 3.3.6 Relevance of predictions for MDR strains

In order to reduce the uncertainties, our computational model was trained on data standardized in three ways, *A. baumannii* strain used, choice of cationic AMPs and measurements by CLSI method. However, considering the threat that *A. baumannii* MDR strains are posing, it is important to ask whether our calculations have any relevance to these clinical variants. The two limitations of this work are the smaller data size used for training, and it was based on ATCC 19606 strain. Interestingly, in the limited studies that we found the activity of cationic AMPs against ATCC 19606 and other MDR strains of *A. baumannii* are comparable[43, 46], thus potentially removing the latter strain specific data limitation for *A. baumannii*, although for other bacteria, such as *S. aureus* the activity changes quite significantly with the strain[69, 70]. Drawing



**Figure 3.3:** The relative importance of the different parameters in Model-1 is shown (blue color). Aliphatic index influences the outcomes of the predictions the most in this model. The green line shows the average and standard deviation value of the variable importance by taking into account all the models of 10 fold cross validation.

confidence from this fact, we used our models to predict the activity for a few MDR strains[71–73]. The results reported in **Table B.4** of **Appendix B** are encouraging at this stage, although more such validations will be helpful in establishing the utility of the screening models we proposed.

### 3.3.7 Conclusions

To our knowledge, the present work is the only QSAR study for predicting AMP activity against *A. baumannii*. The present work is different from the only other QSAR in two different ways, using AMPs instead of small molecules for a better tolerance to antibiotic resistance and a slightly larger set (75 AMPs compared to 29 small molecules). Using the ANN models we developed, we could make quantitative predictions for the entire database of naturally occurring AMPs. We hope that our work will inspire the further studies quantifying the activity of AMPs on *A. baumannii*, some of which may follow the activity predictions and others that differ offer an opportunity to retrain the ANN models.

## Bibliography

- [1] A. Nemeč, M. Musilek, M. Maixnerova, T. De Baere, T. J. van der Reijden, M. Vanechoutte, and L. Dijkshoorn, “Acinetobacter beijerinckii sp. nov. and acinetobacter gyllenbergii sp. nov., haemolytic organisms isolated from humans,” *International*



- Journal of Systematic and Evolutionary microbiology*, vol. 59, no. 1, pp. 118–124, 2009.
- [2] K. Towner, “Acinetobacter: an old friend, but a new enemy,” *Journal of Hospital Infection*, vol. 73, no. 4, pp. 355–363, 2009.
- [3] M. Struelens, E. Carlier, N. Maes, E. Serruys, W. G. Quint, and A. Van Belkum, “Nosocomial colonization and infection with multiresistant acinetobacter baumannii: outbreak delineation using dna macrorestriction analysis and pcr-fingerprinting,” *Journal of Hospital infection*, vol. 25, no. 1, pp. 15–32, 1993.
- [4] K. A. Davis, K. A. Moran, C. K. McAllister, and P. J. Gray, “Multidrug-resistant acinetobacter extremity infections in soldiers,” *Emerging infectious diseases*, vol. 11, no. 8, pp. 1218–1224, 2005.
- [5] L. S. Young, A. L. Sabel, and C. S. Price, “Epidemiologic, clinical, and economic evaluation of an outbreak of clonal multidrug-resistant acinetobacter baumannii infection in a surgical intensive care unit,” *Infection Control & Hospital Epidemiology*, vol. 28, no. 11, pp. 1247–1254, 2007.
- [6] D. M. Sievert, P. Ricks, J. R. Edwards, A. Schneider, J. Patel, A. Srinivasan, A. Kallen, B. Limbago, and S. Fridkin, “Antimicrobial-resistant pathogens associated with healthcare-associated infections summary of data reported to the national healthcare safety network at the centers for disease control and prevention, 2009–2010,” *Infection Control & Hospital Epidemiology*, vol. 34, no. 1, pp. 1–14, 2013.
- [7] L. M. Weiner, A. K. Webb, B. Limbago, M. A. Dudeck, J. Patel, A. J. Kallen, J. R. Edwards, and D. M. Sievert, “Antimicrobial-resistant pathogens associated with healthcare-associated infections: summary of data reported to the national healthcare safety network at the centers for disease control and prevention, 2011–2014,” *Infection Control & Hospital Epidemiology*, vol. 37, no. 11, pp. 1288–1301, 2016.
- [8] N. C. Gordon and D. W. Wareham, “Multidrug-resistant acinetobacter baumannii: mechanisms of virulence and resistance,” *International Journal of Antimicrobial Agents*, vol. 35, no. 3, pp. 219–226, 2010.
- [9] M. E. Falagas and D. E. Karageorgopoulos, “Pandrug resistance (pdr), extensive drug resistance (xdr), and multidrug resistance (mdr) among gram-negative bacilli: need for international harmonization in terminology,” *Clinical Infectious Diseases*, vol. 46, no. 7, pp. 1121–1122, 2008.

- [10] J. Pachón and J. Vila, “Treatment of multiresistant acinetobacter baumannii infections.,” *Current Opinion in Investigational Drugs (London, England: 2000)*, vol. 10, no. 2, pp. 150–156, 2009.
- [11] O. Öncül, Ö. Keskin, H. V. Acar, Y. Küçükardali, R. Evrenkaya, E. M. Atasoyu, C. Top, S. Nalbant, S. Özkan, G. Emekdaş, Ş. Çavuşlu, M. H. Us, A. Pahsa, and M. Gökben, “Hospital-acquired infections following the 1999 marmara earthquake,” *Journal of Hospital Infection*, vol. 51, no. 1, pp. 47–51, 2002.
- [12] J. Davies and D. Davies, “Origins and evolution of antibiotic resistance,” *Microbiology and Molecular Biology Reviews*, vol. 74, no. 3, pp. 417–433, 2010.
- [13] S. Y.-Y. Tan, S. L. Chua, Y. Liu, N. Høiby, L. P. Andersen, M. Givskov, Z. Song, and L. Yang, “Comparative genomic analysis of rapid evolution of an extreme-drug-resistant acinetobacter baumannii clone,” *Genome Biology and Evolution*, vol. 5, no. 5, pp. 807–818, 2013.
- [14] M. I. Hood, B. L. Mortensen, J. L. Moore, Y. Zhang, T. E. Kehl-Fie, N. Sugitani, W. J. Chazin, R. M. Caprioli, and E. P. Skaar, “Identification of an acinetobacter baumannii zinc acquisition system that facilitates resistance to calprotectin-mediated zinc sequestration,” *PLoS Pathogens*, vol. 8, no. 12, p. e1003068, 2012.
- [15] P.-E. Fournier, D. Vallenet, V. Barbe, S. Audic, H. Ogata, L. Poirel, H. Richet, C. Robert, S. Mangenot, C. Abergel, P. Nordmann, J. Weissenbach, D. Raoult, and J.-M. Claverie, “Comparative genomics of multidrug resistance in acinetobacter baumannii,” *PLoS Genetics*, vol. 2, no. 1, p. e7, 2006.
- [16] J. W. Sahl, J. D. Gillece, J. M. Schupp, V. G. Waddell, E. M. Driebe, D. M. Engelthaler, and P. Keim, “Evolution of a pathogen: a comparative genomics analysis identifies a genetic pathway to pathogenesis in acinetobacter,” *PloS One*, vol. 8, no. 1, p. e54287, 2013.
- [17] A. Nemec, L. Krizova, M. Maixnerova, T. J. van der Reijden, P. Deschaght, V. Passet, M. Vaneechoutte, S. Brisse, and L. Dijkshoorn, “Genotypic and phenotypic characterization of the acinetobacter calcoaceticus–acinetobacter baumannii complex with the proposal of acinetobacter pittii sp. nov.(formerly acinetobacter genomic species 3) and acinetobacter nosocomialis sp. nov.(formerly acinetobacter genomic species 13tu),” *Research in Microbiology*, vol. 162, no. 4, pp. 393–404, 2011.
- [18] B. Fan, J. Guan, X. Wang, and Y. Cong, “Activity of colistin in combination with meropenem, tigecycline, fosfomycin, fusidic acid, rifampin or sulbactam against

- extensively drug-resistant acinetobacter baumannii in a murine thigh-infection model,” *PLoS One*, vol. 11, no. 6, p. e0157757, 2016.
- [19] W.-Y. Kim, J.-Y. Moon, J. W. Huh, S.-H. Choi, C.-M. Lim, Y. Koh, Y. P. Chong, and S.-B. Hong, “Comparable efficacy of tigecycline versus colistin therapy for multidrug-resistant and extensively drug-resistant acinetobacter baumannii pneumonia in critically ill patients,” *PLoS One*, vol. 11, no. 3, p. e0150642, 2016.
- [20] G. G. Rao, N. S. Ly, J. Diep, A. Forrest, J. B. Bulitta, P. N. Holden, R. L. Nation, J. Li, and B. T. Tsuji, “Combinatorial pharmacodynamics of polymyxin b and tigecycline against heteroresistant acinetobacter baumannii,” *International Journal of Antimicrobial Agents*, vol. 48, no. 3, pp. 331–336, 2016.
- [21] M. Zasloff, “Antimicrobial peptides of multicellular organisms,” *Nature*, vol. 415, no. 6870, pp. 389–395, 2002.
- [22] R. E. Hancock and H.-G. Sahl, “Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies,” *Nature Biotechnology*, vol. 24, no. 12, pp. 1551–1557, 2006.
- [23] A. Ghosh, R. K. Kar, J. Jana, A. Saha, B. Jana, J. Krishnamoorthy, D. Kumar, S. Ghosh, S. Chatterjee, and A. Bhunia, “Indolicidin targets duplex dna: structural and mechanistic insight through a combination of spectroscopy and microscopy,” *ChemMedChem*, vol. 9, no. 9, pp. 2052–2058, 2014.
- [24] G. Hao, Y.-H. Shi, Y.-L. Tang, and G.-W. Le, “The intracellular mechanism of action on escherichia coli of bf2-a/c, two analogues of the antimicrobial peptide buforin 2,” *Journal of microbiology*, vol. 51, no. 2, pp. 200–206, 2013.
- [25] A. Krizsan, D. Volke, S. Weinert, N. Sträter, D. Knappe, and R. Hoffmann, “Insect-derived proline-rich antimicrobial peptides kill bacteria by inhibiting bacterial protein translation at the 70 s ribosome,” *Angewandte Chemie International Edition*, vol. 53, no. 45, pp. 12236–12239, 2014.
- [26] A. C. Seefeldt, F. Nguyen, S. Antunes, N. Pérébasquine, M. Graf, S. Arenz, K. K. Inampudi, C. Douat, G. Guichard, D. N. Wilson, and C. A. Innis, “The proline-rich antimicrobial peptide onc112 inhibits translation by blocking and destabilizing the initiation complex,” *Nature structural & molecular biology*, vol. 22, no. 6, p. 470, 2015.

- [27] R. N. Roy, I. B. Lomakin, M. G. Gagnon, and T. A. Steitz, "The mechanism of inhibition of protein synthesis by the proline-rich peptide oncocin," *Nature structural & molecular biology*, vol. 22, no. 6, p. 466, 2015.
- [28] A. Dangel, N. Ackermann, O. Abdel-Hadi, R. Maier, K. Önder, P. Francois, C. W. Müller, J. Pané-Farré, S. Engelmann, J. Schrenzel, H. Jürgen, and L. Christian, "A de novo-designed antimicrobial peptide with activity against multiresistant staphylococcus aureus acting on rsbw kinase," *The FASEB Journal*, vol. 27, no. 11, pp. 4476–4488, 2013.
- [29] L. T. Nguyen, E. F. Haney, and H. J. Vogel, "The expanding scope of antimicrobial peptide structures and their modes of action," *Trends in Biotechnology*, vol. 29, no. 9, pp. 464–472, 2011.
- [30] M. R. Yeaman and N. Y. Yount, "Mechanisms of antimicrobial peptide action and resistance," *Pharmacological Reviews*, vol. 55, no. 1, pp. 27–55, 2003.
- [31] K. A. Brogden, "Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria?," *Nature Reviews Microbiology*, vol. 3, no. 3, pp. 238–250, 2005.
- [32] K. Lohner and S. Blondelle, "Molecular mechanisms of membrane perturbation by antimicrobial peptides and the use of biophysical studies in the design of novel peptide antibiotics," *Combinatorial Chemistry & High Throughput Screening*, vol. 8, no. 3, pp. 241–256, 2005.
- [33] H. Jenssen, P. Hamill, and R. E. Hancock, "Peptide antimicrobial agents," *Clinical microbiology reviews*, vol. 19, no. 3, pp. 491–511, 2006.
- [34] A. T. Yeung, S. L. Gellatly, and R. E. Hancock, "Multifunctional cationic host defence peptides and their clinical applications," *Cellular and Molecular Life Sciences*, vol. 68, no. 13, p. 2161, 2011.
- [35] P. Oyston, M. Fox, S. Richards, and G. Clark, "Novel peptide therapeutics for treatment of infections," *Journal of Medical Microbiology*, vol. 58, no. 8, pp. 977–987, 2009.
- [36] K. Poole, "Overcoming multidrug resistance in gram-negative bacteria.," *Current Opinion in Investigational Drugs (London, England: 2000)*, vol. 4, no. 2, pp. 128–139, 2003.

- [37] E. Guaní-Guerra, T. Santos-Mendoza, S. O. Lugo-Reyes, and L. M. Terán, “Antimicrobial peptides: general overview and clinical implications in human health and disease,” *Clinical Immunology*, vol. 135, no. 1, pp. 1–11, 2010.
- [38] H. Jenssen, C. D. Fjell, A. Cherkasov, and R. E. Hancock, “Qsar modeling and computer-aided design of antimicrobial peptides,” *Journal of Peptide Science: An Official Publication of the European Peptide Society*, vol. 14, no. 1, pp. 110–114, 2008.
- [39] H. S. Sayiner, A. A. Abdalrahm, M. A. Basaran, V. Kovalishyn, and F. Kandemirli, “The quantum chemical and qsar studies on acinetobacter baumannii oxphos inhibitors,” *Medicinal Chemistry*, vol. 14, no. 3, pp. 253–268, 2018.
- [40] M. A. Wikler, “Methods for dilution antimicrobial susceptibility tests for bacteria that grow aerobically: approved standard,” *Clinical and Laboratory Standards Institute*, 2006.
- [41] L. Zhao, W. Wang, A. Sedykh, and H. Zhu, “Experimental errors in qsar modeling sets: What we can do and what we cannot do,” *ACS omega*, vol. 2, no. 6, pp. 2805–2812, 2017.
- [42] X. Vila-Farres, C. G. De La Maria, R. López-Rojas, J. Pachón, E. Giralt, and J. Vila, “In vitro activity of several antimicrobial peptides against colistin-susceptible and colistin-resistant acinetobacter baumannii,” *Clinical Microbiology and Infection*, vol. 18, no. 4, pp. 383–387, 2012.
- [43] S.-Y. Peng, R.-I. You, M.-J. Lai, N.-T. Lin, L.-K. Chen, and K.-C. Chang, “Highly potent antimicrobial modified peptides derived from the acinetobacter baumannii phage endolysin lysab2,” *Scientific reports*, vol. 7, no. 1, p. 11477, 2017.
- [44] E. Jayamani, R. Rajamuthiah, J. Larkins-Ford, B. B. Fuchs, A. L. Conery, A. Vilcinskas, F. M. Ausubel, and E. Mylonakis, “Insect-derived cecropins display activity against acinetobacter baumannii in a whole-animal high-throughput caenorhabditis elegans model,” *Antimicrobial agents and chemotherapy*, pp. 1728–1737, 2015.
- [45] M. Jaśkiewicz, D. Neubauer, K. Kazor, S. Bartoszevska, and W. Kamysz, “Antimicrobial activity of selected antimicrobial peptides against planktonic culture and biofilm of acinetobacter baumannii,” *Probiotics and Antimicrobial Proteins*, pp. 1–8, 2018.

- [46] M. F. Mohamed, A. Brezden, H. Mohammad, J. Chmielewski, and M. N. Seleem, “A short d-enantiomeric antimicrobial peptide with potent immunomodulatory and antibiofilm activity against multidrug-resistant *Pseudomonas aeruginosa* and *Acinetobacter baumannii*,” *Scientific Reports*, vol. 7, no. 1, p. 6953, 2017.
- [47] J. Andrä, D. Monreal, G. M. de Tejada, C. Olak, G. Brezesinski, S. S. Gomez, T. Goldmann, R. Bartels, K. Brandenburg, and I. Moriyon, “Rationale for the design of shortened derivatives of the nk-lysin derived antimicrobial peptide nk-2 with improved activity against gram-negative pathogens,” *Journal of Biological Chemistry*, 2007.
- [48] E. M. Kohn, D. J. Shirley, L. Arotzky, A. M. Picciano, Z. Ridgway, M. W. Urban, B. R. Carone, and G. A. Caputo, “Role of cationic side chains in the antimicrobial activity of c18g,” *Molecules*, vol. 23, no. 2, p. 329, 2018.
- [49] D. Li, Y. Yang, Z. Tian, J. Lv, F. Sun, Q. Wang, Y. Liu, and P. Xia, “Synergistic antibiotic effect of looped antimicrobial peptide clp-19 with bactericidal and bacteriostatic agents,” *Oncotarget*, vol. 8, no. 34, pp. 55958–55966, 2017.
- [50] A. Almaaytah, A. Ya’u, A. Abualhaijaa, S. Tarazi, N. Alshar’i, and Q. Al-Balas, “Peptide consensus sequence determination for the enhancement of the antimicrobial activity and selectivity of antimicrobial peptides,” *Infection and drug resistance*, vol. 10, pp. 1–17, 2017.
- [51] T. T. Thomsen, *Peptide Antibiotics for ESKAPE Pathogens: Past, Present and Future Perspectives of Antimicrobial Peptides for the Treatment of Serious Gram-Negative and Gram-Positive Infections*. PhD thesis, Department of Biology, Faculty of Science, University of Copenhagen, 2016.
- [52] A. Almaaytah, M. T. Qaoud, A. Abualhaijaa, Q. Al-Balas, and K. H. Alzoubi, “Hybridization and antibiotic synergism as a tool for reducing the cytotoxicity of antimicrobial peptides,” *Infection and Drug Resistance*, vol. 11, pp. 835–847, 2018.
- [53] O. Conchillo-Solé, N. S. de Groot, F. X. Avilés, J. Vendrell, X. Daura, and S. Ventura, “Aggrescan: a server for the prediction and evaluation of “hot spots” of aggregation in polypeptides,” *BMC Bioinformatics*, vol. 8, no. 1, p. 65, 2007.
- [54] A.-M. Fernandez-Escamilla, F. Rousseau, J. Schymkowitz, and L. Serrano, “Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins,” *Nature Biotechnology*, vol. 22, no. 10, pp. 1302–1306, 2004.

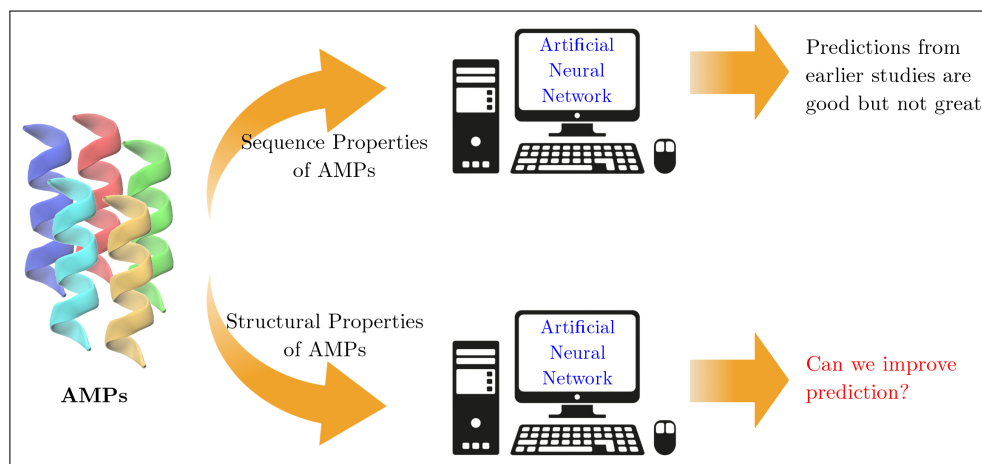
- [55] A. Ikai, “Thermostability and aliphatic index of globular proteins,” *The Journal of Biochemistry*, vol. 88, no. 6, pp. 1895–1898, 1980.
- [56] J. Kyte and R. F. Doolittle, “A simple method for displaying the hydropathic character of a protein,” *Journal of Molecular Biology*, vol. 157, no. 1, pp. 105–132, 1982.
- [57] R. Gautier, D. Douguet, B. Antony, and G. Drin, “Heliquet: a web server to screen sequences with specific  $\alpha$ -helical properties,” *Bioinformatics*, vol. 24, no. 18, pp. 2101–2102, 2008.
- [58] S. Gupta, P. Kapoor, K. Chaudhary, A. Gautam, R. Kumar, G. P. Raghava, O. S. D. D. Consortium, and G. P. S. Raghava, “In silico approach for predicting toxicity of peptides and proteins,” *PLoS One*, vol. 8, no. 9, p. e73957, 2013.
- [59] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, “Scikit-learn: Machine learning in python,” *Journal of Machine Learning Research*, vol. 12, no. Oct, pp. 2825–2830, 2011.
- [60] Z. Wang and G. Wang, “Apd: the antimicrobial peptide database,” *Nucleic Acids Research*, vol. 32, no. suppl\_1, pp. D590–D592, 2004.
- [61] G. Wang, X. Li, and Z. Wang, “Apd2: the updated antimicrobial peptide database and its application in peptide design,” *Nucleic Acids Research*, vol. 37, no. suppl\_1, pp. D933–D937, 2008.
- [62] G. Wang, X. Li, and Z. Wang, “Apd3: the antimicrobial peptide database as a tool for research and education,” *Nucleic Acids Research*, vol. 44, no. D1, pp. D1087–D1093, 2015.
- [63] Y. Shai, “Mode of action of membrane active antimicrobial peptides,” *Peptide Science: Original Research on Biomolecules*, vol. 66, no. 4, pp. 236–248, 2002.
- [64] B. Deslouches, S. M. Phadke, V. Lazarevic, M. Cascio, K. Islam, R. C. Montelaro, and T. A. Mietzner, “De novo generation of cationic antimicrobial peptides: influence of length and tryptophan substitution on antimicrobial activity,” *Antimicrobial Agents and Chemotherapy*, vol. 49, no. 1, pp. 316–322, 2005.
- [65] Z. Oren, J. C. Lerman, G. H. Gudmundsson, B. Agerberth, and Y. Shai, “Structure and organization of the human antimicrobial peptide ll-37 in phospholipid membranes:

- relevance to the molecular basis for its non-cell-selective activity.,” *Biochemical Journal*, vol. 341, no. Pt 3, pp. 501–513, 1999.
- [66] I. Kustanovich, D. E. Shalev, M. Mikhlin, L. Gaidukov, and A. Mor, “Structural requirements for potent versus selective cytotoxicity for antimicrobial dermaseptin s4 derivatives,” *Journal of Biological Chemistry*, vol. 277, no. 19, pp. 16941–16951, 2002.
- [67] R. Feder, A. Dagan, and A. Mor, “Structure-activity relationship study of antimicrobial dermaseptin s4 showing the consequences of peptide oligomerization on selective cytotoxicity,” *Journal of Biological Chemistry*, vol. 275, no. 6, pp. 4230–4238, 2000.
- [68] G. Ehrenstein and H. Lecar, “Electrically gated ionic channels in lipid bilayers,” *Quarterly reviews of biophysics*, vol. 10, no. 1, pp. 1–34, 1977.
- [69] A. Majumder, M. R. Biswal, and M. K. Prakash, “One drug multiple targets: An approach to predict drug efficacies on bacterial strains differing in membrane composition,” *ACS Omega*, vol. 4, no. 3, pp. 4977–4983, 2019.
- [70] N. N. Mishra, J. McKinnell, M. R. Yeaman, A. Rubio, C. C. Nast, L. Chen, B. N. Kreiswirth, and A. S. Bayer, “In vitro cross-resistance to daptomycin and host defense cationic antimicrobial peptides in clinical methicillin-resistant staphylococcus aureus isolates,” *Antimicrobial agents and chemotherapy*, vol. 55, no. 9, pp. 4012–4018, 2011.
- [71] R. C. das Neves, M. R. Mortari, E. F. Schwartz, A. Kipnis, and A. P. Junqueira-Kipnis, “Antimicrobial and antibiofilm effects of peptides from venom of social wasp and scorpion on multidrug-resistant acinetobacter baumannii,” *Toxins*, vol. 11, no. 4, p. 216, 2019.
- [72] R. Hirsch, J. Wiesner, A. Marker, Y. Pfeifer, A. Bauer, P. E. Hammann, and A. Vilcinskis, “Profiling antimicrobial peptides from the medical maggot *lucilia sericata* as potential antibiotics for mdr gram-negative bacteria,” *Journal of Antimicrobial Chemotherapy*, vol. 74, no. 1, pp. 96–107, 2018.
- [73] I. Di Bonaventura, S. Baeriswyl, A. Capecchi, B.-H. Gan, X. Jin, T. N. Siriwardena, R. He, T. Köhler, A. Pompilio, G. Di Bonaventura, *et al.*, “An antimicrobial bicyclic peptide from chemical space against multidrug resistant gram-negative bacteria,” *Chemical communications*, vol. 54, no. 40, pp. 5130–5133, 2018.



# Chapter 4

## Molecular dynamics based antimicrobial activity descriptors for synthetic cationic peptides



### Abstract

There is an urgent need to identify novel antimicrobial drugs in light of the development of resistance by the bacteria for a broad spectrum of antibiotics. Antimicrobial peptides (AMPs) are proving to be an effective remedy to which bacteria have not been able to develop resistance easily. With the goal of progressing towards a rational design of AMPs, we developed a neural network based quantitative model relating their physicochemical properties to their activity. A set of synthetic cationic polypeptides (CAMEL-s) (Mee *et al.* in *J. Peptide Res.* 49:89, 1997) which were studied systematically

This chapter is reprinted by permission from Springer Nature: Journal of Chemical Sciences, Prakash *et al.* Copyright (2019) (<https://doi.org/10.1007/s12039-019-1590-0>).

in experiments was used in the development of our model. Intuitive variables derived from short molecular dynamics simulations in octanol were used as descriptors, resulting in a good prediction of activity and underscoring the possibility of a rational design.

## 4.1 Introduction

Antimicrobial peptides (AMPs) that are part of innate immune system in all multicellular organism are typically of about 12 to 50 amino acids. [1] These AMPs are typically effective against various pathogens like bacteria, fungi, yeast, viruses and even cancer cells.[2] Conventional antibiotics which targets specific bacterial sites sometimes become ineffective by a single mutation at the target site.[3, 4] Unlike these antibiotics, AMPs have an entirely different mode of action by attacking the cell membrane of the bacteria.[5] Various mechanistic models have been proposed for the action of AMPs:[6] barrel stave model, toroidal model and carpet model, which result in the formation of pores or breaking up of the membrane into micelles. Bacteria do not develop resistance to these AMPs easily. Therefore there is an urgent need to identify novel AMPs. There is a growing interest in the development of synthetic AMPs that are able to mimic the natural antimicrobial peptides in their functionality.[7, 8] But owing to the high cost of production, poor pharmacokinetics and susceptibility to proteolysis, large scale pharmaceutical application has not been possible yet.[9] Resolving some of these problems as well as developing AMPs against new pathogens requires a rational design approach. Understanding the properties of AMPs would serve as an important tool to design novel drug molecules with the essential features that make them antimicrobial, with possibly a low rate of haemolysis. In this work, we focus on one such quantitative structure-activity relationship (QSAR) models, relating the physicochemical properties of AMPs to their activity.

We performed our analysis on a series of cationic AMPs, also known as CAMEL-s.[10] CAMEL-s were designed by making substitutions in the 15 amino acid long synthetic design. Their activity was experimentally quantified[10] in terms of the minimum inhibitory concentration (MIC) against 24 Gram-positive and Gram-negative bacteria. QSAR methods have been used for modelling activity in various contexts.[11–15] The activity data of CAMEL-s peptide was previously used by other QSAR models.[10, 16, 17] These earlier models predominantly used either sequence based parameters[18] or in turn depended on other predictive tools such as *in vitro* and *in vivo* aggregation propensities.[19] In our work we explore the possibility of predicting activity of CAMEL-s peptides using their structure and dynamics.

## 4.2 Methods

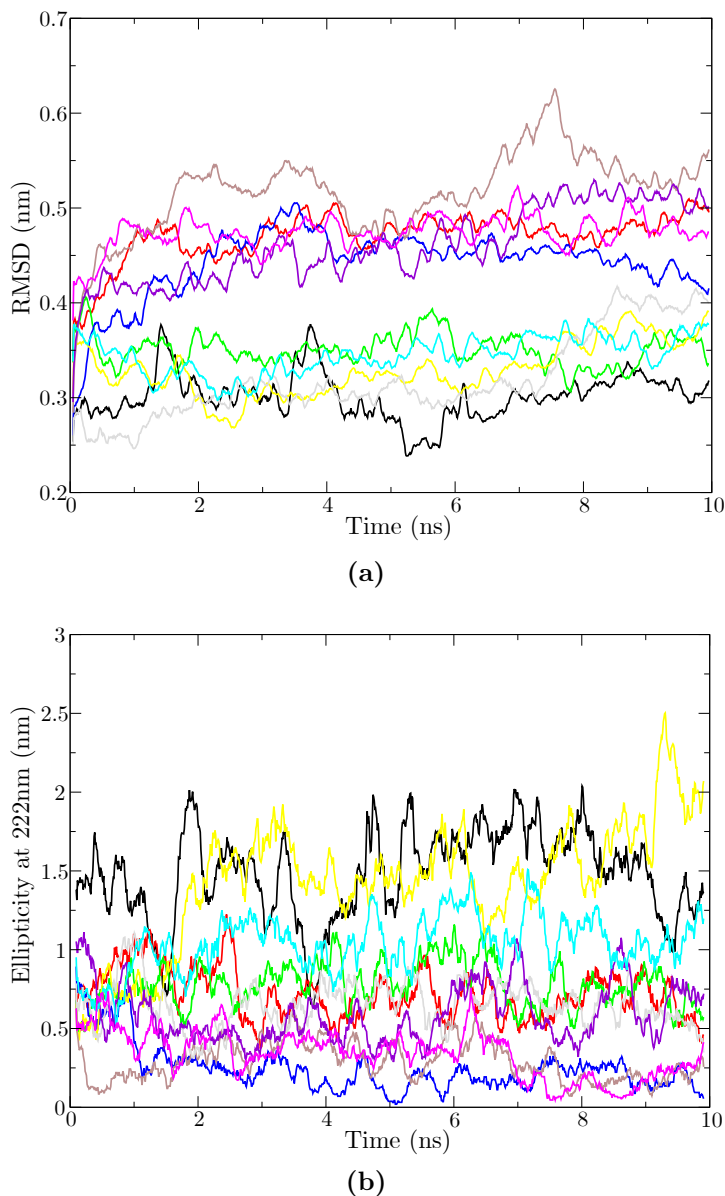
### 4.2.1 Structure and dynamics simulation

The structure of the peptides were generated via homology modelling of the sequence using the Modeller tool.[20] We used  $\alpha$  helix as a template for this modelling. Two different sets of molecular dynamics (MD) simulations were performed using GROMACS 5.1.1,[21] one with the peptides solvated in water and another in octanol. For the simulation in octanol, OPLS-AA/L[22] all-atom force field was used for both peptides and octanol. Instead, for the simulation of peptides in water we used AMBER99SB forcefield[23] for protein and TIP3P[24] for water. Most of the average properties converged within 5 ns of MD simulation. RMSD relative to the starting structure and ellipticity at 222 nm calculated using `gmx rms` and `gmx helix` program were used respectively for monitoring the convergence. The results shown in **Figure 4.1** show that the systems stabilize after 5 ns. The following parameters were obtained by post-processing the last 5 ns of MD trajectories using the tools available with GROMACS 5.1.1:[21] average dipole moment (`gmx dipoles`), average number of hydrogen bonds within the peptide (`gmx hbond`), solvent accessible surface area (SASA) (`gmx sasa`), ellipticity which is proportional to the number of residues in the helix (`gmx helix`), root mean square deviation, volume (`gmx sasa`), density (`gmx sasa`), moment of inertia along major, middle and minor axes (`gmx principal`). In addition to these, averaged root mean square deviation relative to the starting structure was calculated using VMD.[25] These structural properties were then used as input parameters to the artificial neural network study.

### 4.2.2 Neural Network Model

Artificial neural network based model was set up using sci-kit learn module[26] in Python 2.7.6. A two layer feed-forward network using MLPRegressor with logistic sigmoid hidden layer and a linear output layer was used. The network was trained using the low-memory BFGS optimization algorithm. The number of neurons for the hidden layer were varied and the number corresponding to the best  $R^2$  value of the validation set was chosen for further analysis. The calculations were repeated using independent calculations with 3 to 5 neurons in the hidden layer. The results we report correspond to the optimal predictions we obtained with 4 hidden neurons. The data was divided into training sets with 81 peptides, validation sets with 10 and test sets with 10 peptides. 50 different training sets were prepared by randomly choosing them from the data. For every pair of training and validation sets, 50 sets of random inputs for initialization of weights and biases of neural network were used, making them a total of 2500 calibration sets. Of these 12 sets were chosen where the  $R^2$  for the training and validation data was greater

than 0.6. Based on these,  $R^2$  for the test set was calculated.



**Figure 4.1:** (a)RMSD of the peptide was calculated with reference to the starting structure and (b) Ellipticity at 222 nm value calculated for some of the peptides solvated in octanol, shows that the physical parameters stabilizes after 5 ns. We used data from 5 ns to 10 ns in our analysis.

## 4.3 Results

### 4.3.1 Dynamical properties

The sequence and biological activity data of 101 CAMEL-s peptides was obtained from Mee *et al.*[10] Starting from the sequence, a putative helical structure was ascribed to

each of the peptides, as AMPs mostly assume an  $\alpha$  helical structure, especially in the membrane. The structures were generated using Modeller (**Methods Section**). The dynamics of the peptides in octanol medium, which is believed to mimic the properties in lipid membranes, was investigated using molecular dynamics (MD) simulations and ten different dynamical properties were calculated as noted in the **Methods section**. The properties converged with 2 ns as shown in **Figure 4.1** and **Figure C.1** of **Appendix C**. Since biological activity of proteins or peptides usually is associated with their dynamics, we modelled it using these properties derived from MD simulations. We also repeated the MD with the peptides solvated in water, although as we comment later the predictions were better with octanol simulations.

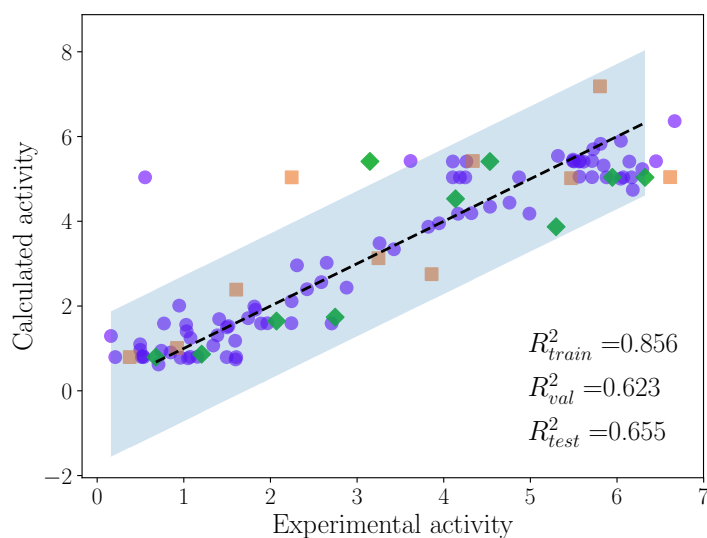
### 4.3.2 Model for CAMEL-s activity prediction

Ten different variables obtained by processing the MD trajectories (**Methods section**) were used as descriptors of the activity. These variables included the descriptors of flexibility of the peptide such as RMSD or properties which could describe how easy or difficult an insertion into the membrane might be – such as the average volume or density of the peptide along the trajectory. The activity data was modelled using artificial neural networks (ANN) on these descriptors. The data was randomly split into several sets – 80% training, 10% validation and tested with remaining 10% test set (**Methods**). While considering the parameters obtained from simulation in water the neural network model prediction was poor (**Figure C.2** of **Appendix C**). But with simulation in octanol, the model was able to predict the biological activity with fairly good accuracy, the coefficient of determination,  $R^2$  for the training set being 0.86, and that for the validation and test set was 0.62 and 0.65, respectively (**Figure 4.2**).

Of the 2500 calibration sets we tried with random initialization (**Methods**), 12 resulted in  $R^2$  of training and validation sets to be greater than 0.6. We reported the data set with the best  $R^2$  for the test set. When we repeated the calculations with 70% training, 10% validation and tested with remaining 20%, the results were not as good (data not shown), possibly because of the size of the data set used in the analysis.

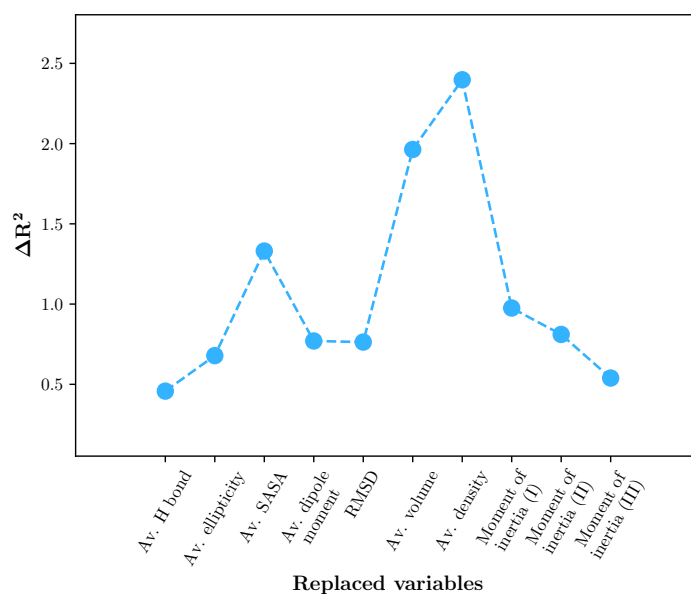
### 4.3.3 Importance of the variables used

In order to understand the relative importance of the ten variables, we compared the  $R^2$  value of the model by replacing variables with their average values. The replaced variables were used as a test data and the  $R^2$  was calculated. The difference between total  $R^2$  of original and modified data was calculated and plotted in **Figure 4.3**. It is observed that the density of AMPs affects their activity the most, followed by volume and SASA. From **Figure 4.3** it is observed that the density is the most relevant variable

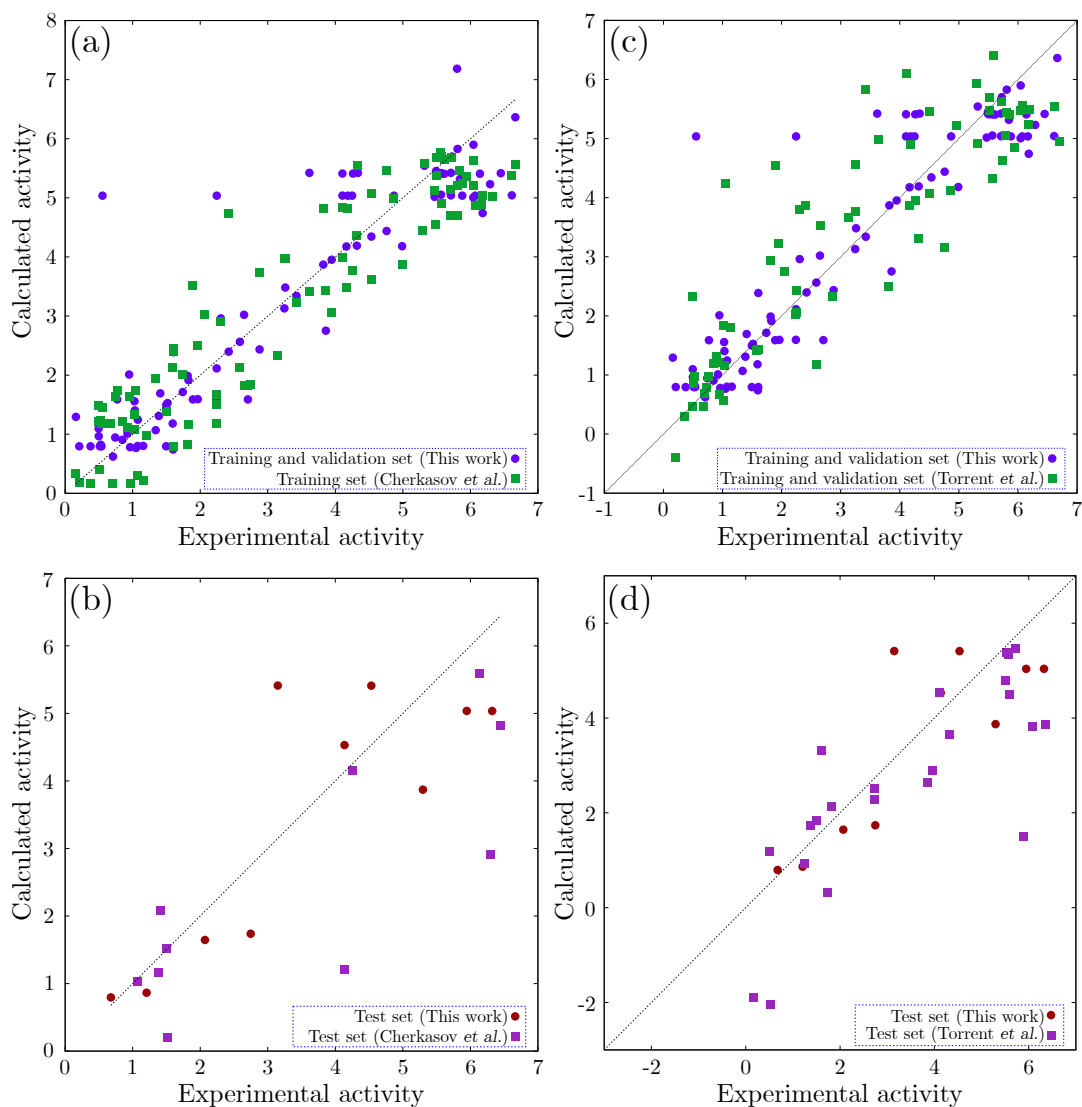


**Figure 4.2:** The experimental vs the calculated values of biological activity for the training (purple circle), validation (orange square) and the test set (green diamond). Blue color area corresponds to 95% confidence interval.

and average hydrogen bond is the least. But overall collective effect of the variables helps to determine the activity of the peptides.



**Figure 4.3:** The relative importance of the different variables used in our model was calculated. The difference in  $R^2$  of predictions when each individual variable is replaced by its average value is considered as a measure of importance. Av. denotes average in the figure above.



**Figure 4.4:** Comparison between the calculations from our work with those from earlier reported works.[18, 19] (a) Calculations from Cherkasov *et al.* training and validation sets ( $R^2 = 0.86$ ) and (b) test set ( $R^2 = 0.46$ ) (c) Calculations from Torrent *et al.* training and validation sets ( $R^2 = 0.75$ ) and (d) Test set ( $R^2 = 0.45$ ). Our calculations had  $R^2$  0.86, 0.62 and 0.65 for training, validation and test sets respectively.

## 4.4 Discussion

Sequence-structure-dynamics-function is the present paradigm in understanding how a sequence translates into a functional peptide or a protein by virtue of its flexible motions in the medium surrounding it. Continuing on this theme, we have used the dynamical properties of AMPs along with artificial neural network models to predict their antibacterial activity. A couple of earlier studies[18, 19] attempted the prediction of the activity. We performed the present research with the goal of addressing the following

concerns: (i) Since protein/peptide dynamics is believed to be the underlying factor determining their activity, can the the activity be captured using molecular dynamics alone? (ii) Can simple and intuitive variables e.g, volume, density, SASA and RMSD of the peptide be sufficient, as compared to the earlier reported works[18, 19] which required derived variables such as aggregation propensities which were in turn predictions from other computational models (iii) Can a secondary dependence of the descriptive parameters on other predictive tools/software which have not been thoroughly validated be avoided? Our results answer these questions in affirmative, demonstrating the utility of MD for the prediction of AMP activity. The comparison between earlier work and our work is plotted in the **Figure 4.4** and it shows that the results of our predictions are at least as good or slightly better than the predictions from the previous models.

Further, we could see from our calculations that the parameters calculated using MD of peptides in octanol performed better than those from water. The simulations in water resulted in structures which were more flexible than in octanol. The variables density, volume and SASA from the simulations in octanol playing an crucial role in the activity of CAMEL-s peptides and the relatively flexible water models not being representative of the activity suggest the importance of the dynamical state of peptides in octanol, which is usually believed to mimic the membrane environment, are more representative of their functional state.

## 4.5 Summary

In summary, we could use parameters based on molecular dynamics in octanol to model the activity of antimicrobial peptides on bacteria. The good quantitative results also suggest the utility of peptide dynamics in understanding their biological activity as well as potential for designing novel AMPs.

## Bibliography

- [1] H. Jenssen, P. Hamill, and R. E. Hancock, "Peptide antimicrobial agents," *Clinical microbiology reviews*, vol. 19, no. 3, pp. 491–511, 2006.
- [2] K. Radek and R. Gallo, "Antimicrobial peptides: natural effectors of the innate immune system," in *Seminars in immunopathology*, vol. 29, pp. 27–43, Springer, 2007.
- [3] J. L. Fox, "Antimicrobial peptides stage a comeback," 2013.
- [4] J. L. Narayana and J.-Y. Chen, "Antimicrobial peptides: possible anti-infective agents," *Peptides*, vol. 72, pp. 88–94, 2015.

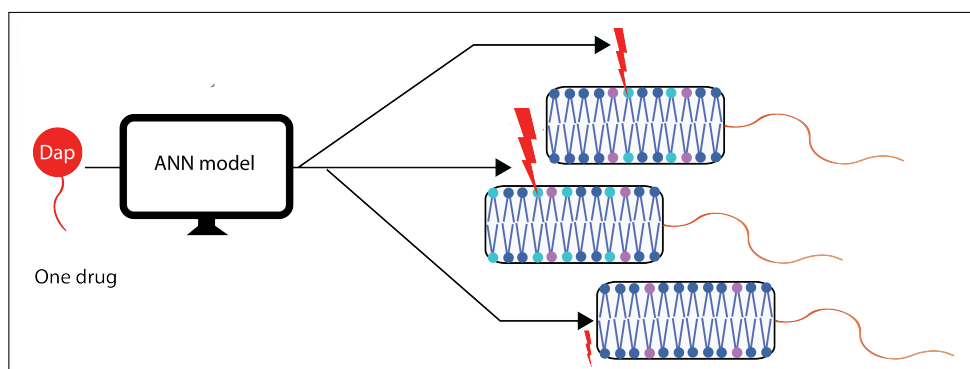


- 
- [5] K. A. Brogden, "Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria?," *Nature reviews microbiology*, vol. 3, no. 3, pp. 238–250, 2005.
- [6] A. A. Bahar and D. Ren, "Antimicrobial peptides," *Pharmaceuticals*, vol. 6, no. 12, pp. 1543–1575, 2013.
- [7] B. P. Mowery, S. E. Lee, D. A. Kissounko, R. F. Epand, R. M. Epand, B. Weisblum, S. S. Stahl, and S. H. Gellman, "Mimicry of antimicrobial host-defense peptides by random copolymers," *Journal of the American Chemical Society*, vol. 129, no. 50, pp. 15474–15476, 2007.
- [8] C. M. Goodman, S. Choi, S. Shandler, and W. F. DeGrado, "Foldamers as versatile frameworks for the design and evolution of function," *Nature chemical biology*, vol. 3, no. 5, pp. 252–262, 2007.
- [9] M. Zasloff, "Antimicrobial peptides of multicellular organisms," *Nature*, vol. 415, no. 6870, pp. 389–395, 2002.
- [10] R. P. MEE, T. R. AUTON, and P. J. MORGAN, "Design of active analogues of a 15-residue peptide using d-optimal design, qsar and a combinatorial search algorithm," *The Journal of peptide research*, vol. 49, no. 1, pp. 89–102, 1997.
- [11] H. Kubinyi, "From narcosis to hyperspace: the history of qsar," *Quantitative Structure-Activity Relationships*, vol. 21, no. 4, pp. 348–356, 2002.
- [12] C. Acharya, A. Coop, J. E Polli, and A. D MacKerell, "Recent advances in ligand-based drug design: relevance and utility of the conformationally sampled pharmacophore approach," *Current computer-aided drug design*, vol. 7, no. 1, pp. 10–22, 2011.
- [13] A. D. Favia, "Theoretical and computational approaches to ligand-based drug discovery," *Front Biosci*, vol. 16, pp. 1276–90, 2011.
- [14] T. Mavromoustakos, S. Durdagi, C. Koukoulitsa, M. Simcic, M. G Papadopoulos, M. Hodoscek, and S. Golic Grdadolnik, "Strategies in the rational drug design," *Current medicinal chemistry*, vol. 18, no. 17, pp. 2517–2530, 2011.
- [15] J. Verma, V. M. Khedkar, and E. C. Coutinho, "3d-qsar in drug design-a review," *Current topics in medicinal chemistry*, vol. 10, no. 1, pp. 95–115, 2010.
- [16] C. Edlund, M. Hedberg, Å. Engström, J.-I. Flock, and D. Wade, "Antianaerobic activity of a cecropin–melittin peptide," *Clinical microbiology and infection*, vol. 4, no. 4, pp. 181–185, 1998.

- [17] H. Oh, M. Hedberg, D. Wade, and C. Edlund, “Activities of synthetic hybrid peptides against anaerobic bacteria: aspects of methodology and stability,” *Antimicrobial Agents and Chemotherapy*, vol. 44, no. 1, pp. 68–72, 2000.
- [18] A. Cherkasov and B. Jankovic, “Application of ‘inductive’ qsar descriptors for quantification of antibacterial activity of cationic polypeptides,” *Molecules*, vol. 9, no. 12, pp. 1034–1052, 2004.
- [19] M. Torrent, D. Andreu, V. M. Nogués, and E. Boix, “Connecting peptide physicochemical and antimicrobial properties by a rational prediction model,” *PloS one*, vol. 6, no. 2, p. e16968, 2011.
- [20] A. Šali and T. L. Blundell, “Comparative protein modelling by satisfaction of spatial restraints,” *Journal of molecular biology*, vol. 234, no. 3, pp. 779–815, 1993.
- [21] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, “Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers,” *SoftwareX*, vol. 1, pp. 19–25, 2015.
- [22] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen, “Evaluation and reparametrization of the opls-aa force field for proteins via comparison with accurate quantum chemical calculations on peptides,” *The Journal of Physical Chemistry B*, vol. 105, no. 28, pp. 6474–6487, 2001.
- [23] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, “Comparison of multiple amber force fields and development of improved protein backbone parameters,” *Proteins: Structure, Function, and Bioinformatics*, vol. 65, no. 3, pp. 712–725, 2006.
- [24] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, “Comparison of simple potential functions for simulating liquid water,” *The Journal of chemical physics*, vol. 79, no. 2, pp. 926–935, 1983.
- [25] W. Humphrey, A. Dalke, and K. Schulten, “Vmd: visual molecular dynamics,” *Journal of molecular graphics*, vol. 14, no. 1, pp. 33–38, 1996.
- [26] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, *et al.*, “Scikit-learn: Machine learning in python,” *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

## Chapter 5

# One drug multiple targets: An approach to predict drug efficacies on bacterial strains differing in membrane composition



### Abstract

The development of drug resistance by bacterial adaptation, or the undesirable damage to gut microbiota raise new challenges in drug design and require one to look beyond the canonical approach of screening multiple drugs against a single target. We develop a complementary approach, one drug multiple targets, and quantitatively illustrate it for a specific case of antibiotics targeting membranes. We curated the data on the activity of daptomycin against *Streptococcus aureus* strains with different membrane compositions, varying mainly in lysylation, and validated the concept. The results of the predictions

---

This chapter is reprinted by permission from ACS: ACS Omega, Prakash *et al.*, 2019 (<https://pubs.acs.org/doi/10.1021/acsomega.8b02862>). Further permission related to the material excerpted should be directed to the ACS.

are good, and within the limits of the scarcely available data, hint at the activity of daptomycin by interacting with the outer leaflet of the membrane.

## 5.1 Introduction

Antibiotic resistance is one of the major health threats in the decades to come. Rational antibiotic design methodologies use structure-activity relationships to predict the activity of the drug candidates against well-defined targets, bacterial proteins or membranes. However, as the knowledge of disease biology and complications of infections increase in the modern times, several newer challenges are emerging in drug design. Many drug candidates are rejected for their toxicity, and others which translate as drugs become redundant as bacteria adapt to them. In the last few decades, several pathogenic bacteria developed resistance to new antibiotics within a few years of their introduction.[1] In what apparently seems unrelated, gut bacteria have been implicated in several important roles in human health and development[2] and a disturbance in their balance can lead to several disorders including Parkinsons' [3]. As such, the effects of commonly used drugs on gut bacteria[4] are being evaluated. Only in the recent times physico-chemical rational intuitions are being used not just for efficacy prediction, but also to understand why most of the drug candidates do not reach advanced stages of clinical trials[5]. The next generation of antibiotic design thus requires one to go beyond the traditional drug design paradigm and address a more comprehensive set of challenges rather than being an efficient drug against a well-defined target.

Computational methodologies such as quantitative structure activity relationships (QSAR) have been traditionally motivated around designing drugs for well defined targets, by binding to the active site,[6, 7] or trapping reaction intermediates[8] or acting allosterically.[9] Several similarity based methods have been developed to predict drug-target interactions using structural or non-structural latent features [10]. Since bacteria have been developing resistance to drugs targeting their enzymes[11, 12], the focus of QSAR has been shifting towards evaluating the activity of cationic antimicrobial peptides (AMP)[13] and their mimics[14] which act on anionic bacterial membranes.

Several mechanisms illustrating the action of AMPs on membrane disruption have been proposed[15, 16]. However, unlike in the case of drugs targeting enzymes, specificity against the membrane of the target bacterium has not been clearly identified. Further, although it is believed that bacteria do not develop resistance to AMPs easily,[17–19] bacteria typically adapt by surface charge reduction using lipid lysylation[20]. It is not immediately apparent how effective the same drug remains when the bacteria adapts by lysylating a fraction of its phospholipids, for example in serial passage experiments.[21] The rational design of AMPs focusing on screening hundreds of drug candidates against

a single bacterial target membrane,[22–24] ignores the potential non-specificity of AMP interactions as well as the plasticity of bacterial genome and thus seems restrictive. Addressing this class of questions needs a complementary strategy of screening drugs against multiple desirable or undesirable targets. Procedurally our approach is the same as a standard QSAR, with a simple shift of focus from considering multiple drugs to multiple targets. We however believe that this simple change in perspective begins a new paradigm in quantitative structure activity predictions for the activity of a drug towards a broader range of targets, membranes in this case. To the best of our knowledge, the systematic data needed to build such model was available only for *Streptococcus aureus* (*S. aureus*). We curated this data on the activity of daptomycin against a variety of *S. aureus* strains from across several pieces of work and used it to develop a quantitative model.

Pathogenic strains of *S. aureus* which can cause severe skin and respiratory infections, develop resistance very fast to new antibiotics, and methicillin resistant *S. aureus* (MRSA) infections are especially problematic due to the lack of a suitable vaccine or antibiotic.[25, 26] Daptomycin is a cyclic lipopeptide antibiotic,[27–29] which binds to the bacterial membranes in a  $\text{Ca}^{2+}$  dependent manner.[30–34] The lipophilic acyl tail of daptomycin interacts with the membrane which then leads to  $\text{K}^+$  leakage and inhibition of protein, DNA or RNA synthesis.[35–37] Daptomycin shows a significant activity against MRSA[38, 39] and vancomycin intermediate *S. aureus*. [30, 40, 41] In this work, with the goal of broadening the scope of a conventional QSAR, we study the activity of daptomycin quantified by its minimum inhibitory concentration (MIC) against *S. aureus* strains characterized by varying compositions of phosphatidylglycerol (PG), cardiolipin (CL), lysyl-PG (LPG).

## 5.2 Results

### 5.2.1 Membrane descriptors

Bacterial membranes are usually composed of lipids such as phosphatidylglycerol (PG), cardiolipin (CL), lysyl-PG (LPG) and zwitterionic phosphatidylethanolamine (PE). We curated this data from 12 different studies [21, 42–52] (**Table D.1** of **Appendix D**), which were mostly obtained using the Etest protocol (AB Biodisk, Dalvagen, Sweden). *S. aureus* does not have PE lipids, and these *S. aureus* strains differing in PG, LPG and CL show varying levels of drug resistance (MIC). The curated data on average reflects that (**Figure D.1** of **Appendix D**) daptomycin binds and oligomerizes in the PG enriched region,[53] as its mechanism of action is believed to be charge driven. The latter might also be the reason for the reduced toxicity to mammalian cells, as they are low in PG

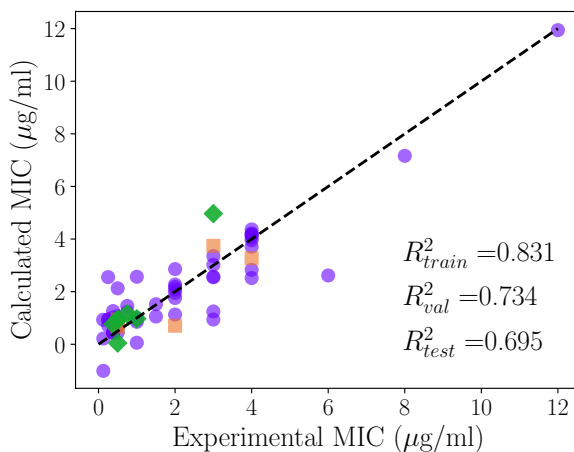
content[15]. CL[54] and LPG[55] on the other hand are negative factors decreasing the activity of daptomycin. However a quantitative relation is non-trivial (**Figure D.1** of **Appendix D**), since it depends on PG, LPG and CL varying simultaneously, and our goal in this work is to develop such relation for both the susceptible ( $\text{MIC} \leq 1 \mu\text{g/ml}$ ) and non-susceptible strains ( $\text{MIC} > 1 \mu\text{g/ml}$ )[38].

### 5.2.2 Membrane bilayer model

Artificial neural network (ANN) model (**Methods** section) was used to obtain a relation for the activity of daptomycin (MIC) on the strains characterized by different membrane compositions (**Methods**). The percentage of different lipids in each strain were the independent variables and the experimentally determined minimum inhibitory concentration (MIC) was the dependent variable in this ANN model. Input and output layers, with one hidden layer with 4 to 10 neurons was used in our model. In order to standardize the procedure, the results presented in this manuscript were trained and tested using the MIC experimentally determined by Etest (**Table D.1** of **Appendix D**). Results obtained from the 10-fold exhaustive cross validation with the total membrane data are given in **Table D.2** and **Figure D.2** of **Appendix D**, and the mean squared error (MSE) and standard deviation (SD) were noted as a function of the number of neurons in the hidden layer **Table B.3**. In our calculations with total membrane composition and MIC from Etest, one hidden layer of 8 neurons with an input and output layer gave the best results. We obtained good results, as indicated by the goodness of fit measure,  $R^2$ , independently for the training, test and validation sets. ( $R^2_{\text{training}} = 0.831$ ,  $R^2_{\text{validation}} = 0.734$ ,  $R^2_{\text{test}} = 0.695$ ). The experimental MIC versus the MIC calculated using the total membrane composition of *S. aureus* is given in **Figure 5.1**. For the sake of completeness, we also predicted the MIC values of the data points which were not measured by Etest method by using the best models obtained from our calculations. All the results are given in **Table D.3** of **Appendix D**, and the results were found to be good.

		Number of neurons in the hidden layer					
		3 neurons	4 neurons	5 neurons	6 neurons	8 neurons	10 neurons
Inner leaflet	$\#R_{test}^2 > 0$	2	7	5	3	4	3
	MSE	64.174	38.183	45.774	47.572	64.688	57.331
	SD	59.622	45.636	67.132	43.895	68.610	79.144
Outer leaflet	$\#R_{test}^2 > 0$			4	6	6	6
	MSE			11.323	20.179	9.042	13.849
	SD			9.670	26.090	9.252	19.915
Total membrane	$\#R_{test}^2 > 0$			8	8	7	7
	MSE			4.396	4.193	3.723	4.994
	SD			5.268	4.744	4.317	4.802

**Table 5.1:** 10-fold cross validation analysis was repeated with a hidden layer between the input and output layers, and by varying the number of neurons in the hidden layer. Here we tabulated the number of times, out of 10,  $R_{test}^2 > 0$ , as well as the mean squared error (MSE) and standard deviation (SD). MSE and SD were used for choosing the optimal number of neurons for each of the membrane compositions studied.

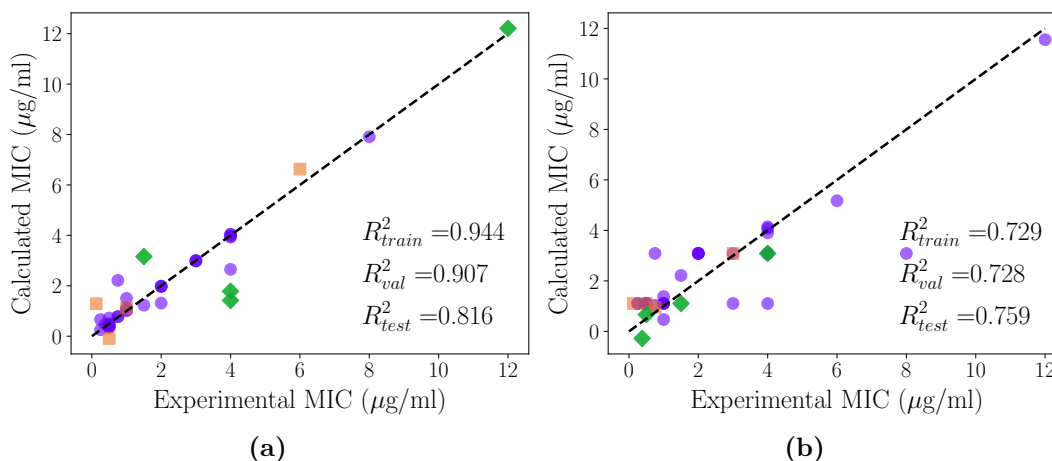


**Figure 5.1:** Comparison of the experimental MIC ( $\mu\text{g}/\text{mL}$ ) and MIC ( $\mu\text{g}/\text{mL}$ ) of daptomycin calculated using 8 neurons in the hidden layer on different total membrane compositions. Training (purple circles), validation (orange squares) and test (green diamonds) sets are shown. The details of the source of the data as well the method by which the MICs were obtained are given in **Table D.1** of **Appendix D**.

### 5.2.3 Inner and outer leaflet models

We derived the lipid compositions in the inner and outer leaflets where the required data was available (**Methods section** and **Table D.4** of **Appendix D**), and performed two independent activity predictions considering these compositions. 10-fold exhaustive cross validations with different number of neurons in the hidden layer were performed

separately using the inner or the outer leaflet composition as the independent variable, and MIC as the dependent variable. As seen from **Table B.3**, calculations with a hidden layer having 8 and 4 neurons gave the best results for the outer and inner leaflet calculations respectively. The results are shown in **Figure 5.2**. The data analysis shows that the test predictions using only the outer leaflet ( $R_{training}^2 = 0.944$ ,  $R_{validation}^2 = 0.907$ ,  $R_{test}^2 = 0.816$ ) were better than those with the inner leaflet compositions ( $R_{training}^2 = 0.729$ ,  $R_{validation}^2 = 0.728$ ,  $R_{test}^2 = 0.759$ ). Results obtained from the 10-fold validation for the outer leaflet composition is given in **Table D.5** and **Figure D.3** of **Appendix D** and for the inner leaflet composition is given in **Table D.6** and **Figure D.4** of **Appendix D**.



**Figure 5.2:** Comparison of the experimental ( $\mu\text{g/mL}$ ) and MIC ( $\mu\text{g/ml}$ ) of daptomycin calculated on different membrane compositions, using the data only from (a) the outer leaflet, with 8 neurons in the hidden layer, or (b) the inner leaflet, with 4 neurons in the hidden layer. Training (purple circles), validation (orange squares) and test (green diamonds) sets are shown. The data used in the analysis is shown in **Table D.4** of **Appendix D**.

## 5.3 Discussion

### 5.3.1 One drug multiple membrane targets

In designing drugs to address antibiotic resistance, newer drugs such as antimicrobial peptide [13] or their mimics[14] which are cationic and effective against anionic bacterial membranes are being developed. As may be expected, bacteria adapt to such drugs, with a surface charge reduction by lysyl modification of the lipids.[20] However, rational design strategies, computational or experimental have focused mainly on designing the activity against a specific target, and effectiveness of the same drug when the bacteria adapts, for



example, by lysylating a fraction of its PG is not clear. From the results of the present work, a direct inference is the efficacy of daptomycin against several strains of *S. aureus* varying in composition, which may be relevant because of the antibiotic resistance. A deeper implication is the scope of addressing the newer challenges arising in rational drug design. The work is an illustration of how when sufficient data is available, the efficacy of a given drug against the adapting pathogen may be established in a rational way. When systematic experimental data is available for training, similar approach can be used to study the effect of any one drug on membranes characterized by differences in lengths, types and saturation levels of lipids.

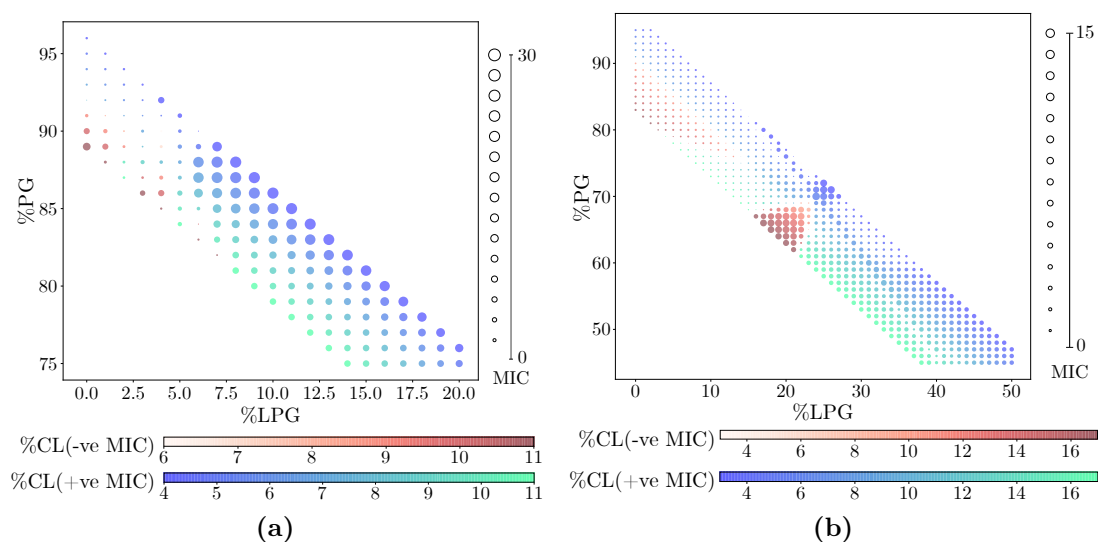
### 5.3.2 Possible mechanism of action

Our calculations suggest that the composition of the total membrane and outer leaflet better predicts the activity of daptomycin on *S. aureus* (**Table B.3**). Although the size of the data is small, the number of data points with the outer or the inner leaflet are both the same, which suggests that the composition in the outer leaflet is a better determinant of the antibiotic activity than the inner leaflet. A possible reason for this could be that the daptomycin activity begins with the charged interactions at the outer leaflet. Earlier experiments have suggested a correlation of the MIC with the LPG in the outer leaflet.[21] However this correlation was based on few data points they obtained in that specific work, and our work demonstrates a similar dependence over a range of membrane compositions where the intuition does not seem trivial. As with the rest of the analysis in this article, the approach is conceptually new and the data available, especially for the individual leaflets, is limited. Of course, these conclusions may have to be re-examined and adapted when more data becomes available.

### 5.3.3 Predicting daptomycin activity

Using the three ANN models we developed for the total membrane, outer and inner leaflets, we calculated the MIC of daptomycin for a systematic variation in the compositions for each of these cases. The results are shown in **Figure 5.3** (and for the inner leaflet in **Figure D.5** of **Appendix D**). In case of total membrane and outer leaflet compositions, MIC value is low when %PG is very high. In the case of outer leaflet composition, for a given value of %PG, MIC increases with increase in LPG content, which is both intuitive and in agreement with the earlier experimental results, as daptomycin activity is believed to be driven by charge interactions. No other trends could be seen in the calculations based on the total membrane composition. But, in case of outer leaflet, when the PG content is below 80%, the MIC values increases with the increase in the %PG. Interestingly, these non-monotonous trends are seen both in the experimental

data (**Figure D.1** of **Appendix D**) as well as prediction from outer leaflet composition (**Figure 5.3**) Physically MIC can not be negative, however our calculations resulted in some negative values, which are artefacts of making predictions with smaller data sets in a landscape with several non-monotonous trends. Continuing our hypothesis that the outer leaflet captures the most relevant interactions, the negative MIC predictions with the outer leaflet data are fewer. The non-monotonous MIC trends in experiments and calculations suggests that there are several ‘local equilibria’ in the membrane compositions that an adapting bacterium may drift towards for improving its resistance (or increasing MIC), depending on the initial conditions or other constraints. A knowledge of how bacteria reach islands of drug resistance with changes in membrane composition may help a drug designer to understand how or how quickly the adaptation may happen.



**Figure 5.3:** Prediction of daptomycin MIC ( $\mu\text{g}/\text{mL}$ ) values for a complete range of variation of lipid composition in the (a) outer leaflet and (b) total membrane. The blue-green color represents the %CL percentage for positive values of MIC and the circle size represents MIC. Since we trained our model on the non-monotonous trends (**Figure D.1** of **Appendix D**), the prediction from the model gives few negative MIC values ( $\mu\text{g}/\text{mL}$ ), represented in red color, over a small parametric region.

## 5.4 Conclusion

To the best of our knowledge a quantitative model that considers the effect of the same drug candidate on multiple membrane compositions had been explored for the first time in this work. Specifically the new approach was used in the present work to characterize the activity of daptomycin on different *S. aureus* strains. With the limited systematic data available, we could build a neural network based model which predicted the activity

and suggested that the composition in the outer leaflet reflects the drug activity better. While practically this ‘one drug multiple targets’ is a trivial extension of the canonical QSAR, conceptually it addresses an entirely new class of problems where the membrane adapts or the drug inadvertently acts on a heterogeneous population of bacteria.

## 5.5 Methods

**Data curation** The experimental data on the membrane phospholipid composition of the different *S. aureus* strains, including methicillin resistant strain and their corresponding minimum inhibitory concentration (MIC) of daptomycin was curated from several published works.[21, 42–52] These works reported the total phospholipid composition - phosphatidylglycerol (PG), lysyl-PG (LPG) and cardiolipin (CL) - in the membrane bilayer (**Table D.1 of Appendix D**). In addition, some works [21, 43, 44, 47–51] also reported the composition of the inner and outer leaflet LPG (referred to as iLPG and oLPG respectively) in the overall composition. The data was then used as follows to obtain the fractions of the phospholipids in the individual leaflets.

Assuming  $2N$  lipid molecules in the bilayer, the composition was used to calculate the number of molecules (PG, LPG and CL) in the inner and outer leaflet. CL is assumed to be equally divided ( $\%CL \times N$ ) between the two layers.  $(\%iLPG \times 2N)$ ,  $(\%PG \times 2N + \%oLPG \times 2N - \%iLPG \times 2N)/4N$ ,  $(\%CL \times N)$  are the number of LPG, PG and CL in the inner leaflet and  $(\%oLPG \times 2N)$ ,  $(\%PG \times 2N + \%iLPG \times 2N - \%oLPG \times 2N)/4N$ ,  $(\%CL \times N)$  are the number of LPG, PG and CL in the outer leaflet. The percentage within each leaflet was then calculated based on these lipid molecule numbers. The phospholipid compositions thus derived for the inner and outer leaflets of the membrane, and the corresponding with daptomycin MIC values are given in **Table D.4 of Appendix D**.

**Dividing data into training and test sets** Artificial neural network (ANN) model was used to obtain a relation between membrane composition and activity. We have performed three separate ANN calculations, using the 66 data points corresponding to total membrane composition, and 38 data points corresponding to the outer and inner leaflet compositions. In all three cases, we tried 10-fold cross validation to establish the robustness of the model. To implement this, we shuffle the data set and divide it into ten different test sets. For each choice of the test set, the rest of the data was randomly split into training and validation sets. 54 and 6 data points were chosen for training and validation respectively with the total membrane. Similarly, 30 and 4 data points were respectively chosen for training and validation with the outer or inner leaflets. The ANN model was based on scikit-learn[56], an open module for machine learning in Python. Logistic function was used as the activation function and low memory BFGS optimization algorithm was used as solver for this neural network.

**Neural Network model** Our ANN model was based on scikit-learn module using MLPRegressor function, had input and output layers, with one hidden layer. The number of neurons in the hidden layer were exhaustively varied from 4 to 10 to determine the optimal number of neurons. With the total membrane composition data, 2500 trial runs were made, using 50 different randomized choices for the input biases in the neural network and 50 different randomized choices for the training and validation set. We used neural network models with an input, output and a hidden layer with 4, 6, 8 or 10 neurons as an additional parameter. We calculated the mean square error (MSE) and standard deviation (SD) for the 10-fold cross validation model. If  $n_i$  is the number of data points present in each test set and  $Y_{exp}$  &  $Y_{predicted}$  are the experimental and predicted value for each data point representing a combination of PG, LPG and CL, and  $CV_{10}^i$  represents the MSE of  $i^{th}$  test set in the 10-fold cross validation model, then MSE is calculated as follows.

$$CV_{10}^i = \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{j,exp} - Y_{j,predicted})^2 \quad (5.1)$$

$$MSE = \frac{1}{10} \sum_{i=1}^{10} CV_{10}^i \quad (5.2)$$

$$SD = SD\{CV_{10}^1, CV_{10}^2, \dots, CV_{10}^{10}\} \quad (5.3)$$

The 10-fold cross validation was used to choose the optimal number of neurons in each of the different membrane composition analyses. Further, for performing detailed predictions, we had to pick one of the 10 data set based models. The best ANN models were selected by screening for two quality criteria,  $R_{training}^2 > 0.7$  and  $R_{validation}^2 > 0.6$ . The same procedure was used when performing the calculations for the 10-fold cross validation with the total membrane composition or with just the inner or outer leaflet compositions. One hidden layer of 8 neurons with an input and output layer gave the best result in our calculation for data corresponding total membrane composition. On the other hand, for the outer leaflet, a model with one hidden layer and 8 neurons gave the best results, while 4 neurons was the optimal choice in modelling the inner leaflet. From this 10-fold calculation, we select the best model for all the three cases, which gives the best  $R^2$  values for all training, validation and test sets. We performed the calculations over the entire range of PG, LPG and CL compositions using this model. The p-values for the variables in the training, validation and test data sets with our neural network model were found to be satisfactory **Table D.7** of **Appendix D**.

## Bibliography

- [1] H. Yoneyama and R. Katsumata, “Antibiotic resistance in bacteria and its future for novel antibiotic development,” *Bioscience, Biotechnology, and Biochemistry*, vol. 70, no. 5, pp. 1060–1075, 2006.
- [2] G. Sharon, T. R. Sampson, D. H. Geschwind, and S. K. Mazmanian, “The central nervous system and the gut microbiome,” *Cell*, vol. 167, no. 4, pp. 915–932, 2016.
- [3] T. R. Sampson, J. W. Debelius, T. Thron, S. Janssen, G. G. Shastri, Z. E. Ilhan, C. Challis, C. E. Schretter, S. Rocha, V. Gradinaru, M.-F. Chesselet, A. Keshavarzian, K. M. Shannon, R. Krajmalnik-Brown, P. Wittung-Stafshede, R. Knight, and S. K. Mazmanian, “Gut microbiota regulate motor deficits and neuroinflammation in a model of parkinson’s disease,” *Cell*, vol. 167, no. 6, pp. 1469–1480, 2016.
- [4] L. Maier, M. Pruteanu, M. Kuhn, G. Zeller, A. Telzerow, E. E. Anderson, A. R. Brochado, K. C. Fernandez, H. Dose, H. Mori, K. R. Patil, P. Bork, and A. Typas, “Extensive impact of non-antibiotic drugs on human gut bacteria,” *Nature*, vol. 555, no. 7698, pp. 623–628, 2018.
- [5] M. J. Waring, J. Arrowsmith, A. R. Leach, P. D. Leeson, S. Mandrell, R. M. Owen, G. Pairaudeau, W. D. Pennie, S. D. Pickett, J. Wang, O. Wallace, and A. Weir, “An analysis of the attrition of drug candidates from four major pharmaceutical companies,” *Nature Reviews Drug Discovery*, vol. 14, pp. 475–486, JUL 2015.
- [6] W.-P. Lu, E. Kincaid, Y. Sun, and M. D. Bauer, “Kinetics of  $\beta$ -lactam interactions with penicillin-susceptible and-resistant penicillin-binding protein 2x proteins from *Streptococcus pneumoniae*. involvement of acylation and deacylation in  $\beta$ -lactam resistance,” *Journal of Biological Chemistry*, vol. 276, no. 34, pp. 31494–31501, 2001.
- [7] C. Goffin and J.-M. Ghuysen, “Multimodular penicillin-binding proteins: an enigmatic family of orthologs and paralogs,” *Microbiology and Molecular Biology Reviews*, vol. 62, no. 4, pp. 1079–1093, 1998.
- [8] M. D. Sintchak and E. Nimmesgern, “The structure of inosine 5’-monophosphate dehydrogenase and the design of novel inhibitors,” *Immunopharmacology*, vol. 47, no. 2-3, pp. 163–184, 2000.
- [9] R. A. Spence, W. M. Kati, K. S. Anderson, and K. A. Johnson, “Mechanism of inhibition of hiv-1 reverse transcriptase by nonnucleoside inhibitors,” *Science*, vol. 267, no. 5200, pp. 988–993, 1995.

- [10] J.-Y. Shi, S.-M. Yiu, Y. Li, H. C. Leung, and F. Y. Chin, “Predicting drug–target interaction for new drugs using enhanced similarity measures and super-target clustering,” *Methods*, vol. 83, pp. 98–104, 2015.
- [11] J. Davies and D. Davies, “Origins and evolution of antibiotic resistance,” *Microbiology and Molecular Biology Reviews*, vol. 74, no. 3, pp. 417–433, 2010.
- [12] R. Benveniste and J. Davies, “Mechanisms of antibiotic resistance in bacteria,” *Annual Review of Biochemistry*, vol. 42, no. 1, pp. 471–506, 1973.
- [13] J. Z. Kubicek-Sutherland, H. Lofton, M. Vestergaard, K. Hjort, H. Ingmer, and D. I. Andersson, “Antimicrobial peptide exposure selects for *Staphylococcus aureus* resistance to human defence peptides,” *Journal of Antimicrobial Chemotherapy*, vol. 72, no. 1, pp. 115–127, 2016.
- [14] X. Vila-Farres, C. G. De La Maria, R. López-Rojas, J. Pachón, E. Giralt, and J. Vila, “In vitro activity of several antimicrobial peptides against colistin-susceptible and colistin-resistant *Acinetobacter baumannii*,” *Clinical Microbiology and Infection*, vol. 18, no. 4, pp. 383–387, 2012.
- [15] M. R. Yeaman and N. Y. Yount, “Mechanisms of antimicrobial peptide action and resistance,” *Pharmacological Reviews*, vol. 55, no. 1, pp. 27–55, 2003.
- [16] T. Ebenhan, O. Gheysens, H. G. Kruger, J. R. Zeevaart, and M. M. Sathekge, “Antimicrobial peptides: their role as infection-selective tracers for molecular imaging,” *BioMed Research International*, vol. 2014, 2014.
- [17] A. Peschel and H.-G. Sahl, “The co-evolution of host cationic antimicrobial peptides and microbial resistance,” *Nature Reviews Microbiology*, vol. 4, no. 7, p. 529, 2006.
- [18] Y. Lai and R. L. Gallo, “Amped up immunity: how antimicrobial peptides have multiple roles in immune defense,” *Trends in Immunology*, vol. 30, no. 3, pp. 131–141, 2009.
- [19] C. D. Fjell, J. A. Hiss, R. E. Hancock, and G. Schneider, “Designing antimicrobial peptides: form follows function,” *Nature Reviews Drug Discovery*, vol. 11, no. 1, p. 37, 2012.
- [20] M. Li, D. J. Cha, Y. Lai, A. E. Villaruz, D. E. Sturdevant, and M. Otto, “The antimicrobial peptide-sensing system ap<sub>6</sub> of *Staphylococcus aureus*,” *Molecular Microbiology*, vol. 66, no. 5, pp. 1136–1147, 2007.

- [21] N. N. Mishra, S.-J. Yang, A. Sawa, A. Rubio, C. C. Nast, M. R. Yeaman, and A. S. Bayer, "Analysis of cell membrane characteristics of in vitro-selected daptomycin-resistant strains of methicillin-resistant *Staphylococcus aureus*," *Antimicrobial Agents and Chemotherapy*, vol. 53, no. 6, pp. 2312–2318, 2009.
- [22] A. Cherkasov, E. N. Muratov, D. Fourches, A. Varnek, I. I. Baskin, M. Cronin, J. Dearden, P. Gramatica, Y. C. Martin, R. Todeschini, V. Consonni, V. E. Kuz'min, R. Cramer, R. Benigni, C. Yang, J. Rathman, L. Terfloth, J. Gasteiger, A. Richard, and A. Tropsha, "Qsar modeling: where have you been? where are you going to?," *Journal of Medicinal Chemistry*, vol. 57, no. 12, pp. 4977–5010, 2014.
- [23] A. Cherkasov and B. Jankovic, "Application of 'inductive' qsar descriptors for quantification of antibacterial activity of cationic polypeptides," *Molecules*, vol. 9, no. 12, pp. 1034–1052, 2004.
- [24] M. Torrent, D. Andreu, V. M. Nogués, and E. Boix, "Connecting peptide physicochemical and antimicrobial properties by a rational prediction model," *PloS one*, vol. 6, no. 2, p. e16968, 2011.
- [25] G. W. Kaatz, T. S. Lundstrom, and S. M. Seo, "Mechanisms of daptomycin resistance in *Staphylococcus aureus*," *International Journal of Antimicrobial Agents*, vol. 28, no. 4, pp. 280–287, 2006.
- [26] J. N. Steenbergen, J. Alder, G. M. Thorne, and F. P. Tally, "Daptomycin: a lipopeptide antibiotic for the treatment of serious gram-positive infections," *Journal of Antimicrobial Chemotherapy*, vol. 55, no. 3, pp. 283–288, 2005.
- [27] R. H. Baltz, "Combinatorial biosynthesis of cyclic lipopeptide antibiotics: a model for synthetic biology to accelerate the evolution of secondary metabolite biosynthetic pathways," *ACS Synthetic Biology*, vol. 3, no. 10, pp. 748–758, 2012.
- [28] M. Debono, M. Barnhart, C. Carrell, J. Hoffman, and R. Hamill, "A21978c, a complex of new acidic peptide antibiotics: factor definition and preliminary chemical characterization," in *Program and Abstracts of 20th Intersci. Conference on Antimicrobial Agents and Chemtherapy*, 1980.
- [29] F. T. Counter, P. W. Ensminger, and L. C. Howard, "A21978c, a complex of new acidic peptide antibiotics: biological activity and toxicity," in *Program and Abstracts of 20th Intersci. Conference on Antimicrobial Agents and Chemtherapy*, 1980.

- [30] C. A. Schriever, C. Fernández, K. A. Rodvold, and L. H. Danziger, “Daptomycin: a novel cyclic lipopeptide antimicrobial,” *American Journal of Health-System Pharmacy*, vol. 62, no. 11, pp. 1145–1158, 2005.
- [31] S. K. Straus and R. E. Hancock, “Mode of action of the new antibiotic for gram-positive pathogens daptomycin: comparison with cationic antimicrobial peptides and lipopeptides,” *Biochimica et Biophysica Acta (BBA)-Biomembranes*, vol. 1758, no. 9, pp. 1215–1223, 2006.
- [32] W. R. Scott, S.-B. Baek, D. Jung, R. E. Hancock, and S. K. Straus, “Nmr structural studies of the antibiotic lipopeptide daptomycin in dhpc micelles,” *Biochimica et Biophysica Acta (BBA)-Biomembranes*, vol. 1768, no. 12, pp. 3116–3126, 2007.
- [33] S. W. Ho, D. Jung, J. R. Calhoun, J. D. Lear, M. Okon, W. R. Scott, R. E. Hancock, and S. K. Straus, “Effect of divalent cations on the structure of the antibiotic daptomycin,” *European Biophysics Journal*, vol. 37, no. 4, pp. 421–433, 2008.
- [34] D. Jung, J. P. Powers, S. K. Straus, and R. E. Hancock, “Lipid-specific binding of the calcium-dependent antibiotic daptomycin leads to changes in lipid polymorphism of model membranes,” *Chemistry and Physics of Lipids*, vol. 154, no. 2, pp. 120–128, 2008.
- [35] J. H. Lakey and M. Ptak, “Fluorescence indicates a calcium-dependent interaction between the lipopeptide antibiotic ly 146032 and phospholipid membranes,” *Biochemistry*, vol. 27, no. 13, pp. 4639–4645, 1988.
- [36] J. A. Silverman, N. G. Perlmutter, and H. M. Shapiro, “Correlation of daptomycin bactericidal activity and membrane depolarization in *Staphylococcus aureus*,” *Antimicrobial Agents and Chemotherapy*, vol. 47, no. 8, pp. 2538–2544, 2003.
- [37] L. J. Wale, A. Shelton, and D. Greenwood, “Scanning electronmicroscopy of *Staphylococcus aureus* and *Enterococcus faecalis* exposed to daptomycin,” *Journal of Medical Microbiology*, vol. 30, no. 1, pp. 45–49, 1989.
- [38] H. W. Boucher and G. Sakoulas, “Perspectives on daptomycin resistance, with emphasis on resistance in *Staphylococcus aureus*,” *Clinical Infectious Diseases*, vol. 45, no. 5, pp. 601–608, 2007.
- [39] J. Joston, C. Grover, C. Downer, T. Pujar, and A. Heidari, “Successful treatment of methicillin-resistant *Staphylococcus aureus* mitral valve endocarditis with sequential linezolid and telavancin monotherapy following daptomycin failure,” *Journal of Antimicrobial Chemotherapy*, vol. 66, no. 9, pp. 2186–2188, 2011.



- [40] G. Sakoulas, “Clinical outcomes with daptomycin: a post-marketing, real-world evaluation,” *Clinical Microbiology and Infection*, vol. 15, pp. 11–16, 2009.
- [41] F. Marco, C. García de la Mària, Y. Armero, E. Amat, D. Soy, A. Moreno, A. del Río, M. Almela, C. A. Mestres, J. M. Gatell, M. T. Jiménez de Anta, and J. M. Miró, “Daptomycin is effective in treatment of experimental endocarditis due to methicillin-resistant and glycopeptide-intermediate *Staphylococcus aureus*,” *Antimicrobial Agents and Chemotherapy*, vol. 52, no. 7, pp. 2538–2543, 2008.
- [42] A. S. Bayer, N. N. Mishra, A. L. Cheung, A. Rubio, and S.-J. Yang, “Dysregulation of *mprF* and *dltABCD* expression among daptomycin-non-susceptible mrsa clinical isolates,” *Journal of Antimicrobial Chemotherapy*, vol. 71, no. 8, pp. 2100–2104, 2016.
- [43] N. N. Mishra, A. S. Bayer, C. Weidenmaier, T. Grau, S. Wanner, S. Stefani, V. Cafiso, T. Bertuccio, M. R. Yeaman, C. C. Nast, and S.-J. Yang, “Phenotypic and genotypic characterization of daptomycin-resistant methicillin-resistant *Staphylococcus aureus* strains: relative roles of *mprF* and *dlt* operons,” *PLoS One*, vol. 9, no. 9, p. e107426, 2014.
- [44] N. N. Mishra and A. S. Bayer, “Correlation of cell membrane lipid profiles with daptomycin resistance in methicillin-resistant *Staphylococcus aureus*,” *Antimicrobial Agents and Chemotherapy*, vol. 57, no. 2, pp. 1082–1085, 2013.
- [45] K.-M. Kang, N. N. Mishra, K. T. Park, G.-Y. Lee, Y. H. Park, A. S. Bayer, and S.-J. Yang, “Phenotypic and genotypic correlates of daptomycin-resistant methicillin-susceptible *Staphylococcus aureus* clinical isolates,” *Journal of Microbiology*, vol. 55, no. 2, pp. 153–159, 2017.
- [46] S.-J. Yang, N. N. Mishra, K.-M. Kang, G.-Y. Lee, J.-H. Park, and A. S. Bayer, “Impact of multiple single-nucleotide polymorphisms within *mprF* on daptomycin resistance in *Staphylococcus aureus*,” *Microbial Drug Resistance*, 2018.
- [47] A. S. Bayer, N. N. Mishra, G. Sakoulas, P. Nonejuie, C. C. Nast, J. Pogliano, K.-T. Chen, S. N. Ellison, M. R. Yeaman, and S.-J. Yang, “Heterogeneity of *mprF* sequences in methicillin-resistant *Staphylococcus aureus* clinical isolates: role in cross-resistance between daptomycin and host defense antimicrobial peptides,” *Antimicrobial Agents and Chemotherapy*, pp. 7462–7467, 2014.
- [48] N. N. Mishra, S.-J. Yang, L. Chen, C. Muller, A. Saleh-Mghir, S. Kuhn, A. Peschel, M. R. Yeaman, C. C. Nast, B. N. Kreiswirth, A.-C. Crémieux, and A. S. Bayer,

- “Emergence of daptomycin resistance in daptomycin-naive rabbits with methicillin-resistant *Staphylococcus aureus* prosthetic joint infection is associated with resistance to host defense cationic peptides and *mprF* polymorphisms,” *PLoS One*, vol. 8, no. 8, p. e71151, 2013.
- [49] N. N. Mishra, G. Y. Liu, M. R. Yeaman, C. C. Nast, R. A. Proctor, J. McKinnell, and A. S. Bayer, “Carotenoid-related alteration of cell membrane fluidity impacts *Staphylococcus aureus* susceptibility to host defense peptides,” *Antimicrobial Agents and Chemotherapy*, vol. 55, no. 2, pp. 526–531, 2011.
- [50] S.-J. Yang, C. C. Nast, N. N. Mishra, M. R. Yeaman, P. D. Fey, and A. S. Bayer, “Cell wall thickening is not a universal accompaniment of the daptomycin nonsusceptibility phenotype in *Staphylococcus aureus*: evidence for multiple resistance mechanisms,” *Antimicrobial Agents and Chemotherapy*, vol. 54, no. 8, pp. 3079–3085, 2010.
- [51] M. H. Murthy, M. E. Olson, R. W. Wickert, P. D. Fey, and Z. Jalali, “Daptomycin non-susceptible methicillin-resistant *Staphylococcus aureus* usa 300 isolate,” *Journal of Medical Microbiology*, vol. 57, no. 8, pp. 1036–1038, 2008.
- [52] V. Cafiso, T. Bertuccio, S. Purrello, F. Campanile, C. Mammina, A. Sartor, A. Raglio, and S. Stefani, “*dltA* overexpression: a strain-independent keystone of daptomycin resistance in methicillin-resistant *Staphylococcus aureus*,” *International Journal of Antimicrobial Agents*, vol. 43, no. 1, pp. 26–31, 2014.
- [53] R. H. Baltz, “Daptomycin: mechanisms of action and resistance, and biosynthetic engineering,” *Current Opinion in Chemical Biology*, vol. 13, no. 2, pp. 144–151, 2009.
- [54] T. Zhang, J. K. Muraih, N. Tishbi, J. Herskovitz, R. L. Victor, J. Silverman, S. Uwumarenogie, S. D. Taylor, M. Palmer, and E. Mintzer, “Cardiolipin prevents membrane translocation and permeabilization by daptomycin,” *Journal of Biological Chemistry*, pp. 11584–11591, 2014.
- [55] T. Jones, M. R. Yeaman, G. Sakoulas, S.-J. Yang, R. A. Proctor, H.-G. Sahl, J. Schrenzel, Y. Q. Xiong, and A. S. Bayer, “Failures in clinical treatment of *Staphylococcus aureus* infection with daptomycin are associated with alterations in surface charge, membrane phospholipid asymmetry, and drug binding,” *Antimicrobial Agents and Chemotherapy*, vol. 52, no. 1, pp. 269–278, 2008.
- [56] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau,

---

M. Brucher, M. Perrot, and É. Duchesnay, “Scikit-learn: Machine learning in python,” *Journal of Machine Learning Research*, vol. 12, no. Oct, pp. 2825–2830, 2011.



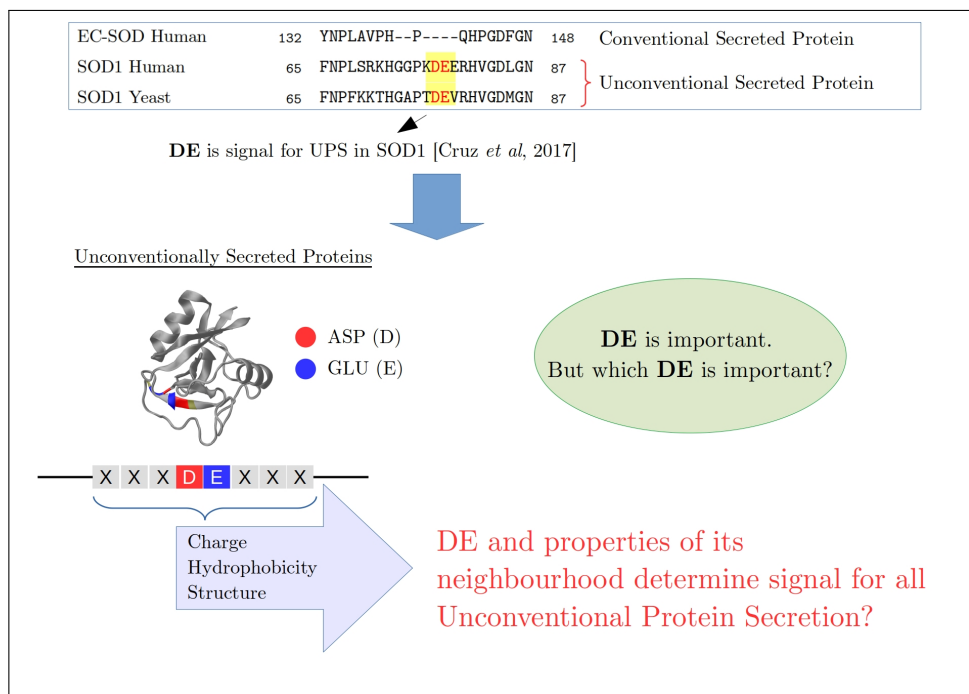
## **PART II**

# **Unconventional protein secretion across membranes**



# Chapter 6

## Exploring the context of diacidic motif DE as a signal for unconventional protein secretion in eukaryotic proteins



### Abstract

Unconventional protein secretion (UPS) is an important phenomenon with fundamental

This chapter is reprinted from "<https://doi.org/10.12688/wellcomeopenres.14914.1>" under the Creative Commons Attribution License.

implications to cargo export. How eukaryotic proteins transported by UPS are recognized without a conventional signal peptide has been an open question. It was recently observed that a diacidic amino acid motif (ASP-GLU or DE) is necessary for the secretion of superoxide dismutase 1 (SOD1) from yeast under nutrient starvation. Taking cue from this discovery, we explore the hypothesis of whether the diacidic motif DE, which can occur fairly ubiquitously, along with its context, can be a generic signal for unconventional secretion of proteins. Four different contexts were evaluated: a physical context encompassing the structural order and charge signature in the neighbourhood of DE, two signalling contexts reflecting the presence of either a phosphorylatable amino acid ('X' in XDE, DXE, DEX) or an LC3 interacting region (LIR) which can trigger autophagy and a co-evolutionary constraint relative to other amino acids in the protein interpreted by examining sequences across different species. Among the 100 proteins we curated from different physiological or pathological conditions, we observe a pattern in the unconventional secretion of heat shock proteins in the cancer secretome, where DE in an ordered structural region has higher odds of being a UPS signal.

## 6.1 Introduction

Proteins need to be secreted outside the cell for several physiologically important reasons.[1–3] In eukaryotes, proteins with an N-terminal signal peptide are known to get conventionally secreted in a vesicular mode through the endoplasmic reticulum (ER) - Golgi secretory pathway (ER-Golgi-secretory vesicles).[4, 5] However, interestingly, many proteins which lack in such well-defined signal peptides are also secreted, mostly under cellular stress.[6] This important class of unconventional protein secretion (UPS) reflects the cellular response to stressors such as inflammation, nutrient stress, ER stress, mechanical stress, etc., and continues to grow in relevance as many such instances are detected in disease associated with dysfunctional autophagy such as neurodegeneration.[7] Secretion of leaderless proteins (without signal peptides) is intriguing, and this non-canonical export mechanism raises several mechanistic questions on their presumably unique secretory pathways.[6] While at least four different UPS mechanisms have been identified so far,[6, 8] even simpler and fundamental questions on how these leaderless proteins are identified remain open.

Conventional signal peptides are 15-50 amino acid tags (“zip code”) at the N-terminus of the proteins.[4, 5, 9] These signal peptides have a characteristic tripartite structure - positively charged N-region, a hydrophobic H-region, and a polar C-region; which makes it easier for the export machinery in the cells, as well as for the computational models, to sort the secretory from the non-secretory proteins. Several predictive models for identifying these signals have been developed.[10, 11] On the other hand, identifying unconventionally



secreted proteins has not been equally intuitive, as they lack in the pattern of canonical leader sequence.[4, 12] Modelling attempts to predict the unconventional secretion have had limited success so far in deciphering the signatures of non-canonical secretion.[13] Using artificial neural networks, which work by recognizing implicit patterns in protein sequences, unconventional secretory proteins were categorized by properties such as amino acid composition, secondary structure, and disordered regions, but do not explicitly reveal the patterns among the unconventional secretory proteins.[14, 15] A rational basis for how non-classically secreted proteins are identified has been missing thus far.

In a recent landmark work, a motif which drives a protein through the non-canonical secretion pathway was identified for the first time. By systematically comparing the superoxide dismutase 1 (SOD1) from human and mouse cells with their extracellular SOD1 counterparts from human, mouse and yeast cells, a diacidic motif (Asp-Glu or DE) had been identified to be responsible for its unconventional secretion. Of the 6 amino acid insertion in SOD1, compared to its extracellular homologs, UPS was found to be sensitive to the mutation of two amino acids aspartate (D) and glutamate (E).[16] As a first and concrete observation, this finding raises the possibility that DE could be a generic UPS signal motif. However, as the short two amino acid length motif could occur in most proteins, in this study, we nucleate a minimal but rational hypothesis for the role of DE and its structural and charge context to act as a trigger for UPS.

## 6.2 Methods

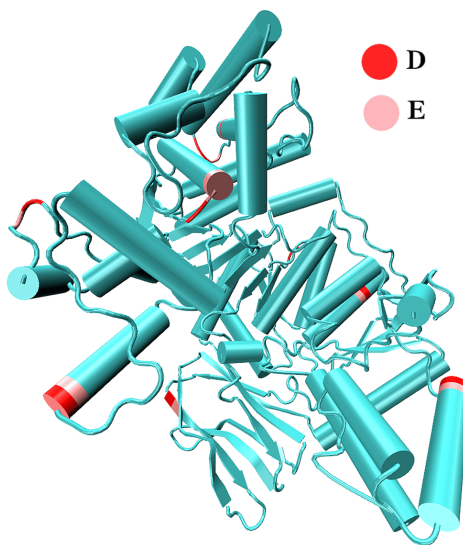
### 6.2.1 Data curation

All the proteins used in this analysis are eukaryotic proteins. Since the hypothesis was centred on the contexts in which DE acts as a UPS trigger, any protein which had no DE motif was discarded from the analysis. The proteins used in the analysis were chosen from four different groups:- Group 1: Since it was recently demonstrated that the presence of DE in UPS of SOD1[16] (cargo involved in neurodegenerative disease, Amyotrophic lateral sclerosis (ALS)), we intended to check the same in all other neurodegeneration causing aggregates such as  $\alpha$ -synuclein,  $\beta$ -amyloid, TDP-43, Tau followed by Group 2 that encompasses all the known UPS cargoes known to date. All these Group 2 secretory proteins were chosen from Keulers et al.,[17] Group 3 were chosen from the dataset of heat shock proteins that are reported to be secreted in an unconventional manner in cancer[18] and the proteins from the breast cancer secretome where UPS is the major contributor.[19] Group 4: While the list of non-secretory proteins could be very large, we did not add any more of them than the number of secretory proteins to keep the data set groups unbiased. Non-secretory proteins (cytoplasmic), were chosen from the

Human Protein Atlas. The complete list of proteins used in the analysis is tabulated in Dataset 1 (<http://www.doi.org/10.17605/OSF.IO/2DMT4>).[20] The final curated set included 100 different eukaryotic proteins, 57 secretory and 43 non-secretory, whose DE sequence details, and the nature of protein are mentioned in **Table E.1** of **Appendix E** and Dataset 1.[20]

### 6.2.2 Variables describing the physical context

For each of the selected proteins, three parameters defining the context in which DE occurs in a protein were defined: (1) Hydrophobicity (H) of the local region, by adding the hydrophobicity score[21] of 3 amino acids on either side of the DE motif (2) Charge (C) of the local region, by summing the charge of 3 amino acids flanking on either side of DE and (3) structural order - the local structure around DE was initially categorically classified as ordered (O) if DE is in  $\alpha$ -helix or  $\beta$ -sheet, disordered (D) if DE is in a loop or disordered region (as illustrated in Figure 6.1). If DE occurred in the border (B) of structured/unstructured regions, given the limited sample size of the data, we considered the option of regrouping B either as O or as D which is discussed below along with other cases.



**Figure 6.1:** Illustration of the definition of structural order in NADP Isocitrate dehydrogenase (PDB:2B0T). The protein has multiple DE motifs, in ordered, loop and border regions.

### 6.2.3 Most relevant DE motif

For every protein used in the analysis, we characterized all DE motifs by their physical descriptors (H,C,O/D/B). DE motif occurs in several places in many protein, and we wanted to identify the most important DE motif for UPS signalling using the analysis

of the curated data. To identify the most relevant DE motif, we generated 16 possible hypotheses listed as cases in **Table E.1** of **Appendix E**. The analyses were performed on all 16 data sets thus created.

#### **6.2.4 Multivariate binary logistic regression**

For each of the 16 hypotheses, different DE motifs were considered to be important, and thus 16 different input files were prepared. Using these different data sets, and the binary prediction (secreted/not-secreted), we performed 16 independent binary logistic regression calculations using scikit-learn module (version 0.19.1) in Python (version 2.7)[22]. Significance quantified by p-values were obtained from these Python analyses. The choice of the hypothesis comes a posteriori, comparing the quality of multivariate binary logistic regression model for the prediction of secretion (**Table E.2** of **Appendix E**).

#### **6.2.5 LIR motif analysis**

We predicted the occurrence of microtubule-associated protein, LC3 interacting motifs in the proteins using the iLIR database[23]. The details of the LIR motifs in the proteins analysed are tabulated in Dataset 1.[20]

#### **6.2.6 Co-evolution analysis**

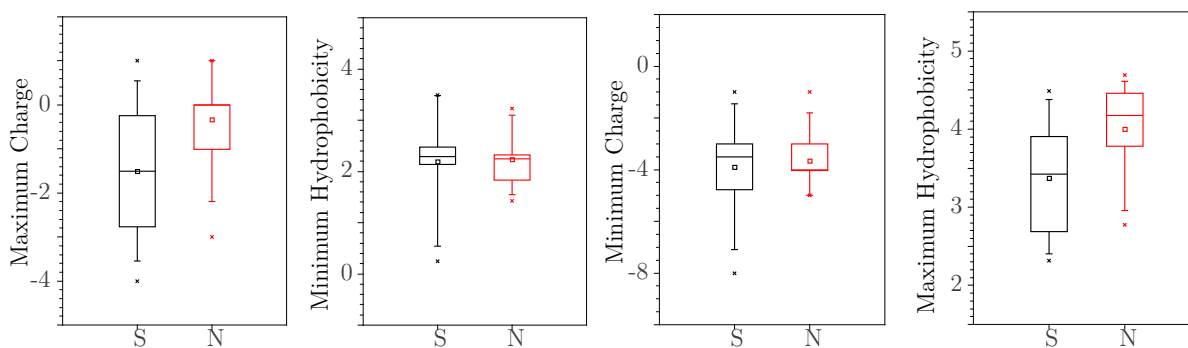
Homologous sequences and their multiple sequence alignments (MSA) for SOD1 and Acb1 proteins were obtained from the Pfam website using their identities Pf00080 and Pf00887 respectively. From the 3904 and 3489 sequences downloaded for the homologs of SOD1 and Acb1, we used a maximum gap frequency of 0.2, and used the selected sequences for co-evolutionary analyses. A consensus sequence was created by choosing the most common amino acid in each position. The entire MSA was then binarized, substituting 1 if the amino acid in a certain position is the same as the one in the consensus sequence, and 0 otherwise. A co-evolution matrix was created with a home-written Python code, following the protocol used in Statistical Coupling Analysis (SCA)[24]. The pairwise co-evolution possibility was used to compare the co-evolutionary couplings different DE motifs have.

## **6.3 Results**

### **6.3.1 Physical nature of the DE neighbourhood as a context**

**Dependency on single variable.** Before building a multivariate predictive model, the curated data was classified as either secretory (UPS) or non-secretory, and their relation to each of the individually chosen variables describing the structural context of where DE

occurs in the protein was examined. The secretory nature has a dependence, although not very strong, on all three physical context variables hydrophobicity (H), charge (C), and (dis)order D/O that we calculated based on the protein sequence and structure data (see Methods). The most significant difference between the physical variables describing secreted and unsecreted proteins was observed in our analysis of the data from the cancer secretome (Figure 6.2). Similar analysis for the other sets are represented in **Figure E.1** of **Appendix E**.



**Figure 6.2:** Dependence of secretory nature on the local properties of DE motif in cancer protein data (Groups 3B, 3C in Dataset 1[20]) on the extreme values of the descriptive parameters. (A) Maximum charge (B) minimum hydrophobicity (C) minimum charge (D) maximum hydrophobicity of the neighbouring amino acids were calculated for all the occurrences of the DE motifs, in each of the proteins appearing in the cancer secretome data.

**Multivariate regression.** Since the three physical parameters showed a relation to the secretory nature, we combined all of them for a multivariate regression model in a binary logistic regression model to predict the odds of secretion. However, there are several DE motifs in most of the proteins, and we developed several hypotheses on how to identify the specific DE which may act as a putative UPS signal. These hypotheses were characterized by the choice of the extremes in the physical parameters used for describing each DE motif. An example of one such hypothesis is a composite assumption that among all the DEs occurring in the protein, DE in the ordered region is more likely to serve as a UPS signal, especially when it is flanked by amino acids with the lowest charge. For the simplicity of the analysis, the border region may be classified as disordered region, making the complete hypothesis as in case 9 in **Table E.1** of **Appendix E**. With 16 different hypotheses (**Table E.1** of **Appendix E**), 16 different data sets were generated which were used for independent binary logistic regressions. The p-values from these analyses, listed in **Table E.2** of **Appendix E**, were used to pick the most plausible hypothesis a posteriori. In each subgroup of data analysis, we

find a few hypotheses which had a p value of  $<0.1$  and possibly even  $<0.05$ . However, one hypothesis that is significant in the total data analysis is that from case 12 (**Table E.2** of **Appendix E**) - that DE in an ordered region with lower charge and higher hydrophobicity context is most relevant for the UPS. Further analysis of the subgroups representing the different pathological conditions suggested a better significance for case 14, but all of them underscored the importance of having DE in the ordered region. In the subgroup comprising the neurodegenerative data set, an ordered region with high hydrophobicity was most relevant.

### 6.3.2 Signalling context via phosphorylation of the motif

The DXE motif in the cytoplasmic domain of the membrane proteins has been implicated as a signal, in a different context, for the cargo transport.[25] It raises questions on whether the same motif could be relevant in other proteins too, especially because DXE has been implicated in the cargo concentration during the protein export from ER. There are several occurrences of DXE or DXXE in our curated data, both in the secretory and non-secretory proteins. Since the unconventional protein secretion is not always active, and it is triggered by cellular stress, we examined the possibility that there is a phosphorylatable amino acid that might act as an on/off signal in the immediate proximity of DE motif. We screened the curated sequences for a motif XDE, DXE, DEX, XED, EXD, EDX where X is one of the following six amino acids that have the propensity for phosphorylation :- three O-phosphorylated amino acids (S, T, Y) or three N-phosphorylated amino acids (H,R,K).[26, 27] The analysis separated the proteins into those which have at least one such motif, and those which do not and studying their relevance for secretion. The contingency tables for the possibility of secretion along with insertions DXE (**Table E.4** of **Appendix E**) or a general proximity of X (**Table E.5** of **Appendix E**) do not suggest any significant conclusions, either for the overall data or for the subgroups considered.

### 6.3.3 Signalling context via LC3 interacting region (LIR) analysis

The LIR is a motif that aids in autophagy.[28] We observed that most of the proteins, whether secretory or non-secretory, in our curated data had an LIR motif and thus the proximity of LIR to the DE motif was used for differentiating the different proteins. We explored the hypothesis that DE motif in the immediate proximity of LIR motif (joined in the primary sequence) triggers UPS. Most secretory proteins had DE and LIR motifs that are discontinuous (Dataset 1[20]). The group I had 4/5, group II with 20/26, group IV - secretory proteins within 7/8 and HSPs with 10/11 discontinuous stretches. No

significant conclusion could be derived either about the presence of LIR or about its proximity to the DE motif.

### 6.3.4 Co-evolutionary context

By using over 3000 homologous sequences (see Methods), pairwise co-evolutionary relationships were constructed among all the amino acids in SOD1 and Acb1. Both these proteins had two DE motifs each, specifically one motif being implicated in UPS (amino acids 77-78 in SOD1 and 23-24 in Acb1).[16] We explored to see if the motifs implicated in UPS signalling have co-evolutionary patterns that are conspicuously different from the other two. D102 in SOD1 had a co-evolutionary relation with three other amino acids (amino acids 58,72,81), but a similar pattern could not be seen in the other DE motifs from SOD1 and Acb1. The identification of the DE signature[16] involved the knowledge of sequences as well an additional information on the protein localization. Thus sequence alignment and coevolution information alone used in the present analysis could not distinguish the DE that acts as a UPS signature.

## 6.4 Discussion

The role of diacidic motif DE as a component of retention or export signals has been noted in several contexts.[29] For example, it had long been known that KDEL or HDEL sequences at the C-terminus act as retention signal to keep the proteins in ER.[29] Further, DXE was identified as the signal for the export of transmembrane proteins such as VSV-G by COPII-coated vesicles from the endoplasmic reticulum to Golgi.[25] Although exporting proteins containing these signals work through conventional secretory pathway, DE signal in SOD1 directs the protein export through the UPS. Thus, it is clear that the same DE motif depending on its flanking residues and insertions can assume a different signalling role.[16] In this work, we examine several eukaryotic proteins to find clues for the patterns leading to UPS.

Since the knowledge of the factors contributing to unconventional secretion is in its nascent stages, the aim of this work was not to predict the possibility of cargo secretion by unconventional pathways but rather to build a hypothesis which may be accepted or rejected as and when more data will be available. The known secreted cargoes that take the UPS route are limited. In this direction, we curated available data which could allow us to make comparisons between the secretory and non-secretory proteins. Towards this, we gathered data from different pathological conditions - cancer secretome, neurodegeneration, and cellular stress. In terms of analysis, we examined the occurrence of DE motif in four different contexts, namely in ordered regions such as helix or sheets, in disordered regions such as loops, or alternatively in the border region between the ordered

and disordered regions. We also considered the possible post-translational modifications of the “X” in the DXE motif, co-evolution of DE motif in Acb1 and SOD1, and charge, hydrophobicity and order of DE motif. Specifically, the physical context defining DE was further dissected into 16 different cases to build a hypothesis to identify the most relevant DE among all that were present in the proteins analyzed.

Our analyses suggest that when the diacidic motif DE appears in an ordered structural environment with lower charge and higher hydrophobicity, it is likely to increase the odds of unconventional protein secretion. While the overall p-values for the binary logistic regression predictions were significant in some of the cases, the dependence on some of the individual variables used was not significant. One of the reasons for this observation might be the limited sample size of UPS cargoes. There are two ways to validate the hypothesis. The first is to curate data on more eukaryotic proteins which certainly go through UPS and ascertain if the statistics improve significantly. Alternatively, since the hypothesis is centred around the structural order and charge on the amino acids flanking DE, SOD1 secretion may be re-examined with mutations which promotes structural order or by increasing the charges, in addition to the substitutions to alanine that have been already explored by the Malhotra group.[16]

Considering the alternative hypothesis proposed in this work that DXE motif could be a signatory sequence for UPS, it is plausible that the post-translational modifications, especially, phosphorylation of the inserted amino acids may act as a signal for UPS. It was observed that among the analysed data, the amino acids with propensity for hydroxyl or amino group phosphorylation (O- or N-phosphorylation) are inserted as DXE in the proteins that undergo UPS. Contrary to this, the DXE motif that was seen in COPII is with aliphatic amino acid insertions.[25] Serine, threonine, tyrosine, as well as arginine,[30] histidine and lysine[28] can undergo phosphorylation under oxidative stress, which is also known to be the external cue that triggers UPS. This hypothesis can be validated by searching for similar patterns in a larger dataset as well as by introducing the DXE motif and observing its chances for unconventional protein secretion.

From the binary logistic regression and odds ratio analysis of the different groups (Figure 2 and Table E.3 of Appendix E), it is evident that the secretion may be favoured by the microenvironment inside and outside the cell in the pathophysiological conditions. For example, altered proteome,[19] intracellular pH,[31] surface charge of the tumour cells,[32] etc., may allow proteins with buried DXE motif to be exposed and accessible for the UPS machinery as has been suggested for the secretion of Acb1 and SOD1.[16] Our analysis that the odds of secretion is significant in terms of minimum charge and maximum hydrophobicity (**Table E.3** and **Table E.6** of **Appendix E**) is in agreement with the propensity of UPS cargoes to be efficiently secreted in an altered microenvironment, as

seen in cancer and neurodegenerative diseases[33] such as Parkinson’s disease, Alzheimer’s disease, Huntington disease and ALS.

We also tested the hypothesis that DE motif in the immediate proximity of LIR motif (joined in the primary sequence) triggers UPS. Most secretory proteins had DE and LIR motifs that are discontinuous (Dataset 1[20]). We did not observe any significance in the proximal position of the LIR motifs with the DE signal. Moreover, it is seen that the sequence alignment and coevolution information alone used in the present analysis could not distinguish the DE that acts as a UPS signature.

## 6.5 Conclusions

In this work, we introduce a hypothesis that the diacidic motif DE, when present in the appropriate context, which may be either structural, evolutionary, or aiding in signalling, acts as a unconventional secretion. Despite the limited availability of data, we find the conditions for the odds of unconventional secretion to be significant. The hypothesis about structural order flanking DE or about the phosphorylation of the insertion DXE or the relevance of the LIR motif that is proximal to the DE motif can be validated either with larger data sets, and/or may be a cue for experimental validation. With this study, we hope to raise testable hypotheses about the recognition of proteins secreted by unconventional pathways, something which remains underexplored as yet.

## Bibliography

- [1] P. Novick, C. Field, and R. Schekman, “Identification of 23 complementation groups required for post-translational events in the yeast secretory pathway,” *Cell*, vol. 21, no. 1, pp. 205–215, 1980.
- [2] J. E. Rothman, “Mechanisms of intracellular protein transport,” *Nature*, vol. 372, no. 6501, pp. 55–63, 1994.
- [3] A. P. Pugsley, “The complete general secretory pathway in gram-negative bacteria,” *Microbiological reviews*, vol. 57, no. 1, pp. 50–108, 1993.
- [4] G. Blobel, “Intracellular protein topogenesis,” *Proceedings of the National Academy of Sciences*, vol. 77, no. 3, pp. 1496–1500, 1980.
- [5] T. A. Rapoport, “Protein transport across the er membrane,” *Trends in biochemical sciences*, vol. 15, no. 9, pp. 355–358, 1990.
- [6] C. Rabouille, “Pathways of unconventional protein secretion,” *Trends in cell biology*, vol. 27, no. 3, pp. 230–240, 2017.



- 
- [7] P. Nilsson, K. Loganathan, M. Sekiguchi, Y. Matsuba, K. Hui, S. Tsubuki, M. Tanaka, N. Iwata, T. Saito, and T. C. Saido, “A $\beta$  secretion and plaque formation depend on autophagy,” *Cell reports*, vol. 5, no. 1, pp. 61–69, 2013.
- [8] C. Rabouille, V. Malhotra, and W. Nickel, “Diversity in unconventional protein secretion,” *Journal of cell science*, vol. 125, no. 22, pp. 5251–5255, 2012.
- [9] K. Nakai, “Protein sorting signals and prediction of subcellular localization,” *Advances in protein chemistry*, vol. 54, pp. 277–344, 2000.
- [10] H. Nielsen, J. Engelbrecht, S. Brunak, and G. V. Heijne, “A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites,” *International journal of neural systems*, vol. 8, no. 05n06, pp. 581–599, 1997.
- [11] K. Hiller, A. Grote, M. Scheer, R. Münch, and D. Jahn, “Predisi: prediction of signal peptides and their cleavage positions,” *Nucleic acids research*, vol. 32, no. suppl\_2, pp. W375–W379, 2004.
- [12] G. Von Heijne, “Patterns of amino acids near signal-sequence cleavage sites,” *European journal of biochemistry*, vol. 133, no. 1, pp. 17–21, 1983.
- [13] A. Lonsdale, M. J. Davis, M. S. Doblin, and A. Bacic, “Better than nothing? limitations of the prediction tool secretomep in the search for leaderless secretory proteins (lsp) in plants,” *Frontiers in plant science*, vol. 7, p. 1451, 2016.
- [14] J. D. Bendtsen, L. J. Jensen, N. Blom, G. Von Heijne, and S. Brunak, “Feature-based prediction of non-classical and leaderless protein secretion,” *Protein Engineering Design and Selection*, vol. 17, no. 4, pp. 349–356, 2004.
- [15] J. D. Bendtsen, L. Kiemer, A. Fausbøll, and S. Brunak, “Non-classical protein secretion in bacteria,” *BMC microbiology*, vol. 5, no. 1, pp. 1–13, 2005.
- [16] D. Cruz-Garcia, N. Brouwers, J. M. Duran, G. Mora, A. J. Curwin, and V. Malhotra, “A diacidic motif determines unconventional secretion of wild-type and als-linked mutant sod1,” *Journal of Cell Biology*, vol. 216, no. 9, pp. 2691–2700, 2017.
- [17] T. G. Keulers, M. B. Schaaf, and K. Rouschop, “Autophagy-dependent secretion: contribution to tumor progression,” *Frontiers in oncology*, vol. 6, p. 251, 2016.
- [18] T. G. Santos, V. R. Martins, and G. N. M. Hajj, “Unconventional secretion of heat shock proteins in cancer,” *International journal of molecular sciences*, vol. 18, no. 5, p. 946, 2017.

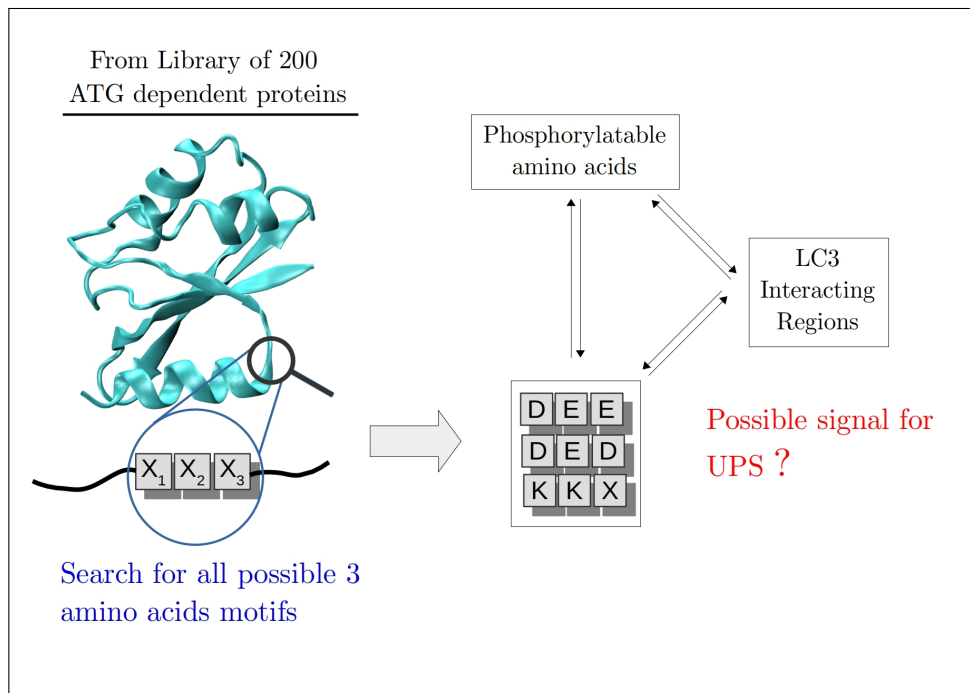
- [19] L. Villarreal, O. Méndez, C. Salvans, J. Gregori, J. Baselga, and J. Villanueva, “Unconventional secretion is a major contributor of cancer cell line secretomes,” *Molecular & Cellular Proteomics*, vol. 12, no. 5, pp. 1046–1060, 2013.
- [20] “<http://www.doi.org/10.17605/osf.io/2dmt4>,”
- [21] R. Cowan and R. G. Whittaker, “Hydrophobicity indices for amino acid residues as determined by high-performance liquid chromatography,” *Peptide research*, vol. 3, no. 2, pp. 75–80, 1990.
- [22] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, *et al.*, “Scikit-learn: Machine learning in python,” *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.
- [23] I. Kalvari, S. Tsompanis, N. C. Mulakkal, R. Osgood, T. Johansen, I. P. Nezis, and V. J. Promponas, “ilir: A web resource for prediction of atg8-family interacting proteins,” *Autophagy*, vol. 10, no. 5, pp. 913–925, 2014.
- [24] S. W. Lockless and R. Ranganathan, “Evolutionarily conserved pathways of energetic connectivity in protein families,” *Science*, vol. 286, no. 5438, pp. 295–299, 1999.
- [25] N. Nishimura, S. Bannykh, S. Slabough, J. Matteson, Y. Altschuler, K. Hahn, and W. E. Balch, “A di-acidic (dx<sub>e</sub>) code directs concentration of cargo during export from the endoplasmic reticulum,” *Journal of Biological Chemistry*, vol. 274, no. 22, pp. 15937–15946, 1999.
- [26] P. Besant, P. Attwood, and M. Piggott, “Focus on phosphoarginine and phospholysine,” *Current Protein and Peptide Science*, vol. 10, no. 6, pp. 536–550, 2009.
- [27] H. R. Matthews, “Protein kinases and phosphatases that act on histidine, lysine, or arginine residues in eukaryotic proteins: a possible regulator of the mitogen-activated protein kinase cascade,” *Pharmacology & therapeutics*, vol. 67, no. 3, pp. 323–350, 1995.
- [28] Á. B. Birgisdóttir, T. Lamark, and T. Johansen, “The lir motif—crucial for selective autophagy,” *Journal of cell science*, vol. 126, no. 15, pp. 3237–3247, 2013.
- [29] M. Capitani and M. Sallese, “The kdel receptor: new functions for an old protein,” *FEBS letters*, vol. 583, no. 23, pp. 3863–3871, 2009.
- [30] J. Fuhrmann, V. Subramanian, D. J. Kojetin, and P. R. Thompson, “Activity-based profiling reveals a regulatory link between oxidative stress and protein arginine phosphorylation,” *Cell chemical biology*, vol. 23, no. 8, pp. 967–977, 2016.

- 
- [31] P. Swietach, R. D. Vaughan-Jones, A. L. Harris, and A. Hulikova, “The chemistry, physiology and pathology of pH in cancer,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 369, no. 1638, p. 20130099, 2014.
- [32] B. Chen, W. Le, Y. Wang, Z. Li, D. Wang, L. Ren, L. Lin, S. Cui, J. J. Hu, Y. Hu, *et al.*, “Targeting negative surface charges of cancer cells by multifunctional nanoprobe,” *Theranostics*, vol. 6, no. 11, p. 1887, 2016.
- [33] A. N. Shrivastava, A. Aperia, R. Melki, and A. Triller, “Physico-pathologic mechanisms involved in neurodegeneration: misfolded protein-plasma membrane interactions,” *Neuron*, vol. 95, no. 1, pp. 33–50, 2017.



# Chapter 7

## Early bioinformatic evidence for a tri-acidic amino acid motif as a possible signature for autophagy dependent unconventional protein secretion



## Abstract

Several proteins are secreted outside the cell, and in many cases they may be identified by a characteristic signal peptide. However, more and more studies point to the evidence for an ‘unconventional’ secretion, where proteins without a hitherto unknown signal are secreted, possibly in conditions of starvation. In this work, we analyse a recently established large data set of 202 RNA binding proteins that are secreted without a signal peptide. Analysis of these proteins secreted by LC3 mediation, the largest unconventionally secreted dataset to our knowledge, identifies the role of KXX motif as well as triacidic amino acid motif, an extension of the recent implication of diacidic amino acids in unconventional secretion. Further data analysis evolves a hypothesis on the sequence or structural proximity of the triacidic or KXX motifs to the LC3 interacting region, and a phosphorylatable amino acid such as serine as a statistically significant feature among these unconventionally secreted proteins. This hypothesis although needs to be validated in experiments that challenge the specific details of each of these aspects, appears to be one of the early steps in defining what may be a plausible signal for unconventional protein secretion.

## 7.1 Introduction

Protein secretion is an essential cellular process. The first step in the translocation of secretory proteins across intracellular membranes and their final localization is the recognition of the ‘address tags’ contained within the amino acid sequences of the proteins. In many cases of protein secretion, a specific configuration of 13 to 36 amino acids in the N-terminal region acts as a ‘signal peptide’, and helps the translocation across the first membrane on the secretory pathway and thus universally controls the entry of all proteins to the secretory pathway in eukaryotes and prokaryotes. In eukaryotes, the signal peptide of a nascent precursor protein (pre-protein) directs the ribosome to the rough endoplasmic reticulum (ER) membrane and initiates the transport of the growing peptide chain across it.[1, 2] In prokaryotes, the signal peptide directs the pre-protein to the cytoplasmic membrane. However, the signal peptide is not responsible for the final destination of the mature protein; secretory proteins devoid of further address tags in their sequence are by default secreted to the external environment. Although signal peptides are not highly conserved, they have a common positively charged n-region, a hydrophobic h-region and a neutral, polar c-region. The c-region contains a weakly conserved cleavage site recognized by membrane-bound signal peptidases. Before the translocation of the pre-protein across the ER membrane, a ribonucleoprotein called signal recognition particle (SRP) binds to the signal peptide emerging from the ribosome.

Then the SRP-signal peptide-ribosome complex binds to the ER membrane via an SRP receptor. The signal peptide is then inserted into the ER membrane via a signal peptide binding protein and the nascent polypeptide then crosses the ER membrane through a transmembrane channel.

The detail in which the signal recognition and secretion process is described above suggests the volumes of research performed on this subject and the clarity on the mechanism of how proteins are ‘conventionally’ secreted. Alternatively, Unconventional protein secretion (UPS) of proteins bypasses the conventional endoplasmic reticulum (ER)-Golgi route. Studies suggest four principal types of UPS that can be further distinguished into non-vesicular and vesicular pathways.[3, 4] The non-vesicular pathways are further classified into Type I (e.g., FGF1) and Type II (e.g., yeast MAT $\alpha$ ). The vesicular pathways are mediated by Type III (e.g., Acb1) and Type IV (e.g., CFTR) mechanisms. Based on a recent classification, Type I is a pore-mediated translocation across the plasma membrane, Type II is an ABC transporter mediated secretion, Type III is an autophagosome/endosome-based secretion and Type IV is a Golgi bypass mechanism.[4] The type III system has a unique feature as the autophagy process has the ability to form de novo vesicles, that have cargo specificity. One such selective form of autophagy that participates in UPS is known as secretory autophagy[5] wherein the cargo is secreted out instead of being degraded.

Unlike the classical secretory proteins that follow the canonical route of secretion (Conventional protein secretion, CPS), the unconventionally secreted protein cargoes follow a plethora of divergent secretory mechanisms. There are no concrete studies on the motif analysis of UPS. Even the signals that may trigger this UPS are not clear. One of the early indications for what may be a possible signal in this fascinating unconventional secretion process, has only recently been discovered. The discovery of the diacidic motif, DE as the signal for UPS[6] along with the context dependence of the presence of this motif in proximity with the charged, unstructured amino acids[7] might provide some clues. With the DE as a potential UPS export signal, the LIR containing proteins possess specific membrane associated receptors and the cells might use this in combination for the type III secretion. This can be resonated with the hypothesis that the UPS cargo containing DE binds to specific binding partner.[8]

Predicting whether or not a protein undergoes a conventional secretion is a relatively well understood phenomenon. Several predictors, such as SecretomeP[9] identify the signal peptide with very high accuracy. There are several other newer predictors such as the OutCyte[10] and ExoPred[11] which are meant to capture the unconventionally secreted proteins as well. These models based on artificial intelligence emphasize the accuracy rather than interpretability in terms of the potential signal-motifs. Further,

the quality of the predictions itself may not be reliable as the models are trained on protein secretion data that is highly inhomogenous. As such, a key to understanding the unconventional secretion signals, and mechanisms is the availability of the relevant high quality data.

Increasing evidence implicates ATG proteins (ATGs) in processes distinct from classical autophagy, such as cellular secretion. Indeed, genetic loss-of-function studies have revealed ATGs are required for the efficient secretion of inflammatory cytokines, extracellular release of bactericidal enzymes and tissue repair factors, EV production and unconventional secretion of proteins lacking amino-terminal leader sequences. The concept of LC3 dependent EV loading and secretion (LDELS) from the secretomic studies has opened up more avenues to ponder upon the autophagy mediated protein secretory cargoes in detail.[12, 13] The recent data on the 202 RNA binding proteins which are unconventionally secreted through an LC3 mediated pathway[13] opens up the possibility of various analyses to understand UPS. We performed bioinformatic analyses on this largest data set till date to explore the possibility of identifying the signals that trigger unconventional secretion.

## 7.2 Methods

**Sequence curation.** The 202 unconventionally secreted set used in the analysis is obtained from the set of proteins proved to be secreted by LC3 mediated mechanism in the analysis of Leidal et al.[13] The set of 1576 conventionally secreted proteins were obtained from the reference set used for training in SecretomeP[9] database ([http://119.3.41.228:8080/SPRomeDB/download\\_enabled.php](http://119.3.41.228:8080/SPRomeDB/download_enabled.php)). For convenience these data sets are also provided in our Supplementary Data set (<https://github.com/malayrb/Thesis/tree/main/Ch7>).

**Discriminatory Motif analysis.** Discriminatory motif (*DiMotif*) analysis[14] of a set relative to the Swiss-Prot reference was performed using the code [https://github.com/ehsanasgari/dimotif/blob/master/notebook/DiMotif\\_step\\_by\\_step\\_example.ipynb](https://github.com/ehsanasgari/dimotif/blob/master/notebook/DiMotif_step_by_step_example.ipynb)

**Motif search.** In addition to the *DiMotif* analysis, we performed a motif search using our script to analyze the differential occurrence of the motifs. The analyses presented in this work are based on 3 amino acid motifs, and which can result in 4200 combinations or 8000 combinations respectively with and without considering the mirror symmetry of the motifs. The proteins from the conventionally secreted and the LC3 mediated groups were scanned for these motif combinations, and the presence or absence of the motif was noted. Similarly, scripts were also used for analysing 4 amino acid motifs as well as the LIR motif (WXXL).[15, 16]



## 7.3 Results

### 7.3.1 Acidic motifs top the differential motif analysis

The presence of a signature signal sequence is common in protein sorting. To identify the presence of a signal sequence in the 202 RNA binding proteins, we performed two different analyses:

#### Discriminatory motif analysis

To capture the unique signature in the 202 LC3 interacting proteins (UPS-ATG dataset), they were compared against 20,117 proteins from the Swiss-Prot database using the *DiMotif* server.[14] This discriminatory motif analysis is meant to identify the motifs which were significantly represented in a chosen set, relative to the all proteins from the Swiss-Prot database. The most significant motifs identified by this analysis are (details shown in **Table 7.1(a)**): EEE, DD, DED, DE, AK, KKE, KK, KT, AKK, KE. These discriminatory motifs have two or three amino acids.

**Table 7.1**

**Table 7.1(a):** Discriminatory motif analysis of the UPS-ATG set relative to the Swiss-Prot database. The E-value denotes the significance level of the hypothesis that the sequence appears selectively in the UPS-ATG dataset, and with these extremely small values of it, it is practically the same as the p-value.

Top motifs	E-value
EEE	4.47E-034
DD	1.40E-029
DED	9.08E-027
DE	2.14E-026
AK	6.32E-026
KKE	4.56E-024
KK	3.65E-023
KT	4.23E-023
AKK	8.80E-023
KE	9.78E-023
DDD	2.98E-022
EK	8.83E-022
RGRG	1.25E-021

**Table 7.1(b):** A summary of the motifs that occur at least 30% more in the UPS-ATG dataset than in the CPS dataset. The total number of proteins in both the sets as well as the percentage of occurrence in them are both shown.

Motifs	%UPS-ATG (202)	%CPS (1576)
EEE	50.990	13.388
KKS	45.545	11.358
AEK	47.525	13.959
AKK	44.554	11.104
KKR	41.584	8.629
KEL	54.950	22.081
DEE	51.980	19.226
KAL	51.980	19.353
EKL	51.980	19.543
KER	42.079	9.772
EKK	43.564	11.802
EEK	49.505	17.830
EKR	43.069	12.183
AEE	51.980	21.129

### Custom motif analysis

In addition to the above-mentioned analysis from *DiMotif* server, we also performed a custom motif search comparing the LC3 interacting proteins with the database of conventionally secreted proteins used for training the SecretomeP.[9] In this analysis, all possible motifs of 2, 3 and 4 amino acids were combinatorially generated and a systematic search for them was performed in the LC3 interacting UPS-ATG dataset (positive-set), and the conventionally secreted proteins (negative-set).[9] The implicit assumption being that the conventionally secreted proteins are not secreted through the LC3 mediated pathway. Comparing the motifs in the positive and the negative sets, the top ten differentiating proteins were identified after imposing a constraint that the motif must occur in at least 30% more often among the proteins in the positive-set than in the negative-set. The occurrence of a motif in the protein, rather than the number of its occurrences in the same protein, was considered important. In this differential analysis, three amino acid motifs had the highest difference between the two sets, while two or four amino acid motifs did not appear to differentiate the two sets significantly to appear among the top differentiators. The three amino acid motifs with the highest difference between the two sets are (details in **Table 7.1(b)**): EEE, KKS, AEK, AKK, KKR, KEL, DEE, KAL, EKL, KER. As may be seen, most of the differentiating motifs are charged, with the triacidic motif at the top.

#### 7.3.2 Acidic motifs appear in the proximity of LIR motifs

The transport of the specific set of proteins analysed in this work is mediated through the LC3 domains. We identified all the LC3 Interacting Regions (LIR) in each protein by performing a search for WXXL motif,[15, 16] and studied the frequency of occurrence of the different 3 amino acid motifs in the proximity of LIR. As the structural information of these proteins is sparse, we restricted the primary analysis to sequence based proximity and wherever the structural information was available, the structural proximity check was subsequently checked for. The most commonly occurring sequences in the proximity of the LIR regions were: KEL, EEL, ALE, KAL, DEE, EKL, AEE, EEE, EEK (details in **Table 7.2**).

#### 7.3.3 Phosphorylatable amino acids occur preferentially in the UPS proteins

Since the unconventional secretion is usually activated under conditions of stress, we explored the possibility that a post-translational modification may be required for its activation. We searched for the presence of serine, threonine or tyrosine within 3 amino acid positions from the differentiating motifs. For almost all the reference motifs we were

**Table 7.2:** Propensity of occurrence of the various motifs in the proximity of LIR motif in the unconventionally secreted proteins as well as in the conventionally secreted proteins.

Motifs	UPS-ATG (202)		CPS (1576)	
	Number	%	Number	%
KEL	20	9.901	92	5.838
EEL	17	8.416	118	7.487
ALE	16	7.921	54	3.426
KAL	15	7.426	57	3.617
DEE	5	2.475	30	1.904
EKL	19	9.406	91	5.774
AEE	17	8.416	32	2.030
EEE	8	3.960	20	1.269
EEK	12	5.941	25	1.586

analysing, the S/T/Y amino acid in the proximity of the motifs occurred preferentially among the proteins from the positive set (**Table 7.3**).

## 7.4 Discussion

### 7.4.1 Triacidic motif is potentially a signal for UPS

The discriminatory motifs identified relative to the Swiss-Prot database and the conventional protein secretion dataset were re-grouped to identify the common patterns among them. Two major patterns emerge among these three amino acid motifs: triacidic motifs (EEE/DDD/DEE/etc) occurring in 160 of the 202 from the positive set, and 625 of the 1576 in the negative set, and basic motifs (KKX) occurring in 187 of the 202 from the positive set and 796 of the 1576 in the negative set. Considering either of these triacidic or KKK motifs as a signal, the difference in the proportion between the positive and the negative data sets is statistically significant ( $p < 0.0002$  in a Z-test).

Of these two statistically significant observations the triacidic motif, by coincidence, happens to be an extension of the observation of the diacidic motif[6] and our earlier attempt to find the context in which the diacidic motif appears.[7] In fact, a quick reanalysis of the multiple sequence alignments from the homologs of SOD1, Acb1[6] by focusing on the mammalian sequences alone shows that they all have a common triacidic motif (**Figure 7.1**). However, despite the statistical evidence over the 202 proteins for the possibility of KKK as a UPS signal, it is present neither in SOD1, nor in Acb1. Since, there is very limited data on unconventionally secreted proteins, we consider the independent finding of the triacidic motif in an already experimentally validated data set

**Table 7.3:** The occurrence of proximal phosphorylatable amino acids close to motifs appearing in the differential analysis is shown here. The motifs are sorted in the decreasing order of occurrence in the UPS-ATG group.

Motifs	UPS-ATG (202)		CPS(1576)	
	Number	%	Number	%
DEE	83	41.089	213	13.515
EEE	79	39.109	140	8.883
KKS	68	33.663	112	7.107
AKK	65	32.178	95	6.028
KES	64	31.683	155	9.835
KAL	62	30.693	205	13.008
GKK	60	29.703	177	11.231
DDE	59	29.208	169	10.723
EKR	59	29.208	132	8.376
KER	56	27.723	96	6.091
KKP	56	27.723	96	6.091
AEK	55	27.228	141	8.947
EAK	54	26.733	130	8.249
KAV	54	26.733	138	8.756
ASK	51	25.248	115	7.297
KVT	51	25.248	174	11.041
DED	47	23.267	99	6.282
GSK	46	22.772	219	13.896
APK	46	22.772	125	7.931
EGE	39	19.307	116	7.360
DKK	36	17.822	111	7.043

as an evidence in support of our finding. Needless to say, the role of which KKK as well as the other features possibly contributing to the signal, as described below require further computational as well as experimental investigations. Further, the positive set which is derived from LC3 mediated secretion (Leidal et al.).[13] The validity of the triacidic or KKK motifs for other types of unconventional secretion will also require investigation.

### 7.4.2 Phosphorylation may be activating the signal

Unlike conventional secretion, the UPS is activated under conditions of stress, suggesting the possibility that post-translational modifications may play a role in activating the signal. In continuation of the hypothesis that the triacidic or KKK motifs may be the ‘signal’, we explored the possibility that the amino acids S/T/Y in the proximity of responsible for activating this signal. S/T/Y amino acids in the proximity of triacidic motifs appeared in 133 of the 202 proteins from the positive-set, and in 422 of the 1576 proteins from the negative set. Similarly, S/T/Y amino acids in the proximity of KKK

EC-SOD Human	132	YNPLAVPH--P---QHPGDFGN	148				
EC-SOD Mouse	146	YNPMEVPH--P---QHPGDFGN	162	Human	11	EEVRHLKTKPSDEEMLFYGHY	32
EC-SOD Zebrafish	114	YNPQEVPH--P---NHPGDMGN	130	Bovine	11	EEVKHLKTKPADEEMLFYSHY	32
SOD1 Human	65	FNPLSRKHGGPKDEERHVGDLGN	87	Zebrafish	11	EEVQLKAKPTDAEMLELYSLY	32
SOD1 Mouse	65	FNPHSKKHGGPADEERHVGDLGN	87	Fruitfly	11	EKVKSLTKRPSDEFLLQLYALF	32
SOD1 Zebrafish	65	FNPHDKTHGGPTDSVRHVGDLGN	87	Dicty	9	QKVKFTKKPSNDELLESLYGLY	30
SOD1 Fruitfly	63	FNPYGKEHGAPVDENRHLGDLGN	85	Yeast	11	KAVNELPTKPSDELLELYALY	32
SOD1 Yeast	65	FNPFKKTGAPTDEVRHVGDMGN	87				

(a) Multiple sequence alignment from homologos of SOD1 (b) Multiple sequence alignment from homologos of ACB1

**Figure 7.1:** A reanalysis of the multiple sequence alignments obtained from Cruz et al.[6]. A comparison among the mammalian sequences, highlighted in purple, shows a common triacidic motif DEE (highlighted in yellow), rather than a diacidic motif (shown in red colored text) when comparing sequences across all species.

appeared in 170 of the 202 proteins from the positive set, and in 603 of the 1576 proteins from the negative set. The statistical significance of the difference between the two sets remains high ( $p < 0.00001$  in a Z-test).

### 7.4.3 LIR motif in the proximity of triacidic motif is discriminatory

The positive set being analysed here, is about the set of proteins where LC3 conjugation machinery is involved in their secretion.[13] However, the LIR motifs are present in abundance in both the positive and the negative sets, making them non-discriminatory. To go beyond the statistical averages from the 202 proteins, and to obtain fine-grained insights into the role of LIR and the triacidic motifs, we analysed the 31 class I proteins from the positive-set which were secreted in all three replicates in a statistically significant way. Among them, 6 proteins had LIR motif within 3 amino acids of the triacidic motif (**Figure 7.2**) along the sequence. From the remaining proteins, structural information was available for 8 of them and in all of them the LIR region was within a structural proximity (**Figure 7.3**), if not a sequential proximity of XYZ Angstrom from triacidic motifs (**Table 7.4**). Coincidentally, in the cases where the structural proximity between the LIR and the triacidic motifs was not seen, it could be seen with the KXX motifs (**Figure 7.4**), underscoring the possible complementarity between the triacidic and KXX motifs in signalling the UPS.

### 7.4.4 Deriving the hypothesis for the signal for UPS

Given the importance of unconventional protein secretion, it is important to identify the signals that trigger it, if such signals exist. The key to building hypotheses is to

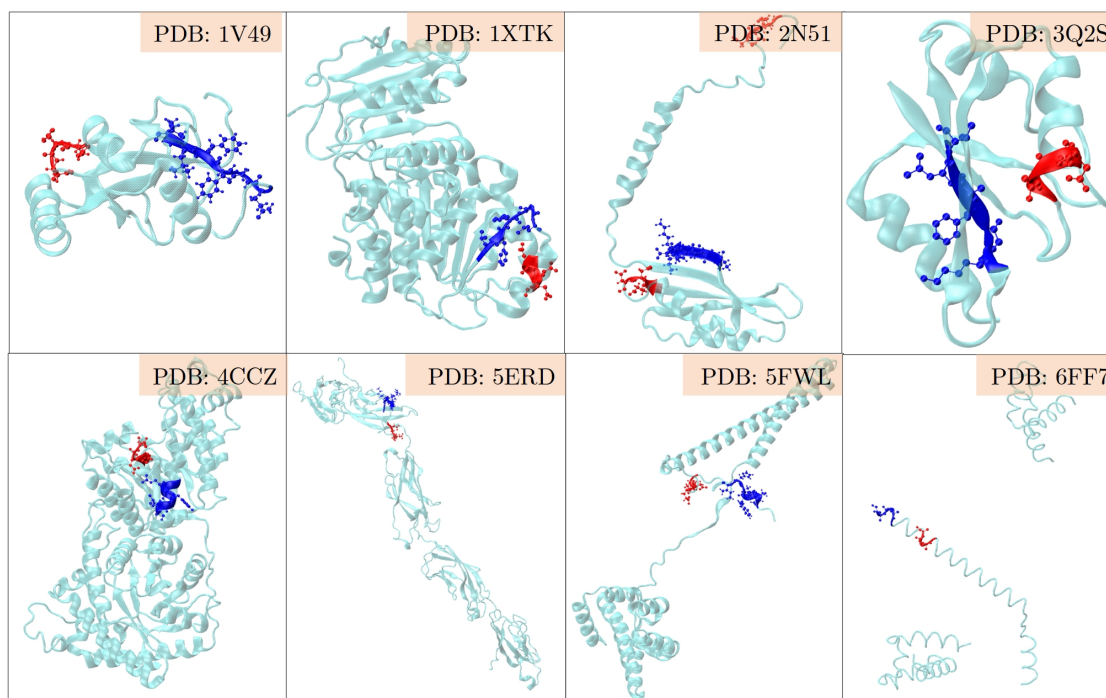
**Table 7.4:** The structural distance between the triacidic motif and the LIR region in some of the class I proteins from the UPS-ATG positive dataset, when there is no sequence proximity, is shown.

Class 1 Protein	Uniprot	PDB	Chain	LIR	Tri-acidic	Distance between LIR and Triacidic Motif (Å)
MAP1LC3B	Q9GZQ8	1V49	A	QAFFLL	DED	11.5
MTR	Q99707	4CCZ	A	RAYHLL	DEE	3
DSG2	Q14126	5ERD	A,B	ARYVKL	DED	8.1
CPSF6	Q16630	3Q2S	C,D	KGFALV	DED	7.1
CDC37	Q16543	5FWL	E	SVWDHI	DDE	6.6
EEF1D	P29692	2N51	A	AAFNKI	EDD	7.3
SF3A1	Q15459	6FF7	EB	VAYAQI	EEE	7.8
DDX39B	Q13838	1XTK	A	VFFGGL	DEE	4.6

IRGQ (Q8WZA9)	...GDSERAAALSPEDETWEVLEEAPPPVFLRPGLPG...	406	441
ECD (O95905)	...HMFVTKFGDNIEDWFIYVIKQITKEFPELVARI...	74	109
SCYL1 (Q96KG9)	...TAEDSSTADRWDEEWGSLQEAEVSLAQDDWSTG...	634	669
STRAP (Q9Y3F4)	...LWQTVVGKTYGLWCVLPEEDSGELAKPKIGFETT...	291	326
SF3B2 (Q15459)	...KTEEEEIDRTPWGELEPSDESSSEEEEEESDEDKP...	697	732
TMPO (P42166)	...SFPFHESILKVIEEWQQVDRQLPSLACKYPVSSRE...	491	526

**Figure 7.2:** An analysis of the sequence-proximity of the triacidic motif with the LIR motif among some of the proteins from the class I of the UPS-ATG data set is shown.

work with highly reliable data, preferably from fewer sources to avoid any biases in the experimental protocols. In this work, we chose to work with a very specific data set from the LC3 machinery driven protein secretion with 202 proteins, and to build a few hypotheses on what may be the signal for the unconventional secretion. The presence of three amino acid motifs, triacidic or KXX, appears recurrently in the set of 202 UPS proteins, significantly more than it occurred either in the conventionally secreted proteins or in the Swiss-Prot database. Although the 202 proteins are believed to be secreted by the LC3 dependent pathway, 5 of these proteins do not have an LIR motif that can interact with the LC3 region. Interestingly even in these proteins, triacidic motif in the proximity of a phosphorylatable amino acid is a common occurrence. Among the proteins that had the LIR motif, it was found mostly in the sequence or a structural

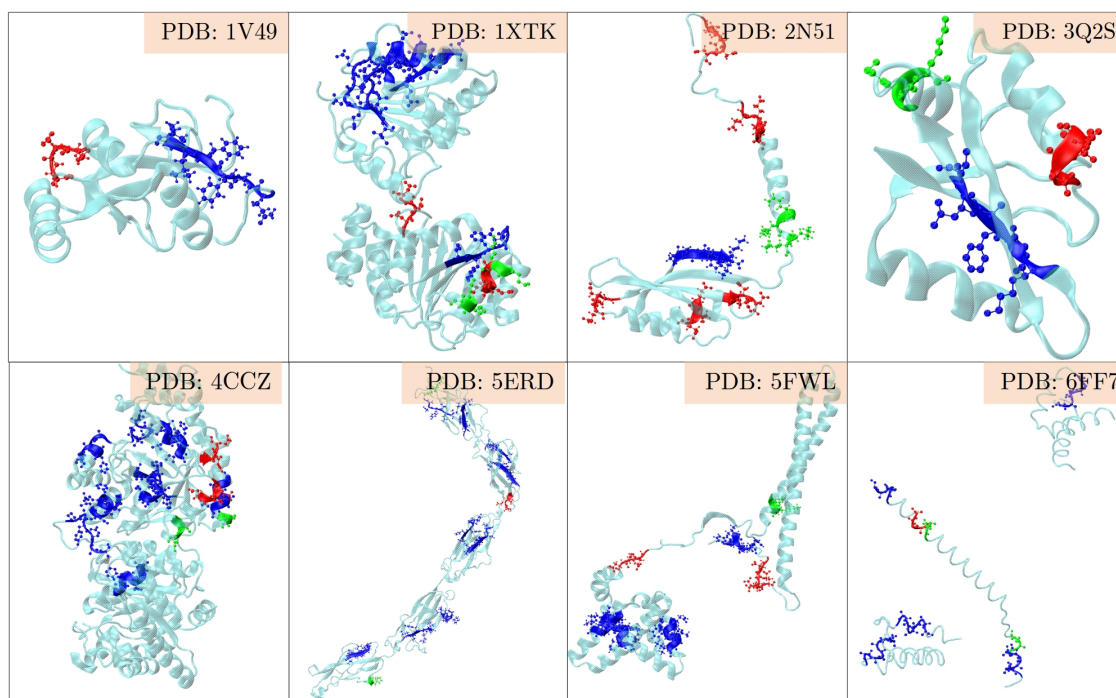


**Figure 7.3:** An analysis of the structural-proximity of the triacidic motif with the LIR motif among the proteins from the class I of the UPS-ATG data set for which structures are known is shown. The blue and red colors indicate the LIR and triacidic motifs. For convenience, only the closest pair is shown and other occurrences of LIR or triacidic motifs are not shown.

proximity from the triacidic or the KXX motifs. Thus, it appears that triacidic or KXX aminoacid motifs in the proximity of LIR and/or phosphorylatable amino acids may play a significant role in triggering the unconventional secretion. This result was also validated in the independently curated dataset of unconventionally secreted proteins from other mammalian cells,[7] where among the 26 mammalian proteins that are secreted unconventionally, 5 of them had triacidic motifs within a 5 amino acid proximity of LIR. 9 of the remaining proteins where there was no sequence proximity, but the structures were available, had LIR motifs within 10 Å of the triacidic motif, and three other structures had them within 15 Å. It will be very interesting to see if this hypothesis can be validated with new experiments in which mutant constructs are designed to challenge each of these aspects of the composite hypothesis – triacidic, KXX, proximity of LIR, proximity of serine amino acid - are developed.

## 7.5 Conclusions

In conclusion, we explored the plausible signals for a very fundamental cellular process - unconventional protein secretion. The field is still in its nascent stages compared to



**Figure 7.4:** An analysis of the structural-proximity of the triacidic as well as the KXX motifs with all LIR motifs occurring among the proteins from the class I of the UPS-ATG data set for which structures are known is shown. The blue, green and red colors indicate the LIR, KXX and triacidic motifs. One may notice that in some structures LIR is close to the triacidic motif, and in others to the KXX motif.

conventional protein secretion where the signals as well as the mechanisms are clearly identified. Exploiting the recent experimental findings of a large set of unconventionally secreted proteins, we could perform bioinformatic analyses as well build hypothesis on the potential role of triacidic amino acids or KXX motif in the proximity of LIR region and phosphorylatable amino acids. As the next steps, we will be exploring collaboration with the relevant experimental groups to validate these hypotheses as well as explore the possibility of deciphering the patterns using interpretable deep-learning methods on the same datasets.

## Supplementary data

The LC3 interacting proteins (positive-set) and the conventionally secreted protein sequences (negative-set) used in the present analysis are all curated at: <https://github.com/malayrb/Thesis/tree/main/Ch7>.



## Bibliography

- [1] A. Devillers-Thiery, T. Kindt, G. Scheele, and G. Blobel, “Homology in amino-terminal sequence of precursors to pancreatic secretory proteins,” *Proceedings of the National Academy of Sciences*, vol. 72, no. 12, pp. 5016–5020, 1975.
- [2] G. von Heijne, “The signal peptide,” *The Journal of membrane biology*, vol. 115, no. 3, pp. 195–201, 1990.
- [3] C. Rabouille, V. Malhotra, and W. Nickel, “Diversity in unconventional protein secretion,” *Journal of cell science*, vol. 125, no. 22, pp. 5251–5255, 2012.
- [4] C. Rabouille, “Pathways of unconventional protein secretion,” *Trends in cell biology*, vol. 27, no. 3, pp. 230–240, 2017.
- [5] S. Jiang, N. Dupont, E. F. Castillo, and V. Deretic, “Secretory versus degradative autophagy: unconventional secretion of inflammatory mediators,” *Journal of innate immunity*, vol. 5, no. 5, pp. 471–479, 2013.
- [6] D. Cruz-Garcia, N. Brouwers, J. M. Duran, G. Mora, A. J. Curwin, and V. Malhotra, “A diacidic motif determines unconventional secretion of wild-type and als-linked mutant *sod1*,” *Journal of Cell Biology*, vol. 216, no. 9, pp. 2691–2700, 2017.
- [7] S. Padmanabhan, M. R. Biswal, R. Manjithaya, and M. K. Prakash, “Exploring the context of diacidic motif de as a signal for unconventional protein secretion in eukaryotic proteins,” *Wellcome open research*, vol. 3, 2018.
- [8] D. Cruz-Garcia, V. Malhotra, and A. J. Curwin, “Unconventional protein secretion triggered by nutrient starvation,” in *Seminars in cell & developmental biology*, vol. 83, pp. 22–28, Elsevier, 2018.
- [9] J. D. Bendtsen, L. J. Jensen, N. Blom, G. Von Heijne, and S. Brunak, “Feature-based prediction of non-classical and leaderless protein secretion,” *Protein Engineering Design and Selection*, vol. 17, no. 4, pp. 349–356, 2004.
- [10] L. Zhao, G. Poschmann, D. Waldera-Lupa, N. Rafiee, M. Kollmann, and K. Stühler, “Outcyte: a novel tool for predicting unconventional protein secretion,” *Scientific reports*, vol. 9, no. 1, pp. 1–9, 2019.
- [11] A. Ras-Carmona, M. Gomez-Perosanz, and P. A. Reche, “Prediction of unconventional protein secretion by exosomes,” *BMC bioinformatics*, vol. 22, no. 1, pp. 1–13, 2021.

- 
- [12] A. M. Leidal and J. Debnath, “Lc3-dependent extracellular vesicle loading and secretion (ldels),” *Autophagy*, vol. 16, no. 6, pp. 1162–1163, 2020.
- [13] A. M. Leidal, H. H. Huang, T. Marsh, T. Solvik, D. Zhang, J. Ye, F. Kai, J. Goldsmith, J. Y. Liu, Y.-H. Huang, *et al.*, “The lc3-conjugation machinery specifies the loading of rna-binding proteins into extracellular vesicles,” *Nature cell biology*, vol. 22, no. 2, pp. 187–199, 2020.
- [14] E. Asgari, A. McHardy, and M. R. K. Mofrad, “Probabilistic variable-length segmentation of protein sequences for discriminative motif mining (dimotif) and sequence embedding (protvecx),” *bioRxiv*, 2018.
- [15] N. N. Noda, Y. Ohsumi, and F. Inagaki, “Atg8-family interacting motif crucial for selective autophagy,” *FEBS letters*, vol. 584, no. 7, pp. 1379–1385, 2010.
- [16] A.-C. Jacomin, S. Samavedam, V. Promponas, and I. P. Nezis, “ilir database: A web resource for lir motif-containing proteins in eukaryotes,” *Autophagy*, vol. 12, no. 10, pp. 1945–1953, 2016.

# Chapter 8

## Conclusions and Future Outlook

In this thesis, I explored aspects of two interesting phenomena that happen across the membrane. Motivated by the new conceptual developments and unfoldings from experiments, and equipped with computational tools such as artificial intelligence, advanced molecular dynamics, I explored the molecular signatures of these phenomena. In the first part of the thesis, I studied various aspects of what makes several novel drug candidates suitable and specific for antibiotic activity against bacterial membranes. In the second part, I performed bioinformatic analyses to search for a potential signal, hitherto unknown, that helps the unconventional protein secretion possible.

### **PART 1: Mechanism of action and design of membrane-active drugs**

*Detailing the Molecular Mechanism of Selective Antibiotic Action Against Bacterial Membranes via Self Assembly and Membrane Penetration.* We used advanced MD simulations to study the action of drug molecules on bacterial membranes. The novelty of the work was the simultaneous consideration of the self-aggregation and penetration of drug molecules on the membrane disruption. Further, it was observed that the activity critically depends on the distance between the cationic functional groups, which defines the “effective length” of the drug molecule, in comparison to the membrane thickness. Using Metadynamics, the drug acting pathway showed how the drug could interact with the membrane then self-aggregating to affect the membrane finally.

*Computational screening of antimicrobial peptides for Acinetobacter baumannii.* Using artificial neural networks, we developed models to screen AMPs for their activity against *A. baumannii*. Although *A. baumannii* is emerging as a global threat, until now, the availability of experimental data in this regard is minimal thus making the computational

efforts important as they can help in prioritizing the experiments from over thousands of possible AMP drug candidates.

*Molecular dynamics based antimicrobial activity descriptors for synthetic cationic peptides.*

We used an artificial neural network and the structural parameters from short MD simulations to understand the biological activity of AMPs. Though performing an MD simulation may appear complicated compared to using sequence properties, it did provide an intuitive meaning to the descriptors as well as served the function of estimating the biological activity.

*One drug multiple targets.* A traditional drug design approach screens several molecules for their activity against one target bacteria. But this approach fails to ask the question about how long the same drug will be effective if the bacteria adapt. In this work, we made an effort to understand the effect of daptomycin against various strains of *S. aureus* with varying degrees of lysine on the phospholipid membranes using an artificial neural network model. Even though this one drug and multiple targets study is a trivial extension of traditional QSAR study, this opens up a new direction of drug development where the drugs can be designed to target one type of bacteria or a specific range of bacteria.

In this part we studied three main important part of a drug development: (a) how drug differentiate between bacteria and human cell, (b) what properties makes a drug effective against bacteria, and (c) can the drug act against a specific type of bacteria taking their membrane composition into account? All these steps are very crucial to understand to develop a drug against bacteria, and we tried to contribute our findings to this ongoing important field of antibiotic resistance.

## **PART 2: Unconventional protein secretion across membranes**

*Exploring the context of diacidic motif DE as a signal for unconventional protein secretion in eukaryotic proteins.*

Unconventional protein secretion is still a very nacent field. Taking a cue from a recent experimental finding where DE is shown to be important for unconventional proteins secretion, we explored how the context of this DE ( the charge, hydrophobicity and structural information) possibly enhances its effectiveness. We found that the odds of secretion is significant in terms of minimum charge, maximum hydrophobicity of the flanking DE and when DE is present in the structural region of the proteins. This hypothesis should be further validated either with a larger data set or experimental validation or structural predictions.

*Early bioinformatic evidence for a tri-acidic amino acid motif as a possible signature for autophagy dependent unconventional protein secretion.* We used the most exhaustive set of unconventionally secreted proteins till date to further expand the search for a possible signal for UPS. We scanned all possible 3-amino acid motifs, and comparing with the conventional secreted protein, we found that the triacidic motif (DEE/EEE/DDD/DED,etc) along with proximal LC3 interacting region (LIR) may be a signal for protein secretion. Also, from discriminatory motif analysis, it was observed that KKK also shows to be significant in the UPS. This hypothesis which is consistent with the diacidic motif discovery and identified in a larger data set of over 200 unconventionally secreted proteins, needs to be examined in experiments.

The two topics chosen in this work, combating the threat from antibiotic resistance and understanding the molecular steps that lead to secretory autophagy, continue to gain importance. Many fascinating discoveries continue offering much more scope for a theoretical and computational investigation.

From the specifics of the details of membrane-disrupting action, general aspects of how the length of the peptide or peptide-like molecules affects the membrane should be derived. This knowledge should also be coupled with the development of other predictive models for quantifying the undesirable effects of these drug candidates on human tissues to avoid potential rejections during the clinical trials. However, the bigger question still lingers. Will the bacteria eventually develop resistance also to membrane-active drugs? Theoretical models which study the mechanisms of membrane modification, and their repercussions on bacterial growth need to be developed to understand this.

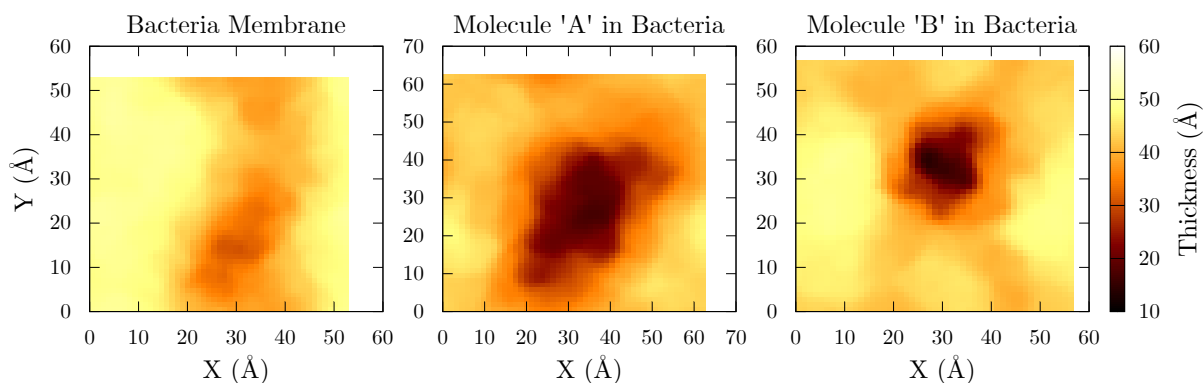
In this work the continuous amino acid motifs, and their structural context were explored to search for the signals or triggers for the unconventional secretion. However, newer deep learning methods may be helpful in identifying patterns that are non-local, and not continuous. This understanding may further be enhanced using structural predictors such as *alphafold 2*.

In summary, membrane separates the internal and the external of the cell and the phenomena that happen across the membrane while maintaining or destroying the cellular activity are extremely critical. The need and scope for detailed investigations continues to remain a challenge and inspires us for further advanced theoretical and computational analyses.



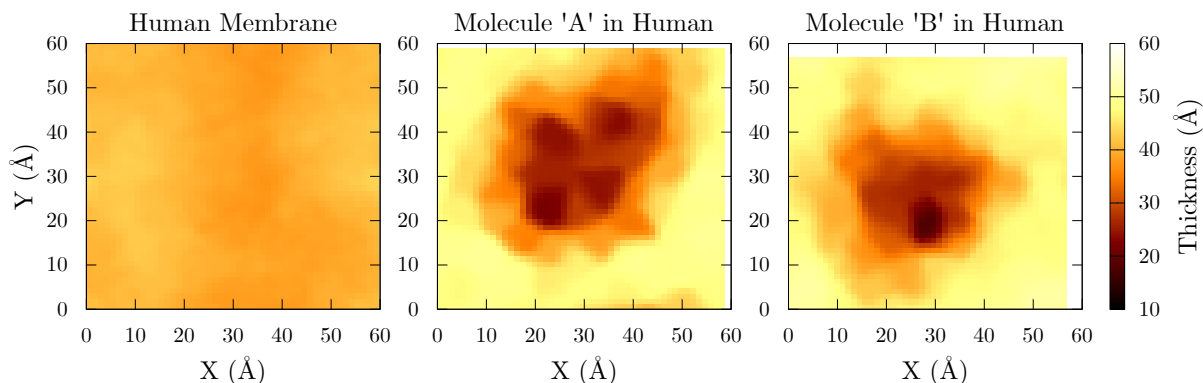
# Appendix A

## Detailing the Molecular Mechanism of Selective Antibiotic Action Against Bacterial Membranes via Self Assembly and Membrane Penetration



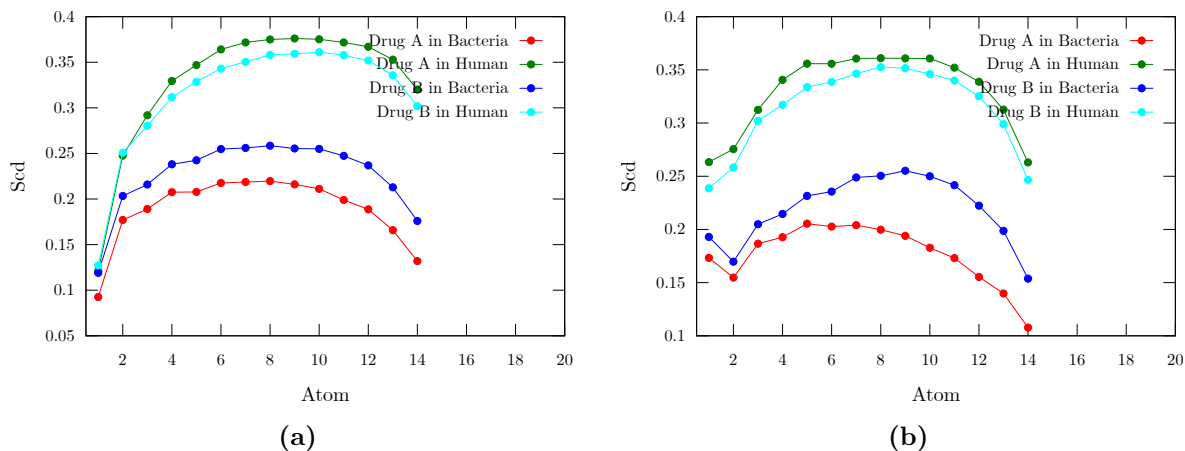
**Figure A.1:** Thickness profile of the membrane (A) Bacterial Membrane (Averaged over last 20ns) (B) Molecule 'A' in bacteria (C) Molecule 'B' in bacteria; All data corresponding to the 7 Å distance between the center of mass of the aggregates for both molecule 'A' and 'B'. The plots are averaged over last 6 ns of the simulation for molecules 'A' and 'B'.

The thickness profile, across the plane of the membrane was analyzed in the presence and absence of the molecules. The plane of the lipid bilayer along the z-axis was divided into bins of  $1 \text{ \AA} \times 1 \text{ \AA}$ , and the distance between the C2 atoms of the upper and lower leaflets was calculated. As depicted in **Figure A.1**, lateral inhomogeneity was observed



**Figure A.2:** Thickness profile of the membrane (A) Human Membrane (Averaged over last 20ns) (B) Molecule ‘A’ in human (C) Molecule ‘B’ in human; All data is corresponding to the 7 Å distance between the center of mass of the aggregates for both molecule ‘A’ and ‘B’. The plots are averaged over last 6 ns of the simulation for molecules ‘A’ and ‘B’.

due to the insertion of the molecules. It is evident from the plots that a pinching of the membrane is taking place owing to the insertion of the molecules. However the pinching is much more localized in the case of molecule ‘B’ than in molecule ‘A’. This suggest that the molecule ‘B’ is more lethal than its other counterpart.



**Figure A.3:** Deuterium order parameter (a) Chain 1 (b) Chain 2; data averaged over 6ns

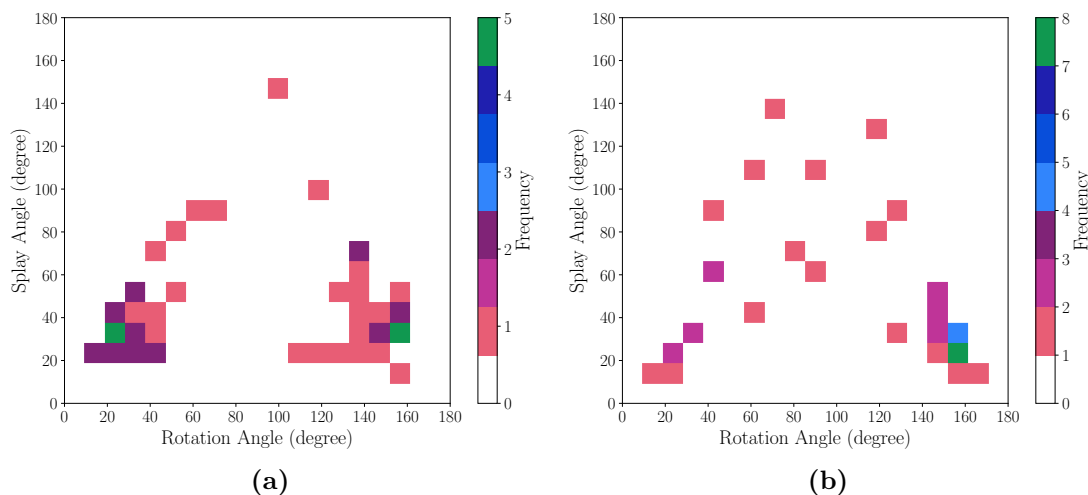
In order to get deeper insights into the lateral inhomogeneity and reduction in the bilayer thickness value, the deuterium order parameter ( $S_{cd}$ ) for the lipid acyl tail was calculated, which is defined as:[1, 2]

$$S_{cd} = \frac{1}{2} \langle 3\cos^2\theta - 1 \rangle \quad (\text{A.1})$$

where  $\theta$  is the angle between a C–H bond and the membrane normal of the lipid. Higher



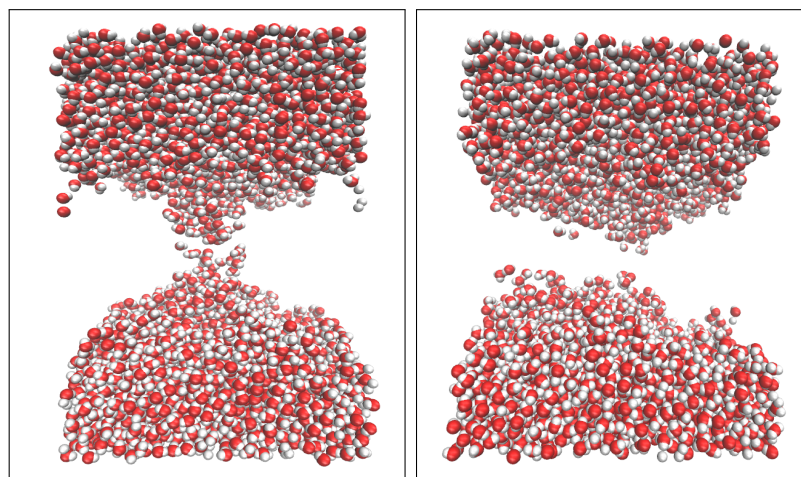
is the  $S_{cd}$  value more is the ordering in the lipid bilayer. The values have been computed for both the chains in the upper and the lower leaflet and are shown in **Figure A.3**. The figure clearly suggests that molecule ‘A’ in bacteria causes maximum reduction in the ordering of the lipid molecules. In addition to this, a rotation angle was defined as the



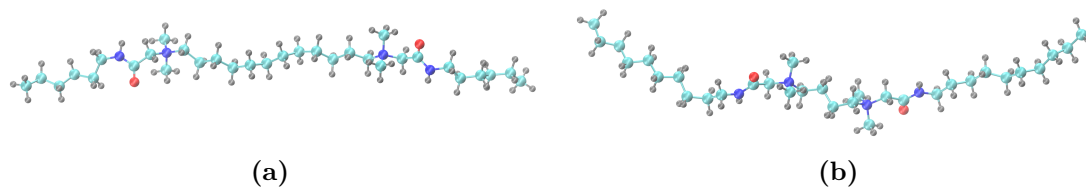
**Figure A.4:** Comparison of the rotation of the lipid molecules to the splay between the lipid acyl chains **(A)** Molecule ‘A’ in bacteria **(B)** Molecule ‘B’ in bacteria; for lipids within 5 Å of the molecules.

angle between the membrane normal and the vector connecting C2 atoms of the lipid head groups and the C216/C316 of the lipid tail. The splay angle was defined as the angle between the two vectors, each connecting the C2 atom of the lipid to the C216 and C316 atoms of the lipid tail. The results are shown in **Figure A.4**. The rotation and the splay angles are more widely distributed for molecule ‘B’. This suggests that the inhomogeneity is higher for the case of molecule ‘B’, suggesting a more impactful destruction of the membrane.

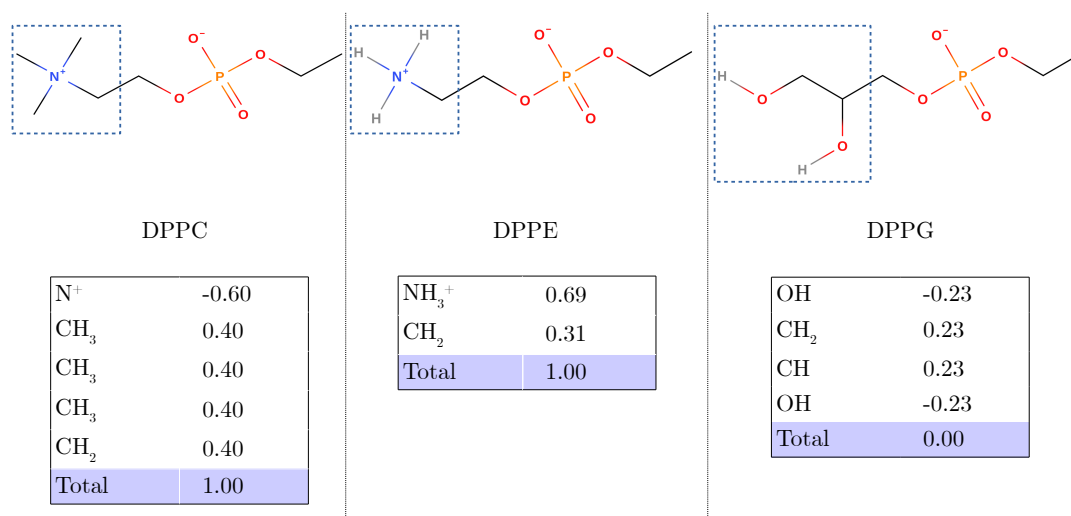
The simulation results depict the formation of a water channel all through the membrane in case of molecule ‘B’ which was absent in the absence of the molecules (**Figure A.5**). However, the effect of molecule ‘B’ is more drastic on the membrane in comparison to molecule ‘A’. The molecules in particular interact with the phosphate group of the membrane, thereby causing a pinching in the membrane. Thus, it is clear that the mechanism of action of these small amphiphilic molecules is different from the conventional antimicrobial peptides.[3]



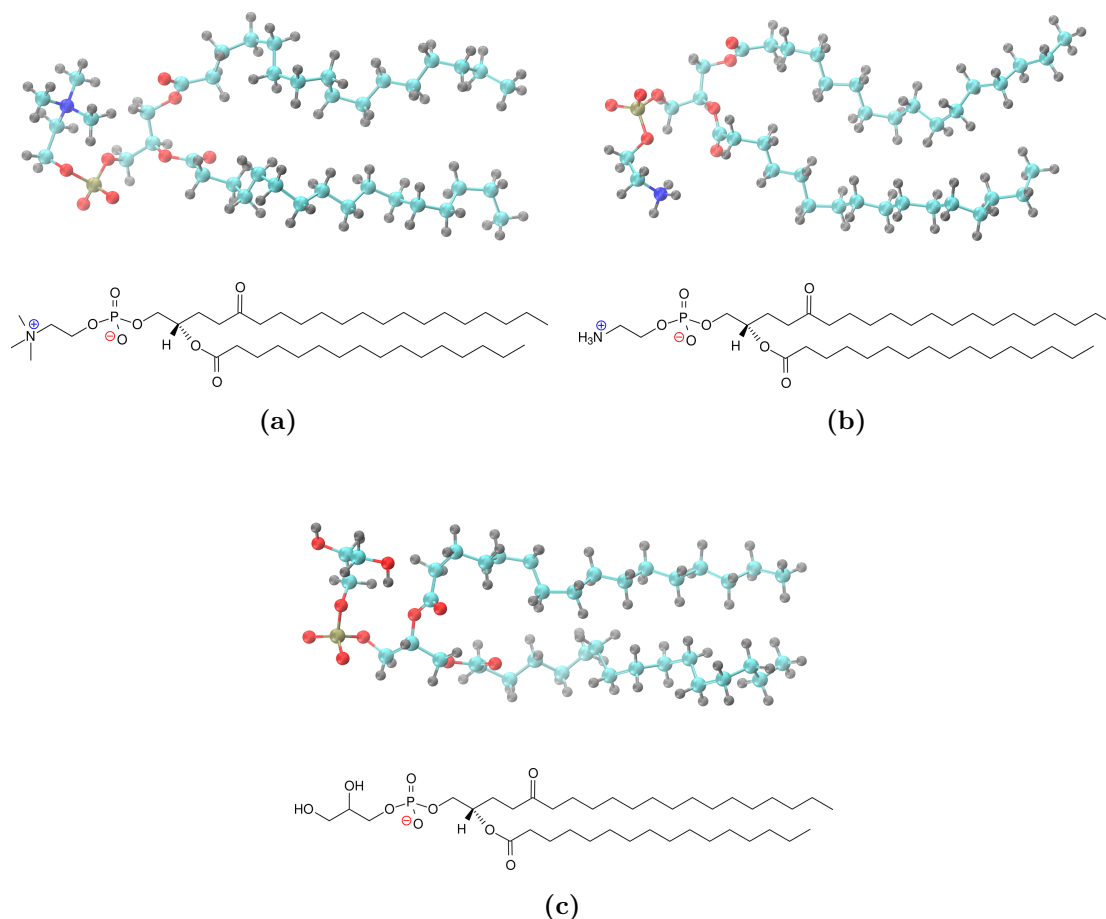
**Figure A.5:** Formation of a water channel across the membrane (A) Molecule ‘B’ in bacteria (B) Molecule ‘A’ in bacteria



**Figure A.6:** 3D structure of (a) Molecule ‘A’ and (b) Molecule ‘B’.



**Figure A.7:** Partial charges on lipid head group shown for DPPC, DPPE and DPPG lipid molecules.



**Figure A.8:** 3D and 2D structure of (a) DPPC, (b) DPPE and (c) DPPG are shown.

## Bibliography

- [1] H. Schindler and J. Seelig, "Deuterium order parameters in relation to thermodynamic properties of a phospholipid bilayer. statistical mechanical interpretation," *Biochemistry*, vol. 14, no. 11, pp. 2283–2287, 1975.
- [2] L. S. Vermeer, B. L. De Groot, V. Réat, A. Milon, and J. Czaplicki, "Acyl chain order parameter profiles in phospholipid bilayers: Computation from molecular dynamics simulations and comparison with 2h nmr experiments," *Eur. Biophys. J.*, vol. 36, no. 8, pp. 919–931, 2007.
- [3] K. Matsuzaki, "Control of cell selectivity of antimicrobial peptides," *Biochim. Biophys. Acta, Biomembr.*, vol. 1788, no. 8, pp. 1687–1692, 2009.



# Appendix B

## Computational screening of antimicrobial peptides for *Acinetobacter baumannii*

**Table B.1:** Details of all the 75 antimicrobial peptides and their experimental protocol used in the present work. Qualitative test set is indicated by blue color, and with a lower bound in MIC.

Sl. No	Peptide	Sequence	MIC ( $\mu\text{g/ml}$ )	Ref.	Methods
1	Bactenecin	RLCRIVVIRVCR	64	[1]	CLSI
2	Cecropin A	KWKLFKKIEKVGQNIRDGIIKAGPAVAVVGQATQIAK	32	[1]	CLSI
3	Cecropin B	KWKIFKKIEKVGRNIRNGIIKAGPAVAVLGEAKAL	32	[1]	CLSI
4	Cecropin P1	SWLSKTAKKLENSAKKRISGIIAIAIQGGPR	1.6	[1]	CLSI
5	Histatin 8	KFHEKHSHRGY	32	[1]	CLSI
6	HNP-1	ACYCRIPACIAGERRYGTCTIYQGRWAFCC	50	[1]	CLSI
7	HNP-2	CYCRIPACIAGERRYGTCTIYQGRWAFCC	50	[1]	CLSI
8	Indolicidin	ILPWKPWWPWR	8	[1]	CLSI
9	Magainin I	GIGKFLHSAGKFGKAFVGEIMKS	64	[1]	CLSI
10	Magainin II	GIGKFLHSAGKFGKAFVGEIMNS	256	[1]	CLSI
11	Mastoparan	INLKALAALAKKIL	4	[1]	CLSI
12	Melittin	GIGAVLKVLTTGLPALISWIKRKRQQ	4	[1]	CLSI
13	$\beta$ -Defensin	DHYNCSVSSGGQCLYSACPIFTKIQTGTCYRGAACKCK	256	[1]	CLSI
14	LysAB2 P0 <sup>1</sup>	NPEKALEPLIAIQIAIKGMLNGWFTGVGFRKR	237.66	[2]	Conlon et.al. [3]
15	LysAB2 P1 <sup>1</sup>	EKALEKLIQKAIKMLNGWFTGVGFRKR	28.48	[2]	Conlon et.al. [3]

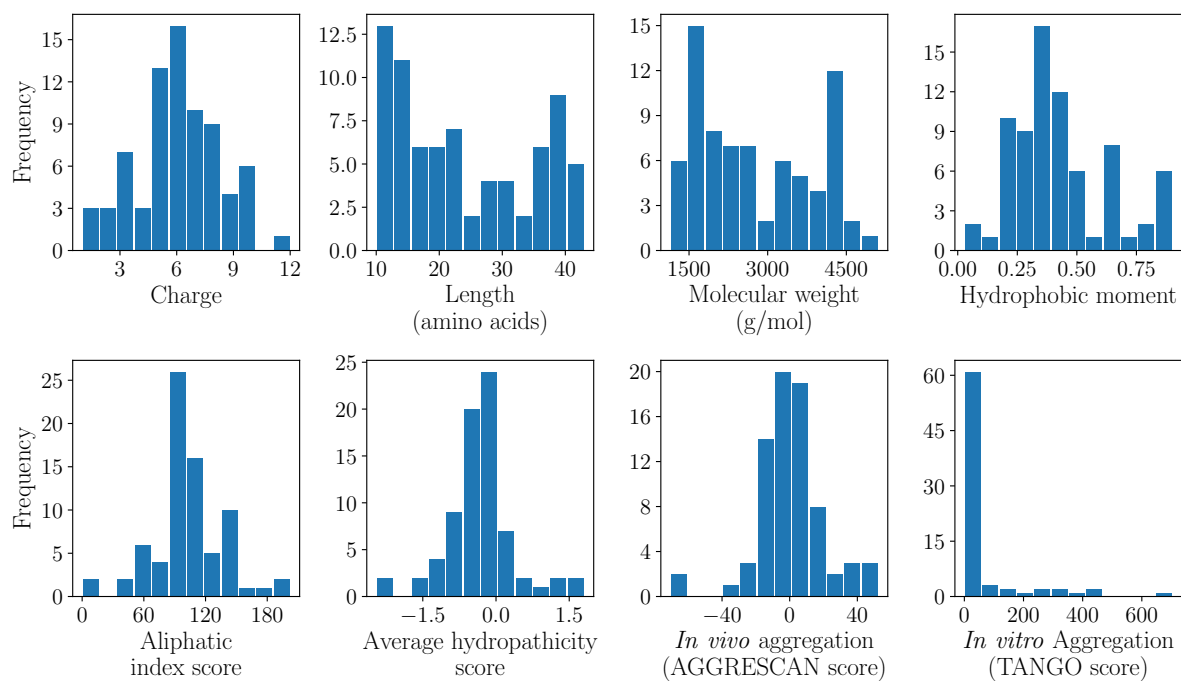
Sl. No	Peptide	Sequence	MIC ( $\mu\text{g/ml}$ )	Ref.	Methods
16	LysAB2 P2 <sup>1</sup>	EKALEKLIQKAIKGLAGWFTGVGARRKR	55.06	[2]	Conlon et.al. [3]
17	LysAB2 P3 <sup>1</sup>	NPEKALEKLIQKAIKGLNGWFTGVGFRKR	30.12	[2]	Conlon et.al. [3]
18	LL-37	LLGDFFRKSKEKIGKEFKRIVQRIKDFLRNLVPRTE	4	[4]	CLSI
19	Aurein 1.2	GLFDIIKKIAESF -NH <sub>2</sub>	16	[4]	CLSI
20	CAMEL	KWKLFKKIGAVLKV -NH <sub>2</sub>	2	[4]	CLSI
21	Citropin 1.1	GLFDVIKKVASVIGGL -NH <sub>2</sub>	16	[4]	CLSI
22	Omiganan	ILRWPWPWRRK -NH <sub>2</sub>	32	[4]	CLSI
23	r-Omiganan	KRRWPWPWRLI -NH <sub>2</sub>	16	[4]	CLSI
24	Pexiganan	GIGKFLKKAKKFGKAFVKILKK -NH <sub>2</sub>	2	[4]	CLSI
25	Temporin A	FLPLIGRVLSGIL -NH <sub>2</sub>	128	[4]	CLSI
26	rr	WLRRIKAWLRR	24.86	[5]	CLSI
27	rr2	WIRRIKKWIRRVHK	3.95	[5]	CLSI
28	rr3	WLRRIKAWLRRKRK	31.46	[5]	CLSI
29	rr4	WLRRIKAWLRRIKA	3.73	[5]	CLSI
30	CLP-19	CRKPTFRRLKWKIKFKFKC	80.36	[6]	CLSI
31	C18G <sup>2</sup>	ALWKKLLKLLKSACKLG -NH <sub>2</sub>	3.87	[7]	Wiegand et.al. [8]
32	C18G-Arg <sup>2</sup>	ALWRRLRLRLRSARRLG -NH <sub>2</sub>	8.48	[7]	Wiegand et.al. [8]
33	NK-2	KILRGVCKKIMRTFLRRISKDILTGGK -NH <sub>2</sub>	2	[9]	CLSI
34	NK27	KILRGVSKKIMRTFLRRISKDILTGGK -NH <sub>2</sub>	2	[9]	CLSI
35	N17	KILRGVSKKIMRTFLRR -NH <sub>2</sub>	8	[9]	CLSI
36	C20	KKIMRTFLRRISKDILTGGK -NH <sub>2</sub>	8	[9]	CLSI
37	C20-DK	KKIMRTFLRRISKKILTGGK -NH <sub>2</sub>	8	[9]	CLSI
38	NK23a	KISKKIMRTFLRRISKDILTGGK -NH <sub>2</sub>	4	[9]	CLSI
39	NK23b	KILRGVSKKIMRRISKDILTGGK -NH <sub>2</sub>	8	[9]	CLSI
40	NK22b	KILGVSKKIMRRISKDILTGGK -NH <sub>2</sub>	16	[9]	CLSI
41	NK23c	KILRGVSKKIMRTFLRRILTGGK -NH <sub>2</sub>	4	[9]	CLSI
42	NK19a	KISKKIMRTFLRRILTGGK -NH <sub>2</sub>	4	[9]	CLSI
43	NK19b	KILRGVSKKIMRRILTGGK -NH <sub>2</sub>	2	[9]	CLSI
44	NK19b-KR	RILRGVSRIMRRILTGR -NH <sub>2</sub>	8	[9]	CLSI
45	NK13	KISKKIMRTFLRR -NH <sub>2</sub>	256	[9]	CLSI
46	Pepcon <sup>2</sup>	FLFSLIPSAIGGLISAFK	37.62	[10]	Wiegand et.al. [8]
47	BP100	KKLFKKILKYL -NH <sub>2</sub>	8.5	[11]	CLSI
48	RW-BP10	RRLFRRILRWL -NH <sub>2</sub>	8.5	[11]	CLSI
49	BP202	KRLFRKILKYL -NH <sub>2</sub>	4.5	[11]	CLSI
50	BP203	KKLFKKILRYL -NH <sub>2</sub>	8.5	[11]	CLSI

Sl. No	Peptide	Sequence	MIC ( $\mu\text{g/ml}$ )	Ref.	Methods
51	H4	KFKKLFKKLSPVIGKEFKRIVERIKRFLR	36.34	[12]	CLSI
52	BR001 <sup>3</sup>	KWKLFFKKIEKVGQNIRDGIIKAGPAVAVVGGATQIAK -NH <sub>2</sub>	10	[13]	Jorgensen et.al.[14]
53	BR002 <sup>3</sup>	GWLKKGKGIKIERVGGHTRDATTIQLGIAQQAANVAATAR -NH <sub>2</sub>	10	[13]	Jorgensen et.al.[14]
54	BR003 <sup>3</sup>	GGLKKGKLEGAGKRVFNAAEKALPVAAGAKALRK	5	[13]	Jorgensen et.al.[14]
55	BR005 <sup>3</sup>	RGFRKHFNKLKVKVHTISETAHVAKDTAVIAGSGAAVVAAT -NH <sub>2</sub>	20	[13]	Jorgensen et.al.[14]
56	BR029 <sup>3</sup>	KWKIFKKIEKAGRNIRDGIIKAGPAVSVVGEAATTYKTG	20	[13]	Jorgensen et.al.[14]
57	BR031 <sup>3</sup>	GWLRFDFGKRIERVGGHTRDATTIQAIGVAQQAANVAATVRG	20	[13]	Jorgensen et.al.[14]
58	BR032 <sup>3</sup>	GWLKKGKGIKIERVGGHTRDATTIQLVGVAAQQAANVPATARG	20	[13]	Jorgensen et.al.[14]
59	BR033 <sup>3</sup>	GWLKKGKGIKIERVGGHTRDATTITIGVAQQAANVAATLKG	20	[13]	Jorgensen et.al.[14]
60	BR034 <sup>3</sup>	GWLKKGKGIKIERVGGHTRDATTIQAIGVAQQAANVAATLKG	20	[13]	Jorgensen et.al.[14]
61	BR035 <sup>3</sup>	GWLKKGKGIKIERVGGHTRDASIQAIQIAQQAANVAATARG	10	[13]	Jorgensen et.al.[14]
62	BR036 <sup>3</sup>	GWLKKGKGIKIERVGGHTRDATTIQLVGVAAQQAANVAATARG	20	[13]	Jorgensen et.al.[14]
63	BR037 <sup>3</sup>	GLVKKKGKGIKIERVGGHTRDASIQAIQIAQQAANVAATARG	20	[13]	Jorgensen et.al.[14]
64	Buforin I	AGRGKQGGKVRKAKTRSSRAGLQFPVGRVHLLRKGNY	>256	[1]	CLSI
65	Histatin 5	DSHAKRHHGYKRFHEKHSHSRGY	>256	[1]	CLSI
66	rr1	WKRRIKIWKIR	>438.06	[5]	CLSI
67	C18G-His	ALWHLHLHLLHSAHHLG -NH <sub>2</sub>	>31.93	[7]	Wiegand et.al.[8]
68	I10	KKIMRTFLRR -NH <sub>2</sub>	>128	[9]	CLSI
69	NK15	KILRGVSKRILTGKK -NH <sub>2</sub>	>128	[9]	CLSI
70	NK14	KILGVSKRILTGKK -NH <sub>2</sub>	>256	[9]	CLSI
71	NK11	KISKRILTGKK -NH <sub>2</sub>	>256	[9]	CLSI
72	NK10	ISKRILTGKK -NH <sub>2</sub>	>256	[9]	CLSI
73	BR030 <sup>3</sup>	KWKFYKKIERVGGQNIRDGIIKAGPAVQVGGQPPRYIKENRFYS	>20	[13]	Jorgensen et.al.[14]
74	BR043 <sup>3</sup>	AGFRKRFNKLKVKVHTIKETANVSKDVAIVAGSGVAVGAAMG	>20	[13]	Jorgensen et.al.[14]
75	BR044 <sup>3</sup>	GFRKRFNKLKVKVHTIKETANVSKDVAIVAGSGVAVGAAMG	>20	[13]	Jorgensen et.al.[14]

<sup>1</sup>These MIC values are calculated according to CLSI guideline.

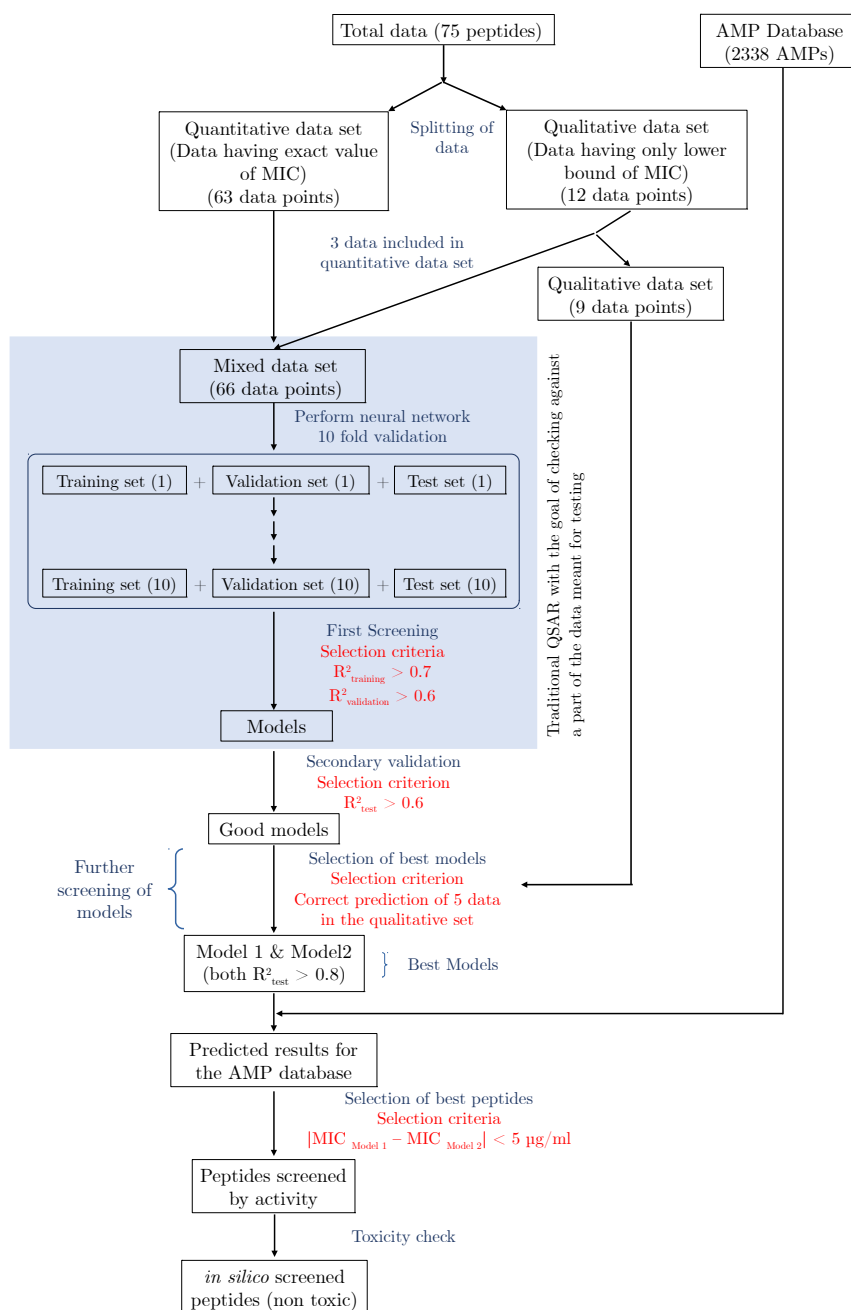
<sup>2</sup>These MIC values are calculated according to CLSI and EUCAST guideline.

<sup>3</sup>MIC values obtained from this procedure is comparable with the MIC values obtained according to CLSI guideline.



**Figure B.1:** Histogram of all the parameters corresponding to the AMPs shown in **Table B.1**. The raw data of these parameters is in **Data File in GitHub** ([https://github.com/malayrb/Thesis/blob/main/Ch3/File\\_S2.xlsx](https://github.com/malayrb/Thesis/blob/main/Ch3/File_S2.xlsx)).





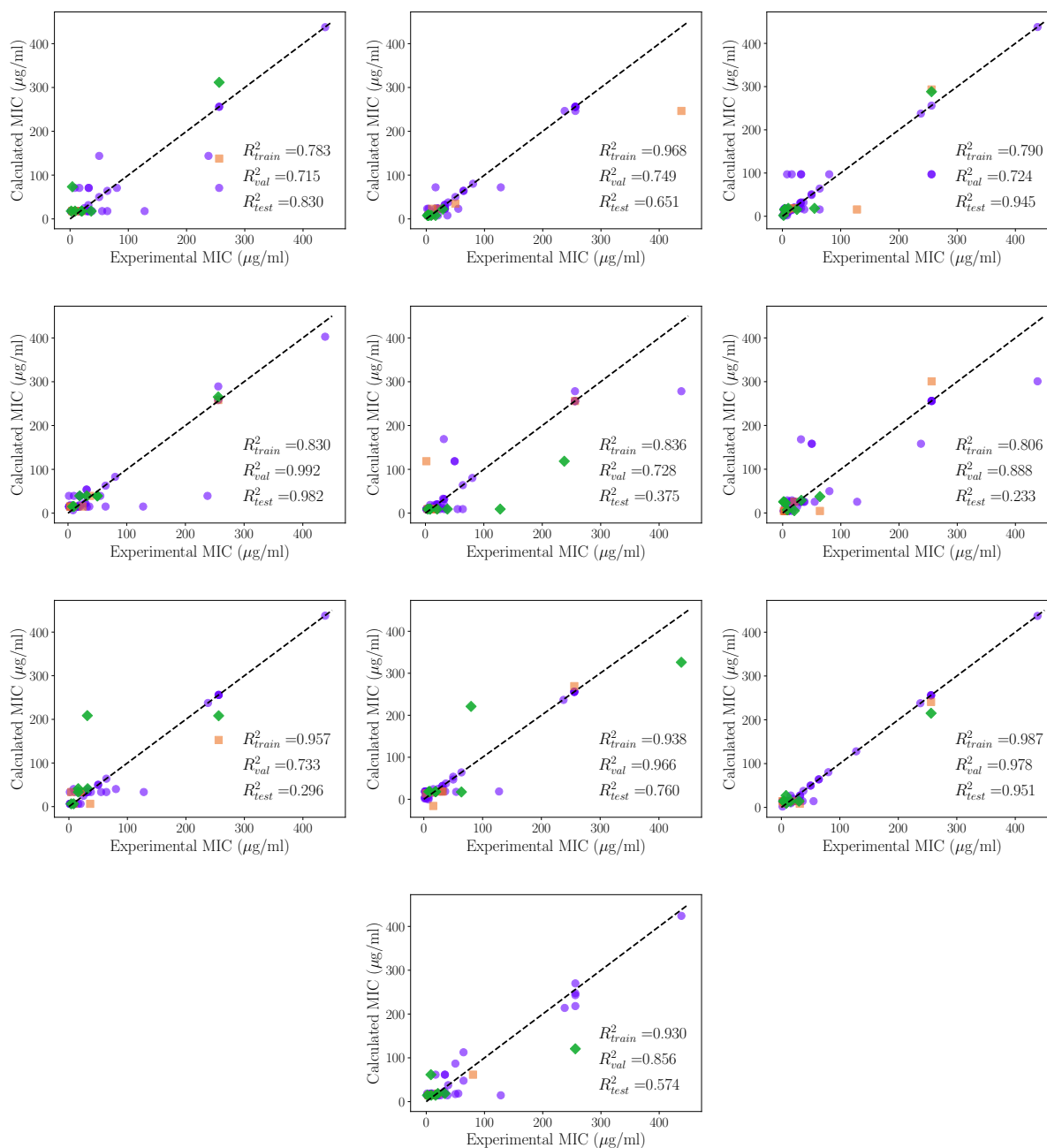
**Figure B.2:** Flow chart showing the schematic of how the ANN models were developed. The part embedded in the dotted box is the logic of selecting the best models in traditional QSAR method, if the goal had been to be content with the prediction for the test set with 7 peptides. In such methods, typically a person conducting the test or the person developing the model but chooses to stay blind to this additional test data, eventually compares the performance of the predictions on this small test set. However, in our case, the goal is to make predictions for the 2338 AMPs for which no activity measurements are yet available, a secondary validation criteria was used to screen the models further.

**Table B.2:** Experimental and predicted MIC values of the 9 qualitative AMP data that were used for an additional test. Before accepting a model, at least 5 of these 9 results were verified to be more than the lower bound suggested by the experiments (to within a factor of 2). An additional condition  $R^2 > 0.6$  was used with the test data set from the quantitative data.

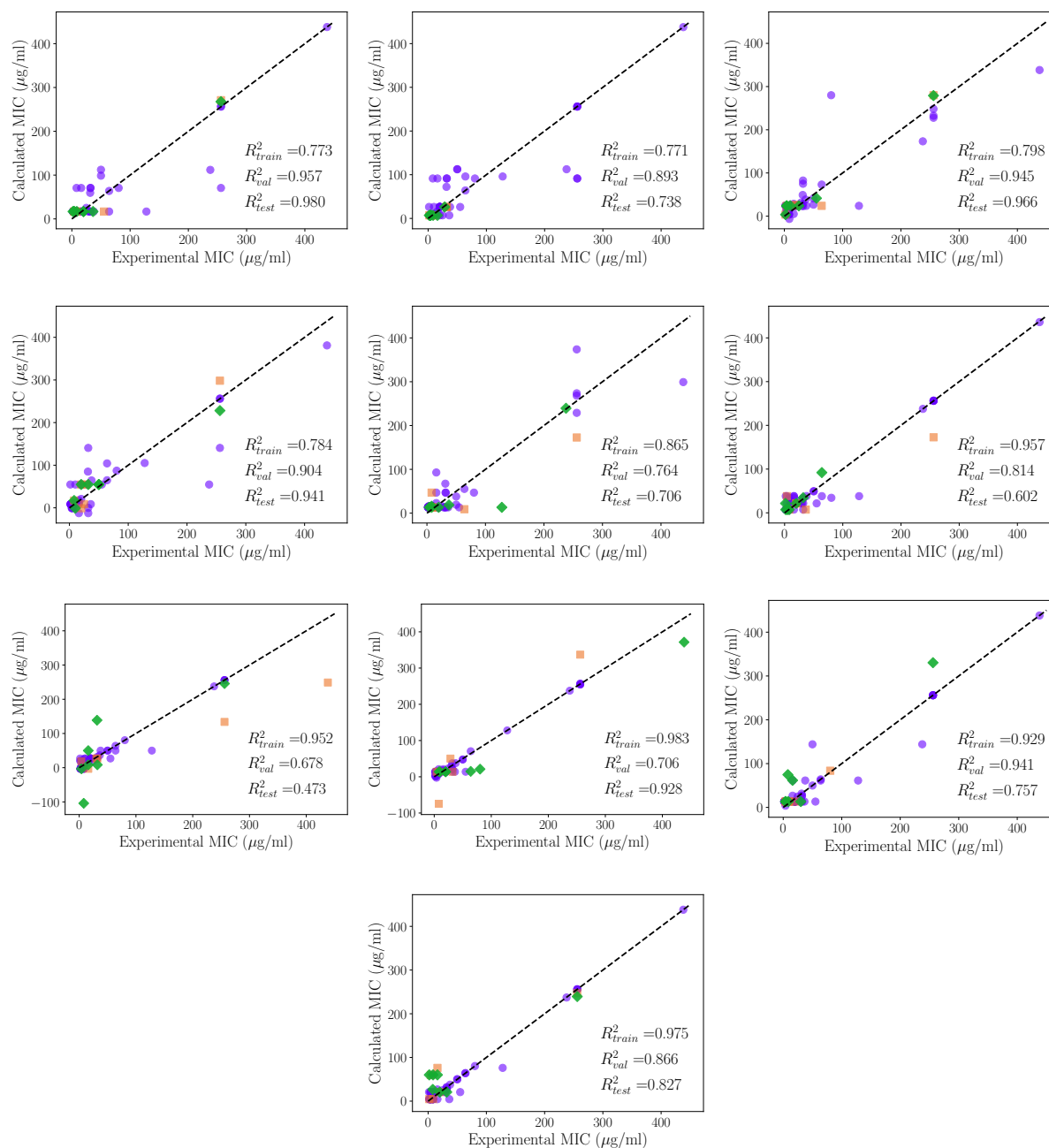
Name	Expt. MIC	Model-1 MIC	Model-2 MIC
	$\mu\text{g/ml}$	$\mu\text{g/ml}$	$\mu\text{g/ml}$
C18G-His	>31.9275	75.9	9.4
I10	>128	482.1	437.1
NK15	>128	4.6	14.2
NK14	>256	60.1	14.2
NK11	>256	239.17	215.1
NK10	>256	86.1	14.6
BR030	>20	255.5	214.8
BR043	>20	22.4	14.2
BR044	>20	62.6	14.2

**Table B.3:** 10-fold cross validation analysis was performed with a hidden layer between the input and output layers. The hidden layer architecture with 6, 8, and 10 neurons respectively were independently modelled. Here we tabulated the number of times, out of 10,  $R^2_{test} > 0.6$ , as well as the mean squared error (MSE) and standard deviation (SD). The overall error was optimal with the choice of 8 neurons, although all three architectures performed satisfactorily.

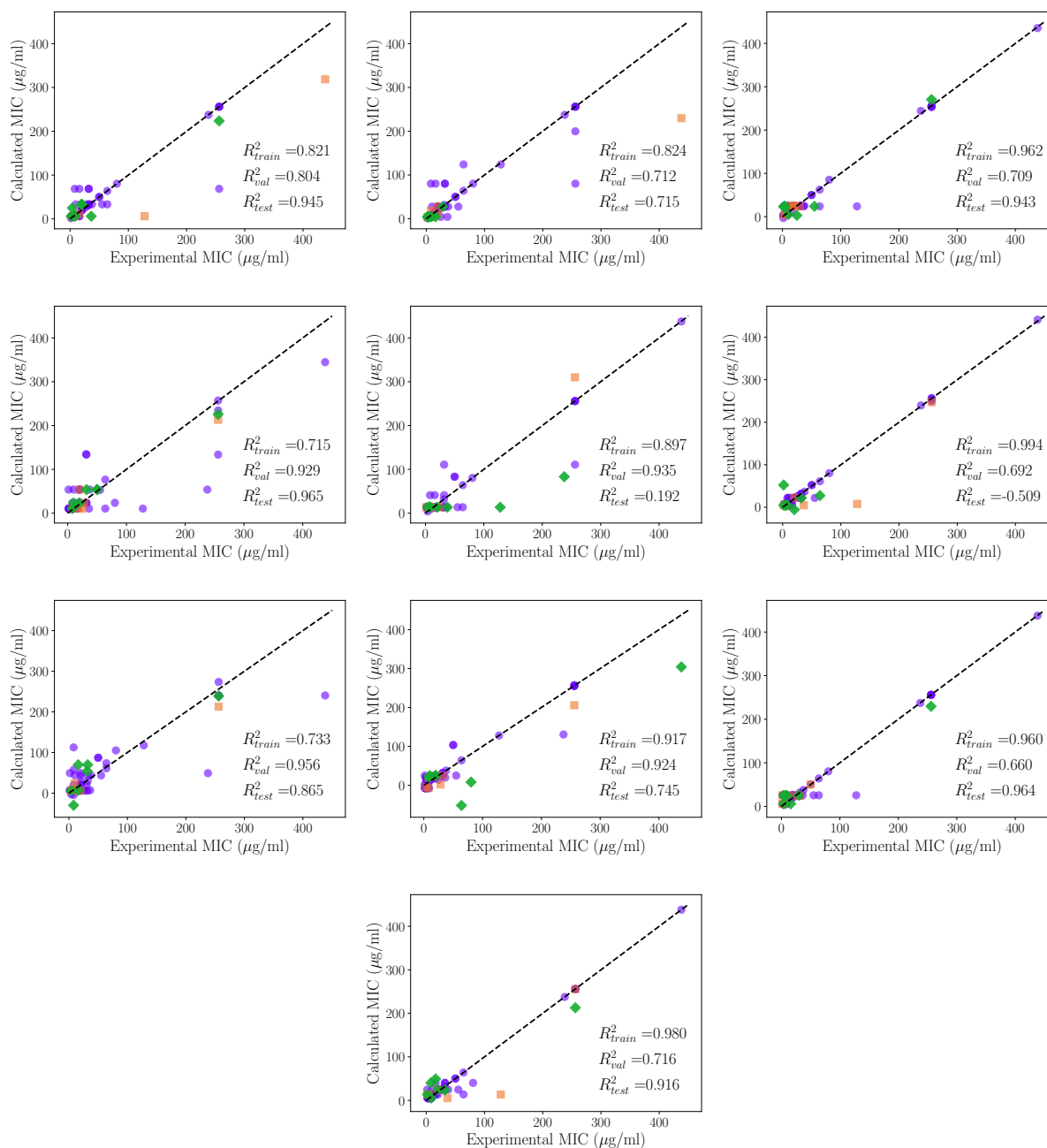
	Number of neurons in the hidden layer		
	6 neurons	8 neurons	10 neurons
$\#R^2_{test} > 0.6$	6	9	8
MSE	1963.979	1119.023	1423.350
SD	1984.295	1112.944	1972.955



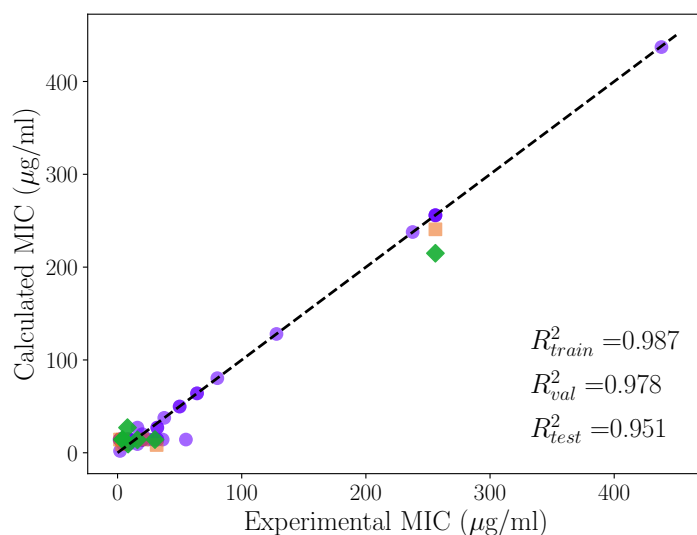
**Figure B.3:** Comparison of the experimental MIC ( $\mu\text{g/ml}$ ) and predicted MIC ( $\mu\text{g/ml}$ ) values of AMP obtained from the 10 fold cross validation calculation. The neural network is trained with one hidden layer consisting of 6 neurons.



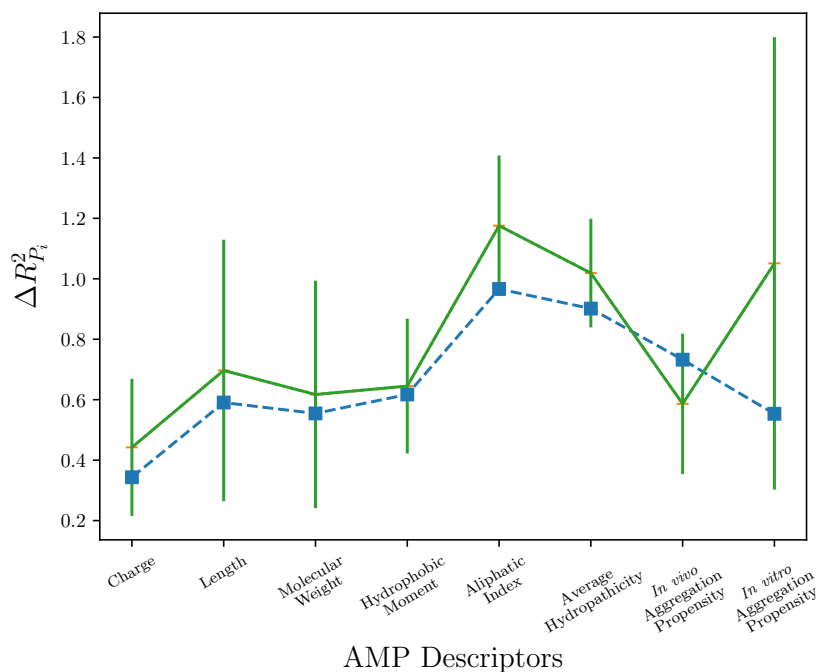
**Figure B.4:** Comparison of the experimental MIC ( $\mu\text{g/ml}$ ) and predicted MIC ( $\mu\text{g/ml}$ ) values of AMP obtained from the 10 fold cross validation calculation. The neural network is trained with one hidden layer consisting of 8 neurons.



**Figure B.5:** Comparison of the experimental MIC ( $\mu\text{g/ml}$ ) and predicted MIC ( $\mu\text{g/ml}$ ) values of AMP obtained from the 10 fold cross validation calculation. The neural network is trained with one hidden layer consisting of 10 neurons.



**Figure B.6:** Comparison of the experimental and calculated MIC ( $\mu\text{g/ml}$ ) of curated AMPs on *A. baumannii* obtained from Model-2, calculated by using 6 hidden neurons. Training (purple circles), validation (orange squares) and test (green diamonds) sets are shown. The data used in the analysis for the peptides given in **Table B.1**. The raw data is in **Data File in GitHub** ([https://github.com/malayrb/Thesis/blob/main/Ch3/File\\_S2.xlsx](https://github.com/malayrb/Thesis/blob/main/Ch3/File_S2.xlsx)).



**Figure B.7:** The relative importance of the different parameters is shown for Model-2 (blue color). Aliphatic index continues to be the most relevant variable, similar to **Figure 3.3**. The green line shows the average and standard deviation value of the variable importance by taking into account all the models of 10 fold cross validation.

**Table B.4:** Comparison of the experimental and predicted MIC values for the MDR *A. baumannii* strains. The results were satisfactory for most of these cases. Polydim-I has extremely low values in one of the parameter used in the model ( $\mu_H$ ), which is not falling in the training range.

Peptide	Expt. MIC $\mu\text{g/ml}$	Model-1 MIC $\mu\text{g/ml}$	Model-2 MIC $\mu\text{g/ml}$	Reference
Agelaia-MPI	39.2	76.1	9.4	[15]
Polybia-MPII	40.4	76.1	9.4	[15]
Polydim-I	> 61.1	-110.9	-172.2	[15]
Con10	70.6	74.8	-174.7	[15]
NDBP-5.8	> 37.8	76.0	9.4	[15]
LS-sarcotoxin	4	20.7	14.2	[16]
LS-stomoxyn	4	76.1	14.2	[16]
BP56	4	4.6	14.2	[17]

## Bibliography

- [1] X. Vila-Farres, C. G. De La Maria, R. López-Rojas, J. Pachón, E. Giralt, and J. Vila, “In vitro activity of several antimicrobial peptides against colistin-susceptible and colistin-resistant *Acinetobacter baumannii*,” *Clinical Microbiology and Infection*, vol. 18, no. 4, pp. 383–387, 2012.
- [2] S.-Y. Peng, R.-I. You, M.-J. Lai, N.-T. Lin, L.-K. Chen, and K.-C. Chang, “Highly potent antimicrobial modified peptides derived from the *Acinetobacter baumannii* phage endolysin lysab2,” *Scientific reports*, vol. 7, no. 1, p. 11477, 2017.
- [3] J. M. Conlon, E. Ahmed, T. Pal, and A. Sonnevend, “Potent and rapid bactericidal action of alyteserin-1c and its [e4k] analog against multidrug-resistant strains of *Acinetobacter baumannii*,” *Peptides*, vol. 31, no. 10, pp. 1806–1810, 2010.
- [4] M. Jaśkiewicz, D. Neubauer, K. Kazor, S. Bartoszewska, and W. Kamysz, “Antimicrobial activity of selected antimicrobial peptides against planktonic culture and biofilm of *acinetobacter baumannii*,” *Probiotics and Antimicrobial Proteins*, pp. 1–8, 2018.
- [5] M. F. Mohamed, A. Brezden, H. Mohammad, J. Chmielewski, and M. N. Seleem, “A short d-enantiomeric antimicrobial peptide with potent immunomodulatory and antibiofilm activity against multidrug-resistant *Pseudomonas aeruginosa* and *Acinetobacter baumannii*,” *Scientific Reports*, vol. 7, no. 1, p. 6953, 2017.

- [6] D. Li, Y. Yang, Z. Tian, J. Lv, F. Sun, Q. Wang, Y. Liu, and P. Xia, “Synergistic antibiotic effect of looped antimicrobial peptide clp-19 with bactericidal and bacteriostatic agents,” *Oncotarget*, vol. 8, no. 34, pp. 55958–55966, 2017.
- [7] E. M. Kohn, D. J. Shirley, L. Arotzky, A. M. Picciano, Z. Ridgway, M. W. Urban, B. R. Carone, and G. A. Caputo, “Role of cationic side chains in the antimicrobial activity of c18g,” *Molecules*, vol. 23, no. 2, p. 329, 2018.
- [8] I. Wiegand, K. Hilpert, and R. E. Hancock, “Agar and broth dilution methods to determine the minimal inhibitory concentration (mic) of antimicrobial substances,” *Nature protocols*, vol. 3, no. 2, p. 163, 2008.
- [9] J. Andrä, D. Monreal, G. M. de Tejada, C. Olak, G. Brezesinski, S. S. Gomez, T. Goldmann, R. Bartels, K. Brandenburg, and I. Moriyon, “Rationale for the design of shortened derivatives of the nk-lysin derived antimicrobial peptide nk-2 with improved activity against gram-negative pathogens,” *Journal of Biological Chemistry*, 2007.
- [10] A. Almaaytah, A. Ya’u, A. Abualhaijaa, S. Tarazi, N. Alshar’i, and Q. Al-Balas, “Peptide consensus sequence determination for the enhancement of the antimicrobial activity and selectivity of antimicrobial peptides,” *Infection and drug resistance*, vol. 10, pp. 1–17, 2017.
- [11] T. T. Thomsen, *Peptide Antibiotics for ESKAPE Pathogens: Past, Present and Future Perspectives of Antimicrobial Peptides for the Treatment of Serious Gram-Negative and Gram-Positive Infections*. PhD thesis, Department of Biology, Faculty of Science, University of Copenhagen, 2016.
- [12] A. Almaaytah, M. T. Qaoud, A. Abualhaijaa, Q. Al-Balas, and K. H. Alzoubi, “Hybridization and antibiotic synergism as a tool for reducing the cytotoxicity of antimicrobial peptides,” *Infection and Drug Resistance*, vol. 11, pp. 835–847, 2018.
- [13] E. Jayamani, R. Rajamuthiah, J. Larkins-Ford, B. B. Fuchs, A. L. Conery, A. Vilcinskas, F. M. Ausubel, and E. Mylonakis, “Insect-derived cecropins display activity against *Acinetobacter baumannii* in a whole-animal high-throughput *Caenorhabditis elegans* model,” *Antimicrobial agents and chemotherapy*, pp. 1728–1737, 2015.
- [14] L. B. Reller, M. Weinstein, J. H. Jorgensen, and M. J. Ferraro, “Antimicrobial susceptibility testing: A review of general principles and contemporary practices,” *Clinical infectious diseases*, vol. 49, no. 11, pp. 1749–1755, 2009.

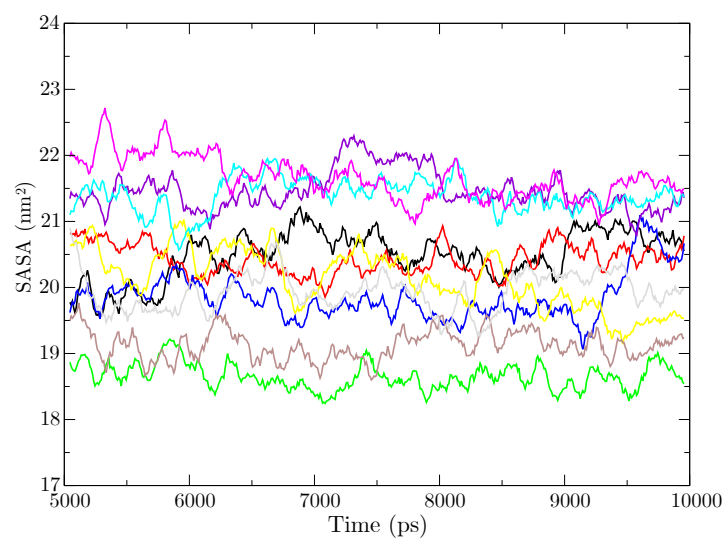


- 
- [15] R. C. das Neves, M. R. Mortari, E. F. Schwartz, A. Kipnis, and A. P. Junqueira-Kipnis, “Antimicrobial and antibiofilm effects of peptides from venom of social wasp and scorpion on multidrug-resistant *Acinetobacter baumannii*,” *Toxins*, vol. 11, no. 4, p. 216, 2019.
- [16] R. Hirsch, J. Wiesner, A. Marker, Y. Pfeifer, A. Bauer, P. E. Hammann, and A. Vilcinskas, “Profiling antimicrobial peptides from the medical maggot *Lucilia sericata* as potential antibiotics for mdr gram-negative bacteria,” *Journal of Antimicrobial Chemotherapy*, vol. 74, no. 1, pp. 96–107, 2018.
- [17] I. Di Bonaventura, S. Baeriswyl, A. Capecchi, B.-H. Gan, X. Jin, T. N. Siriwardena, R. He, T. Köhler, A. Pompilio, G. Di Bonaventura, *et al.*, “An antimicrobial bicyclic peptide from chemical space against multidrug resistant gram-negative bacteria,” *Chemical communications*, vol. 54, no. 40, pp. 5130–5133, 2018.

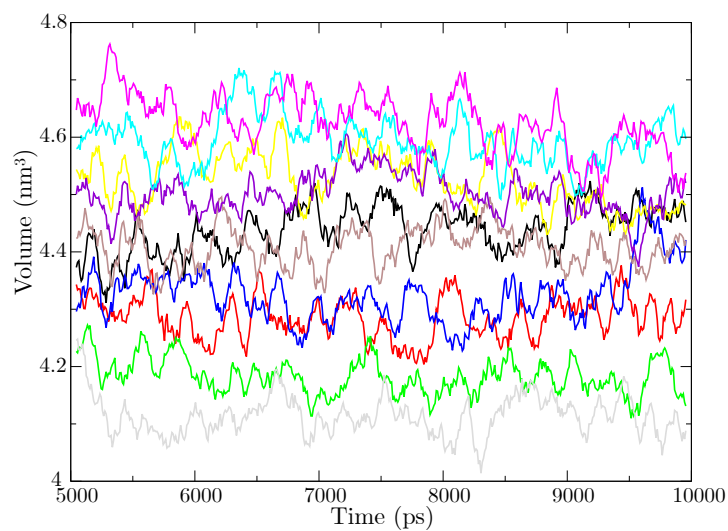


## Appendix C

# Molecular Dynamics Based Antimicrobial Activity Descriptors for Synthetic Cationic Peptides

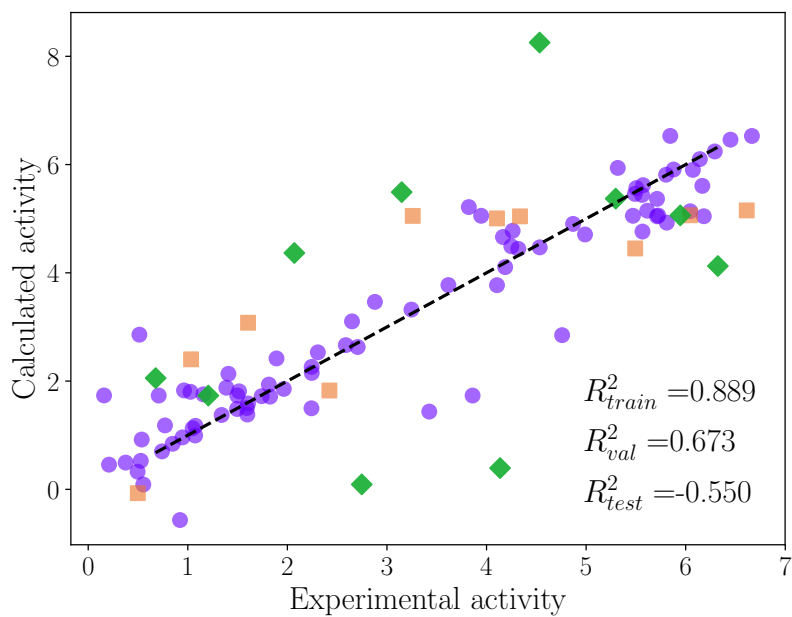


(a)



(b)

**Figure C.1:** The SASA and volume over 10ns simulation shows the system gets stabilized within 10ns. Only 10 peptides are shown as an example.



**Figure C.2:** The model predicted using the parameters calculated from the last 5ns of total 10ns simulation. It shows very poor prediction for the test set.



## Appendix D

### One drug multiple targets: An approach to predict drug efficacies on bacterial strains differing in membrane composition

Strain	MIC	%LPG	%PG	%CL	%iLPG	%oLPG	References
C3	0.5	23.5	68.5	8	-	-	[1]
C4	4	25.6	65.7	8.7	-	-	[1]
C5	0.25	14.2	81.8	4	-	-	[1]
C6	3	19.2	77.7	3	-	-	[1]
C9	0.5	19.3	76	4.7	-	-	[1]
C10	3	23.9	71.5	4.7	-	-	[1]
C19	0.38	14.9	75.1	10	-	-	[1]
C21	4	30.9	50.5	18.6	-	-	[1]
C26	0.38	31.8	62.9	5.3	-	-	[1]
C27	2	28.7	68.9	2.4	-	-	[1]
C32	0.5	21.9	71.2	6.9	-	-	[1]
C33	2	24.1	73.1	2.7	-	-	[1]
C36	0.5	15.1	81.5	3.4	-	-	[1]
C37	3	26.1	69	5	-	-	[1]
C40	0.25	26	68.4	5.6	-	-	[1]
C41	3	30	59.7	10.3	-	-	[1]

Strain	MIC	%LPG	%PG	%CL	%iLPG	%oLPG	References
CB1118 <sup>1</sup>	1	12.37	83.96	5.38	11.16	1.21	[2]
CB2201	1.5	12.29	86.37	3.6	10.92	1.37	[2]
CB2202	3	13.34	80.95	8.59	11.64	1.7	[2]
CB2203	6	17.96	72.93	12.7	15.41	2.56	[2]
CB2205	12	24.61	70.02	7.43	18.61	6	[2]
CB1483	0.25	15.3	77.21	7.49	13.39	1.91	[4]
CB185	4	35.96	52.31	11.73	32.1	3.85	[4]
CB5079	0.5	15.43	72.12	12.44	13.63	1.8	[4]
CB5080	2	26.18	64.08	9.75	24.78	1.39	[4]
CB5083	0.25	12.07	83.3	4.63	10.15	1.92	[4]
CB5082	4	21.61	73.58	4.82	19.24	2.37	[4]
CB5088	0.5	15.36	77.66	6.97	13.61	1.75	[4]
CB5089	2-4	24.91	65.93	9.16	22.62	2.29	[4]
CB1631	0.5	12.08	80.41	7.51	10.25	1.83	[4]
CB1634	4	20.61	71.75	7.63	18.68	1.93	[4]
CB1663	0.5	11.96	83.2	4.84	10.24	1.72	[4]
CB1664	4	16.05	81.33	2.62	14.69	1.36	[4]
CB5057	0.5	15.91	79.25	4.85	14.32	1.59	[4]
CB5059	4	27.22	69.92	2.87	24.77	2.44	[4]
CB5062	0.5	12.71	79.9	7.4	11.49	1.21	[4]
CB5063	8	31.55	59.01	9.44	29.23	2.33	[4]
CB5015	1	14.06	83.31	2.63	12.73	1.32	[4]
CB5016	4	19.27	77.2	3.53	17.61	1.66	[4]
C11	0.38	18	78	5	-	-	[5]
C12	3	25	68	7	-	-	[5]
C28	0.12	15	79	6	-	-	[5]
C29	2	23	70	7	-	-	[5]
C44	0.38	20	74	5	-	-	[5]
C45	4	26	68	6	-	-	[5]
Newman	0.5	17	72	11	-	-	[6]
Newman mprF	0.125	0	86	14	-	-	[6]

<sup>1</sup>CB1118 strain was used as a parental strain in two independent serial passage studies [2, 3]. Novel mutations evolved in the latter study and led to two notable changes - a decreased LPG after certain time and an increased production of carotenoid. Since the influence of carotenoids was an additional parameter for which no other study we considered provided the data, we excluded the latter study from our analysis.

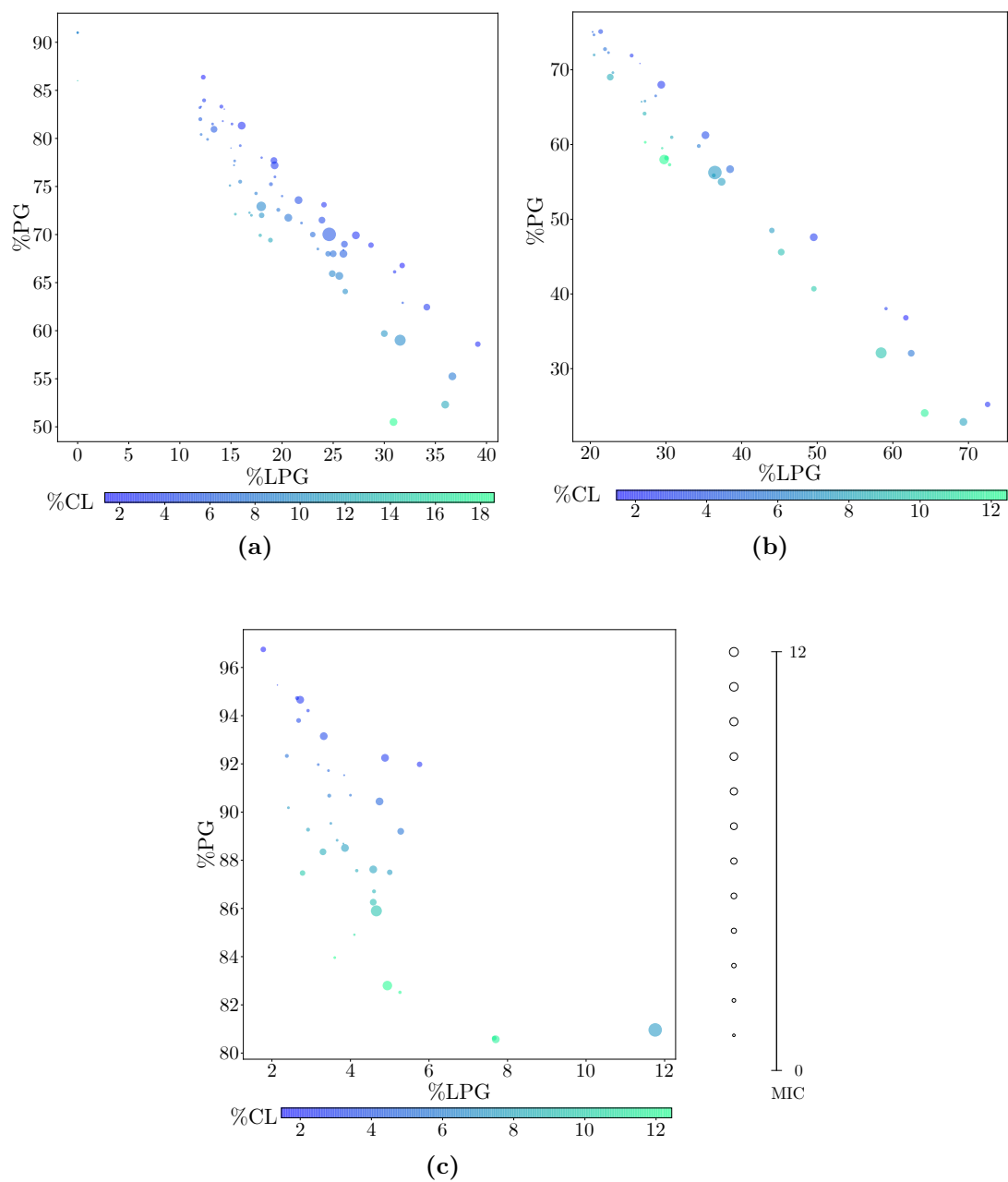


Strain	MIC	%LPG	%PG	%CL	%iLPG	%oLPG	References
CS295L	1	12	82	7	-	-	[6]
CS295L+L826F	0.38	0	91	9	-	-	[6]
CT345A	2	18	72	10	-	-	[6]
CT345A+L826F	0.38	0	91	9	-	-	[6]
None	0.5	13.2	81.5	5.3	11.2	2	[7]
None	1	15.9	75.5	8.7	13.6	2.3	[7]
None	2	24.5	68	7.5	22	2.5	[7]
L271	0.125	14.35	83.05	2.59	13.28	1.07	[8]
L8	2	31.75	66.78	1.47	30.86	0.89	[8]
L16	0.75	31.01	66.12	2.87	29.55	1.46	[8]
L56	2	39.15	58.6	2.26	36.27	2.88	[8]
L76	0.38	16.81	72.27	11	14.76	2.05	[8]
SA144	0.75	17.45	74.28	8.27	15.38	2.08	[9]
SA145	0.75	17.86	69.92	12.22	15.24	2.63	[9]
SA147	1.5	18.86	69.42	11.72	15.04	3.83	[9]
MRSA 11/11	1	18.91	75.24	5.86	17.17	1.73	[10, 11]
MRSA 11/17	1	19.63	72.57	7.81	18.16	1.46	[10, 11]
MRSA 11/21	3	34.16	62.45	5.68	32.15	2.72	[10, 11]
REF2145	4	36.66	55.25	7.8	34.67	2.29	[10, 11]
1A <sup>2</sup>	0.25	13	81	6	11	2	[12, 13]
1C <sup>2</sup>	4	26	68	6	22	5	[12, 13]
2A <sup>2</sup>	0.5	11	84	5	9	2	[12, 13]
2C <sup>2</sup>	2	24	67	9	19	4	[12, 13]
3A <sup>2</sup>	0.5	22	69	9	19	3	[12, 13]
3B <sup>2</sup>	4	16	80	4	13	3	[12, 13]

**Table D.1:** Summary of the data curated from multiple experimental sources on different *S. aureus* strains is shown. Total membrane phospholipid composition is described in terms of phosphatidylglycerol (PG), lysyl-PG (LPG), cardiolipin (CL), inner leaflet LPG (iLPG) and outer leaflet LPG (oLPG) and the daptomycin MIC values are given in  $\mu\text{g}/\text{mL}$ . All the MIC values are evaluated according to standard Etest, except when mentioned.

All the curated experimental data are plotted in Figure D.1. Bacterial membrane phospholipid is mainly composed of anionic phosphatidylglycerol (PG), cardiolipin (CL) and lysyl-phosphatidylglycerol (LPG).

<sup>2</sup>MIC values obtained from CLSI guideline, and excluded from our ANN model.



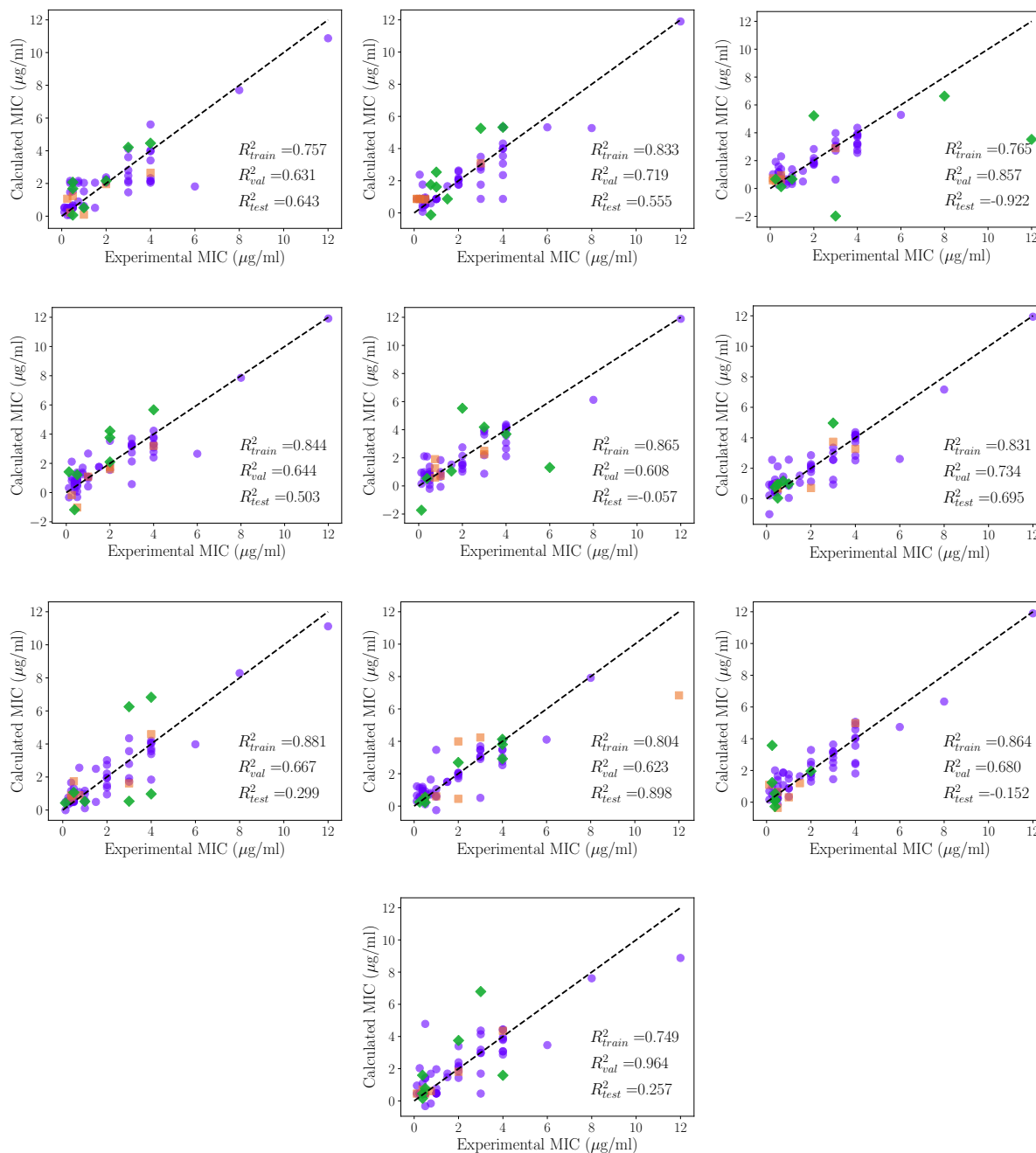
**Figure D.1:** Dependence of the experimentally determined daptomycin MIC values on the phospholipid compositions in a) total membrane, b) inner leaflet and c) outer leaflet. The radius of the circle represents MIC ( $\mu\text{g}/\text{ml}$ ) and the change in color represents CL concentration. The data shows ‘non-monotonous’ trends, as the magnitude of MIC goes through variations, rather than a straightforward increase with the parameters.

---

Serial no.	$R_{training}^2$	$R_{validation}^2$	$R_{test}^2$
1	0.757	0.631	0.643
2	0.833	0.719	0.555
3	0.765	0.857	-0.922
4	0.844	0.644	0.503
5	0.865	0.608	-0.057
6	0.831	0.734	0.695
7	0.881	0.667	0.299
8	0.804	0.623	0.898
9	0.864	0.680	-0.152
10	0.749	0.964	0.257

---

**Table D.2:**  $R_{training}^2$ ,  $R_{validation}^2$  and  $R_{test}^2$  values obtained from the 10 fold validation calculation for total membrane composition. The neural network is trained with one hidden layer consisting of 8 neurons. A summary of mean square error and standard deviation when the number of neurons is varied is shown in **Table D.3**.



**Figure D.2:** Comparison of the experimentally determined MIC ( $\mu\text{g/ml}$ ) of daptomycin with the MIC calculated using the 10-fold cross validation for total membrane composition. The neural network is trained with one hidden layer consisting of 8 neurons. The choice of the 8-neurons was made after trying calculations with different number of neurons. The mean square error and standard deviation for these choices of neurons are shown in **Table D.3**.

Strain	MIC( $\mu\text{g/ml}$ )	MIC with best model		
		Total	outer	inner
1A	0.25	0.943	-0.76	0.855
1C	4	3.271	22.701	3.064
2A	0.5	0.943	0.648	0.855
2C	2	2.716	10.49	3.018
3A	0.5	-0.533	2.76	2.901
3B	4	0.152	0.256	0.838

**Table D.3:** Predictions for the data points that were not measured by Etest. The best model developed based on Etest results was used for making this cross-prediction for the data measured by other methods. All the data used for this calculation is given in **Table D.1 & D.4**.

Strain	MIC	Inner leaflet			Outer leaflet			References
		%LPG	%PG	%CL	%LPG	%PG	%CL	
CB1118	1	21.94	72.77	5.29	2.38	92.33	5.29	[2]
CB2201	1.5	21.36	75.12	3.52	2.68	93.8	3.52	[2]
CB2202	3	22.63	69.02	8.35	3.3	88.35	8.35	[2]
CB2203	6	29.75	57.99	12.26	4.94	82.8	12.26	[2]
CB2205	12	36.47	56.25	7.28	11.76	80.96	7.28	[2]
CB1483	0.25	26.78	65.73	7.49	3.82	88.69	7.49	[4]
CB185	4	64.21	24.06	11.73	7.7	80.57	11.73	[4]
CB5079	0.5	27.26	60.3	12.44	3.6	83.96	12.44	[4]
CB5080	2	49.56	40.69	9.75	2.78	87.47	9.75	[4]
CB5083	0.25	20.3	75.07	4.63	3.84	91.53	4.63	[4]
CB5082	4	38.48	56.7	4.82	4.74	90.44	4.82	[4]
CB5088	0.5	27.22	65.81	6.97	3.5	89.53	6.97	[4]
CB5089	2-4	45.24	45.6	9.16	4.58	86.26	9.16	[4]
CB1631	0.5	20.5	71.99	7.51	3.66	88.83	7.51	[4]
CB1634	4	37.36	55.01	7.63	3.86	88.51	7.63	[4]
CB1663	0.5	20.48	74.68	4.84	3.44	91.72	4.84	[4]
CB1664	4	29.38	68	2.62	2.72	94.66	2.62	[4]
CB5057	0.5	28.64	66.51	4.85	3.18	91.97	4.85	[4]
CB5059	4	49.54	47.59	2.87	4.88	92.25	2.87	[4]
CB5062	0.5	22.98	69.62	7.4	2.42	90.18	7.4	[4]
CB5063	8	58.45	32.11	9.44	4.66	85.9	9.44	[4]
CB5015	1	25.46	71.91	2.63	2.64	94.73	2.63	[4]

Strain	MIC	Inner leaflet			Outer leaflet			References
		%LPG	%PG	%CL	%LPG	%PG	%CL	
CB5016	4	35.22	61.25	3.53	3.32	93.15	3.53	[4]
None	0.5	22.40	72.30	5.30	4.00	90.70	5.30	[7]
None	1	27.17	64.14	8.69	4.60	86.71	8.69	[7]
None	2	44.00	48.50	7.50	5.00	87.50	7.50	[7]
L271	0.125	26.56	70.85	2.59	2.14	95.27	2.59	[8]
L8	2	61.72	36.81	1.47	1.78	96.75	1.47	[8]
L16	0.75	59.10	38.03	2.87	2.92	94.21	2.87	[8]
L56	2	72.53	25.21	2.26	5.76	91.98	2.26	[8]
L76	0.38	29.50	59.51	10.99	4.10	84.91	10.99	[8]
SA144	0.75	30.76	60.97	8.27	4.16	87.57	8.27	[9]
SA145	0.75	30.48	57.30	12.22	5.26	82.52	12.22	[9]
SA147	1.5	30.08	58.20	11.72	7.66	80.62	11.72	[9]
MRSA 11/11	1	34.34	59.80	5.86	3.46	90.68	5.86	[10, 11]
MRSA 11/17	1	36.32	55.87	7.81	2.92	89.27	7.81	[10, 11]
MRSA 11/21	3	62.43	32.06	5.51	5.28	89.20	5.51	[10, 11]
REF2145	4	69.33	22.87	7.80	4.58	87.62	7.80	[10, 11]
1A <sup>2</sup>	0.25	22	72	6	4	90	6	[12, 13]
1C <sup>2</sup>	4	43.56	50.5	5.94	9.9	84.16	5.94	[12, 13]
2A <sup>2</sup>	0.5	18	77	5	4	91	5	[12, 13]
2C <sup>2</sup>	2	38.38	52.53	9.09	8.08	82.83	9.09	[12, 13]
3A <sup>2</sup>	0.5	38	53	9	6	85	9	[12, 13]
3B <sup>2</sup>	4	26	70	4	6	90	4	[12, 13]

**Table D.4:** Inner and outer leaflet of membrane phospholipid compositions of the different *S. aureus* strains in terms of lysl-phosphatidylglycerol (lysl-PG), phosphatidylglycerol (PG), cardiolipin (CL) and the daptomycin MIC values in  $\mu\text{g}/\text{ml}$ . These numbers were derived as discussed in the **Methods section** of the article. All the MIC values are evaluated according to standard Etest, except when noted otherwise.

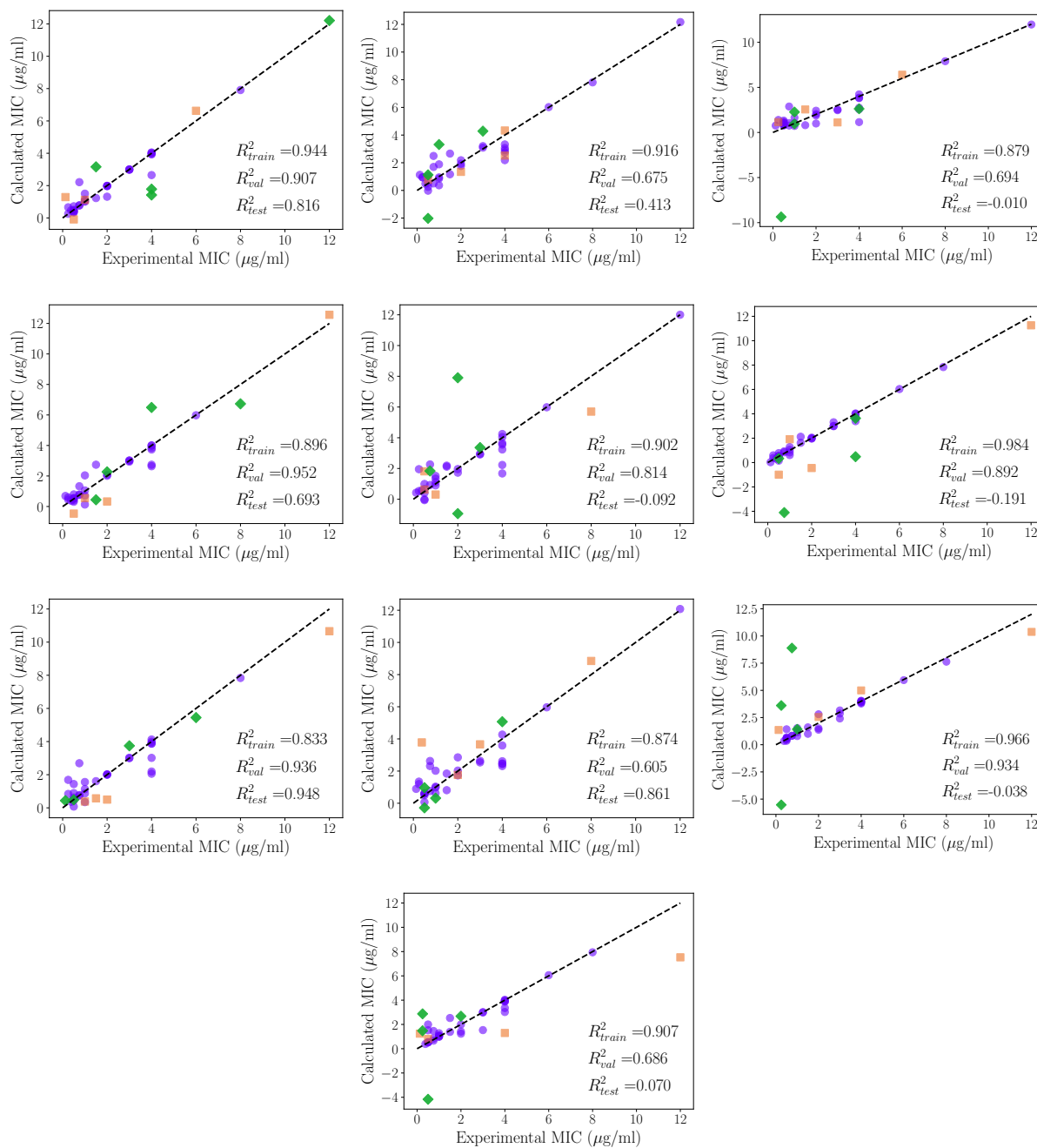
<sup>2</sup>MIC determined using CLSI guideline, and was excluded in the 10-fold cross validation ANN model analysis.

---

Serial no.	$R_{training}^2$	$R_{validation}^2$	$R_{test}^2$
1	0.944	0.907	0.816
2	0.916	0.675	0.413
3	0.879	0.694	-0.010
4	0.896	0.952	0.693
5	0.902	0.814	-0.092
6	0.984	0.892	-0.191
7	0.833	0.936	0.948
8	0.874	0.605	0.861
9	0.966	0.934	-0.038
10	0.907	0.686	0.070

---

**Table D.5:**  $R_{training}^2$ ,  $R_{validation}^2$  and  $R_{test}^2$  values obtained from the 10 fold validation for outer leaflet composition. The neural network is trained with one hidden layer consisting of 8 neurons.



**Figure D.3:** Comparison of the experimentally determined MIC of daptomycin (µg/ml) and the MIC obtained from the 10-fold cross validation using the outer leaflet composition. The neural network is trained with one hidden layer consisting of 8 neurons.

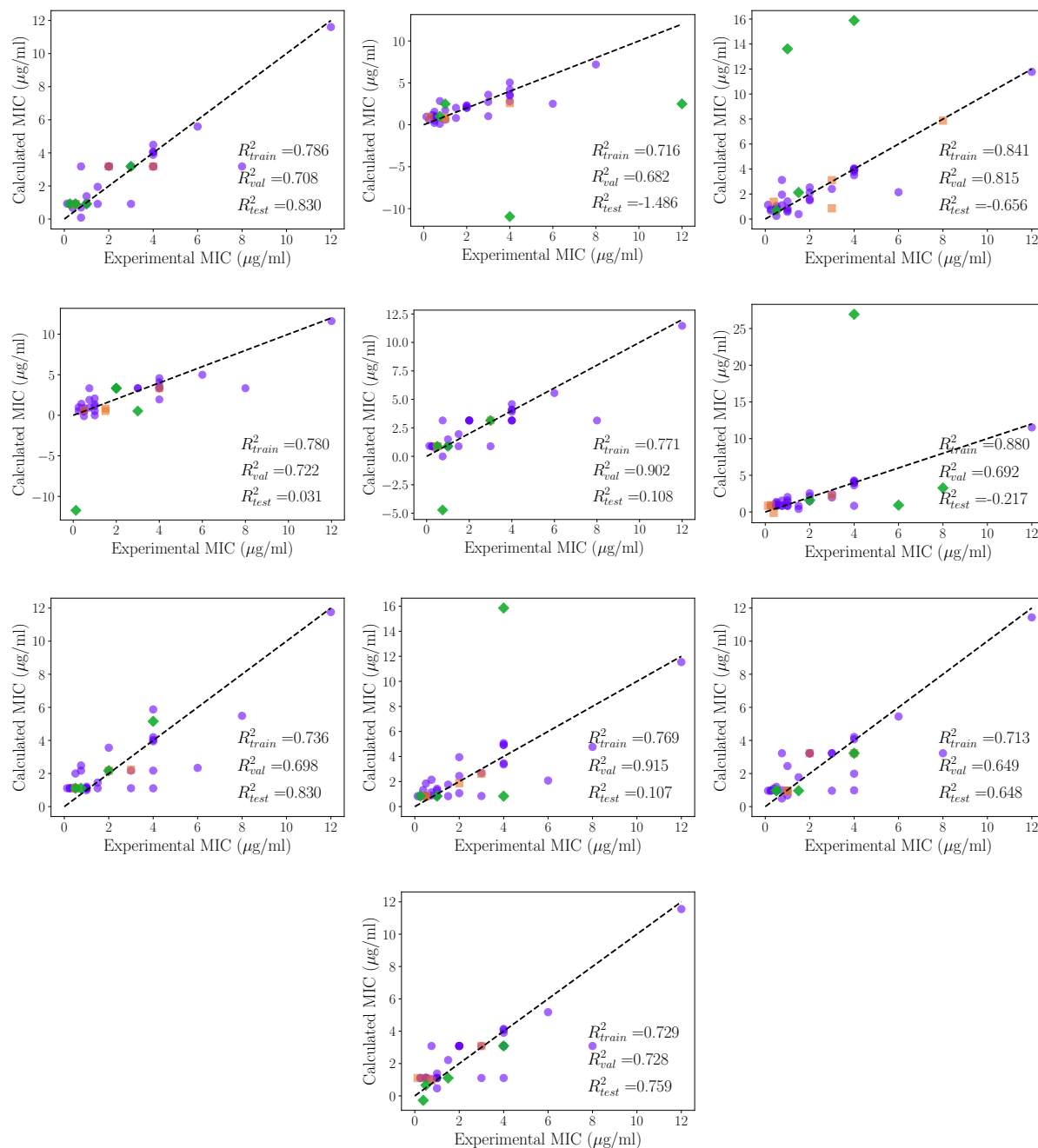


---

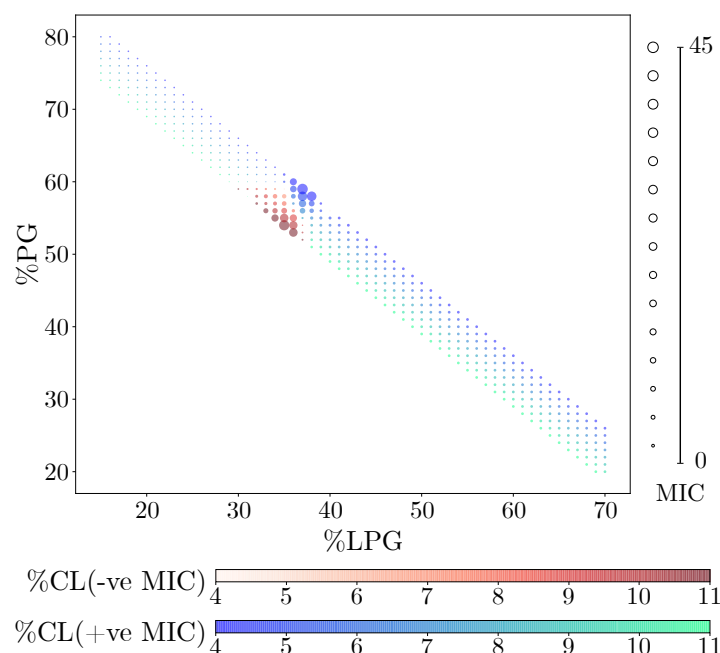
Serial no.	$R_{training}^2$	$R_{validation}^2$	$R_{test}^2$
1	0.786	0.708	0.830
2	0.716	0.682	-1.486
3	0.841	0.815	-0.656
4	0.780	0.722	0.031
5	0.771	0.902	0.108
6	0.880	0.692	-0.217
7	0.736	0.698	0.830
8	0.769	0.915	0.107
9	0.713	0.649	0.648
10	0.729	0.728	0.759

---

**Table D.6:**  $R_{training}^2$ ,  $R_{validation}^2$  and  $R_{test}^2$  values obtained from the 10-fold cross validation calculation for inner leaflet composition. The neural network is trained with one hidden layer consisting of 4 neurons.



**Figure D.4:** Comparison of the experimentally determined MIC of daptomycin ( $\mu\text{g/ml}$ ) with the MIC calculated with 10-fold cross validation for inner leaflet composition. The neural network is trained with one hidden layer consisting of 4 neurons.



**Figure D.5:** Dependence of daptomycin MIC ( $\mu\text{g}/\text{ml}$ ) values in our predictions across the entire range of %PG and %LPG compositions of the inner leaflet. The blue-green and red color represents the change in CL percentage for positive and negative MIC values respectively and radius of the circle increases with increase in absolute value of MIC. Negative MIC is not physically meaningful and we believe it is an artefact of training a non-monotonous variation of MIC on a limited data.

	total		outer		inner	
	LPG	PG	oLPG	oPG	iLPG	iPG
Training set	0.002	0.002	0.029	0.395	0.108	0.069
Validation set	0.037	0.072	0.146	0.112	0.007	0.001
Test set	0.056	0.245	0.302	0.697	0.113	0.198

**Table D.7:** P-values of the variables used in the analysis, obtained from our neural network analysis performed using total membrane and outer leaflet composition with 8 neurons, and inner leaflet composition with 4 neurons. We report for two independent variables, since the percentage of the third lipid (CL) is 100-PG-LPG. The p-values for the training set are good.

## Bibliography

- [1] A. S. Bayer, N. N. Mishra, A. L. Cheung, A. Rubio, and S.-J. Yang, “Dysregulation of *mprF* and *dltABCD* expression among daptomycin-non-susceptible mrsa clinical isolates,” *Journal of Antimicrobial Chemotherapy*, vol. 71, no. 8, pp. 2100–2104, 2016.
- [2] N. N. Mishra, S.-J. Yang, A. Sawa, A. Rubio, C. C. Nast, M. R. Yeaman, and A. S. Bayer, “Analysis of cell membrane characteristics of in vitro-selected daptomycin-resistant strains of methicillin-resistant staphylococcus aureus,” *Antimicrobial agents and chemotherapy*, vol. 53, no. 6, pp. 2312–2318, 2009.
- [3] N. N. Mishra, A. Rubio, C. C. Nast, and A. S. Bayer, “Differential adaptations of methicillin-resistant staphylococcus aureus to serial in vitro passage in daptomycin: evolution of daptomycin resistance and role of membrane carotenoid content and fluidity,” *International journal of microbiology*, vol. 2012, 2012.
- [4] N. N. Mishra and A. S. Bayer, “Correlation of cell membrane lipid profiles with daptomycin resistance in methicillin-resistant staphylococcus aureus (mrsa).,” *Antimicrobial agents and chemotherapy*, pp. AAC-02182, 2012.
- [5] K.-M. Kang, N. N. Mishra, K. T. Park, G.-Y. Lee, Y. H. Park, A. S. Bayer, and S.-J. Yang, “Phenotypic and genotypic correlates of daptomycin-resistant methicillin-susceptible staphylococcus aureus clinical isolates,” *Journal of Microbiology*, vol. 55, no. 2, pp. 153–159, 2017.
- [6] S.-J. Yang, N. N. Mishra, K.-M. Kang, G.-Y. Lee, J.-H. Park, and A. S. Bayer, “Impact of multiple single-nucleotide polymorphisms within *mprf* on daptomycin resistance in staphylococcus aureus,” *Microbial Drug Resistance*, 2018.
- [7] A. S. Bayer, N. N. Mishra, G. Sakoulas, P. Nonejuie, C. C. Nast, J. Pogliano, K.-T. Chen, S. N. Ellison, M. R. Yeaman, and S.-J. Yang, “Heterogeneity of *mprf* sequences in methicillin-resistant staphylococcus aureus clinical isolates: role in cross-resistance between daptomycin and host defense antimicrobial peptides,” *Antimicrobial agents and chemotherapy*, pp. AAC-03422, 2014.
- [8] N. N. Mishra, S.-J. Yang, L. Chen, C. Muller, A. Saleh-Mghir, S. Kuhn, A. Peschel, M. R. Yeaman, C. C. Nast, B. N. Kreiswirth, A.-C. Crémieux, and A. S. Bayer, “Emergence of daptomycin resistance in daptomycin-naive rabbits with methicillin-resistant staphylococcus aureus prosthetic joint infection is associated with resistance

- to host defense cationic peptides and mprf polymorphisms,” *PLoS One*, vol. 8, no. 8, p. e71151, 2013.
- [9] N. N. Mishra, G. Y. Liu, M. R. Yeaman, C. C. Nast, R. A. Proctor, J. McKinnell, and A. S. Bayer, “Carotenoid-related alteration of cell membrane fluidity impacts staphylococcus aureus susceptibility to host defense peptides,” *Antimicrobial agents and chemotherapy*, vol. 55, no. 2, pp. 526–531, 2011.
- [10] S.-J. Yang, C. C. Nast, N. N. Mishra, M. R. Yeaman, P. D. Fey, and A. S. Bayer, “Cell wall thickening is not a universal accompaniment of the daptomycin nonsusceptibility phenotype in staphylococcus aureus: evidence for multiple resistance mechanisms,” *Antimicrobial agents and chemotherapy*, vol. 54, no. 8, pp. 3079–3085, 2010.
- [11] M. H. Murthy, M. E. Olson, R. W. Wickert, P. D. Fey, and Z. Jalali, “Daptomycin non-susceptible methicillin-resistant staphylococcus aureus usa 300 isolate,” *Journal of medical microbiology*, vol. 57, no. 8, pp. 1036–1038, 2008.
- [12] N. N. Mishra, A. S. Bayer, C. Weidenmaier, T. Grau, S. Wanner, S. Stefani, V. Cafiso, T. Bertuccio, M. R. Yeaman, C. C. Nast, and S.-J. Yang, “Phenotypic and genotypic characterization of daptomycin-resistant methicillin-resistant staphylococcus aureus strains: relative roles of mprf and dlt operons,” *PLoS One*, vol. 9, no. 9, p. e107426, 2014.
- [13] V. Cafiso, T. Bertuccio, S. Purrello, F. Campanile, C. Mammina, A. Sartor, A. Raglio, and S. Stefani, “dltA overexpression: a strain-independent keystone of daptomycin resistance in methicillin-resistant staphylococcus aureus,” *International journal of antimicrobial agents*, vol. 43, no. 1, pp. 26–31, 2014.



# Appendix E

## Exploring the context of diacidic motif DE as a signal for unconventional protein secretion in eukaryotic proteins

**Table E.2:** Probability of various cases considering the three factors (Order, charge and hydrophobicity) affecting the secretion process and the p-value for each case with individual p-value for each variable

	CASE	NN	NS	SN	SS	p-Value	p-Value Charge	p-Value Hydrophobicity	p-Value Ordered or dis-ordered	Odds Ratio
	1	5	5	34	45	0.7430	0.5234	0.7412	0.4599	1.3235
	2	10	8	29	42	0.2963	0.5355	0.2447	0.4599	1.8103
	3	15	10	24	40	0.0620	0.2017	0.4791	0.4599	2.5
	4	5	4	34	46	0.4974	0.9483	0.9435	0.4599	1.6912
	5	4	8	35	42	0.5393	0.6623	0.2526	0.6801	0.6
	6	12	8	27	42	0.1266	0.3702	0.3558	0.6801	2.3333
	7	13	9	26	41	0.1371	0.1439	0.3871	0.6801	2.2778
Total	8	13	6	26	44	0.0195	0.9116	0.1294	0.6801	3.6667
data	9	17	14	22	36	0.1784	0.9694	0.9351	0.1030	1.987
	10	19	14	20	36	0.0504	0.8179	0.9141	0.1030	2.4429
	11	16	14	23	36	0.2593	0.5252	0.8126	0.1030	1.7888
	12	19	11	20	39	0.0125	0.8703	0.4019	0.1030	3.3682
	13	15	10	24	40	0.0620	0.7699	0.8581	0.0921	2.5
	14	15	11	24	39	0.1048	0.8016	0.7573	0.0921	2.2159
	15	15	10	24	40	0.0620	0.9585	0.9043	0.0921	2.5
	16	15	11	24	39	0.1048	0.7480	0.6659	0.0921	2.2159
	1	4	4	5	6	1.0000	0.6710	0.6148	0.3574	1.2
	2	7	3	2	7	0.0698	0.1023	0.4396	0.3574	8.1667





---

7	4	1	5	20	0.0195	0.0932	0.5881	0.3551	16
8	0	0	9	21	1.0000	0.8562	0.7565	0.3551	INFINITY
9	0	1	9	20	1.0000	0.7687	0.5010	0.7552	0
10	5	2	4	19	0.0139	0.3427	0.0395	0.7552	11.875
11	4	2	5	19	0.0492	0.7600	0.0741	0.7552	7.6
12	4	2	5	19	0.0492	0.1545	0.0953	0.7552	7.6
13	2	2	7	19	0.5632	0.9180	0.3877	0.2739	2.7143
14	5	1	4	20	0.0046	0.2981	0.0553	0.2739	25
15	3	2	6	19	0.1432	0.6566	0.2073	0.2739	4.75
16	4	1	5	20	0.0195	0.1748	0.2009	0.2739	16

---

**Table E.1:** All possible cases considered for analysis. If at least one DE motif is present in disordered region (or ordered region) then disordered (or ordered) is important. There are cases where the DE motif is in the border between ordered and disordered region. We have looked those as both ordered (O) and disordered (D) cases.

<b>CASE</b>	<b>Important</b>	<b>Bordered region in D/O</b>	<b>Parameters sorted for analysis</b>
1	Disordered is important	Border as D	min C (max H)
2	Disordered is important	Border as D	max C (min H)
3	Disordered is important	Border as D	min H (max C)
4	Disordered is important	Border as D	max H (min C)
5	Disordered is important	Border as O	min C (max H)
6	Disordered is important	Border as O	max C (min H)
7	Disordered is important	Border as O	min H (max C)
8	Disordered is important	Border as O	max H (min C)
9	Ordered is important	Border as D	min C (max H)
10	Ordered is important	Border as D	max C (min H)
11	Ordered is important	Border as D	min H (max C)
12	Ordered is important	Border as D	max H (min C)
13	Ordered is important	Border as O	min C (max H)
14	Ordered is important	Border as O	max C (min H)
15	Ordered is important	Border as O	min H (max C)
16	Ordered is important	Border as O	max H (min C)

**Table E.3:** Binary logistic regression – odds ratio of the groups analyzed**Total data**

	B	Sig.	Odds=Exp(B)
Charge	0.016	0.931	1.016
Hydrophobicity	-0.031	0.908	0.97
Order	-0.808	0.105	0.446

**Neuro degenerative and known UPS vs non-secretory**

	B	Sig.	Odds=Exp(B)
Charge	0.636	0.037	1.889
Hydrophobicity	1.062	0.011	2.892
Order	-0.143	0.834	0.867

**HSP vs non-secretory**

	B	Sig.	Odds=Exp(B)
Charge	-0.098	0.79	0.907
Hydrophobicity	-0.903	0.097	0.405
Order	-2.014	0.083	0.134

**Cancer data from Groups 3B, 3C**

	B	Sig.	Odds=Exp(B)
Charge	0.011	0.984	1.011
Hydrophobicity	-4.526	0.05	0.011
Order	-5.629	0.106	0.004

**Cancer secretion data including HSP secretion (Groups 3A, 3B, 3C)**

	B	Sig.	Odds=Exp(B)
Charge	-0.232	0.568	0.793
Hydrophobicity	-2.37	0.031	0.093
Order	-3.185	0.056	0.041

**Table E.4: Propensity of the amino acid insertion at DXE sites among secretory and non-secretory proteins:** Among the secretory and non-secretory proteins with a DXE motif, we analyzed to see how often X is one of the three O-phosphorylated amino acids (S, T, Y) or one of the three N-phosphorylated amino acids (H, R, K)

**Total data**

	Insertion X=S,T,Y,H,R,K	Other insertions
Secreted	37	10
Unsecreted	28	10

**Neurodegenerative and known UPS vs non-secretory**

	Insertion X=S,T,Y,H,R,K	Other insertions
Secreted	18	9
Unsecreted	23	5

**Cancer data from Groups 3B, 3C**

	Insertion X=S,T,Y,H,R,K	Other insertions
Secreted	10	0
Unsecreted	5	5

**HSP vs non-secretory**

	Insertion X=S,T,Y,H,R,K	Other insertions
Secreted	9	1
Unsecreted	23	5

**Cancer secretion data including HSP secretion (Groups 3A, 3B, 3C)**

	Insertion X=S,T,Y,H,R,K	Other insertions
Secreted	19	1
Unsecreted	5	5

- 15 proteins do not have DXE
- Only 85 proteins are considered for calculation.

**Table E.5: Propensity of the amino acid insertion at DEX/EDX sites among secretory and non-secretory proteins:** Among the secretory and non-secretory proteins with a DEX/EDX motif, we analyzed to see how often X is one of the three O-phosphorylated amino acids (S, T, Y) or one of the three N-phosphorylated amino acids (H, R, K)

**Total data**

	Insertion X=S,T,Y,H,R,K	Other insertions
Secreted	42	8
Unsecreted	29	10

**Neurodegenerative and known UPS vs non-secretory**

	Insertion X=S,T,Y,H,R,K	Other insertions
Secreted	23	6
Unsecreted	20	10

**Cancer data from Groups 3B, 3C**

	Insertion X=S,T,Y,H,R,K	Other insertions
Secreted	8	2
Unsecreted	9	0

**HSP vs non-secretory**

	Insertion X=S,T,Y,H,R,K	Other insertions
Secreted	11	0
Unsecreted	20	10

**Cancer secretion data including HSP secretion (Groups 3A, 3B, 3C)**

	Insertion X=S,T,Y,H,R,K	Other insertions
Secreted	19	2
Unsecreted	9	0

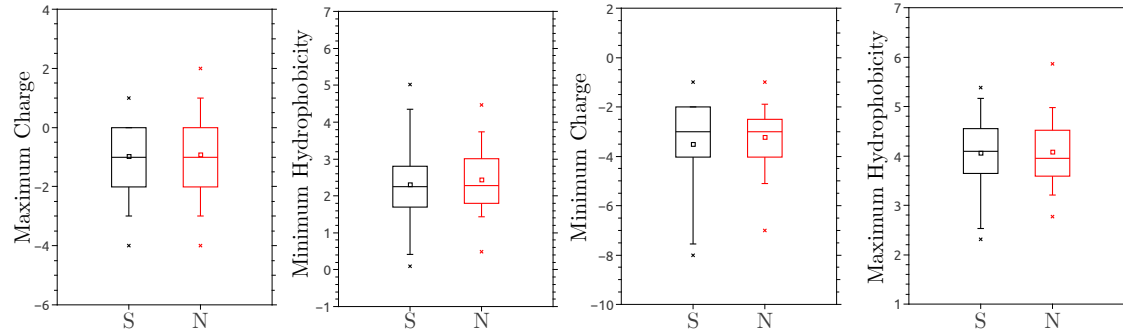
- 11 proteins do not have DE
- Only 89 proteins are considered for calculation.

**Table E.6:** Secondary structure of DE motif in secretory and non-secretory proteins. The DE motif is ordered (O) if they are in  $\alpha$ -helix or  $\beta$ -sheets and disordered (D) if they are in loop or missing residues. Separate cases has been considered if they are in between ordered and disordered

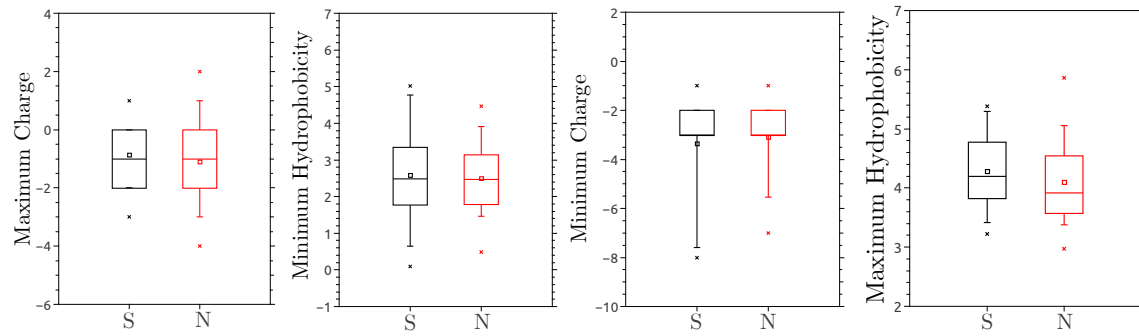
	Disordered is important & B is Disordered		Disordered is important & B is ordered		Ordered is important & B is Disordered		Ordered is important & B is ordered	
	O	D	O	D	O	D	O	D
<b>Total data</b>								
S	4	46	5	45	33	17	39	11
N	5	34	5	34	19	20	24	15
<b>Neurodegenerative and known UPS vs non-secretory</b>								
	O	D	O	D	O	D	O	D
S	2	27	3	26	20	9	23	6
N	5	25	5	25	14	16	19	11
<b>Cancer data from Groups 3B, 3C</b>								
	O	D	O	D	O	D	O	D
S	1	9	1	9	4	6	6	4
N	0	9	0	9	5	4	5	4
<b>HSP vs non-secretory</b>								
	O	D	O	D	O	D	O	D
S	1	10	1	10	9	2	10	1
N	5	25	5	25	14	16	19	11
<b>Cancer secretion data including HSP secretion (Groups 3A, 3B, 3C)</b>								
	O	D	O	D	O	D	O	D
S	1	10	1	10	9	2	10	1
N	5	25	5	25	14	16	19	11

**Figure E.1:** Comparison between the secretory and non-secretory proteins with their charges and hydrophobicity for various groups

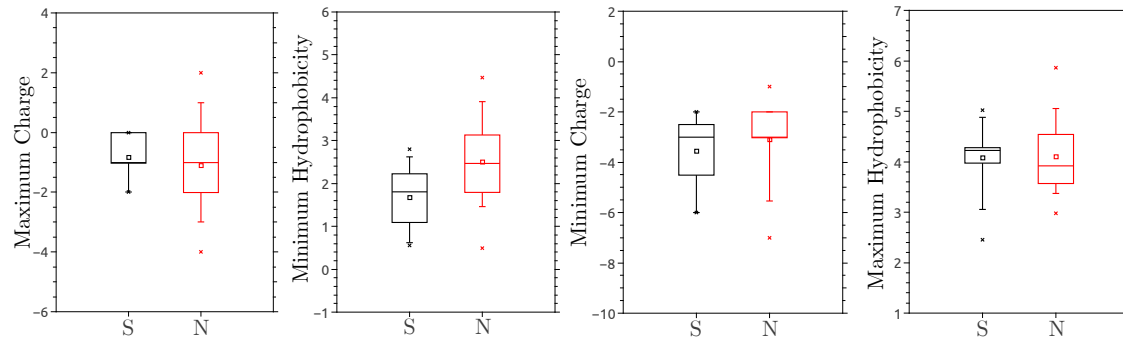
**Total data**



**Neurodegenerative and known UPS vs non-secretory**



**HSP vs non-secretory**



**Cancer secretion data including HSP secretion (Groups 3A, 3B, 3C)**

