DESIGNING NOVEL MEMBRANE ACTIVE PEPTIDES AND UNDERSTANDING THEIR MECHANISM OF ACTION

SAURABH SARASWAT

KUSUMA SCHOOL OF BIOLOGICAL SCIENCES

INDIAN INSTITUTE OF TECHNOLOGY DELHI

April 2024

DESIGNING NOVEL MEMBRANE ACTIVE PEPTIDES AND UNDERSTANDING THEIR MECHANISM OF ACTION

by

SAURABH SARASWAT

Kusuma School of Biological Sciences

Submitted

In fulfilment of the requirements of the degree of Doctor of Philosophy

to the

INDIAN INSTITUTE OF TECHNOLOGY DELHI

April 2024

CERTIFICATE

This is to certify that the thesis titled **"Designing novel membrane active peptides and understanding their mechanism of action",** being submitted by **Mr. Saurabh Saraswat** to the Indian Institute of Technology Delhi for the award of the degree of "**Doctor of Philosophy**" is a record of the bonafide research carried out by her, which has been prepared under our supervision and guidance in conformity with the rules and regulations of the Indian Institute of Technology Delhi, India. The results prescribed in it have not been submitted in part or full to any other University for the award of any degree or diploma.

Dr. Archana Chugh

Professor

Kusuma School of Biological Sciences

Indian Institute of Technology Delhi

New Delhi 110016, India

Date:

New Delhi

ACKNOWLEDGEMENTS

I am delighted to convey my sincere appreciation to everyone who has assisted me in navigating through the challenges of my PhD journey. Above all, I would like to extend my heartfelt gratitude to my PhD mentor, **Dr. Archana Chugh**, for her unwavering support throughout this academic expedition. I hold great respect for her and am truly grateful for her guidance, which has been a constant source of inspiration for excelling not only as a researcher but also as a compassionate individual. Her provision of laboratory facilities and essential research materials has been instrumental in my work, and I am profoundly thankful for the trust and confidence.

I extend my deep gratitude to the members of my SRC members, the chairperson, Prof. Vivekanandan Perumal, internal expert Prof. Bishwajit Kundu, and external expert Prof. Pramit Kumar Chowdhury, for their invaluable insights and suggestions during the progress presentations.

I would like to express my gratitude to the faculty members of KSBS for their generosity in providing me access to essential instruments and facilities crucial for my study. I would like to thank Prof. Tapan K Chaudhuri for facilitating the use of the fluorescence spectrometer and CD spectrometer.

I would like to extend my acknowledgment to the staff members at KSBS, namely Ms. Mini Sharma, Mr. Praveen, Mr. Pradeep, Mr. Vijaypal, Pushpalta Ma'am, and Garg Sir, for their invaluable support during Teaching Assistant (TA) duties and administrative proceedings. Their assistance has been instrumental in facilitating a smooth and efficient academic environment.

I am also grateful to IIT Delhi for providing financial assistance through the GATE fellowship and a conducive campus environment. Special thanks to IIT Delhi for supporting my travel to present my work at an international conference in Cologne University, Germany, in 2019.

I would like to express my gratitude to Prof. Gaurav Goel (Chemical Engineering) and Akshay Chenna for extending the opportunity to collaborate with them.

I extend my gratitude to Dr. Parvesh Wadhwani of KIT, Karlsruhe, Germany, for generously providing the peptides that played a significant role in advancing my research. His support has been invaluable to the completion of my work.

The enriching lab environment has been pivotal to the success of my journey. I thank Dr. Anupama for fostering a friendly atmosphere and guiding me into the lab. I would also like to appreciate the support from my lab mates, including Dr Nirupama Gangopadhyay, Dr Manobala, Dr Deepthi Poornima, Dr. Vivek Kumar, Dr. Harsha Rohira, Dr. Pankhuri Narula, Dr. Sujithra Shankar, Dr. Anjali Priya, Saiguru S., Prasanjeet Kaur, Aditi Arora, Gagan Kaur, Malay Ranjan Nayak, Mayank Sinha, Jyoti Sood, A. Sivalingam, Koyena Kundu.

I extend my special appreciation to Central Research Facility, Department of Textile Technology and Mr. Kuldeep for generously providing access to the Scanning Electron Microscopy facility.

With immense love, I want to express my heartfelt thanks to my family for their unwavering support throughout my tenure. I am eternally grateful to my parents, who not only believed in my work but also took joy in it as much as I did. Their unwavering belief in me always

remained be my greatest source of strength, serving as the inspiration behind every achievement in my life.

Finally, I want to thank myself for maintaining self-belief and not allowing adversities to deter me from completing this challenging journey.

Saurabh Saraswat

ABSTRACT

Membrane active peptides (MAPs) have attracted considerable attention due to their impressive ability to interact with lipid bilayers, holding great promise for revolutionary applications in the field of medicine. This thesis conducts a comprehensive exploration of MAPs, focusing primarily on cell-penetrating peptides (CPPs) and antimicrobial peptides (AMPs). It unravels their intricate design, cellular uptake mechanisms, and potential applications in targeted drug delivery.

In the first section, the research concentrates on developing a novel CPP, Engraulisin, sourced from marine origins. This peptide demonstrates effective cellular penetration and cargo delivery without significant cytotoxicity, particularly showcasing uptake in HeLa cells at a concentration of 5uM. Notably, Engraulisin exhibits potent antimicrobial properties against MRSA at a concentration of 10uM, indicating its potential for addressing challenging infections. The section concludes by delving into the fundamental mechanisms underlying CPP's membrane penetration, emphasizing electrostatic interactions, hydrophobic forces, and specific peptide sequences.

The second section investigates novel short Latarcin-derived peptides (sLtcs) chosen for their unique characteristics as potential antimicrobial agents. Systematic experiments reveal sLtc 4 and 5 as promising candidates, displaying broad-spectrum antimicrobial activity at micromolar concentrations, both at the membrane and intracellular levels. The antimicrobial potential and mechanism were analysed, revealing multi-target activities, including changes in membrane potential, ROS production, and membrane disruption. Compared to parent sequences, the pattern

of activity also appeared to change considerably, enhancing our understanding of sLtcs multifaceted antimicrobial capabilities.

The final chapter introduces a novel cell-penetrating prediction tool developed through large language models and machine learning techniques. Recognizing the crucial need for accurate prediction of peptide cell-penetrating capabilities, CPPsite-1 and CPPsite-3 are presented as predictive models capable of generalizing well to diverse peptide characteristics. Rigorous testing and validation ensure the tool's accuracy, providing a reliable method for predicting the cell-penetrating potential of peptides.

Overall, these objectives collectively advance the field by introducing novel peptides like Engraulisin and sLtcs, showcasing their potential in drug delivery and antimicrobial applications. The innovative predictive tool contributes to streamlining the identification of promising candidates, paving the way for the rational design of peptides with enhanced cellular uptake capabilities.

सार

मेम्ब्रेन सक्रिय पेप्टाइड्स (एमएपी) ने लिपिड बाइलेयर्स के साथ इंटरैक्ट करने की अपनी प्रभावशाली क्षमता के कारण काफी ध्यान आकर्षित किया है, जो चिकित्सा के क्षेत्र में क्रांतिकारी अनुप्रयोगों के लिए बड़ा वादा है। यह थीसिस मुख्य रूप से सेल-पेनेट्रेटिंग पेप्टाइड्स (सीपीपी) और रोगाणुरोधी पेप्टाइड्स (एएमपी) पर ध्यान केंद्रित करते हुए एमएपी की व्यापक खोज करती है। यह उनके जटिल डिजाइन, सेलुलर अपटेक तंत्र और लक्षित दवा वितरण में संभावित अनुप्रयोगों को उजागर करता है।

पहले खंड में, अनुसंधान समुद्री मूल से प्राप्त एक नवीन सीपीपी, एंगरौलीसिन विकसित करने पर केंद्रित है। यह पेप्टाइड महत्वपूर्ण साइटोटॉक्सिसिटी के बिना प्रभावी सेलुलर प्रवेश और कार्गो डिलीवरी को प्रदर्शित करता है, विशेष रूप से 5uM की एकाग्रता पर हेला कोशिकाओं में अवशोषण को प्रदर्शित करता है। विशेष रूप से, एंगरौलीसिन 10uM की सांद्रता पर एमआरएसए के खिलाफ शक्तिशाली रोगाणुरोधी गुण प्रदर्शित करता है, जो चुनौतीपूर्ण संक्रमणों को संबोधित करने की इसकी क्षमता को दर्शाता है। यह खंड सीपीपी की झिल्ली पैठ के अंतर्निहित मूलभूत तंत्रों पर प्रकाश डालकर, इलेक्ट्रोस्टैटिक इंटरैक्शन, हाइड्रोफोबिक बलों और विशिष्ट पेप्टाइड अनुक्रमों पर जोर देकर समाप्त होता है।

दूसरा खंड संभावित रोगाणुरोधी एजेंटों के रूप में उनकी अनूठी विशेषताओं के लिए चुने गए उपन्यास लघु लैटरसिन-व्युत्पन्न पेप्टाइड्स (एसएलटीसी) की जांच करता है। व्यवस्थित प्रयोगों से एसएलटीसी 4 और 5 को आशाजनक उम्मीदवारों के रूप में पता चलता है, जो झिल्ली और इंट्रासेल्युलर दोनों स्तरों पर माइक्रोमोलर सांद्रता में व्यापक स्पेक्ट्रम रोगाणुरोधी गतिविधि प्रदर्शित करते हैं। रोगाणुरोधी क्षमता और तंत्र का विश्लेषण किया गया, जिसमें झिल्ली क्षमता, आरओएस उत्पादन और झिल्ली व्यवधान में परिवर्तन सहित बहु-लक्ष्य गतिविधियों का खुलासा हुआ। मूल अनुक्रमों की तुलना में, गतिविधि का पैटर्न भी काफी बदलता हुआ दिखाई दिया, जिससे एसएलटीसी की बहुआयामी रोगाणुरोधी क्षमताओं के बारे में हमारी समझ में वृद्धि हुई।

अंतिम अध्याय बड़े भाषा मॉडल और मशीन सीखने की तकनीकों के माध्यम से विकसित एक उपन्यास सेल-मर्मज्ञ भविष्यवाणी उपकरण का परिचय देता है। पेप्टाइड सेल-मर्मज्ञ क्षमताओं की सटीक भविष्यवाणी की महत्वपूर्ण आवश्यकता को पहचानते हुए, सीपीपीसाइट-1 और सीपीपीसाइट-3 को विभिन्न पेप्टाइड विशेषताओं को अच्छी तरह से सामान्य बनाने में सक्षम पूर्वानुमानित मॉडल के रूप में प्रस्तुत किया गया है। कठोर परीक्षण और सत्यापन उपकरण की सटीकता सुनिश्चित करते हैं, पेप्टाइड्स की कोशिका-मर्मज्ञ क्षमता की भविष्यवाणी करने के लिए एक विश्वसनीय तरीका प्रदान करते हैं।

कुल मिलाकर, ये उद्देश्य सामूहिक रूप से एंगरौलीसिन और एसएलटीसी जैसे नवीन पेप्टाइड्स को पेश करके, दवा वितरण और रोगाणुरोधी अनुप्रयोगों में अपनी क्षमता का प्रदर्शन करके क्षेत्र को आगे बढ़ाते हैं। नवोन्मेषी पूर्वानुमानित उपकरण आशाजनक उम्मीदवारों की पहचान को सुव्यवस्थित करने में योगदान देता है, जिससे बढ़ी हुई सेलुलर ग्रहण क्षमताओं के साथ पेप्टाइड्स के तर्कसंगत डिजाइन का मार्ग प्रशस्त होता है।

TABLE OF CONTENTS

Title	Page No.
Certificate	i
Acknowledgements	ii
Abstract	v
Table of contents	vii
List of Figures	xi
List of Tables	xv
Abbreviations	xvi
Symbols	xxi
Chapter 1, Introduction and Objectives	**1-24**
1.1 Membrane active peptides	2
1.2 Cell Penetrating Peptides	3
1.2.1 Classification of cell penetrating peptides	4
1.2.1.1 Cationic CPPs	4
1.2.1.2 Primary amphipathic	5
1.2.1.3 Secondary amphipathic	6
1.2.2 Mechanisms of. CPPs Internalization	7
1.2.2.1 Direct translocation	8
1.2.2.2 Uptake through endocytosis	10
1.3 Antimicrobial Peptides	13
1.3.1 Classification of AMPs	14
1.3.1.1 Based on sources	15

1.3.1.2 Based on activity 17

1.3.2 AMPs as membrane targeting agents 19

1.3.3 Bioactive antimicrobial peptides 20

1.4 Objectives 23

Chapter 2, Review of Literature 25-39

2.1 Application of CPPs 26

2.1.1 CPP in medical imaging 26

2.1.2 CPP-assisted siRNA downregulation 27

2.1.3 CPPs for anticancer drug delivery 28

2.1.4 CPP-functionalized nanocarriers for intracellular delivery 29

2.2 Marine and toxins are a rich source of membrane active peptides 31

2.2.1 Marine derived peptides 32

2.2.2 Toxin derived peptides 34

2.2.2.1 Snake venom derived peptides 34

2.2.2.2 Scorpion peptides 35

2.2.2.3 Spider toxin derived peptides 36

2.2.2.3.1 Latarcins 36

2.3 Machine learning for designing of novel CPPs 37

Chapter 3, Designing a novel cell penetrating peptide from a marine source 40-63

3.1 Introduction 41

	3.2	Material and methods	43
	3.3	Results	50
	3.4	Discussion	60
	3.5	Conclusion	63
Chapter 4,	**Understanding the antimicrobial action of short Latarcin peptides (sLtcs)**		**64-78**
	4.1	Introduction	65
	4.2	Material and methods	69
	4.3	Results	71
	4.4	Discussion	76
	4.5	Conclusion	80
Chapter 5,	**Designing Cell penetrating peptide prediction tool with Machine learning**		**81-105**
	5.1	Introduction	82
	5.2	Material and methods	84
	5.3	Results	91
	5.4	Discussion	104
	5.5	Conclusion	107
Summary			**108-111**
References			**112-134**
Annexures			**135-142**
	Annexure 1	List of chemicals	134
	Annexure 2	Preparation of buffers	137

List of publications and conferences **143-144**

Author's resume **145-148**

LIST OF FIGURES

Figure No.	Title	Page No.
1.1	Schematic representation of mechanism of internalisation by cell penetrating peptides.	4
1.2	Schematic representation of mechanism of action of antimicrobial peptides.	14
3.1	Cellular uptake of Engraulisin peptide through (a) flow cytometry and (c) confocal laser scanning micrograph. Image (a) - i, ii, iii and iv represent fluorescence shift at 5, 10, 15 and 20μM peptide concentration. (b) depicts the uptake percentage of the population	51
3.2	Cellular cytotoxicity of Engraulisin peptide investigated in HeLa cells by MTT assay. No significant reduction in cell viability was observed till 100 μM concentration of peptide treatment.	52
3.3	Cargo delivery and mechanistic characterization of peptide. Cargo delivery (a) analysed through X-Gal staining under various treatment conditions. (a); a-d represents untreated, Engraulisin/cargo complex, tat2/cargo complex (positive control), and β-galactosidase only respectively. (b) Engraulisin uptake in presence of inhibitors of endocytosis. Significant reduction was observed in case of only 4^0C as physical inhibitor of endocytosis.	53
3.4	Circular dichroism analysis of Engraulisin peptide in the presence of water, SDS micelles and 2,2,2 trifluoroethanol.	55

3.5	Membrane lytic effect of Engraulisin peptide on MRSA. A. untreated cells. B and C show effect of 10 and 20 μM peptide concentration on cell morphology.	56
3.6	Peptide in water simulation; (a) transition of peptide from alpha helix at 0 ns (b) to random coil conformation at 100 ns can be observed in aqueous environment, (c) Secondary structure transformation over the course of simulation was calculated through the DSSP program	57
3.7	Peptide membrane interaction in POPC and POPG membrane over a course of 100 ns. Left, middle and right columns depict snapshots captured at 0, 50 and 100 ns respectively. POPC (upper) and POPG (lower) depict mammalian and bacterial membranes respectively	58
3.8	(a) Average density and (b) distance plot of phosphate atoms, n-terminal (RLFR) and c-terminal (VLRL) residues across Z-axis in POPC and POPG membrane. Both Density and distance profile of n-terminal residues was found to be higher at the deeper selection of the membrane as compared to the c-terminal indicating peptide interaction though n-terminal residues.	59
3.9	Residue wise distance heatmap of N-terminal and C-terminal residues from phosphate atoms along z-axis in (a) POPC and (b) POPG membrane. Blue and red colour indicates higher and lower distance between amino acid residue and membrane phosphate atoms.	60

4.1	sLtcs mediated analysis of concentration-dependent membrane depolarization in MRSA cells.	73
4.2	sLtcs induced membrane permeabilization analysis using PI uptake	74
4.3	Time kill kinetics studies of sLtc 4 and 5 on MRSA cells	75
4.4	ROS production in MRSA cells by sLtcs	76
5.1	Comparison between CPPsite-1 and CPPsite-3 datasets. (a) showcases variances in the amino acid composition within both datasets. (b) demonstrates the relative distribution of different physicochemical characteristics among positive and negative examples and between the two datasets under investigation.	83
5.2	The architectural framework of the proposed model consists of four blocks: data collection, feature extraction, classification and uncertainty estimation. Data collection and feature extraction represent Module 1, while Module 2 comprises prediction and uncertainty estimation blocks.	86
5.3	Violin plot showing the effect of different feature extractors (a) and model architecture (b) on model performance.	90
5.4	Effect of training dataset size on model performance.	94
5.5	Uncertainty thresholding based on the J index. (a) represents thresholding using J (youden's) index on test dataset. (b) zoomed image for threshold representation. (c) represents density estimate with positive in blue and negative (non-cpps) in red separated by dotted threshold line.	97

5.6 Uncertainty-aware estimation of model performance on cppsite-1 and cppsite-3 dataset. Training dataset size is reprented on x-axis with accuracy on y-axis. Accuracy with high confidence(red), without uncertainty(green) and low-confidence(blue) are compared showing increase in cpp prediction accuracy due to thresholding in both datasets. 98

5.7 Comparison of CPP prediction accuracy on CPPsite-1 and CPPsite-3 datasets using different parameters. Each circle corresponds to an individual model configuration, with the color indicating the quantity of hidden layers and the circle's size reflecting the number of neurons. 101

5.8 Violin plot showing comparison of CNN-based model architecture on model performance on both cppsite-1 and cppsite-3 dataset. 101

5.9 Representative TSNE representation of CPPsite-1 dataset from the pre-trained model. 102

LIST OF TABLES

Table No.	Title	Page No.
3.1	Detailed information regarding components	50
3.2	MIC concentration of Engraulisin peptide against various microorganism	55
4.1	List of Latarcin peptides found in the venom of *Lachesana tarabaevi*	68
4.2	List of short Latarcin peptides used in this study.	68
4.3	Minimal Inhibitory concentration of sLtc against different microbes.	72
5.1	Configuration details of various pre-trained models used in this study.	84
5.2	Embedding-wise mean percent scores.	91
5.3	Model-wise mean percent mean scores.	92-93
5.4	Uncertainty-aware estimation of model performance.	95-97
5.5	Comparison of the proposed method with previous state-of-the-art methods.	99-100
5.6	Top 10 final performers segregated based on CPPsite-1 and CPPsite-3 datasets.	100

ABBREVIATIONS

1D	One Dimensional
1D-CNN	one-Dimensional Convolutional Neural Networks
ACC	Accuracy Correlation Coefficient
ACPs	Anticancer Peptides
APD3	Antimicrobial Peptide Databases
AMP	Antimicrobial Peptides
ATR-FTIR	Attenuated Total Reflection Fourier Transform Infrared Spectroscopy
AUC-ROC	Area Under the Receiver Operating Characteristic Curve
BERT	Bidirectional Encoder Representations from Transformers
CAMP	Collection of Antimicrobial Peptides
CCR2	Chemokine Receptor type 2
CD	Circular Dichroism
CFU	Colony Forming Unit
CLSM	Confocal Laser Scanning Microscopy
CME	Clathrin-Mediated Endocytosis
CPP	Cell Penetrating Peptide
CSαβ	Cysteine-Stabilized alpha/beta
DBAASP	Database of Antimicrobial Activity And Structure of Peptide
dCTP	Deoxycytidine Triphosphate
dGTP	2'-Deoxyguanosine 5'-Triphosphate
DISC 3	3,3′-Dipropylthiadicarbocyanine iodide

DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic Acid
DSSP	Database of Secondary Structure Assignments
EDTA	Etheylene Diamine Tetraacetic Acid
ERT	Extremely Randomized Tree
ESM	Evolutionary Scale Modeling
F	Phenylalanine
FACS	Fluorescence Activated Cell Sorting
FBS	Fetal Bovine Serum
FDA	Food and Drug Administration
FITC	Fluorescein isothiocyanate
fGNPs	Functionalized GNPs
FGS	Fluorescence-Guided Surgery
GNPs	Gold Nanoparticles
HATF-3	Hepatocyte ATF3
HβD3	Human β-Defensin 3
H_2DCF	2′,7′-dichlorodihydrofluorescein
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)
HIV	Human immunodeficiency viruses
HSPG	Heparan Sulfate Proteoglycan
hTERT	Human Telomerase Reverse Transcriptase
IC50	Half-maximal inhibitory concentration

KELM	Kernel Extreme Learning Machine
LanCL1	Lanthionine Synthetase Component C-like protein
LSTM	Long Short-Term Memory
Ltc	Latarcin
MAP	Membrane Active Peptides
MCC	Matthews Correlation Coefficient
MIC	Minimum Inhibitory Concentration
MD	Molecular Dynamic
MDR	Multidrug Resistant
MHB	Mueller Hinton Broth
ML	Machine Learning
MMDP	Marine Metagenome-Derived Peptide
MMP-2	Matrix Metalloproteinase-2
MO	Morpholino
MRSA	Methicillin-Resistant *Staphylococcus Aureus*
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
NF-κB	Nuclear Factor k-Light-Chain-Enhancer of Activated B cells
NLSs	Nuclear localization sequences
NN	Neural Network
Oct-6	**Optical Coherence Tomography *6***
PBS	Phosphate Buffer Saline
PCA	Principal Component Analysis
PDB	Potato Dextrose Broth

pDNA	Plasmid DNA
pH	Potential of Hydrogen
PI	Propidium Iodide
PME	Particle Mesh Ewald
POPC	2-oleoyl-1-pamlitoyl-sn-glyecro-3-phosphocholine
POPG	2-oleoyl-1-pamlitoyl-sn-glyecro-3-glycerol
Pro	Proline
R	Arginine
RACPPs	Ratiometric Activable CPPs
RF	Random Forest
RNA	Ribonucleic Acid
ROS	Reactive Oxygen Species
SDC3	Syndecan 3
SDS	Sodium Dodecyl Sulfate
siRNA	Small Interfering RNAs
sLtcs	short Latarcin peptides
SMA	Spinal Muscular Atrophy
SP	Specificity
SV40	Simian Virus 40
SVM	Support Vector Machine
TAT2	*Transactivating transcriptional activator 2*
TEM	Transmission Electron Microscopy
TFE	2,2,2 Trifluoroethanol

TFIIE-beta	Transcription Initiation Factor IIE subunit beta
TP10	Transportan
t-SNE	t-Distributed Stochastic Neighbor Embedding
VRE	Vancomycin-Resistant *Enterococcus*
W	Tryptophan
YADAMP	Yet Another Database of Antimicrobial Peptides

SYMBOLS

α	alpha
β	Beta
˚C	Degree Celsius
cm	Centimeter
fs	Femtoseconds
hr	Hour/ Hours
g	RCF
k	Kappa
K	Kilo
kJ	Kilo Joule
kDa	Kilodalton
µg	Microgram
µl	Microliter
µm	Micrometer
µM	Micromolar
ml	Milliliter
mg	Milligram
mM	Millimolar
min	Minute/Minutes
ng	Nanogram
nm	Nanometer

nM	Nanomolar
ns	Nanosecond
%	Percent
ps	Picosecond
rpm	Revolutions per minute
sec	Second/seconds
v/v	Volume by volume