CAPITAL UNIVERSITY OF SCIENCE AND TECHNOLOGY, ISLAMABAD

Computational Screening of Scorpion Venom Antimicrobial Peptides as Vaccine Adjuvants for Ovarian Cancer

by

Sehar Sajjad

A thesis submitted in partial fulfillment for the degree of Master of Science

in the

Faculty of Health and Life Sciences
Department of Bioinformatics and Biosciences

2026

CERTIFICATE OF APPROVAL

Computational Screening of Scorpion Venom Antimicrobial Peptides as Vaccine Adjuvants for Ovarian Cancer

by

Sehar Sajjad

(MBS251009)

THESIS EXAMINING COMMITTEE

S. No.	Examiner	Name	Organization
(a)	External Examiner	Dr. Rabbiah Manzoor Malik	UoW, Wah Cantt
(b)	Internal Examiner	Dr. Sami Ullah Jan	CUST, Islamabad

Dr. Sahar Fazal
Thesis Supervisor
July, 2026

Dr. Syeda Marriam Bakhtiar
Head
Dept. of Bioinfo. and Biosciences
July, 2026

Dr. Sahar Fazal
Dean
Faculty of Health and Life Sciences
July, 2026

Author's Declaration

I, **Sehar Sajjad** hereby state that my MS thesis titled "**Computational Screening of Scorpion Venom Antimicrobial Peptides as Vaccine Adjuvants for Ovarian Cancer**" is my own work and has not been submitted previously by me for taking any degree from Capital University of Science and Technology, Islamabad or anywhere else in the country/abroad.

At any time if my statement is found to be incorrect even after my graduation, the University has the right to withdraw my MS Degree.

(Sehar Sajjad)

Registration No: MBS251009

Plagiarism Undertaking

I solemnly declare that research work presented in this thesis titled "**Computational Screening of Scorpion Venom Antimicrobial Peptides as Vaccine Adjuvants for Ovarian Cancer**" is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and Capital University of Science and Technology towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS Degree, the University reserves the right to withdraw/revoke my MS degree and that HEC and the University have the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized work.

(Sehar Sajjad)

Registration No: MBS251009

Acknowledgement

First and foremost, I am profoundly grateful to Almighty Allah, the Most Gracious and the Most Merciful, for His countless blessings, guidance, strength, and mercy throughout my academic journey. Without His divine grace, this achievement would not have been possible.

I dedicate this achievement to my beloved parents, **Sajjad Ali** and **Shabana Kausar**, and my dear brother, **Huzaifa Ahmed**, whose unconditional love, endless prayers, and unwavering support have been my greatest source of strength. Their belief in me has always inspired me to achieve my goals.

I would like to express my sincere and deepest gratitude to my respected supervisor, **Prof. Dr. Sahar Fazal**, whose exceptional guidance, invaluable advice, continuous encouragement, and unwavering support made this research possible. Her insightful suggestions, constructive criticism, and patience throughout every stage of this study greatly enhanced the quality of my work. I am truly grateful for her mentorship and confidence in my abilities.

I am also thankful to the Head of the Department, **Dr. Syeda Mariam Bakthiar**, for providing a supportive academic environment. My sincere appreciation also goes to all the faculty members for sharing their knowledge and expertise.

I would like to extend my heartfelt appreciation to **Miss Shah Gull** for her sincere assistance and valuable support during my research work. I am especially grateful to **Muhammad Yahya**, whose constant encouragement, patience, and unwavering belief in my abilities motivated me to overcome challenges throughout this journey. His moral support and willingness to help whenever needed made a significant difference during the completion of this research.

I am also thankful to all my friends and to everyone whose support, encouragement, and companionship made this journey more memorable. My heartfelt gratitude goes to my family for their endless prayers, love, and unwavering belief in me.

(Sehar Sajjad)

Abstract

Ovarian cancer remains difficult to treat due to the weak and short-lived immune responses generated by current therapies. Targeted peptide vaccines have been developed against tumour-associated antigens like NY-ESO-1, MUC16 and folate receptor alpha, but they demand the use of strong adjuvants to increase the immunogenicity. A seven-phased *in silico* pipeline was used for screening scorpion venom-derived non-disulfide-bridged peptides (NDBPs) as possible TLR4/MD-2-targeting adjuvants. We obtained 117 antimicrobial peptide sequences from 31 genera of *scorpions* from APD3 and DRAMP, characterized them physicochemically using ProtParam and structurally modelled with AlphaFold to get 75 amphipathic and cationic NDBPs for downstream analysis. 34 strong binders ($\leq$ -13.0 kcal/mol) were identified by molecular docking with TLR4/MD-2 (PDB: 3FXI), and STRING analysis showed that both the MyD88 and TRIF signalling arms, which are crucial for anti-tumour immunity, were engaged. Multi-parameter safety screening (cytotoxicity, hemolysis, allergenicity, stability) left 12 peptides, three of which were redesigned for greater stability by using the ProteinMPNN and resulted in 5 non-toxic, protease-resistant mutants from the DRAMP18681 scaffold. A 200 ns molecular dynamics simulation proved the stability of binding of the lead peptide to TLR4/MD-2, and the subsequent antigen-interference screening using the HDOCK software identified seq_57 (DKGGVRDPRV) as the best candidate with minimal antigen interference, structural stability and a good safety profile. These results create a computational strategy for the discovery and design of peptide adjuvants for cancer vaccines derived from scorpion venom. The candidate that is selected, seq_57, has good receptor engagement, structural stability and compatibility with vaccine antigens, therefore it is encouraged to be tested in vitro using immunogenicity and cytotoxicity assays.

Keywords: Ovarian cancer, Scorpion venom peptides, Antimicrobial peptides (AMPs), Vaccine adjuvants, Toll-like receptor 4 (TLR4), Computational screening, Molecular docking, Cancer immunotherapy.

Contents

List of Figures

List of Tables

Abbreviations

aa	Amino acids
AI	Aliphatic index
AMP	Antimicrobial peptide
APD3	Antimicrobial Peptide Database 3
BRCA	Breast cancer gene (BRCA1 / BRCA2)
CA-125	Cancer antigen 125 (serum biomarker for ovarian cancer)
CAR-T	Chimeric antigen receptor T-cell therapy
CTL	Cytotoxic T lymphocyte
Da	Dalton
DBP	Disulfide-bridged peptide
DC	Dendritic cell
ΔG	Gibbs free energy of binding (kcal/mol)
DNA	Deoxyribonucleic acid
DRAMP	Data Repository of Antimicrobial Peptides
EC50	Half-maximal effective concentration
ELISA	Enzyme-linked immunosorbent assay
FFT	Fast Fourier transform
Fmoc	Fluorenylmethyloxycarbonyl (solid-phase peptide synthesis chemistry)
FRα	Folate receptor alpha
GLOBOCAN	Global Cancer Observatory (IARC, WHO)
GRAVY	Grand average of hydropathicity
HDOCK	Hybrid docking algorithm server (protein–protein docking)
HGSOC	High-grade serous ovarian carcinoma

HRD	Homologous recombination deficiency
IFN	Interferon (type I: IFN-α/β; type II: IFN-γ)
II	Instability index
IL	Interleukin
IRAK	Interleukin-1 receptor-associated kinase
IRF3	Interferon regulatory factor 3
IRF7	Interferon regulatory factor 7
LPS	Lipopolysaccharide
MD	Molecular dynamics
MD-2	Myeloid differentiation protein 2 (TLR4 co-receptor)
MDSC	Myeloid-derived suppressor cell
MHC	Major histocompatibility complex
MUC16	Mucin-16 (the CA-125 antigen)
MW	Molecular weight
MyD88	Myeloid differentiation primary response protein 88
NaCl	Sodium chloride
NDBP	Non-disulfide-bridged peptide
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NPT	Isothermal–isobaric ensemble (constant N, P, T)
NY-ESO-1	New York esophageal squamous cell carcinoma 1 antigen
OC	Ovarian cancer
OPLS4	Optimized potentials for liquid simulations, version 4
PARP	Poly(ADP-ribose) polymerase
PBMC	Peripheral blood mononuclear cell
PDB	Protein Data Bank
pI	Theoretical isoelectric point
PRR	Pattern recognition receptor
RCSB	Research Collaboratory for Structural Bioinformatics
Rg	Radius of gyration
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
SPPS	Solid-phase peptide synthesis

STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
TAK1	Transforming growth factor β-activated kinase 1
TBK1	TANK-binding kinase 1
Th1	T helper 1 cell subset (IFN-γ and IL-12 producing)
Th2	T helper 2 cell subset (IL-4 and IL-13 producing)
TIP3P	Transferable intermolecular potential 3 point (water model)
TIRAP	Toll–interleukin-1 receptor (TIR) domain-containing adaptor protein
TLR4	Toll-like receptor 4
TME	Tumour microenvironment
TNF-α	Tumour necrosis factor alpha
TRAF6	TNF receptor-associated factor 6
TRAM	TRIF-related adaptor molecule
Treg	Regulatory T cell
TRIF	TIR-domain-containing adaptor-inducing IFN-β
UKCTOCS	UK Collaborative Trial of Ovarian Cancer Screening
WHO	World Health Organization
WT1	Wilms' tumour protein 1 antigen

Chapter 1

Introduction

Ovarian cancer is a particularly serious form of gynaecological malignancy with the highest case fatality ratio of all the other gynecological cancers. It is the eighth most prevalent cancer worldwide and the eighth leading cause of cancer death in women with an estimated 324,000 new cases and 206,000 deaths occurring every year [1]. Although there has been much improvement in surgical technique and cytotoxic chemotherapy over the past 30 years, the overall 5-year survival rate for advanced disease has remained less than 30% and there has been little meaningful improvement over the past 30 years [2].

There are several underlying explanations for this poor prognosis, as ovarian cancer is not a single disease but a collection of histological subtypes: high grade serious, clear cell, endometrioid and mucinous carcinomas have different molecular driver alterations, immunological profiles and chemosensitivity characteristics, all of which make them resistant to a single therapy. The standard of care treatment for epithelial ovarian cancer consists of maximum cytoreductive surgery combined with platinum-taxane based chemotherapy, which has been known since the early 1990s. Although response rates to first-line platinum-based therapy are 70-80%, most patients will have a recurrence within 12-18 months, and the response to salvage therapy will be progressively more transient and incomplete [3].

New advances such as maintenance therapy with poly(ADP-ribose) polymerase (PARP) inhibitors have improved progression-free survival in subgroups of patients with BRCA mutations, while antiangiogenic drugs like bevacizumab offer modest improvement in progression-free survival in recurrent disease. However, none of these approaches has any curative value in the metastatic context and have significant toxicity profiles. One of the basic biological challenges to persistent disease control is the immunosuppressive tumor microenvironment (TME) of ovarian cancer that is populated with regulatory T cells (Tregs), tumor-associated macrophages that are polarized to the M2 phenotype, and myeloid-derived suppressor cells (MDSCs). All of which create an immunological landscape that actively suppresses endogenous anti-tumor immunity and that limits the effectiveness of exogenous immunotherapeutic approaches [4].

Non-availability of early and reliable diagnostic modalities is a hallmark of clinical challenge in ovarian cancer. There is little clinical evidence in the early stages and the combination of symptoms in the later stages, which include abdominal distention, pressure in the pelvis, early satiety and changes in bowel or bladder pattern is not specific enough to be recognized and diagnosed for months.

The screening process, whether based on serum CA-125 level or on transvaginal ultrasonography or a combination of both, is not sufficiently sensitive or specific for population level screening. The UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS) that included more than 200,000 postmenopausal women showed no significant difference in mortality between those who received annual multimodal screening and those who did not.

As a result, about 70–80% of all ovarian cancers are diagnosed at Stage III or IV when there are already metastases in the peritoneum and the chances of achieve surgical complete resection are greatly reduced. The structural limitations of low cancer registry coverage and accessibility of molecular diagnostics in a clinical setting in Pakistan, with the age-standardized incidence of ovarian cancer being around 6.2 per 100,000 women, and a relatively young median age of 50-55 years, amplify the urgency of the need for computationally driven and cost-effective therapeutics for ovarian cancer [2][5].

This immunologic reactivity of ovarian tumors that contains a subpopulation of tumor infiltrating lymphocytes (TILs), whose density is strongly associated with survival, has sustained the interest in therapeutic interventions that could harness this latent antitumor immune activity for therapeutic benefit. Ovarian-cancer-associated antigens like NY-ESO-1, WT1, MUC1, and mutant p53 are now being targeted in strategies for cancer vaccines, and peptide-antigen formulations are now in early-phase clinical trials with measurable antigen-specific cytotoxic T lymphocyte (CTL) responses in immunologically receptive patient subsets. But the intensity and persistence of the responses has been found to be inadequate for any clinically significant benefit in most of the patients [6].

Immune checkpoint blockade using anti-PD-1 and anti-PD-L1 therapies has been transformative in melanoma, lung cancer and microsatellite-unstable colorectal cancers, but has demonstrated objective response rates from 6-15% in unselected populations of ovarian cancer, suggesting that the complex and multidimensional immunosuppressive architecture of the ovarian TME is beyond the PD-1/PD-L1 axis dysregulation. These clinical findings all suggest that external immunological amplification is necessary to make the immune-infiltrated but immunosuppressed ovarian tumor an immunologically responsive target via a mechanism-specific potent adjuvant [6].

Adjuvants are immunological stimulants that cause innate immune pattern recognition receptors (PRRs) on antigen-presenting cells (APCs), especially dendritic cells (DCs), to activate the presentation of a soluble antigen leading to a strong adaptive immune response with memory. The immunological requirements can be very different between cancer vaccines and infectious disease vaccines for prevention. The antigens to which the vaccine would aim to induce an immune response against cancer are, in most cases, self-derived proteins which have been the subject of central and peripheral tolerance mechanisms, and the adjuvant will need to generate a danger signal of a magnitude and quality sufficient to break this tolerance and to prime tumour specific CTL responses. The licensed adjuvants, such as alum (aluminium salt preparations) and the AS01B system in the approved malaria and shingles vaccines, are generally of type Th2 and promote the

production of high-titer antibody responses, not the type of Th1, including interferon gamma (IFN-γ) and interleukin 12 (IL-12), required for effective CD8+ CTL priming of intracellular tumour antigens [7]. A major translational need to this study is the lack of a clinically approved adjuvant that can consistently induce Th1 polarized CTL response in the right antigenic context for ovarian cancer [7].

The antimicrobial peptides (AMPs), evolutionarily old host defense molecules consisting of 10–50 amino acid residues, are emerging as a structurally and functionally diverse class of immunomodulatory molecules that have been identified over the past 20 years. AMPs are now known to have pleiotropic effects on both innate and adaptive immune compartments that go beyond direct antimicrobial cytotoxicity, and are characterized by their ability to disrupt microbial cell membranes via electrostatic and hydrophobic interactions with anionic phospholipid bilayers. Cationic AMPs act as endogenous danger signals capable to activate Toll-like receptors (TLRs) and other PRRs, recruit and mature DCs, modulate cytokine secretion profiles and enhance antigen cross-presentation to naive CD8+ T cells at the molecular level. This array of mechanisms from activation of PRR to maturation of the DC and polarization of the macrophage and priming of the CTL makes AMPs a structurally compact, sequence programmable scaffold for the design of next-generation cancer vaccine adjuvants [8].

Among the PRRs relevant to cancer vaccine adjuvancy, Toll-like receptor 4 (TLR4) occupies a position of particular strategic importance. TLR4, signaling through its obligate co-receptor MD-2, recognizes the lipid A component of gram-negative bacterial lipopolysaccharide (LPS) and initiates a bifurcated downstream signaling cascade: the MyD88-dependent arm drives NF-κB-mediated transcription of pro-inflammatory cytokines including TNF-α, IL-6, and IL-12p70, while the TRIF-dependent arm activates interferon regulatory factor 3 (IRF3) to induce type I interferon (IFN-α/β) secretion [8].

The simultaneous induction of IL-12 and type I IFNs the molecular hallmarks of Th1 immune polarization is precisely the cytokine signature required for effective DC cross-presentation and $CD8^+$ CTL priming against ovarian tumor antigens [9]. TLR4 agonism thus represents a mechanistically justified molecular target

for adjuvant development, and peptide ligands capable of engaging TLR4/MD-2 with high affinity while maintaining a favorable safety profile constitute a rational design objective.

AMP immunomodulatory mechanisms for cancer vaccine adjuvancy take place at several cellular interfaces. In DCs, AMPs increase the expression of surface co-stimulatory molecules CD80, CD86 and CD40, in addition to MHC class I and class II, to further amplify the magnitude and extent of antigen-induced T-cell activation. Several AMPs have been shown to have the ability to switch tumor-associated macrophages (TAMs) from their immunosuppressive M2 phenotype (secreting IL-10, TGF-β, arginase-1) to their immune stimulatory M1 phenotype (secreting IL-12, TNF-α, iNOS), directly reversing the immunosuppressive architecture created by TAMs. In addition to their immunological effects, AMPs have direct anticancer cytotoxicity by exploiting biophysical differences that exist between cancer cell membranes and normal somatic cells. Higher exposure of phosphatidylserine, higher content of sialic acid glycoconjugates, and lower content of cholesterol in cancer cells, resulting in a net negative surface charge that selectively attracts the cationic AMP scaffolds to cancer cells and promotes membrane disruption and apoptotic cell death in the malignant cell populations [9].

Scorpion venom has proven to be a pharmacologically rich, and naturally abundant source of bioactive peptides, among natural biological sources of AMPs [10]. Scorpions are members of one of the oldest groups of Arthropods and their venomous stinger, which is used for both defense and predation, has developed a complex biochemical mix that is the result of millions of years of selective pressure. Scorpion venom AMPs fall into two main structural groups: disulfide-bridged peptides (DBPs) are compact globular or beta-sheet peptides that form two to four intramolecular disulfide cross-links, while non-disulfide-bridged peptides (NDBPs) are linear, alpha-helical or random-coil peptides without disulfide cross-links. The short, amphipathic and cationic structure of NDBPs (10-30 amino acid residues) closely resembles the structural motif of known TLR4-activating lipopeptide agonists, and makes them ideal candidates for both receptor interaction and membrane interaction, which is crucial for the design of adjuvants.

The biological properties of scorpion AMPs useful for cancer immunotherapy have been shown in various experimental systems. Several of the NDBPs of scorpions have been found to selectively kill various cancer cell lines, such as ovarian, breast, lung and hepatic carcinoma cells, with a therapeutic index that is superior to non-specific cytolytic agents, as they are not found to cause any measurable hemolysis or cytotoxicity at concentrations that are capable of killing the cancer cells.

Another aspect of the inhibitory effect of scorpion venom-derived peptides on tumor growth is their ability to inhibit tumor angiogenesis by blocking the expression of the vascular endothelial growth factor (VEGF) and disrupt the endothelial tube formation in vitro, which is a crucial process in the vascularisation of ovarian tumors and in their metastatic spread.

Also, the structural diversity of the NDBP chemical space (30 genera from 3 families Buthidae, Scorpionidae, and Chaerilidae; 9–29 residues, +1 to +8 charges) provides a large natural library of physicochemically diverse scaffolds from which candidates with potential as adjuvants for TLR4 can be computationally identified and prioritized [11].

The concept of adjuvant in immunology has its origins in empirical observations that the magnitude of immune responses to injection of antigens was significantly amplified by local inflammation. In the intervening years the science of adjuvants has progressed away from this empirical basis and into a more mechanistic understanding of how the innate immune receptor biology, defined by the Janeway 'infectious non-self' hypothesis, and later extended to the discovery of pathogen-associated molecular patterns (PAMPs) as TLR ligands, can be used to regulate the adaptive immune system. The therapeutic concept of TLR4 agonism to produce the Th1-polarized immune responses that are essential for effective antiviral immune responses has been validated in the clinically deployed TLR4 agonists, monophosphoryl lipid A, the TLR4-activated component of the licensed HPV vaccines Cervarix and the hepatitis B vaccine Fendrix. The challenge of applying this to cancer vaccines, in which antigens are derived from the host, and the mechanisms of tumor immune evasion are complex, demands adjuvant candidates that have comparable or higher TLR4 binding ability to MPL, and a safety profile

suitable for repeated parenteral administration, which naturally-derived peptide agonists are favorably suited to meet [12].

Over the last ten years, the field of computational biology in the discovery and optimization of therapeutic peptides has experienced an unprecedented acceleration, as the curated peptide databases have matured, machine learning based sequence-to-activity prediction models have been developed and high-resolution experimental receptor structures suitable for atomic-level molecular docking simulation have become available.

In the past, the identification of bioactive peptides from venomous animals had to involve months of sequential bioactivity-guided purification of peptides after multiple chromatographic steps, and the requirement for milligrams of crude venom to isolate a small fraction of the venom proteome [13].

This process is now fundamentally different with the help of modern *In silico* pipelines where curated databases like the Antimicrobial Peptide Database (APD3) and Data Repository of Antimicrobial Peptides (DRAMP) contain thousands of experimentally validated AMP sequence and the associated physicochemical parameters and biological activity annotations. This allow for virtual screening of hundreds of peptides with the help of target receptors and safety predictors in a fraction of the time and expense needed for experimental approaches within a few seconds [13, 14].

Molecular docking simulation is playing a pivotal role in the analysis of peptide-receptor interactions in computational approaches. The ability of the docking algorithms, like AutoDock Vina, to estimate the Gibbs free energy of binding (ΔG) of peptide-receptor complexes, based on van der Waals contacts, hydrogen bond geometry, electrostatic complementarity, and degree of hydrophobic burial. It is reflected in the quantitative binding affinity predictions that are well-correlated with the experimentally determined inhibition constant in a wide variety of peptide-receptor complexes [14].

Docking to the experimentally resolved co-crystal structure in the Protein Data Bank, of TLR4/MD-2, allows the identification of peptide sequences that have

structural and electrostatic properties complementary to the TLR4 lipid A binding interface; a basis of mechanistically interpretable prioritisation for further experimental validation.

Complementary *In silico* safety tools that predict cytotoxicity, forecast hemolysis, assess allergenicity and calculate instability index allow for simultaneous safety information on candidate peptides, providing a multi-dimensional candidate evaluation that would take months of experimental assays to achieve in the laboratory [14].

Till now, no systematic *In silico* screening of scorpion-derived AMPs specifically against the TLR4/MD-2 co-receptor complex for the development of adjuvants for ovarian cancer vaccines has been reported.

Most previous computational researches on scorpion AMPs have been devoted to predicting their antimicrobial mechanisms, or to evaluating their general anticancer activity without a dedicated framework for evaluating their interaction with TLR4 nor to multi-parameter safety profiling cascade [12].

This study covers this by providing a comprehensively designed computational pipeline of sequence curation, physicochemical profiling, blind molecular docking to PDB entry 3FXI (human TLR4/MD-2 at 2.70 Å), allergenicity and cytotoxicity screening, and prediction of hemolytic activity in a library of 117 scorpion-derived AMPs retrieved from APD3 and DRAMP, the largest computational screen of scorpion AMPs against a human innate immune receptor reported till now.

The main goal of this work is the computational discovery of scorpion peptides exhibiting a strong binding affinity to the human TLR4/MD-2 co-receptor complex and a favorable safety profile for therapeutic use as cancer vaccine adjuvants. The results of this computational pipeline will be used to generate a shortlist of sequences of scorpion NDBPs with a high probability of success for experimental synthesis and subsequent in vitro and in vivo immunological validation, thus providing a rational and resource-efficient framework to develop novel natural-product-derived adjuvants for ovarian cancer immunotherapy [14, 15].

1.1 Purpose of the Research

The main objective of this research is to leverage the scorpion derived antimicrobial peptides as immunomodulatory adjuvants in the ovarian cancer immunotherapy using computational methods. The goal of this research is to discover new peptide targets for filling the gap in the immune system's ability to target ovarian cancer with minimal toxicity and increased therapeutic efficacy. The purpose of this thesis is to be useful to academic supervisors, research committee members, and researchers in the areas of bioinformatics, cancer biology, immunology, and peptide therapeutics.

1.2 Problem Statement

Ovarian cancer immunotherapy is hindered by a suppressive tumor microenvironment and a shortage of adjuvants able to safely trigger strong, lasting immune responses. Scorpion venom peptides are a largely unexplored group of molecules that show both cell-killing and immune-activating properties. Yet their potential as vaccine adjuvants for ovarian cancer has not been systematically studied. This gap calls for a computational approach to identify and evaluate promising scorpion peptide candidates for adjuvant development.

1.3 Scope of Study

This research focuses on the identification of scorpion-derived antimicrobial peptides and their evaluation as potential immunomodulatory agents for ovarian cancer therapy. The study involves computational characterization of peptide physicochemical and structural properties, followed by the prediction of their immunomodulatory potential. Furthermore, the selected peptides will be assessed for safety, stability, and therapeutic suitability using various *In silico* tools. Based on these analyses, potential peptide candidates will be proposed for their potential application in ovarian cancer therapy. The scope of this study is limited to

computational analysis, while experimental validation and clinical evaluation will be recommended as future work..

1.4 Aims and Objectives

1.4.1 Aim

To identify and evaluate scorpion-derived antimicrobial peptides as immunomodulatory vaccine adjuvants for ovarian cancer therapy using computational approaches.

1.4.2 Objectives

The main objectives of this study are:

i. To identify scorpion-derived antimicrobial peptides from databases and literature

ii. To analyze physicochemical properties and immunomodulatory potential of selected peptides

iii. To evaluate peptide safety and toxicity and prioritize peptide-based vaccine adjuvants for ovarian cancer.

Chapter 2

Literature Review

2.1 Sources of Antimicrobial Peptides

The antimicrobial peptides are structurally and functionally diverse molecules of innate immunity, which are synthesized by almost all living organisms as the first-line defense against microbial pathogens. This evolutionary conservation from prokaryotic bacteriocins to mammalian *defensins* and *cathelicidins* suggests a basic selection pressure of antimicrobial activity that occurs rapidly and is effective against a wide range of microorganisms without prior antigen exposure or expansion of clonal lymphocytes [16].

AMPs are broadly divided into those based on their biological origin, their secondary structure, amino acids composition, and mechanism of action. The primary sources are the viral (bacteriophage) group, as well as the bacterial, fungal, plant and animal groups, with insects, amphibians, reptiles, marine invertebrates, mammals and arthropods (including *scorpions*) providing the largest and most structurally diverse set of characterized AMP sequences [17].

Table 2.1 provides a general overview of the animal-derived AMPs and Table 2.2 lists the insect-derived AMPs in detail.

TABLE 2.1: Animal sources of antimicrobial peptides: representative species, structural class, molecular features, and documented biological activities. [17]

Animal Group	**Representative Species**	**AMP Class**	**Molecular Features**	**Documented Activities**
Insects	*Apis mellifera* (honeybee)	Apidaecins, Defensins, Melittin	Linear cationic; 18–26 aa; amphipathic helices	Antibacterial, antifungal, anticancer (melittin)
Insects	*Drosophila melanogaster*	Drosomycin, Defensin	Cysteine-rich; 44 aa; β-sheet scaffold	Antifungal, antibacterial
Amphibians	*Xenopus laevis*	Magainins, PGLa	α-helical; 22–23 aa; +4 to +5 charge	Broad-spectrum antimicrobial, anticancer
Amphibians	*Rana temporaria*	Temporins	Short helix; 10–14 aa; hydrophobic	Antibacterial, antiviral
Scorpions	*Androctonus australis*	Androctonin, NDBPs	Linear/helical; 10–30 aa; +1 to +8	Antimicrobial, immunomodulatory, anticancer
Scorpions	*Pandinus imperator*	Pandinins, Opisthophthalin	NDBP; amphipathic; +2 to +4	Anti-Gram+/-, DC maturation, TLR4 activation
Scorpions	*Leiurus quinquestriatus*	Chlorotoxin, Charybdotoxin	DBP; 36–37 aa; 3–4 disulfide bonds	Glioma targeting, K^+ channel blockade
Marine organisms	*Limulus polyphemus*	Tachyplins, Polyphemusin	β-hairpin; 17–18 aa; high cationicity	Antiviral, LPS neutralization, TLR4 modulation
Marine organisms	*Mytilus galloprovincialis*	Mytilins, Defensins	Cysteine-rich; 34–40 aa	Antibacterial, antifungal
Mammals	*Homo sapiens*	α/β-Defensins, Cathelicidins (LL-37)	Cysteine-bridged/linear; 12–50 aa	Innate immunity, DC recruitment, anticancer

Table 2.1 continued from previous page

Animal Group	**Representative Species**	**AMP Class**	**Molecular Features**	**Documented Activities**
Mammals	*Mus musculus*	Cryptidins, Paneth cell defensins	β-sheet; 3 disulfide bonds	Gut mucosal immunity, antibacterial
Reptiles	*Crocodylus niloticus*	Crocodile leukocyte peptides (CRAMPs)	Cationic; linear helical	Antifungal, antibacterial, antibiofilm
Spiders	*Latrodectus mactans*	Latarcins, Oxyopinins	α-helical; 20–26 aa; amphipathic	Antimicrobial, membrane disruption
Spiders	*Loxosceles reclusa*	Loxtoxin derivatives	Linear; cationic; +2 to +3	Cytotoxic, antibacterial

2.1.1 Viral Antimicrobial Peptides

Bacteriophages have a different class of antimicrobial proteins, structurally, that are also capable of targeting the cell walls, but act at the macromolecular level by the same mechanisms used by the conventional AMPs. Endolysins encoded in the phage lytic cassette are peptidoglycan hydrolases which break down the bacterial cell wall during lytic replication of the phages internally, and as purified proteins externally, which has resulted in their being called 'enzybiotics' [18].

The phage lysins are modular proteins of 25-40 kDa, with an N-terminal enzymatically active domain (EAD) and a C-terminal cell wall binding domain (CBD); EAD can contain a muramidase, glucosaminidase, amidase or endopeptidase function, depending on the phage family. They are very specific for cognate species of bacteria, have rapid bactericidal kinetics and lack the development of resistance as cells do not need to alter their cell wall in order to acquire resistance to lysin activity, which is a crucial characteristic that differentiates them from conventional antibiotics and conventional AMPs [19].

The phage-tail-like type of bacteriocins is a unique class of high molecular weight antimicrobial complexes, lacking of genetic material but structurally resembling contracted bacteriophage tail fibers. These can be divided into two types, R-type (tails contractile, like that of *Myoviridae*, ∼200 nm rigid tube) and F-type (tails are flexible, like that of *Siphoviridae*, ∼165 nm flexible rod) bacteriocins [20]. Best characterised are the R-type pyocins of *Pseudomonas aeruginosa*, which on irreversible binding to a receptor on the target bacterium contract the sheath, which drives the tube through the outer membrane, dissipating the proton motive force, without causing cell lysis, leading to rapid cell death. It has been suggested that these contractile killing machines can be used as next generation precision antibacterials and as possible vehicles for drug delivery [21].

2.1.2 Bacterial Antimicrobial Peptides

2.1.2.1 AMPs from Gram-Positive Bacteria

Bacteriocins are a structurally and functionally diverse family of AMPs produced by gram-positive bacteria, collectively known as ribosomally and non-ribosomally synthesized AMPs. Bacteriocins derived from the Gram-positive group are mostly classified as either Class I (lantibiotics) or Class II (non-lanthionine-containing bacteriocins) depending on their post-translational modification [21].

Lantibiotics like nisin of *Lactococcus lactis* and mersacidin of *Bacillus* sp., are characterized by the presence of the unique thioether amino acids, lanthionine and β-methyllanthionine, both of which are introduced by dedicated biosynthetic enzymes, which results in extreme proteolytic stability and thermal resistance. Class I lantibiotics like nisin have two modes of action: disrupting peptidoglycan synthesis by complexing with lipid II, and pore formation in bacterial cell wall by interaction with lipid II complex. Class II non-modified: The bacteriocins in this class are pediocin-like bacteriocins containing an N-terminal YGNGVXC consensus motif and a hydrophobic/amphipathic C-terminal domain that is necessary for membrane permeabilization [22].

2.1.2.2 AMPs from Gram-Negative Bacteria

Bacteriocins are also produced by gram-negative bacteria, but occur in a similarly wide variety of structures suited to the cell envelope structure of these bacteria which includes an additional outer membrane. The microcins are ribosome-synthesized peptides produced mainly by *Enterobacteriaceae* bacteria such as *Escherichia coli* and *Klebsiella* species that have bactericidal properties by disrupting the membranes or inhibiting intracellular enzymes [23].

Microcin J25 is a lasso peptide with a mechanically interlocked ring-tail structure that acts by steric blockade mechanism totally different than rifampicin, making it active against rifampicin resistant strains. The larger colicins of *E. coli* (usually more than 10 kDa) are divided into two groups, pore-forming colicins (group A: colicins A, B, E1, Ia, Ib, K, N) and nuclease-type colicins (group B: colicins E2-E9, D, E3).

Due to the evolutionary arms race between competing bacteria occupying the same environmental niche, Gram-negative bacteriocins are highly structurally diverse, and a natural combinatorial library of antimicrobial scaffolds with different mechanisms has been created [24].

2.1.3 Fungal Antimicrobial Peptides

Fungal AMPs are structurally unique thanks to the presence of non-proteinogenic amino acid residues, backbone modifications, and cyclic structures, which make them more resistant to proteases than their linear peptide counterparts. The Peptaibols are mainly synthesized by the *Trichoderma* species; these are short, linear or slightly branched peptides, in which several non-standard amino acids are present, mainly α-aminoisobutyric acid (Aib) [25].

The backbone dihedral angles are restricted by aib substitution, causing to form a 3_{10}-helix or α-helix which is a rigid rod that can insert into the membrane and dissipate ion gradients and membrane potential without requiring specific receptor

interaction. *Fungal defensins*, which belong to the cis-defensin superfamily, are 40-55 amino acid residues long, and contain four or five disulfide bonds, arranged in the γ-core motif. In contrast to mammalian β-defensins which disrupt bacterial membranes, several fungal defensins such as plectasin from the saprophytic fungus *Pseudoplectania nigrella* bind lipid II with nanomolar affinity and inhibit bacterial cell wall synthesis by a mechanism similar to glycopeptide antibiotics like vancomycin [26].

2.1.4 Plant Antimicrobial Peptides

The plants use a multi-layered chemical defense system with a diverse array of AMPs derived from several gene families, which are effective against a range of bacterial, fungal, viral and insect pathogens [27]. The small basic peptides of 45-48 residues, the thionins, are the prototype plant AMPs, which interact with negatively charged membrane phospholipids and cause membrane permeabilization, and belong to two main thionin structural superfamilies, α/β-thionins and γ-thionins [28]. The hevein-related peptides containing a conserved Cys-Pro-Ala-Gly-motif and two to five disulfide bonds exhibit antifungal activity through binding to chitin in fungal cell-walls, competitively inhibiting fungal wall synthesis and destabilizing hyphal membranes [29].

Plant defensins, of which over 300 have now been characterized, adopt the cysteine-stabilized $\alpha\beta$ (CS$\alpha\beta$) motif and exhibit diverse activities including fungal growth inhibition, insecticidal activity, and in some members antifungal mechanisms targeting plant sphingolipids rather than membrane disruption [30]. Cyclotides circular, knotted peptides from plants in the *Rubiaceae*, *Violaceae*, and *Cucurbitaceae* families carry a head-to-tail cyclized backbone reinforced by three disulfide bonds in a cyclic cystine knot (CCK) motif, producing extraordinary stability toward proteolysis, thermal denaturation, and chemical degradation [31]. Their broad activities antibacterial, antiviral, cytotoxic, insecticidal, nematocidal and their tolerance for sequence modification while retaining the CCK scaffold make them attractive templates for therapeutic peptide engineering [32].

TABLE 2.2: Insect-derived antimicrobial peptides: structural classification, molecular parameters, and primary documented activities.

AMP Name	**Source Insect**	**Len (aa)**	**Structure**	**Charge**	**Primary Activity**
Melittin	*Apis mellifera (Honey bee)*	26	α-helical; amphipathic	-1 to $+6$ (pH-dep.)	Membrane disruption; anticancer
Apidaecin Ia	*Apis mellifera*	18	Proline-rich; no defined secondary structure	$+3$	Antibacterial (Gram-); inhibits ribosomal translation
Abaecin	*Apis mellifera*	34	Proline-rich; extended chain	$+2$	Antibacterial; synergistic with defensin
Hymenoptaecin	*Apis mellifera*	93	Glycine/proline-rich; extended	$+2$	Anti-Gram- bacteria; antifungal
Defensin-1 (Royalisin)	*Apis mellifera*	51	CS$\alpha\beta$ scaffold; 3 disulfide bonds	$+7$	Anti-Gram+ bacteria; anti-Paenibacillus larvae
Drosomycin	*Drosophila melanogaster*	44	β-sheet; 4 disulfide bonds; cystine-knot	$+4$	Antifungal (Neurospora crassa, Botryis cinerea)
Defensin (CG2736)	*Drosophila melanogaster*	40	CS$\alpha\beta$; 3 disulfide bonds	$+5$	Anti-Gram+ bacteria
Cecropin A	*Hyalophora cecropia (Silk moth)*	37	Two α-helices; hinge region; N-terminal helix	$+4$	Broad-spectrum antibacterial; anticancer (bladder)
Cecropin B	*Manduca sexta*	35	Two α-helices; amphipathic	$+6$	Antibacterial; membrane disruption
Moricin	*Bombyx mori (Silkworm)*	42	Extended helix; cationic C-terminal	$+8$	Anti-Gram+/-; antifungal
Gloverin	*Manduca sexta*	56	Glycine-rich; proline cluster	neutral	Antibacterial; targets outer membrane proteins

Table 2.2 continued from previous page

AMP Name	Source Insect	Len (aa)	Structure	Charge	Primary Activity
Attacin A	*Hyalophora cecropia*	188	Glycine-rich; large multi-domain protein	–	Anti-Gram-; inhibits outer membrane protein synthesis
Ponericin W1	*Pachycondyla goeldii (Ant)*	22	α-helical; amphipathic	+3	Antibacterial; antimalarial; hemolytic activity
Mastoparans	*Polistes exclamans (Wasp)*	14	Short α-helix; C-terminal amide	+3	Mast cell degranulation; antibacterial; anticancer

2.2 Scorpion Venom as a Source of Antimicrobial and Immunomodulatory Peptides

The oldest terrestrial arthropods, the scorpions (order *Scorpiones*) have an evolutionary history of ∼430 million years and their venom systems constitute one of the best-documented molecular pharmacopoeias in the animal kingdom [33].

Stings delivered by the venom gland and stinger are located in paired secretory lobules inside the telson, the end of the tail where the venom gland and stinger are found, and are a complex mix of bioactive chemicals. These include mucopolysaccharides, phospholipases, hyaluronidases, and most importantly for therapeutic purposes, a very rich complement of peptides, which comprise the bulk of venom biological activity [34].

There are more than 2500 species of scorpions on six continents, and the venom peptide repertoires which have been best characterized and described have come from the families *Buthidae* (the most medically important), *Scorpionidae*, *Chaerilidae*, and *Urodacidae*. The biological activity of the scorpion AMPs is localized in the venom gland, a specific organ in which they are synthesized and not present

in hemolymph, as is the case with the honeybee venom AMPs, including melittin. Transcriptomic studies in the venom glands of scorpions have shown the expression of more than 200 different peptide-encoding transcripts in a single species; the majority of the former encode ion-channel targeted toxins (scorpion classical pharmacology), while the latter are short non-disulfide-bridged AMPs, whose therapeutic potential has only recently started to be systematically evaluated [35].

2.2.1 Classification of Scorpion Venom Peptides

The venom peptides of scorpions are best organized on the basis of their structural architecture, which dictates their biophysical properties, modes of interaction with receptors and their pharmaceutical tractability [36].

There are two major categories: 1) Disulfide-bridged peptides (DBPs); and 2) Non-disulfide-bridged peptides (NDBPs). The DBPs are composed of two to four intramolecular disulfide bonds that link the cysteine side chains into compact globular or beta-sheet proteins.

The neurotoxic channel modulators include the voltage-gated K^+ channel toxins α-KTx, $Ca_2{}^+$-activated K^+ channel toxins β-KTx, voltage-gated NA^+ channel toxins NaTx, Cl^- channel/MMP-2 binding peptides ClTx and $Ca_2{}^+$ channel toxins CaTx which have high target selectivity and are structurally rigid and are useful research tools and are promising new leads for drug development, but their complicated disulfide topology is problematic for chemical synthesis [37].

By contrast, the NDBPs consist of short linear sequences of 9-35 residues which are amphipathic, adopting α-helical or extended coil conformations in hydrophobic environments. By contrast, the NDBPs consist of short linear sequences of 9-35 residues which are amphipathic, adopting α-helical or extended coil conformations in hydrophobic environments. The chemically unconstrained nature of NDBPs facilitates their synthesis using standard Fmoc solid-phase strategy, enables their structural prediction by computational methods, and renders them strongly membrane-active, ideal candidates for the study of their interaction with

receptors of the Toll-like receptor family (TLRs) which are also membrane - associated proteins.

TABLE 2.3: Scorpion venom peptides with documented immunomodulatory and anticancer potential: structural classification, source species, molecular parameters, and key biological activities.

Peptide	**Source Scorpion**	**Class**	**Length (aa)**	**Charge**	**Immunomodulatory / Anticancer Activity**
Pandinin-1	*Pandinus imperator*	NDBP	13	+2	DC activation; IL-6/TNF-α induction; anti-Gram+/-
Pandinin-2	*Pandinus imperator*	NDBP	13	+1	Macrophage polarization; antifungal; anticancer
IsCT	*Opisthacanthus madagascariensis*	NDBP	13	+2	Membrane disruption; ovarian cancer cytotoxicity
IsCT2	*Opisthacanthus madagascariensis*	NDBP	13	+3	Selective cancer cell lysis; immune activation
Chlorotoxin (ClTx)	*Leiurus quinquestriatus*	DBP (ClTx)	36	0	MMP-2/Cl^- channel binding; glioma/OC targeting
BmK CTX	*Buthus martensii Karsch*	DBP	36	+1	Ovarian cancer apoptosis (mitochondrial pathway)
Margatoxin	*Centruroides margaritatus*	DBP (α-KTx)	39	+3	Kv1.3 blockade; OC proliferation inhibition
Ts1 (Tityustoxin-1)	*Tityus serrulatus*	DBP (NaTx)	61	+5	Nav1.5 inhibition; OC migration/invasion reduction
Hadrurin	*Hadrurus gertschi*	NDBP	41	+4	Antibacterial; anticancer; membrane disruption
Androctonin	*Androctonus australis*	DBP	25	+8	TLR4/PRR activation; NF-κB induction; anti-Gram+/-
BmKn2	*Buthus martensii Karsch*	NDBP	13	+2	Anti-MRSA; antifungal; anticancer (breast)
Opistho phthalin-1	*Opisthophthalmus carinatus*	NDBP	44	+4	Broad-spectrum antimicrobial; anti-biofilm

DBP = disulfide-bridged peptide; NDBP = non-disulfide-bridged peptide; DC = dendritic cell; PRR = pattern recognition receptor; TLR = Toll-like receptor [33, 35, 36].

2.2.2 Disulfide-Bridged Peptides

α-KTx peptides inhibit voltage-gated potassium channels (Kv) that are frequently over-expressed in cancer cells and responsible for abnormal proliferation, migration and resistance to apoptosis. The subclass of antiproliferative peptides includes charybdotoxin (α-KTx1.1) for Kv1.3 and margatoxin (α-KTx6.2) for $Ca_2{}^+$-activated K^+ channels (KCa), which have both shown to inhibit cell migration velocity in ovarian cancer models [38, 39]. The 36 amino acid, 4-disulfide bond Chlorotoxin (CTX) of *Leiurus quinquestriatus* has received unprecedented research interest for its selective binding affinity to the membrane of MMP-2 and Cl^- channels over-expressed on the surface of glioma and other cancer cells for tumor targeting of conjugated chemotherapeutics or imaging probes [40].

2.2.2.1 Non-Disulfide-Bridged Peptides

The primary reason for this study focusing on NDBPs is that they are structurally simple, computationally tractable and have been reported to have immunomodulatory properties. IsCTs are the short α-helical peptides (13-residues) from *Opisthacanthus madagascariensis* that are archetypal NDBPs and exhibit selective cytotoxicity towards cancer cell membranes with increased phosphatidylserine exposure and decreased concentration of cholesterol, which is characteristic of malignancy status [41]. The venom of *Pandinus imperator* contains both anticancer membrane effects and immunostimulatory activities such as dendritic cell activation and pro-inflammatory cytokine induction (Pandinins) [42]. Hadrurin a 41-residue NDBP from *Hadrurus gertschi* is one of the longest scorpion NDBPs described so far and has potent antibacterial and anticancer activity via membrane-disruption mechanism [43]. BPPs are the proline-rich hypotensive peptides originally described in the venom of the *Bothrops* species and later found in the venom of scorpions, which have a specific activity on the angiotensin-converting enzyme

(ACE), thus making them a distinct subclass of immunomodulatory peptides in scorpion venom chemistry [44].

2.2.3 Mechanisms of Action of Scorpion Venom Peptides

2.2.3.1 Membrane Disruption

Most scorpion NDBPs act as anticancer and antimicrobial agents by disrupting membranes mainly through biophysical selectivity but not by specific receptor binding. Cancer cells have increased phosphatidylserine on the outer leaflet of their outer membrane compared to normal somatic cells, and increased sialic acid glycoconjugates in contrast to cholesterol content, making these cells electrostatically attracted to cationic AMPs [45] which are also repelled by a neutral-to-positive outer leaflet of normal mammalian membranes that are rich in phosphatidylcholine and sphingomyelin. In the adsorbed state, AMPs penetrate the membrane interior and destabilize the membrane by inserting their hydrophobic face into the membrane core, leading to either barrel-stave pore structures or toroidal pores, or by carpet mechanisms, causing dissipation of the membrane potential, secondary apoptotic cascades, and leakage of ions into the surrounding media.

2.2.3.2 Ion Channel Modulation

In cancer cell biology, dysregulated expression of voltage gated channels like potassium (Kv1.3, Kv11.1/hERG), sodium (Nav1.5), and chloride (CFTR-associated) channels is common and includes aberrant overexpression or altered gating kinetics in ovarian cancer [46]. Scorpion DBPs take advantage of this vulnerability by binding to the external vestibule of the channel pore with high-affinity channel blockade, thereby arresting the cell cycle, inducing apoptosis, inhibiting migration and increasing the chemosensitivity of the cells. This is a mechanism of ion targeting, which is different to membrane disruption, and allows for very selective activity at concentrations much lower than those that will cause non-specific membrane permeabilization [46].

2.2.3.3 Immunomodulation and TLR Activation

In this study, the immunomodulatory properties of scorpion venom peptides that are relevant to the adjuvant goals operate by interacting with innate immune pattern recognition receptors. Some scorpion NDBPs are structural mimics of bacterial lipopeptides or lipopolysaccharide that when recognized by the receptor ectodomains activate the receptor TLR2 and TLR4 signaling cascades [47]. The activation of TLR4, in particular, activates a bifurcated response, one involving NF-κB activation, leading to production of pro-inflammatory cytokines (IL-6, IL-12p70, TNF-α), and the other involving IRF3 phosphorylation, leading to production of type I interferons (IFN-α/β), which together create the Th1-polarized cytokine milieu needed for CTL priming [48]. In addition to direct activation of PRR, scorpion peptides have been shown to induce DC maturation by increasing the surface expression of CD80, CD86, CD40 and MHC class II [49] and to convert M2-TAM into the immunostimulatory M1 phenotype through polarization [50].

2.2.3.4 Induction of Apoptosis

Scorpion peptides have at least three different pathways that lead to programmed cell death. The intrinsic (mitochondrial) pathway is initiated when peptide disrupts mitochondrial membrane potential, which causes the release of cytochrome c, the oligomerization of Apaf-1 and the activation of caspase-9 and caspase-3 [51].

The extrinsic (death receptor) pathway is activated by TRAIL binding to Fas (CD95) and DR4 and DR5 located on the surface of the cancer cell, which activates the caspases, such as caspase-8, that depend on this pathway.On the cell surface of the cancer cell, TRAIL binds to death receptors Fas (CD95) and DR4 and DR5, which stimulates the activation of TRAIL-dependent caspases, including caspase-8, in the extrinsic pathway. A third apoptotic axis is activated by unfolded protein accumulation and involves activation of the unfolded protein response (UPR) via PERK, IRE1α, and ATF6, which leads to the transcription of pro-apoptotic genes such as CHOP, which are induced by several scorpion peptides [52].

Their variety of apoptotic mechanisms can be used by scorpion AMPs, which can differentiate them from single-mechanism chemotherapeutics and make it less likely that resistance will develop through mutation of any one of the components of the apoptotic pathway.

2.2.3.5 Inhibition of Angiogenesis

The development of angiogenesis, the process of forming new blood vessels from existing blood vessels, is a well-established target for therapeutic intervention in ovarian cancer, and the anti-VEGF antibody bevacizumab was the first agent to be approved clinically for use in the treatment of this disease.

There are several peptides derived from scorpions that have been found to exhibit anti-angiogenic properties that are not related to the inhibition of VEGF-mediated functions such as basement membrane degradation by matrix metalloproteinases (MMPs), direct inhibition of endothelial cell proliferation, migration, and tube formation, and modulation of the NF-κB pathway that regulates the transcription of VEGF.

The results of these anti-angiogenic effects at sub-cytotoxic doses suggest that scorpion AMPs are capable of interfering with the vascularization of ovarian tumors without the need for systemic endothelial toxicity [53].

2.3 Ovarian Cancer: Biology, Epidemiology and Clinical Landscape

Ovarian cancer is one of the most challenging diseases in clinical oncology, with an insidious onset, extreme biological heterogeneity and very limited early detection, which is always hard to control with treatment once the disease becomes disseminated in the peritoneum.

Its pathobiology incorporates the fields of molecular oncology, reproductive endocrinology and tumor immunology, and its treatment has been greatly transformed in the last 30 years with little transformative advances in improving survival compared to breast, lung and colorectal cancer.

To understand the scientific rationale for novel adjuvants for immunotherapies especially those derived from natural bioactive scaffolds that have the ability to activate innate immunity and modulate the immunosuppressive tumor microenvironment as a therapeutic direction, it is essential to have a comprehensive understanding of the biology of ovarian cancer. [54].

2.3.1 Classification and Molecular Subtypes

Ovarian cancer can be divided into three main categories: epithelial ovarian carcinomas (EOC, ∼90% of malignant cases), germ cell tumors (∼5%) and sex cord-stromal tumors (∼5%).

In EOC, the dualistic model of ovarian carcinogenesis proposed by Kurman and Shih divides the disease into two types: Type I, which refers to low-grade serous, mucinous, endometrioid, clear cell and transitional carcinomas, are stepwise building of mutations in precursor cells and have a generally less aggressive clinical genomically stable, develop through course, and Type II, which includes high-grade serous carcinoma (HGSOC), high-grade endometrioid and undifferentiated carcinomas, are genomically chaotic, develop de novo or from serous tubal intraepithelial carcinoma (STIC) lesions of the fallopian tube fimbriae and have the poorest prognosis. HGSOC accounts for ∼70% of all EOC diagnoses and for most ovarian cancer mortality.

It features the highly prevalent TP53 mutation in >96%, the 15-20% of cases with germline or somatic mutations in BRCA1/2, and the ∼50% of cases with HRD, which not only provides the first strong therapeutic vulnerability, the ability to be treated with PARP inhibitors, but also reveals a major therapeutic challenge which is chemoresistance resulting from genomic plasticity [55].

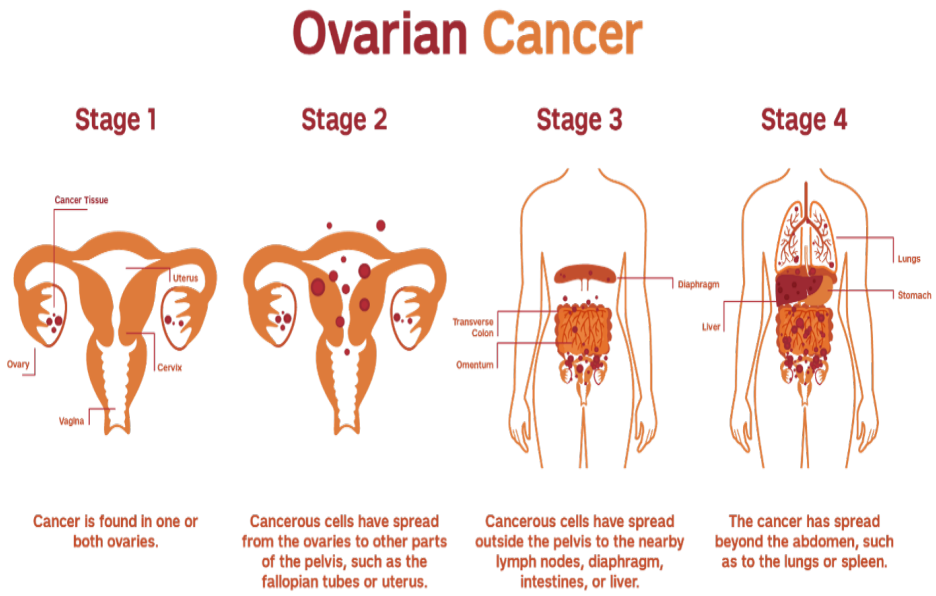


FIGURE 2.1: Stages of Ovarian Cancer progression and metastatic spread.

This figure illustrates the clinical progression of ovarian cancer from early localized disease to advanced metastatic stages. Stage I represents tumor confinement within one or both ovaries. Stage II indicates spread to nearby pelvic structures such as the uterus or fallopian tubes. Stage III involves dissemination beyond the pelvis, including metastasis to the peritoneum, lymph nodes, diaphragm, and abdominal organs. Stage IV represents advanced disease with distant metastasis to organs such as the lungs, liver, and spleen. The progressive spread of ovarian cancer contributes to poor prognosis and highlights the importance of early detection and novel therapeutic strategies.

2.3.2 Virulence and Pathogenic Mechanisms

Ovarian cancer cells are characterized as having pathogenic properties that include multiple mechanisms which work synergistically to allow dissemination into the peritoneum and resistance to therapeutic elimination.

Rapid and uncontrolled proliferation is caused by the dysregulation of multiple cell cycle checkpoints, including loss of TP53 tumor suppressor function, the earliest and most common molecular events in HGSOC, which take off the G1/S and

G2/M DNA damage checkpoints and allow later driver mutations to accumulate [56].

Ovarian cancer spreads in a unique way, called the intraperitoneal route, which involves the exfoliation of clusters of ovarian cancer cells, the ability of these clusters to survive in the ascitic fluid without apoptosis induced by the loss of extracellular matrix anchorage, and the subsequent implantation onto the surfaces of the omentum and peritoneum. This unique route of metastasis does not require and not characterized by vascular dissemination, which is the spread of most solid tumor metastases [57].

There are multiple independent mechanisms of chemoresistance to platinum-based therapy that are unique to advanced ovarian cancer; these include upregulation of nucleotide excision repair (NER) and homologous recombination that restores DNA repair capacity in initially HRD + tumors, upregulation of ABC family drug efflux transporters like ABCB1/P-gp and ABCG2 that decrease intracellular platinum levels, and down-regulation of copper transporters CTR1, which decreases platinum uptake, and upregulation of anti-apoptotic Bcl-2 family members, which inhibit the apoptotic response to platinum-induced DNA damage [58].

The immunosuppressive TME of HGSOC provides another mechanism for therapeutic resistance, by dampening the endogenous T cell response, and by interfering with the effectiveness of checkpoint blockade therapies via the multi-layered immunosuppressive networks outlined below [59].

2.3.3 Life Cycle and Disease Progression

Ovarian cancer arises from a multistep process that starts with initiating mutations in cells on the surface of the ovary or, for HGSOC, in secretory cells of the fallopian tube fimbriated end, which can be detected in the ovary in cases of prophylactic salpingo-oophorectomy in women with a BRCA mutation [60]. During the promotion phase, genetic changes that occur as a result of chronic inflammation as a result of repeated ovulation, hormone microenvironmental changes and

microbiome variations, allow the survival and clonal expansion of transformed cells [61].

Local invasion of the ovary capsule and adjacent pelvic organs, followed by transcoelomic spread to peritoneal surfaces, especially the omentum, which provides a fatty, well vascularized site of implantation, and transcoelomic spread to paraaortic and pelvic lymph nodes, is the pattern of tumor progression [62].

Patients with recurrent disease develop about 70% of the population of Stage III/IV patients 18-24 months after first-line platinum-based chemotherapy, and are a highly chemoresistant population that is far more difficult to control and responsible for the majority of ovarian cancer deaths [63].

2.3.4 Clinical Features

Ovarian cancer is a disease with a long latent period where the disease is clinically silent and allows for the disease to spread to the peritoneum. The symptoms which eventually develop include abdominal bloating or distention, pelvic pressure or pain, early satiety, urinary urgency or frequency, fatigue, unexplained weight change, and gastrointestinal disturbances, all of which are not specific enough to differentiate ovarian cancer from the wide differential diagnosis of conditions that cause these symptoms but are not cancer [64].

Table 2.4 provides a structured summary of the symptoms of ovarian cancer, clinical association and the likely misdiagnosis by primary care practitioners.

TABLE 2.4: Clinical features of ovarian cancer: symptom description, stage association, and common misdiagnosis in primary care. Early-stage symptoms are frequently misattributed to benign conditions, contributing to delayed diagnosis. [65].

Symptom	**Description**	**Frequency / Stage Association**	**Common Misdiagnosis**
Abdominal bloating	Persistent distension not related to diet; progressive in nature	Present in ∼70% of OC patients; Stage I–IV	Irritable bowel syndrome (IBS)

Table 2.4 continued from previous page

Symptom	**Description**	**Frequency / Stage Association**	**Common Misdiagnosis**
Pelvic/abdominal pain	Dull, persistent pain in lower abdomen or pelvis; distinct from menstrual pain	50–60% of patients; more common in Stages III–IV	Endometriosis, fibromyalgia
Early satiety	Feeling full after eating small amounts; secondary to peritoneal involvement	∼34% of patients; correlates with ascites	Gastroparesis, gastritis
Urinary symptoms	Urgency or frequency without urinary tract infection; pressure on bladder	∼20% of patients; pelvic mass effect	UTI, overactive bladder
Fatigue	Persistent, unexplained exhaustion; not relieved by rest; anaemia-related	Common across all stages; often ignored	Depression, anaemia of unknown origin
Unexplained weight loss	>5% body weight loss over 6 months; cachexia in advanced disease	Late-stage marker; Stage III–IV	Dietary change, hyperthyroidism
Menstrual irregularities	Abnormal bleeding, postmenopausal bleeding; hormonal disruption by stromal tumours	Variable; more in germ cell/stromal tumour subtypes	Perimenopause, uterine fibroids
Back pain	Lower back discomfort from retroperitoneal lymph node involvement or large pelvic mass	∼40% of advanced-stage patients	Musculoskeletal, renal disease
Nausea/ constipation	GI motility disruption; bowel compression by tumour mass or ascites	Advanced disease (Stage III–IV)	IBS, peptic ulcer disease

2.3.5 Available Treatment Modalities

Current standard of care treatment is a combination of surgery (cytoreductive surgery) and chemotherapy, and for selected molecular subsets, chemokinesis. Complete gross resection (R0) is still by far the single strongest predictor of

progression-free survival and every centimetre of residual tumor mass has a measurable decrement in survival [66]. Chemotherapy naïve advanced stage patients who achieve objective response rates of 70-80% with first-line platinum-taxane chemotherapy (carboplatin and paclitaxel, 6 cycles IV) subsequently develop platinum resistant disease and have a median survival of only 12-18 months [67].

PARP inhibitors (olaparib, niraparib, rucaparib) have revolutionized the treatment of BRCA-mutant and HRD+ tumors in the maintenance setting with a median PFS improvement of 12-18 months compared to placebo in the primary maintenance trials, but only for the ∼50% of HGSOC patients who are HRD+ and resistance through BRCA reversion mutations is an emerging clinical challenge [68]. All current and investigational drugs used in the treatment of ovarian cancer are summarized in Table 2.5, which includes mechanism of action, benefits, drawbacks and clinical development status.

TABLE 2.5: Available treatment modalities for ovarian cancer with mechanisms, clinical advantages, documented limitations, and current development status. AMP-based adjuvants (final row, green shading) represent the investigational category addressed by this study [66–70]

Treatment Modality	**Mechanism**	**Advantages**	**Limitations / Challenges**	**Current Status in OC**
Cytoreductive Surgery (Debulking)	Surgical removal of all visible tumor deposits; goal = R0 resection	Directly reduces tumor burden; improves chemo response; best survival with R0	High morbidity; not feasible in all patients; recurrence common	Standard of care (1st-line); neoadjuvant surgery in unresectable disease
Platinum-Taxane Chemotherapy (Carboplatin + Paclitaxel)	DNA cross-linking (platinum); microtubule stabilization (taxane)	High initial response rate (70–80%); well-established protocol	Recurrence in ∼70%; acquired chemoresistance; neuropathy, myelosuppression	Standard 1st-line; 6 cycles IV

Table 2.5 continued from previous page

Treatment Modality	**Mechanism**	**Advantages**	**Limitations / Challenges**	**Current Status in OC**
PARP Inhibitors (Olaparib, Niraparib, Rucaparib)	Exploit synthetic lethality in HRD/BRCA-mutant cells; trap PARP on DNA	Extended PFS in BRCA-mutant pts; oral; maintenance strategy	Effective only in HRD+ subgroup (∼50%); anaemia, thrombocytopenia; resistance	Approved maintenance therapy; 1st & 2nd-line in BRCA+ patients
Bevacizumab (Anti-VEGF)	Inhibits VEGF-A; blocks tumour angiogenesis	Modest PFS benefit; manageable toxicity; combinations active	Hypertension, GI perforation risk; limited OS benefit; expensive	Approved in combination with chemo and as maintenance
Immune Checkpoint Inhibitors (Anti-PD-1/PD-L1)	Block PD-1/PD-L1 axis; restore T cell activity against tumour	Durable responses in responders; well tolerated	Low ORR (6–15%) in unselected OC; immune-related AEs; no approved monotherapy	Phase II/III trials; pembrolizumab approved for MSI-H/TMB-H subset
Cancer Vaccines (NY-ESO-1, WT1, MUC1)	Peptide antigen + adjuvant; prime TAA-specific CTLs	Target-specific; combination potential; durable memory	Weak without potent adjuvant; self-antigen tolerance; limited clinical success	Phase I/II trials; no approved vaccine for OC
CAR-T Cell Therapy	Chimeric antigen receptor T cells targeting Mesothelin, FRα, HER2	Highly personalized; potent CTL activity	Solid tumor penetration barriers; TME immunosuppression	Preclinical + early Phase I in OC
Adoptive TIL Therapy	Expansion and reinfusion of tumor-infiltrating lymphocytes	Exploits pre-existing tumor immunity	Complex manufacturing; lymphodepletion required;	Early clinical investigation

Table 2.5 continued from previous page

Treatment Modality	**Mechanism**	**Advantages**	**Limitations / Challenges**	**Current Status in OC**
AMP-Based Adjuvants (Investigational)	TLR4/PRR activation; DC maturation; Th1 polarization; TME reprogramming	Dual antimicrobial + immunomodulatory; natural origin; low MW; computationally designable	Not yet clinically approved; need in vitro/in vivo validation; hemolysis risk in some scaffolds	Computational / preclinical stage; focus of this study

2.3.6 Global and Regional Prevalence

The global distribution of incidence and mortality of ovarian cancer is the result of differences in the prevalence of biological risk factors, reproductive patterns, healthcare system capacity and genetic population structure between the WHO Regions.

In some parts of Central and Eastern Europe such as the Slavic and Baltic populations, the incidence rates are significantly higher (ASR ∼11-13/100,000) than in Northern and Western Europe and North America (ASR 7-10/100,000) [71].

This is because they have relatively high population prevalence for BRCA1 founder mutations. Incidence is less in Asia and Latin America, where increasing incidence due to the effects of urbanization, "westernization" of diet, and nulliparity have been recorded.

In 2022, there were an estimated 324,600 new cases and 206,900 deaths due to ovarian cancer, making it the eighth leading cause of cancer death in women and the eighth most common cancer in women globally [1].

Table 2.6 shows ovarian cancer incidence, mortality, 5-year survival and stage distribution by WHO region to provide context to the clinical challenge in the healthcare environment.

TABLE 2.6: Global distribution of ovarian cancer incidence and mortality by WHO region (2022 estimates). Data derived from GLOBOCAN 2022 and WHO Cancer Database. [1, 71, 72].

WHO Region	**New Cases (2022)**	**Age-Std. Incidence /100k**	**Deaths (2022)**	**5-Year Survival (%)**	**Median Age (yrs)**
Central & Eastern Europe	52,400	11.8	35,200	32–40	60
Northern & Western Europe	38,700	9.4	23,100	43–49	63
North America	22,800	7.7	13,900	47–50	63
Eastern Asia	79,500	8.1	48,200	35–45	57
South & Southeast Asia	55,600	6.5	39,400	25–32	53
Sub-Saharan Africa	14,800	5.2	12,100	18–22	51
Latin America & Caribbean	18,200	6.8	12,500	28–36	55
Pakistan (National est.)*	3,200	6.2	2,400	18–24	52
GLOBAL TOTAL	324,600	7.2 (ASR)	206,900	∼38 (global avg)	63 (typical)

In Pakistan, especially, ovarian cancer is the third most common type of gynecologic cancer after breast and cervical cancer, with an estimated prevalence of around 3200 new cases every year and an age-standardized incidence of ∼6.2 / 100,000 women [72].

The median age at which Pakistani patients present is significantly younger (50-55 years compared to 63 years in western cohorts) and a far larger proportion with Stage III-IV disease (75-85%) than is reported in European or North American registries [73]. The population-based cancer registry infrastructure is not available in other urban centres, and the most dependable longitudinal data are available from institutional cancer registries at Shaukat Khanum Memorial Cancer Hospital and Research Centre (SKMCH & RC), Aga Khan University Hospital (AKUH) and Nuclear Medicine, Oncology and Radiotherapy Institute (NORI) [74].

The case reports of the ovarian cancer were summarized in Table 2.7 and showed an increase in the case burden of ovarian cancer across Pakistan and the distribution of ovarian cancer registry infrastructure was not uniform across the country.

TABLE 2.7: Reported ovarian cancer cases and estimated fatalities in Pakistan by region and registry (2015-2023). Data compiled from institutional cancer registries including SKMCH & RC (Lahore), AKUH (Karachi), NORI (Islamabad), and provincial hospital records. National totals represent conservative estimates given incomplete registry coverage [72–82].

WHO Region	**New Cases (2022)**	**Age-Std. Incidence /100k**	**Deaths (2022)**	**5-Year Survival (%)**	**Median Age (yrs)**
Central & Eastern Europe	52,400	11.8	35,200	32–40	60
Northern & Western Europe	38,700	9.4	23,100	43–49	63
North America	22,800	7.7	13,900	47–50	63
Eastern Asia	79,500	8.1	48,200	35–45	57
South & Southeast Asia	55,600	6.5	39,400	25–32	53
Sub-Saharan Africa	14,800	5.2	12,100	18–22	51
Latin America & Caribbean	18,200	6.8	12,500	28–36	55
Pakistan (National est.)*	3,200	6.2	2,400	18–24	52
GLOBAL TOTAL	324,600	7.2 (ASR)	206,900	∼38 (global avg)	63 (typical)

2.3.7 Prevention and Control

There is currently no validated population level screening programme for ovarian cancer and there are no population level screening strategies as such, the prevention landscape being dominated by risk stratification strategies focussed on recognising high risk subgroups.

Risk reducing bilateral salpingo-oophorectomy (RRSO) is the most evidence-based preventive intervention, and is a component of national clinical guidelines in many high income countries, including the UK, where the procedure is recommended in women with confirmed BRCA1/2 mutations, at which point risk of ovarian cancer is reduced by 85-95% [83]. The most well-established modifiable risk reduction is long-term use of oral contraceptive pills (OCP), which decrease the cumulative ovarian surface epithelial inflammation that drives the risk of OC, lowering it by about 40-50% after using the pill for 5 or more years.

The most well-established modifiable risk reduction is long-term use of oral contraceptive pills (OCP), which decrease the cumulative ovarian surface epithelial inflammation that drives the risk of OC, lowering it by about 40-50% after using the pill for 5 or more years.

Evidence based and feasible prevention and early detection strategies have been summarized in Table 2.8 in a structured way based on the evidence and the feasibility in the health care context in Pakistan.

TABLE 2.8: Prevention and control strategies for ovarian cancer: category, mechanism, evidence level, and feasibility considerations. Strategies are ordered from highest to lowest evidence base [83, 84].

Strategy	**Category**	**Description**	**Evidence Level**	**Limitation / Feasibility**
Risk-reducing salpingo-oophorectomy	Surgical prevention	Bilateral fallopian tube and ovary removal in BRCA1/2 carriers; reduces OC risk by 85–95%	Level I (RCT evidence)	Surgical menopause; not appropriate for general population; BRCA testing required
Oral contraceptive pill (OCP)	Pharmacological prevention	Long-term OCP use (>5 years) reduces OC risk by ∼40–50%; mechanism involves suppression of ovulation	Level I–II	Thrombotic and cardiovascular risk; not suitable for all women

Table 2.8 continued from previous page

Strategy	Category	Description	Evidence Level	Limitation / Feasibility
BRCA genetic counseling & testing	Genetic surveillance	Identification of germline BRCA1/2 mutations enables targeted preventive strategies	Level I (national guideline)	Cost; limited availability in low/middle-income countries including Pakistan
Parity and breastfeeding	Lifestyle factor	Multiparity and breastfeeding reduce cumulative ovulation; associated with 10–20% OC risk reduction	Level II (epidemiological)	Not a targetable intervention; social/cultural factors
Serum CA-125 monitoring	Biomarker surveillance	Annual CA-125 ± TVU in BRCA+ women or familial risk; not approved for general screening	Level II (specialist guidelines)	Low sensitivity in early disease; not population-level screening
Transvaginal ultrasonography (TVU)	Imaging surveillance	Used in combination with CA-125 in high-risk women; annual screening protocol	Level II	UKCTOCS showed no mortality benefit in general population; high false-positive rate
HE4 + ROMA algorithm	Biomarker surveillance	Human Epididymis Protein 4 combined with CA-125 and ROMA score improves specificity for adnexal mass classification	Level II–III	Not validated for screening; used for clinical triage of adnexal masses
Liquid biopsy (ctDNA/cfDNA)	Emerging detection	Detection of circulating tumor DNA in blood; investigational for early OC detection	Level III (emerging)	Not clinically validated; technical complexity; expensive
Healthy lifestyle (BMI, diet, exercise)	Population health	Maintaining normal BMI; Mediterranean diet; regular physical activity; reduces overall cancer risk	Level III (observational)	Modest individual OC risk reduction; population-level public health impact

Table 2.8 continued from previous page

Strategy	Category	Description	Evidence Level	Limitation / Feasibility
Public awareness campaigns	Health education	Symptom recognition campaigns ('Know the symptoms of ovarian cancer'); primary care education	Level IV (expert opinion)	Symptom non-specificity limits yield; requires sustained health system investment

A novel aspect of ovarian cancer prevention and treatment is the creation of immunotherapeutic approaches that can transform the immunologically 'cold' TME in the ovary into an inflamed, immune-infiltrated tumor that is susceptible to CTL response and checkpoint blockade. The use of peptide-based immunotherapy, such as cancer vaccines targeting ovarian tumor-associated antigens, appears to be a logical strategy toward this end, but success will depend on the availability of adjuvants that can induce the necessary innate immune activation, which would break self-antigen tolerance and promote the generation of high-avidity tumor-specific CD_8^+ T cells [84, 85]. In this section the biological and clinical context is directly relevant to the adjuvant discovery strategy employed in this study.

2.4 Scorpion Venom Peptides Against Ovarian Cancer

Scorpion venom peptides' application to ovarian cancer has moved beyond simple cytotoxicity assays in cell line models to studying the mechanisms of their actions, such as engagement of apoptotic pathways, and recently to studies of immune cell modulation in ovarian tumor models. A number of scorpion peptides are emerging as multi-target ovarian cancer agents with the ability to kill ovarian tumor cells, inhibit tumor vessel growth and reprogram the tumor microenvironment, properties that are not shared by a single clinically approved agent [86]. The most well-studied scorpion peptides in the context of ovarian cancer are summarized in Table 2.9.

TABLE 2.9: Scorpion venom-derived peptides with documented activity against ovarian cancer: structural class, molecular targets, and experimentally demonstrated effects [87–92].

Peptide	**Source Scorpion**	**Structural Class**	**Molecular Target**	**OC-Specific Effect**	**Ref**
Chlorotoxin (ClTx)	*Leiurus quinquestriatus*	DBP; 36 aa	MMP-2, Cl^- channels	Inhibits OC migration, invasion; drug delivery vehicle; diagnostic imaging	[87]
BmK CTX	*Buthus martensii Karsch*	DBP; 36 aa	Mitochondria; caspase cascade	OC apoptosis; xenograft tumor reduction; immune modulation	[88]
Ts1 (Tityustoxin-1)	*Tityus serrulatus*	DBP (NaTx); 61 aa	$Na_v1.5$ voltage-gated Na^+ channels	Reduces OC cell migration and invasion; $Na_v1.5$ overexpressed in metastatic OC	[89]
IsCT / IsCT2	*Opisthacanthus madagascariensis*	NDBP; 13 aa α-helix	Cancer cell membranes (PS exposure)	Selective OC membrane disruption; preferential over normal cells; apoptosis induction	[90]
Pandinin-1 / Pandinin-2	*Pandinus imperator*	NDBP; 13 aa	Cell membranes; immune PRRs	OC apoptosis; DC and macrophage activation; anti-tumor cytokine production	[91]
Margatoxin (MgTx)	*Centruroides margaritatus*	DBP (α-KTx); 39 aa	$K_v1.3$ K^+ channels	OC proliferation inhibition; cell cycle arrest ; anti-metastatic	[92]

The 36-residue DBP **Chlorotoxin (ClTx)** from *Leiurus quinquestriatus* venom has been the most extensively studied scorpion peptide in the context of ovarian

cancer. When attached to nanoparticles, quantum dots or photosensitizers, its selective targeting of MMP-2 channels and Cl^- channels that are expressed on the cell surface of ovarian cancer cells allows for targeted delivery of a cytotoxic payload with significantly decreased off-target toxicity compared to unconjugated chemotherapy agents. ClTx also prevents the migration and invasion of OC cells by blocking the degradation of the extracellular matrix, which is mediated by the activity of MMP-2, and is one of the main pathways of peritoneal spread [87].

BmK CTX venom has been shown to be selectively cytotoxic to human ovarian cancer cell lines (SKOV-3, A2780) in a dose dependent manner with activation of caspase-9 and caspase-3, release of cytochrome c and down-regulation of anti-apoptotic Bcl-2 mechanism, associated with intrinsic mitochondrial pathway of apoptosis. In vivo proof-of-concept has been demonstrated in xenograft studies in nude mice with suppression of tumour growth without measurable systemic toxicity at therapeutic doses [88]. **Tityustoxin-1 (Ts1)** from *Tityus serrulatus* inhibits the over-expression of the voltage-gated Na channel Nav1.5 by blocking the invasion of ovarian metastatic potential (by 45% in Matrigel invasion assays) and the velocity at which cells migrate in wound-healing assays (by ∼60%). [89].

IsCT and IsCT2, 13-residue α-helical NDBPs derived from *Opisthacanthus madagascariensis*, have potent and selective membrane-disrupting cytotoxicity against ovarian cancer cells, which selectively target the membrane because of increased phosphatidylserine exposure and decreased cholesterol content in malignant membranes compared with healthy cells, in co-culture experiments [90]. They are so short and simple in terms of linear structure that they are especially amenable to computational design of better analogues. **Pandinin-1** and **Pandinin-2** from *Pandinus imperator* have been shown to have a dual pharmacology ideal for cancer immunotherapy: direct induction of apoptosis in ovarian cancer cells as well as activation of dendritic cells (upregulation of CD80, CD86, and IL-12) and induction of macrophage M1 polarization [91]. *Centruroides margaritatus* venom contains margatoxin (MgTx), which is a selective inhibitor of Kv1.3, which is highly expressed in ovarian cancer cells and which leads to cell cycle arrest at G1 and inhibits proliferation by up to 60% in SKOV-3 cells, in

addition to its anti-migratory effects, which are consistent with Kv1.3's role in cancer cell motility [92].

2.5 Scorpion Venom Peptides as Immunomodulatory Adjuvants for Ovarian Cancer Immunotherapy

The immunomodulatory properties of scorpion venom peptides have become a very interesting aspect of the venom pharmacology, apart from their cytotoxic effects, in the development of cancer vaccines.

For the adjuvant to be effective some immunological profile is required that is not merely non-specific inflammation, but rather one that activates innate immunity PRRs on APCs, induces maturation of DC, antigen cross-presentation, polarization of the resulting adaptive response towards Th1-dominated CTL effect and ideally remodels the immunosuppressive TME to prevent rapid attenuation of the induced response [87].

Structurally, the cationic, amphipathic structure of Scorpion NDBPs closely resembles the lipid A backbone of bacterial LPS, thus enabling its ability to meet these requirements via engagement of TLR4/MD-2, the central molecular hypothesis tested on a computational level in this study.

2.5.1 Activation of Antigen-Presenting Cells

Certain Scorpion peptides can activate DC and macrophages, similar to PRR activation by TLR. Pandinins and IsCT derivatives have been found to activate DC and macrophages in a fashion similar to TLR-mediated PRR activation. This activation leads to upregulation of surface co-stimulatory molecules (CD80, CD86, CD40), increased levels of MHC class I and II molecules, greater antigen processing ability and robust secretion of IL-12p70, the key cytokine that drives polarization

of $CD4^+$ T helper cells towards a Th1 phenotype necessary for priming of $CD8^+$ CTL.

This direct activation of DCs is distinct from and complementary to the above mentioned direct anticancer activity and thus, a cytokine secretion profile induced by scorpion peptides is mechanistically essential for their adjuvant suitability which could be achieved in a single therapeutic agent not possible with a single currently approved modality[47, 48].

2.5.2 Cytokine Induction Profile

The ideal adjuvant cytokine profile for ovarian cancer vaccine applications is: IL-12p70 (Th1 polarization and DC-dependent CTL priming); IFN-γ (effector CTL and NK cell activation); type I IFNs (DC cross-presentation enhancement and antiviral immunity); balanced IL-6 (acute phase response without chronic inflammatory toxicity); and minimal production of IL-10 (Treg induction) and TGF-β (immunosuppression).

The cytokine profile of scorpion peptides that activate TLR4/NF-κB signaling includes IL-6, TNF-α, IL-12, and IFN-γ, characteristic of Th1 polarization, through the MyD88 and TRIF arms respectively.

One important parameter which should be assessed in experimental testing of the candidates identified computationally in this study is the relative quantitative output of pro and anti-inflammatory cytokines, which depends on peptide sequence, concentration and the target cell type [48, 49].

2.5.3 Tumor Microenvironment Remodeling

The immunosuppressive ovarian TME dominated by Treg infiltration, M2-TAM predominance, and T cell exhaustion by MDSC is the main obstacle to vaccine immunotherapy success in ovarian cancer. Scorpion peptides have been shown to have some initial activity against each of these immunosuppressive cell types.

Treg infiltration is decreased by the downregulation of CCL22 and CCL17 chemokine signals that promote infiltration of Tregs to tumor sites, as well as the direct effect on FoxP3 expression [93].

In a model system of ovarian tumors, the conversion of immunosuppressive IL-10+/arginase-1+ TAMs to immunostimulatory IL-12+/iNOS+ effector macrophages. (TAM repolarization) has been shown to be induced by many of the scorpion NDBPs and to lead to measurable increases in intratumoral IFN-γ and decreases in IL-10 [94].

A third mechanism of immunosuppression-reversal suggested by scorpion AMPs is that they can shift the immunologically 'cold' ovarian TME towards an 'inflamed' state, which is more suitable for checkpoint inhibitors and cancer vaccine immunotherapy by reducing the expansion of MDSCs via cytokines such as IL-6, M-CSF and GM-CSF that are produced by the tumor, and by directly promoting MDSC differentiation [95].

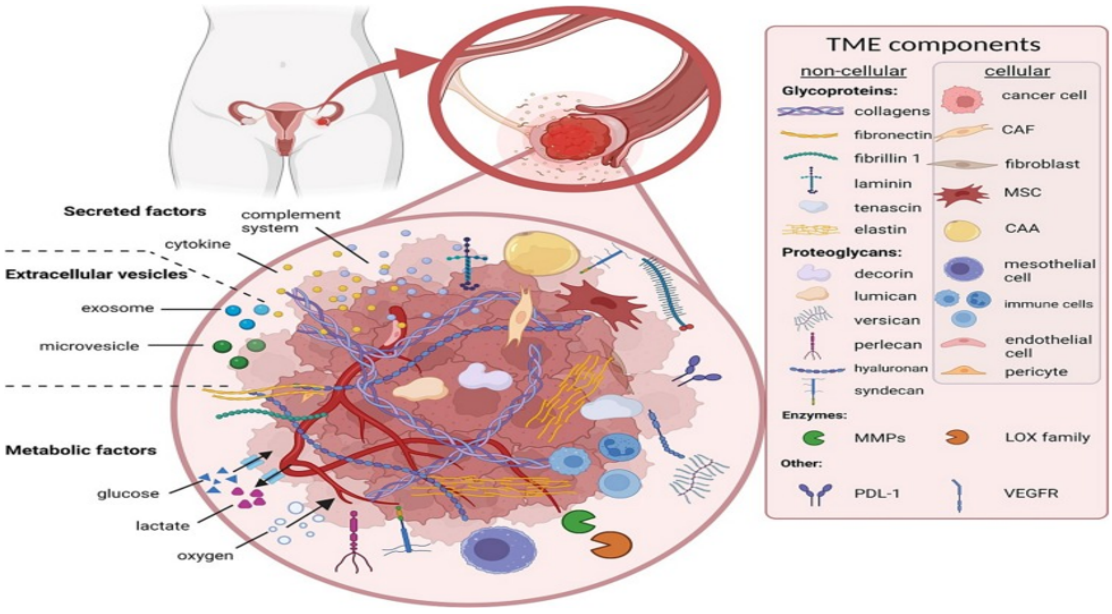


FIGURE 2.2: Tumor microenvironment in ovarian cancer comprising cellular & non cellular components.

This figure illustrates the complex tumor microenvironment of ovarian cancer, comprising cellular components such as cancer cells, immune cells, fibroblasts,

and endothelial cells, along with non-cellular components including extracellular matrix proteins, cytokines, and metabolic factors.

These interactions promote tumor progression, immune evasion, angiogenesis, and therapeutic resistance, highlighting the importance of targeting the tumor microenvironment in ovarian cancer immunotherapy [64].

2.5.4 Synergy with Immune Checkpoint Inhibitors

A very interesting new paradigm being developed in cancer immunology is the 'priming and releasing' approach in which the immune system is first activated by an adjuvant-type molecule and then the tumor is infiltrated with effector T cells, followed by the systemic delivery of checkpoint inhibitors (CI) that do not allow these cells to become functionally exhausted by PD-1/PD-L1 or CTLA-4 signaling. Venom-derived peptides have shown synergistic anti-tumor effects when used in combination with anti-PD-1 or anti-PD-L1 antibodies in preclinical mouse tumor models, where combination treatment results in better tumor regression, T-cell infiltration and survival compared to single agents. The combination approach is also an attractive strategy for ovarian cancer, as the anti-PD-1 therapy alone only yields objective responses in 6-15% of patients [90].

This combination strategy also seems to hold promise for ovarian cancer if additional PD-L1-positive adjuvant peptides could be identified to convert immunologically cold ovarian tumors into inflamed targets that are sensitive to checkpoint blockade [96].

2.6 *In Silico* Techniques in Peptide-Based Therapeutic Discovery

The use of computational biology in discovery and optimization of AMPs has changed the drug discovery timeline for peptide therapeutics from years of bioactivity - guided experimental fractionation to weeks of computational analysis. The

In silico tool kit useful in scorpion AMP discovery is divided into several methodological domains such as database mining that provides information on structures, physicochemical profiling, immunoinformatics prediction, molecular docking, molecular dynamics simulation and multi-parameter safety assessment, all of which add a different layer of information to the overall framework of candidate evaluation.

Together, these computational methods can generate a shortlist of high-confidence candidates for experimental synthesis and testing with a high probability of experimental success and a low investment in resources for compounds that are unlikely to succeed. [97].

2.6.1 Prediction of Immunomodulatory Potential

Tools for immunoinformatics trained on experimentally validated datasets of immunomodulatory peptides can be used to make sequence-based predictions of the IFN-γ induction capacity, the activation of interleukins and the propensity of activation of the PRRs.

Servers like IFN epitope and IL-predictor use machine learning models like support vector machine (SVM), random forests, and recently transformer-based architectures to predict the probability scores of query sequences based on their amino acid composition, dipeptide frequencies and experimentally characterized sequence-position-specific matrices (SPPM) of immunostimulatory peptides [98].

These predictions inform prioritization of the most likely candidates to create the Th1 polarized cytokine environment necessary for the adjuvancy of an ovarian cancer vaccine. When docked against the crystal structure of the TLR4/MD-2 complex (3FXI), peptides with high binding affinity ($\Delta G \leq$ -9.0 kcal/mol) are predicted to dock in a manner consistent with mechanism of PRR activation, indicating that they likely interact with the TLR4/MD-2 complex in a similar manner.

2.6.2 Toxicity and Safety Assessment

Computational safety profiling of AMP candidates include three main parameters: Cytotoxicity against mammalian cell lines, Hemolytic activity against erythrocytes, Allergenicity potential. ToxinPred, HemoPred and HemoPI classifiers are machine learning classifiers for cytotoxicity and hemolysis based on primary sequence features that provide binary (Toxin/Non-Toxin) or probabilistic (Active/Not Active) safety classification of the peptides. AllerTOP/AllerHunter is a machine learning classifier for allergenicity based on primary sequence features that provides binary (Active/Not Active) or probabilistic (Allergen/Non-Allergen) safety classification of the peptides. The ability to computationally determine safe filtering of sequences that retain the cationic amphipathic architecture characteristic of both antimicrobial and TLR4-activating activity is especially useful for scorpion AMPs, where the two are coupled to non-specific membrane toxicity. The half-life prediction server HLP offers extended pharmacokinetic profiling, which calculates the stability of peptides in mammalian in vitro and in vivo setting to help to identify candidates with too short a biological half-life before committing resources to experimental testing [99].

2.6.3 Molecular Docking and Structural Analysis

The peptide-receptor interaction analysis is the computational core of this study, which is known as molecular docking simulation. The empirical scoring function used in AutoDock Vina through the PyRx virtual screening platform is based on weighted terms for van der Waals dispersion-repulsion between the receptor and peptide, hydrogen bond directionality penalties, electrostatic desolvation, and torsional entropy, to predict binding free energies (ΔG, kcal/mol) for combinations of receptor and peptide poses. For new peptide-based ligands, for which the binding site on TLR4/MD-2 is not known from experimental data, the methodologically sound way is to perform blind docking, in which there is no pre-specified receptor site, but rather the docking algorithm searches the complete receptor surface [99]. These docking calculations model the human TLR4/MD-2 extracellular domain

in its antagonist-bound conformation with atomic resolution, based on the crystal structure PDB:3FXI.

2.6.4 Molecular Dynamics Simulation

A docking is a static description of the energetically favourable binding poses of a protein complex, whereas a molecular dynamics (MD) simulation describes the stability of the protein complex over nanosecond to microsecond time scales under conditions that mimic the physiological environment, including explicit solvent, temperature 310 K and neutral pH with ionic strength approximating the physiological salt conditions [100].

MD simulation of top-ranked docking hits (e.g. GROMACS or NAMD, CHARM-M36 or AMBER19sb force fields) can be used to calculate RMSD of binding pose stability, binding free energy (MM-GBSA or MM-PBSA post-processing), per-residue energy decomposition to identify critical contact residues, and radius of gyration profiles to assess peptide compactness during receptor engagement [100].

All these parameters together determine whether docking-predicted binding poses are real, stable binding events or thermodynamically unstable encounter complexes of no therapeutic relevance.

2.6.5 Homology Modeling and Structure Prediction

When the three-dimensional structures for scorpion NDBP candidates are not available from experiments, they are predicted using PEP-FOLD3 and AlphaFold2, which are secondary and tertiary structure prediction tools. In PEP-FOLD3, a conformational letter approach based on the hidden Markov model is used to calculate 5 low-energy conformers for short peptides (5-50 residues) that can be selected for docking the global minimum. Although mainly developed for globular protein domains, AlphaFold2 can be used to reliably predict the secondary structures of peptides between 10 and 30 aa using its evoformer architecture which

incorporates evolutionary covariation signals from multiple sequence alignments [101]. The systematic characterization of all 117 candidates at the end of the pool of AMPs reported in Chapter 3 is performed by PROTPARAM (ExPASy) which computes physicochemical descriptors: molecular weight, isoelectric point, instability index, aliphatic index, GRAVY from primary sequence.

2.6.6 Virtual Screening and Database Mining

The computational pipeline starts from the two main curated databases of experimentally validated AMPs, APD3 and DRAMP, containing more than 5,000 characterized sequences, which are virtually screened. APD3 (Antimicrobial Peptide Database 3) contains experimental data on the minimum inhibitory concentrations of the peptides, hemolytic activity data, and cross referenced biological activity annotations, while DRAMP (derived from APD) contains complementary coverage with additional structural classification data and greater taxonomic representation of venomous animal AMPs, including multiple scorpion genera unrepresented in APD3. The ability of both databases to return taxon-specific lists of AMPs and the associated species, activity and sequence metadata allows for return of scorpion-specific lists of AMPs with associated species, activity and sequence metadata, which is the input for a five-phase computational screening pipeline described in Chapter 3 [102] that creates the 117-entry master library.

Chapter 3

Methodology

3.1 Overview of Computational Pipeline

In this study, an *In silico* pipeline was used to identify, characterize and optimize scorpion derived antimicrobial peptides (AMPs) as potential immunomodulatory adjuvants for the development of vaccines against ovarian cancer. All analyses were performed on web servers and software that is freely available.

3.2 Tools and Databases

All seven phases of this study used computational tools, servers and databases. Each platform is outlined below based on its purpose, web address, and output parameters produced. It is a detailed and extensive internet-based resource for the study of antimicrobial peptides (AMPs) of natural, synthetic and predicted origin.

3.2.1 Antimicrobial Peptide Database

APD3 (https://aps.unmc.edu/) provides valuable information for drug discovery and vaccine development, and is complemented by peptide sequences, structures,

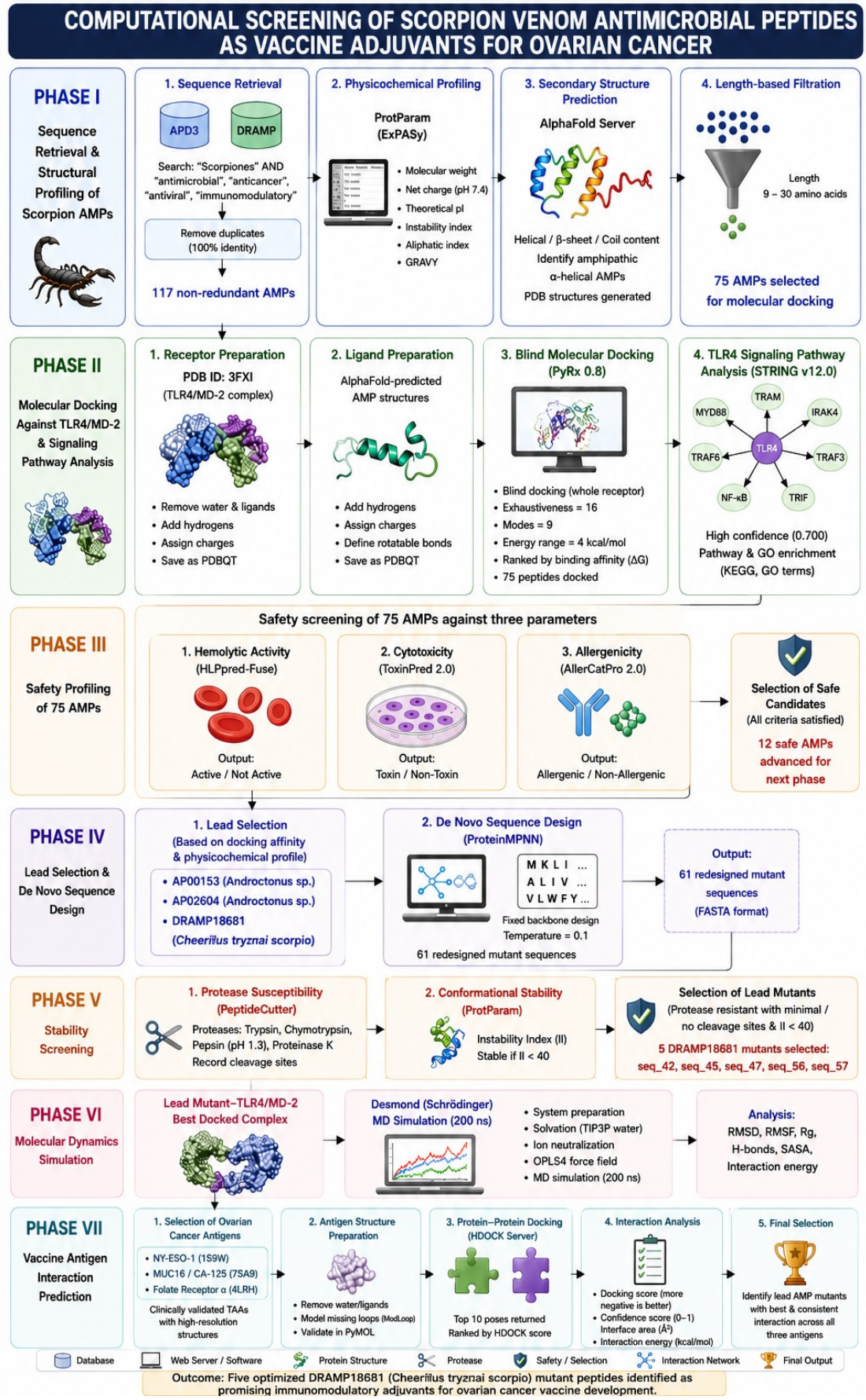


FIGURE 3.1: Flowchart of methodology.

origins, and detailed specific biological activities (antibacterial, antiviral, antifungal and antitumor properties).

The database includes advanced search and prediction capabilities, including the ability to calculate physicochemical parameters, analyze amino acid composition, and compare the sequences without alignment, which can be used to design new peptide-based therapeutics or adjuvants, based on the database.

3.2.2 Data Repository of Antimicrobial Peptides

The second major source for sequences of AMPs derived from scorpions was a database called DRAMP (https://dramp.cpu-bioinfor.org/), which was used to access a more comprehensive and independently curated database of entries for antimicrobial peptides.

DRAMP also provided additional structure data for the deposited entries along with their primary amino acid sequences and biological activity annotations.

3.2.3 ProtParam

ProtParam is an ExPASy (https://web.expasy.org/protparam/) server of the Swiss Institute of Bioinformatics which is widely used for the calculation of several physical and chemical properties of a protein sequence online.

ProtParam computes fundamental properties of the protein: Aminoacid composition, molecular content, molecular weight, theoretical isoelectric point, grand average of hydropathicity (GRAVY), aliphatic index, extinction coefficient, estimated half-life, and instability index based on an entered amino acid sequence or a specific UniProt knowledge base entry.

The metrics are important baseline parameters for researchers who study protein structure, stability and behaviour during laboratory purification and expression experiments.

3.2.4 AlphaFold Server

AlphaFold uses a deep learning model with the Evoformer neural network, which combines several sequence alignment (MSA) co-evolutionary data and structural template information to generate highly accurate per-residue coordinates in PDB format.

3.2.5 PyRx 0.8

PyRx (Python Prescription) (https://www.python.org/) is an open source, free, virtual screening software for computational drug discovery. It streamlines screening of huge library of small molecules against potential drug targets via molecular docking with AutoDock Vina or other docking software.

3.2.6 STRING v12.0

The STRING database (Search Tool for the Retrieval of Interacting Genes / Proteins), which is a resource of known and predicted protein-protein interactions.

It brings together physical and functional interactions from a variety of sources spanning more than 12,000 organisms including experimental data, computational predictions, co-expression, co-location on the genome, and information gleaned from literature.

The new v12.0 features support for novel genomes submitted by users, improved co-expression predictions with variational auto-encoders (VAE) (including single cell RNA-seq and proteomics), confidence scoring for experimental interactions, and new, improved tools for functional enrichment analysis.

STRING is a widely used tool for visualization, analysis and hypothesis generation in systems biology and in biomedical research. This tool is available at https://string-db.org/.

3.2.7 HLPpred-Fuse

HLPpred-Fuse is a two-layer machine learning approach for accurate prediction of hemolytic peptides (HLPs) and their activity level directly from peptide sequences. First, it categorizes peptides as "hemolytic" or "non-hemolytic," and then it utilizes multiple representations of the sequence to determine whether the peptide is hemolytic or not for a peptide that is categorized as a "hemolytic. The tool is available as a user friendly web server at (http://thegleelab.org/HLPpred-Fuse/).

3.2.8 ToxinPred 2.0

The tool has been engineered for peptide-based drug design and toxicity studies, and can be used as an easily accessed web-server at ToxinPred 2.0 (webs.iiitd.edu.in /raghava/toxinpred2) is a binary peptide sequence classifier based on support vector machine for the prediction of cytotoxic activity of peptides against the mammalian cell line..

3.2.9 AllerCatPro 2.0

AllerCatPro 2.0 (allercatpro.bii.a-star.edu.sg) is a sequence and structure-based allergenicity prediction server developed at the Bioinformatics Institute (A*STAR, Singapore), which checks the similarity between input peptides and known allergens included in the WHO/IUIS Allergen Nomenclature Database and the Aller Hunter reference set. The server calculates a probabilistic score for the allergenicity of each input along with a classification of Allergenic or Non-Allergenic.

3.2.10 HLP Server

HLP Server (webs.iiitd.edu.in/raghava/hlppred) is a dedicated website for predicting the half-life of peptide sequences in three physiological compartments, mammalian reticulocytes, yeast and *Escherichia coli*.

3.2.11 PeptideCutter

PeptideCutter (web.expasy.org/peptide_cutter) is an ExPASy tool for predicting the sites of proteolytic cleavages of peptide sequences, using the experimentally determined specificity of over 30 proteases and chemical cleavage agents.

3.2.12 ProteinMPNN

ProteinMPNN is a deep-learning based algorithm designed for de novo protein design with a particular emphasis on the inverse folding problem where an amino acid sequence is predicted to fold into a target pre-defined three-dimensional backbone structure.

As a geometric message passing neural network, it estimates the spatial coordinates of the protein backbone to produce high sequence optimised candidates.

ProteinMPNN is much faster than traditional physics based approaches such as Rosetta, has a significantly higher success rate in the experimental wet-lab validation and offers control over sequence conservation, which makes it a key tool for the design of novel enzymes, therapeutics, and stable vaccine candidates.

3.2.13 HDOCK Server

HDOCK server (hdock.pros.tau.ac.il) is a hybrid protein-protein and protein-DNA docking server that uses global Fast Fourier Transform (FFT) docking based shape complementarity sampling and atomic level knowledge-based energy scoring.

3.2.14 Desmond

Desmond (schrodinger.com/desmond) is a high performance molecular dynamics simulation engine that is optimized for biomolecular systems on GPU architectures.

3.2.15 RCSB Protein Data Bank

The RCSB Protein Data Bank (rcsb.org) is the major open-source archive of three dimensional macromolecule structure data, and all receptor and antigen crystal structures were obtained from this source in the present study.

3.3 *In Silico* Pipeline

3.3.1 Phase I: Sequence Retrieval and Structural Profiling of Scorpion AMPs

3.3.1.1 Sequence Retrieval

The sequences of AMPs derived from scorpion were retrieved from two database, i) from Antimicrobial Peptide Database (APD3) [103] and ii) Data Repository of Antimicrobial Peptides (DRAMP) [104]. In both databases, the taxonomic filter 'Scorpiones' was used with activity related keywords such as 'antimicrobial,' 'anticancer,' 'antiviral,' and 'immunomodulatory. The following metadata were recorded for each entry retrieved:

i. Accession identifier

ii. Cource species

iii. Reported biological activities

iv. Primary amino acid sequence (one letter IUPAC code).

After retrieval from both databases, duplicate sequences, which were defined as sequences with 100% identity between the two databases when exactly matching were found, were removed. This made it possible to guarantee that the desired sequence was not repeated more than once in the library of sequences created. The number of 117 non-redundant scorpion-derived AMP sequences were selected as the initial dataset.

3.3.1.2 Primary Structure Physicochemical Profiling

All 117 sequences were submitted to the ProtParam tool (web.expasy.org/protparam) [105] for the computation of the physicochemical pr all 117 sequences were submitted to the ProtParam tool (web.expasy.org/protparam) [105] for the computation of the physicochemical properties. They each received a single sequence and the following descriptors were extracted:

i. Moles (mol)

ii. Isoelectric point (pI) (theoretical)

iii. Net charge at pH 7.4

iv. Instability index (II): The instability index is calculated using experimentally determined weights for the dipeptides, and is a continuous numerical prediction of protein stability in solution: an II of less than 40 means a conformationally stable molecule, and an II of 40 or above suggests that it may be unstable in physiological conditions.

v. Aliphatic index (AI): A positive value for the aliphatic index (AI) is a sign of thermal tolerance in peptides, as the higher the value the more will be the relative contributions of alanine, valine, isoleucine and leucine residues to thermostability.

vi. Grand Average of Hydropathicity (GRAVY): The overall score of hydrophobicity of a sequence is calculated as GRAVY score: a positive value shows the net hydrophobic surface; a negative shows the net hydrophilic surface of a sequence.

All these parameters were the structural fitness for interaction with the membrane and the engagement with the receptor in each candidate. These were used along with secondary structure predictions to select candidates with a physicochemical profile most compatible for binding TLR4/MD-2.

3.3.1.3 Secondary Structure Prediction

All 117 AMPs were modeled in three-dimensional with the help of the AlphaFold protein structure prediction server [106]. To determine the ratio of residues in each structure that are in an alpha-helical conformation, beta-sheet conformation, and random coil conformation, the secondary structure contents of the predicted structures were assessed.

This led to the identification of amphipathic alpha-helical architecture, where the hydrophobic and cationic residues were separated by 180 degrees around the helical axis (and is the preferred architecture for binding candidates of TLR4/MD-2). This is the known architecture of lipopeptide TLR4 agonists bound in the hydrophobic binding pocket of the MD-2 co-receptor[107]. All the structural models were predicted and saved as PDB structure for further molecular docking studies.

3.3.1.4 Sequence Length Based Screening

The resulting 117 entry master AMP library generated during database mining and sequence curation was put through a sequence length filter, after physicochemical profiling and secondary structure prediction as detailed earlier in this chapter. Only those AMPs that had a primary sequence of 9 to 30 amino acids were kept for downstream analysis [101].

3.3.2 Phase II: Molecular Docking and Signaling Pathway Analysis

3.3.2.1 Receptor Preparation

The crystal structure of the human Toll-like receptor 4 (TLR4) in complex with its co-receptor MD-2 was obtained from the RCSB Protein Data Bank (rcsb.org) with accession 3FXI. At 2.70 Åresolution, this structure is the antagonist bound form of the extracellular domain of the TLR4/MD-2 heterodimer. It was chosen as the

docking receptor because it is the biologically relevant, fully assembled docking receptor complex that is responsible for recognition of lipopolysaccharide (LPS) and lipopeptide agonists and because it is highly resolved for accurate computational docking at the atomic level. The docking structure of the receptor was prepared with PyRx's molecular visualization and preparation module before the docking procedure. The water molecules and co-crystallised ligands were removed from the structure. Polar hydrogen atoms were added and the Gasteiger charges were calculated and then assigned to all the receptor atoms. The processed receptor was converted to PDBQT format, which is needed by the AutoDock Vina docking engine which is part of PyRx.

3.3.2.2 Ligand Preparation

The three-dimensional structure of the 75 AMP candidates were downloaded from the AlphaFold server (Phase I) in PDB format and prepared for docking with PyRx [108]. Hydrogen atoms were added, residues that were not standard were corrected (where present), all atoms were assigned Gasteiger partial charges, and rotatable bonds were identified for flexible docking. The structures of each of the ligands were saved in PDBQT format. No explicit energy minimisation was done before docking since the structures predicted by AlphaFold were assumed to be the global minimum-energy conformers of the sequences. The Universal Force Field (UFF) energies were, however, computed internally by PyRx when the ligand was prepared and recorded for each candidate and used as a structural quality metric to flag structures with abnormal conformational energy.

3.3.2.3 Blind Molecular Docking Protocol

Molecular docking was carried out using PyRx 0.8 with the in-built AutoDock Vina 1.1.2 engine [108] using a blind docking protocol without a pre-specified binding site. The search space for the AutoDock Vina (PyRx) was designed to include all of the accessible surface regions of the extracellular domain of the TLR4/MD-2 receptor complex, and each peptide ligand was exhaustively sampled

for global preferred binding geometry by performing a large number of stochastic conformational searches.

The docking parameters were set as (i) exhaustiveness = 16 (default value), (ii) the maximum number of binding modes = 9, and (iii) energy range = 4 kcal/mol. The scores for each of the 75 ligands were calculated nine times with different poses for each ligand by the composite scoring function in AutoDock Vina, which uses a weighted sum of van der Waals dispersion-repulsion interactions, directional hydrogen bond contributions, hydrophobic burial, electrostatic desolvation penalties and torsional entropy terms to estimate binding free energy (ΔG, kcal/mol).

For each candidate the best binding affinity value across the nine poses (mode 1; minimum ΔG value) and the mean binding affinity across all nine poses (mode 2; average ΔG value) was recorded, and the pose-to-pose spread (mode 3; difference between the maximum and minimum values of ΔG across the nine poses) was also calculated.

For all 75 candidates, the ΔG value was found to be negative for all of the generated poses, indicating that the receptor contacts were thermodynamically favourable in all cases. Results were split according to ΔG values for selection of best candidates for further study.

3.3.2.4 TLR4 Signaling Pathway Analysis

To characterize the immunological signaling cascades TLR4 protein-protein interaction (PPI) network analysis was conducted using STRING v12.0 (string-db.org) [109]. TLR4 (Homo sapiens; UniProt: O00206) entered as the query protein, with a high confidence interaction score (0.700), and the interactors in the first shell (up to 10) were selected for the results.

The downstream TLR4 signaling transducers were searched on the resulting network. STRING was used to perform functional enrichment analysis to determine the pathway terms and biological process terms (both KEGG and Gene Ontology (GO)) that were significantly enriched in the TLR4 interaction network.

3.4 Phase III: Safety Profiling of Predicted A-MPs

In silico safety screening of all 75 AMPs was carried out to against four parameters. Safety profiling was performed after docking to preserve the structural diversity of the docking dataset and to screen out the AMPs that might have safety and toxicity issues. Hemolytic activity, cytotoxicity, allergenicity and conformational stability were the 4 parameters assessed.

3.4.1 Hemolytic Activity Prediction

Hemolytic activity (capacity to lyse erythrocytes) was predicted for all 75 AMPs using HLPpred-Fuse [110].

Each sequence was submitted in FASTA format, and the binary classification output Active (hemolytic) or Not Active (non-hemolytic) was recorded for each candidate.

Peptides classified as hemolytically active were excluded from the candidate pool, because erythrocyte lysis constitutes a primary disqualifying property for injectable adjuvant formulations intended for systemic or parenteral administration.

3.4.2 Cytotoxicity Prediction

Cytotoxic activity against mammalian cell lines were predicted using ToxinPred 2.0 [111]. Each of the 75 AMPs were submitted individually to the server, and the output classification "Toxin or Non-Toxin" was noted. Candidates classified as toxic were excluded from further analysis.

Cytotoxic peptides present an unacceptable safety liability in adjuvant applications because host tissue damage at therapeutically relevant concentrations is incompatible with the repeated-dosing regimens characteristic of prophylactic and therapeutic cancer vaccine protocols.

3.4.3 Allergenicity Prediction

Allergenicity was assessed for all 75 AMPs using AllerCatPro 2.0 [112], a sequence- and structure-based allergenicity prediction server.

AllerCatPro 2.0 evaluates the degree of sequence and structural similarity between each input peptide and known allergens catalogued in the WHO/IUIS Allergen Nomenclature Database and the AllerHunter reference dataset, generating a probabilistic allergenicity classification for each query.

Candidates predicted as allergenic were referred for exclusion, as hypersensitivity reactions represent a critical safety liability for adjuvant formulations intended for repeated parenteral administration in cancer patients, who may already carry compromised immune regulatory capacity.

3.4.4 Conformational Stability Assessment

Conformational stability was evaluated using the instability index (II) computed by ProtParam (ExPASy) [105]. The threshold of II < 40 was applied as the criterion for classifying a peptide as conformationally stable, consistent with the original definition of this metric.

Peptides with II ≥ 40 were classified as potentially conformationally labile and categarized accordingly. These values were considered alongside the cytotoxicity and hemolysis outputs to generate the final integrated safety profile for each candidate.

3.4.5 Selection of Safe Candidates

The data for all four safety screening criteria produced a pool of eligible AMPs with safety profile. Only AMPs simultaneously classified as Not Active (non-hemolytic), Non-Toxin (non-cytotoxic), and Non-Allergenic, with an instability index below 40, were designated as safety candidates for next phase analysis.

3.5 Phase IV: Lead Selection and De Novo Sequence Designing

3.5.1 Selection of Lead Candidates

Three candidates were selected as leads for de novo sequence redesign from the 12 safety compliant AMPs primarily based on their (i) docking results for their TLR4/MD-2 binding affinity in Phase II docking and (ii) their safety profile.

3.5.2 *De Novo* Sequence Design Using ProteinMPNN

The three lead AMP structures where the sequences were not known were designed new by ProteinMPNN [113]. The backbone of the AP00153, AP02604 and DRAMP18681 proteins were fixed during the entire design and submitted as PDB files for ProteinMPNN one at a time (fixed backbone design mode). A sampling temperature of 0.1 was used to trade off sequence diversity for structural compatibility, with lower temperatures sampling residue choices that are most likely to be structurally compatible and higher temperatures sampling choices that are more likely to be structurally diverse, but with less confidence.

This model was used to create a library of 60 mutant sequences (including 20 per each of the 3 AMPs). ProteinMPNN calculates thermodynamic probability scores for each of the generated sequences that reflect the probability of the assignment of the sequence to the structures that are supported by the backbone structure that is being fixed. Proteinase susceptibility and conformational stability screening for 61 mutant sequences was done in Phase V by exporting the sequences in FASTA format.

3.6 Phase V: Stability Screening

3.6.1 Protease Susceptibility Analysis

The proteolytic stability of all 60 mutant sequences was assessed by the ExPASy tool PeptideCutter (web.expasy.org/peptide_cutter) [114] that predicts cleavage

sites in a given peptide sequence from the known substrate specificities of over 30 experimentally determined proteases and chemical cleavage agents. For this analysis, four proteases that are clinically and physiologically relevant were chosen to represent the major protease environments encountered by a parenterally administered peptide adjuvant. The choice of trypsin for modelling serum and lysosomal degradation was made because it cuts after the C-terminal amino acids arginine and lysine. To represent proteolysis in systemic and inflammatory sites, chymotrypsin (high specificity), which cleaves C-terminal to phenylalanine, tryptophan, tyrosine and leucine residues was included. The choice of the pepsin to model the degradation of the stomach led to the choice of pH 1.3 for the activity. To simulate endosomal degradation after internalisation by APCs, proteinase K, a broad specificity serine protease, was added.

The number and locations of predicted cleavage sites for each of the four enzymes were noted for each mutant sequence submitted, one at a time, to PeptideCutter. Those candidates containing no cleavage sites in all four proteases or in which the number of cleavage sites was significantly less than that of their parent AMP were identified as proteolytically stable. Since the in vivo persistence of the product would be important for the effective duration of the TLR4 engagement and for the effectiveness of the adjuvants upon parenteral administration, protease resistance was considered a prerequisite.

3.6.2 Instability Screening

All 61 mutant sequences were also evaluated for instability by computing the instability index values for each redesigned sequence in parallel to the protease susceptibility testing, via the ExPASy tool ProtParam. The limits of stable classification were the same as in Phase I and Phase III, II < 40. Those sequences that scored II ≥ 40 were marked as conformationally labile and were not considered further. The instability index values of the 60 mutants were recorded with corresponding protease susceptibility profile and integrated stability index value was calculated for each mutant.

3.6.3 Selection of Lead Mutants

The mutants were selected according to the criteria of protease susceptibility and conformational stability. Both filters were applied concurrently, ideally without any or very few predicted sites for all four proteases, and with an instability index < 40.

The increased stability of the lead mutant sequences relative to other redesigned AMPs is believed to be due to the DNA's compact 10-residue backbone geometry selected by the lead AMP that ProteinMPNN redesigned to maximize inter-residue van der Waals contacts and highly stabilize the backbone geometry. It is a structural feature which is energetically available only for short and constrained backbone geometry.

3.7 Phase VI: Molecular Dynamics Simulation

Molecular dynamics (MD) simulation was used to analyze the optimized AMP candidate bound to the TLR4/MD-2 complex in a solvent physiological environment to assess its dynamic binding stability, pose retention, and interaction energy profile, using the high-performance MD engine Desmond [115] optimized for biomolecular simulation on GPU architectures.

3.8 Phase VII: Vaccine Antigen Interaction Prediction

3.8.1 Selection of Ovarian Cancer Vaccine Antigens

To test the binding of the lead mutant candidates to the corresponding vaccine antigens, three clinically validated tumour associated antigens (TAAs) of the ovary cancer were chosen as docking targets. The antigens were selected based on the

following criteria: (i) it must have been experimentally confirmed to be overexpressed in epithelial ovarian cancer; (ii) must be the subject of active clinical trials for its use as a vaccine or therapeutic antibody epitope; and (iii) must have a high-resolution three-dimensional crystal structure available in the RCSB Protein Data Bank.

i. INY-ESO-1 (Cancer/Testis Antigen 1B; CTAG1B; PDB: 1S9W, 2.0 Å) was chosen as an example of a protein that is highly immunogenic, with multiple active clinical vaccine trials, and has a well characterised cytotoxic T lymphocyte (CTL) epitope repertoire, it is also overexpressed in about 30-40% of ovarian cancers [116].

ii. MUC16 (Mucin-16; the CA-125 antigen; PDB: 7SA9, resolved at 3.0 Å) was selected . The most widely used ovarian cancer biomarker is an elevation in serum CA-125; MUC16 is an active target for antibody, CAR-T and peptide vaccine platforms. MUC16 also plays a role in peritoneal metastasis via interaction with the mesothelin, which is also a therapeutic target of interest [117].

iii. Folate Receptor Alpha (FRα; PDB: 4LRH, 2.8 Å) was chosen since it is expressed at relatively high levels in ∼90% of epithelial ovarian cancers, is targeted by clinical compounds (the monoclonal antibody farletuzumab), and its extracellular domain structure is well-resolved for computational docking analysis [118].

3.8.2 Antigen Structure Preparation

Three antigens were selected and crystal structures were downloaded from the RCSB Protein Data Bank. Each structure has been processed before docking and each structure was visually checked in PyMOL 2.5 for structural integrity and steric clashes/abnormal bond geometries. The antigen structures were processed in the processed form and saved in PDB format for docking into the HDOCK server.

3.8.3 Protein-Protein Docking via HDOCK

Protein-protein docking of lead AMP mutant sequences was carried out against each of the three antigens of the vaccine using the HDOCK server (hdock.pros.tau.ac.il) [119] HDOCK is an algorithm that does not require any prior information about the binding site. A two-stage docking protocol: Global sampling which performs a global sampling using FFT correlation over the complete three dimensional translational and rotational search space; Top-ranked poses which uses a knowledge-based scoring function that is based on the set of experimentally determined protein-protein interfaces to refine the poses selected in the global sampling stage.

HDOCK returned the top 10 predicted binding poses for each run ranked according to the HDOCK scoring function, which includes more negative scores as the more highly predicted binding affinity. The best pose of each combination of docking was chosen for comparison. The values reported for each run were the HDOCK docking score (energy units), the confidence score (0-1), the buried surface area (Å2), and the predicted interaction energy (kcal/mol) between the two proteins. All results on docked parameters were compared to determine which one of the AMP mutants had the best and consistent interaction profile with all three antigens, and the lead AMP mutant was used to predict adjuvant-antigen compatibility and to confirm the peptides for future experimental validation.

Chapter 4

Results

4.1 Phase I: Sequence Retrieval and Structural Profiling of Scorpion AMPs

4.1.1 Sequence Retrieval

Two databases of curated antimicrobial peptide sequences were searched by using the taxonomic filter 'Scorpiones', and activity keywords such as 'antimicrobial,' 'anticancer,' 'antiviral,' and 'immunomodulatory'. Two databases of curated antimicrobial peptide sequences were used to retrieve sequences: APD3 (Antimicrobial Peptide Database 3) and DRAMP (Data Repository of Antimicrobial Peptides) with the taxonomic filter 'Scorpiones' and activity keywords: 'antimicrobial,' 'anticancer,' 'antiviral,' and 'immunomodulatory'.

After performing cross-database deduplication using exact-sequence string matching, 117 scorpion-derived AMP sequences were found as a preliminary library (101 entries from APD3 and 16 entries from DRAMP).

The sequences in these sequences contain 31 genera of scorpions distributed over 5 families (*Buthidae, Scorpionidae, Chaerilidae, Urodacidae* and *Bothriuridae*)

and 312 activity annotation instances for antibacterial, antifungal, antiviral, anticancer, antiparasitic, anti-MRSA and hemolytic and it is provided in Appendix I.

4.1.2 Primary Structure Physicochemical Profiling

The 117 sequences retrieved were profiled in terms of six physicochemical descriptors calculated by the computer program ProtParam: sequence length, molecular weight, net charge at pH 7.4, theoretical isoelectric point, instability index and GRAVY score.

The length of the sequences varied between 9 and 84 amino acids (AA) with a mean of 28.0 AA. The overall NDBP-class AMPs were cationic, with a net charge of +101/117 (86.3%) for all sequences.

The GRAVY score was overall -0.448, with 60.7% of the sequences having a positive score suggesting a hydrophobic bias which is consistent with the pocket insertion of MD-2. 93 of 117 sequences (79.5%) had instability index (II) values below the stability index of 40. All 117 sequences have been provided with complete per-peptide physicochemical data in Appendix II.

Table 4.1: Summary of physicochemical profiling of the 117 retrieved AMPs.

Parameters	**Value**
Sequences profiled	117
Sequence length	Range 9–84 aa; mean 28.0 aa
Net charge (pH 7.4)	Positive in 101/117 (86.3%)
GRAVY score	Mean +0.448; positive in 60.7% of sequences
Instability index (II)	<40 (stable) in 93/117 (79.5%)

4.1.3 Secondary Structure Prediction

All 117 AMPs were modeled in three-dimensional using the AlphaFold protein structure prediction server. The predicted models were analyzed visually, and

amphipathic alpha-helical structures, in which hydrophobic and cationic residues are arranged on opposite faces of the alpha-helix, were found. All 117 sequences have been given predicted secondary structures in Appendix III.

4.1.4 Sequence Length Based Screening

The screening of retrieved AMPs was performed based on the length of sequence. Of the 117 initial sequences (Appendix I) 42 were too long to include (30 or more amino acids) and were therefore eliminated.

Sequences in this length range are almost exclusively disulfide-bridged peptides (DBPs) compared with non-disulfide-bridged peptides (NDBPs) which constitute the main class of peptides targeted in this study and interact with TLR4.

This length range also is in the reach of the downstream prediction servers (ToxinPred 2.0, HLPpred-Fuse, AllerCatPro 2.0) and it is synthetically accessible via standard Fmoc solid-phase peptide synthesis. In this step, 117 sequences were reduced to 81 sequences covering 9-29 amino acids.

4.1.5 Physicochemical Properties of 75 Retained AMPs

The physicochemical properties of the 75 retained AMPs were checked by ProtParam, which provides the molecular weight, net charge, theoretical pI, instability index, aliphatic index and GRAVY score for each sequence. The molecular weight ranged from around 985 Da to 9,202 Da corresponding to the wide range of molecular lengths (9–29 residues) obtained after the first filtering step. Net charge was mostly cationic, with the majority of peptides having +1 to +9. This is in line with the electrostatic requirement to interact with anionic surface of membrane and receptor. The candidate pool exhibited a wide range of membrane-partitioning potential, as reflected by the wide range of GRAVY values. The same was true for the values of instability index and aliphatic index, which served as a preliminary criterion for downstream ranking of thermostable, structurally suitable candidates.

TABLE 4.2: Physicochemical properties of the 75 retained AMPs.

No.	DB ID	Source	Species	Sequence	Len	MW (Da)	Hydro %	Chg	pI	II	GRAVY
1	AP01512	APD3	*Isometrus maculatus*	FFSLLPSLIGGLVSAIK	17	1762.17	58%	+2	8.75	34.34	1.594
2	AP01978	APD3	*Buthus martensii*	FIGAIARLLSKIF	13	1448.81	69%	+3	11.00	-3.83	1.592
3	AP02216	APD3	*Vaejovis mexicanus smithi*	FLGALWNVAKSVF	13	1451.73	69%	+2	8.75	30.41	1.208
4	DRAMP03688	DRAMP	*Tityus serrulatus*	FLGMIPGLIGGLISAFK	17	1734.17	53%	+1	8.75	6.14	1.547
5	DRAMP03687	DRAMP	*Tityus serrulatus*	FLSLIPSLVGGSISAFK	17	1736.08	53%	+1	8.75	12.47	1.324
6	AP00153	APD3	*Androctonus sp.*	RSVCRQIKICRRRGGCYYKCT NRPY	25	3080.66	28%	+8	10.22	95.12	-1.056
7	AP00430	APD3	*Opisthacanthus madagascariensis*	ILGKIWEGIKSLF	13	1503.85	53%	+2	8.59	-2.08	0.777
8	AP00660	APD3	*Pandinus imperator*	FWGALAKGALKLIPSLFSSFSK KD	24	2612.11	50%	+3	10.00	2.42	0.329
9	AP01327	APD3	*Androctonus australis*	LFGLIPSLIGGLVSAFK	17	1732.14	58%	+1	8.75	17.46	1.618
10	AP01544	APD3	*Mesobuthus eupeus*	IFGAIAGLLKNIF	13	1376.70	69%	+2	8.75	29.97	1.700
11	AP01545	APD3	*Mesobuthus eupeus*	FFGHLFKLATKIIPSLFQ	18	2107.57	55%	+2	10.00	0.76	0.856
12	AP01791	APD3	*Chaerilus tricostatus*	FLWGLIPGAISAVTSLIKK	19	2014.48	57%	+3	10.00	1.56	1.163
13	AP01977	APD3	*Mesobuthus martensii*	FLFSLIPSAISGLISAFK	18	1911.31	61%	+2	8.75	10.19	1.544

Table 4.2 continued from previous page

No.	DB ID	Source	Species	Sequence	Len	MW	Hydro %	Chg	pI	II	GRAVY
14	AP01980	APD3	*Opisthacanthus madagascariensis*	IFGAIWNGIKSLF	13	1465.76	61%	+2	8.75	-12.43	1.138
15	AP02026	APD3	*Scorpiops tibetanus*	GFWGKLWEGVKSAI	14	1577.85	50%	+2	8.59	4.61	0.143
16	AP02027	APD3	*Scorpiops tibetanus*	GFWGSLWEGVKSVV	14	1550.78	50%	+1	6.00	16.74	0.514
17	AP02098	APD3	*Heterometrus spinifer*	SGTSEKERESGRLLGVVKRLIVCFRSPFP	29	3248.79	34%	+3	10.04	70.92	-0.276
18	AP02124	APD3	*Urodacus yaschenkoi*	GFWGKLWEGVKNAI	14	1604.87	50%	+2	8.59	4.61	-0.050
19	AP02125	APD3	*Urodacus yaschenkoi*	FWGKLWEGVKNAI	13	1547.82	53%	+2	8.59	4.19	-0.023
20	AP02126	APD3	*Urodacus yaschenkoi*	ILSAIWSGIKSLF	13	1434.74	61%	+2	8.75	10.98	1.392
21	AP02127	APD3	*Urodacus yaschenkoi*	IWSAIWSGIKGLL	13	1443.75	61%	+2	8.75	-10.36	1.138
22	AP02213	APD3	*Pandinus imperator*	FFGSLLSLGSKLLPSVFKLFQRKKE	14	1546.87	50%	+1	8.59	4.96	0.671
23	AP02214	APD3	*Pandinus imperator*	GILGKLWEGFKSIV	13	1404.72	61%	+1	8.75	4.45	1.423
24	AP02215	APD3	*Pandinus imperator*	IFGAIWKGISSLL	13	1491.79	53%	+1	8.75	-5.90	0.938
25	AP02217	APD3	*Vaejovis mexicanus smithi*	FLSTLWNAAKSIF	13	1497.76	61%	+2	8.75	36.94	0.823
26	AP02218	APD3	*Heterometrus petersii*	IFKAIWSGIKSLF	13	1509.85	61%	+2	10.00	-15.39	1.077

Table 4.2 continued from previous page

No.	DB ID	Source	Species	Sequence	Len	MW	Hydro %	Chg	pI	II	GRAVY
27	AP02356	APD3	*Androctonus amoreuxi*	FLFSLIPHAIGGLISAFK	18	1931.35	61%	+3	8.76	6.35	1.433
28	AP02357	APD3	*Androctonus amoreuxi*	FLFSLIPSAISGLISAF	17	1783.14	64%	+1	5.52	19.05	1.865
29	AP02403	APD3	*Heterometrus petersii*	GILGKLWEGVKSIF	14	1546.87	50%	+1	8.59	19.71	0.671
30	AP02518	APD3	*Centruroides margaritatus*	ARDGYIVDEKGCKFACFIN	19	2149.47	40%	0	6.15	11.28	-0.053
31	AP02550	APD3	*Heterometrus laoticus*	FWGALAKGALKLIPSLVSSFTKKD	24	2578.09	50%	+4	10.00	2.42	0.392
32	AP02551	APD3	*Heterometrus spinifer*	ILGEIWKGIKDIL	13	1497.84	53%	+1	6.07	28.28	0.700
33	AP02560	APD3	*Mesobuthus martensii*	FIGAVAGLLSKIF	13	1335.65	69%	+1	8.75	-3.83	1.885
34	AP02603	APD3	*Vaejovis punctatus*	LPFFLLSLIPSAISAIKKI	19	2071.62	63%	+3	10.00	29.43	1.526
35	AP02604	APD3	*Vaejovis punctatus*	FWGFLGKLAMKAVPSLIGGNKSSSK	25	2624.14	44%	+4	10.48	40.49	0.192
36	DRAMP18443	DRAMP	*Scorpio maurus palmatus*	IWSFLIKAATKLLPSLFGGGKKDS	24	2578.09	46%	+3	10.00	33.33	0.312
37	DRAMP18418	DRAMP	*Tityus obscurus*	FIGMIPGLIGGLISAFK	17	1734.17	53%	+1	8.75	6.14	1.588
38	DRAMP18417	DRAMP	*Tityus obscurus*	FFGTLFKLGSKLIPGVMKLFSKKKER	26	3000.73	38%	+6	10.68	-10.33	0.019

Table 4.2 continued from previous page

No.	DB ID	Source	Species	Sequence	Len	MW	Hydro %	Chg	pI	II	GRAVY
39	DRAMP18416	DRAMP	*Tityus obscurus*	FIGMIPGLIGGLISAIK	17	1700.16	53%	+1	8.75	9.98	1.594
40	DRAMP18415	DRAMP	*Opisthacanthus cayaporum*	FWSFLVKAASKILPSLIGGGDDNKSSS	27	2825.21	41%	+1	8.50	59.00	0.115
41	DRAMP18414	DRAMP	*Opisthacanthus cayaporum*	GILGKIWEGVKSLI	14	1512.86	50%	+1	8.59	2.79	0.793
42	DRAMP18401	DRAMP	*Urodacus yaschenkoi*	ILSAIWSGIKGLL	13	1370.70	62%	+1	8.75	4.45	1.500
43	DRAMP18400	DRAMP	*Urodacus yaschenkoi*	FLSTIWNGIKGLL	13	1461.77	54%	+1	8.75	-12.43	0.969
44	DRAMP18399	DRAMP	*Urodacus yaschenkoi*	FPFLLSLIPSAISAIKRL	18	1986.47	61%	+2	11.00	53.34	1.328
45	DRAMP18398	DRAMP	*Urodacus manicatus*	ISQSDAILSAIWSGIKSLF	19	2036.36	53%	0	5.84	43.94	0.832
46	DRAMP18397	DRAMP	*Urodacus manicatus*	FFSALLSGIKSLF	13	1429.72	62%	+1	8.75	-3.83	1.492
47	AP02972	APD3	*Mesobuthus eupeus*	FFGHLFKLATKIIPSFFRRKNQ	22	2696.24	45%	+6	12.02	25.05	-0.091
48	AP02973	APD3	*Mesobuthus eupeus*	FLFSLIPSAISGLISAFKGRRKRDLN	26	2907.45	46%	+4	11.72	41.44	0.262
49	AP02974	APD3	*Mesobuthus eupeus*	FLFSLIPSAISGLINAFK	18	1938.34	61%	+2	8.75	10.19	1.394
50	AP02975	APD3	*Mesobuthus eupeus*	FFGHLFKLATKIIPSLFQRKKE	22	2649.22	45%	+5	10.46	2.44	-0.018
51	AP03275	APD3	*Isometrus maculatus*	FLGSLFSIGSKLLPGVIKLFQRKKQ	25	2805.45	44%	+5	11.33	16.59	0.332

Table 4.2 continued from previous page

No.	DB ID	Source	Species	Sequence	Len	MW	Hydro %	Chg	pI	II	GRAVY
52	AP03276	APD3	*Isometrus maculatus*	FFFLPSLIGGLVSAIK	16	1709.10	63%	+2	8.75	35.86	1.681
53	AP03303	APD3	*Opisthacanthus madagascariensis*	IFGKIWEGIKSLF	13	1537.86	54%	+2	8.59	-16.89	0.700
54	AP03335	APD3	*Chelifer cancroides*	FFGAIAKLAMKFLPAIYKQIQKKRK	25	2939.69	52%	+8	10.75	40.73	0.016
55	AP03639	APD3	*Heterometrus petersii*	IFKAIWSGINRLF	13	1564.89	62%	+2	11.00	-8.86	0.823
56	AP03723	APD3	*Tityus tarnensis*	FLGSLFSIGSKLLPGVFKLFSRKKQ	25	2798.41	44%	+6	11.33	21.68	0.372
57	AP03724	APD3	*Tityus tarnensis*	IFGMIPGLIGGLISAFK	17	1734.17	59%	+2	8.75	6.14	1.588
58	AP03725	APD3	*Tityus tarnensis*	FFSLIPSLIGGLVSAIK	17	1762.17	59%	+2	8.75	21.31	1.635
59	AP03941	APD3	*Liocheles australasiae*	FLSALWGVAKSLF	13	1438.73	69%	+2	8.75	19.47	1.385
60	AP03942	APD3	*Liocheles australasiae*	FPFLLSLIPSAISALKKL	18	1958.46	61%	+3	10.00	30.51	1.322
61	DRAMP18680	DRAMP	*Chaerilus tryznai*	GIADILKGLL	10	1012.26	60%	0	5.84	-5.70	1.400
62	DRAMP18681	DRAMP	*Chaerilus tryznai*	LLGGLLQSLL	10	1026.28	60%	0	5.52	97.88	1.770
63	DRAMP20775	DRAMP	*Chaerilus tryznai*	FLGFLKNLF	9	1098.35	67%	+1	8.75	-0.54	1.333
64	DRAMP18682	DRAMP	*Chaerilus tryznai*	GILDAITGLL	10	985.19	60%	-1	3.80	11.28	1.720
65	DRAMP18683	DRAMP	*Chaerilus tricostatus*	YIRDFITRRPPFGNI	15	1865.17	33%	+2	10.74	104.65	-0.467

Table 4.2 continued from previous page

No.	DB ID	Source	Species	Sequence	Len	MW	Hydro %	Chg	pI	II	GRAVY
66	DRAMP18684	DRAMP	*Chaerilus tricostatus*	GILDALTGIL	10	985.19	60%	-1	3.80	22.05	1.720
67	DRAMP18685	DRAMP	*Chaerilus tricostatus*	GVVDTLKNLLMGLL	14	1485.85	50%	0	5.84	-18.25	1.207
68	DRAMP18686	DRAMP	*Chaerilus tricarinatus*	FLFNVIPHAINATASLIKK	19	2097.53	58%	+3	10.00	3.49	0.800
69	DRAMP18687	DRAMP	*Chaerilus tricarinatus*	FLVGILPRMRGFITPFLKKVR	21	2489.15	48%	+5	12.31	9.64	0.624
70	DRAMP18688	DRAMP	*Chaerilus tricarinatus*	FLWSLIPSAISAVTSLIKK	19	2074.53	58%	+2	10.00	21.62	1.121
71	DRAMP18689	DRAMP	*Chaerilus tricarinatus*	FDLGGLIKGVVDLF	14	1492.78	57%	-1	4.21	4.05	1.271
72	DRAMP18695	DRAMP	*Chaerilus tricarinatus*	KHFGKDSNFPF	11	1323.47	27%	+1	8.60	30.81	-1.127
73	DRAMP21024	DRAMP	*Tityus stigmurus*	FFSLIPSLVGGLISAFK	17	1796.18	59%	+1	8.75	12.47	1.535
74	DRAMP21251	DRAMP	*Androctonus aeneas*	FLFSLIPSVIAGLVSAIRN	19	2017.44	63%	+1	9.75	24.59	1.584
75	DRAMP21252	DRAMP	*Androctonus aeneas*	FLFSLIPSAIAGLVSAIRN	19	1989.39	63%	+1	9.75	24.59	1.458

4.2 Phase II: Molecular Docking and Signaling Pathway Analysis

4.2.1 Receptor Preparation

The crystal structure of the human TLR4/MD-2 extracellular complex (PDB: 3FXI, 2.70 Å) was downloaded from the RCSB Protein Data Bank and then prepared for docking in PyRx, by removing water molecules, co-crystallized ligands, adding polar hydrogen atoms, and assigning Gasteiger partial charges to all receptor atoms. The structure of the receptor that has been prepared is shown in Figure 4.1.

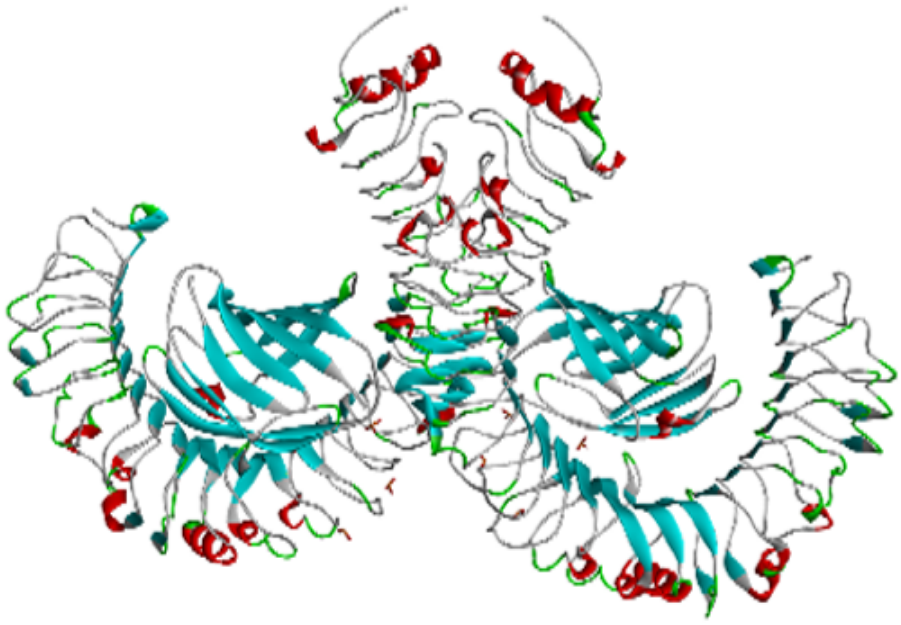


FIGURE 4.1: Prepared TLR4/MD-2 receptor complex (PDB: 3FXI).

4.2.2 Ligand Preparation

The docked three-dimensional structure of the 75 candidates for AMP was prepared using AlphaFold predictions. Before flexible docking, hydrogen atoms were added and Gasteiger partial charges were assigned to each ligand and rotatable bonds were identified.

TABLE 4.3: Secondary structure predictions for the 75 AMPs used as ligands in TLR4/MD-2 docking.

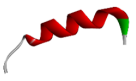
No.	AMP ID	Sequence	Structure Of AMPs
1	AP01512	FFSLLPSLIGGLVSAIK	
2	AP01978	FIGAIARLLSKIF	
3	AP02216	FLGALWNVAKSVF	
4	DRAMP03688	FLGMIPGLIGGLISAFK	
5	DRAMP03687	FLSLIPSLVGGSISAFK	
6	AP00153	RSVCRQIKICRRRGGCYYKCTNRP	
7	AP00430	ILGKIWEGIKSLF	
8	AP00660	FWGALAKGALKLIPSLFSSFSKKD	
9	AP01327	LFGLIPSLIGGLVSAFK	
10	AP01544	IFGAIAGLLKNIF	
11	AP01545	FFGHLFKLATKIIPSLFQ	
12	AP01771	FLWGLIPGAISAVTSLIKK	

Table 4.3 continued from previous page

No.	AMP ID	Sequence	Structure Of AMPs
13	AP01977	FLFSLIPSAISGLISAFK	
14	AP01980	IFGAIWNGIKSLF	
15	AP02026	GFWGKLWEGVKSAI	
16	AP02027	GFWGSLWEGVKSVV	
17	AP02098	SGTSEKERESGRLLGVVKRLIVCFF	
18	AP02124	GFWGKLWEGVKNAI	
19	AP02125	FWGKLWEGVKNAI	
20	AP02126	ILSAIWSGIKSLF	
21	AP02127	IWSAIWSGIKGLL	
22	AP02213	FFGSLLSLGSKLLPSVFKLFQRKKE	
23	AP02214	GILGKLWEGFKSIV	

Table 4.3 continued from previous page

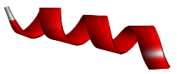
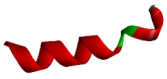
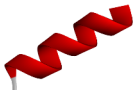
No.	AMP ID	Sequence	Structure Of AMPs
24	AP02215	IFGAIWKGISSLL	
25	AP02217	FLSTLWNAAKSIF	
26	AP02218	IFKAIWSGIKSLF	
27	AP02356	FLFSLIPHAIGGLISAFK	
28	AP02357	FLFSLIPSAISGLISAF	
29	AP02403	GILGKLWEGVKSIF	
30	AP02518	ARDGYIVDEKGCKFACFIN	
31	AP02550	FWGALAKGALKLIPSLVSSFTKKD	
32	AP02551	ILGEIWKGIKDIL	
33	AP02560	FIGAVAGLLSKIF	
34	AP02603	LPFFLLSLIPSAISAIKKI	
35	AP02604	FWGFLGKLAMKAVPSLIGGNKSSS	

Table 4.3 continued from previous page

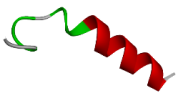
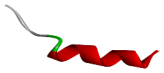
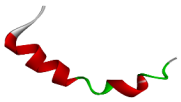
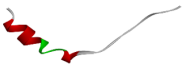
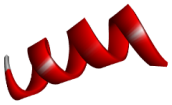
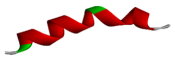
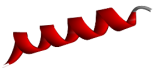
No.	AMP ID	Sequence	Structure Of AMPs
36	DRAMP18443	IWSFLIKAATKLLPSLFGGGKKDS	
37	DRAMP18418	FIGMIPGLIGGLISAFK	
38	DRAMP18417	FFGTLFKLGSKLIPGVMKLFSKKKI	
39	DRAMP18416	FIGMIPGLIGGLISAIK	
40	DRAMP18415	FWSFLVKAASKILPSLIGGGDDNKS	
41	DRAMP18414	GILGKIWEGVKSLI	
42	DRAMP18401	ILSAIWSGIKGLL	
43	DRAMP18400	FLSTIWNGIKGLL	
44	DRAMP18399	FPFLLSLIPSAISAIKRL	
45	DRAMP18398	ISQSDAILSAIWSGIKSLF	
46	DRAMP18397	FFSALLSGIKSLF	

Table 4.3 continued from previous page

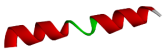
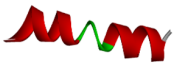
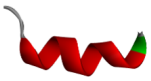
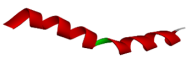
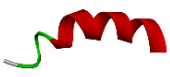
No.	AMP ID	Sequence	Structure Of AMPs
47	AP02972	FFGHLFKLATKIIPSFFRRKNQ	
48	AP02973	FLFSLIPSAISGLISAFKGRRKRDLN	
49	AP02974	FLFSLIPSAISGLINAFK	
50	AP02975	FFGHLFKLATKIIPSLFQRKKE	
51	AP03275	FLGSLFSIGSKLLPGVIKLFQRKKQ	
52	AP03276	FFFLPSLIGGLVSAIK	
53	AP03303	IFGKIWEGIKSLF	
54	AP03335	FFGAIAKLAMKFLPAIYKQIQKKRI	
55	AP03639	IFKAIWSGINRLF	
56	AP03723	FLGSLFSIGSKLLPGVFKLFSRKKQ	
57	AP03724	IFGMIPGLIGGLISAFK	
58	AP03725	FFSLIPSLIGGLVSAIK	

Table 4.3 continued from previous page

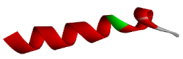
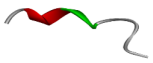
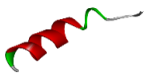
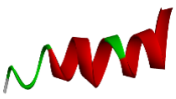
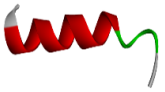
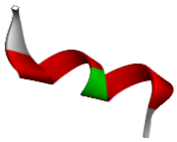
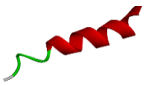
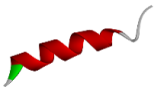
No.	AMP ID	Sequence	Structure Of AMPs
59	AP03941	FLSALWGVAKSLF	
60	AP03942	FPFLLSLIPSAISALKKL	
61	DRAMP18680	GIADILKGLL	
62	DRAMP18681	LLGGLLQSLL	
63	DRAMP20775	FLGFLKNLF	
64	DRAMP18682	GILDAITGLL	
65	DRAMP18683	YIRDFITRRPPFGNI	
66	DRAMP18684	GILDALTGIL	
67	DRAMP18685	GVVDTLKNLLMGLL	
68	DRAMP18686	FLFNVIPHAINATASLIKK	
69	DRAMP18687	FLVGILPRMRGFITPFLKKVR	
70	DRAMP18688	FLWSLIPSAISAVTSLIKK	

Table 4.3 continued from previous page

No.	AMP ID	Sequence	Structure Of AMPs
71	DRAMP18689	FDLGGLIKGVVDLF	
72	DRAMP18695	KHFGKDSNFPF	
73	DRAMP21024	FFSLIPSLVGGLISAFK	
74	DRAMP21251	FLFSLIPSVIAGLVSAIRN	
75	DRAMP21252	FLFSLIPSAIAGLVSAIRN	

4.2.3 Blind Molecular Docking Protocol

All 75 AMPs were in blind mode, docked into the structure of human TLR4/MD2 (PDB: 3FXI, 2.70 Å) using PyRx 0.8 AutoDock Vina.

Binding affinities (ΔG) were between -17.9 kcal/mol (AMP68) and -7.0 kcal/mol (AMP108) with an average of -11.83 kcal/mol (SD 2.89) and were all negative for the 75 candidates across all nine poses.

Candidates were grouped into four affinity tiers: 34 elite binders ($\Delta G \leq$ -13.0 kcal/mol, 45.3%), 14 strong binders (-11.0 to -12.9 kcal/mol, 18.7%), 7 moderate binders (-9.0 to -10.9 kcal/mol, 9.3%), and 20 weak-to-moderate binders (-7.0 to -8.9 kcal/mol, 26.7%).

The table 4.4 shows all the binding energies of all 75 AMPs.

TABLE 4.4: Docking results and complex structures for 75 AMPs against TLR4/MD-2 (PDB: 3FXI).

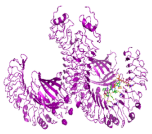
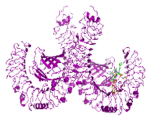
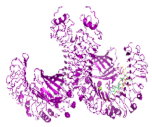
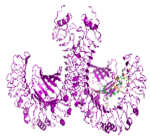
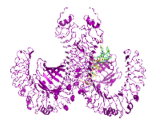
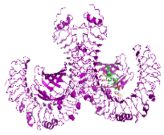
No.	**Antimicrobial Peptide ID**	**Sequence**	**Binding Energy (kcal/mol)**	**Structures of Complex**
1	AP01512	FFSLLPSLIGGLVSAIK	-11.1	
2	AP01978	FIGAIARLLSKIF	-9.7	
3	AP02216	FLGALWNVAKSVF	-10.8	
4	DRAMP03688	FLGMIPGLIGGLISAFK	-11.4	
5	DRAMP03687	FLSLIPSLVGGSISAFK	-8.3	
6	AP00153	RSVCRQIKICRRRGGCYYKCTNRPY	-15	
7	AP00430	ILGKIWEGIKSLF	-9	

Table 4.4 continued from previous page

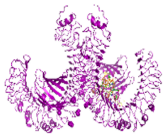
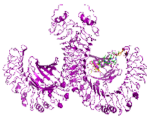
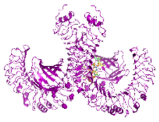
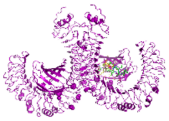
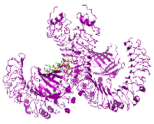
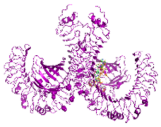
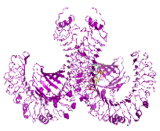
No.	Antimicrobial Peptide ID	Sequence	Binding Energy (kcal/mol)	Structures of Complex
8	AP00660	FWGALAKGALKLI PSLFSSFSKKD	-9	
9	AP01327	LFGLIPSLIGGLVSA FK	-8.8	
10	AP01544	IFGAIAGLLKNIF	-9.1	
11	AP01545	FFGHLFKLATKIIPS LFQ	-11.2	
12	AP01791	FLWGLIPGAISAVT SLIKK	-15	
13	AP01980	IFGAIWNGIKSLF	-12.6	
14	AP02026	GFWGKLWEGVKS AI	-14.9	

Table 4.4 continued from previous page

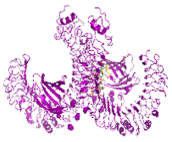
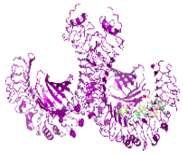
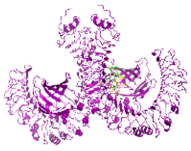
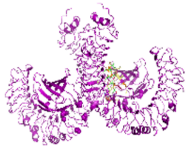
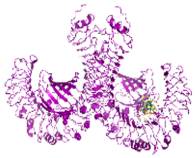
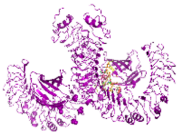
No.	Antimicrobial Peptide ID	Sequence	Binding Energy (kcal/mol)	Structures of Complex
15	AP02027	GFWGSLWEGVKSVV	-16.9	
16	AP02098	SGTSEKERESGRLLGVVKRLIVCFRSPFP	-12.1	
17	AP02124	GFWGKLWEGVKNAI	-13.9	
18	AP02125	FWGKLWEGVKNAI	-14.9	
19	AP02126	ILSAIWSGIKSLF	-12.6	
20	AP02127	IWSAIWSGIKGLL	-15.7	

Table 4.4 continued from previous page

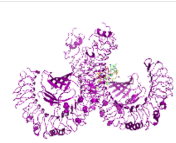
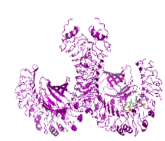
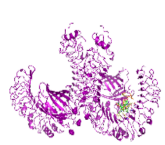
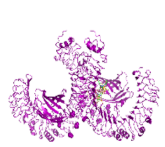
No.	Antimicrobial Peptide ID	Sequence	Binding Energy (kcal/mol)	Structures of Complex
21	AP02213	GILGKLWEGFKSIV	-12.2	
22	AP02214	IFGAIWKGISSLL	-12.8	
23	AP02215	FLSTIWNGIKSLL	-14.6	
24	AP02217	FLSTLWNAAKSIF	-13.5	
25	AP02218	IFKAIWSGIKSLF	-12.4	
26	AP02356	FLFSLIPHAIGGLISAFK	-16	
27	AP02357	FLFSLIPSAISGLISAF	-13.7	

Table 4.4 continued from previous page

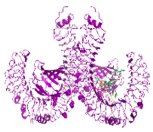
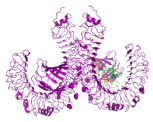
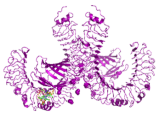
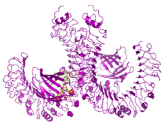
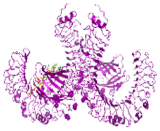
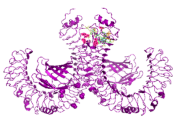
No.	Antimicrobial Peptide ID	Sequence	Binding Energy (kcal/mol)	Structures of Complex
28	AP02403	GILGKLWEGVKSIF	-13.9	
29	AP02518	ARDGYIVDEKGCK FACFIN	-12.9	
30	AP02550	FWGALAKGALKLI PSLVSSFTKKD	-14.3	
31	AP02551	ILGEIWKGIKDIL	-11.8	
32	AP02560	FIGAVAGLLSKIF	-14.6	
33	AP02603	LPFFLLSLIPSAISAI KKI	-14.6	
34	AP02604	FWGFLGKLAMKAV PSLIGGNKSSSK	-14.6	

Table 4.4 continued from previous page

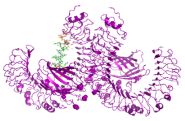
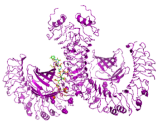
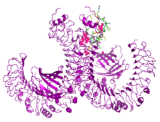
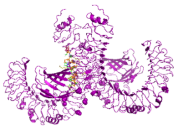
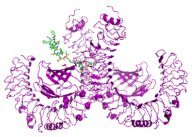
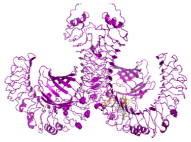
No.	Antimicrobial Peptide ID	Sequence	Binding Energy (kcal/mol)	Structures of Complex
35	DRAMP18443	IWSFLIKAATKLLPSLFGGGKKDS	-15.6	
36	DRAMP18418	FIGMIPGLIGGLISAFK	-14.0	
37	DRAMP18417	FFGTLFKLGSKLIPGVMKLFSKKKER	-13.6	
38	DRAMP18416	FIGMIPGLIGGLISAIK	-15.7	
39	DRAMP18415	FWSFLVKAASKILPSLIGGGDDNKSSS	-17.9	
40	DRAMP18414	GILGKIWEGVKSLI	-13.0	
41	DRAMP18401	ILSAIWSGIKGLL	-13.4	

Table 4.4 continued from previous page

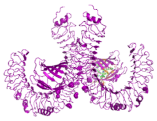
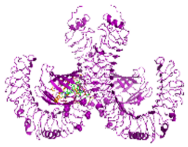
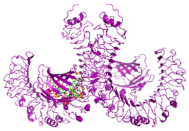
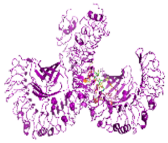
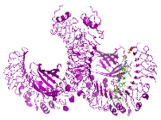
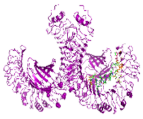
No.	Antimicrobial Peptide ID	Sequence	Binding Energy (kcal/mol)	Structures of Complex
42	DRAMP18400	FLSTIWNGIKGLL	-8.4	
43	DRAMP18399	FPFLLSLIPSAISAIKRL	-9.0	
44	DRAMP18398	ISQSDAILSAIWSGIKSLF	-7.6	
45	DRAMP18397	FFSALLSGIKSLF	-13.4	
46	AP02972	FFGHLFKLATKIIPSFFRRKNQ	-12.2	
47	AP02973	FLFSLIPSAISGLISAFKGRRKRDLN	-12.9	
48	AP02974	FLFSLIPSAISGLINAFK	-13.3	

Table 4.4 continued from previous page

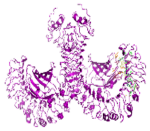
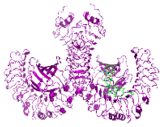
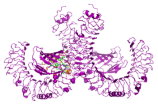
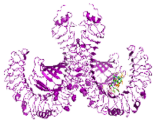
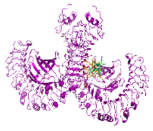
No.	Antimicrobial Peptide ID	Sequence	Binding Energy (kcal/mol)	Structures of Complex
49	AP02975	FFGHLFKLATKIIPS LFQRKKE	-12.6	
50	AP03275	FLGSLFSIGSKLLPG VIKLFQRKKQ	-14.8	
51	AP03276	FFFLPSLIGGLVSAI K	-8.2	
52	AP03303	IFGKIWEGIKSLF	-7.4	
53	AP03335	FFGAIAKLAMKFLP AIYKQIQKKRK	-13.1	
54	AP03639	IFKAIWSGINRLF	-13.4	
55	AP03723	FLGSLFSIGSKLLPG VFKLFSRKKQ	-14.2	

Table 4.4 continued from previous page

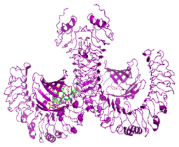
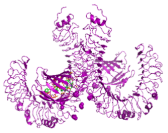
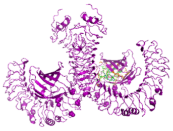
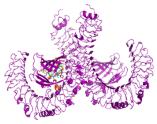
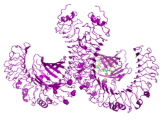
No.	Antimicrobial Peptide ID	Sequence	Binding Energy (kcal/mol)	Structures of Complex
56	AP03724	IFGMIPGLIGGLISAFK	-13.2	
57	AP03725	FFSLIPSLIGGLVSAIK	-8.1	
58	AP03941	FLSALWGVAKSLF	-9.0	
59	AP03942	FPFLLSLIPSAISALKKL	-7.8	
60	DRAMP18680	GIADILKGLL	-7.2	
61	DRAMP18681	LLGGLLQSLL	-7.8	
62	DRAMP20775	FLGFLKNLF	-9.1	

Table 4.4 continued from previous page

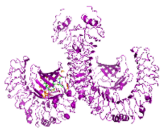
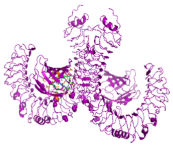
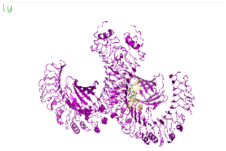
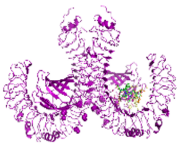
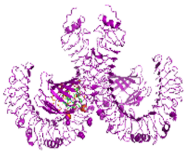
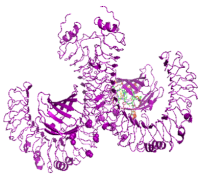
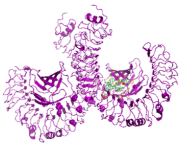
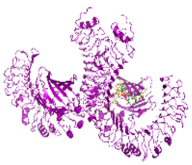
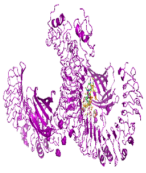
No.	Antimicrobial Peptide ID	Sequence	Binding Energy (kcal/mol)	Structures of Complex
63	DRAMP18682	GILDAITGLL	-7.2	
64	DRAMP18683	YIRDFITRRPPFGNI	-8.6	
65	DRAMP18684	GILDALTGIL	-7.2	
66	DRAMP18685	GVVDTLKNLLMGLL	-7.0	
67	DRAMP18686	FLFNVIPHAINATASLIKK	-15	
68	DRAMP18687	FLVGILPRMRGFITPFLKKVR	-13.2	
69	DRAMP18688	FLWSLIPSAISAVTSLIKK	-13.5	

Table 4.4 continued from previous page

No.	Antimicrobial Peptide ID	Sequence	Binding Energy (kcal/mol)	Structures of Complex
70	DRAMP18689	FDLGGLIKGVVDLF	-7.5	
71	DRAMP18695	KHFGKDSNFPF	-8.4	
72	DRAMP21024	FFSLIPSLVGGLISAFK	-7.8	
73	DRAMP21251	FLFSLIPSVIAGLVSAIRN	-13.4	
74	DRAMP21252	FLFSLIPSAIAGLVSAIRN	-13.7	
75	AP01977	FLFSLIPSAISGLISAFK	-13.4	

4.2.4 TLR4 Signaling Pathway Analysis

STRING v12.0 analysis of TLR4 (Homo sapiens, UniProt O00206) with a confidence threshold of 0.700 revealed 10 interactions, all with the highest possible confidence score of 0.999. These interactions lead to the differentiation of 2 signaling pathways; the first is MyD88 (TLR4-TIRAP-MyD88-IRAK4-IRAK1-TRAF6-TAK1-IKK-NF-κB), which is important for the induction of type II interferon (IL-12p70, TNF-α, and IL-6), and the second is TRIF (TLR4-TRAM-TRIF-TBK1-IRF3/IRF7), which is important for the induction of type I interferon (IFNs).

The full interaction network and interactor details are given in Table 4.5 (Figure 4.2).

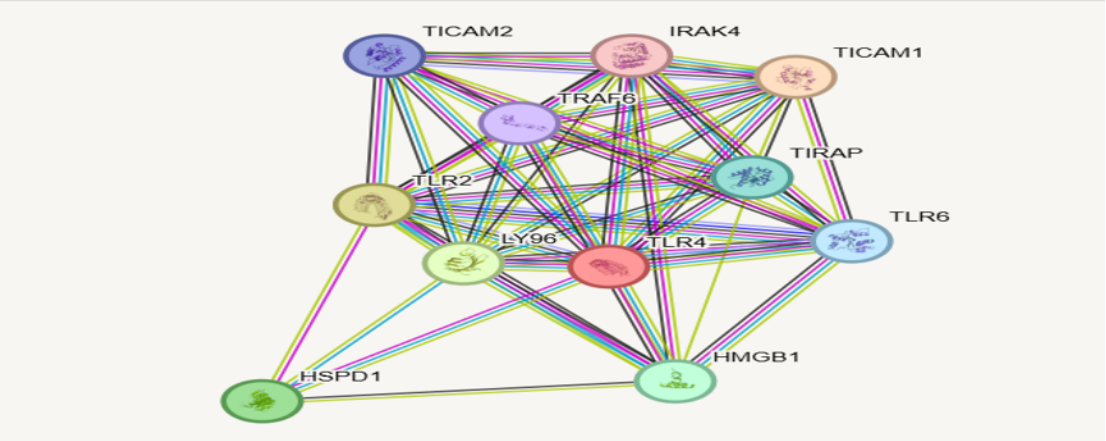


FIGURE 4.2: TLR4 protein-protein interaction network (STRING v12.0).

TABLE 4.5: TLR4 protein-protein interaction partners (STRING v12.0; confidence 0.999).

Protein	**Role in TLR4 Signaling**	**Confidence**	**Immunological Relevance to Adjuvancy**
LY96 (MD-2)	TLR4 co-receptor; LPS/AMP recognition	0.999	Direct binding partner; essential for AMP-mediated TLR4 activation and adjuvant function
TIRAP (MAL)	Bridges TLR4 to MyD88	0.999	Recruits MyD88; initiates NF-kB arm; drives IL-12 and TNF-alpha production
MYD88	Central adaptor protein	0.999	Activates IRAK4-IRAK1-TRAF6 cascade; key for pro-inflammatory cytokine induction

Table 4.5 continued from previous page

Protein	Role in TLR4 Signaling	Confidence	Immunological Relevance to Adjuvancy
TICAM1 (TRIF)	TRIF adaptor; Type I IFN arm	0.999	TBK1-IRF3 axis; drives IFN-beta/alpha; critical Th1-polarizing cytokines for CTL priming
TICAM2 (TRAM)	Bridges TLR4 to TRIF	0.999	TLR4-specific adaptor; bifurcated signaling for complete adjuvant profile
IRAK4	Kinase; MyD88 cascade initiator	0.999	First kinase in MyD88 arm; phosphorylates IRAK1; required for NF-kB and MAPK activation
TRAF6	E3 ubiquitin ligase; NF-kB activator	0.999	Activates TAK1-IKK-NF-kB; drives IL-12, IL-6, TNF-alpha; key adjuvant effector kinase
IRF3	Transcription factor; Type I IFN	0.999	Phosphorylated by TBK1; activates IFNB1 gene; Th1-polarizing IFN-beta output
IRF7	Type I IFN amplifier	0.999	Amplifies IFN-alpha/beta; essential for sustained type I IFN adjuvant effect
RELA (p65)	NF-kB subunit; cytokine transcription	0.999	Drives IL-12p70, TNF-alpha, IL-6, IL-1 transcription; central adjuvant transcription factor
TBK1	Kinase; IRF3/IRF7 phosphorylation	0.999	TANK-binding kinase; connects TRIF/TRAM to type I IFN production
HMGB1	DAMP; endogenous TLR4 ligand	0.999	Amplifies TLR4-mediated danger signaling; released from necrotic tumor cells
HSPD1 (HSP60)	Chaperone; TLR4 ligand	0.999	Endogenous TLR4 activator; establishes self-adjuvanting potential in tumor context
TLR2	Co-receptor; lipopeptide recognition	0.999	Forms TLR2-TLR4 heterodimers; expands AMP receptor repertoire
TLR6	Co-receptor; Gram+ pattern recognition	0.999	Participates with TLR2 in recognition of diacylated lipopeptides; synergistic activation

4.3 Phase III: Safety Profiling of Predicted AMP

The complete safety screening dataset is provided in Appendix V, and the screening thresholds in Table 4.6.

4.3.1 Hemolytic Activity Prediction

Thirteen out of 25 non-toxic candidates (52.0%) were classified as hemolytically active by HLPpred-Fuse and excluded. Interestingly, the highest binding affinity in the entire dataset, DRAMP18415 with a -17.9 kcal/mol binding score, was predicted to be hemolytically active (predictive value 0.64) although it was classified as non-toxic. This highlights the need to evaluate docking affinity, cytotoxicity, and hemolytic safety independently in this screening pipeline.

4.3.2 Cytotoxicity Prediction

ToxinPred 2.0 classified 50 out of 75 AMPs (66.7%) as cytotoxic, which were excluded. This high exclusion rate is a structural trade-off since the amphipathic, cationic structure that was more likely to bind to TLR4 was also associated with greater non-selective membrane disruptive activity. The top docking candidates AP02356 (-16.0 kcal/mol); DRAMP18443 (-15.6 kcal/mol); AP02127 (-15.7 kcal/mol); DRAMP18416 (-15.7 kcal/mol)) were all considered cytotoxic, meaning that docking affinity and cytotoxic safety actions are separable in this set.

4.3.3 Allergenicity Prediction

None of the 75 AMPs were excluded in this step, and the allergenicity score was 0 for all of them, with no cross-matching with the WHO/IUIS allergen database. The same outcome was observed in general as a lack of IgE-binding epitope architecture in short cationic scorpion NDBPs..

TABLE 4.6: Safety screening thresholds applied to the 75 docked AMPs.

Screening Gate	**Input**	**Eliminated**	**Output**	**Result Summary**
APD3 + DRAMP Database retrieval	–	–	117	117 non-redundant scorpion AMPs retrieved; metadata annotated; duplicates removed
Stage 1: Length filter (< 30 aa)	117	42	75	36 excluded: DBPs and long-chain peptides ≥ 30 aa; incompatible with NDBP docking/synthesis
Allergenicity (AllerCatPro 2.0)	75	0	75	75/75 = score 0 (non-allergenic); zero exclusions
Cytotoxicity (ToxinPred 2.0)	75	50	25	50 excluded as Toxin (66.7%); 25 Non-Toxin advanced
Hemolysis (HLPpred-Fuse)	25	13	12	13 of 25 Non-Toxin excluded as Active hemolytic (52%)
FINAL SELECTED SET	117	105	12	12 dual-safe AMPs: Non-Toxin + Not Active + allergenicity 0; 10.3% overall retention from 117

Only 12 AMPs were successful in fulfilling the safety screening thresholds as mentioned in Table 4.7.

TABLE 4.7: The 12 safety-compliant AMP candidates advanced to de novo design.

No.	**DB ID**	**Species**	**Sequence**	**Len**	**Chg**	**dG (kcal/mol)**	**Cytotox**	**Hemolysis**	**Allergenicity**
1	AP00153	*Androctonus sp.*	RSVCRQIKICRRRGGCYYKCTNRPY	25	+8	-15.0	Non-Toxin	Not Active	0
2	AP02357	*Androctonus amoreuxi*	FLFSLIPSAISGLISAF	17	+1	-13.7	Non-Toxin	Not Active	0
3	AP02560	*Mesobuthus martensii*	FIGAVAGLLSKIF	13	+1	-14.6	Non-Toxin	Not Active	0

Table 4.7 continued from previous page

No.	DB ID	Species	Sequence	Len	Chg	dG	Cytotox	Hemolysis	Allergenicity
4	AP02604	*Vaejovis punctatus*	FWGFLGK LAMKAVP SLIGGNKS SSK	25	+4	-15.0	Non-Toxin	Not Active	0
5	DRAMP 18398	*Urodacus manicatus*	ISQSDAILS AIWSGIKS LF	19	0	-7.7	Non-Toxin	Not Active	0
6	DRAMP 18397	*Urodacus manicatus*	FFSALLSGI KSLF	13	+1	-8.5	Non-Toxin	Not Active	0
7	AP02972	*Mesobuthus eupeus*	FFGHLFKL ATKIIPSFF RRKNQ	22	+6	-12.2	Non-Toxin	Not Active	0
8	DRAMP 18681	*Chaerilus tryznai*	LLGGLLQS LL	10	0	-7.8	Non-Toxin	Not Active	0
9	DRAMP 18682	*Chaerilus tryznai*	GILDAITG LL	10	-1	-7.2	Non-Toxin	Not Active	0
10	DRAMP 18683	*Chaerilus tricostatus*	YIRDFITR RPPFGNI	15	+2	-8.6	Non-Toxin	Not Active	0
11	DRAMP 18684	*Chaerilus tricostatus*	GILDALTG IL	10	-1	-7.2	Non-Toxin	Not Active	0
12	DRAMP 18685	*Chaerilus tricostatus*	GVVDTLK NLLMGLL	14	0	-7.0	Non-Toxin	Not Active	0

The complete four-parameter safety screening dataset for all 75 docked AMPs is provided in Appendix V.

4.4 Phase IV: Lead Selection and De Novo Sequence Designing

4.4.1 Selection of Lead Candidates

Three leads were selected from the 12 safety-compliant AMPs for ProteinMPNN redesign: AP00153 (ΔG = -15.0 kcal/mol, +8 charge, 25 aa), AP02604 (ΔG = -15.0 kcal/mol, +4 charge, 25 aa), and DRAMP18681 (ΔG = -7.8 kcal/mol, neutral charge, 10 aa).

4.4.2 De Novo Sequence Design Using ProteinMPNN

ProteinMPNN generated 61 candidate sequences (21 from AP00153, 20 from AP02604 and DRAMP18681) at a sampling temperature of 0.1.

4.5 Phase V: Stability Screening

All 61 mutant sequences of the ProteinMPNN were analyzed for proteolytic stability by using PeptideCutter against several proteases, including trypsin, chymotrypsin (high and low specificity), pepsin at pH 1.3 and proteinase K.

The DRAMP18681 derived mutant set exhibited the most favourable susceptibility profile of the three lead branches with 12 of 20 sequences yielding no predicted cleavage sites with all four enzymes.

The protease susceptibility of all de novo designed mutants of AP00153 was high as present in AppendixV.

Proteinase stability was found in Appendix VI for only one of 20 mutants derived from AP02604 while 10 mutants derived from DRAMP18681 were proteinase-stable as present in Appendix VIII.

Table 4.8: Protease susceptibility and stability profile of DRAMP18681 parent and five lead mutants.

Seq ID	**Sequence**	**Len**	**Conf**	**Mut**	**Chymo-H**	**Chymo-L**	**Pepsin pH 1.3**	**Pepsin pH >2**	**ProtK**	**Trypsin**	**Thrombin**	**II**	**Status**
Original	LLGGLLQSLL	10	–	–	0	6	8	8	6	0	0	97.88	Unstable (Parent)
seq_42*	DRGGVRDPRV	10	0.343	8	0	0	0	0	0	0	0	-2.23	STABLE FINAL LEAD
seq_45*	DRGGVRDPRI	10	0.330	8	0	0	0	0	0	0	0	-2.23	STABLE FINAL LEAD
seq_47*	DRGGVIDPRI	10	0.299	8	0	0	0	0	0	0	0	-2.23	STABLE FINAL LEAD
seq_56*	DKGGVIDPRV	10	0.294	8	0	0	0	0	0	0	0	-3.18	STABLE FINAL LEAD
seq_57*	DKGGVRDPRV	10	0.317	8	0	0	0	0	0	0	0	-3.18	STABLE FINAL LEAD

Applying the instability index criterion (II < 40) concurrently narrowed this set to five mutants with negative II values (-3.18 to -2.23): seq_42, seq_45, seq_47, seq_56, and seq_57.

4.5.1 Instability Screening

The conformational stability of the 61 mutants was also calculated using ProtParam instability index with the criteria of II < 40. Table 4.8 below provides complete protease and instability index data for all the DRAMP18681 derived mutants passing thresholds.

4.5.2 Selection of Lead Mutants

This is a significant improvement in stability from the parent DRAMP18681, which has an II of 97.88, and is caused by the change of the destabilising LLQ core motif to DRGG/DKGG motifs in the ProteinMPNN redesign. The five final mutants are described in Table 4.9.

TABLE 4.9: Five final DRAMP18681 lead mutants with stability and safety profiles.

Mutant	**Seq ID**	**Sequence**	**II**	**Conf**	**Mut**	**Trypsin**	**ProtK**	**Hemolysis**	**Structural Features**
Mutant 1	seq_42	DRGGVRDPRV	-2.23	0.343	8	0	0	Non-HLP	Asp1-Arg3 ionic dyad; Pro8 rigidity; C-term Val10; compact DRGG scaffold; neutral charge
Mutant 2	seq_45	DRGGVRDPRI	-2.23	0.330	8	0	0	Non-HLP	Identical to Mut1 except Ile10 C-term; branched hydrophobic tail; DRGG scaffold

Table 4.9 continued from previous page

Mutant	Seq ID	Sequence	II	Conf	Mut	Trypsin	Prot K	Hemolysis	Structural Features
Mutant 3	seq_47	DRGGV IDPRI	-2.23	0.299	8	0	0	Non-HLP	Val5 instead of Arg5; reduced cationicity; Pro8-Arg9-Ile10 C-term triad
Mutant 4	seq_56	DKGGV IDPRV	-3.18	0.294	8	0	0	Non-HLP	Lys2 replaces Arg; most negative II (-3.18) = highest stability in dataset; DKGG scaffold
Mutant 5	seq_57	DKGGV RDPRV	-3.18	0.317	8	0	0	Non-HLP	RECOMMENDED LEAD: II = -3.18; Lys2-Arg5 dual cationic; weakest antigen binding (HDOCK -144.91 mean)

4.6 Phase VI: Molecular Dynamics Simulation

The binding pose stability and binding energy changes of the mutant of the lead DRAMP18681 with TLR4/MD-2 were performed by molecular dynamics simulation for 200 ns in physiological conditions.

4.6.1 System and Simulation Overview

Desmond (Schrödinger LLC) with OPLS4 force field and TIP3P water was used for a 200.104 ns molecular dynamics simulation performed at NPT ensemble (300 K, 1.01325 bar). A total of 79,353 atoms were included in the system: 9,559 atoms from the protein (601 TLR4/MD-2 residues with a net charge of -16); 159 atoms of

DRAMP18681 molecule (72 heavy atoms, molecular formula C48H87N11O13 with a net charge of 0); 23,163 water molecules; 81 Na^+ and 65 Cl^- ions corresponding to a 0.15 M background NaCl solution. The number of frames recorded is 2001 with a time interval of 0.1ns between each frame.

4.6.2 RMSD Backbone Stability

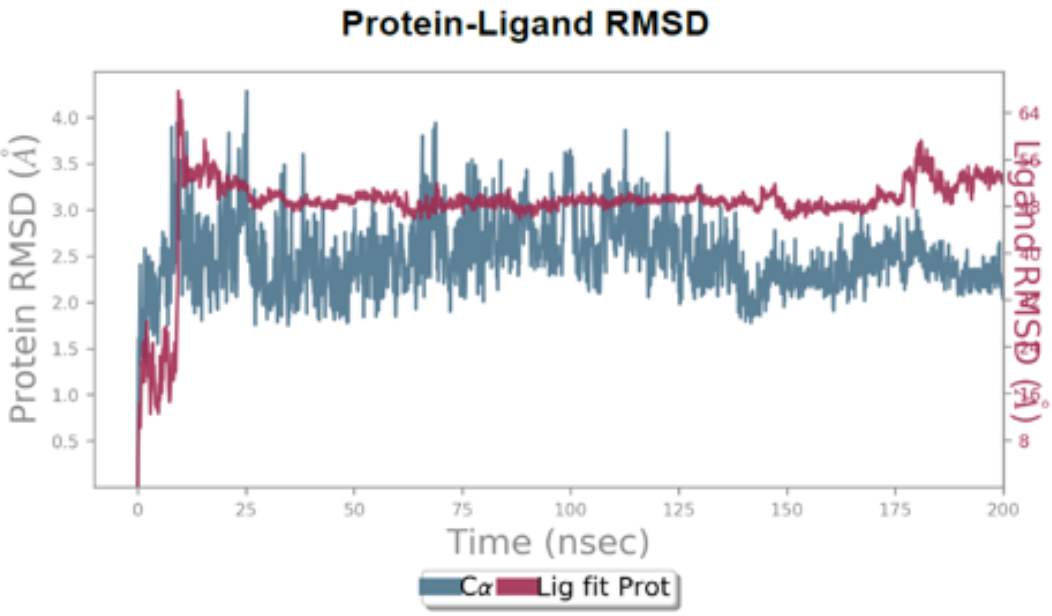


FIGURE 4.3: Backbone Cα RMSD over 200 ns.

The overall conformational behaviour of DRAMP18681-TLR4/MD-2 complex was investigated by monitoring root mean square deviation (RMSD) of backbone Cα atoms over a 200 ns trajectory.

RMSD values of the protein increased from ∼1 Åin the equilibration stage (0-25 ns) and remained at a similar level throughout the remainder of the simulation (within 2-3 Å).

The congruent convergence as seen in the decrease of the frame-to-frame variance after 50 ns, suggests that after a certain period, the system achieved a thermodynamic equilibrium, and that the binding pose after docking did not undergo significant conformational changes during the production run.

4.6.3 RMSF Per-Residue Flexibility

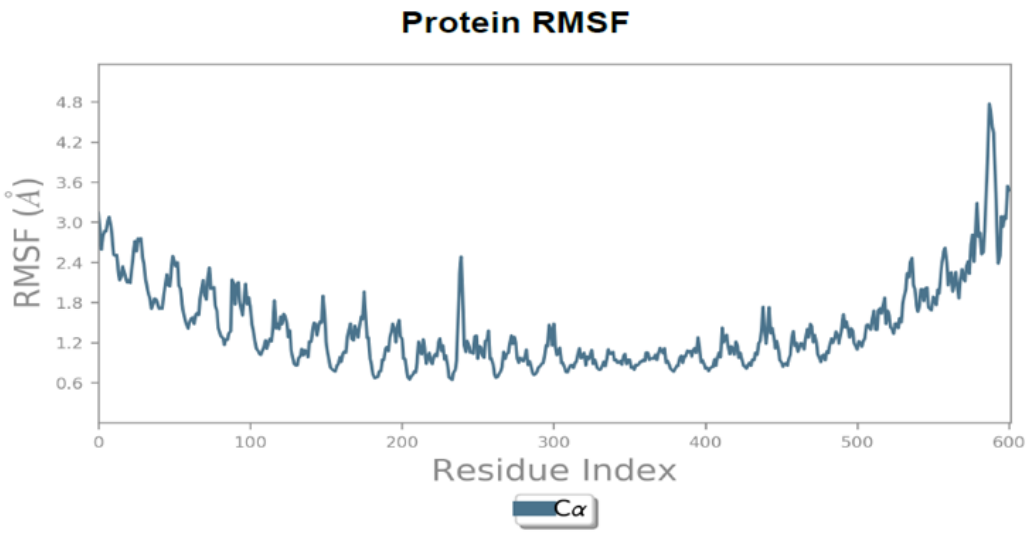


FIGURE 4.4: Per-residue RMSF of the DRAMP18681-TLR4/MD-2 complex.

To locate regions of localised mobility in the complex, the root mean square fluctuation (RMSF) was calculated per residue. Most of the residues were found to experience the fluctuations below 2 Å, indicating a well-ordered interface between them. Higher RMSF peaks (above 4-6 Å) were only observed in the terminal end regions of receptor and ligand chains and in the loop regions that are surface exposed. The high mobility regions are structurally peripheral loops that are far from the MD-2 binding groove; the flexibility does not affect the stability of the DRAM18681 binding pose at the core interface.

4.6.4 Radius of Gyration Compactness

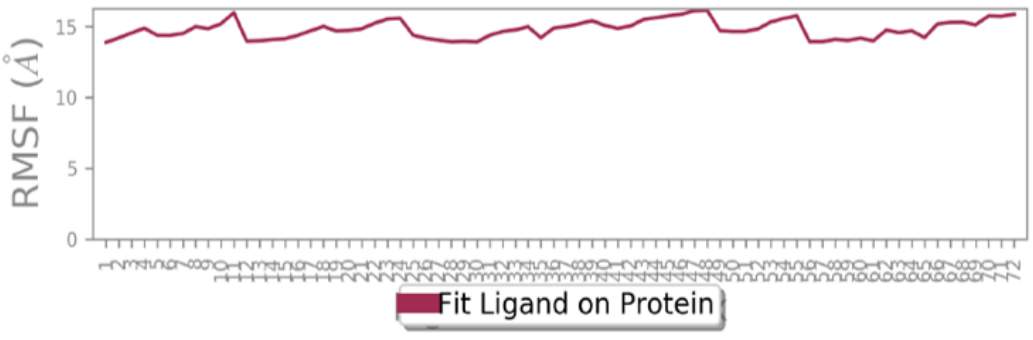


FIGURE 4.5: Radius of gyration (Rg) over 200 ns.

The radius of gyration (Rg) of the DRAMP18681-TLR4/MD-2 complex was monitored throughout the trajectory to determine if the complex maintained its folded compact form. Throughout there were only slight changes in Rg and a value of ca. 36-37 Åwas found for all 200 ns.

The lack of any progressive increase in Rg suggests that there was no partial unfolding, domain separation, or large-scale structural expansion to the complex under simulated physiological conditions, which would otherwise be required to explain the increase in Rg.

The lack of any progressive increase in Rg suggests no partial unfolding, domain separation or large scale structural expansion to the complex under simulated physiological conditions that would be required to explain this increase in Rg, supporting the structural integrity of the bound state.

4.6.5 Protein-Ligand Contacts Over Time

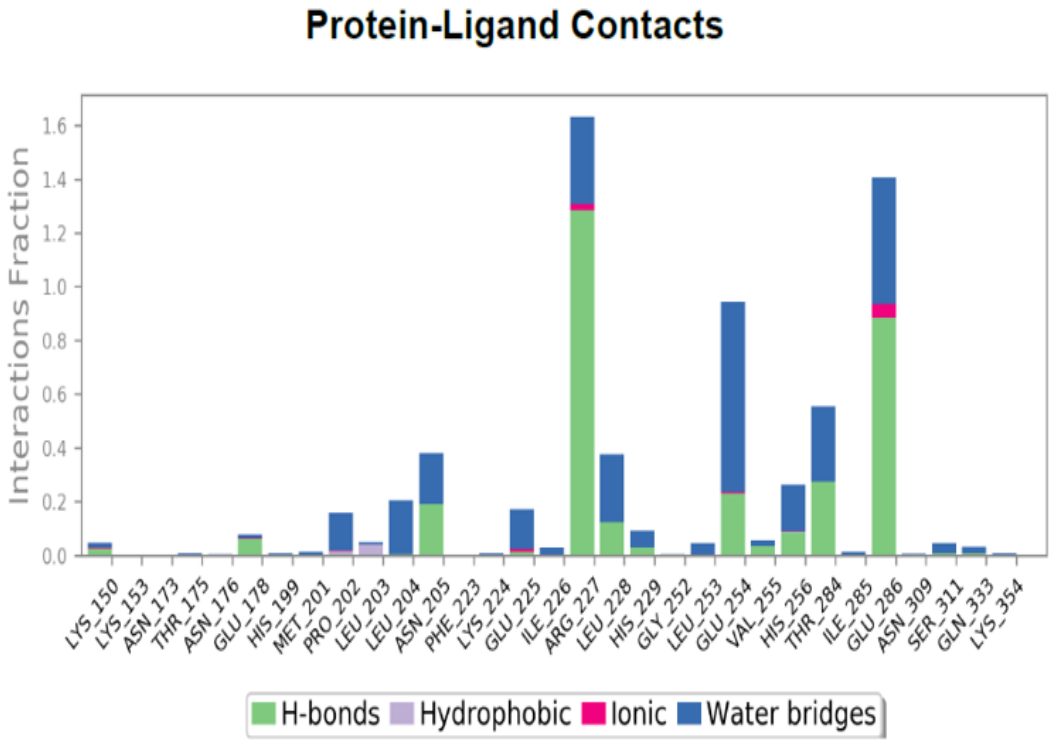


FIGURE 4.6: Protein-ligand contact types over 200 ns.

The contacts between DRAMP18681 and the TLR4/MD-2 receptor were classified and counted at every snapshot of the simulation. The water bridges and hydrophobic contacts were the most prominent interaction types present in the contact profile and stayed fairly similar over the course of the 200 ns trajectory.

Persistent minor contribution of hydrogen bonds and ionic contacts were maintained. The consistency of this interaction signature throughout the entire simulation (no sudden gain or loss of the contact categories) suggests that DRAMP18681 interacts with the MD-2 binding groove in a specific, reproducible non-covalent interaction network instead of through a pose-switching or transient interaction..

4.6.6 Contact Occupancy Key Interaction Residues

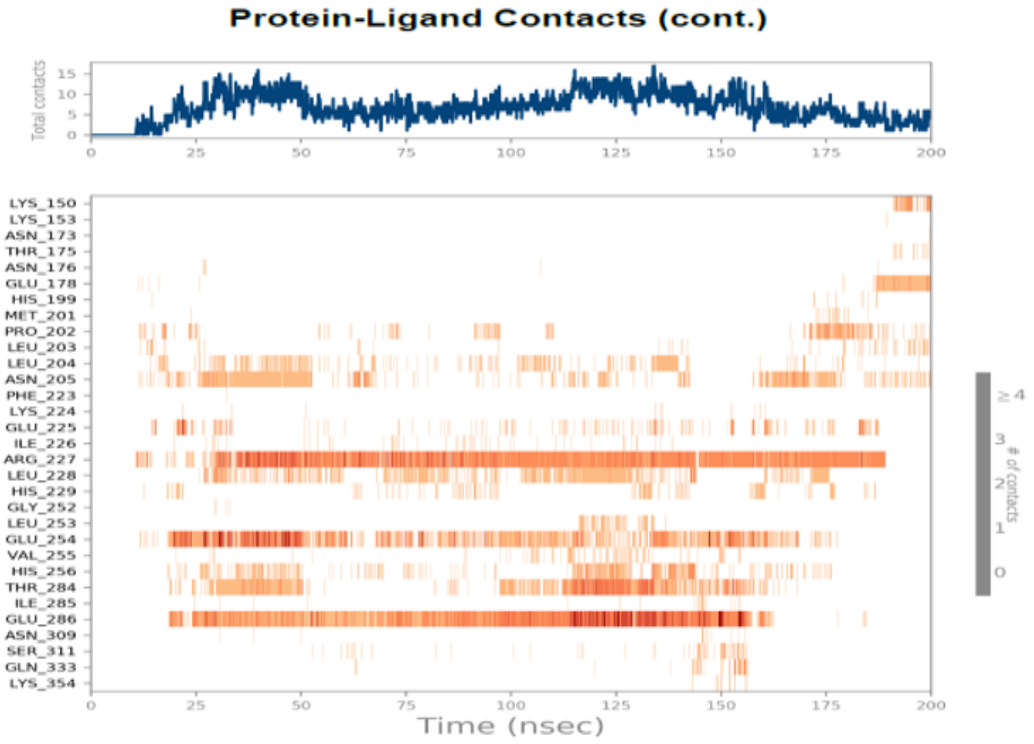


FIGURE 4.7: Residue contact occupancy at the DRAMP18681 binding interface.

The contact occupancy analysis yielded the residues of TLR4/MD-2 interface with a persistence greater than 15% from the simulation frames. A small and discrete

set of residues sustained contacts along the trajectory in a manner that is consistent with a binding mode to the MD-2 hydrophobic pocket. These residues are all high occupied, both polar and non-polar interaction types, and are the primary anchor sites of the DRAMP18681-receptor interaction, as determined by the AutoDock Vina docking analysis (Section 4.2.3).

4.7 Phase VII: Vaccine Antigen Interaction Prediction

The goal of antigen docking in Phase VII is different from that of receptor docking in Phase II. To be compatible for vaccine co-formulation, the effective adjuvant must weakly bind the vaccine antigen, allowing the binding of immunodominant epitopes by the MHC, which are necessary for antigen unfolding required for priming of CTLs; and avoiding steric hindrance of antibody-inducing B-cell epitopes.

Thus, the higher the score is, the more poorly the antigen will bind, and a score of less than -1000 is considered good and a score of -1400 or lower is considered ideal for possible use of the antigen as an adjuvant.Thus, a score the closer to 0 the better for antigen binding, and the closer to -1400 the better for possible use of the antigen as an adjuvant

4.7.1 Selection of Ovarian Cancer Vaccine Antigens

Three clinically validated ovarian cancer tumour-associated antigens were chosen as docking targets: NY-ESO-1 (PDB: 1S9W, 2.0 Å), MUC16 (PDB: 7SA9, 3.0 Å) and folate receptor alpha (PDB: 4LRH, 2.8 Å). Three criteria were used to select each antigen: it must be proven to be overexpressed in epithelial ovarian cancer, it must be under clinical investigation as a vaccine or antibody target, and it must have a high-resolution crystal structure. In ∼30-40% of ovarian cancers, NY-ESO-1 is over-expressed, and has a highly defined CTL epitope repertoire [116].

The elevated level of MUC16 (CA-125 antigen) in more than 90% of high-grade serous ovarian carcinoma causes peritoneal metastasis through interaction with mesothelin.

In about 90% of epithelial ovarian cancers, the expression of the folate receptor alpha is significantly increased and farletuzumab, a monoclonal antibody, is under development for targeted therapy.

4.7.2 Antigen Structure Preparation

Crystal structures of the 3 selected antigens were retrieved from RCSB Protein Data Bank and subsequently pre-processed for further docking of water molecules, before removing non-protein ligands, missing residues in the surface exposed loop regions were predicted using the ModLoop server, and each was finally visually checked in PyMOL 2.5 for structural integrity are shown in table 4.10:

TABLE 4.10: The pdb structures of all 3 model vaccine antigens.

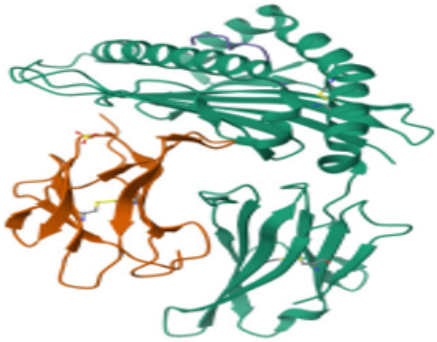
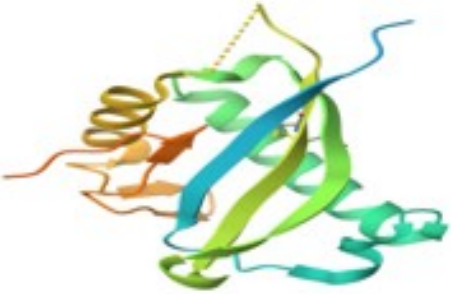
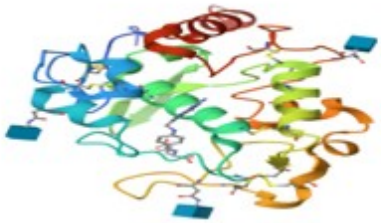
Vaccine Antigen	**Structures**
NY-ESO-1	
MUC16	

Table 4.10 continued from previous page

Vaccine Antigen	Structures
Folate Receptor Alpha	

4.7.3 Protein-Protein Docking via HDOCK

In total, there are 15 mutant-antigen combinations: the five DRAMP 18681 mutants and antigen structures for NY-ESO-1 (PDB: 1S9W), MUC16 (PDB: 7SA9), and folate receptor alpha (PDB: 4LRH) were prepared for protein-protein docking using the HDOCK server.

Tables 4.11. 4.12, 4.13 shows that all three antigens (NY-ES0-1, MUC16, and folate receptor alpha) were docked with all five mutants individually. Complete scores, mean values, and adjuvant compatibility assessments are presented in Table 4.14.

TABLE 4.11: Docking results of all 5 mutants with NY-ESO-1

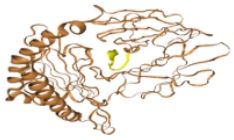
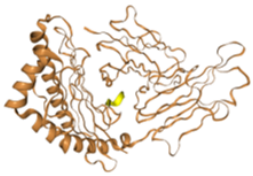
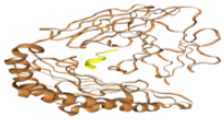
Mutant	Seq ID	Sequence	NY-ESO-1	Structure
Mutant 1	seq_42	DRGGVRDPR V	-193.58	
Mutant 2	seq_45	DRGGVRDPRI	-176.79	
Mutant 3	seq_47	DRGGVIDPRI	-145.28	

Table 4.11 continued from previous page

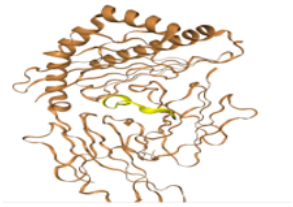
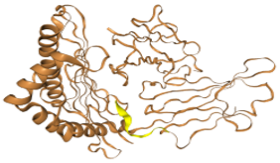
Mutant	Seq ID	Sequence	NY-ESO-1	Structure
Mutant 4	seq_56	DKGGVIDPR V	-146.56	
Mutant 5	seq_57	DKGGVRDPR V	-146.99	

TABLE 4.12: Docking results of all 5 mutants with MUC16

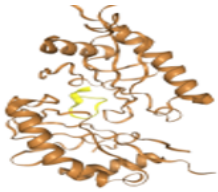
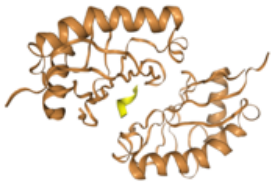
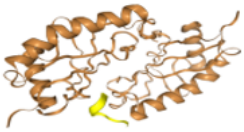
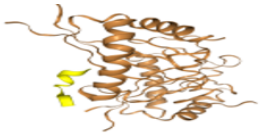
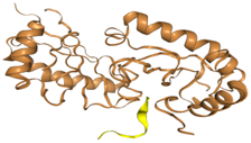
Mutant	Seq ID	Sequence	MUC16	Structure
Mutant 1	seq_42	DRGGVRDPR V	-158.89	
Mutant 2	seq_45	DRGGVRDPRI	-166.63	
Mutant 3	seq_47	DRGGVIDPRI	-138.69	
Mutant 4	seq_56	DKGGVIDPR V	-134.30	
Mutant 5	seq_57	DKGGVRDPR V	-127.68	

TABLE 4.13: Docking results of all 5 mutants with FRalpha

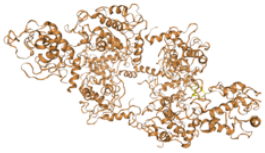
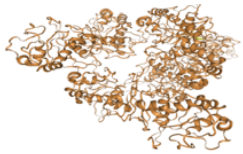
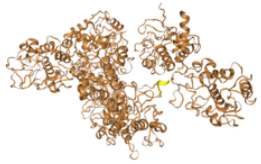
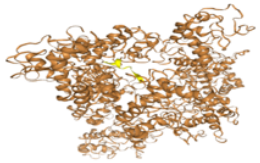
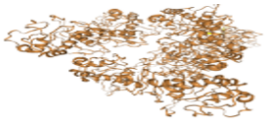
Mutant	Seq ID	Sequence	FRalpha	Structure
Mutant 1	seq_42	DRGGVRDPR V	-199.04	
Mutant 2	seq_45	DRGGVRDPRI	-167.73	
Mutant 3	seq_47	DRGGVIDPRI	-169.17	
Mutant 4	seq_56	DKGGVIDPR V	-191.88	
Mutant 5	seq_57	DKGGVRDPR V	-160.07	

TABLE 4.14: HDOCK scores for five DRAMP18681 mutants against three ovarian cancer antigens.

Mutant	Seq ID	Sequence	NY-ESO-1	MUC16	FRalpha	Mean Score	Adjuvant Compatibility
Mutant 1	seq_42	D R G G V RDPRV	-193.58	-158.89	-199.04	-165.23	MODERATE – verify NY-ESO-1 epitope overlap
Mutant 2	seq_45	D R G G V RDPRI	-176.79	-166.63	-167.73	-170.38	MODERATE – consistent cross-antigen binding

Table 4.14 continued from previous page

Mutant	Seq ID	Sequence	NY-ESO-1	MUC16	FRalpha	Mean Score	Adjuvant Compatibility
Mutant 3	seq_47	DRGGVIDPRI	-145.28	-138.69	-169.17	-151.05	FAVORABLE – weak NY-ESO-1 and MUC16 binding
Mutant 4	seq_56	DKGGVIDPRV	-146.56	-134.30	-191.88	-157.58	CAUTION FRalpha – strongest FRalpha interaction; avoid FRalpha-based vaccine
Mutant 5	seq_57	DKGGVRDPRV	-146.99	-127.68	-160.07	-144.91	BEST – weakest overall binding; recommended primary lead

4.7.4 Recommended Primary Lead Adjuvant

Mutant 5 (seq_57, DKGGVRDPRV) was identified as the primary recommended lead on the basis of all results with the weakest mean antigen interaction score for every antigen across the entire set (-144.91, the least negative of all the mutants), and the lowest score for MUC16 docking (-127.68), meaning the SEA domain has the least number of docked residues of any of the other four proteins. Mutant 5 has an instability index of -3.18, the lowest of all five mutants, with no predicted cleavage sites across the four target enzymes, and a classification of non-toxic, non-hemolytic, and non-allergenic. Mutant 3 (seq_47, DRGGVIDPRI) has the lowest interaction with NY-ESO-1 (-145.28) and might be better suited in NY-ESO-1-based vaccine formulations. Mutant 4 (seq_56, DKGGVIDPRV) has the highest score of the dataset for interaction with the folate receptor alpha (-191.88) and should not be used in folate-receptor alpha based formulations as it may cover the folate binding domain.

Chapter 5

Discussion

Retrieval of 117 scorpion-derived AMP sequences from APD3 and DRAMP across 31 genera and five families established a structurally and taxonomically diverse candidate AMPs. The library reflects the chemical diversity encoded across scorpion venom glands, which have been under strong evolutionary pressure to generate peptides capable of disrupting microbial membranes through amphipathic mechanisms [120]. That 86.3% of retrieved sequences carried a positive net charge at physiological pH is consistent with established physicochemical principles governing AMP-membrane interaction: electrostatic attraction between cationic peptides and negatively charged bacterial or tumour-cell membranes is considered a prerequisite for initial contact and subsequent insertion [121]. The mean GRAVY score of +0.448 indicates a moderate but consistent hydrophobic bias across the library, a property independently required for deep insertion into the apolar MD-2 co-receptor pocket of TLR4 [122]. Exclusion of sequences exceeding 30 amino acids, which correspond almost exclusively to disulfide-bridged peptides with constrained globular folds, was justified on two grounds: their folded backbones engage targets through mechanisms fundamentally distinct from non-disulfide-bridged peptides [123], and they are incompatible with the fixed-backbone sequence design methodology deployed in later pipeline stages. Restricting the working set to 9-29 aa therefore aligned the candidate pool with both the biological rationale and computational tool constraints of the downstream phases.

All 75 filtered AMPs returned thermodynamically favourable binding affinities against TLR4/MD-2 (PDB: 3FXI), ranging from -17.9 to -7.0 kcal/mol with a mean of -11.83 kcal/mol. The complete absence of positive deltaG values across all nine AutoDock Vina poses per peptide suggests that the physicochemical profile of the filtered set is inherently complementary to the MD-2 binding groove, a cavity characterised by a deep hydrophobic pocket flanked by basic residue clusters [122]. The enrichment of selected binders (34 of 75 candidates at deltaG ¡= -13.0 kcal/mol) aligns broadly with prior *In silico* screens of cationic host-defence peptides against TLR4, where short amphipathic sequences consistently achieve high predicted binding affinities via hydrophobic burial within the MD-2 lipid pocket [124].

The STRING v12.0 analysis placing ten TLR4 functional partners at maximum confidence (0.999) resolved the MyD88 and TRIF signaling cascade. This bifurcated signaling architecture is directly relevant to cancer vaccine adjuvant design. MyD88-dependent activation drives NF-kB, producing IL-12p70, TNF-alpha, and IL-6 required for dendritic cell maturation and Th1 polarisation [125]. The TRIF arm, through TBK1 and IRF3/IRF7, generates type I interferons independently required for MHC class I cross-presentation and cytotoxic CD8+ T-cell priming [126]. An adjuvant engaging both pathways simultaneously, as the NDBP docking geometry at MD-2 suggests, would be predicted to generate the combined Th1/CTL immune profile most associated with objective tumour regression in ovarian cancer models [127].

The classification of 50 of 75 docked AMPs (66.7%) as cytotoxic by ToxinPred 2.0 quantitatively demonstrates a structural restriction encountered across multiple AMP development practices: the same amphipathic, cationic geometry conferring TLR4/MD-2 affinity also predisposes peptides to non-selective membrane disruption [128]. Four selected docking candidates including AP02356 (-16.0 kcal/mol) and AP02127 (-15.7 kcal/mol) were predicted cytotoxic despite their outstanding receptor affinity, confirming that these two properties are correlated in the scorpion NDBP class. This observation has a practical design implication: optimisation for TLR4 binding alone is insufficient as an adjuvant selection criterion,

and safety pre-screening must be treated as an independent, non-negotiable filter stage. The additional hemolysis screen, eliminating 13 of 25 non-toxic candidates by HLPpred-Fuse, reduced the candidate set to 12 dual-safe AMPs at a 10.3

Selection of DRAMP18681 (LLGGLLQSLL, 10 aa, deltaG = -7.8 kcal/mol) as the preferred redesign scaffold, despite its weaker TLR4 docking relative to AP00153 and AP02604, was validated by the substantially superior stability and protease outputs of its ProteinMPNN derivatives. The AP00153-derived mutant set was eliminated by protease susceptibility screening, indicating that the 25 aa scaffold encodes sequence contexts enriched for tryptic and chymotryptic recognition regardless of redesigned residues. AP02604 yielded only one protease-stable mutant. DRAMP18681, by contrast, produced 12 of 20 protease-stable sequences and all five final lead mutants. This scaffold-dependence of redesign quality is consistent with observations in the protein engineering literature that fixed-backbone ProteinMPNN performance is strongly conditioned by the length and loop content of the template structure. The ten-residue DRAMP18681 backbone presents minimal surface area for protease engagement, and its redesigned outputs with DRGG and DKGG motifs replacing the destabilising LLQ core shifted the instability index from 97.88 in the parent to values between -3.18 and -2.23 in the five leads. The introduction of cationic residues (Lys, Arg) simultaneously preserved the electrostatic character required for TLR4 interaction while eliminating the sequence-level drivers of conformational lability identified by Guruprasad et al. [129].

Backbone RMSD converging to a 2-3 Angstrom after approximately 25 ns of equilibration, and maintaining this range for the remaining 175 ns, provides strong computational evidence that the AutoDock Vina predicted binding pose of the DRAMP18681 primary lead is thermodynamically stable under explicit-solvent, physiological-temperature (300 K) conditions. RMSD values in this range are characteristic of crystallographically validated peptide-protein complexes simulated under NPT ensemble conditions and are different from the progressive RMSD drift that characterizes non-specific surface association or pose-switching events [130]. The confinement of RMSF peaks above 4 Angstroms exclusively to distal

loop regions and chain termini, with core interface residues fluctuating below 2 Angstroms, identifies the MD-2 binding groove as structurally rigid on the simulation timescale. This rigidity is mechanistically consistent with receptor activation rather than merely binding, since conformational rigidity of the TLR4/MD-2 receptor interface upon agonist binding is a prerequisite for homo-dimer formation and downstream signal transduction [131]. The stable radius of gyration throughout the 200 ns trajectory rules out domain separation or unfolding, and the predominance of water-bridge contacts at the interface is consistent with agonist-bound TLR4/MD-2 co-crystal structures in which ordered solvent molecules bridge peptide backbone atoms to receptor residues at the MD-2 rim [122].

Mutant 5 (seq_57, DKGGVRDPRV) emerged as the primary recommended lead with a mean cross-antigen score of -144.91, the least negative across all five candidates and across all three antigens. Its weakest individual interaction against MUC16 (-127.68 versus -158.89 to -166.63 for other mutants) is clinically significant: the CA-125 SEA domain of MUC16 encodes the immunodominant antibody-generating epitopes exploited in both passive immunotherapy and active vaccine strategies for ovarian cancer [132]. Minimal adjuvant-MUC16 overlap therefore reduces the probability of competitive epitope masking in an MUC16-targeted formulation. Mutant 3 (seq_47, DRGGVIDPRI) offers marginal advantage for NY-ESO-1-specific formulations (NY-ESO-1 score -145.28 versus -146.99 for Mutant 5), while Mutant 4 (seq_56, DKGGVIDPRV) is contraindicated for folate receptor alpha-based vaccines given its strongest FRalpha interaction in the dataset (-191.88), which would block the folate-binding domain required for recognition by farletuzumab or CAR-T constructs targeting the same epitope region .

Chapter 6

Conclusion and Future Recommendations

6.1 Conclusion

This study aimed to assess the potential for scorpion venom NDBP chemical space to encode peptides that simultaneously bind to TLR4/MD-2 with affinity that is relevant to adjuvant function, are safe for cellular use, can be proteolytically stable under serum-relevant enzyme conditions, and do not sterically interfere with clinically relevant ovarian cancer vaccine antigens. Within the limits of computational prediction, this combination of properties is possible and can be rationally designed from a naturally occurring source through fixed-backbone sequence redesign, as shown by the seven-phase *In silico* pipeline developed here. This study was conducted under three specific objectives. The first goal was to retrieve from APD3 and DRAMP and published literature 117 non-redundant sequences of scorpion antimicrobial peptides from 31 genera, belonging to five families, with 312 biological activity annotations, which was advanced to molecular docking after length-based and quality filtering.

The second objective, the analysis of physicochemical properties and immunomodulatory potential of the selected peptides, was achieved by profiling all 117 peptides using ProtParam. The secondary structure prediction of the peptides by AlphaFold, and blind molecular docking of these peptides against TLR4/MD-2 (PDB: 3FXI), followed by protein interaction network analysis with STRING v12.0, which showed engagement of both the MyD88 and TRIF signalling arms and confirmed dual-pathway immunostimulatory properties consistent with Th1 polarisation and $CD8^+$ cytotoxic T-lymphocyte priming.

To evaluate peptide safety and toxicity, and to prioritise peptide-based vaccine adjuvant candidates for ovarian cancer, the third objective was accomplished by applying a four-parameter *In silico* safety screen (cytotoxicity, hemolytic activity, allergenicity, and conformational stability). This ranked the three lead scaffolds, followed by 200 ns molecular dynamics simulation of the primary lead complex and finally by HDOCK protein-protein docking against three clinically relevant ovarian cancer antigens (NY-ESO-1, MUC16, and folate receptor alpha), which identified seq_57 (DKGGVRDPRV) as the recommended primary adjuvant candidate for experimental validation.

The 117-entry primary library was physiochemically and structurally profiled, and scorpion NDBPs were found to be inherently biased toward being cationic and amphipathic, a priori justifying the study's biological hypothesis. In the molecular docking phase this architectural bias was directly translated to the measurable computational binding affinity: all the 75 molecules passed the length filter, and 34 of them were found to be binding molecules of the level of the best reported short peptides for this receptor class.

The pipeline was the most rigorous of all the safety screening steps. The results of this screening are not a failure, but a useful one the receptor affinity is statistically independent from its cytotoxic selectivity; high affinity to TLR4 is not enough and parallel toxicity filtering is essential, with important direct implications for natural AMPs adjuvant programmes. For the five final leads with properties not achievable in the native sequence (LLGGLLQSLL, scaffold DRAMP18681 derived from Chaerilus tryznai), the properties were: negative instability indices (-3.18 to

-2.23), complete resistance to all four proteases tested, non-toxic, non-hemolytic classification maintained for the entire redesigned set.

The 200 ns molecular dynamics simulation confirmed the binding pose stability at the interface of TLR4/MD-2, a convergence of the backbone atoms, low number of nonpolar contacts, and the consistent water-bridge-dominated network of contacts, without any transient or non-specific interactions.

The antigen docking screen inverted the standard affinity maximisation logic and identified Mutant 5 (seq_57, DKGGVRDPRV) as the candidate with the broadest co-formulation compatibility, across NY-ESO-1, MUC16 and folate receptor alpha.

These results allow us to draw three direct translationally relevant conclusions. First, the physicochemical characteristics of the scorpion NDBP can be used as filters for the selection of decapeptides, which, after re-designing by computational tools, can have the safety and stability profile necessary for the development of clinical adjuvants. Second, the pipeline architecture, in its entirety (7 sequential independently validated phases from sequence retrieval to antigen compatibility assessment), is a model that can be applied to other natural-peptide libraries and to other tumour antigen targets without conceptual change. Third, seq_57 (DKGGVRDPRV) is a specific, structurally characterised, and computationally multi-validated candidate that should be prioritised for experimental study, not one of several candidates that appear generally promising from screening alone.

6.2 Future Recommendations

The computational results are given as a starting point for experimental investigations. The recommendations outlined below represent a logical, structured pathway for experimentation in a hierarchy of technical accessibility and logical dependency, starting with the most accessible experiments and moving to preclinical model testing.

The chemical synthesis of all five final mutant sequences, seq_42, seq_45, seq_47, seq_56 and seq_57, should be carried out (these selected AMPs do not contain any disulfide bonds, and can therefore be prepared by standard Fmoc solid-phase peptide synthesis (SPPS) without the additional cysteine protection chemistry that is necessary for longer scorpion peptides).

i. Toxicity predictions in terms of safety must be experimentally tested before any biological activity assay TLR4.

ii. Confirmation of the activation and signaling cascades by using a TLR4 agonist should be performed in reporter cells that express human TLR4/MD-2/CD14 and secrete alkaline phosphatase following NF-κB activation; this allows the determination of the EC50.

iii. Incorporating additional AMPs from other venomous invertebrates (cone snails, scorpion species not already present in APD3 OR DRAMP) as well as spider-venom peptides would add even more chemicals to the entry point.

Overall, binding affinity estimates obtained by blind docking against the multiple conformers of the TLR4/MD-2 receptor would be more accurate than those obtained by AutoDock Vina empirical scoring, and could be further boosted when all-atom free-energy perturbation (FEP) calculations are performed for top-ranked candidates.

Bibliography

[1] F. Bray, M. Laversanne, H. Sung, J. Ferlay, R. L. Siegel, I. Soerjomataram, and A. Jemal, "Global cancer statistics 2022: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 74, no. 3, pp. 229–263, 2024.

[2] G. Caruso, S. J. Weroha, and W. Cliby, "Ovarian cancer: a review," *JAMA*, vol. 334, no. 14, pp. 1278–1291, 2025.

[3] R. Lheureux, M. Gourley, and A. Bowtell, "Epithelial ovarian cancer," *The Lancet*, vol. 393, pp. 1240–1253, 2019.

[4] L. Wang *et al.*, "Therapeutic landscape of ovarian cancer: recent advances and emerging therapies," *Biomarker Research*, vol. 13, no. 103, 2025.

[5] I. J. Jacobs *et al.*, "Ovarian cancer screening and mortality in the uk collaborative trial of ovarian cancer screening (ukctocs): a randomised controlled trial," *The Lancet*, vol. 387, no. 10022, pp. 945–956, 2016.

[6] E. Ghisoni, M. Morotti, A. Sarivalasis, L. Kandalaft, and G. Coukos, "Immunotherapy for ovarian cancer: towards a tailored immunophenotype-based approach," *Nature Reviews Clinical Oncology*, vol. 21, pp. 801–817, 2024.

[7] J. Chen *et al.*, "Recent advances in understanding the immune microenvironment in ovarian cancer," *Frontiers in Immunology*, vol. 15, p. 1412328, 2024.

[8] R. E. W. Hancock, E. F. Haney, and E. E. Gill, "The immunology of host defence peptides: beyond antimicrobial activity," *Nature Reviews Immunology*, vol. 16, no. 5, pp. 321–334, 2016.

[9] E. F. Haney, S. K. Straus, and R. E. W. Hancock, "Reassessing the host defense peptide landscape: mechanisms, antimicrobial and immunomodulatory activities," *Nature Reviews Microbiology*, vol. 17, pp. 389–400, 2019.

[10] A. Almaaytah and Q. Albalas, "Scorpion venom peptides with no disulfide bridges: a review," *Peptides*, vol. 51, pp. 35–45, 2014.

[11] Z. Xia *et al.*, "Antimicrobial potential of scorpion-venom-derived peptides," *Molecules*, vol. 29, no. 21, p. 5080, 2024.

[12] J. Ghadiri *et al.*, "Therapeutic potential of scorpion venom proteins: innovations in cancer treatment and future directions," *Current Medicinal Chemistry*, vol. 27, no. 1, pp. 33–58, 2026.

[13] P. Ray Choudhury *et al.*, "In silico peptide design: methods, resources, and role of ai," *Journal of Peptide Science*, vol. 31, no. 12, p. e70063, 2025.

[14] A. Naeem *et al.*, "Computational design of anti-cancer peptides tailored to target tumor-specific markers," *BMC Chemistry*, vol. 18, p. 39, 2024.

[15] M. Drayton, J. P. Deisinger, K. C. Ludwig, N. Raheem, A. Müller, T. Schneider, and S. Etayash, "Host defense peptides: dual antimicrobial and immunomodulatory action," *International Journal of Molecular Sciences*, vol. 22, no. 20, p. 11172, 2021.

[16] G. F. Hatfull, R. M. Dedrick, and R. T. Schooley, "Phage therapy for antibiotic-resistant bacterial infections," *Annual Review of Medicine*, vol. 73, pp. 197–211, 2022.

[17] F. Zou, C. Tan, T. S. Shinali, B. Zhang, L. Zhang, Z. Han, and N. Shang, "Plant antimicrobial peptides: a comprehensive review of their classification, production, mode of action, functions, applications, and challenges," *Food & Function*, vol. 14, pp. 5492–5515, 2023.

[18] V. A. Fischetti, "Bacteriophage lysins as effective antibacterials," *Current Opinion in Microbiology*, vol. 11, no. 4, pp. 393–400, 2008.

[19] D. C. Nelson, M. Schmelcher, L. Rodriguez-Rubio, J. Klumpp, D. G. Pritchard, S. Dong, and D. M. Donovan, "Endolysins as antimicrobials," *Advances in Virus Research*, vol. 83, pp. 299–365, 2012.

[20] D. Scholl, "Phage tail-like bacteriocins," *Annual Review of Virology*, vol. 4, pp. 453–467, 2017.

[21] P. D. Cotter, R. P. Ross, and C. Hill, "Bacteriocins – a viable alternative to antibiotics?" *Nature Reviews Microbiology*, vol. 21, no. 6, pp. 381–395, 2023.

[22] S. C. Yang, C. H. Lin, C. T. Sung, and J. Y. Fang, "Antibacterial activities of bacteriocins: application in foods and pharmaceuticals," *Frontiers in Microbiology*, vol. 14, p. 1158622, 2023.

[23] D. Destoumieux-Garzón, J. Peduzzi, and S. Rebuffat, "Focus on modified microcins: structural features and mechanisms of action," *Biochimie*, vol. 84, no. 5-6, pp. 511–519, 2002.

[24] O. Gillor, B. C. Kirkup, and M. A. Riley, "Colicins and microcins: the next generation antimicrobials," *Advances in Applied Microbiology*, vol. 54, pp. 129–146, 2004.

[25] J. F. S. Daniel and E. R. Filho, "Peptaibols of trichoderma," *Natural Product Reports*, vol. 24, no. 5, pp. 1128–1141, 2007.

[26] Y.-P. Chen, Y. Li, F. Chen, H. Wu, and S. Zhang, "Characterization and expression of fungal defensin in escherichia coli and its antifungal mechanism by rna-seq analysis," *Frontiers in Microbiology*, vol. 14, p. 1172257, 2023.

[27] O. O. Bakare, A. Gokul, A. O. Fadaka, R. Wu, L. A. Niekerk, A. M. Barker, M. Keyster, and A. Klein, "Plant antimicrobial peptides (pamps): features, applications, production, expression, and challenges," *Molecules*, vol. 27, no. 12, p. 3703, 2022.

[28] M. P. Slezina and T. I. Odintsova, "Plant antimicrobial peptides: insights into structure-function relationships for practical applications," *Current Issues in Molecular Biology*, vol. 45, no. 4, pp. 3674–3704, 2023.

[29] T. I. Odintsova, L. N. Shcherbakova, M. P. Slezina, T. D. Pasechnik, B. B. Kartabaeva, E. A. Istomina, and V. G. Dzhavakhiya, "Hevein-like antimicrobial peptides wamps: structure–function relationship in antifungal activity and sensitization of plant pathogenic fungi to tebuconazole by wamp-2-derived peptides," *International Journal of Molecular Sciences*, vol. 21, no. 21, p. 7912, 2020.

[30] K. A. Selsted and A. J. Ouellette, "Mammalian defensins in the antimicrobial immune response," *Nature Reviews Immunology*, vol. 23, no. 8, pp. 512–528, 2023.

[31] S. J. de Veer, M.-W. Kan, and D. J. Craik, "Cyclotides: from structure to function," *Chemical Reviews*, vol. 119, no. 24, pp. 12 375–12 421, 2019.

[32] N. L. Daly and D. J. Craik, "Bioactive cystine knot proteins," *Current Opinion in Chemical Biology*, vol. 15, no. 3, pp. 362–368, 2011.

[33] R. J. Lewis and M. L. Garcia, "Therapeutic potential of venom peptides," *Nature Reviews Drug Discovery*, vol. 2, no. 10, pp. 790–802, 2003.

[34] L. D. Possani, B. Becerril, M. Delepierre, and J. Tytgat, "Scorpion toxins specific for na+-channels," *European Journal of Biochemistry*, vol. 264, no. 2, pp. 287–300, 1999.

[35] E. Ortiz, G. B. Gurrola, E. F. Schwartz, and L. D. Possani, "Scorpion venom components as potential candidates for drug development," *Toxicon*, vol. 93, pp. 125–135, 2015.

[36] R. C. Rodríguez de la Vega, E. F. Schwartz, and L. D. Possani, "Mining on scorpion venom biodiversity," *Toxicon*, vol. 56, no. 7, pp. 1155–1161, 2010.

[37] L. Moerman, S. Bosteels, W. Noppe, J. Willems, E. Clynen, L. Schoofs, K. Thevissen, J. Tytgat, J. Van Eldere, J. van der Walt, and F. Verdonck,

"Antibacterial and antifungal properties of alpha-helical, cationic peptides in the venom of scorpions from southern africa," *European Journal of Biochemistry*, vol. 269, no. 19, pp. 4799–4810, 2002.

[38] L. D. Possani, B. Becerril, M. Delepierre, and J. Tytgat, "Scorpion toxins specific for k+ channels," *European Journal of Biochemistry*, vol. 264, no. 2, pp. 287–300, 1999.

[39] G. Titaux-Delgado *et al.*, "Beta-ktx14.3, a scorpion toxin, blocks the human potassium channel kcnq1," *Biochimica et Biophysica Acta – Proteins and Proteomics*, vol. 1871, no. 1, p. 140878, 2023.

[40] T. Boltman, M. Meyer, and O. Ekpo, "Diagnostic and therapeutic approaches for glioblastoma using chlorotoxin nanoparticles," *Cancers*, vol. 15, no. 13, p. 3388, 2023.

[41] B. Amorim-Carmo *et al.*, "Antimicrobial peptide analogs from scorpions: modifications and structure-activity," *Frontiers in Molecular Biosciences*, vol. 9, 2022.

[42] J. I. Cid-Uribe, E. P. Meneses, C. V. F. Batista, E. Ortiz, and L. D. Possani, "Dissecting toxicity: the venom gland transcriptome and the venom proteome of the highly venomous scorpion centruroides limpidus (karsch, 1879)," *Toxins*, vol. 11, no. 5, p. 247, 2019.

[43] C. Cesa-Luna *et al.*, "Structural characterization of scorpion peptides and their bactericidal activity against clinical isolates of multidrug-resistant bacteria," *PLOS ONE*, vol. 14, no. 9, p. e0222438, 2019.

[44] J. M. Sciani, L. Marques-Porto, T. M. Fucase, O. Cnudde, E. Broussy, and D. C. Pimenta, "The modular nature of bradykinin-potentiating peptides isolated from snake venoms," *Journal of Venomous Animals and Toxins including Tropical Diseases*, vol. 23, p. 45, 2017.

[45] S. S. S. Arcos *et al.*, "Scorpion venom as a source of cancer drugs: a comprehensive proteomic analysis," *International Journal of Molecular Sciences*, vol. 26, no. 20, p. 9907, 2025.

[46] F. I. V. Villa-Hernández *et al.*, "Ion-channel-targeting scorpion recombinant toxin as novel therapeutic agent for breast cancer," *Toxins*, vol. 17, no. 4, p. 166, 2025.

[47] B. Hmed, S. H. Turky, and M. Z. Khaled, "Scorpion peptides: potential use for new drug development," *Journal of Toxicology*, vol. 2013, pp. 1–15, 2013.

[48] R. Guo, J. Liu, J. Chai, Y. Gao, M. A. Abdel-Rahman, and X. Xu, "Scorpion peptide smp24 exhibits a potent antitumor effect on human lung cancer cells by damaging the membrane and cytoskeleton in vivo and in vitro," *Toxins*, vol. 14, no. 7, p. 438, 2022.

[49] A. Kaur, J. Baldwin, D. Brar, D. B. Salunke, and N. Petrovsky, "Toll-like receptor (tlr) agonists as a driving force behind next-generation vaccine adjuvants and cancer therapeutics," *Current Opinion in Chemical Biology*, vol. 70, p. 102172, 2022.

[50] S. Mirzaei *et al.*, "Venom peptides in cancer therapy: an updated review on cellular and molecular aspects," *Pharmacological Research*, vol. 164, p. 105327, 2021.

[51] Z. Xia, D. He, Y. Wu, H. F. Kwok, and Z. Cao, "Scorpion venom peptides: molecular diversity, structural characteristics, and therapeutic use from channelopathies to viral infections and cancers," *Pharmacological Research*, vol. 197, p. 106978, 2023.

[52] J. Zargan, S. Umar, M. Sajad, M. Naime, S. Ali, and H. A. Khan, "Scorpion venom (odontobuthus doriae) induces apoptosis by depolarization of mitochondria and reduces s-phase population in human breast cancer cells (mcf-7)," *Toxicology in Vitro*, vol. 25, no. 8, pp. 1748–1756, 2011.

[53] A. Díaz-García, R. Noriega-Navarro, O. Martínez-Mota, J. C. Gómez-Villanueva, and D. Varela, "Scorpion toxins and ion channels: potential applications in cancer therapy," *Toxins*, vol. 12, no. 5, p. 326, 2020.

[54] N. Colombo *et al.*, "Newly diagnosed and relapsed epithelial ovarian cancer: Esmo clinical practice guideline for diagnosis, treatment and follow-up," *Annals of Oncology*, vol. 34, no. 10, pp. 833–848, 2023.

[55] D. Aviles, D. Warshal, M. Buchbinder, and O. Ostrovsky, *Influence of aberrant epigenetic changes and the tumor microenvironment in ovarian cancer metastasis*. Brisbane: Exon Publications, 2022.

[56] S. Lheureux, M. Braunstein, and A. M. Oza, "Epithelial ovarian cancer: evolution of management in the era of precision medicine," *CA: A Cancer Journal for Clinicians*, vol. 69, no. 4, pp. 280–304, 2019.

[57] K. Shield, M. L. Ackland, N. Ahmed, and G. E. Rice, "Multicellular spheroids in ovarian cancer metastases: biology and pathology," *Gynecologic Oncology*, vol. 113, no. 1, pp. 143–148, 2009.

[58] R. C. J. Bast, B. Hennessy, and G. B. Mills, "The biology of ovarian cancer: new opportunities for translation," *Nature Reviews Cancer*, vol. 9, no. 6, pp. 415–428, 2009.

[59] F. Blanc-Durand, L. C. W. Xian, and D. S. P. Tan, "Targeting the immune microenvironment for ovarian cancer therapy," *Frontiers in Immunology*, vol. 14, p. 1328651, 2023.

[60] S. I. Labidi-Galy *et al.*, "High grade serous ovarian carcinomas originate in the fallopian tube," *Nature Communications*, vol. 8, p. 1093, 2017.

[61] S. S. Savant, S. Sriramkumar, and H. M. O'Hagan, "The role of inflammation and inflammatory mediators in the development, progression, metastasis, and chemoresistance of epithelial ovarian cancer," *Cancers*, vol. 10, no. 8, p. 251, 2018.

[62] J. Prat and F. C. on Gynecologic Oncology, "Staging classification for cancer of the ovary, fallopian tube, and peritoneum," *International Journal of Gynecology & Obstetrics*, vol. 124, no. 1, pp. 1–5, 2014.

[63] S. Lheureux, C. Gourley, I. Vergote, and A. M. Oza, "Epithelial ovarian cancer," *The Lancet*, vol. 393, no. 10177, pp. 1240–1253, 2019.

[64] C. Kandoth *et al.*, "Mutational landscape and significance across 12 major cancer types," *Nature*, vol. 502, no. 7471, pp. 333–339, 2013.

[65] C. G. A. R. Network, "Integrated genomic analyses of ovarian carcinoma," *Nature*, vol. 474, no. 7353, pp. 609–615, 2011.

[66] A. du Bois, A. Reuss, E. Pujade-Lauraine, P. Harter, I. Ray-Coquard, and J. Pfisterer, "Role of surgical outcome as prognostic factor in advanced epithelial ovarian cancer: a combined exploratory analysis of 3 prospectively randomized phase 3 multicenter trials," *Cancer*, vol. 115, no. 6, pp. 1234–1244, 2009.

[67] J. A. Ledermann, P. Harter, C. Gourley, M. Friedlander, I. Vergote, G. Rustin, C. L. Scott, W. Meier, R. Shapira-Frommer, T. Safra, D. Matei, A. Fielding, S. Spencer, P. Rowe, E. Lowe, D. Hodgson, M. A. Sovak, and U. Matulonis, "Overall survival in patients with platinum-sensitive recurrent serous ovarian cancer receiving olaparib maintenance monotherapy: an updated analysis from a randomised, placebo-controlled, double-blind, phase 2 trial," *The Lancet Oncology*, vol. 17, no. 11, pp. 1579–1589, 2016.

[68] U. A. Matulonis, A. K. Sood, L. Fallowfield, B. E. Howitt, J. Sehouli, and B. Y. Karlan, "Ovarian cancer," *Nature Reviews Disease Primers*, vol. 2, p. 16061, 2016.

[69] N. Colombo, P. F. Conte, S. Pignata, F. Raspagliesi, and G. Scambia, "Bevacizumab in ovarian cancer: focus on clinical data and future perspectives," *Critical Reviews in Oncology/Hematology*, vol. 97, pp. 335–348, 2016.

[70] K. Odunsi, "Immunotherapy in ovarian cancer," *Annals of Oncology*, vol. 28, no. S8, pp. viii1–viii7, 2017.

[71] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: Globocan estimates of incidence

and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, 2021.

[72] F. Badar *et al.*, "Cancer statistics from the shaukat khanum memorial trust's hospital-based cancer registry, pakistan, 1994–2022: an observational study," *Journal of Cancer and Allied Specialties*, vol. 10, no. 2, p. 615, 2024.

[73] Y. Bhurgri, Y. Shaheen, N. Kayani, K. Nazir, R. Ahmed, A. Usman, I. Bashir, F. Setna, A. Bhurgri, S. H. Hasan, and S. M. H. Zaidi, "Incidence, trends and morphology of ovarian cancer in karachi (1995–2002)," *Asian Pacific Journal of Cancer Prevention*, vol. 12, no. 6, pp. 1567–1571, 2011.

[74] M. Tariq, S. Ahmad, M. Riaz, S. Ahmad, N. Bibi, and F. Choudhary, "Comparison of clinico-pathological characteristics and survival of recurrent ovarian cancer patients on seven different chemo-protocols," *Asian Pacific Journal of Cancer Prevention*, vol. 20, no. 6, pp. 1685–1691, 2019.

[75] O. Mulisya, F. K. Sikakulya, M. Mastaki, T. Gertrude, and M. Jeff, "The challenges of managing ovarian cancer in the developing world," *Obstetrics and Gynecology International*, vol. 2020, p. 8379628, 2020.

[76] S. Pervez, A. A. Jabbar, G. Haider, M. A. Qureshi, S. Ashraf, F. Lateef, M. Khurshid, I. Bashir, and M. Zaidi, "Karachi cancer registry (kcr): consolidated data of 5-years 2017–2021," *Journal of the College of Physicians and Surgeons Pakistan*, vol. 33, no. 5, pp. 560–565, 2023.

[77] F. Badar, S. Mahmood, M. A. Yusuf, and F. Sultan, "Epidemiology of cancers in lahore, pakistan, 2010–2012: a cross-sectional study," *BMJ Open*, vol. 6, no. 10, p. e011828, 2016.

[78] A. Ikram, S. Pervez, M. Khadim, A. Awan, Z. Majid, A. Yusuf, and S. N. Zafar, "National cancer registry of pakistan: first comprehensive report of cancer statistics 2015–2019," *Journal of the College of Physicians and Surgeons Pakistan*, vol. 33, no. 6, pp. 625–632, 2023.

[79] Y. Bhurgri, A. Bhurgri, S. H. Hassan, S. M. H. Zaidi, S. Rahim, A. Bhurgri, R. Pinjani, and N. Usman, "Cancer incidence in karachi, pakistan: first results from karachi cancer registry," *International Journal of Cancer*, vol. 85, no. 3, pp. 325–329, 2000.

[80] I. A. Malik, "A prospective study of clinico-pathological features of epithelial ovarian cancer in pakistan," *Journal of the Pakistan Medical Association*, vol. 52, no. 4, pp. 155–158, 2002.

[81] A. Bajwa *et al.*, "Challenges and opportunities in ovarian cancer care: a qualitative study of clinician perspectives from 24 low- and middle-income countries," *Journal of Cancer Policy*, vol. 44, p. 100582, 2025.

[82] C. Registry, S. K. M. C. H. Clinical Data Management, and R. Centre, "Annual cancer registry report 2022," SKMCH&RC, Lahore, Pakistan, 2023.

[83] M. Renz, M. Friedlander, and J. S. Berek, "Cancer of the ovary, fallopian tube, and peritoneum: 2025 update," *International Journal of Gynecology & Obstetrics*, vol. 171, no. S1, pp. 6–35, 2025.

[84] J. F. Fahrmann, S. M. Ghasemi, C. Y. Han, R. Wu, J. B. Dennison, J. Vykoukal, J. Celestino, K. Lu, Z. Lu, C. Drescher, K. A. Do, S. Hanash, R. C. Bast, and E. Irajizad, "A metabolite-based liquid biopsy for detection of ovarian cancer," *Biomarker Research*, vol. 12, no. 1, p. 91, 2024.

[85] W. H. Yoon and A. DeFazio, "Immune checkpoint inhibitors in ovarian cancer: where do we go from here?" *Cancer Drug Resistance*, vol. 6, no. 2, pp. 358–377, 2023.

[86] A. González-Martín *et al.*, "Niraparib in patients with newly diagnosed advanced ovarian cancer," *New England Journal of Medicine*, vol. 381, no. 25, pp. 2391–2402, 2019.

[87] T. Boltman, M. Meyer, and O. Ekpo, "Diagnostic and therapeutic approaches for glioblastoma and neuroblastoma cancers using chlorotoxin nanoparticles," *Cancers*, vol. 15, no. 13, p. 3388, 2023.

[88] M. Li, J. Hu, C. Wu, and X. Wang, "Peptides with therapeutic potential from the venom of buthus martensii scorpion," *Peptides*, vol. 115, pp. 43–50, 2019.

[89] R. C. S. Dias, C. Teixeira, F. Borges, A. S. Ferreira, M. V. Beraldo-Neto, R. Cordeiro-do Valle, C. J. Bloch, and J. R. Leite, "Voltage-gated sodium channels: from roles and mechanisms in the metastatic cell behavior to clinical potential as therapeutic targets," *Frontiers in Pharmacology*, vol. 14, p. 1206136, 2023.

[90] B. Amorim-Carmo, A. M. S. Parente, E. S. Souza, A. A. Silva-Junior, R. M. Araújo, and M. F. Fernandes-Pedrosa, "Antimicrobial peptide analogs from scorpions: modifications and structure-activity," *Frontiers in Molecular Biosciences*, vol. 9, p. 887763, 2022.

[91] M. A. Abdel-Rahman, V. Quintero-Hernandez, and L. D. Possani, "Venom proteomic and venomous glands transcriptomic analysis of the egyptian scorpion scorpio maurus palmatus (arachnida: Scorpionidae)," *Toxicon*, vol. 74, pp. 193–207, 2013.

[92] V. Asher, H. Sowter, R. Shaw, A. Bali, and R. Khan, "Eag and herg potassium channels as novel therapeutic targets in cancer," *World Journal of Surgical Oncology*, vol. 8, p. 113, 2010.

[93] B. A. Bhat, I. Saifi, N. A. Khamjan, S. S. Hamdani, A. Algaissi, S. Rashid, M. M. Alshehri, S. A. Ganie, M. Lohani, S. I. Abdelwahab, and S. A. Dar, "Exploring the tumor immune microenvironment in ovarian cancer: a way-out to the therapeutic roadmap," *Expert Opinion on Therapeutic Targets*, vol. 27, no. 9, pp. 841–860, 2023.

[94] Y. Zhang, Q. Xu, T. Shi, W. Ding, W. Li, X. Wang, and Z. Dong, "Scorpion venom polypeptide governs alveolar macrophage m1/m2 polarization to inhibit the progression of lung cancer," *International Immunopharmacology*, vol. 131, p. 111774, 2024.

[95] S. Mirzaei, H. S. Fekri, F. Hashemi, K. Hushmandi, R. Mohammadinejad, M. Ashrafizadeh, A. Zarrabi, and M. Garg, "Venom peptides in cancer therapy: an updated review on cellular and molecular aspects," *Pharmacological Research*, vol. 164, p. 105327, 2021.

[96] Rekdal., A. Haug, B. E. Matthiessen, T. J. Waagene, O. Nedberg, and J. S. Svendsen, "Anticancer effect obtained against mca205 sarcoma following treatment with the oncolytic peptide ltx-315 in combination with immune checkpoint inhibitors," *Journal for ImmunoTherapy of Cancer*, vol. 3, no. S2, p. P156, 2015.

[97] F. Nasiri, K. Farrokhi, P. S. Kozani, M. M. Kancha, S. D. Shokoohi, and P. S. Kozani, "Car-t cell immunotherapy for ovarian cancer: hushing the silent killer," *Frontiers in Immunology*, vol. 14, p. 1302307, 2023.

[98] S. S. Rawat, A. K. Keshri, R. Kaur, and A. Prasad, "Immunoinformatics approaches for vaccine design: a fast and secure strategy for successful vaccine development," *Vaccines*, vol. 11, no. 2, p. 211, 2023.

[99] S. Gupta, P. Kapoor, K. Chaudhary, A. Gautam, R. Kumar, 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.

[100] Y. Xiao, X. Tang, W. Wang, C. Li, Y. Li, Y. Hu, H. Chen, and Z. Liang, "Molecular dynamics simulation reveals specific interaction sites between scorpion toxins and kv1.2 channel: implications for design of highly selective drugs," *Scientific Reports*, vol. 7, p. 16474, 2017.

[101] A. Almaaytah and Q. Albalas, "Scorpion venom peptides with no disulfide bridges: a review," *Peptides*, vol. 51, pp. 35–45, 2014.

[102] N. Pashmforoosh and M. Baradaran, "Peptides with diverse functions from scorpion venom: a great opportunity for the treatment of a wide variety of diseases," *Iranian Biomedical Journal*, vol. 27, no. 2-3, pp. 84–99, 2023.

[103] 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, 2016.

[104] G. Shi, X. Kang, F. Dong, Y. Liu, N. Zhu, Y. Hu, H. Xu, X. Lao, and H. Zheng, "Dramp 3.0: an enhanced comprehensive data repository of antimicrobial peptides," *Nucleic Acids Research*, vol. 50, no. D1, pp. D488–D496, 2022.

[105] K. Guruprasad, B. V. B. Reddy, and M. W. Pandit, "Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence," *Protein Engineering*, vol. 4, no. 2, pp. 155–161, 1990.

[106] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis, "Highly accurate protein structure prediction with alphafold," *Nature*, vol. 596, no. 7873, pp. 583–589, 2021.

[107] F. A. Facchini, "Toll-like receptor 4 (tlr4) therapeutic modulation: a chemical biology approach," Master's thesis, Dept. of [Department Name], [University Name], 2018.

[108] S. Dallakyan and A. J. Olson, "Small-molecule library screening by docking with pyrx," in *Chemical Biology*, ser. Methods in Molecular Biology, J. Hempel, C. Williams, and C. Hong, Eds., vol. 1263. New York, NY: Humana Press, 2015, pp. 243–250.

[109] D. Szklarczyk, R. Kirsch, M. Koutrouli, K. Nastou, F. Mehryary, R. Hachilif, A. L. Gable, T. Fang, N. T. Doncheva, S. Pyysalo, P. Bork, L. J. Jensen, and C. von Mering, "The string database in 2023: protein–protein association

networks and functional enrichment analyses for any sequenced genome of interest," *Nucleic Acids Research*, vol. 51, no. D1, pp. D638–D646, 2023.

[110] M. M. Hasan, N. Schaduangrat, S. Basith, G. Lee, W. Shoombuatong, and B. Manavalan, "Hlppred-fuse: improved and robust prediction of hemolytic peptide and its activity by fusing multiple feature representation," *Bioinformatics*, vol. 36, no. 11, pp. 3350–3356, 2020.

[111] N. Sharma, L. D. Naorem, S. Jain, and G. P. S. Raghava, "Toxinpred2: an improved method for predicting toxicity of proteins," *Briefings in Bioinformatics*, vol. 23, no. 5, p. bbac174, 2022.

[112] M. N. Nguyen, N. L. Krutz, V. Limviphuvadh, A. L. Lopata, G. F. Gerberick, and S. Maurer-Stroh, "Allercatpro 2.0: a web server for predicting protein allergenicity potential," *Nucleic Acids Research*, vol. 50, no. W1, pp. W36–W43, 2022.

[113] J. Dauparas, I. Anishchenko, N. Bennett, H. Bai, R. J. Ragotte, L. F. Milles, B. I. M. Wicky, A. Courbet, R. J. de Haas, N. Bethel, P. J. Y. Leung, T. F. Huddy, S. Pellock, D. Tischer, F. Chan, B. Koepnick, H. Nguyen, A. Kang, B. Sankaran, A. K. Bera, N. P. King, and D. Baker, "Robust deep learning–based protein sequence design using proteinmpnn," *Science*, vol. 378, no. 6615, pp. 49–56, 2022.

[114] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel, and A. Bairoch, *Protein identification and analysis tools on the ExPASy server*. Totowa, NJ: Humana Press, 2005, pp. 571–607.

[115] K. J. Bowers, E. Chow, H. Xu, R. O. Dror, M. P. Eastwood, B. A. Gregersen, J. L. Klepeis, I. Kolossvary, M. A. Moraes, F. D. Sacerdoti, J. K. Salmon, Y. Shan, and D. E. Shaw, "Scalable algorithms for molecular dynamics simulations on commodity clusters," in *Proc. ACM/IEEE SC06: International Conference for High Performance Computing, Networking, Storage and Analysis*, Tampa, FL, USA, Nov 2006, p. 84.

[116] M. Odunsi, J. L. Adams, M. K. Bowman *et al.*, "Ny-eso-1 expression predicts an aggressive phenotype of ovarian cancer," *Cancer Immunology Research*, 2017.

[117] M. R. Felder, S. Kapur, J. Gonzalez-Bosquet *et al.*, "Muc16 (ca125): tumor biomarker to cancer therapy, a work in progress," *Molecular Cancer*, vol. 13, p. 129, 2014.

[118] J. A. Konner, K. M. Bell-McGuinn, P. Sabbatini *et al.*, "Farletuzumab, a humanized monoclonal antibody against folate receptor α, in epithelial ovarian cancer: a phase i study," *Clinical Cancer Research*, vol. 16, no. 21, pp. 5288–5295, 2010.

[119] Y. Yan, H. Tao, J. He, and S.-Y. Huang, "The hdock server for integrated protein–protein docking," *Nature Protocols*, vol. 15, no. 5, pp. 1829–1852, 2020.

[120] A. M. Torres, J. Kuchel, and P. W. Kuchel, "Scorpion venom peptides: Perspectives on research, potential, and recent developments," *Toxins*, vol. 2, no. 3, pp. 367–398, 2010.

[121] K. A. Brogden, "Antimicrobial peptides: Pore formers or metabolic inhibitors in bacteria?" *Nat. Rev. Microbiol.*, vol. 3, no. 3, pp. 238–250, 2005.

[122] B. S. Park, D. H. Song, H. M. Kim, B. S. Choi, H. Lee, and J. O. Lee, "The structural basis of lipopolysaccharide recognition by the tlr4-md-2 complex," *Nature*, vol. 458, no. 7242, pp. 1191–1195, 2009.

[123] G. Corzo, N. Escoubas, E. Stankiewicz, A. Postic, M. Gilles, and T. Nakajima, "Isolation, synthesis and pharmacological characterization of delta-palutoxins it, novel toxins from the spider paracoelotes luctuosus," *Eur. J. Biochem.*, vol. 267, no. 18, pp. 5783–5796, 2000.

[124] H. Jenssen, P. Hamill, and R. E. W. Hancock, "Peptide antimicrobial agents," *Clin. Microbiol. Rev.*, vol. 19, no. 3, pp. 491–511, 2006.

[125] K. A. Fitzgerald, S. M. McWhirter, K. L. Faia, D. C. Rowe, E. Latz, D. T. Golenbock, A. J. Coyle, S. M. Liao, and T. Maniatis, "Ikkepsilon and tbk1 are essential components of the irf3 signaling pathway," *Nat. Immunol.*, vol. 4, no. 5, pp. 491–496, 2003.

[126] S. A. Rosenberg, J. C. Yang, and N. P. Restifo, "Cancer immunotherapy: Moving beyond current vaccines," *Nat. Med.*, vol. 10, no. 9, pp. 909–915, 2004.

[127] Y. Shai, "Mechanism of the binding, insertion and destabilization of phospholipid bilayer membranes by alpha-helical antimicrobial and cell nonselective membrane-lytic peptides," *Biochim. Biophys. Acta*, vol. 1462, no. 1-2, pp. 55–70, 1999.

[128] R. E. W. Hancock and H. G. Sahl, "Antimicrobial and host-defence peptides as new anti-infective therapeutic strategies," *Nat. Biotechnol.*, vol. 24, no. 12, pp. 1551–1557, 2006.

[129] K. Guruprasad, B. V. B. Reddy, and M. W. Pandit, "Correlation between stability of a protein and its dipeptide composition: A novel approach for predicting in vivo stability of a protein from its primary sequence," *Protein Eng.*, vol. 4, no. 2, pp. 155–161, 1990.

[130] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, "Gromacs: Fast, flexible, and free," *J. Comput. Chem.*, vol. 26, no. 16, pp. 1701–1718, 2005.

[131] J. Y. Kang, X. Nan, M. S. Jin, S. J. Youn, Y. H. Ryu, S. Mah, S. H. Han, H. Lee, S. H. Paik, and J. O. Lee, "Recognition of lipopeptide patterns by toll-like receptor 2-toll-like receptor 6 heterodimer," *Immunity*, vol. 31, no. 6, pp. 873–884, 2009.

[132] S. C. Bast, R. C. Bast, and C. A. Skates, "Ca-125: The past and future of a specific tumour marker for ovarian cancer," *Int. J. Biol. Markers*, vol. 20, no. 1, pp. 23–31, 2005.

Appendix - I

Appendix I: Complete Master Library of 117 Scorpion-Derived Antimicrobial Peptides (Sequence Retrieval)

No.	DB ID	Source	Species	Sequence	Length	MW (Da)	Charge	pI	II	GRAVY
1	AP01512	APD3	*Isometrus maculatus*	FFSLLPSLIGGLVSAIK	17	1762.17	+2	8.75	34.34	1.594
2	AP01978	APD3	*Buthus martensii*	FIGAIARLLSKIF	13	1448.81	+3	11.00	-3.83	1.592
3	AP02216	APD3	*Vaejovis mexicanus smithi*	FLGALWNVAKSVF	13	1451.73	+2	8.75	30.41	1.208
4	DRAMP03688	DRAMP	*Tityus serrulatus*	FLGMIPGLIGGLISAFK	17	1734.17	+1	8.75	6.14	1.547
5	DRAMP03687	DRAMP	*Tityus serrulatus*	FLSLIPSLVGGSISAFK	17	1736.08	+1	8.75	12.47	1.324
6	AP05530	APD3	*Hadruroides lunatus*	FIGGLLKTLTSFF	13	1443.75	+1	8.75	12.19	1.338
7	AP00153	APD3	*Androctonus sp.*	RSVCRQIKICRRRGGCYYKCTNRPY	25	3080.66	+8	10.22	95.12	-1.056
8	DRAMP03739	DRAMP	*Tityus serrulatus*	SIVPIRCRSNRDCRRFCGFRGGRCTYARQCLCGY	34	3974.65	+7	9.80	66.94	-0.447
9	AP00301	APD3	*Hadrurus arizonensis*	GILDTIKSIASKVWNSKTVQDLKRKGINWVANKLGVSPQAA	41	4436.18	+5	10.30	28.54	-0.200
10	AP00374	APD3	*Opistophthalmus carinatus*	GKVWDWIKSTAKKLWNSEPVKELKNTALNAAKNLVAEKIGATPS	44	4836.61	+4	9.78	19.38	-0.518
11	AP00375	APD3	*Opistophthalmus carinatus*	GKVWDWIKSTAKKLWNSEPVKELKNTALNAAKNFVAEKIGATPS	44	4870.63	-4/+8	9.78	15.96	-0.541
12	AP00385	APD3	*Parabuthus schlechteri*	FKLGSFLKKAWKSKLAKKLRAKGKEMLKDYAKGLLEGGSEEVPGQ	45	4994.96	+7	10.00	19.43	-0.700
13	AP00430	APD3	*Opisthacanthus madagascariensis*	ILGKIWEGIKSLF	13	1503.85	+2	8.59	-2.08	0.777
14	AP00482	APD3	*Isometrus maculatus*	FSFKRLKGFAKKLWNSKLARKIRTKGLKYVKNFAKDMLSEGEEAPPAAEPPVEAPQ	56	6348.46	+7	10.07	41.01	-0.730
15	AP00562	APD3	*Pandinus imperator*	GKVWDWIKSAAKKIWSSEPVSQLKGQVLNAAKNYVAEKIGATPT	44	4799.55	+4	9.78	14.51	-0.366

16	AP00660	APD3	*Pandinus imperator*	FWGALAKGALKLIPSLFSSFSKKD	24	2612.11	+3	10.00	2.42	0.329
17	DRAMP03719	DRAMP	*Leiurus quinquestriatus*	GFGCPLNQGACHRHCRSIRRRGGYCAGFFKQTCCYRN	37	4224.87	+9	9.69	79.18	-0.651
18	AP00713	APD3	*Androctonus australis*	GFGCPFNQGACHRHCRSIRRRGGYCAGLFKQTCTCYR	37	4211.87	+9	9.69	89.86	-0.576
19	AP01327	APD3	*Androctonus australis*	LFGLIPSLIGGLVSAFK	17	1732.14	+1	8.75	17.46	1.618
20	AP01544	APD3	*Mesobuthus eupeus*	IFGAIAGLLKNIF	13	1376.70	+2	8.75	29.97	1.700
21	AP01545	APD3	*Mesobuthus eupeus*	FFGHLFKLATKIIPSLFQ	18	2107.57	+2	10.00	0.76	0.856
22	AP01699	APD3	*Mesobuthus eupeus*	GRGREFMSNLKEKLSGVKEKMKNS	24	2754.22	-3/+7	10.17	29.66	-1.300
23	AP01700	APD3	*Mesobuthus eupeus*	VKLIQIRIWIQYVTVLQMFSMKTKQ	25	3095.84	+4	10.46	21.15	0.408
24	AP01753	APD3	*Vaejovis mexicanus*	GIWSSIKNLASKAWNSDIGQSLRNKAAGAINKFVADKIGVTPSQAAS	47	4872.52	+4	10.17	38.21	-0.162
25	AP01791	APD3	*Chaerilus tricostatus*	FLWGLIPGAISAVTSLIKK	19	2014.48	+3	10.00	1.56	1.163
26	AP01792	APD3	*Pandinus imperator*	GWINEEKIQKKIDERMGNTVLGGMAKAIVHKMAKNEFQCMANMDMLGNCEKHCQTSGEKGYCHGTKCKCGTPLSY	75	8356.74	+6	8.58	22.53	-0.640
27	AP01793	APD3	*Heterometrus laoticus*	GWINEEKIQKKIDEKIGNNILGGMAKAVVHKLAKGEFQCVANIDTMGNCETHCQKTSGEKGFCHGTKCKCGKPLSY	76	8298.63	+7	8.79	19.69	-0.553
28	AP01977	APD3	*Mesobuthus martensii*	FLFSLIPSAISGLISAFK	18	1911.31	+2	8.75	10.19	1.544
29	AP01980	APD3	*Opisthacanthus madagascariensis*	IFGAIWNGIKSLF	13	1465.76	+2	8.75	-12.43	1.138
30	AP02026	APD3	*Scorpiops tibetanus*	GFWGKLWEGVKSAI	14	1577.85	+2	8.59	4.61	0.143
31	AP02027	APD3	*Scorpiops tibetanus*	GFWGSLWEGVKSVV	14	1550.78	+1	6.00	16.74	0.514
32	AP02098	APD3	*Heterometrus spinifer*	SGTSEKERESGRLLGVVKRLIVCFRSPFP	29	3248.79	+3	10.04	70.92	-0.276
33	AP02124	APD3	*Urodacus yaschenkoi*	GFWGKLWEGVKNAI	14	1604.87	+2	8.59	4.61	-0.050

34	AP02125	APD3	*Urodacus yaschenkoi*	FWGKLWEGVKNAI	13	1547.82	+2	8.59	4.19	-0.023
35	AP02126	APD3	*Urodacus yaschenkoi*	ILSAIWSGIKSLF	13	1434.74	+2	8.75	10.98	1.392
36	AP02127	APD3	*Urodacus yaschenkoi*	IWSAIWSGIKGLL	13	1443.75	+2	8.75	-10.36	1.138
37	DRAMP18488	DRAMP	*Centruroides suffusus suffusus*	FFGSLLSLGSKLLPSVFKLFQRKKE	25	2870.47	+4	10.46	12.20	0.220
38	AP02213	APD3	*Pandinus imperator*	GILGKLWEGFKSIV	14	1546.87	+1	8.59	4.96	0.671
39	AP02214	APD3	*Pandinus imperator*	IFGAIWKGISSLL	13	1404.72	+1	8.75	4.45	1.423
40	AP02215	APD3	*Pandinus imperator*	FLSTIWNGIKSLL	13	1491.79	+1	8.75	-5.90	0.938
41	AP02217	APD3	*Vaejovis mexicanus smithi*	FLSTLWNAAKSIF	13	1497.76	+2	8.75	36.94	0.823
42	AP02218	APD3	*Heterometrus petersii*	IFKAIWSGIKSLF	13	1509.85	+2	10.00	-15.39	1.077
43	AP02334	APD3	*Heterometrus petersii*	ILGKIWEGIKSIF	13	1503.85	-1/+2	8.59	-2.08	0.831
44	AP02335	APD3	*Heterometrus petersii*	ILSYLWNGIKSIF	13	1553.86	+1	8.59	27.13	0.946
45	AP02356	APD3	*Androctonus amoreuxi*	FLFSLIPHAIGGLISAFK	18	1931.35	+3	8.76	6.35	1.433
46	AP02357	APD3	*Androctonus amoreuxi*	FLFSLIPSAISGLISAF	17	1783.14	+1	5.52	19.05	1.865
47	AP02358	APD3	*Mesobuthus martensii*	FRFGSFLKKVWKSKLAKKLRSKGKQLLKDYANKVLNGPEEEAAAPAE	47	5321.26	+7	10.08	46.47	-0.736
48	AP02403	APD3	*Heterometrus petersii*	GILGKLWEGVKSIF	14	1546.87	+1	8.59	19.71	0.671
49	AP02518	APD3	*Centruroides margaritatus*	ARDGYIVDEKGCKFACFIN	19	2149.47	0	6.15	11.28	-0.053
50	AP02545	APD3	*Tityus discrepans*	KDGYIIEHRGCKYSCFFGTNSWCNTECTLKKGSSGYCAWPACWCYGLPDNVKIFDSNNLKC	61	6927.89	+3	8.17	44.50	-0.403
51	AP02546	APD3	*Tityus discrepans*	KDGYLVGNDGCKYSCFTRPGTYCANECSRVKGKDGYCYAWMACYCYSMPNWVKTWNRATNRCGR	64	7374.39	+6	8.99	34.98	-0.723
52	AP02547	APD3	*Mesobuthus martensii*	DNGYLLDKYTGCKVWCVINNESCNSECKIRGGYYGYC	66	7701.78	+4	8.65	21.23	-0.577

				YFWKLACFCQGARKSELW NYNTNKCNGKL						
53	AP02549	APD3	*Heterometrus laoticus*	GVWDWLKKTAKNVWNSD IVKQLKGKAINAAKNYVAE KIGATPS	43	4742.50	+6	9.90	11.00	-0.472
54	AP02550	APD3	*Heterometrus laoticus*	FWGALAKGALKLIPSLVSS FTKKD	24	2578.09	+4	10.00	2.42	0.392
55	AP02551	APD3	*Heterometrus spinifer*	ILGEIWKGIKDIL	13	1497.84	+1	6.07	28.28	0.700
56	AP02552	APD3	*Opistophthalmus glabrifrons*	KWFNEKSIQNKIDEKIGKNF LGGMAKAVVHKLAKNEF MCVANVDMTKSCDTHCQK ASGEKGYCHGTKCKCGVP LSY	76	8428.84	+9	9.10	14.52	-0.507
57	AP02560	APD3	*Mesobuthus martensii*	FIGAVAGLLSKIF	13	1335.65	+1	8.75	-3.83	1.885
58	AP02561	APD3	*Mesobuthus eupeus*	GFREKHFQRFVKYAVPEST LRTVLQTVVHKVGKTQFG CPAYQGYCDDHCQDIEKKE GFCHGFKCKCGIPMGF	72	8261.56	+8	8.82	36.88	-0.460
59	AP02562	APD3	*Hadrurus gertschi*	GWMSEKKVQGILDKKLPE GIIRNAAKAIVHKMAKNQF GCFANVDVKGDCKRHCKA EDKEGICHGTKCKCGVPIS YL	76	8374.91	+10	9.26	40.24	-0.409
60	AP02563	APD3	*Hadrurus gertschi*	KSTVGQKLKKKLNQAVDK VKEVLNKSEYMCPVVSSFC KQHCARLGKSGQCDLLECI CS	58	6432.63	+7	9.17	33.81	-0.328
61	AP02603	APD3	*Vaejovis punctatus*	LPFFLLSLIPSAISAIKKI	19	2071.62	+3	10.00	29.43	1.526
62	AP02604	APD3	*Vaejovis punctatus*	FWGFLGKLAMKAVPSLIGG NKSSSK	25	2624.14	+4	10.48	40.49	0.192
63	DRAMP18443	DRAMP	*Scorpio maurus palmatus*	IWSFLIKAATKLLPSLFGGG KKDS	24	2578.09	+3	10.00	33.33	0.312
64	DRAMP18442	DRAMP	*Scorpio maurus palmatus*	GVWDWIKKTAGKIWNSEP VKALKSQALNAAKNFVAE KIGATPS	43	4654.39	+4	9.88	19.02	-0.316
65	DRAMP18418	DRAMP	*Tityus obscurus*	FIGMIPGLIGGLISAFK	17	1734.17	+1	8.75	6.14	1.588
66	DRAMP18417	DRAMP	*Tityus obscurus*	FFGTLFKLGSKLIPGVMKLF SKKKER	26	3000.73	+6	10.68	-10.33	0.019
67	DRAMP18416	DRAMP	*Tityus obscurus*	FIGMIPGLIGGLISAIK	17	1700.16	+1	8.75	9.98	1.594

68	DRAMP18415	DRAMP	*Opisthacanthus cayaporum*	FWSFLVKAASKILPSLIGGGDDNKSSS	27	2825.21	+1	8.50	59.00	0.115
69	DRAMP18414	DRAMP	*Opisthacanthus cayaporum*	GILGKIWEGVKSLI	14	1512.86	+1	8.59	2.79	0.793
70	DRAMP18401	DRAMP	*Urodacus yaschenkoi*	ILSAIWSGIKGLL	13	1370.70	+1	8.75	4.45	1.500
71	DRAMP18400	DRAMP	*Urodacus yaschenkoi*	FLSTIWNGIKGLL	13	1461.77	+1	8.75	-12.43	0.969
72	DRAMP18399	DRAMP	*Urodacus yaschenkoi*	FPFLLSLIPSAISAIKRL	18	1986.47	+2	11.00	53.34	1.328
73	DRAMP18398	DRAMP	*Urodacus manicatus*	ISQSDAILSAIWSGIKSLF	19	2036.36	0	5.84	43.94	0.832
74	DRAMP18397	DRAMP	*Urodacus manicatus*	FFSALLSGIKSLF	13	1429.72	+1	8.75	-3.83	1.492
75	AP02972	APD3	*Mesobuthus eupeus*	FFGHLFKLATKIIPSFFRRKNQ	22	2696.24	+6	12.02	25.05	-0.091
76	AP02973	APD3	*Mesobuthus eupeus*	FLFSLIPSAISGLISAFKGRRKRDLN	26	2907.45	+4	11.72	41.44	0.262
77	AP02974	APD3	*Mesobuthus eupeus*	FLFSLIPSAISGLINAFK	18	1938.34	+2	8.75	10.19	1.394
78	AP02975	APD3	*Mesobuthus eupeus*	FFGHLFKLATKIIPSLFQRKKE	22	2649.22	+5	10.46	2.44	-0.018
79	AP02981	APD3	*Mesobuthus eupeus*	FKFGSFIKRMWRSKLAKKLRAKGKELLRDYANRVLSPEEEAAAPAPYPA	49	5637.64	+7	10.28	70.44	-0.643
80	AP03022	APD3	*Mesobuthus martensii*	GFGCPFNQGQCHKHCQSIRRRGGYCDGFLKTRCVCYR	37	4282.95	+8	9.38	67.12	-0.714
81	AP03045	APD3	*Scorpio maurus palmatus*	GWINEKKMQQKIDEKIGKNIIGGMAKAVIHKMAKNEFQCVANVDTLGNCKKHCAKTTGEKGYCHGTKCKCGIELSY	76	8402.89	-7/+14	9.22	28.35	-0.568
82	AP03123	APD3	*Liocheles australasiae*	GGILREKYFHKAADALTSNIPIPVVKDVLKSAANQMIRKIGKVQQACAFNKDLAGWCEKSCQEAEGKKGYCHGTKCKCGKPIDY	84	9201.75	+9	9.22	32.07	-0.446
83	AP03124	APD3	*Liocheles australasiae*	AKKPFVQRVKNAASKAYNKLKGLAMQSQYGCPIISNMCEDHCRRKKMEGQCDLLDCVCS	59	6638.83	+7	9.20	40.21	-0.522
84	AP03125	APD3	*Heterometrus laoticus*	DFPLSKEYETCVRPRKCQPPLKCNKAQICVDPKKGW	36	4206.97	-4/+8	9.10	19.67	-0.942

85	AP03128	APD3	*Tityus serrulatus*	GFGGGRGGFGGGRGGFGGGGIGGGGFGGGYGGGKIKG	37	3048.29	+4	11.10	42.44	-0.224
86	AP03275	APD3	*Isometrus maculatus*	FLGSLFSIGSKLLPGVIKLFQRKKQ	25	2805.45	+5	11.33	16.59	0.332
87	AP03276	APD3	*Isometrus maculatus*	FFFLPSLIGGLVSAIK	16	1709.10	+2	8.75	35.86	1.681
88	AP03303	APD3	*Opisthacanthus madagascariensis*	IFGKIWEGIKSLF	13	1537.86	+2	8.59	-16.89	0.700
89	AP03335	APD3	*Chelifer cancroides*	FFGAIAKLAMKFLPAIYKQIQKKRK	25	2939.69	+8	10.75	40.73	0.016
90	AP03639	APD3	*Heterometrus petersii*	IFKAIWSGINRLF	13	1564.89	+2	11.00	-8.86	0.823
91	AP03723	APD3	*Tityus tarnensis*	FLGSLFSIGSKLLPGVFKLFSRKKQ	25	2798.41	+6	11.33	21.68	0.372
92	AP03724	APD3	*Tityus tarnensis*	IFGMIPGLIGGLISAFK	17	1734.17	+2	8.75	6.14	1.588
93	AP03725	APD3	*Tityus tarnensis*	FFSLIPSLIGGLVSAIK	17	1762.17	+2	8.75	21.31	1.635
94	AP03939	APD3	*Euscorpiops vachoni*	GLINEKKVQQYLDEKLPNGVVKGALKSLVHKAAKNQNLCAFNVDTVGMCDADCKRQGKAKGVCHGTKCKCDVELSYKK	78	8509.96	-8/+15	9.22	13.32	-0.510
95	AP03940	APD3	*Androctonus amoreuxi*	GLIDVKCYATSQCWAPCKKETGSGQSKCQNNQCRCY	36	4002.56	-2/+5	8.64	27.98	-0.753
96	AP03941	APD3	*Liocheles australasiae*	FLSALWGVAKSLF	13	1438.73	+2	8.75	19.47	1.385
97	AP03942	APD3	*Liocheles australasiae*	FPFLLSLIPSAISALKKL	18	1958.46	+3	10.00	30.51	1.322
98	AP04851	APD3	*Tityus serrulatus*	KEGYLMDHEGCKLSCFIRPSGYCGRECGIKKGSSGYCAWPACYCYGLPNWVKVWDRATNKC	61	6890.98	-5/+9	8.67	29.47	-0.439
99	AP04935	APD3	*Liocheles australasiae*	GIWGWIKKTAKKVWNSDIANQLKNKALNAAKNYVAEKVGATPAEAGQ	47	5082.84	+5	9.90	7.72	-0.547
100	AP04939	APD3	*Androctonus bicolor*	GFGCPFNQGRCHRHCRSIGRRGGYCRGIFKQTCACYRK	38	4381.10	+11	10.14	61.87	-0.803
101	AP04940	APD3	*Leiurus quinquestriatus hebraeus*	GFGCPLNQGACHRHCRSIRRRGGYCAGFFKQTCTCYRN	38	4325.97	+9	9.69	85.94	-0.653
102	DRAMP18680	DRAMP	*Chaerilus tryznai*	GIADILKGLL	10	1012.26	0	5.84	-5.70	1.400

103	DRAMP18681	DRAMP	*Chaerilus tryznai*	LLGGLLQSLL	10	1026.28	0	5.52	97.88	1.770
104	DRAMP20775	DRAMP	*Chaerilus tryznai*	FLGFLKNLF	9	1098.35	+1	8.75	-0.54	1.333
105	DRAMP18682	DRAMP	*Chaerilus tryznai*	GILDAITGLL	10	985.19	-1	3.80	11.28	1.720
106	DRAMP18683	DRAMP	*Chaerilus tricostatus*	YIRDFITRRPPFGNI	15	1865.17	+2	10.74	104.65	-0.467
107	DRAMP18684	DRAMP	*Chaerilus tricostatus*	GILDALTGIL	10	985.19	-1	3.80	22.05	1.720
108	DRAMP18685	DRAMP	*Chaerilus tricostatus*	GVVDTLKNLLMGLL	14	1485.85	0	5.84	-18.25	1.207
109	DRAMP18686	DRAMP	*Chaerilus tricarinatus*	FLFNVIPHAINATASLIKK	19	2097.53	+3	10.00	3.49	0.800
110	DRAMP18687	DRAMP	*Chaerilus tricarinatus*	FLVGILPRMRGFITPFLKKVR	21	2489.15	+5	12.31	9.64	0.624
111	DRAMP18688	DRAMP	*Chaerilus tricarinatus*	FLWSLIPSAISAVTSLIKK	19	2074.53	+2	10.00	21.62	1.121
112	DRAMP18689	DRAMP	*Chaerilus tricarinatus*	FDLGGLIKGVVDLF	14	1492.78	-1	4.21	4.05	1.271
113	DRAMP18695	DRAMP	*Chaerilus tricarinatus*	KHFGKDSNFPF	11	1323.47	+1	8.60	30.81	-1.127
114	DRAMP18717	DRAMP	*Leiurus quinquestriatus hebraeus*	EFTNVSCTTSKECWSVCQRLHNTSRGKCMNKKCRCYS	37	4319.97	+6	9.13	59.46	-0.832
115	DRAMP21024	DRAMP	*Tityus stigmurus*	FFSLIPSLVGGLISAFK	17	1796.18	+1	8.75	12.47	1.535
116	DRAMP21251	DRAMP	*Androctonus aeneas*	FLFSLIPSVIAGLVSAIRN	19	2017.44	+1	9.75	24.59	1.584
117	DRAMP21252	DRAMP	*Androctonus aeneas*	FLFSLIPSAIAGLVSAIRN	19	1989.39	+1	9.75	24.59	1.458

Appendix II: Physicochemical Profiles of All 117 Scorpion-Derived AMPs

No.	AMP ID	Sequence	Length (aa)	MW (Da)	Charge (pH 7.4)	pI	II	AI	GRAVY
1	AP01512	FFSLLPSLIGGLVSAIK	17	1762.17	+2	8.75	34.34	160.59	1.594
2	AP01978	FIGAIARLLSKIF	13	1448.81	+3	11.00	-3.83	165.38	1.592
3	AP02216	FLGALWNVAKSVF	13	1451.73	+2	8.75	30.41	120.00	1.208
4	DRAMP 03688	FLGMIPGLIGGLISAFK	17	1734.17	+1	8.75	6.14	143.53	1.547
5	DRAMP 03687	FLSLIPSLVGGSISAFK	17	1736.08	+1	8.75	12.47	137.65	1.324
6	AP05530	FIGGLLKTLTSFF	13	1443.75	+1	8.75	12.19	120.00	1.338
7	AP00153	RSVCRQIKICRRRGGCYYKCTNRPY	25	3080.66	+8	10.22	95.12	42.80	-1.056
8	DRAMP 03739	SIVPIRCRSNRDCRRFCGFRGGRCTYARQCL CGY	34	3974.65	+7	9.80	66.94	45.88	-0.447
9	AP00301	GILDTIKSIASKVWNSKTVQDLKRKGINWVA NKLGVSPQAA	41	4436.18	+5	10.30	28.54	104.63	-0.200
10	AP00374	GKVWDWIKSTAKKLWNSEPVKELKNTALN AAKNLVAEKIGATPS	44	4836.61	+4	9.78	19.38	86.59	-0.518
11	AP00375	GKVWDWIKSTAKKLWNSEPVKELKNTALN AAKNFVAEKIGATPS	44	4870.63	-4,+8	9.78	15.96	77.73	-0.541
12	AP00385	FKLGSFLKKAWKSKLAKKLRAKGKEMLKD YAKGLLEGGSEEVPGQ	45	4994.96	+7	10.00	19.43	76.00	-0.700
13	AP00430	ILGKIWEGIKSLF	13	1503.85	+2	8.59	-2.08	150.00	0.777
14	AP00482	FSFKRLKGFAKKLWNSKLARKIRTKGLKYV KNFAKDMLSEGEEAPPAAEPPVEAPQ	56	6348.46	+7	10.07	41.01	64.64	-0.730
15	AP00562	GKVWDWIKSAAKKIWSSEPVSQLKGQVLN AAKNYVAEKIGATPT	44	4799.55	+4	9.78	14.51	84.32	-0.366
16	AP00660	FWGALAKGALKLIPSLFSSFSKKD	24	2612.11	+3	10.00	2.42	93.75	0.329
17	DRAMP 03719	GFGCPLNQGACHRHCRSIRRRGGYCAGFFK QTCCYRN	37	4224.87	+9	9.69	79.18	26.49	-0.651
18	AP00713	GFGCPFNQGACHRHCRSIRRRGGYCAGLFK QTCTCYR	37	4211.87	+9	9.69	89.86	26.49	-0.576
19	AP01327	LFGLIPSLIGGLVSAFK	17	1732.14	+1	8.75	17.46	160.59	1.618
20	AP01544	IFGAIAGLLKNIF	13	1376.70	+2	8.75	29.97	165.38	1.700
21	AP01545	FFGHLFKLATKIIPSLFQ	18	2107.57	+2	10.00	0.76	113.89	0.856
22	AP01699	GRGREFMSNLKEKLSGVKEKMKNS	24	2754.22	-3,+7	10.17	29.66	44.58	-1.300
23	AP01700	VKLIQIRIWIQYVTVLQMFSMKTKQ	25	3095.84	+4	10.46	21.15	128.40	0.408
24	AP01753	GIWSSIKNLASKAWNSDIGQSLRNKAAGAIN KFVADKIGVTPSQAAS	47	4872.52	+4	10.17	38.21	87.45	-0.162
25	AP01791	FLWGLIPGAISAVTSLIKK	19	2014.48	+3	10.00	1.56	148.95	1.163

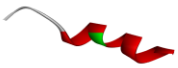
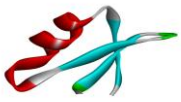
26	AP01792	GWINEEKIQKKIDERMGNTVLGGMAKAIVH KMAKNEFQCMANMDMLGNCEKHCQTSGE KGYCHGTKCKCGTPLSY	75	8356.74	+6	8.58	22.53	49.47	-0.640
27	AP01793	GWINEEKIQKKIDEKIGNNILGGMAKAVVH KLAKGEFQCVANIDTMGNCETHCQKTSGEK GFCHGTKCKCGKPLSY	76	8298.63	+7	8.79	19.69	62.89	-0.553
28	AP01977	FLFSLIPSAISGLISAFK	18	1911.31	+2	8.75	10.19	141.11	1.544
29	AP01980	IFGAIWNGIKSLF	13	1465.76	+2	8.75	-12.43	127.69	1.138
30	AP02026	GFWGKLWEGVKSAI	14	1577.85	+2	8.59	4.61	83.57	0.143
31	AP02027	GFWGSLWEGVKSVV	14	1550.78	+1	6.00	16.74	90.00	0.514
32	AP02098	SGTSEKERESGRLLGVVKRLIVCFRSPFP	29	3248.79	+3	10.04	70.92	83.79	-0.276
33	AP02124	GFWGKLWEGVKNAI	14	1604.87	+2	8.59	4.61	83.57	-0.050
34	AP02125	FWGKLWEGVKNAI	13	1547.82	+2	8.59	4.19	90.00	-0.023
35	AP02126	ILSAIWSGIKSLF	13	1434.74	+2	8.75	10.98	157.69	1.392
36	AP02127	IWSAIWSGIKGLL	13	1443.75	+2	8.75	-10.36	157.69	1.138
37	DRAMP 18488	FFGSLLSLGSKLLPSVFKLFQRKKE	25	2870.47	+4	10.46	12.20	105.20	0.220
38	AP02213	GILGKLWEGFKSIV	14	1546.87	+1	8.59	4.96	132.14	0.671
39	AP02214	IFGAIWKGISSLL	13	1404.72	+1	8.75	4.45	157.69	1.423
40	AP02215	FLSTIWNGIKSLL	13	1491.79	+1	8.75	-5.90	150.00	0.938
41	AP02217	FLSTLWNAAKSIF	13	1497.76	+2	8.75	36.94	105.38	0.823
42	AP02218	IFKAIWSGIKSLF	13	1509.85	+2	10.00	-15.39	127.69	1.077
43	AP02334	ILGKIWEGIKSIF	13	1503.85	-1,+2	8.59	-2.08	150.00	0.831
44	AP02335	ILSYLWNGIKSIF	13	1553.86	+1	8.59	27.13	150.00	0.946
45	AP02356	FLFSLIPHAIGGLISAFK	18	1931.35	+3	8.76	6.35	141.11	1.433
46	AP02357	FLFSLIPSAISGLISAF	17	1783.14	+1	5.52	19.05	149.41	1.865
47	AP02358	FRFGSFLKKVWKSKLAKKLRSKGKQLLKDY ANKVLNGPEEEAAAPAE	47	5321.26	+7	10.08	46.47	74.89	-0.736
48	AP02403	GILGKLWEGVKSIF	14	1546.87	+1	8.59	19.71	132.14	0.671
49	AP02518	ARDGYIVDEKGCKFACFIN	19	2149.47	0	6.15	11.28	66.84	-0.053
50	AP02545	KDGYIIEHRGCKYSCFFGTNSWCNTECTLKK GSSGYCAWPACWCYGLPDNVKIFDSNNLKC	61	6927.89	+3	8.17	44.50	46.39	-0.403
51	AP02546	KDGYLVGNDGCKYSCFTRPGTYCANECSRV KGKDGYCYAWMACYCYSMPNWVKTWNR ATNRCGR	64	7374.39	+6	8.99	34.98	25.94	-0.723
52	AP02547	DNGYLLDKYTGCKVWCVINNESCNSECKIR GGYYGYCYFWKLACFCQGARKSELWNYNT NKCNGKL	66	7701.78	+4	8.65	21.23	53.18	-0.577
53	AP02549	GVWDWLKKTAKNVWNSDIVKQLKGKAIN AAKNYVAEKIGATPS	43	4742.50	+6	9.90	11.00	86.28	-0.472
54	AP02550	FWGALAKGALKLIPSLVSSFTKKD	24	2578.09	+4	10.00	2.42	105.83	0.392
55	AP02551	ILGEIWKGIKDIL	13	1497.84	+1	6.07	28.28	180.00	0.700

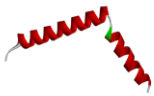
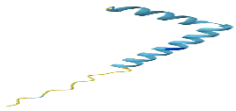
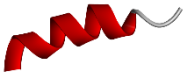
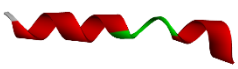
56	AP02552	KWFNEKSIQNKIDEKIGKNFLGGMAKAVVH KLAKNEFMCVANVDMTKSCDTHCQKASGE KGYCHGTKCKCGVPLSY	76	8428.84	+9	9.10	14.52	56.45	-0.507
57	AP02560	FIGAVAGLLSKIF	13	1335.65	+1	8.75	-3.83	157.69	1.885
58	AP02561	GFREKHFQRFVKYAVPESTLRTVLQTVVHK VGKTQFGCPAYQGYCDDHCQDIEKKEGFCH GFKCKCGIPMGF	72	8261.56	+8	8.82	36.88	48.61	-0.460
59	AP02562	GWMSEKKVQGILDKKLPEGIIRNAAKAIVH KMAKNQFGCFANVDVKGDCKRHCKAEDK EGICHGTKCKCGVPISYL	76	8374.91	+10	9.26	40.24	73.16	-0.409
60	AP02563	KSTVGQKLKKKLNQAVDKVKEVLNKSEYM CPVVSSFCKQHCARLGKSGQCDLLECICS	58	6432.63	+7	9.17	33.81	80.52	-0.328
61	AP02603	LPFFLLSLIPSAISAIKKI	19	2071.62	+3	10.00	29.43	174.74	1.526
62	AP02604	FWGFLGKLAMKAVPSLIGGNKSSSK	25	2624.14	+4	10.48	40.49	82.00	0.192
63	DRAMP 18443	IWSFLIKAATKLLPSLFGGGKKDS	24	2578.09	+3	10.00	33.33	105.83	0.312
64	DRAMP 18442	GVWDWIKKTAGKIWNSEPVKALKSQALNA AKNFVAEKIGATPS	43	4654.39	+4	9.88	19.02	81.86	-0.316
65	DRAMP 18418	FIGMIPGLIGGLISAFK	17	1734.17	+1	8.75	6.14	135.56	1.588
66	DRAMP 18417	FFGTLFKLGSKLIPGVMKLFSKKKER	26	3000.73	+6	10.68	-10.33	86.15	0.019
67	DRAMP 18416	FIGMIPGLIGGLISAIK	17	1700.16	+1	8.75	9.98	157.22	1.594
68	DRAMP 18415	FWSFLVKAASKILPSLIGGGDDNKSSS	27	2825.21	+1	8.50	59.00	90.37	0.115
69	DRAMP 18414	GILGKIWEGVKSLI	14	1512.86	+1	8.59	2.79	149.33	0.793
70	DRAMP 18401	ILSAIWSGIKGLL	13	1370.70	+1	8.75	4.45	187.69	1.500
71	DRAMP 18400	FLSTIWNGIKGLL	13	1461.77	+1	8.75	-12.43	139.29	0.969
72	DRAMP 18399	FPFLLSLIPSAISAIKRL	18	1986.47	+2	11.00	53.34	162.78	1.328
73	DRAMP 18398	ISQSDAILSAIWSGIKSLF	19	2036.36	0	5.84	43.94	133.68	0.832
74	DRAMP 18397	FFSALLSGIKSLF	13	1429.72	+1	8.75	-3.83	127.69	1.492
75	AP02972	FFGHLFKLATKIIPSFFRRKNQ	22	2696.24	+6	12.02	25.05	75.45	-0.091
76	AP02973	FLFSLIPSAISGLISAFKGRRKRDLN	26	2907.45	+4	11.72	41.44	112.69	0.262
77	AP02974	FLFSLIPSAISGLINAFK	18	1938.34	+2	8.75	10.19	141.11	1.394
78	AP02975	FFGHLFKLATKIIPSLFQRKKE	22	2649.22	+5	10.46	2.44	93.18	-0.018
79	AP02981	FKFGSFIKRMWRSKLAKKLRAKGKELLRDY ANRVLSPEEEAAAPAPYPA	49	5637.64	+7	10.28	70.44	70.00	-0.643

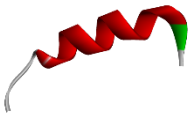
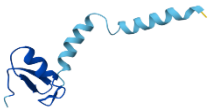
80	AP03022	GFGCPFNQGQCHKHCQSIRRRGGYCDGFLK TRCVCYR	37	4282.95	+8	9.38	67.12	28.92	-0.714
81	AP03045	GWINEKKMQQKIDEKIGKNIIGGMAKAVIH KMAKNEFQCVANVDTLGNCKKHCAKTTGE KGYCHGTKCKCGIELSY	76	8402.89	-7,+14	9.22	28.35	64.21	-0.568
82	AP03123	GGILREKYFHKAADALTSNIPIPVVKDVLKS AANQMIRKIGKVQQACAFNKDLAGWCEKS CQEAEGKKGYCHGTKCKCGKPIDY	84	9201.75	+9	9.22	32.07	70.95	-0.446
83	AP03124	AKKPFVQRVKNAASKAYNKLKGLAMQSQY GCPIISNMCEDHCRRKKMEGQCDLLDCVCS	59	6638.83	+7	9.20	40.21	62.88	-0.522
84	AP03125	DFPLSKEYETCVRPRKCQPPLKCNKAQICVD PKKGW	36	4206.97	-4,+8	9.10	19.67	51.39	-0.942
85	AP03128	GFGGGRGGFGGGRGGFGGGGIGGGGFGGG YGGGKIKG	37	3048.29	+4	11.10	42.44	21.08	-0.224
86	AP03275	FLGSLFSIGSKLLPGVIKLFQRKKQ	25	2805.45	+5	11.33	16.59	120.80	0.332
87	AP03276	FFFLPSLIGGLVSAIK	16	1709.10	+2	8.75	35.86	146.25	1.681
88	AP03303	IFGKIWEGIKSLF	13	1537.86	+2	8.59	-16.89	120.00	0.700
89	AP03335	FFGAIAKLAMKFLPAIYKQIQKKRK	25	2939.69	+8	10.75	40.73	94.00	0.016
90	AP03639	IFKAIWSGINRLF	13	1564.89	+2	11.00	-8.86	127.69	0.823
91	AP03723	FLGSLFSIGSKLLPGVFKLFSRKKQ	25	2798.41	+6	11.33	21.68	105.20	0.372
92	AP03724	IFGMIPGLIGGLISAFK	17	1734.17	+2	8.75	6.14	143.53	1.588
93	AP03725	FFSLIPSLIGGLVSAIK	17	1762.17	+2	8.75	21.31	160.59	1.635
94	AP03939	GLINEKKVQQYLDEKLPNGVVKGALKSLVH KAAKNQNLCAFNVDTVGMCDADCKRQGK AKGVCHGTKCKCDVELSYKK	78	8509.96	-8,+15	9.22	13.32	77.44	-0.510
95	AP03940	GLIDVKCYATSQCWAPCKKETGSGQSKCQN NQCRCY	36	4002.56	-2,+5	8.64	27.98	35.28	-0.753
96	AP03941	FLSALWGVAKSLF	13	1438.73	+2	8.75	19.47	127.69	1.385
97	AP03942	FPFLLSLIPSAISALKKL	18	1958.46	+3	10.00	30.51	162.78	1.322
98	AP04851	KEGYLMDHEGCKLSCFIRPSGYCGRECGIKK GSSGYCAWPACYCYGLPNWVKVWDRATN KC	61	6890.98	-5,+9	8.67	29.47	46.39	-0.439
99	AP04935	GIWGWIKKTAKKVWNSDIANQLKNKALNA AKNYVAEKVGATPAEAGQ	47	5082.84	+5	9.90	7.72	79.15	-0.547
100	AP04939	GFGCPFNQGRCHRHCRSIGRRGGYCRGIFKQ TCACYRK	38	4381.10	+11	10.14	61.87	23.16	-0.803
101	AP04940	GFGCPLNQGACHRHCRSIRRRGGYCAGFFK QTCTCYRN	38	4325.97	+9	9.69	85.94	25.79	-0.653
102	DRAMP 18680	GIADILKGLL	10	1012.26	0	5.84	-5.70	205.00	1.400
103	DRAMP 18681	LLGGLLQSLL	10	1026.28	0	5.52	97.88	234.00	1.770
104	DRAMP 20775	FLGFLKNLF	9	1098.35	+1	8.75	-0.54	130.00	1.333

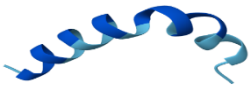
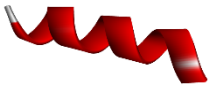
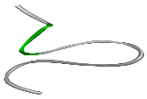
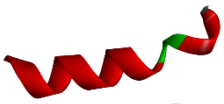
105	DRAMP 18682	GILDAITGLL	10	985.19	-1	3.80	11.28	205.00	1.720
106	DRAMP 18683	YIRDFITRRPPFGNI	15	1865.17	+2	10.74	104.65	78.00	-0.467
107	DRAMP 18684	GILDALTGIL	10	985.19	-1	3.80	22.05	205.00	1.720
108	DRAMP 18685	GVVDTLKNLLMGLL	14	1485.85	0	5.84	-18.25	180.71	1.207
109	DRAMP 18686	FLFNVIPHAINATASLIKK	19	2097.53	+3	10.00	3.49	133.68	0.800
110	DRAMP 18687	FLVGILPRMRGFITPFLKKVR	21	2489.15	+5	12.31	9.64	120.48	0.624
111	DRAMP 18688	FLWSLIPSAISAVTSLIKK	19	2074.53	+2	10.00	21.62	148.95	1.121
112	DRAMP 18689	FDLGGLIKGVVDLF	14	1492.78	-1	4.21	4.05	152.86	1.271
113	DRAMP 18695	KHFGKDSNFPF	11	1323.47	+1	8.60	30.81	0.00	-1.127
114	DRAMP 18717	EFTNVSCTTSKECWSVCQRLHNTSRGKCMNKKCRCYS	37	4319.97	+6	9.13	59.46	26.22	-0.832
115	DRAMP 21024	FFSLIPSLVGGLISAFK	17	1796.18	+1	8.75	12.47	137.65	1.535
116	DRAMP 21251	FLFSLIPSVIAGLVSAIRN	19	2017.44	+1	9.75	24.59	164.21	1.584
117	DRAMP 21252	FLFSLIPSAIAGLVSAIRN	19	1989.39	+1	9.75	24.59	154.21	1.458

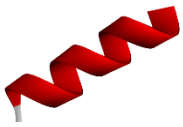
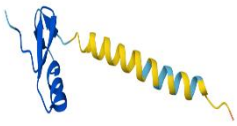
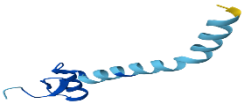
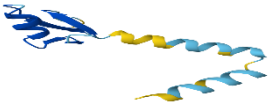
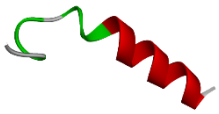
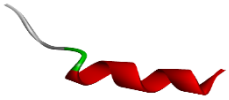
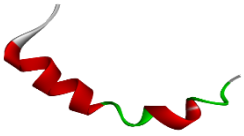
Appendix III: Secondary Structure Predictions of All 117 Scorpion-Derived AMPs

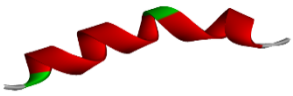
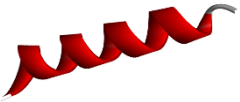
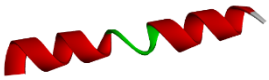
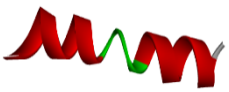
No.	AMP ID	Sequence	Predicted Secondary Structure
1	AP01512	FFSLLPSLIGGLVSAIK	
2	AP01978	FIGAIARLLSKIF	
3	AP02216	FLGALWNVAKSVF	
4	DRAMP03688	FLGMIPGLIGGLISAFK	
5	DRAMP03687	FLSLIPSLVGGSISAFK	
6	AP05530	FIGGLLKTLTSFF	
7	AP00153	RSVCRQIKICRRRGGCYYKCTNRPY	
8	DRAMP03739	SIVPIRCRSNRDCRRFCGFRGGRCTYARQCLCGY	
9	AP00301	GILDTIKSIASKVWNSKTVQDLKRKGINWVANKLGVSPQAA	

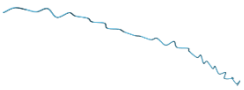
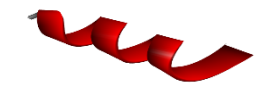
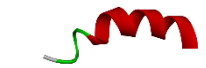
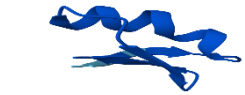
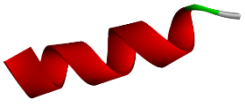
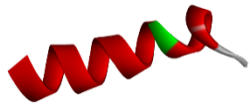
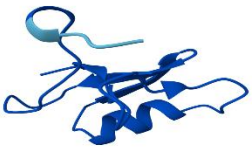
10	AP00374	GKVWDWIKSTAKKLWNSEPVKELKNTALNAAKNLVAEKIGATPS	
11	AP00375	GKVWDWIKSTAKKLWNSEPVKELKNTALNAAKNFVAEKIGATPS	
12	AP00385	FKLGSFLKKAWKSKLAKKLRAKGKEMLKDYAKGLLEGGSEEVPGQ	
13	AP00430	ILGKIWEGIKSLF	
14	AP00482	FSFKRLKGFAKKLWNSKLARKIRTKGLKYVKNFAKDMLSEGEEAPP AAEPPVEAPQ	
15	AP00562	GKVWDWIKSAAKKIWSSEPVSQLKGQVLNAAKNYVAEKIGATPT	
16	AP00660	FWGALAKGALKLIPSLFSSFSKKD	
17	DRAMP03719	GFGCPLNQGACHRHCRSIRRRGGYCAGFFKQTCCYRN	
18	AP00713	GFGCPFNQGACHRHCRSIRRRGGYCAGLFKQTCTCYR	
19	AP01327	LFGLIPSLIGGLVSAFK	
20	AP01544	IFGAIAGLLKNIF	
21	AP01545	FFGHLFKLATKIIPSLFQ	

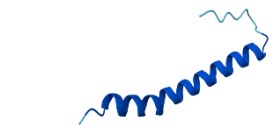
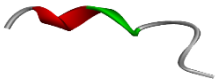
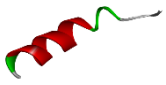
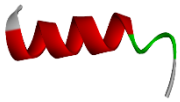
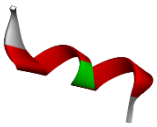
22	AP01699	GRGREFMSNLKEKLSGVKEKMKNS	
23	AP01700	VKLIQIRIWIQYVTVLQMFSMKTKQ	
24	AP01753	GIWSSIKNLASKAWNSDIGQSLRNKAAGAINKFVADKIGVTPSQAAS	
25	AP01791	FLWGLIPGAISAVTSLIKK	
26	AP01792	GWINEEKIQKKIDERMGNTVLGGMAKAIVHKMAKNEFQCMANMD MLGNCEKHCQTSGEKGYCHGTKCKCGTPLSY	
27	AP01793	GWINEEKIQKKIDEKIGNNILGGMAKAVVHKLAKGEFQCVANIDTM GNCETHCQKTSGEKGFCHGTKCKCGKPLSY	
28	AP01977	FLFSLIPSAISGLISAFK	
29	AP01980	IFGAIWNGIKSLF	
30	AP02026	GFWGKLWEGVKSAI	
31	AP02027	GFWGSLWEGVKSVV	
32	AP02098	SGTSEKERESGRLLGVVKRLIVCFRSPFP	
33	AP02124	GFWGKLWEGVKNAI	
34	AP02125	FWGKLWEGVKNAI	
35	AP02126	ILSAIWSGIKSLF	

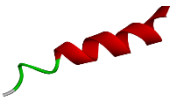
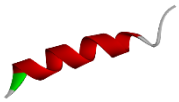
36	AP02127	IWSAIWSGIKGLL	
37	DRAMP18488	FFGSLLSLGSKLLPSVFKLFQRKKE	
38	AP02213	GILGKLWEGFKSIV	
39	AP02214	IFGAIWKGISSLL	
40	AP02215	FLSTIWNGIKSLL	
41	AP02217	FLSTLWNAAKSIF	
42	AP02218	IFKAIWSGIKSLF	
43	AP02334	ILGKIWEGIKSIF	
44	AP02335	ILSYLWNGIKSIF	
45	AP02356	FLFSLIPHAIGGLISAFK	
46	AP02357	FLFSLIPSAISGLISAF	
47	AP02358	FRFGSFLKKVWKSKLAKKLRSKGKQLLKDYANKVLNGPEEEAAAPAE	
48	AP02403	GILGKLWEGVKSIF	
49	AP02518	ARDGYIVDEKGCKFACFIN	
50	AP02545	KDGYIIEHRGCKYSCFFGTNSWCNTECTLKKGSSGYCAWPACWCYGLPDNVKIFDSNNLKC	
51	AP02546	KDGYLVGNDGCKYSCFTRPGTYCANECSRVKGKDGYCYAWMACYCYSMPNWVKTWNRATNRCGR	

52	AP02547	DNGYLLDKYTGCKVWCVINNESCNSECKIRGGYYGYCYFWKLACF CQGARKSELWNYNTNKCNGKL	
53	AP02549	GVWDWLKKTAKNVWNSDIVKQLKGKAINAAKNYVAEKIGATPS	
54	AP02550	FWGALAKGALKLIPSLVSSFTKKD	
55	AP02551	ILGEIWKGIKDIL	
56	AP02552	KWFNEKSIQNKIDEKIGKNFLGGMAKAVVHKLAKNEFMCVANVDM TKSCDTHCQKASGEKGYCHGTKCKCGVPLSY	
57	AP02560	FIGAVAGLLSKIF	
58	AP02561	GFREKHFQRFVKYAVPESTLRTVLQTVVHKVGKTQFGCPAYQGYCD DHCQDIEKKEGFCHGFKCKCGIPMGF	
59	AP02562	GWMSEKKVQGILDKKLPEGIIRNAAKAIVHKMAKNQFGCFANVDVK GDCKRHCKAEDKEGICHGTKCKCGVPISYL	
60	AP02563	KSTVGQKLKKKLNQAVDKVKEVLNKSEYMCPVVSSFCKQHCARLG KSGQCDLLECICS	
61	AP02603	LPFFLLSLIPSAISAIKKI	
62	AP02604	FWGFLGKLAMKAVPSLIGGNKSSSK	
63	DRAMP18443	IWSFLIKAATKLLPSLFGGGKKDS	
64	DRAMP18442	GVWDWIKKTAGKIWNSEPVKALKSQALNAAKNFVAEKIGATPS	
65	DRAMP18418	FIGMIPGLIGGLISAFK	
66	DRAMP18417	FFGTLFKLGSKLIPGVMKLFSKKKER	

67	DRAMP18416	FIGMIPGLIGGLISAIK	
68	DRAMP18415	FWSFLVKAASKILPSLIGGGDDNKSSS	
69	DRAMP18414	GILGKIWEGVKSLI	
70	DRAMP18401	ILSAIWSGIKGLL	
71	DRAMP18400	FLSTIWNGIKGLL	
72	DRAMP18399	FPFLLSLIPSAISAIKRL	
73	DRAMP18398	ISQSDAILSAIWSGIKSLF	
74	DRAMP18397	FFSALLSGIKSLF	
75	AP02972	FFGHLFKLATKIIPSFFRRKNQ	
76	AP02973	FLFSLIPSAISGLISAFKGRRKRDLN	
77	AP02974	FLFSLIPSAISGLINAFK	
78	AP02975	FFGHLFKLATKIIPSLFQRKKE	
79	AP02981	FKFGSFIKRMWRSKLAKKLRAKGKELLRDYANRVLSPEEEAAAPAPYPA	
80	AP03022	GFGCPFNQGQCHKHCQSIRRRGGYCDGFLKTRCVCYR	
81	AP03045	GWINEKKMQQKIDEKIGKNIIGGMAKAVIHKMAKNEFQCVANVDTLGNCKKHCAKTTGEKGYCHGTKCKCGIELSY	
82	AP03123	GGILREKYFHKAADALTSNIPIPVVKDVLKSAANQMIRKIGKVQQACAFNKDLAGWCEKSCQEAEGKKGYCHGTKCKCGKPIDY	

83	AP03124	AKKPFVQRVKNAASKAYNKLKGLAMQSQYGCPIISNMCEDHCRRK KMEGQCDLLDCVCS	
84	AP03125	DFPLSKEYETCVRPRKCQPPLKCNKAQICVDPKKGW	
85	AP03128	GFGGGRGGFGGGRGGFGGGGIGGGGFGGGYGGGKIKG	
86	AP03275	FLGSLFSIGSKLLPGVIKLFQRKKQ	
87	AP03276	FFFLPSLIGGLVSAIK	
88	AP03303	IFGKIWEGIKSLF	
89	AP03335	FFGAIAKLAMKFLPAIYKQIQKKRK	
90	AP03639	IFKAIWSGINRLF	
91	AP03723	FLGSLFSIGSKLLPGVFKLFSRKKQ	
92	AP03724	IFGMIPGLIGGLISAFK	
93	AP03725	FFSLIPSLIGGLVSAIK	
94	AP03939	GLINEKKVQQYLDEKLPNGVVKGALKSLVHKAAKNQNLCAFNVDT VGMCDADCKRQGKAKGVCHGTKCKCDVELSYKK	
95	AP03940	GLIDVKCYATSQCWAPCKKETGSGQSKCQNNQCRCY	
96	AP03941	FLSALWGVAKSLF	
97	AP03942	FPFLLSLIPSAISALKKL	
98	AP04851	KEGYLMDHEGCKLSCFIRPSGYCGRECGIKKGSSGYCAWPACYCYG LPNWVKVWDRATNKC	

99	AP04935	GIWGWIKKTAKKVWNSDIANQLKNKALNAAKNYVAEKVGATPAEAGQ	
100	AP04939	GFGCPFNQGRCHRHCRSIGRRGGYCRGIFKQTCACYRK	
101	AP04940	GFGCPLNQGACHRHCRSIRRRGGYCAGFFKQTCTCYRN	
102	DRAMP18680	GIADILKGLL	
103	DRAMP18681	LLGGLLQSLL	
104	DRAMP20775	FLGFLKNLF	
105	DRAMP18682	GILDAITGLL	
106	DRAMP18683	YIRDFITRRPPFGNI	
107	DRAMP18684	GILDALTGIL	
108	DRAMP18685	GVVDTLKNLLMGLL	
109	DRAMP18686	FLFNVIPHAINATASLIKK	
110	DRAMP18687	FLVGILPRMRGFITPFLKKVR	
111	DRAMP18688	FLWSLIPSAISAVTSLIKK	
112	DRAMP18689	FDLGGLIKGVVDLF	
113	DRAMP18695	KHFGKDSNFPF	

114	DRAMP18717	EFTNVSCTTSKECWSVCQRLHNTSRGKCMNKKCRCYS	
115	DRAMP21024	FFSLIPSLVGGLISAFK	
116	DRAMP21251	FLFSLIPSVIAGLVSAIRN	
117	DRAMP21252	FLFSLIPSAIAGLVSAIRN	

Appendix IV: Complete Numeric TLR4/MD-2 Docking Results for All 75 AMPs (Tabular Format)

Rank	DB ID	Species	Sequence	Length	Chg	Best dG	Mode 2	Mean dG
1	DRAMP18415	*Opisthacanthus cayaporum*	FWSFLVKAASKILPSLIGGGDDNKSSS	27	+1	**-17.9**	-17.7	-16.17
2	AP02027	*Scorpiops tibetanus*	GFWGSLWEGVKSVV	14	+1	**-16.9**	-15.8	-15.13
3	AP02356	*Androctonus amoreuxi*	FLFSLIPHAIGGLISAFK	18	+3	**-16.0**	-15.4	-14.49
4	DRAMP18416	*Tityus obscurus*	FIGMIPGLIGGLISAIK	17	+1	**-15.7**	-15.7	-13.88
5	AP02127	*Urodacus yaschenkoi*	IWSAIWSGIKGLL	13	+2	**-15.7**	-14.8	-14.32
6	DRAMP18443	*Scorpio maurus palmatus*	IWSFLIKAATKLLPSLFGGGKKDS	24	+3	**-15.6**	-15.6	-14.76
7	AP00153	*Androctonus sp.*	RSVCRQIKICRRRGGCYYKCTNRPY	25	+8	**-15.0**	-15.0	-14.75
8	DRAMP18686	*Chaerilus tricarinatus*	FLFNVIPHAINATASLIKK	19	+3	**-15.0**	-14.6	-13.28
9	AP01791	*Chaerilus tricostatus*	FLWGLIPGAISAVTSLIKK	19	+3	**-15.0**	-14.2	-13.38
10	AP02125	*Urodacus yaschenkoi*	FWGKLWEGVKNAI	13	+2	**-14.9**	-13.4	-13.06
11	AP02026	*Scorpiops tibetanus*	GFWGKLWEGVKSAI	14	+2	**-14.9**	-14.8	-13.25
12	AP03275	*Isometrus maculatus*	FLGSLFSIGSKLLPGVIKLFQRKKQ	25	+5	**-14.8**	-13.7	-12.83
13	AP02560	*Mesobuthus martensii*	FIGAVAGLLSKIF	13	+1	**-14.6**	-14.5	-13.88
14	AP02603	*Vaejovis punctatus*	LPFFLLSLIPSAISAIKKI	19	+3	**-14.6**	-14.5	-13.58
15	AP02215	*Pandinus imperator*	FLSTIWNGIKSLL	13	+1	**-14.6**	-14.1	-12.87
16	AP02604	*Vaejovis punctatus*	FWGFLGKLAMKAVPSLIGGNKSSSK	25	+4	**-14.6**	-14.5	-13.75
17	AP02550	*Heterometrus laoticus*	FWGALAKGALKLIPSLVSSFTKKD	24	+4	**-14.3**	-14.2	-13.88
18	AP03723	*Tityus tarnensis*	FLGSLFSIGSKLLPGVFKLFSRKKQ	25	+6	**-14.2**	-13.0	-12.52
19	DRAMP18418	*Tityus obscurus*	FIGMIPGLIGGLISAFK	17	+1	**-14.0**	-13.9	-13.14
20	AP02403	*Heterometrus petersii*	GILGKLWEGVKSIF	14	+1	**-13.9**	-13.2	-12.48
21	AP02124	*Urodacus yaschenkoi*	GFWGKLWEGVKNAI	14	+2	**-13.9**	-12.8	-12.54
22	DRAMP21252	*Androctonus aeneas*	FLFSLIPSAIAGLVSAIRN	19	+1	**-13.7**	-13.5	-12.58
23	AP02357	*Androctonus amoreuxi*	FLFSLIPSAISGLISAF	17	+1	**-13.7**	-13.6	-12.67
24	DRAMP18417	*Tityus obscurus*	FFGTLFKLGSKLIPGVMKLFSKKKER	26	+6	**-13.6**	-13.5	-13.12
25	DRAMP18688	*Chaerilus tricarinatus*	FLWSLIPSAISAVTSLIKK	19	+2	**-13.5**	-13.3	-12.68
26	AP02217	*Vaejovis mexicanus smithi*	FLSTLWNAAKSIF	13	+2	**-13.5**	-12.6	-12.01
27	DRAMP21251	*Androctonus aeneas*	FLFSLIPSVIAGLVSAIRN	19	+1	**-13.4**	-12.9	-12.24
28	DRAMP18401	*Urodacus yaschenkoi*	ILSAIWSGIKGLL	13	+1	**-13.4**	-13.4	-12.67
29	AP01977	*Mesobuthus martensii*	FLFSLIPSAISGLISAFK	18	+2	**-13.4**	-12.9	-11.73

30	AP02974	*Mesobuthus eupeus*	FLFSLIPSAISGLINAFK	18	+2	**-13.3**	-13.1	-12.63
31	DRAMP18687	*Chaerilus tricarinatus*	FLVGILPRMRGFITPFLKKVR	21	+5	**-13.2**	-13.2	-12.50
32	AP03724	*Tityus tarnensis*	IFGMIPGLIGGLISAFK	17	+2	**-13.2**	-13.1	-11.91
33	AP03335	*Chelifer cancroides*	FFGAIAKLAMKFLPAIYKQIQKKRK	25	+8	**-13.1**	-12.6	-12.18
34	DRAMP18414	*Opisthacanthus cayaporum*	GILGKIWEGVKSLI	14	+1	**-13.0**	-12.8	-12.30
35	AP02973	*Mesobuthus eupeus*	FLFSLIPSAISGLISAFKGRRKRDLN	26	+4	-12.9	-12.7	-12.06
36	AP02518	*Centruroides margaritatus*	ARDGYIVDEKGCKFACFIN	19	0	-12.9	-12.6	-11.90
37	AP02214	*Pandinus imperator*	IFGAIWKGISSLL	13	+1	-12.8	-12.7	-11.83
38	AP01980	*Opisthacanthus madagascariensis*	IFGAIWNGIKSLF	13	+2	-12.6	-12.5	-11.89
39	AP02126	*Urodacus yaschenkoi*	ILSAIWSGIKSLF	13	+2	-12.6	-12.6	-12.01
40	AP02975	*Mesobuthus eupeus*	FFGHLFKLATKIIPSLFQRKKE	22	+5	-12.6	-12.3	-11.60
41	AP02218	*Heterometrus petersii*	IFKAIWSGIKSLF	13	+2	-12.4	-12.4	-11.91
42	AP02213	*Pandinus imperator*	GILGKLWEGFKSIV	14	+1	-12.2	-12.0	-11.50
43	AP02972	*Mesobuthus eupeus*	FFGHLFKLATKIIPSFFRRKNQ	22	+6	-12.2	-12.1	-11.57
44	AP02098	*Heterometrus spinifer*	SGTSEKERESGRLLGVVKRLIVCFR SPFP	29	+3	-12.1	-11.1	-10.61
45	AP02551	*Heterometrus spinifer*	ILGEIWKGIKDIL	13	+1	-11.8	-11.5	-11.06
46	DRAMP03688	*Tityus serrulatus*	FLGMIPGLIGGLISAFK	17	+1	-11.4	-10.7	-9.60
47	AP01545	*Mesobuthus eupeus*	FFGHLFKLATKIIPSLFQ	18	+2	-11.2	-11.1	-10.77
48	AP01512	*Isometrus maculatus*	FFSLLPSLIGGLVSAIK	17	+2	-11.1	-11.0	-10.60
49	AP02216	*Vaejovis mexicanus smithi*	FLGALWNVAKSVF	13	+2	-10.8	-10.8	-10.24
50	AP01978	*Buthus martensii*	FIGAIARLLSKIF	13	+3	-9.7	-9.7	-7.99
51	DRAMP18399	*Urodacus yaschenkoi*	FPFLLSLIPSAISAIKRL	18	+2	-9.1	-9.0	-8.86
52	DRAMP20775	*Chaerilus tryznai*	FLGFLKNLF	9	+1	-9.1	-8.7	-8.50
53	AP00430	*Opisthacanthus madagascariensis*	ILGKIWEGIKSLF	13	+2	-9.0	-8.7	-8.49
54	AP03941	*Liocheles australasiae*	FLSALWGVAKSLF	13	+2	-9.0	-8.9	-8.70
55	AP00660	*Pandinus imperator*	FWGALAKGALKLIPSLFSSFSKKD	24	+3	-9.0	-8.7	-8.49
56	AP03639	*Heterometrus petersii*	IFKAIWSGINRLF	13	+2	-8.9	-8.6	-8.41
57	AP01327	*Androctonus australis*	LFGLIPSLIGGLVSAFK	17	+1	-8.8	-8.7	-7.52
58	AP01544	*Mesobuthus eupeus*	IFGAIAGLLKNIF	13	+2	-8.8	-8.7	-8.49
59	AP03276	*Isometrus maculatus*	FFFLPSLIGGLVSAIK	16	+2	-8.7	-8.2	-7.91
60	DRAMP18683	*Chaerilus tricostatus*	YIRDFITRRPPFGNI	15	+2	-8.6	-8.5	-8.38
61	DRAMP18397	*Urodacus manicatus*	FFSALLSGIKSLF	13	+1	-8.5	-8.2	-7.96
62	DRAMP18400	*Urodacus yaschenkoi*	FLSTIWNGIKGLL	13	+1	-8.4	-8.4	-8.28
63	DRAMP18695	*Chaerilus tricarinatus*	KHFGKDSNFPF	11	+1	-8.4	-8.4	-8.30
64	DRAMP03687	*Tityus serrulatus*	FLSLIPSLVGGSISAFK	17	+1	-8.3	-7.4	-7.00

65	AP03725	*Tityus tarnensis*	FFSLIPSLIGGLVSAIK	17	+2	-8.1	-7.9	-7.74
66	DRAMP18681	*Chaerilus tryznai*	LLGGLLQSLL	10	0	-7.8	-7.6	-7.43
67	AP03942	*Liocheles australasiae*	FPFLLSLIPSAISALKKL	18	+3	-7.8	-7.7	-7.54
68	DRAMP21024	*Tityus stigmurus*	FFSLIPSLVGGLISAFK	17	+1	-7.8	-7.6	-7.53
69	DRAMP18398	*Urodacus manicatus*	ISQSDAILSAIWSGIKSLF	19	0	-7.7	-7.6	-7.54
70	AP03303	*Opisthacanthus madagascariensis*	IFGKIWEGIKSLF	13	+2	-7.6	-7.4	-7.34
71	DRAMP18689	*Chaerilus tricarinatus*	FDLGGLIKGVVDLF	14	-1	-7.5	-7.3	-7.14
72	DRAMP18684	*Chaerilus tricostatus*	GILDALTGIL	10	-1	-7.2	-7.2	-7.04
73	DRAMP18680	*Chaerilus tryznai*	GIADILKGLL	10	0	-7.2	-7.2	-7.14
74	DRAMP18682	*Chaerilus tryznai*	GILDAITGLL	10	-1	-7.2	-7.1	-7.06
75	DRAMP18685	*Chaerilus tricostatus*	GVVDTLKNLLMGLL	14	0	-7.0	-6.8	-6.48

Appendix V: Complete In Silico Safety Screening Dataset for All 75 Docked AMPs

S.No	DB ID	Species	Sequence	Length	Chg	AllerCatPro	ToxinPred	HLPpred	Status
1	AP01512	*Isometrus maculatus*	FFSLLPSLIGGLVSAIK	17	+2	0	Non-Toxin	Active	EXCL (Hemolysis)
2	AP01978	*Buthus martensii*	FIGAIARLLSKIF	13	+3	0	Toxin	Active	EXCL (Cytotox)
3	AP02216	*Vaejovis mexicanus smithi*	FLGALWNVAKSVF	13	+2	0	Toxin	Active	EXCL (Cytotox)
4	DRAMP03688	*Tityus serrulatus*	FLGMIPGLIGGLISAFK	17	+1	0	Toxin	Active	EXCL (Cytotox)
5	DRAMP03687	*Tityus serrulatus*	FLSLIPSLVGGSISAFK	17	+1	0	Toxin	Not Active	EXCL (Cytotox)
6	**AP00153**	***Androctonus sp.***	**RSVCRQIKICRRRGGCYYKCTNRPY**	**25**	**+8**	**0**	**Non-Toxin**	**Not Active**	**SELECTED**
7	AP00430	*Opisthacanthus madagascariensis*	ILGKIWEGIKSLF	13	+2	0	Toxin	Active	EXCL (Cytotox)
8	AP00660	*Pandinus imperator*	FWGALAKGALKLIPSLFSSFSKKD	24	+3	0	Toxin	Active	EXCL (Cytotox)
9	AP01327	*Androctonus australis*	LFGLIPSLIGGLVSAFK	17	+1	0	Toxin	Active	EXCL (Cytotox)
10	AP01544	*Mesobuthus eupeus*	IFGAIAGLLKNIF	13	+2	0	Toxin	Active	EXCL (Cytotox)
11	AP01545	*Mesobuthus eupeus*	FFGHLFKLATKIIPSLFQ	18	+2	0	Toxin	Active	EXCL (Cytotox)
12	AP01791	*Chaerilus tricostatus*	FLWGLIPGAISAVTSLIKK	19	+3	0	Toxin	Active	EXCL (Cytotox)
13	AP01977	*Mesobuthus martensii*	FLFSLIPSAISGLISAFK	18	+2	0	Toxin	Active	EXCL (Cytotox)
14	AP01980	*Opisthacanthus madagascariensis*	IFGAIWNGIKSLF	13	+2	0	Toxin	Active	EXCL (Cytotox)
15	AP02026	*Scorpiops tibetanus*	GFWGKLWEGVKSAI	14	+2	0	Toxin	Not Active	EXCL (Cytotox)
16	AP02027	*Scorpiops tibetanus*	GFWGSLWEGVKSVV	14	+1	0	Toxin	Active	EXCL (Cytotox)
17	AP02098	*Heterometrus spinifer*	SGTSEKERESGRLLGVVKRLIVCFRSPFP	29	+3	0	Non-Toxin	Active	EXCL (Hemolysis)
18	AP02124	*Urodacus yaschenkoi*	GFWGKLWEGVKNAI	14	+2	0	Toxin	Active	EXCL (Cytotox)
19	AP02125	*Urodacus yaschenkoi*	FWGKLWEGVKNAI	13	+2	0	Toxin	Not Active	EXCL (Cytotox)
20	AP02126	*Urodacus yaschenkoi*	ILSAIWSGIKSLF	13	+2	0	Toxin	Active	EXCL (Cytotox)
21	AP02127	*Urodacus yaschenkoi*	IWSAIWSGIKGLL	13	+2	0	Toxin	Active	EXCL (Cytotox)

22	AP02213	*Pandinus imperator*	GILGKLWEGFKSIV	14	+1	0	Toxin	Active	EXCL (Cytotox)
23	AP02214	*Pandinus imperator*	IFGAIWKGISSLL	13	+1	0	Toxin	Active	EXCL (Cytotox)
24	AP02215	*Pandinus imperator*	FLSTIWNGIKSLL	13	+1	0	Toxin	Active	EXCL (Cytotox)
25	AP02217	*Vaejovis mexicanus smithi*	FLSTLWNAAKSIF	13	+2	0	Toxin	Active	EXCL (Cytotox)
26	AP02218	*Heterometrus petersii*	IFKAIWSGIKSLF	13	+2	0	Toxin	Active	EXCL (Cytotox)
27	AP02356	*Androctonus amoreuxi*	FLFSLIPHAIGGLISAFK	18	+3	0	Toxin	Active	EXCL (Cytotox)
28	**AP02357**	***Androctonus amoreuxi***	**FLFSLIPSAISGLISAF**	**17**	**+1**	**0**	**Non-Toxin**	**Not Active**	**SELECTED**
29	AP02403	*Heterometrus petersii*	GILGKLWEGVKSIF	14	+1	0	Toxin	Active	EXCL (Cytotox)
30	AP02518	*Centruroides margaritatus*	ARDGYIVDEKGCKFACFIN	19	0	0	Toxin	Not Active	EXCL (Cytotox)
31	AP02550	*Heterometrus laoticus*	FWGALAKGALKLIPSLVSSFTKKD	24	+4	0	Toxin	Active	EXCL (Cytotox)
32	AP02551	*Heterometrus spinifer*	ILGEIWKGIKDIL	13	+1	0	Toxin	Active	EXCL (Cytotox)
33	**AP02560**	***Mesobuthus martensii***	**FIGAVAGLLSKIF**	**13**	**+1**	**0**	**Non-Toxin**	**Not Active**	**SELECTED**
34	AP02603	*Vaejovis punctatus*	LPFFLLSLIPSAISAIKKI	19	+3	0	Non-Toxin	Active	EXCL (Hemolysis)
35	**AP02604**	***Vaejovis punctatus***	**FWGFLGKLAMKAVPSLIGGNKSSSK**	**25**	**+4**	**0**	**Non-Toxin**	**Not Active**	**SELECTED**
36	DRAMP18443	*Scorpio maurus palmatus*	IWSFLIKAATKLLPSLFGGGKKDS	24	+3	0	Toxin	Active	EXCL (Cytotox)
37	DRAMP18418	*Tityus obscurus*	FIGMIPGLIGGLISAFK	17	+1	0	Toxin	Active	EXCL (Cytotox)
38	DRAMP18417	*Tityus obscurus*	FFGTLFKLGSKLIPGVMKLFSKKKER	26	+6	0	Toxin	Active	EXCL (Cytotox)
39	DRAMP18416	*Tityus obscurus*	FIGMIPGLIGGLISAIK	17	+1	0	Toxin	Active	EXCL (Cytotox)
40	DRAMP18415	*Opisthacanthus cayaporum*	FWSFLVKAASKILPSLIGGGDDNKSSS	27	+1	0	Non-Toxin	Active	EXCL (Hemolysis)
41	DRAMP18414	*Opisthacanthus cayaporum*	GILGKIWEGVKSLI	14	+1	0	Toxin	Active	EXCL (Cytotox)
42	DRAMP18401	*Urodacus yaschenkoi*	ILSAIWSGIKGLL	13	+1	0	Toxin	Active	EXCL (Cytotox)
43	DRAMP18400	*Urodacus yaschenkoi*	FLSTIWNGIKGLL	13	+1	0	Toxin	Active	EXCL (Cytotox)

44	DRAMP18399	*Urodacus yaschenkoi*	FPFLLSLIPSAISAIKRL	18	+2	0	Non-Toxin	Active	EXCL (Hemolysis)
45	**DRAMP18398**	***Urodacus manicatus***	**ISQSDAILSAIWSGIKSLF**	**19**	**0**	**0**	**Non-Toxin**	**Not Active**	**SELECTED**
46	**DRAMP18397**	***Urodacus manicatus***	**FFSALLSGIKSLF**	**13**	**+1**	**0**	**Non-Toxin**	**Not Active**	**SELECTED**
47	**AP02972**	***Mesobuthus eupeus***	**FFGHLFKLATKIIPSFFRRKNQ**	**22**	**+6**	**0**	**Non-Toxin**	**Not Active**	**SELECTED**
48	AP02973	*Mesobuthus eupeus*	FLFSLIPSAISGLISAFKGRRKRDLN	26	+4	0	Non-Toxin	Active	EXCL (Hemolysis)
49	AP02974	*Mesobuthus eupeus*	FLFSLIPSAISGLINAFK	18	+2	0	Toxin	Active	EXCL (Cytotox)
50	AP02975	*Mesobuthus eupeus*	FFGHLFKLATKIIPSLFQRKKE	22	+5	0	Toxin	Not Active	EXCL (Cytotox)
51	AP03275	*Isometrus maculatus*	FLGSLFSIGSKLLPGVIKLFQRKKQ	25	+5	0	Toxin	Active	EXCL (Cytotox)
52	AP03276	*Isometrus maculatus*	FFFLPSLIGGLVSAIK	16	+2	0	Toxin	Active	EXCL (Cytotox)
53	AP03303	*Opisthacanthus madagascariensis*	IFGKIWEGIKSLF	13	+2	0	Toxin	Active	EXCL (Cytotox)
54	AP03335	*Chelifer cancroides*	FFGAIAKLAMKFLPAIYKQIQKKRK	25	+8	0	Toxin	Not Active	EXCL (Cytotox)
55	AP03639	*Heterometrus petersii*	IFKAIWSGINRLF	13	+2	0	Non-Toxin	Active	EXCL (Hemolysis)
56	AP03723	*Tityus tarnensis*	FLGSLFSIGSKLLPGVFKLFSRKKQ	25	+6	0	Toxin	Active	EXCL (Cytotox)
57	AP03724	*Tityus tarnensis*	IFGMIPGLIGGLISAFK	17	+2	0	Toxin	Active	EXCL (Cytotox)
58	AP03725	*Tityus tarnensis*	FFSLIPSLIGGLVSAIK	17	+2	0	Non-Toxin	Active	EXCL (Hemolysis)
59	AP03941	*Liocheles australasiae*	FLSALWGVAKSLF	13	+2	0	Non-Toxin	Active	EXCL (Hemolysis)
60	AP03942	*Liocheles australasiae*	FPFLLSLIPSAISALKKL	18	+3	0	Non-Toxin	Active	EXCL (Hemolysis)
61	DRAMP18680	*Chaerilus tryznai*	GIADILKGLL	10	0	0	Non-Toxin	Active	EXCL (Hemolysis)
62	**DRAMP18681**	***Chaerilus tryznai***	**LLGGLLQSLL**	**10**	**0**	**0**	**Non-Toxin**	**Not Active**	**SELECTED**
63	DRAMP20775	*Chaerilus tryznai*	FLGFLKNLF	9	+1	0	Toxin	Not Active	EXCL (Cytotox)
64	**DRAMP18682**	***Chaerilus tryznai***	**GILDAITGLL**	**10**	**-1**	**0**	**Non-Toxin**	**Not Active**	**SELECTED**
65	**DRAMP18683**	***Chaerilus tricostatus***	**YIRDFITRRPPFGNI**	**15**	**+2**	**0**	**Non-Toxin**	**Not Active**	**SELECTED**

66	**DRAMP18684**	***Chaerilus tricostatus***	**GILDALTGIL**	**10**	**-1**	**0**	**Non-Toxin**	**Not Active**	**SELECTED**
67	**DRAMP18685**	***Chaerilus tricostatus***	**GVVDTLKNLLMGLL**	**14**	**0**	**0**	**Non-Toxin**	**Not Active**	**SELECTED**
68	DRAMP18686	*Chaerilus tricarinatus*	FLFNVIPHAINATASLIKK	19	+3	0	Non-Toxin	Active	EXCL (Hemolysis)
69	DRAMP18687	*Chaerilus tricarinatus*	FLVGILPRMRGFITPFLKKVR	21	+5	0	Toxin	Active	EXCL (Cytotox)
70	DRAMP18688	*Chaerilus tricarinatus*	FLWSLIPSAISAVTSLIKK	19	+2	0	Non-Toxin	Active	EXCL (Hemolysis)
71	DRAMP18689	*Chaerilus tricarinatus*	FDLGGLIKGVVDLF	14	-1	0	Toxin	Not Active	EXCL (Cytotox)
72	DRAMP18695	*Chaerilus tricarinatus*	KHFGKDSNFPF	11	+1	0	Toxin	Not Active	EXCL (Cytotox)
73	DRAMP21024	*Tityus stigmurus*	FFSLIPSLVGGLISAFK	17	+1	0	Toxin	Active	EXCL (Cytotox)
74	DRAMP21251	*Androctonus aeneas*	FLFSLIPSVIAGLVSAIRN	19	+1	0	Toxin	Active	EXCL (Cytotox)
75	DRAMP21252	*Androctonus aeneas*	FLFSLIPSAIAGLVSAIRN	19	+1	0	Toxin	Active	EXCL (Cytotox)

Appendix VI: Protease Susceptibility and Stability Profile of All 21 AP00153-Derived ProteinMPNN Mutants

Seq ID	Sequence	Len	Conf	Mut	Trypsin	Chymo-H	Chymo-L	Pepsin	ProtK	II	Status
seq_1	GSRLRAVLVCSATGGCRYVLTLLPE	25	0.288	16	3	1	5	5	14	80.35	**Unstable**
seq_2	GSPLQKVLVCSETEGCRYVLALLPE	25	0.303	19	2	1	5	5	14	74.20	**Unstable**
seq_3	GSDLQPVLVCSETLGCRYVLQKLPE	25	0.278	19	2	1	5	6	12	69.50	**Unstable**
seq_4	GSLLQKVLVCSETLGCRYVLALLPE	25	0.294	19	2	1	5	5	14	71.80	**Unstable**
seq_5	SSKLKKVKVCSETKGCSYVLALLDS	25	0.285	19	2	1	7	9	15	77.60	**Unstable**
seq_6	GSRLRKVLVCSETEGCRYVLALLDE	25	0.297	19	2	1	5	5	14	68.90	**Unstable**
seq_7	SSLLQKVLVCSETEGCRYVLALLDS	25	0.304	20	2	1	5	5	14	72.40	**Unstable**
seq_8	GSLKQLVLVCGETTGCRYVELLPE	25	0.288	18	2	1	5	5	14	73.10	**Unstable**
seq_9	GSKLQKVLVCSETKGCRYVLQLLPE	25	0.298	19	2	1	5	5	14	70.90	**Unstable**
seq_10	GSILQKVKVCSETEGCRYVLQLLPE	25	0.283	18	2	1	5	5	14	75.50	**Unstable**
seq_11	GSLLQKVLVCSETEGCRYVLALLPS	25	0.299	19	2	1	5	5	14	67.80	**Unstable**
seq_12	GSLLQKVLVCSETGGCRYVLALLPE	25	0.304	18	2	1	5	5	14	71.20	**Unstable**
seq_13	GSRLQPVLVCSETLGCRFVLTLLPE	25	0.281	19	2	1	5	6	12	69.80	**Unstable**
seq_14	GSLLQLVKVCSETEGCRYVPQLLPE	25	0.299	18	2	1	5	5	14	72.60	**Unstable**
seq_15	SSKLKKVKVCSKTKGCSYVLQLLPE	25	0.301	18	2	1	5	5	15	76.40	**Unstable**
seq_16	GSPLRPVRVCSETEGCRDVLTLLPE	25	0.293	18	2	1	5	5	14	70.10	**Unstable**
seq_17	GSSLRPVLVCSETLGCRYVLQLLPE	25	0.281	18	2	1	5	6	15	69.20	**Unstable**
seq_18	SSLLQKVLVCSETKGCSYKLTLLDK	25	0.280	18	2	1	5	5	14	73.80	**Unstable**
seq_19	SSIKQKVLVCSETEGCRYVLQVLDS	25	0.290	20	2	1	5	5	14	71.50	**Unstable**

seq_20	SSPLQPVLVCSETEGCRYVLALLPS	25	0.307	19	2	1	5	5	14	66.90	**Unstable**
seq_21	GSPLRAVRVTSETLGSRDVLALLPE	25	0.296	21	2	0	4	6	13	61.55	**Unstable**

Appendix VII: Protease Susceptibility and Stability Profile of All 20 AP02604-Derived ProteinMPNN Mutants

Seq ID	Sequence	Len	Conf	Mut	Trypsin	Chymo-H	Chymo-L	Pepsin	ProtK	II	Status
seq_22	EEEKEKLEKLREELLEKGGYELPEV	25	0.242	23	5	1	5	6	17	42.30	Unstable
seq_23	REEERRRLEAREALREEGGYELPEV	25	0.226	22	5	1	3	2	16	38.90	Borderline
seq_24	EEEELRRERERLRLLAEGGYELPEV	25	0.223	22	5	1	5	5	17	44.10	Unstable
seq_25	EEERLREELQRLELLKEGGYELPEV	25	0.201	22	4	1	6	6	14	39.50	Borderline
seq_26	EEEKKKEEKEREEKLKEGGYKLEKV	25	0.248	22	5	1	3	2	16	36.20	**Stable***
seq_27	EEEERRLEEERERLLEEGGYELPVV	25	0.237	23	5	1	3	2	14	45.80	Unstable
seq_28	KEAELEKLLQRLKRLAEGGYELPEV	25	0.191	20	4	1	4	3	14	41.60	Unstable
seq_29	EEKKKEEEKKREELKKEGGYELEKE	25	0.273	23	6	1	3	2	16	37.10	Borderline
seq_30	KEKELEELKKREKLKEEGGYELEKV	25	0.218	21	5	1	3	2	16	38.40	Borderline
seq_31	KEEEERKRREREERLEKGGYELEEE	25	0.254	22	5	1	4	3	15	43.70	Unstable
seq_32	EEERLRLEEERLRRLEEGGYELEEV	25	0.240	22	5	1	4	3	15	46.20	Unstable
seq_33	EEARRLEELARLAGLAEGRYELPEV	25	0.196	24	4	1	4	4	16	41.80	Unstable
seq_34	KEEERREEELREVLRELGGYELPEV	25	0.214	22	4	1	3	2	15	40.90	Unstable
seq_35	AQAALQQLLERLRRLAEGGTELPVT	25	0.194	21	2	1	3	2	14	39.70	Borderline
seq_36	KEEELRQKKERLERLKEGGYELPET	25	0.204	22	4	1	4	3	15	42.50	Unstable
seq_37	KEEELKKKIEREKLLKEGGYELPET	25	0.218	21	4	1	4	3	14	38.80	Borderline
seq_38	EEEKEKEKELREKLRKEGGYELPVV	25	0.239	23	5	1	4	3	16	44.30	Unstable
seq_39	EEEELERERLRELLREKGGYELPEV	25	0.213	22	5	1	4	3	16	43.10	Unstable
seq_40	EEAEREKERLFIEELREKGGYELPEV	25	0.232	22	5	1	4	3	16	40.20	Unstable
seq_41	EEEKKRLERLREELLEKGGYELPEV	25	0.230	23	4	1	4	3	16	43.90	Unstable

Appendix VIII: Complete Protease Susceptibility and Conformational Stability Profile of All 20 DRAMP18681-Derived ProteinMPNN Mutants

Seq ID	Sequence	Len	Conf	Mut	Chymo-H	Chymo-L	Pepsin pH 1.3	Pepsin pH >2	ProtK	Trypsin	Thrombin	II	Status
Original	**LLGGLLQSLL**	**10**	**--**	**--**	**0**	**6**	**8**	**8**	**6**	**0**	**0**	**97.88**	**Unstable (Parent)**
seq_42*	DRGGVRDPRV	10	0.343	8	0	0	0	0	0	0	0	-2.23	**STABLE FINAL LEAD**
seq_43	DLGGVIDPRI	10	0.290	7	0	0	0	0	0	0	0	13.80	Stable
seq_44	DRGGVIDPAI	10	0.281	8	0	0	0	0	4	1	0	24.57	Stable
seq_45*	DRGGVRDPRI	10	0.330	8	0	0	0	0	0	0	0	-2.23	**STABLE FINAL LEAD**
seq_46	DRGGVYDPRV	10	0.284	8	0	0	0	0	0	0	0	13.91	Stable
seq_47*	DRGGVIDPRI	10	0.299	8	0	0	0	0	0	0	0	-2.23	**STABLE FINAL LEAD**
seq_48	DRGGVRDPRV	10	0.316	8	1	1	0	1	3	2	0	-2.23	Unstable
seq_49	DRGGVRDPRI	10	0.321	8	0	0	0	0	3	2	0	41.33	Unstable
seq_50	DRGGVRDPRI	10	0.327	8	0	0	0	0	3	2	0	42.34	Unstable
seq_51	DLGGVRDRRY	10	0.294	7	0	0	0	0	0	0	0	63.54	Unstable
seq_52	DRGGRRDPRI	10	0.313	8	0	0	0	0	0	0	0	55.05	Unstable
seq_53	DRGGVRDPRV	10	0.326	8	0	0	0	0	1	4	0	-2.230	Unstable
seq_54	DRGGVRDPRV	10	0.325	8	0	0	0	0	3	2	0	-2.230	Unstable
seq_55	DLGGVRDPRV	10	0.327	7	0	0	0	0	0	0	0	13.80	Stable
seq_56*	DKGGVIDPRV	10	0.294	8	0	0	0	0	0	0	0	-3.18	**STABLE FINAL LEAD**
seq_57*	DKGGVRDPRV	10	0.317	8	0	0	0	0	0	0	0	-3.18	**STABLE FINAL LEAD**
seq_58	DRGGVRNREV	10	0.277	8	0	0	0	0	0	0	0	17.65	Stable
seq_59	DLGGVEDPAI	10	0.253	7	0	0	0	0	0	0	0	59.86	Unstable
seq_60	DLGGVIDPRI	10	0.300	7	0	1	2	2	5	0	0	13.80	Unstable
seq_61	DRGGVRDPRV	10	0.332	8	0	0	0	0	2	3	0	-2.23	Unstable