Identification and Molecular Validation of Biomarkers for the Accurate and Sensitive Diagnosis of Bacterial and Viral Pneumonia

UNIVERSITY *of the*
WESTERN CAPE

A thesis submitted in fulfillment of the requirement for the degree of

Philosophiae Doctor

In the Department of Biotechnology

Faculty of Natural Sciences, University of the Western Cape

By

Olalekan Olanrewaju Bakare

Supervisor: Dr. Ashley Pretorius

Co supervisor: Dr. Marshall Keyster

November, 2019

Declaration

I declare that this thesis is my original research work. I have written herewith the PhD Thesis, '*The Identification and Molecular Validation of Biomarkers for the Accurate and Sensitive Diagnosis of Bacterial and Viral Pneumonia'* completely by myself and have documented all sources and material used. This thesis has not been previously presented for another examination board in any other university and has not been published.

O. O. Bakare October, 2019

Signature

Acknowledgement

Glory be to Almighty Allah, the most beneficent and most merciful, for His unlimited source of inspiration and protection throughout the journey of this entire research work. It is also due to his benevolent I have a life worthy of living.

Special appreciation and recognition goes to my supervisor, Dr. A. Pretorius who is the Principle Investigator of the Bioinformatics Research Group (BRG) on Disease Diagnostics at the University of the Western Cape, South Africa for his love, care, support, constructive correction, impactful contribution, motivation and guidance. This is laudable because it is the reason this research work meets the standard requirement. I must concede that I have benefitted immensely from his wealth of experience. Words alone cannot express my gratitude as your path in my life will remain indelible.

In addition, my appreciation goes to the entire Bioinformatics Research Group (BRG) for your encouragement, training and friendly mannerism throughout this program. I must confess that it is with your support that acclimatization on arrival was smooth and loneliness was eliminated. The overall impact of this is that you indirectly enhanced my research work by facilitating an enabling environment for its accomplishment. For future reference, it is imperative to mention names of my colleagues in this group, in persons of Marius Tincho Belmundo, Adewale Oluwaseun Fadaka, Chipampe Patricia Lombe, Gadijah Abdullah, Mohammad Hassan, Sarah Latib, and Naser Eoshinebo. Adewale Oluwaseun Fadaka, particularly, has displayed a behavior worthy of being termed “ä brother from another mother”. Overall, you are all very wonderful.

This page will be incomplete if I fail to recognize the effort of my parents Mr and Late Mrs S. Bakare for their prayers, love, encouragement, support and concern for my total well-being. I am eternally indebted in thanks. A special appreciation goes to my wife,

Hafsoh Flora Bakare, for her patience and unrelenting prayers and also to my brothers and sister, Monsoor, Rosheedah, Abeeb, and Hammed.

I am limited by the length of this page to express the depth of my gratitude to Jumoke Adebisi Aboyewa, Riziki Darius, Dr Nicole Sibuyi, Phumusile Dube, Dr Wusu-Dorcas Adedoja and all who made my stay in the laboratory at third floor, Biotechnology Department, Life Science Building, fulfilling. I want to say my words cannot be enough to say “thank you”. I also want to acknowledge the entire staff of the Department of Biotechnology, University of the Western Cape for their assistance one way or the other towards the accomplishment of this work.

My profound appreciation goes to National Research Foundation (NRF) and South African National Zakat Foundation for their financial assistance and support towards completion of this research work.

Dedication

This research project work is dedicated to The Almighty Allah.

Also, to my parents, Mr and late Mrs Bakare Salami.

Abstract

Identification of Molecular Validation Biomarkers for the Accurate and Sensitive Diagnosis of Bacterial and Viral Pneumonia

O. O. Bakare

PhD. thesis, Department of Biotechnology, Faculty of Natural Sciences, University of the Western Cape

Background

Pneumonia remains the major cause of death in children and the elderly and several efforts have been intensified to reduce the rate of pneumonia infection. The major breakthrough has been the discovery of certain biomarkers for the diagnosis of pneumonia through immunogenic techniques. However, the use of these biomarkers has been associated with low specificity and sensitivity as they are implicated in other diseases such as atherosclerotic plaques; also the biomarkers cannot differentiate between bacterial and viral types of the pneumonia pathogens and the use of these biomarkers often constitutes a problem by resulting in false positive or false negative results and are time consuming. The need for a sensitive and specific diagnosis has brought about the search for new biomarkers to lower mortality rates. Antimicrobial Peptides (AMPs) are promising diagnostic candidates due to their success in the detection of HIV.

The aim of this research was to identify AMPs, using *in silico* technologies, which would bind bacterial and viral pneumonia pathogen receptors, to detect the following causative agents: *Streptococcus pneumoniae, Acinetobacter baumannii, Klebsiella pneumoniae, Respiratory Syncytial Virus, Influenza A and B.*

Methods

Experimentally validated AMPs were identified for bacterial and viral pneumonia from databases such as CAMP, APD BACTIBASE and AVPDB. Subsequently, *in silico* mathematical algorithms, such as profile Hidden Markov Models (HMMER) and CD-HIT were used to identify additional potentially novel anti-pneumonia AMPs. The identified AMPs 3-D structure prediction was carried out using I-TASSER and the modelled AMPs were docked against various pneumonia protein receptors using PATCHDOCK with Pymol and Rasmol used as visualization tools respectively. Subsequently, derivative AMPs were produced from the parental peptides, using site directed mutagenesis (SDM), which bound the protein receptors of bacterial and viral pneumonia pathogens with increased binding affinity.

Results

The results showed that 45 putative anti-pneumonia AMPs were identified with 27 anti-viral AMPs against *Respiratory Syncytial Virus, Influenza A* and *B* while the 18 AMPs identified were anti-bacterial against *Streptococcus pneumoniae*, *Acinetobacter baumannii* and *Klebsiella pneumoniae*. Using the parental AMPs, 33 derivative AMPs were produced following SDM, 13 of which were anti-bacterial (2 against *Acinetobacter baumannii*, 4 against *Klebsiella pneumoniae* and 7 against *Streptococcus pneumoniae*) and the remaining 20 were anti-viral (8 against *Influenza A*, 1 against *Influenza B* and 11 against *Respiratory Syncytial* Virus). Both parental and derivative AMPs showed high binding affinity against their respective receptors with greater specificity (binding affinity) displayed by the derivative AMPs.

Conclusion

The putative Ap-AMPs against each pneumonia pathogen showed the expected physicochemical properties inherent to this particular class of molecules using APD and Bactibase and bound their respective protein receptors with great affinity as observed using PatchDock. The AMPs could serve as diagnostic molecules for use in a lateral-flow-device (LFD) in the detection of bacterial and viral pneumonia from patient samples. Also, this is the first report where *in silico* analysis were used to design putative AMPs that could have potential use in a LFD construction for the differential detection of viral and bacterial pneumonia at an early stage to prevent serious complications following infection.

Keywords: Antimicrobial Peptides, Bacteria, Viruses, Databases, Algorithms, Pathogens, Diagnostics, Receptors, Protein and Ligands.

TABLE OF CONTENTS

Declaration……..ii

Acknowledgement……..iii

Dedication…….....v

Abstract…….....vi

Table of Contents…….....ix

List of Abbreviations…….....xix

List of Figures……xxv

List of Tables…….....xxvi

List of Appendices…….xxviii

Chapter 1.0…….……..1

1.0. Introduction and Literature Review..... 1

1.1. Pneumonia 1

1.2. Causes of Pneumonia from Anatomical and Immunological Perspectives 4

1.3. Causes of Pneumonia from Clinical Perspective..... 7

1.3.1. *Klebsiella pneumoniae*..... 9

1.3.2. *Streptococcus pneumoniae* 10

1.3.3. *Acinetobacter baumannii*..... 12

1.3.4. *Respiratory Syncytial Virus* 14

1.3.5. *Influenza* Viruses 15

1.4. Symptoms of Pneumonia..... 16

1.5. Vaccines for Pneumonia 17

1.6. Available Methods for Pneumonia Diagnosis 18

1.6.1. Blood Culture Method 18

1.6.2. Chest X-ray Scan 19

1.6.3. Bronchoscopy analysis using sputum sample 19

1.6.4. Polymerase Chain Reaction (PCR) Assays 20

1.6.5. Matrix-Assisted Laser Desorption Ionization Time of Flight 20

1.6.6. Immunoassay…….21

1.7. Treatment of Pneumonia…….....21

1.8. Use of Host-Mediated Biomarkers in the diagnosis of Pneumonia 22

1.8.1. C-Reactive Protein 22

1.8.2. Procalcitonin 23

1.8.3. Soluble-Triggering Receptor Expressed on Myeloid Cells-1 24

1.9. Biomarkers used in prognosis 25

1.9.1. Copeptin and Mid-Regional and Atrial-Natriuretic Peptide (MR-proANP…….....25

1.9.2. Cortisol 26

1.9.3. Genetic and Proteomic Biomarkers 26

1.9.4. Use of Biomarkers in Diagnostic Kits……27

1.10. Antimicrobial Peptides (AMPs) 27

1.10.1. Categories of AMPs 29

1.10.2. Modes of Action of AMPs 30

1.10.3. Techniques used to Elucidate AMP Functions……..31

1.10.4. Natural and Artificial Production of AMPs ..... ….33

1.10.5. Computational tools used for Prediction of Antimicrobial Peptides 34

1.10.6. HMMER as a Prediction Tool for AMPs…….36

1.10.7. CD-HIT as a Prediction Tool for AMPs……..37

1.11. Rationale behind this Research Project……..37

1.12. Significance of this Study…….……..38

1.13. Limitations and Strength of this Research work…….39

1.14. Aim…….40

1.15. Objectives…….....40

Chapter 2.0…….41

2.0. In silico identification of anti-pneumonia antimicrobial peptides AMPs……..….41

2.1. Introduction……...41

2.2. Data Mining 42

2.2.1. Antimicrobial peptides Databases……..42

2.2.1.1. Collection of Antimicrobial peptides Database……..42

2.2.1.2. Antimicrobial peptides Database (APD) 43

2.3. Text Mining 43

2.4. *In silico* tool for Peptide Discovery 44

2.4.1. Hidden Markov Models (HMMER) 44

2.5.1. Aim……...…….44

2.5.2. Objectives of this chapter……..44

2.6. Materials and Methods 46

2.6.1. Data retrieval 47

2.6.2. Training and testing data sets 48

2.6.3. Construction of AMPs profiles..... 48

2.6.4. Independent model testing..... 49

2.6.5. Performance measures of each profile based on a prediction of the positive and the negative testing set……...50

2.6.6. Identification of novel putative anti-Pneumonia AMPs from proteome sequences…….50

2.6.7. Final list of Anti-pneumonia AMPs……..51

2.6.8. Physicochemical properties of the Putative anti-pneumonia AMPs…….....51

2.7. Results and Discussion……52

2.7.1. Retrieval of AMPs……52

2.7.2. Profile Creation using HMMER……53

2.7.3. Independent testing of the Models…….54

2.7.4. Performance measurement of the target specific profiles..... 57

2.7.5. Proteome sequence databases query and discovery of putative anti- pneumonia AMPs……...59

2.7.6. Physicochemical properties of the AMPs…….62

2.7.7. Similarity Analysis using BLAST…….66

2.8. Conclusion 69

Chapter 3.0…….70

3.0. *In silico* Comparison of HMMER with CD-HIT Identification of Anti-Pneumonia Antimicrobial Peptides (AP-AMPs) 70

3.1. Introduction 70

3.1.1. CD-HIT Algorithm 71

3.2 Aim……71

3.3. Objectives 72

3.4. Materials and Methods 72

3.4.1. Data retrieval 72

3.4.2. Identification of novel putative anti-Pneumonia AMPs from proteome sequences……73

.4.3. AMPs Prediction using CD-HIT 73

3.5. Results and Discussion…….....73

3.5.1. CD-HIT Analysis of db1 and db2……..73

3.5.2. AMPs Identification using CD-HIT 74

3.6. Conclusion……77

Chapter 4.0……78

4.0. Identification of Receptors as Targets for Detection of Bacterial and Viral Pneumonia……...78

4.1. Introduction 78

4.2. Characteristics of Receptors to be used in Diagnosis of Pneumonia……78

4.3. Types of Receptors 78

4.3.1. Intracellular Receptors..... 79

4.3.2. Cell Surface Receptors 80

4.3.3. Ligand Gated Ion Channel Receptors..... 80

4.3.4. G-Protein Coupled Receptors (GPCR)..... 81

4.3.5. Receptor Tyrosine Kinases 82

4.4. Aim……..82

4.5. Objectives 82

4.6. Materials and Methods 83

4.6.1. Identification of Receptors 83

4.6.2. Protein Retrieval 83

4.6.3. Determination of the protein sequences 84

4.6.4. Selection of Final list of Pneumonia Proteins……84

4.7. Results and Discussion 84

4.7.1. Retrieval of pneumonia protein receptors of the pathogens……..84

4.7.2. Physicochemical properties of the proteins……84

4.7.2.1. *Acinetobacter baumannii*…….....85

4.7.2.2. *Streptococcus pneumoniae*……..85

4.7.2.3. *Klebsiella pneumoniae*……86

4.7.2.4. *Respiratory syncytial* virus……..86

4.7.2.5. *Influenza* A virus…….87

4.7.2.6. *Influenza* B virus…….88

4.7.3. Final list of the pneumonia protein receptors…….....89

4.8. Conclusion 89

Chapter 5.0…….91

5.0. Structure Prediction and Docking Analysis of the Anti-pneumonia Putative AMPs and Pneumonia Ligands 91

5.1. Introduction 91

5.2. Structure Prediction using ITASSER 91

5.3. Docking Analysis using PATCHDOCK 92

5.3.1. Visualization of the Structures..... 93

5.4. Aim…….....93

5.5. Objectives 93

5.6. Materials and Methods 94

5.6.1. *De novo* structure predictions of the putative anti-Pneumonia AMPs and Pneumonia proteins (receptors) using I-TASSER..... 94

5.6.2. PyMol Analysis of the I-TASSER 3-D Structures 94

5.6.3. Docking Analysis of the Putative Anti-pneumonia AMPs and Pneumonia Proteins using PATCHDOCK 95

5.7. Results and Discussion 95

5.7.1. Structure Prediction of the putative anti-pneumonia AMPs and Pneumonia protein receptors……..95

5.7.1.1. C score……97

5.7.1.2. TM score…….....97

5.7.1.3. RMSD……98

5.7.2. Docking Interaction analysis of the anti-pneumonia AMPs and their receptors……101

5.7.2.1. Anti-bacterial AMPs……104

5.7.2.2. Anti-viral AMPs……104

5.8. Conclusion……105

Chapter 6.0 106

6.0. *In Silico* Site-Directed Mutagenesis (SDM) Analysis of Putative Anti-pneumonia Antimicrobial Peptides 106

6.1. Introduction 106

6.1.1. Knowledge-based FADE and Contacts (KFC)..... 107

6.2. Aim……108

6.3. Objectives 108

6.4. Materials and Methods 108

6.4.1. Identification of hotspot amino acids 108

6.4.2. *In silico* site-directed mutagenesis:..... 109

6.4.3. Physicochemical properties: 109

6.4.4. *In silico* 3-D structure prediction:..... 109

6.4.5. *In silico* protein-protein interaction study 109

6.5. Results and Discussion 109

6.5.1. Knowledge-based FADE and Contacts (KFC) analysis..... 110

6.5.2. *In silico* SDM of the AMPs……..118

6.5.3. Physicochemical properties of the AMPs……119

6.5.3.1. Physicochemical properties of the Anti-bacterial AMPs……..119

6.5.3.2. Physicochemical properties of the Anti-viral AMPs……119

6.5.4. Structure Prediction of the mutated AMPs……..123

6.5.4.1. C score……..123

6.5.4.2. TM score and RSMD……....123

6.5.5. Docking Interaction analysis of the mutated AMPs and the pneumonia protein receptors……125

6.5.5.1. Anti-bacterial mutated AMPs……125

6.5.5.2. Anti-viral mutated AMPs…….125

6.6. Conclusion ……129

Chapter 7.0……...130

7.0. General Discussion 130

7.1. Brief Overview 130

7.1.1. Summary of Chapters 131

7.2. Conclusion 134

7.3. Future Work 135

7.3.1. Peptides synthesis 135

7.3.2. Molecular study 135

7.3.3. Construction of a lateral flow device 135

7.4. Future perspectives 135

REFERENCES 137

APPENDICES 159

List of abbreviations

Acronyms	Full Meaning
ABG	Arterial blood gas
ADIT	AutoDep Input Tool
AIDS	Acquired Immune Deficiency Syndrome
AMP	Antimicrobial peptide
ANN	Artificial Neural Network
AntiBP	Antibacterial peptide prediction
AP	Anti-pneumonia
APD	Antimicrobial Peptides Database
ATCC	American Type Culture Collection
AVP	Arginine vasopressin
AVPDB	Antiviral peptides databases
BAA	Billiard Association of America
BACTIBASE	Bacterial antimicrobial peptides database
BAL	Bronchoalveolar lavage
BLASTP	Basic local alignment search tool for protein
CAP	Community acquired pneumonia
CAMP	Collection of antimicrobial peptide
C3b-C9	Complement factors
CD	Cluster of differentiation
CD-HIT	Cluster database at high identity with tolerance
CDC's ACIP	Centre for Disease Control's Advisory Committee on Immunization Practices

CNS	Central Nervous System
COPD	Chronic obstructive pulmonary disorder
CRAMP	Cathelicidin-related antimicrobial peptide
CRP	C-reactive proteins
CT	Computed tomography
CT-proAVP	C-terminal pro-vasopressin- Copeptin
CX3CR1	Chemokine/fractalkine/ G-protein coupled receptor
DA	Discriminant analysis
DNA	Deoxyribonucleic acid
D2R	Dopamine Receptor D2
EVC	Exhaled ventilator condensate
Eno	Enolase
F	Type I integral membrane fusion protein
FDA	Food and Drug Administration
F-kNN	Fuzzy K-nearest neighbor algorithm
FN	False negative
FP	False positive
G	a type II integral membrane attachment glycoprotein
GABA-A	Gamma Amino-butyric acid-A
GERD	Gastroesophageal reflux disease
GLAM2	Gap Local Alignment of Motifs 2
GPCR	G-Protein Coupled Receptors
GPR2	G-protein Coupled Receptor 2
HnNn	Influenza A subtypes

HA	Hemagglutinin
HBD	Human **β** defensing
hCAP18	Human cationic antibacterial protein of 18 kDa
HGF	Hepatocyte growth factor
HIV	Human immunodeficiency virus
HMGB-1	High Mobility Group Box-1
HMMER	Hidden Markov Models
HPAI	Highly Pathogenic Avian Influenza
HPUB	Held until publication
HSP	High scoring segment pairs
ICU	Intensive care unit
IgA	Immunoglobulin A
IRAK-4	Interleukin-1-receptor associated kinase-4
I-TASSER	Iterative threading assembly refinement
KFC	Knowledge-based FADE and Contacts
L	Large protein, the major polymerase
LD	Linear Discriminant Analysis
LFD	Lateral flow device
LL-37	Interleukin
CAP-18	Cathelicidin antimicrobial peptides
LOMETS	Local meta-threading server
LPAI	Low Pathogenic Avian Influenza
LPS	Lipopolysaccharide
LTS	Long term support

Mp	Matrix protein
MAP	Multiple antigenic peptide
MAXIT	Macromolecular Exchange Input Tool
MCC	Mathew's correlation coefficient
MD	Molecular Dynamics
mGluR6	Metabotropic Glutamatereceptor-type 6
MIC	Minimum inhibitory concentration
MmCIF	Macromolecular Crystallographic Information File
mRMR	Maximum relevance minimum redundancy
mRNA	Messenger ribonucleic acid
MR-proANP	Midregional pro-atrial natriuretic peptide
MyD88	Myeloid differentiation marker 88
MySQL	Michael Widenius's Structured Query Language
NA	Neuraminidase
NCBI	National Centre for Biotechnology Information
NLR	Neutrophil-like receptors
NMR	Nuclear Magnetic Resonance
NNA	Nearest neighbor algorithm
NS	Nonstructural
OMP	Outer membrane protein
OS	Operating system
OTC	Over-the-counter
PAF	Platelet activating factor
PavA	Pneumococcal adhesion and virulence factor A

PCR	Polymerase Chain Reaction
PCT	Procalcitonin
PCV	Pneumococcal conjugate vaccine
PDB	Protein Data Bank
PH	Hydrogen potency
PI	Isoelectric point
PHP	Hypertext preprocessor
PHYRE	Protein Homology/Analogy Recognition Engine
POC	Point of Care
proADM	Proadrenomedullin
PROC	Processing
pseAAC	Pseudo amino acid composition
PspA	Pneumococcal surface protein A
Psa	Pneumococcal surface antigen
PTMs	Post-translational modifications
PubMed	Public/Publisher MEDLINE
QSAR	Quantitative Structure Activity Relationship
RCBS-PDB	Research Collaborators for Structural Bioinformatics Protein Data Bank
RCT	Randomized controlled trials
RF	Random Forest
RIDT	Rapid influenza diagnostic tests
RMSD	Root mean square deviation
RLR	RIG-1-like Receptors

RNA	Ribonucleic acid
RSVM	Anti-*Respiratory Syncytial* Virus model
SARS	Severe acute respiratory syndrome
SDM	Site directed mutagenesis
SNP	Single Nucleotide Polymorphs
STREM-1	Soluble triggering receptor expressed on myeloid cells-1
SVM	Support vector machine
SW	Sliding Window
TIRAP	Toll interleukin-1 receptor domain containing adaptor protein
TLR	Toll-like receptors
TM	Template modelling score
TN	True negative
TP	True positive
TRAM	Toll interleukin receptor-domain-containing-adaptor-inducing-Interferon **β**
UniProtKB	Universal Protein Resource Knowledgebase
VAP	Ventilator acquired pneumonia

List of Figures

Chapter 1

Figure 1. 1: Diagram of lungs infected with pneumonia 1

Figure 1. 2: Mechanisms of antimicrobial peptides.. 32

Chapter 2

Figure 2. 1: Schematic representation of the methodology used for creation of various profiles using HMMER. 46

Figure 2. 2: Diagrammatic representation of the discovery of potential Anti-Pneumonia (AP) AMPs through proteome scanning 47

Chapter 3

Figure 3. 1: Outline of the CD-HIT *in silico* methodology used for AMPs identification……72

Chapter 4

Figure 4. 1: Diagrammatic representation of the process for receptors identification ...83

Chapter 5

Figure 5. 1: Schematic representation of the methodology used within this chapter. 94

Figure 5.2: 3-D Structures of the AMPs and Pneumonia Proteins as determined by I-TASSER……..96

Figure 5.3: Representative of the Docking Interaction of a Protein Receptor and Putative anti-pneumonia AMPs as determined by PATCHDOCK……..102

Chapter 6

Figure 6. 1: 3-D Structure of the Derivative anti-Pneumonia AMPs using I-TASSER……121

Figure 6. 1: 3-D structures of the Derivative anti pneumonia AMPs using I-TASSER……123

Figure 6. 2: Representative sample of the docking analysis of a mutated Ap-AMP with its receptor using PATCHDOCH and visualized using PyMOL.. 128

List of Tables

Chapter 2.0

Table 2.1: Retrieval and Diversity of the experimentally validated AMPs from the databases……..53

Table 2.2: Profile Creation by HMMER……..54

Table 2.3: Independent testing of the profiles……..57

Table 2.4: Performance measurement of the target profiles…….57

Table 2.5: Final list of AMPs……62

Table 2.6: Physicochemical parameters of the AMPs……..67

Chapter 3.0

Table 3.1: Outline of the CD-HIT In silico Methodology used for AMPs Identification…….....74

Table 3.2: Comparison of the Percentage Similarity of the Anti-pneumonia AMPs and Experimentally Validated AMPs…….75

Chapter 4.0

Table 4.1: Physicochemical properties of the pneumonia protein receptors……88

Chapter 5.0

Table 5. 1: Quality assessment scores of the predicted 3-D structures of the pneumonia receptors……..98

Table 5. 2: Quality assessment scores of the predicted 3-D structures of the putative anti-*Acinetobacter baumannii* AMPs. 99

Table 5. 3: Quality assessment scores of the predicted 3-D structures of the putative anti-Influenza A Virus AMPs. 99

Table 5. 4: Quality assessment scores of the predicted 3-D structures of the putative anti-Influenza B Virus AMPs. 100

Table 5. 5: Quality assessment scores of the predicted 3-D structures of the putative anti-*Klebsiella pneumoniae* AMPs. 100

Table 5. 6: Quality assessment scores of the predicted 3-D structures of the putative anti-Respiratory Syncytial Virus AMPs. 100

Table 5. 7: Quality assessment scores of the predicted 3-D structures of the putative anti-*Streptococcus pneumoniae* AMPs. 101

Table 5. 8: Quality assessment scores of the docking analysis for the anti-pneumonia putative AMPs and the pneumonia receptors 102

Chapter 6.0

Table 6. 1: Identification of mutation sensitive amino acid residues "hotspots" using KFC……...111

Table 6. 2: Physicochemical properties of the mutated anti-bacterial and anti-viral pneumonia AMPs……..121

Table 6. 3: Structure prediction scores of the anti-pneumonia AMPs 124

Table 6. 4: The docking analysis of only derivative viral and bacterial anti-pneumonia AMPs and their receptors which showed increased binding scores 127

List of Appendices

Appendix A: Supplementary Materials for Chapter 2.0

Table A1:Anti-*Acinetobacter baumannii* AMPs……159

Table A2:*Anti-Influenza A* virus AMPs…….159

Table A3:*Anti-Influenza B* virus AMPs…….161

Table A4:Anti-*Klebsiella pneumoniae* AMPs……...162

Table A5:Anti-*Respiratory Syncytial Virus* AMPs……...168

Table A6:Anti-*Streptococcus pneumoniae* AMPs……173

Appendix B: Supplementary Materials for chapter 4.0

Table B1: Iron-regulated outer membrane proteins [*Acinetobacter baumannii*]….175

Table B2: Iron-regulated outer membrane protein [*Klebsiella pneumoniae]* ……175

Table B3: Pneumolysin [*Streptococcus pneumoniae*]…….....176

Table B4: Membrane protein M1 [*Influenza A* Virus……176

Table B5: Chain C, R416a Monomeric Nucleoprotein of *Influenza A* Virus……..176

Table B6: Nucleoprotein [*Influenza B* virus]…….. 176

Table B7: RSV Chain A, Rsv Matrix Protein…….....176

Table B8: Chain X, Human *Respiratory Syncytial* Virus Fusion Protein Core…...177

Chapter 1

1.0. Introduction and Literature Review

1.1. Pneumonia

Pneumonia is an inflammation of the lung parenchyma (Curbelo *et al.*, 2017). It is derived from a Greek word, Pneuma, which means "breath". Parenchyma is the functional part of any organ whereas stroma means the structural part made of connective tissue. This disease starts when the lungs parenchyma become inflamed, and the tiny air sacs, or alveoli, inside them, fills up with fluid. At this point, pneumonia-causing germs then settle in the alveoli and multiply, causing the lung sacs to be filled with pus (Ngari *et al.*, 2014). This contagious disease is then spread through contact with or touch of shared objects, sneezing, and coughing or even inhaled. It is caused by either a bacterial, viral, or fungal agent. It is most dangerous for older adults, infants, people with other diseases, and those with an impaired immune system. Pneumonia affects children and families everywhere in the world but it's most prevalent in Africa (Kalil *et al.*, 2016).

Figure 1. 1: Diagram of lungs infected with pneumonia (Copyright© 2007 Wolters Kluwer Health Care, Inc. All rights reserved) (Andersson-Gäre and Neuhauser, 2007)

Each year, approximately 1.6 million children die from pneumonia (Benjamin *et al.*, 2017). 920,136 children under the age of five were killed by pneumonia in 2015, accounting for sixteen percent of all deaths of children in that year. In the United States, for instance, around one million people are treated in hospital for pneumonia each year, and around 50,000 die from the disease. In 2013, more than $19.9 billion was spent on pneumonia and influenza health care. Pneumonia expenses accounted for eighty-one percent of this total ($16.2 billion) (Li *et al.*, 2017a). Pneumonia is ranked as one of the leading cause of death (Ramirez *et al.*, 2017).

Significant morbidity and mortality is associated with pneumonia in older patients compared to younger adult (Son *et al.*, 2017). Those most at risk include people who are aged under 5 years or over 65 years; smoke tobacco, consume large amounts of alcohol, or both; have underlying conditions such as cystic fibrosis, chronic obstructive pulmonary disorder (COPD), asthma, or conditions that affect the kidneys, heart, or liver; have a weakened or impaired immune system, for example, due to AIDS, HIV, or cancer; take medicines for gastroesophageal reflux disease (GERD); have recently recovered from a cold or influenza infection; experience malnutrition; have been recently hospitalized in an intensive care unit; and have been exposed to certain chemicals or pollutants (Caruana *et al.*, 2015). Some groups are more prone than others to pneumonia, including Native Alaskan or certain Native American ethnicities (Kim *et al.*, 2017). Despite considerable effort by health agencies and organizations, mortality rates due to pneumonia in most developing countries are growing at an alarming increase (Liu *et al.*, 2018).

Due to this, efforts have been made to bring several articles together that could fill the remaining gaps in understanding this acute respiratory infection in children and the elderly. Almeida and Boattini (2017) revised national-level estimates of the morbidity, severe

morbidity, mortality, etiology and risk factors for acute lower respiratory infections. This provided an assessment of comorbidity between childhood pneumonia and diarrhea with the use of surveys to study the perception of people towards oral rehydration sachets, antibiotics, and other therapies for diarrhea and pneumonia in India and Kenya; Momosaki *et al.* (2016) analyzed scaling up access to oral rehydration solution for diarrhea and pneumonia, learning from historical experience in low– and high–performing countries; Wazny *et al.* (2013) added to this work by surveying caregivers in Kenya to assess perceptions of zinc as a treatment for diarrhea and pneumonia. Finally, Bhutta *et al.* (2013) reported on the global action plan for childhood diarrhea and pneumonia and showed how research priorities were developed resulting from this effort, among others. Despite all these, experts have called for more diagnostic research on the pathogens causing pneumonia and the routes through which it is transmitted, as this is of critical importance for treatment and prevention (Rajpurkar *et al.*, 2017).

Furthermore, the Alexandre Project, a worldwide surveillance study found that the global rate of pneumococcal macrolide resistance had risen from 16.5_% to 21.9_% in 1996-1997 to 29_% in 1998-2000 (Rudan *et al.*, 2005). Resistance to macrolides and fluoroquinolones which first became available with approval of levofloxacin in 1996 have been reported with increasing frequency, with worldwide rate of resistance to *S. pneumonia* strains ranging from 4_% to 100_% (Zilberberg *et al.*, 2017). In the face of growing bacterial resistance, the utility of conventional antibiotics for the treatment of pneumonia is strained and development of new biomarkers for accurate and sensitive diagnosis to aid proper treatment and understanding of the pathogens is essential (Fernández-Barat *et al.*, 2017).

1.2. Causes of Pneumonia from Anatomical and Immunological Perspectives

Pneumonia occurs with or without predisposing factors. However, the presence of these factors increases the chance of contracting pneumonia (Hong *et al.*, 2017). These factors are divided into two, namely: (i) impaired defense mechanism and (ii) decreased host resistance. To understand the cause of pneumonia from an impaired-defense-mechanism point of view, the knowledge of the anatomy of the lung in the area of cough reflex function and mucociliary apparatus or reflex of the lung is crucial (Alsubaie *et al.*, 2015). It is noteworthy to say that the lung does not only take air in or out but also particles as well. Examples of these particles are chemicals, smoke, dust, micro-organisms, just to mention a few. If the particles reach the alveoli, alveolar macrophages and tissue, histiocytes engulf them. The respiratory tract also has upper airway filters that rid these particles through cough reflex, thereby preventing aspiration. However, these defenses can be overcome by large numbers of organisms or by compromised effectiveness resulting from air pollutants (e.g. cigarette smoke) or interference with protective mechanisms (e.g. endotracheal intubation, tracheostomy) resulting in pneumonia (Mehl, 2018). Also, loss of cough reflex can result in pneumonia in cases like unconsciousness or coma, administration of anesthesia during surgery, neuromuscular disorder, certain drugs etc (Suntrup-Krueger *et al.*, 2017).

Furthermore, to better understand the mucociliary apparatus or reflex, it is imperative to divide the lung into the conducting and respiratory zones (Chinnapaiyan *et al.*, 2017). The conducting parts include the nose, pharynx, trachea, bronchi, smaller bronchi (bronchioles). The respiratory parts also include respiratory bronchioles, alveolar ducts, and alveoli. It is relevant to mention that the right bronchus is shorter, wider and more vertical than the left. This bronchus (either left or right including the bronchioles) is lined by pseudostratified ciliated columnar epithelium that is responsible for propelling different particles taken into the lung to

the outside (Kojima *et al.*, 2017). Stratification means multiple layers but when this part is viewed under a microscope, it shows a single layer with multiple nuclei, giving rise to the word pseudo-stratified. However, the alveoli do not have these epithelial linings; rather it has very thin cells known as pneumocysts (Liu *et al.*, 2017c). Pneumocysts are of two types: those ones that are thin, squamous which are responsible for gaseous exchange and the non-squamous which is used to make important substances called surfactants, for surface tension reduction and prevention of collapse of the alveoli (Sweet *et al.*, 2017). If invading organisms reach the tracheobronchial tree, the mucociliary epithelium transports them away from the lung. However, the loss of mucociliary apparatus or reflex function either in the pseudostratified ciliated columnar epithelium or pneumocysts results in micro-organisms and other particles becoming stuck which can affect the lung parenchyma. Activities that disrupt the proper functioning of the mucociliary apparatus through the bronchial pseudo-stratified ciliated columnar epithelium include smoking, inhalation of poisonous gases, genetic disorder like immotile cilia syndrome and viral infections (Chinnapaiyan *et al.*, 2017). The accumulation of secretion causing cystic fibrosis which brings about infection is the only form of pneumonia that affects the alveoli (Rutter *et al.*, 2017).

In addition, immunodeficiency or decreased host resistance of the innate or adaptive immune system can cause pneumonia disorders (Messiaen *et al.*, 2017). The immune system is a host's defense system comprising many biological structures and processes within an organism that protects against diseases. Pathogens can rapidly evolve and adapt, and thereby avoid detection and neutralization by the immune system; however, multiple defense mechanisms such as phagocytosis, use of antimicrobial peptides, and the complement system have also evolved to recognize and neutralize pathogens (Lee, 2017). The ability of the immune system to respond to pathogens is diminished in both the young and the elderly, with immune responses beginning

to decline at around 50 years of age due to immuno-senescence (Rajpurkar *et al.*, 2017). Factors such as obesity, alcoholism, drug abuse, and malnutrition are common causes of poor immune function that can make an individual vulnerable to pneumonia infection (Nguyen *et al.*, 2017).

The first response of the immune system of an individual to pneumonia infection is inflammation (Chakraborty *et al.*, 2017). Inflammation is initiated by the release of cytokines which include: interleukins that are responsible for communication between white blood cells; chemokines that promote chemotaxis; and interferons that have anti-viral effects, such as shutting down protein synthesis in the host cell. Growth factors and cytotoxic factors may also be released. These cytokines and other chemicals recruit immune cells to the site of infection and promote healing of any damaged tissue following the removal of pathogens (Martin and Leibovich, 2005). However, during pneumonia infection, inflammatory response of an individual is compromised by the bacterial and viral agents.

Moreover, the host resists further attack of pneumonia pathogens by using the toll-like receptors 1-9 (TLR1-9) of the leukocytes in a cascade system to trap bacterial pneumonia through binding to their lipopolysaccharide with high specificity. TLR signaling has very high specificity due to the recruitment of adaptor molecules such as myeloid differentiation marker 88 MyD88, TIR domain-containing adaptor protein TIRAP, TIR domain-containing adaptor inducing beta interferon TRIF, TRIF-related adaptor molecule TRAM) (Tagami *et al.*, 2004) Leukocytes also use neutrophil-like receptors, NLR, which are cytoplasmic proteins with three domains and have more than 22 members in humans. The leucine-rich repeats recognizes bacterial cell wall components. Of all NLRs, it is the neutrophil-like receptors (NLR) oligomerization domain 1 and 2 proteins that bind peptidoglycan and muramyl dipeptide of the pneumonia bacterial cell wall. However, after neutrophils capture bacteria, they die.

Subsequently, they become apoptotic as they release injurious substances after phagocytosing bacteria.

Persistent presence of dying neutrophils causes tissue damage and eventually organ dysfunction. This is the cause of lung failure in acute pneumonia. Thereafter, accumulated macrophages phagocytose apoptotic neutrophils to form Hepatocyte Growth Factor, HGF, a potent growth factor for pulmonary epithelial cells (Boven *et al.*, 2005). Therefore, an early and sensitive diagnostic biomarker is imperative that can detect the viral and bacterial pneumonia pathogens before it establishes lung dysfunction.

1.3. Causes of Pneumonia from a Clinical Perspective

Etiology of pneumonia, for easy understanding, could be viewed from a clinical settings (Murdoch *et al.*, 2009). Based on this, pneumonia is divided into (i) community-acquired acute pneumonia which can be caused by *Streptococcus pneumoniae, Haemophilus influenzae, Moraxella catarrhalis, Staphylococcus aureus, Legionella pneumophila, Enterobacteriaceae (Klebsiella pneumoniae and Pseudomonas spp),* (ii) community-acquired atypical pneumonia which is caused by *Mycoplasma pneumonia, Chlymadia spp, Coxiella burnetti and viruses (Respiratory Syncytial Virus, Influenzae* types A and B, Adenovirus known as SARS virus)), (iii) hospital-acquired pneumonia which is caused by Gram-negative rod, *Enterobacteriaceae spp, Staphylococcus aureus and Pseudomonas spp*, (iv) aspiration pneumonia means loss of cough reflex occurring in alcoholics which is caused by anaerobic oral flora such as *Bacteroides, prevotella, Fusobacterium, Pepto-streptococcus,* admixed with aerobic bacteria), (v) chronic pneumonia which is caused by *Nocardia, Actinomyces, granulomatous, Mycobacterium tuberculosis, atypical mycobacteria, Histoplasma capsulatum, Coccidiodes immitis, Blastomyces, Dermatidis* and (vi) necrotizing pneumonia and lung abscesses caused by anaerobic bacteria with/or without mixed infection that occurs in immune-compromised

patients as a result of *Cytomegalovirus, Pneumocystis jiroveci, Mycobacterium avium-intracellulare*, invasive Aspergillosis, invasive candidiasis infection.

Worldwide, the most common bacterial cause of community-acquired pneumonia (CAP) is *S. pneumoniae* (Bogaert *et al.*, 2004). This can be strengthened by a recent state-of-the-art diagnostic techniques study for bacterial, viral, and fungal infections in which 38_% of a specific pathogen was detected for community-acquired pneumonia (CAP) cases. The detection of one or more viruses was seen in 23_% of cases and 11_% of other bacteria. Bacterial and viral pathogens combinations were seen in 3% of cases with fungal and mycobacterial organisms accounting for 1_% of cases. Human rhinoviruses were isolated in 9_% of cases and influenza virus in 6_% of cases (Jain *et al.*, 2015).

In summary, bacterial pneumonia is most commonly caused by *Streptococcus pneumoniae, Klebsiella pneumoniae and Acinetobacter baumannii* (Palacios *et al.*, 2009). Viral pneumonia can result, most commonly from the *Respiratory Syncytial* virus (RSV), *Influenzae* types A and B, known as the flu and *Pneumocystis jiroveci* in people infected with HIV, responsible for at least one-quarter of all pneumonia deaths in HIV infected individuals (Ruuskanen *et al.*, 2011). Fungal pneumonia can result from a condition such as a valley fever, caused by the *Coccidioides* fungus (Huang and Crothers, 2009). Aspiration pneumonia which is not contagious, can manifest when a person breathes food, liquids, or stomach contents into the lungs (Marik and Kaplan, 2003). Hospital-acquired pneumonia can occur in patients being treated for other conditions, for example, those attached to a respirator, or a breathing machine (Society and America, 2005). Regardless of the cause, the signs and symptoms will be similar. However, only pneumonia caused by bacteria can be apparently diagnosed, necessitating

differentiation of bacterial from viral causative agents of the disease, with the effectiveness of the drugs and vaccines against the disease being controversial (Liu *et al.*, 2015).

1.3.1. *Klebsiella pneumoniae*

Klebsiella pneumoniae is a lactose fermenting, non-motile, rod shaped bacterium, Gram-negative, encapsulated, facultative anaerobic, that appears as a mucoid on MacConkey agar (Siu *et al.*, 2012). It live harmlessly inside the intestine and faeces but can be destructive when found in other organs such as the lungs (Munoz-Price *et al.*, 2013). *Klebsiella pneumoniae* pathogenicity can be initiated by lipopolysaccharides (LPS) through activation of the C3b complement for selective deposition onto LPS molecules at sites distant from the bacterial cell membrane. This results in the inhibition of the membrane attack complex (C5b-C9) causing membrane damage and cell death (Cortés *et al.*, 2002). The hosts vulnerability to *K pneumoniae* infection is also increased by iron availability where bacteria are able to compete effectively for host iron proteins through secretion of siderophores, a high-affinity, low molecular weight iron chelators (Weinberg, 2009). These bypass processes make its treatment at the later stages of its infection difficult.

Recent preclinical studies of *Klebsiella pneumonia* infection have shown that neutrophil myeloperoxidase and lipopolysaccharide-binding protein of the host play a protective role in defense against *K pneumoniae* infection (Hirche *et al.*, 2005). This is achieved through (i) oxidative inactivation of elastase of this pathogen which is mediated by neutrophil myeloperoxidase, a tissue-destroying diseases enzyme that plays a vital role in the organism's pathogenesis; and (ii) the transfer of bacterial cell wall components to inflammatory cells which is enhanced by lipopolysaccharide-binding protein of the host. There is higher rates of infection in experimental mice deficient in the genes that control the expression of these 2 agents: neutrophil myeloperoxidase and lipopolysaccharide-binding protein. The innate host immunity

overcomes *Klebsiella pneumoniae* invasion through the possession of a polysaccharide capsule with additional layer of protection from the bacterium through phagocytosis of polymorphonuclear granulocytes. However, the capsule also prevents bacterial death caused by bactericidal serum factors which inhibit the activation or uptake of complement components, especially C3b. Multiple adhesins are also produced by these bacteria which may be fimbrial or nonfimbrial, each with distinct receptor specificity (Strieter *et al.*, 2003). These enhance the microbes to stick to host cells, which is critical to the infection process. Outer membrane protein (OMP) is an highly conserved protein throughout evolution in this bacterium. It is, therefore, not surprising that antibody-mediated detection against iron regulated outer membrane protein of *Klebsiella pneumoniae* has been used for the diagnosis of this bacterium (Todhunter *et al.*, 1991).

1.3.2. *Streptococcus pneumoniae*

Streptococcus pneumoniae, also called pneumococcus, is a non-spore forming, non-motile and Gram positive bacterium, which is alpha haemolytic under aerobic condition, and beta haemolytic under anaerobic condition (Cherazard *et al.*, 2017). It is a facultative anaerobe of the genus "*Streptococcus*" that are usually found in pairs (Liu *et al.*, 2017b). *S pneumoniae* causes bacteremia, meningitis and otitis media, sinusitis, septic arthritis, osteomyelitis, peritonitis, and endocarditis. Fourteen and a half million invasive pneumococcal disease was reported in children under 5 years of age, which correlates to more than 800,000 estimated deaths (O'Brien *et al.*, 2009). Childhood pneumococcal conjugate vaccine (PCV) which was introduced in the United States in 2000, has led to a decreased invasive pneumococcal infections (>90_%) as well as an overall decrease in invasive diseases (45_% overall; 75_% in children < 5 years) (Cutts *et al.*, 2005).

Apart from this, many studies have shown increased rates of invasive and non-invasive disease caused by serotypes 15, 19A, and 33F not covered by the vaccine (van der Linden *et al.*, 2015). Invasive disease analysis of over 700 cases, in completely immunized children with PCV7 revealed that 96_% of infections were due to non-vaccine serotypes. Another 6 serotypes accounted for almost two-thirds of invasive infections in this age group. Before and after the implementation of PCV7 immunization, 653 invasive pneumococcal infections analysis in the Spanish population showed an increased incidence in the post-vaccine period, which was primarily due to non-vaccine serotypes and was associated with higher rates of complications, such as septic shock. Similar studies in the United States and other European countries have shown similar results, introducing the concept of replacement disease and its effects (O'Brien *et al.*, 2009).

Two distinct states determine the virulence of this bacterium: opaque and transparent colony types which influence the invasion of host defenses. The transparent phenotype is predominantly found in the nasopharynx whereas the opaque type predominates in the lung, central nervous system (CNS), and bloodstream (Henriques-Normark and Tuomanen, 2013). In addition, *in vitro* and *in vivo* studies of clinical isolates have shown that pneumococci have the tendency to acquire DNA from other pneumococci (or other bacteria) through transformation, thus switching serotypes to distinct capsular types. For this reason, *S pneumoniae* can adhere strictly to the respiratory epithelium and mucus by exhibiting different surface proteins that recognize and attach to human cells (Izoré *et al.*, 2010). Pneumolysin, and the ABC transporter have been implicated in adherence of this bacterium to the host (Kadioglu *et al.*, 2008). Detection of pneumolysin in antibody titers has been used for positive diagnostic results of this bacterium (Andrade *et al.*, 2016). *S pneumoniae* produces biofilm which is regulated by external factors such as temperature after binding to host cells.

Increased invasion of this bacterium is enhanced by phospho-cholines after interacting with the PAF receptor, where it is inserted into the host cell through endocytosis, causing translocation of bacteria via the endothelium. This is essential for translocation over the blood-brain barrier during meningitis development. Surface coat of this bacterium is changed through gene expression to bypass host defenses and complete translocation. Invasion is also mediated by adhesins and pneumolysin; pneumolysin is a cytotoxin that induces apoptosis of epithelial cells by membrane pore formation, resulting in access to subendothelium. Intra-alveolar replication of pneumococci, penetration into the interstitium, and dissemination into the bloodstream are among other functions of pneumolysin (Kadioglu *et al.*, 2008). Serious clinical complications from pneumococcal disease results from the activation of the complement pathways and cytokine release, which induce a significant inflammatory response. Activation of the alternative complement pathway is enhanced by cell wall components and pneumococcal capsule whereas polysaccharide cell wall antibodies activate the classical complement pathway. Cell wall proteins, autolysin, and DNA released from *S. pneumoniae* breakdown all contribute to the production of cytokines, inducing further inflammation (Mitchell and Mitchell, 2010).

1.3.3. *Acinetobacter baumannii*

Acinetobacter baumannii is a pleomorphic, non-motile, Gram negative, opportunistic, aerobic rod-shaped bacillus (Lee *et al.*, 2017). It is a member of the ACB complex (*A. baumannii, A. calcoaceticus* and *Acinetobacter* genomic species 13TU) which makes its diagnosis very difficult (Schooley *et al.*, 2017). *A. baumannii* specifically targets moist tissues with mucous membranes or areas of the skin that are exposed through accident or injury. Skin and soft tissues infected with *A. baumannii* initially has a characteristic orange appearance followed by a sandpaper-like presentation which gives room for the appearance of clear vesicles on the skin

(Lee *et al.*, 2017). This infection can lead to septicemia and death if left untreated (Howard *et al.*, 2012). Despite extensive research into the virulence potential of *A. baumannii,* the understanding of its true pathogenic potential or virulence repertoire is still limited.

There is evidence that OmpA, a member of the Outer membrane proteins (OMPs), has been known to contribute significantly to the disease-causing potential of the pathogen (Howard *et al.*, 2012). *A. baumannii* OMPs bind to the host epithelial and mitochondria to induce mitochondrial dysfunction and swelling. Cytochrome c and heme proteins are then released which lead to the formation of the apoptosome resulting in apoptosis of the cell (Choi *et al.*, 2005). OMP, being the most abundant surface protein on the pathogen, is also involved in resistance to complement and the formation of biofilms, which are two key stress survival strategies and potentially important virulence-associated factors that help to promote bacterial survival both inside and outside the host (Wang *et al.*, 2017b). Iron regulated outer membrane protein, for instance, has been used to raise an antibody for diagnosis since it plays a role in siderophore-independent iron acquisition from hemoglobin of the host and has an important role in the survival of the bacterium inside the host to establish infection (Goel and Kapil, 2001).

Other key proteins that have been shown to contribute to *A. baumannii* virulence include phospholipase D and C. While phospholipase D is important for resistance to human serum, epithelial cell evasion and pathogenesis (Jacobs *et al.*, 2010), phospholipase C enhances toxicity to epithelial cells (Camarena *et al.*, 2010). Along with OmpA, fimbria, also expressed on the surface of the bacterial cell, contribute to the adhesion of the pathogen to host epithelia.

1.3.4. *Respiratory Syncytial Virus*

The Respiratory syncytial virus is a negative-strand non-segmented RNA virus of the family *Paramyxoviridae*, the subfamily *Pneumovirinae* and the genus *Pneumovirus* (Wright *et al.*, 2007). There are two subtypes, A and B. Virus size is pleiomorphic, typically 100–350 nm yet filaments may be as long as 10 µm. There are ten viral genes that are transcribed sequentially into separate mRNAs by the viral polymerase. Genes are linearly ordered from the 3′ to 5′ end of the negative-sense RNA genome. On the 3′ end are genes *NS1*, *NS2* (putative nonstructural proteins 1 and 2, anti-inflammatory proteins), *N* (nucleoprotein) and *P* (phosphoprotein, a cofactor in RNA synthesis). These are followed by *M* (matrix protein), *SH* (small hydrophobic protein, a nonessential protein that may function as a pentameric ion channel (Collins and Graham, 2008, Gan *et al.*, 2008), *G* (a type II integral membrane attachment glycoprotein), and *F* (a type I integral membrane fusion protein). Most 5′ are *M2* (a gene that encompasses two open-reading frames encoding M2-1 and M2-2, supportive of mRNA transcription and the balance between RNA replication and transcription) and *L* (large protein, the major polymerase) (Collins and Graham, 2008, *Wright et al., 2007*).

RSV replicates almost exclusively in the mammalian respiratory tract (Graham, 2011) where the furin-like protease is synthesized. Apart from facilitating virus fusion, the RSV F protein has isoform chains which are also capable of signaling through Toll-like receptor (TLR)-4 and CD14 (Kurt-Jones *et al.*, 2000, Blanco *et al.*, 2010). This signaling may induce proinflammatory cytokines including interferons with antiviral activities. By contrast, the G, NS-1 and NS-2 proteins may in some cases downregulate inflammatory events (Lo *et al.*, 2005).

The virus life cycle typically begins by attachment of the G glycoprotein to the mammalian host plasma cell membrane (F can also facilitate attachment), followed by F-mediated fusion. Some siRNA studies have suggested that clathrin-mediated endocytosis and early endosomes may also play a role in RSV infectivity (Kolokoltsov *et al.*, 2007). The cellular targets of G and F proteins are not fully defined. G protein can bind glycosaminoglycans such as heparin sulfate and chondroitin sulfate B as well as C-type lectins such as surfactant protein (LeVine and Whitsett, 2001). F protein also has heparin-binding domains (Feldman *et al.*, 2000) and like G, can bind glycosaminoglycans (Techaarpornkul *et al.*, 2002, Hallak *et al.*, 2000). The use of fusion protein isoform chains for the diagnosis of this virus in host blood samples is becoming a common practice because G proteins are less stable (Falsey *et al.*, 2002).

1.3.5. Influenzae Viruses

Influenza A virus is divided into subtypes based on two proteins on the surface of the virus hemagglutinin (HA) and neuraminidase (NA) (Taubenberger *et al.*, 2005). Influenza A virus has been named in accordance with their HA and NA surface proteins. An "H7N2 virus", for instance, designates an influenza A subtype that has an HA7 protein and an NA2 protein. In a similar manner, an "H5N1" virus has an HA5 protein and an NA1 protein. Unlike Influenza A virus, Influenza B virus is found only in humans but it is not classified according to subtypes. Although influenza type B virus can cause human epidemics, it has not caused pandemics (Wrammert *et al.*, 2011).

Drift and shift processes exist for new strains of influenza virus to replace older strains (Bouvier and Palese, 2008). This reduces the sensitivity of antibodies for their detection, therapeutics or vaccination as new strains emerge. Thus, the influenza vaccine is updated on a yearly basis to keep up with the changes in influenza viruses.

Drift occurs when an infected person with a particular flu virus strain develops an antibody against that virus. However, the antibodies against this virus strains are no longer recognized as "newer" virus strains appear. This is the main reasons why people can get flu more than one time and virus strains in the influenza vaccine are updated to keep up with the changes in the circulating flu viruses. For this reason, people who want to be immunized against influenza need to receive a flu vaccination every year.

The diagnosis of Influenza A using matrix protein M1 has been very promising because it is the most abundant protein across strains and subtypes within the viral particle (Eierhoff *et al.*, 2009). Also, there is evidence for the use of Influenza A and B viruses nucleoproteins against specific IgA antibodies which yielded positive results in 41.2_% of Influenza A infected patients and 66.7_% of Influenza B infected patients on day 6 after the onset of clinical symptoms (Voeten *et al.*, 1998).

1.4. Symptoms of Pneumonia

The first symptoms of pneumonia usually resemble that of a cold or flu. The person then develops a high fever, chills, and cough with sputum. Other symptoms include rusty or green phlegm/sputum coughed up from lungs, fast breathing and shortness of breath, chest pain that usually worsens when taking a deep breath, known as pleuritic pain, fast heartbeat, fatigue and weakness, nausea and vomiting, diarrhea, sweating, headache, muscle pain, confusion or delirium, especially in older adults, dusky or purplish skin color, or cyanosis, from poorly oxygenated blood (Thomas Jr and Limper, 2004). These symptoms can vary depending on other underlying conditions, the type of pneumonia and medical history (Weinberger *et al.*, 2009). An X-ray can show if there is any damage to the lungs. It is usual to suspect pneumonia if coarse breathing, wheezing, crackling, or decreased breath sounds is heard from listening to

the chest through a stethoscope. The oxygen levels in the blood, with a painless monitor on the finger, called a pulse oximeter may also suspect pneumonia (Marik, 2010). However, it is always advised to carry out a diagnostic test to establish the status of a patient rather than suspicion-based judgement.

1.5. Vaccines for Pneumonia

Two different vaccines exist to prevent pneumonia which cover a wide variety of pneumococcal infections and are recommended for both children and adults, depending on their health conditions. Pneumococcal conjugate vaccine (also known as Prevnar) and pneumococcal polysaccharide vaccine (known as Pneumovax) are used for prevention of pneumonia (Ngari *et al.*, 2014). Prevnar (PCV13) is normally included as part of an infant's routine immunizations. It is recommended for children under 2 years, adults over 65 years, and those between the ages of 2 and 64 years with certain medical conditions. Pneumovax (PPSV23) is recommended for children and adults who are at increased risk of developing pneumococcal infections. This includes adults aged 65 years or older, people with diabetes, those with chronic heart, lung, or kidney disease, people who consume large amounts of alcohol or who smoke, and those without a spleen. Those aged between 2 and 64 years with certain other medical conditions may be advised to have this vaccine (O'Brien *et al.*, 2009).

However, the vaccines may not completely protect older adults from pneumonia, but it can significantly reduce the risk of developing pneumonia and other infections caused by *S. pneumoniae),* including blood and brain infections. Also, there is a limitation to the use of vaccines as their efficacy is concentrated on pneumococcal pneumonia but not viral pneumonia. Along with vaccinations, physicians recommend regular hand washing, covering the mouth and nose when coughing or sneezing, refraining from smoking, eating healthfully,

exercising 5 days a week and staying away from the sputum or cough particles of pneumonia patient. This goes to say that vaccination might not fully protect an individual from the disease.

1.6. Available Methods for Pneumonia Diagnosis

Pneumonia is difficult to diagnose because it may cause symptoms commonly seen in people with colds or the flu. In the last years, the methods of detection of the biomarkers implicated in pneumonia disease have been constantly advanced, ranging from blood cultures, polymerase chain reaction, matrix-assisted laser desorption or ionization time of flight, immunofiltration, turbidimetric immunoassay based on latex agglutination, just to mention a few (Marik and Kaplan, 2003). With these methods, it is possible to detect locally confined bacterial infections like pneumonia. The treatment of pneumonia would greatly benefit from tests that could rapidly detect infection due to these organisms.

1.6.1. Blood Culture Method

Blood culture is a phenotypic method that has been used for the detection of bacterial pneumonia. It is noteworthy to mention that blood tests measure the white blood cell count which helps to determine how severe the infection is, and whether a bacteria, virus, or fungus is the likely cause (Davis *et al.*, 2017). Blood cultures may reveal whether the microorganism from the lungs has spread into the bloodstream. An arterial blood gas (ABG) test may provide a more accurate reading of the body's oxygen and carbon dioxide levels and other factors (Kalil *et al.*, 2016). Biochemical tests such as bile solubility, optochin sensitivity, counter current immunoelectrophoresis and coagglutination using laboratory prepared reagents also exist which can be used for pneumonia diagnosis. In a test performed by Sychantha and Clarke (2018) using 83 capsular types of *Streptococcus pneumoniae* strains, only bile solubility and

optochin tests were reliable, as others produced cross-reactivity. However, the bile solubility test is based on visual interpretation (Slotved *et al.*, 2017) while optochin can be susceptible to atmospheric carbon (IV) oxide and result obtained can be considered presumptive (Aguiar *et al.*, 2006), thus necessitating further testing. Chromogenic agar media test can be used to analyze the bacterial type causing pneumonia but it is time-consuming (Malhotra-Kumar *et al.*, 2010) while automated system like VITEK 2 can detect bacteria such as *Streptococcus pneumoniae* in an enriched growth medium but has very low sensitivity (Jorgensen *et al.*, 2000).

In general, blood cultures has shortcoming such as poor sensitivity with further limitations observed in the strength of the agglutination, which is not indicative of the concentration of the biomarkers and patients have to wait for days before getting a result. Blood culture may also produce false negative and false positive results as seen in patients with high titers of rheumatoid factors which may give false positive results (O'Brien *et al.*, 2009).

1.6.2. Chest X-ray Scan

Chest X-rays is a phynotypic method that can be used for pneumonia diagnosis and will also show which areas of the lungs are affected. A CT scan of the chest may provide more detailed information (Marik, 2010). However, this test cannot tell what kind of germ is causing the pneumonia (Rajpurkar *et al.*, 2017).

1.6.3. Bronchoscopy analysis using sputum sample

A sputum analysis is a phenotypic method that can determine which organism is causing the pneumonia. A bronchoscopy is sometimes used for this investigation (Ngari *et al.*, 2014). A thin, flexible, and lighted tube called a bronchoscope is passed down into the lungs. This enables the doctor to examine directly the infected parts of the airways and lungs. However,

sputum analysis has shortcoming such as poor sensitivity and is time-consuming (Cilloniz *et al.*, 2016).

1.6.4. Polymerase Chain Reaction (PCR) Assays

PCR is a molecular method that also includes quantitative (qPCR) and reverse transcriptase (RT-PCR) assays which have been developed to detect majority of pneumonia pathogens such as *Mycobacteria* and *Chlamydia trachomatis* in blood samples (Fischer *et al.*, 2001). Although research has shown the use of PCR for the diagnosis of *Streptococcus pneumoniae,* it has not been explored for commercial purpose. PCR is also prone to carry over contamination during manipulation and amplification of its products and it cannot provide anything other than supplemental data for the clinician; and there is an overwhelming lack of data on the efficacy of this method as a diagnostic tool for other bacterial pneumonia (Ahn *et al.*, 2016). This is coupled with the fact that bacteria are not the only causative agents of pneumonia.

1.6.5. Matrix-Assisted Laser Desorption Ionization Time of Flight

Matrix-assisted laser desorption ionization time of flight mass spectrometry has been introduced as a new microbiological diagnostic tool as a routine assay for microbial identification in laboratories (Wenzler *et al.*, 2016). The method is cheap, fast relies on the measurement of the molecular mass of all cellular proteins to determine the unique global protein profile that is characteristic of the pathogen (Mok *et al.*, 2016). There is an overwhelming lack of data on the efficacy of this method as a diagnostic tool for some bacteria including mycobacterial, with further limitation of an insufficient data for viral and fungal pneumonia pathogens (Kriegsmann *et al.*, 2016).

1.6.6. Immunoassay

Immunoassay such as enzyme-linked immunosorbent **assay** (**ELISA**) is carried out for the diagnosis of pneumococcal pneumonia using a receptor protein such as highly purified pneumolysin as an antigen for its serological detection using antibodies (Blanco-Covián *et al.*, 2017). However, this method is time-consuming, expensive and lacks sensitivity towards the detection of certain bacteria such as *Haemophilus influenza, Legionella pneumophila* and *Influenza* A virus (Schiffner-Rohe *et al.*, 2016).

1.7. Treatment of Pneumonia

The treatment of pneumonia depends on the type and severity of pneumonia. Bacterial types of pneumonia are usually treated with antibiotics while viral types have no definitive treatment but are usually treated with plenty of fluids, however, antiviral medications can be used in influenza (Loeffler *et al.*, 2001). Fungal types of pneumonia are usually treated with antifungal medications. Doctors commonly prescribe over-the-counter (OTC) medications to help manage the symptoms of pneumonia (Murdoch *et al.*, 2009). These include treatments for reducing fever, reducing aches and pains, and suppressing coughs. In addition, it is crucial to rest and drink plenty of fluids. Staying hydrated helps to thin out thick phlegm and mucus, making it easier to cough up. Hospitalization for pneumonia may be required if symptoms are especially bad or if an individual has a weakened immune system or other serious illnesses. In the hospital, patients are generally treated with intravenous antibiotics and fluids which may require a supplemental oxygen supply (O'Brien *et al.*, 2009). However, effective treatment is limited by sensitive and accurate diagnosis.

1.8.0. Use of Host-Mediated Biomarkers in Diagnosis and their Shortcomings

Till now, there is no universal definition of a biomarker, but it can be understood to be any biomolecule that is associated with a particular pathological or physiological state. Ideally, a biomarker should be one which cannot be detected or whose value is very low in the absence of inflammation; it should rise with increasing inflammatory processes and should decrease with resolving inflammation.

So far, five diagnostic biomarkers have been linked to pneumonia, namely: C-reactive protein (CRP), Procalcitonin (PCT), a Soluble triggering receptor expressed on myeloid cells-1 (STREM-1), CD163 and High Mobility Group Box-1(HMGB-1). It has also been revealed that CRP and PCT have been proven useful in diagnosis as it is produced in considerably high concentration but there is ambiguity in their specificity towards pneumonia as they can be produced in response to other inflammatory stimuli in neuron, atherosclerotic plaques, myocytes, and lymphocytes (Mandell, 1999); whilst, the mechanism regulating their syntheses at these sites is not clearly understood (Naughton *et al*., 2000 There are some other biomarkers which are still being studied for their probable use in pneumonia diagnosis; these include copeptin, cortisol, endotoxin, proadrenomedullin, amongst others.

1.8.1. C-Reactive Protein

CRP, even though being nonspecific, has proven to be helpful in establishing the etiology of some infections. A high CRP value (100 mg/L) can indicate a severe bacterial infection (Melbye and Stocks, 2006, Flanders *et al.*, 2004). *Streptococcus pneumoniae*, viruses, *Chlamydia pneumoniae*, *Mycoplasma pneumoniae*, *Legionella pneumophila*, and *Coxiella burnetii* were detected from bronchoalveolar lavage (BAL) culture of the patients and the serum CRP values revealed: (1) high CRP values in pneumonia patients with *S. pneumoniae*

or *L. pneumophila*; (2) increased CRP values with the severity of disease showing the use of CRP in pneumonia prognosis; (3) the use of CRP for the commencement of treatment procedure (Almirall *et al.*, 2004). CRP values are higher in certain pathogens than others and cannot differentiate viral from bacterial pneumonia (Garcia et al., 2003).

Also CRP does not allow for the distinguishing between all the types of pneumonia, its sensitivity is questionable. In a systematic review published in 2005, it was stated that "testing for C reactive protein is neither sufficiently sensitive to rule out nor sufficiently specific to rule in an infiltrate on chest radiograph and bacterial etiology of lower respiratory tract infection (van der Meer *et al.*, 2005). However, the advantage of CRP is that its levels can be used for the follow-up of pneumonia as well as to evaluate patient management and response to antibiotic therapy (Coelho *et al.*, 2007).

1.8.2. Procalcitonin

Procalcitonin (PCT) is synthesized from the thyroid cells and has a low serum concentration in healthy individuals. Increased CALC-I gene expression causes its release from parenchymal tissues and differentiated cell types of the liver and blood mononuclear cells during microbial infection (Chastre *et al.*, 2006). The inflammatory release of induced PCT through inflammation occurs due to the release of endotoxins by microbes and pro-inflammatory cytokines through a cell-mediated host response (e.g. interleukins 1b and 6, tumor necrosis factor-alpha). PCT has been used as biomarker for sepsis and other infection since it satisfies all the required criteria such as (1) high specificity and sensitivity; (2) ready measurability; (3) less expensive and ready availability; (4) reproducibility; and (5) half-life of 24 hours (Shehabi and Seppelt, 2008). However, reduced sensitivity of PCT is observed due to its elevated concentration during non-infectious conditions such as trauma, surgery, cardiogenic shock,

burns, heat stroke, acute respiratory distress syndrome, infectious necrosis after acute pancreatitis, and rejection after transplantation (Jung *et al.*, 2008).

Difficulty in diagnosis of pneumonia is due to its nonspecific symptoms in children and delay in treatment may have fatal outcomes in these patients (Shehabi and Seppelt, 2008). PCT sensitivity has been compared with CRP, interleukin-6 and interferon-alpha to detect possible bacterial CAP infection with PCT, showing high reliability more than others in the detection of CAP (Sager *et al.*, 2017). In another study, two years later, where PCT concentration was investigated in 101 children and the result showed that elevated PCT of over 1.0 ng/L can help in detecting bacterial pneumonia in children who are above 5 years of age (Korppi *et al.*, 2008). A study conducted by Franzin and Cabodi (2005) showed that PCT is increased in *Legionella pneumoniae* and, despite its non-specificity, it can be used as a prognostic marker. These results are in concordance with a study carried out by Haeuptle *et al.* (2009).

1.8.3. Soluble-Triggering Receptor Expressed on Myeloid Cells-1 (STREM-1)

Another biomarker with great potential for pneumonia detection is Triggering receptor expressed on myeloid cells-1 (TREM-1). TREM-1 is a member of the immunoglobin superfamily that takes part in inflammation. Its expression on neutrophils, monocytes, and macrophages during acute inflammation is through signal transduction of the transmembrane adapter protein DAP12 (Bouchon *et al.*, 2000). The concentration of TREM-1 increases during inflammation such as sepsis and ventilator acquired pneumonia (VAP) but not in noninfectious inflammatory conditions like psoriasis, ulcerative colitis, and vasculitis (Phua *et al.*, 2006). Tejera *et al.* (2007) found that serum STREM-1 is high in patients with CAP, and that the prognostic value of STREM-1 is independent of age, other inflammatory markers such as IL-6, pneumonia severity index, sepsis, and nutritional status. The study found that patients who

had increased STREM-1 levels on admission had the worst prognostic values.

1.9. Biomarkers used in prognosis

Many biomarkers such as endotoxin, proadrenomedullin, natriuretic peptides, endothelin-1 precursor peptides, copeptin and cortisol have been linked with the diagnosis and prognosis of pneumonia (Christ-Crain *et al.*, 2008). These biomarkers are useful in guiding culture sampling, empirical antibiotics prescription, following the clinical course of the condition and identify those who do not respond to therapy (Ito *et al.*, 2017). It is good to point out that while endotoxin is a diagnostic marker just like CRP, PCT, and sTREM-1, the other biomarkers mentioned (i.e., proadrenomedullin, natriuretic peptides, endothelin-1 precursor peptides, as well as copeptin and cortisol levels) have, up till now, only been found to be useful as prognostic markers and may be of great help in the risk stratification of patients.

1.9.1. Copeptin and Mid-Regional and Atrial-Natriuretic Peptide (MR-proANP)

Copeptin, and antidiuretic hormone arginine vasopressin (AVP) are derived from a pre-pro-vasopressin, which consists of a signal peptide, AVP, neurophysin II, and copeptin (Curbelo *et al.*, 2017). Copeptin is very stable when withdrawn from blood for days and its level can be measured quickly and easily (Curbelo *et al.*, 2017). It has been found that its level is significantly different between the two sexes, being higher in men than in women and also vary with exercise, fasting, and water load (Morgenthaler *et al.*, 2006). Müller *et al.*, 2007 found that copeptin has a tendency to increase as the severity of lower respiratory tract infection increases and this increase has been found to be more pronounced in patients with CAP. During investigation of the correlation of copeptin with serious sepsis in patients with VAP, Seligman *et al.* (2008) also found that copeptin increases progressively with the severity of sepsis.

Atrial natriuretic peptide (ANP) is synthesized from the cardiac atria and its mid-region MR-proANP (midregional pro-atrial natriuretic peptide) has been found to increase with severity of sepsis and VAP (Seligmann *et al*., 2008). The use of PCT together with Mid-regional Pro-atrial Natriuretic Peptide (MR-proANP) supports the differential diagnosis of pneumonia and congestive heart failure (Alan *et al.*, 2015). Both MR-proANP and CT-proAVP (C-terminal pro-vasopressin- Copeptin) can be used to predict the severity of disease in patients with CAP (Krüger *et al.*, 2007).

1.9.2. Cortisol

In early postoperative, post-trauma, or sepsis patients, serum cortisol level rises with the benefit of effective energy redistribution to the vital organs (Julián-Jiménez *et al.*, 2017). High cortisol levels were already known to be a good prognostic marker in sepsis. However, whether or not cortisol levels can be used as a marker of prognosis in CAP are still unknown. Evidence exists from the work of Christ-Crain *et al*., (2007) where 278 patients with CAP were found to have high cortisol levels during prognosis (Davi *et al.*, 2017, Nickler *et al.*, 2017).

1.9.3. Genetic and Proteomic Biomarkers

Genetic markers such as single nucleotide polymorphs (SNP) and haplotypes such as CD143 rs4340 have been linked to increased risk of community acquired pneumonia (CAP) (Solé-Violán *et al.*, 2010). Comparison of gene expression profiles of infected patients with systemic inflammatory response syndrome leading to sepsis, have been carried out with those whose systemic inflammatory response syndrome was due to other causes, confirming the use of genetic biomarkers in prognosis (Johnson *et al.*, 2007). These patients have distinguished inflammatory response of several hundred genes in either an up or downregulated manner due to infection from those whose respiratory syndrome was non-infective in origin. Despite evidence of differential gene expression in gram-negative and gram-positive sepsis in a murine

model (Yu *et al.*, 2004), gene expression profiles in critically ill patients with gram-positive and gram-negative sepsis do not differ (Tang *et al.*, 2008).

Genetic factor outcomes related to cytokines expression may be influenced where patients with pulmonary sepsis were found to have specific IL-10 haplotype production with significantly higher mortality than patients who had alternative haplotypes (Wattanathum *et al.*, 2005). Gene polymorphisms in human during regulated signal transduction events of the innate immune response, including Mal (Von Bernuth *et al.*, 2008), IRAK-4 (Ku *et al.*, 2007) and MyD88 (Khor *et al.*, 2007), can increase the risk of invasive bacterial pneumonia. Human gene mutations and polymorphisms of the complement systems (Walport, 2001), mannose-binding lectin systems (Roy et al., 2002), coagulation and fibrinolytic systems (Kerlin *et al.*, 2003, Opal, 2003) also alter the risk and prognosis of invasive bacterial pneumonia. Undoubtedly, more evidence of disease risk, treatment response and prognosis in CAP will become available as functional genomics transcriptomics become the standard of care in critical care units in the future.

1.9.4. Use of Biomarkers in Diagnostic Kits

Several biomarkers have been used for the development of diagnostic kits in pharmaceutical and biotechnology industries because they have advantages such as ease of operation, ready availability, portability and cost effectiveness (Cho *et al.*, 2013). The use of commercial immunoassay kits are used to support drug development but their validation can be laborious (Nerenberg *et al.*, 2018). Several studies have developed lateral flow devices for the detection of pathogens such as invasive Aspergillus (Thornton, 2008), malaria parasites (Tahar *et al.*, 2013) just to mention a few. Six commercially available immuno-chromatographic lateral flow device (LFD) kits were recently tested for rabies of the brain (Eggerbauer *et al.*, 2016). However, the only known research work where antimicrobial peptides for p24 detection,

discovered using HMMER (Tincho *et al.*, 2016) were validated for use within a lateral flow device in a specific and sensitive manner (Williams *et al.*, 2016). The details of the feasibility of the use of AMPs for in diagnostics is provided in the "future work" section of chapter seven.

1.10. Antimicrobial Peptides (AMPs)

Antimicrobial peptides (AMPs) are small molecular weight proteins which have broad spectrum antimicrobial action against bacteria, viruses and fungi (Torres *et al.*, 2019). These peptides have been conserved throughout evolution and are usually positively charged with both hydrophobic and hydrophilic side chains that enhance them to traverse aqueous environments and lipid-rich membranes. Once in a target microbial membrane, the peptide kills target cells through diverse mechanisms. (Demaria *et al.*, 2017). In addition, growing evidence also indicates that AMPs alter the host immune response through receptor-dependent interactions and have been shown to be important in such diverse functions as angiogenesis, wound healing, and chemotaxis (Roy *et al.*, 2011). As the knowledge of AMP biology expands, the precise role and relevance of these peptides will be better elucidated. These peptides interact with the host itself, triggering events that complement their role as antibiotics (Lai and Gallo, 2009).

Combined, the functions of the AMPs suggest that they are an important and previously underappreciated component of the human immunology. Several research work have been carried out on the diverse use of Antimicrobial peptides. Li *et al.* (2017b)worked on the effect of positive and negative amino acids substitution on the structure and biological activities of antimicrobial peptides after proteolytic digestion with trypsin which resulted in differential activities. Yang *et al.* (2018) provided structural insights and putative transport mechanism of a MacAB-like efflux pump in gram-positive bacteria. Wu *et al.* (2005) worked on the use of

antimicrobial peptides family of bombings as a potential source of drug candidates for anti-infection and anti-cancer therapy. Demaria *et al.* (2017) outlined the evidence linking antimicrobial peptides expression with a range of human inflammatory and auto-immune diseases as well as giving promising perspectives to endogenous antimicrobial peptides induction for treatment of diseases with certain druggable antimicrobial peptides undergoing clinical trials. Currently, Xuan (2018) worked on the synergistic mode of action of two antimicrobial peptides against *Staphylococcus cereus* and *Staphylococcus aureus* strains through permeabilization of the cell membrane causing cell death through influx and efflux of molecules across the membrane.

1.10.1. Categories of AMPs

AMPs have been divided on the basis of their amino acid composition and structures into sub-groups. The secondary structural architecture of these vital molecules include α-helical, β-stranded, β-hairpin or loop and extended structures. Their β strandedness is due to the presence of 2 or more disulfide bonds whilst the β-hairpin or loop is contributed by a single disulfide bond and/or cyclization of the peptide chain (Dhople *et al.*, 2006). Although in free solution, these peptides do not have defined structures but acquire a final configuration upon traversing the biological membranes. However, some AMPs do not belong to any of these aforementioned groups (McManus *et al.*, 2000) with some being composed of two different structural components (Uteng *et al.*, 2003). As mentioned, certain AMPs form active structures only during reaction with membranes of target cells. An example of this, is indolicin that shows a changed conformation during interaction with DNA, observable with reduced fluorescence intensity and a slight shift in wavelength of maximum emission (Hsu *et al.*, 2005), globular and amphipathic conformation in an aqueous solution and a wedge-shaped conformation in a lipid bilayer environment (Rozek *et al.*, 2000). Generally, the hydrophilic amino acid residues

of these peptides align along one side of their structure whilst their hydrophobic amino acid residues are formed along the opposite side of a helical molecule (Yeaman and Yount, 2003). These structural characteristics contribute to their amphipathicity that allows for partitioning into the membrane lipid bilayer, thereby enhancing their antimicrobial activities, thus impacting effective membrane permeabilization on many cytoplasmic targets (Hancock and Rozek, 2002).

1.10.2. Modes of Action of AMPs

Antimicrobial peptides have a varying mode of actions by which they kill target microbes (Nguyen *et al.*, 2011, O'Driscoll *et al.*, 2013). AMPs possess multiple activities including anti-gram-positive bacterial, anti-gram-negative bacterial, anti-fungal, anti-viral, anti-parasitic, and anticancer activities. These molecules interfere primarily with the cytoplasmic membrane of the target but can also interact with DNA, protein synthesis, protein folding, and cell wall synthesis Nguyen *et al*., 2011). For bacteria, the initial contact with the peptide is electrostatic, as most bacterial surfaces are anionic, or hydrophobic. Antimicrobial peptides attach to and insert into membrane bilayers forming pores by 'barrel-stave', 'carpet' or 'toroidal-pore' mechanisms due to their amino acid composition, amphipathicity, cationic charge, and size. They may also bind intracellular molecules which are crucial to cell living causing inhibition of cell wall synthesis, alteration of the cytoplasmic membrane, activation of autolysin, inhibition of DNA, RNA, and protein synthesis, and inhibition of certain enzymes (Brogden, 2005). For instance, buforin II can permeate cells and bind to DNA and RNA without destroying the cell membrane (Park *et al.*, 1998). Drosocin, pyrrhocoricin, and apidaecin are other examples of such AMPs. These AMPs have 18–20 amino acid residues with an active site for their intracellular target (Otvos *et al.*, 2000). Other AMPs can block cell wall synthesis such

as nicin, gramicidin S, and PGLa (Brumfitt *et al.*, 2002, Hartmann *et al.*, 2010). Generally, their antimicrobial activity is determined by measuring the minimal inhibitory concentration (MIC), which is the lowest concentration of drug that inhibits the growth of the pathogen (Meher *et al.*, 2017). It is therefore not surprising that AMPs have been used for their therapeutic abilities such as antibacterial diseases (Hartmann *et al.*, 2010) and diagnostic purposes (Williams *et al.*, 2016).

1.10.3. Techniques Used to Elucidate AMP Function

The exact mechanism by which antimicrobial peptides exact their effect remains unclear. Three major mechanisms by which they establish their actions include the formation of toroidal pore, carpet and barrel stave structures. Although the specificity of these mechanisms differs, they work by allowing peptide-induced membrane rupture and this gives rise to cytoplasmic leakage which ultimately leads to death of the target organisms. Many methods have been employed to evaluate the mechanisms of antimicrobial peptide activity (O'Driscoll *et al.*, 2013, Hein *et al.*, 2015). To this end, the use of dual polarization interferometry is becoming increasingly popular (Hirst *et al.*, 2011, Lee *et al.*, 2010). Other methods include solid-state NMR which has provided an atomic-level resolution of membrane destruction by antimicrobial peptides (Hallock *et al.*, 2003); X-ray crystallography has been employed to delineate how the family of plant defensins rupture membranes by identifying key phospholipids in the cell membranes of the pathogen (Järvå *et al.*, 2018, Poon *et al.*, 2014); Microscopy is used to visualize the effects of antimicrobial peptides on microbial cells (Brogden, 2005); Atomic emission spectroscopy is used to detect loss of intracellular potassium (an indication that bacterial membrane integrity has been compromised) (Cushnie and Lamb, 2005); Fluorescent dyes are used to measure the ability of antimicrobial peptides to permeabilize membrane vesicles

(Breukink *et al.*, 1999); Ion channel formation are used to assess the formation and stability of an antimicrobial-peptide-induced pore; and Circular dichroism and orientated circular dichroism are used to measure the orientation and secondary structure of an antimicrobial peptide bound to a lipid bilayer (Brogden, 2005).

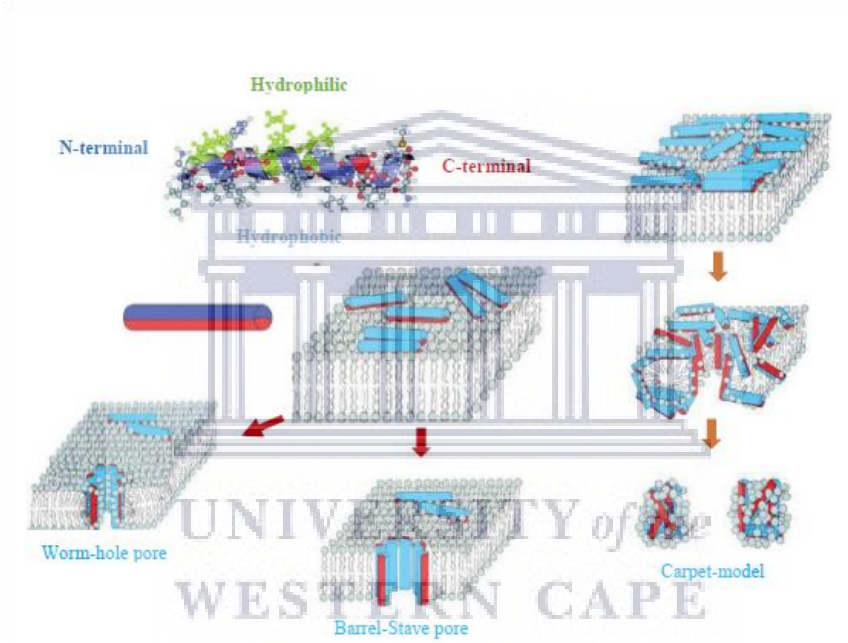


Figure 1. 2: Mechanisms of antimicrobial peptides. Several AMPs hold the promise to fold into amphipathic a-helices with hydrophilic and hydrophobic sides. This conformation is represented schematically as an amphiphilic cylinder, with hydrophobic (red) and hydrophilic (blue) halves. (centre) Antimicrobial peptides bind to the membrane surface with the hydrophobic side groups anchored in the hydrophobic lipid core of the bilayer, leading to different outcomes. (Bottom, left) The wormhole (or toroidal) pore model as proposed for magainin (Bottom, center) Barrel-stave pore model, as proposed for alamethicin, a peptide antibiotic produced by the soil fungus *Trichoderma viridis*. (Bottom, right) Carpet model: antimicrobial peptides crowd on the membrane's surface and lead subsequently to micellation. Modified from Salditt *et al.* (2006).

1.10.4. Natural and Artificial Production of AMPs

Most natural AMPs are synthesized by viruses and specific cells both within prokaryotic and eukaryotic organisms whilst; others are inducible. AMPs are produced either by ribosomal translation of mRNA through genetic coding in all life forms or by nonribosomal peptide synthesis mainly by bacteria (Hancock and Chapple, 1999). Certain eukaryotic cells such as lymph, epithelial cells in gastrointestinal and genitourinary systems (Ganz, 2003, Niyonsaba *et al.*, 2002), phagocytes (Hancock and Scott, 2000) and lymphocytes (Oppenheim *et al.*, 2003) of the immune system are involved in the production of AMPs. Apart from the production in innate immune response, AMPs can be produced by the influence of the host's inflammatory responses during an infection (Scott *et al.*, 2000). This is evident from the treatment of mammals with bacterial lipopolysaccharide (LPS) antibiotics for host immunity where AMP production was induced, reducing the inflammatory response (Hancock and Scott, 2000). Certain AMPs such as CAP18 (Larrick *et al.*, 1995), CAP35 (Brackett *et al.*, 1997), and a lactoferrin-derivative (Zhang *et al.*, 1999) block lipopolysaccharide (LPS)-induced cytokine release by macrophages. In contrast to this inflammatory regulation, antibiotics do not have this regulation on the inflammatory response of the host immune system; and lipopolysaccharide (LPS) secretion which makes them cause over-reaction of the host immune system, resulting in sepsis (Loppnow *et al.*, 1990).

It is relatively easy to modify the structure of AMPs, including library construction and screening, as well as immobilize AMPs on surfaces because of their amino acid composition (Costa *et al.*, 2011). This is necessary for modification of existing AMPs and designing new synthetic ones for a potential change of targets and improved stability against proteases (Papo *et al.*, 2002). Artificial synthesis is carried out by chemical synthesis using fully synthetic peptides (Wade *et al.*, 2012) or by using recombinant expression systems (Piers *et al.*, 1993).

Despite these advantageous features of AMPs, these methods have shortcomings due to inability to construct AMPs that are free from some challenges such as potential toxicity to humans (Matsuzaki, 2009), sensitivity to harsh environmental conditions such as proteases and extreme pH (Sieprawska-Lupa *et al.*, 2004), lack of selectivity against specific strains (Eckert *et al.*, 2006), high production costs (Bommarius *et al.*, 2010), folding issues of some large AMPs (Porto *et al.*, 2012), reduced activity when used for surface coating (Bagheri *et al.*, 2009), and bacterial resistance to some AMPs (Bader *et al.*, 2005). Hence there is a need to explore more sensitive options such as the use of *in silico* technologies that take into consideration size, charge, hydrophobicity, amphipathicity, and solubility to modify the activity and target spectrum of AMPs for the production of novel AMPs and mutagenesis of existing ones.

1.10.5. Computational tools used for Prediction of Antimicrobial Peptides

In an attempt to establish a new method for predicting antimicrobial peptides (AMPs) which are important to the immune system, new insights are being discovered, leading to a more sensitive and accurate methods for diseases diagnosis (Tincho *et al.*, 2016). Researchers are also interested in designing alternative drugs using AMPs because a large number of bacterial strains have become resistant to available antibiotics. However, obstacles have been encountered in the AMPs designing process as experiments to extract AMPs from protein sequences are costly and require a long set-up time.

A computational tool for AMPs prediction is currently being explored to resolve this problem (Hammami and Fliss, 2010). Effort is gradually being intensified to develop a high accuracy computational method which is able to predict these AMPs sequences effectively. Integrated algorithms have been introduced to predict AMPs such as integrated sequence alignment and support vector machine- (SVM-) LZ complexity pairwise algorithm. For this

reason, most researchers are concentrating on discovering new *in silico* tools for antimicrobial peptide prediction as computational approaches can accelerate the process of antimicrobial drug discovery and design.

Many computational methods have been introduced to predict AMPs based on the different features such as the binary pattern method using AntiBP server, which has been used to predict antibacterial peptides (Lata *et al.*, 2007), the CAMP methods (Thomas *et al.*, 2009), the combination of sequence alignment and the feature selection method (Wang *et al.*, 2011), and the pseudo amino acid composition method (Xiao *et al.*, 2013, Hajisharifi *et al.*, 2014). Other methods for predicting peptides include support vector machine (SVM), quantitative matrices (QM), and artificial neural network (ANN).

Unfortunately, AMPs have many variations in size with the aforementioned machine learning methods only working well on AMPs of fixed lengths (Lata *et al.*, 2007). For the CAMP methods (Thomas *et al.*, 2009), the AMPs prediction is performed using Random Forests (RF), SVM, and Discriminant Analysis (DA) based on all classes of full AMPs sequences. The sequence alignment method (Wang *et al.*, 2011) enjoys high prediction accuracy but it is not able to predict all sequences. This is because the classification concept used in the sequence alignment relies on High Scoring Segment Pairs (HSPs) scores which represent the similarity scores between two sequences using BLASTP. If the test sequence has no relationship with any training sequence, an HSP score cannot be generated; thus the classification concept cannot be performed on that particular sequence.

For the purpose of this study, HIMMER and CD-HIT will be discussed in more detail below as these two algorithms will be employed to identify novel AMPs to sensitively and specifically differentiate viral from bacterial pneumonia.

1.10.6. Hidden Markov Model (HMMER) as a prediction tool for AMPs

HMMER uses profile method for AMPs prediction in which each sequence is expressed as a set of similarities (probabilities) with a model or a family of sequences (Porto *et al.*, 2017). The construction of a profile is normally limited to the use of positive data (functional sequences) without discrimination. HMMER is frequently used for the statistical analysis of multiple DNA sequence alignments; to identify genomic features such as Open Reading Frames (ORFs); insertions, deletions, substitutions and protein domains amongst many others; to identify homologies, for example, the database of protein families (Altschul *et al.*, 1990, 1997). An important property of HMMER is the ability to capture information about the degree of conservation at various positions in a sequence alignment, and the varying degree to which insertions and deletions are permitted. This explains why HMMER can detect considerably more homologs than simple pairwise comparison (Liu *et al.*, 2017a). Large molecular structures, such as protein domains, are most often defined using the profile approach because of the incorporation of positions specific residues in such models (Waghu *et al.*, 2015). In general, whole proteins and proteins domains may have insertions at any sequence position exposed on the surface of the protein in many families.

The building of a probabilistic model using HIMMER requires the use of a training and a testing dataset. The training dataset is represented by three-quarters of the retrieved sequences while the remaining one-quarter represents the testing set. The retrieved sequences contain the activity of the class of AMPs for which a model is to be build. The activity for the class of AMP, for which a model is to be build, has been demonstrated using experimentation.

The Hidden Markov Models (HMMER) algorithm is then used to construct specific models/profiles using the training datasets (Xiao *et al.*, 2013). The testing dataset is used for validating the robustness of each profile created. This independent testing of a created profile

are performed in a step called “Query profiles”. The query of profiles is performed to confirm that the profile trained have the desired activity of the input training set (Hammer *et al.*, 2017). The use of the HMMER algorithm is deemed an appropriate tool, which enables a more sophisticated search for novel peptides through the proteome sequence scanning.

In the work of Tincho *et al.* (2016), a list of experimentally validated anti-HIV AMPs were retrieved and divided into the training and testing datasets, with HMMER being used to construct several species specific models for peptide discovery using proteome data. This study resulted in the identification of several AMPs that bound the HIV p24 protein and subsequently demonstrated using a lateral flow device employing these AMPs that detected both HIV 1 and 2 (Williams *et al.*, 2016).

1.10.7. Cluster Database at High Identity with Tolerance (CD-HIT) as a Tool for Prediction of AMPs

CD-HIT is an algorithm used for reducing homology bias and redundancy in gene or peptide sequences that have equal or greater pairwise identity sequences to any other in a subset (Wang *et al.*, 2017a). For AMPs, the bandwidth is set as an equivalent of 5 to the shortest sequence of the AMPs. The performance of CD-HIT is higher for redundancy removal as the number of AMP sequences increase. This redundancy-check task of CD-HIT is to ensure that each subset has enough samples for statistical analysis and to ensure that all categories (antifungal, antibacterial, antiviral and anticancer peptides) are covered (Fernandes *et al.*, 2012).

1.11. Rationale behind this Research Project

Pneumonia remains the leading infectious cause of death among children under five years of age, killing 2500 a day (Ginsburg *et al.*, 2015); recent research has also shown that mortality is higher in the elderly (Menendez *et al.*, 2004). Several biomarkers, which hitherto have been identified for its diagnosis lack specificity, as they could not differentiate viral from bacterial pathogens of

the disease; these biomarkers also fail to establish a distinction between pneumonia and other associated diseases such as pulmonary tuberculosis and Human Immunodeficiency Virus (HIV). There is a growing consensus, globally, of an improved understanding of the use of new biomarkers, which are produced in response to pneumonia infection for precision diagnosis to completely overcome these aforementioned limitations (Chopra *et al.*, 2013). In line with this, the discovery of more biomarkers is therefore imperative, using bioinformatics for example, for more accurate diagnosis of pneumonia. In addition, the use of AMPs rather than antibodies to bind and recognize receptors for diagnostics, offer several compensatory advantages over shortcomings such as non-specificity and cost to mention only a few. With more accurate and sensitive diagnostics of pneumonia, the problem of indiscriminate overuse of antibiotics, toxicity due to the wrong prescription, bacterial resistance, scarcity and the high cost of existing antibiotics can be reduced or even eliminated. The use of AMPs for HIV diagnostics on patient samples has been demonstrated by Williams *et al.* (2016), with a commercial kit currently in production.

1.12. Significance of this Study

The research would have an impact on the lives of South Africans and other Africans through modelling AMPs for potential facilitation of pneumonia diagnosis in an easy, fast, and sensitive manner. It would also provide information for an improved awareness and understanding of the mechanism of AMPs that would eventually reduce the incidence of pneumonia by optimizing the use of antibiotics. There is a promising perspective to the discovery of new biomarkers after the completion of the future work of this study which would facilitate the development of collaborations with industries for sustainable investment to develop the technologies developed from the project that offer the prospect of job creation and economic growth in South Africa. Training of post-graduate students would also take effect in a scarce-skilled field such as Bioinformatics.

1.13. Limitations and Strength of this study

Designing and modelling of novel AMPs for therapeutics and diagnostics is an active area of research explored to circumvent the shortcomings of the conventional antibiotics and the non-specificity of the current diagnostic and prognostic biomarkers (Veltri *et al.*, 2018). One of the strengths of this research work is that it would offer an insight into the modular architecture of AMPs using *in silico* analysis for potential pneumonia diagnosis. However, despite the success of HMMER in the prediction of AMPs, it still suffers some limitations in its model prediction. One of these limitations is the higher level information or correlation with the amino acid residues of AMPs which is difficult to capture by this tool because of the linear nature of HMMER (Yuan, 2018). An example of the higher level information is the prediction of the actual distance between the folding of proteins and its spreading out as well as the prediction between the electrical and chemical interactions (Tamposis *et al.*, 2019), necessitating combination of tools to address the inherent problems associated with the tools. Another limitation is the low sensitivity of HMMER to the use of small datasets because of the available number of AMPs in the databases to specific targets, justifying the combination of tools. Apart from this, one of the drawbacks reported for the use of AMPs is their susceptibility to proteolytic degradation due to the presence of L-amino acids, the target of proteases (Sinha and Shukla, 2019). However, the putative AMPs discovered from this analysis would be used for the *in vitro* construction of a lateral flow device, thus circumventing this drawback. The incorporation of certain amino acid residues during site directed mutagenesis analysis accounts for the changes observed in the binding potential as well as improved efficacy of the anti-pneumonia putative AMPs towards their targets respectively.

1.14. Aims

The aim of this research work was to generate data for the identification of novel AMPs, using *in silico* techniques, to accurately differentiate between viral and bacterial associated pneumonia.

1.15. Objectives

The objectives of this research work were to:

1. Identify novel anti-pneumonia AMPs using HMMER and CD-HIT algorithms;
2. Identify some bacterial and viral receptors as targets for the putative AMPs;
3. Predict the 3-D structure of the identified anti-pneumonia AMPs using in-silico methods;
4. *In-silico* binding studies of the anti-pneumonia AMPs and pneumonia proteins; and
5. Perform a site-directed mutagenesis analyses of the anti-pneumonia AMPs and protein receptors for improved performance.

Chapter 2

2.0. *In silico* Identification of Putative Anti-Pneumonia Antimicrobial Peptides (AMPs)

2.1. Introduction

Several *in silico* approaches to identify AMPs have been established and promised to be less time consuming, more cost effective and less labor intensive thus speeding up the discovery process. These qualitative and quantitative activity prediction tools have proven useful in the biological experimental design to identify AMPs including tools such as Support Vector Machine (SVM) (Thomas *et al.*, 2009), profile Hidden Markov Models (HMMER) (Fjell *et al.*, 2007, Brahmachary *et al.*, 2004), Gap Local Alignment of Motifs 2 (GLAM2) (Fjell *et al.*, 2009), Quantitative Structure Activity Relationship (QSAR) (Torrent *et al.*, 2011), Linear Discriminant Analysis (LD), Random Forest (RF) (Thomas *et al.*, 2009) and Sliding Window (SW) (Juretić *et al.*, 2011). Such approaches have allowed for the systematic mining of genomic expressed sequence tag data, with the aim of discovering undescribed natural AMP sequences (Eddy, 1998).

Furthermore, the growing interest in system biology approaches towards the discovery of new biomarkers offers compensatory advantages such as early diagnosis and prognosis, classification of disease subtypes, prediction of treatment response, and identification of potential targets for drug therapy (Iorio and Croce, 2012). In line with the growing problems of diagnosis of diseases and difficulty of resistance to conventional antibiotics, more robust laboratory and pharmacological applications of antimicrobial peptides can be achieved with the use of *in silico* technologies to develop AMP blueprints in order to speed up their discovery. Also, *in silico* technologies offer potential promise of increasing the potency and specificity of the AMPs so that they are specific to target pathogens, despite a broad spectrum of these peptides being identified and discussed in the literature (Randou *et al.*, 2013).

2.2. Data Mining

Data mining is a process of extracting information from databases and transforming the information into a comprehensible structure for further use. This can be followed with the use of specific machine learning and statistical tools to predict and discover patterns within the extracted data (Wang *et al.*, 2005). Furthermore, data mining is used to estimate the value of a particular target attribute where data is divided into groups without prior group knowledge for pattern discovery (Tang and Zhang, 2013). The words "antimicrobial peptide databases" are searched on Google and the relevant database is opened to retrieve the experimentally validated AMPs specific for the pathogens of interest.

2.2.1. Antimicrobial Peptides Databases

2.2.1.1. Collection of Antimicrobial Peptides Database (CAMP)

CAMP (http://www.camp.bicnirrh.res.in) has been designed to improve the understanding of antimicrobial peptide family-based studies since it incorporates many tools relevant for discovery and design of AMPs. This database contains information on the conserved sequence signatures captured as patterns and Hidden Markov Models (HMMs) for 1386 AMPs represented by 45 families (Swart, 2012). Information related to sequence, protein definition, accession numbers, activity, source organism, target organisms, protein family descriptions and links to databases like UniProt, PubMed, and other antimicrobial peptide databases are provided for the benefit of users. The database also provides tools for sequence alignment, pattern creation, pattern and HMMER-based searches for novel AMPs identification (Senthilkumar and Rajasekaran, 2016).

2.2.1.2. Antimicrobial peptides Database (APD)

The APD (http://aps.unmc.edu/AP/main.html) focuses only on natural AMPs with known amino acid sequence, biological activity and long amino acid peptides. It has an easy-to-search interface that can sort information based on net charge, length and hydrophobic content. The APD contains 2169 anti-bacterial, 172 anti-viral, 105 anti-HIV, 959 antifungal, 80 anti-parasitic and 185 anti-cancer peptides (Wang *et al.*, 2016).

2.2.1.3. Antiviral Peptides Database (AVPdb)

AVPdb (http://www.crdd.osdd.net/servers/avpdb/) is the first comprehensive database of antiviral peptides (AVPs) targeting about 60 medically important viruses (Qureshi *et al.*, 2013). It is a manually curated, open source archive having 2683 experimentally verified AVPs including 624 modified antiviral peptides (Polanco *et al.*, 2014). Exhaustive information on AVPs like sequence, source, targeted virus, cell line, assays, experimental efficacy (qualitative/quantitative) and references is provided. Also incorporated are the physicochemical properties of the peptides and their predicted structures. This database is easy to operate through browsing and field / advanced/rule-based searches with sorting and filtering functions. It incorporates analytical tools such as Basic Local Alignment Search Tool (BLAST), MAP as well as links to major peptide resources (Sencanski *et al.*, 2015).

2.3. Text Mining

Text mining is the process of deriving high-quality information from text, which is achieved by using statistics for patterns and trends. This involves using the input text for deriving patterns within the structured data, while evaluating and interpreting the output. A well-established text mining approach gives relevance, novelty and interest to the output result (Aggarwal and Zhai, 2012). The key words for HMMER include “hmmcluster”, “hmmbuild”,

"hmmcalibrate" and "hmmsearch" and these would be explained further in the methodology section.

2.4. *In silico* tool for Peptide Discovery

2.4.1. Hidden Markov Models (HMMER)

The Hidden Markov Models (HMMER) algorithm is a probabilistic system designed to model sequences by analyzing and proposing an evolutionary information contained in their profiles and subsequently enriching them with an increased performance using computer programming language in the form of command lines. HMMER only requires the sequences of experimentally validated biomolecules for the construction of models and will display their features based on the motifs of the input sequences with desired activity against the specific target (Fjell *et al.*, 2012, Boman, 1995). The high accuracy of the HMMER profiles is due to the combination of the scoring system and the low cut-off E-value calculated during the proteome sequence scanning step of the algorithm. The calculated E-value provides accurate information about the probability of the predicted biomolecules to be true a positive or a false negative (Fjell *et al.*, 2009, Brahmachary *et al.*, 2004, Boman, 1995). Several AMPs discovered through this process (Tincho *et al.*, 2016), have been proven to detect certain diseases such as HIV (Williams *et al.*, 2016).

2.5.1. Aim

The aim of this chapter is to use HMMER for prediction of novel AMPs with high sensitivity, accuracy and specificity for potential differential detection of viral and bacterial pneumonia.

2.5.2. Objectives of this chapter

- Retrieve experimentally validated anti-pneumonia AMPs from various databases;
- Build predictive profiles using the HMMER software using retrieved anti-pneumonia AMPs as input;

- Test the created profiles using the testing sets (positive and negative) and calculate the performance measures of the created profiles using several parameters such as accuracy, specificity, sensitivity and Mathew's correlation coefficients.
- Identify novel putative anti-pneumonia (AP) AMPs by scanning the created profiles against various proteome sequences; and
- Determine the physicochemical properties of the identified peptides.

2.6. Materials and Methods

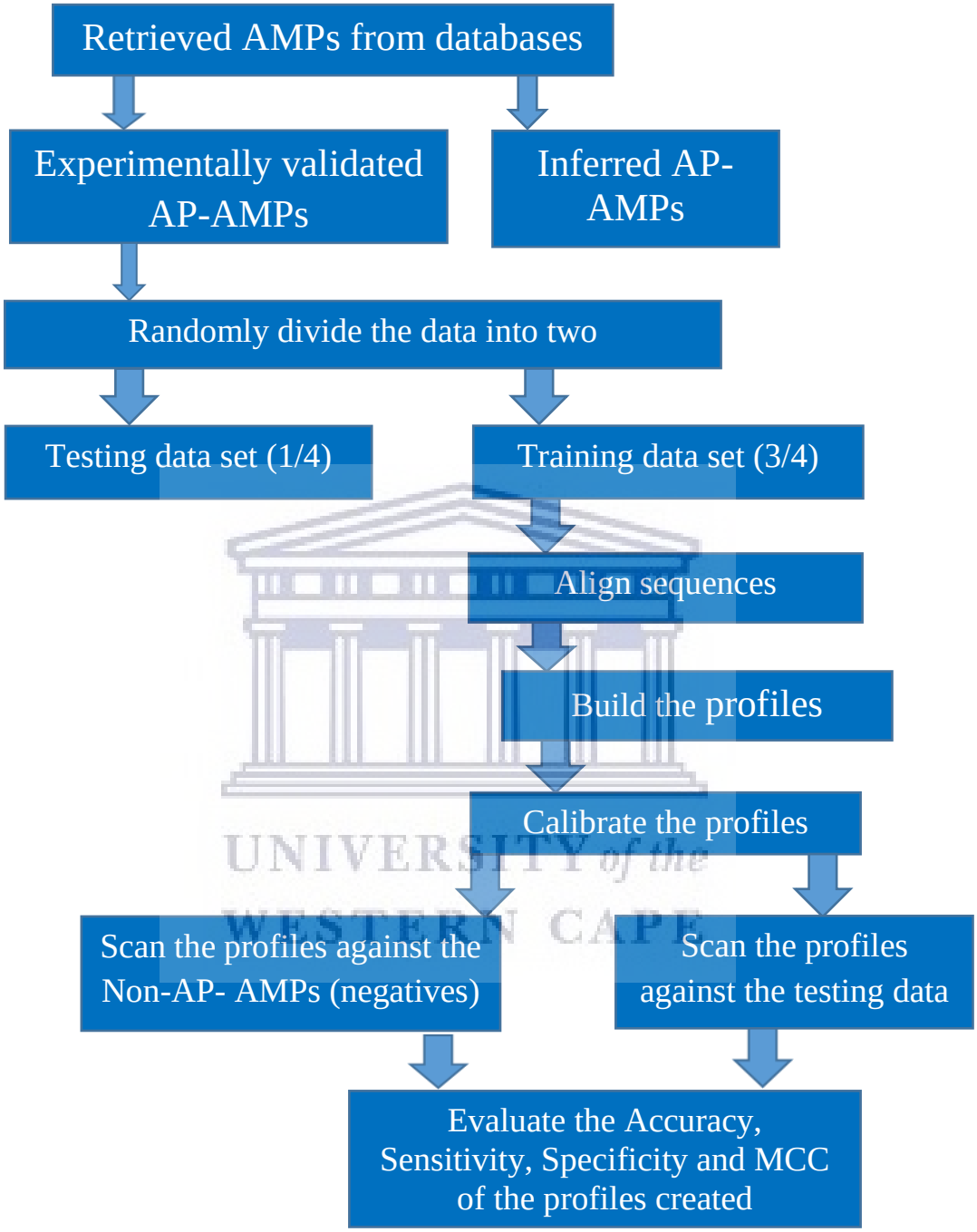


Figure 2. 1: Schematic representation of the methodology used for creation of various profiles using HMMER.

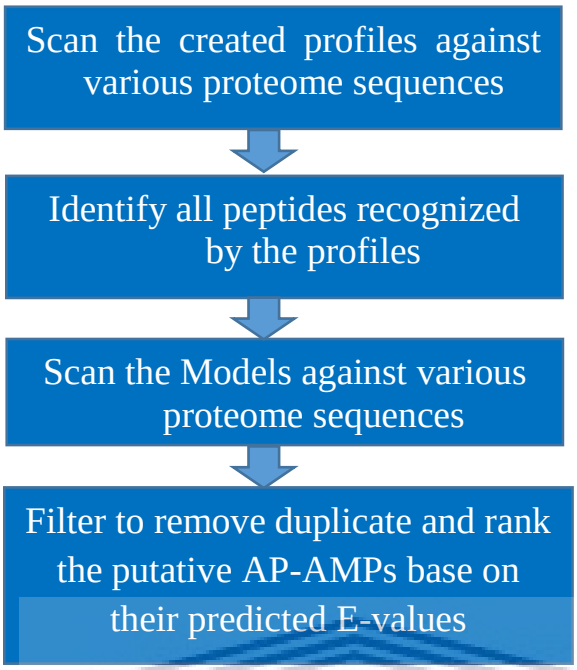


Figure 2. 2: Diagrammatic representation of the discovery of potential Anti-Pneumonia (AP) AMPs through proteome scanning

2.6.1. Data retrieval

The experimentally validated anti-pneumonia AMPs for the bacterial causative agents (*Streptococcus pneumoniae*, *Acinetobacter baumannii* and *Klebsiella pneumoniae*) and viral causative agents (*Respiratory syncytial virus*, *Influenza A* and *B* viruses) were retrieved from antimicrobial peptide databases such as Antimicrobial Peptides Database (APD) (Wang and Wang, 2004, Wang *et al.*, 2008), Collection of Antimicrobial Peptides (CAMP) (Thomas *et al.*, 2009), and Antiviral peptides databases (AVPDB) (Sencanski *et al.*, 2015). Thereafter, curation was carried out through literature mining to confirm that all the retrieved AMPs were either experimentally validated or predicted. Duplicate experimentally validated AMPs were then discarded from the retrieved list using the Cluster Database at High Identity with Tolerance (CD-HIT) (Li and Godzik, 2006).

2.6.2. Training and testing datasets

After retrieval of the experimentally validated AMPs as indicated in section 2.6.1, the AMPs were categorized according to their activities against specific pathogenic agents with AB denoting anti-*Acinetobacter baumannii*; SP-anti-*Streptococcus pneumoniae*; KP-anti-*Klebsiella pneumoniae*; INFA-anti-*Influenza A*; INFB-anti-*Influenza B*; and RSV-anti-*Respiratory Syncytial Virus* (Mulvenna *et al.*, 2006, Gennaro and Zanetti, 2000, Vizioli and Salzet, 2002). Each category of the aforementioned datasets was randomly divided into two portions: three-quarters of each dataset was utilized as the training set (to build each profile), whilst one-quarter was used as the testing dataset (for optimization/calibration of the created profiles).

2.6.3. Construction of AMPs profiles

The profile Hidden Markov Models (HMMER) algorithm version 2.3.2 (Potter *et al.*, 2018) was used to construct specific pathogen-targeted models/profiles using the respective training datasets indicated in section 2.6.2. All the HMMER profiles were built on the Ubuntu 12.04 LTS operating system. The task was accomplished on a terminal and the command lines used to build each profile was written in accordance with the corresponding algorithm and the steps involved in their construction were as below:

For the first step, the training datasets of each target class were aligned using the ClustalW alignment tool (Thompson *et al.*, 1994). The alignment was carried out using the command line:

```
clustalw–align–output=gcg–case=upper–sequos=off–outorder=aligned–
infile=targetclass.fasta ……(i)
```

The command line simply states <<do an alignment of the sequences which are in the upper case found in the input file "target class.fasta" with the FastA, using ClustalW as multiple

alignment tools and GCG Postscript output for graphical printing>>. The output of the command results in the construction of aligned sequences, called “target class.msf”. The aligned sequences was used as input in the next step.

The next step enhances the construction of the profiles of the target class sequences by showing the common motifs/signatures within the profiles. To achieve this, the “Build profiles” was run using the following command:

hmmbuild target class profile.hmm target class profile.msf……(ii)

To enhance the sensitivity of the profiles, the file generated (target class. hmm) from the profile building step was calibrated by using the command line:

hmmcalibrate target class profile.hmm ……(iii)

The resulting profiles “target class.hmm” was used in evaluating the profiles performance by testing the created profiles on an independent AMP dataset.

2.6.4. Independent profile testing

The independent testing of each created profile was performed in a step called “Query profiles”. The testing dataset were queried against the created profiles using the command line, with an E-value threshold of 95% or 0.05:

hmmsearch–E 5e-2 target class profile.hmm target class profile query.txt>resultfile.txt …… ……..(iv)

2.6.5. Performance measurement prediction of the profiles using the testing datasets

Statistical performance measures were then calculated using sensitivity, specificity, accuracy and Mathew Correlation Coefficient as parameters. The measures used are described as follows:

Sensitivity is the percentage of anti-pneumonia AMPs against a particular pathogen (testing sets) correctly predicted as anti-pneumonia AMPs (positive). The sensitivity (recall) is defined in equation (1):

$$\text{Sensitivity} = \left(\frac{TP}{TP+FN}\right)\times 100 \quad (1)$$

Specificity is the percentage of non-anti-pneumonia AMPs (negative sets) correctly predicted as non-anti-pneumonia AMPs (negative). The specificity is defined in equation (2):

$$\text{Specificity} = \left(\frac{TN}{TN+FP}\right)\times 100 \quad (2)$$

Accuracy is the percentage of correctly predicted peptides (anti-pneumonia AMPs and non-anti-pneumonia AMPs). The accuracy is defined in equation (3):

$$\text{Accuracy} = \left(\frac{TP+TN}{TP+FP+TN+FN}\right)\times 100 \quad (3)$$

Mathew's correlation coefficient (MCC) is a measure of both sensitivity and specificity. MCC=0 indicates completely random prediction, while MCC=1 indicates perfect prediction. It is defined in equation (4):

$$\text{MCC} = \left(\frac{(TP\times TN)-(\text{FN}\times\text{FP})}{\sqrt{(TP+FN)\times(\text{TN}+\text{FP})\times(\text{TP}+\text{FP})\times(\text{TN}+\text{FN})}}\right) \quad (4)$$

2.6.6. Identification of novel putative anti-Pneumonia AMPs from proteome sequences

Proteome sequences were queried by the respective profiles with the list of all proteome sequences (in the fasta format) retrieved, from the Ensembl database (http:// www.ensembl.org/index.html) and the UniProt database (http://www.uniprot.org/). A cut-off

E-value was set to be 0.05 for the search of putative anti-pneumonia AMPs. This was accomplished using the "hmmsearch" module of the HMMER algorithm with the command line employed stated below:

hmmsearch–E 5e-2 target class.hmm target pathogen query.txt>resultfile.txt ….....……(v)

Where specific target class.hmm in one of the six profiles, target class query.txt representing the species scanned against the profile and resultfile.txt is the result file saved after querying that species against a particular pathogen profile.

2.6.7. Final List of anti-Pneumonia AMPs

The final list of putative anti-pneumonia AMPs were retrieved following the proteome scanning using BLASTp analysis of the Uniprot database (https://www.uniprot.org/blast/) (Wu *et al.*, 2006).

2.6.8. Physicochemical properties of the putative anti-Pneumonia AMPs

Physicochemical properties of the putative anti-pneumonia AMPs were calculated using the prediction interface of Bactibase (http://bactibase.pfba-lab-tun.org/physicochem) (Hammami *et al.*, 2007, Hammami *et al.*, 2010) and APD ((http://aps.unmc.edu/AP/design/design_improve.php).) (Wang and Wang, 2004, Wang *et al.*, 2008) using the amino acid sequences of the putative peptides as input.

2.7. Results and Discussion

2.7.1. Retrieval of AMPs

Anti-bacterial AMPs (BAP-AMPs)

In this section, experimentally validated AMPs were retrieved from various databases where literature mining revealed that CAMP, APD and BACTIBASE had 155, 9 and 4 experimentally validated bacterial anti-pneumonia antimicrobial peptides (BAP-AMPs) respectively. BAP-AMPs against the pathogens *Klebsiella pneumonia* totalled 140 peptides, *Streptococcus pneumoniae* totalled 16, and *Acinetobacter baumannii* totalled 12 peptides combined from the various databases (see table 2.1).

These experimentally validated anti-pneumonia AMPs were derived from bacteria, arthropods, Mammalia Amphibia, nematode, Pisces, Arachnida, aves, plants, reptilian, fungi, and viruses with the numbers for each division indicated in table 2.1. The peptide total in table 2.1 represents only unique peptides for each organism following removal of duplicates.

Anti-viral AMPs (VAP-AMPs)

Literature mining revealed 176 experimentally validated anti-viral pneumonia antimicrobial peptides (VAP-AMPs) combined for the CAMP, APD and AVPDB databases against the pathogens *Respiratory Syncytial Virus, Influenza A and B* in the order 112, 52 and 12 respectively. These AMPs were derived from Mammalia, plants, amphibian Arthropoda and virus.

Proof of AMPs Activity

Only experimentally validated AMPs were retrieved from the various databases since they have proven activity against the target pathogens with minimum inhibitory concentration (MIC) as indicator using the agar dilution or broth microdilution methods (Kim *et al.*, 2012). Experimentally validated AMPs were used since their activities have been proven. These

activities would be retained in the newly retrieved AMPs since HMMER creates a profile by retaining the functionally significant amino acid residues.

Table 2 1: Retrieval and Diversity of the experimentally validated AMPs from the databases against the pneumonia pathogens

S/N	Micro-organisms	Databases	Division
1	*Klepsiella pneumoniae*	CAMP = 131 APD = 6 BACTIBASE = 3	Bacteria=68, Arthropoda=16, Mammalia=19, Amphibia=28, Nematoda=1, Fish=2, Arachnida=1, Aves=3, Plant=1. Reptilia=1
2	*Streptococcus Pneumoniae*	CAMP = 14 APD = 1 BACTIBASE = 1	Bacteria=5, Amphibia=2, Fungus=2, Mammalia=4, Arthropoda=1, Fish=1, Plant=1
3	*Acetobacter baumannii*	CAMP=10 APD=2	Coelentrata=1, Bacteria=8, Arthropoda=2, Mammalia=1
4	*Respiratory Syncytial Virus*	APD = 28 AVPDB = 84	Mammalia=3, Virus=88
5	*Influenza type A*	CAMP = 1, APD = 2, AVPDB = 49	Plant=1, Arthropoda=2, Amphibia=1, Virus=48
6	*Influenza type B*	AVPDB=12	Virus=12

The list of anti-pneumonia AMPs was retained within their respective pathogenic target groups as retrieved from the various databases to allow for specific species/pathogen profile creation.

2.7.2. Profile creation using HMMER

The first step in the profile creation pipeline was the random grouping of the different classes into ¾ and ¼ of the experimentally validated AMPs. The ¾ was used as the training dataset, for which multiple alignments were generated using HMMER ClustalW. This step prevents the profile being sensitive to small misalignments and to report reliable E-values in order to capture the diversity of the sequences since the AMPs were derived from different organisms (Madera and Gough, 2002)

A total of six AMP profiles were created for each of the following classes (anti-*Streptococcus pneumoniae* (SP), anti-*Acinetobacter baumannii* (AB) *and* anti-*Klebsiella pneumoniae* (KP)), (anti-*Respiratory syncytial virus* (RSVM) and, anti-*Influenza A* and *B* (INFA and INFB). See table 2.2 for the number of AMPs used to create each profile.

Table 2. 2: Profile creation by HMMER

Profiles	Training Datasets	Testing Datasets	Total
AB	9	3	12
SP	12	4	16
KP	105	35	140
INFB	9	3	12
INFA	39	13	52
RSVM	84	28	112

Legend: AB: Anti-*Acinetobacter baumannii*; SP: Anti-*Streptococcus pneumoniae*; KP: Anti-*Klebsiella pneumoniae*; INFA: Anti-*Influenza A virus*; INFB: Anti-*Influenza B virus*; and RSVM: Anti-*Respiratory Syncytial Virus.*

2.7.3. Independent testing of the created profiles

Positive Datasets

Each created profile was tested against a positive dataset which represented about a quarter of the the retrieved experimentally validated AMPs (see table 2.2), from which the training dataset used for the construction of the respective profiles was derived as well. The profiles created using the training dataset must have the ability to recognize and identify this subset of AMPs i.e. testing dataset. Since experimentally verified AMPs were used, the assumption is that the profiles constructed should be able to identify sequences with the exact same activity and discriminate those that do not have this activity. The strength of the created profiles lies in its high discriminatory power to identify novel sequences.

Negative Datasets

The trained profiles were scanned against a negative control dataset, made up of random fragments of 17236 neuropeptides, which had no recorded anti-pneumonia activity.

Independent testing of the created profiles was done with the negative dataset (neuro-peptides) to determine whether the trained profiles would discriminate against non-anti-pneumonia peptides i.e. the profiles should thus not select any of these peptides within the negative dataset. This method of taking random sequences as a negative dataset is a routinely used procedure (Sharma *et al.*, 2013) and this is based on the assumption that probability of finding random sequences with a highly discriminative profile is very low.

Evaluation of the Independent testing results

The independent testing of the profiles was carried out using the true positive (TP), false positive (FP), true negative (TN) and false negative (FN) measures. A cut-off E-value of 0.05 was applied to the HMMER algorithm to strengthen the ability of the profile to discriminate between the true positive anti-pneumonia AMP and false negative anti-pneumonia AMPs.

TP (True positive) represents correctly predicted positive sequences (anti-pneumonia AMPs), TN (True negative) denotes correctly predicted negative sequences (non-anti-pneumonia AMPs), FP (False positive) is the number of non-anti-pneumonia AMPs wrongly predicted as anti-pneumonia AMPs (AP-AMPs), FN (False negative) is the number of anti-pneumonia AMPs wrongly predicted as non-anti-pneumonia AMPs. It was possible to calculate the number of TP AMPs from the total number of input sequences, thus the FP number could be extrapolated with the results shown in Table 2.3, reflecting the capacity of each profile to distinguish a true anti-pneumonia AMPs from false anti-pneumonia AMPs.

Anti-bacterial Profiles

In table 2.3 below, the AB profile identified all sequences within its testing dataset as true positives, the KP profile identified 10 out of the 35 sequences within its testing dataset as true positive while the SP profile identified one of the four sequences within its testing dataset as true positive. All the anti-bacterial profiles discriminated against the

neuropeptides as was expected identifying all of the sequences as non-AMPs. The lower hits observed for the KP and SP profiles is as a result of an overlap of homologous relationship in the AMPs used in the creation of their respective profiles (Conlon *et al.*, 2004).

HMMER used a default E value of 0.05 for every hit considered to be true positive. All anti-bacterial profiles yielded true positives with E-values lower than 0.05 indicating that there was only a 5% chance that the hit was false or random, i.e. true positive with an E value less than 0.05 is considered ideal, indicating an extremely high prediction confidence.

Anti-viral Profiles

In table 2.3 below, the INFB profile identified all of the sequences within its testing dataset as true positive while the RSVM profile identified 22 out of the 28 sequences within its testing dataset as true positive. The INFA profile identified 6 out of 13 sequences within its testing datasets as true positive which could also be due to an overlap of homologous relationship in the AMPs used in the creation of the profiles as well related to the number of input sequences to create these profiles (Conlon *et al.*, 2004).. All anti-viral profiles discriminated against the neuropeptides.

The evaluation of independent testing was also determined using E-value and "hits" from positive and negative datasets as measured. From the results, HMMER assigned E-values for the true positives that were hit by the profiles. The anti-viral profiles had hit with significant E-value less than 0.05 which are considered ideal, indicating the likelihood of the profiles giving false "hits' and positive "hits" at an extremely high confidence.

Table 2. 3*:* Independent testing of profiles against test and negative datasets

Profiles	TP	FN	TN	FP
AB	3	0	17236	0
SP	1	3	17236	0
KP	10	25	17236	0
INFB	3	0	17236	0
INFA	6	7	17236	0
RSVM	22	6	17236	0

Legend: AB: Anti-*Acinetobacter baumannii*; SP: Anti-*Streptococcus pneumoniae*; KP: Anti-*Klebsiella pneumoniae*; INFA: Anti-*Influenza A virus*; INFB: Anti-*Influenza B virus*; and RSVM: Anti-*Respiratory Syncytial Virus.*

2.7.4. Performance measurement of the target specific profiles

After evaluating the ability of the optimized profiles, the performance was calculated with the aim of determining the robustness of each profile, using specificity, sensitivity, accuracy and Matthew's Correlation Coefficient (MCC), introduced by biochemist Brian W. Matthews in 1975 (Meher *et al.*, 2017). The sensitivity, specificity, accuracy and MCC were calculated as reported in Table 2.4.

Table 2. 4: Summary of Performance measurement of the profiles

Profiles	Sensitivity (%)	Specificity (%)	Accuracy (%)	MCC
AB	100	100	100	1
SP	25	100	99.6	0.50
KP	40	100	99.9	0.63
INFB	100	100	100	1
INFA	46	100	99.1	0.68
RSVM	79	100	99.96	0.89

Legend: AB: Anti-*Acinetobacter baumannii*; SP: Anti-*Streptococcus pneumoniae*; KP: Anti-*Klebsiella pneumoniae*; INFA: Anti-*Influenza A virus*; INFB: Anti-*Influenza B virus*; and RSVM: Anti-*Respiratory Syncytial Virus.*

From the result shown in table 2.4, sensitivity values were high in AB, INFB and RSVM of anti-bacterial and anti-viral profiles tested. KP and INFA had 40% and 46% respectively while there was a relatively low sensitivity observed for the SP profile. All profiles showed good values for specificity and accuracy with significant MCC values. The MCC is considered to give the best performance measurement of models since it incorporates sensitivity, specificity and accuracy (Łuksza *et al.*, 2017).

Specificity

The specificity results for all profiles were 100%. Specificity estimates the true negative rate by calculating the proportion of the negative datasets that were correctly predicted as negative (Tyrer *et al.*, 2015). A poor specificity results when AMP sequences within a profile have closely related functions with sequences within another profile. The specificity results for the profiles thus indicated that their sequences do not have closely related functions with sequences making up any of the other profiles.

Sensitivity

High sensitivity values of INFB, AB and RSVM profiles showed correct prediction. The moderate sensitivity of KP and INFA could be attributed to the significant overlap in the conserved domain of the AMPs used for their profile construction (Liu *et al.*, 2017a). However, the relatively low sensitivity of SP could be due to low available AMPs against *Streptococcus pneumoniae* from the databases coupled with serious overlap in the conserved consensus sequences (Southey *et al.*, 2006). Sensitivity is the measure of the true positive rate which represents the ratio of the correctly predicted positive datasets to the total number of the positive datasets (Lázár *et al.*, 2018). Low or moderate sensitivity is only possible where a false-positive might score higher than a true-positive. In this case, however, the model discriminated against all false-positive with some true-positive sequences within the KP, INFA and SP profiles behaving as true-negative, further strengthening the argument of overlapping consensus in the conserved domains of these sequences.

Accuracy

The profiles showed very significant accuracy results. Accuracy is a commonly used predictive profile parameter to reduce errors by establishing number of misclassified AMPs from both positive and negative datasets. Accuracy considers interactions between features of AMPs by

discriminating them against non-AMPs, even from similar accuracy values in other machine learning methods (Fernandes *et al.*, 2012). The accuracy results were very high for all the profiles indicating elimination of errors by nullifying misclassified AMPs from both positive and negative datasets.

Mathew's Correlation Coefficient (MCC)

MCC values for all the profiles showed very significant results with the lowest value recorded for SP (0.50). The MCC value '0.5 to 1' corresponds to the perfect prediction, whereas '0' points to a completely random prediction. Thus all profiles indicated correct prediction (AB = INFB > RSVM > INFA > KP > SP). MCC is considered as the most robust parameter for evaluating the prediction of profiles. This is because it offers an advantage by increasing the understanding of the relationship between sensitivity, specificity and accuracy; reduces uncertainty through identification of profiles that are defective significantly; and searches for errors in the model (Farnaaz and Jabbar, 2016).

This result agrees with the work of Bhadra *et al.* (2018) where performance was compared in terms of accuracy, precision, Mathew's Correlation Coefficient (MCC) using benchmark datasets as inputs.

2.7.5. Proteome sequence databases query and discovery of putative anti-pneumonia AMPs

The discovery stage was to search for novel bacterial and viral anti-pneumonia AMPs against *Acinetobacter baumannii, Streptococcus pneumoniae, Klebsiella pneumoniae* and viral targets *Influenza A, B* as well as *Respiratory Syncytial Virus* with the aim of identifying peptides that had the same signatures/motifs and properties as the input sequences used to build the profiles AB, SP, KP, RSVM, INFA and INFB. The matches of the respective profiles to the proteome sequences were also shown with E-values (table

2.5). The E-value of 0.05 cut-off was applied to search for putative AMPs.

Anti-bacterial pneumonia AMPs

Scanning the profiles to identify novel anti-bacterial AMPs, profile AB identified 7 AMP sequences that adhered to the E-value set, which was reduced to 5 after removing of duplicates. Profile SP identified 58 AMP sequences that adhered to the E-value set, which was reduced to 7 after removing duplicate entries, while profile KP identified 7 peptide sequences without duplicates. These anti-bacterial peptide sequences were all single domains.

A final list of AMPs was identified and the AMPs were ranked according to their E-values with those having the smallest E-values considered the most likely putative anti-pneumonia AMPs. There was a very low probability that these peptides were wrongly predicted to be anti-pneumonia AMPs due to adhering to the E-value cut-off. These putative AMPs were named by adding BOPAM to the trained profiles from which they were derived (BOPAM-AB, BOPAM-SP, and BOPAM-KP). Table 2.5 shows top 2 predicted anti-bacterial pneumonia AMPs against each target class of bacterial pneumonia pathogens.

Anti-viral pneumonia AMPs

Scanning the profiles to identify novel anti-viral AMPs, profile INFA identified 13 AMP sequences which was reduced to 8 after removing duplicates. Profile INFB identified 8 AMP sequences which was reduced to 6 after removing duplicates while RSVM identified 14 AMP sequences which yielded 13 after removing duplicate sequences. All anti-viral peptide sequences had multiple repeats/domains, which when considered together, had E-values below the cut-off with each domain E-value above the default 0.05. Multiple repeats are common features during prediction of anti-viral peptides. This repeat of anti-viral peptides is important for the host immunity to restrain viruses by enhancing the peptides

availability during activity (Chang and Yang, 2013).

However, these multiple repeat of motifs could not be considered even though they had significant E-values since most AMPs are short with a low molecular weight. Also, only the individual repeats (domains) of the sequences were predicted to have anti-pneumonia activity from the anti-viral pneumonia profiles but not the entire sequences. These domains were then considered based on their similarity to the target profiles and thus could be termed putative anti-pneumonia AMPs due to the fact that most active AMP sequences range from 10 to 100 amino acids in length (Tincho *et al.*, 2016).

A final list of anti-viral AMPs was ranked according to their E-values with those having the smallest E-values considered the most likely putative anti-viral pneumonia AMP.

These putative AMPs were named by adding BOPAM to the trained profiles from which they were derived (BOPAM-RSV, BOPAM-INFA and BOPAM-INFB). Table 2.5 shows top 2 predicted anti-viral pneumonia AMPs against each target class of viral pneumonia pathogens.

Table 2. 5: Final list of the Anti-bacteria and Anti-viral pneumonia AMPs with their sources.

S/N	TARGET	ORGANISMS	E-VALUE	SEQUENCE
1	BOPAM-AB1	*Amphibian predicted*	0.0076	FLPI----LKSGLSGLL
2	BOPAM-AB 2	*Amphibian predicted*	0.0091	FFP----LLKSGLSGLL
3	BOPAM-KP 2	*Atta cephalotes*	0.007	MWK----GGLNNGVQCKIDNC
4	BOPAM-KP 3	*Atta cephalotes*	0.009	MWKHVVKLGNNFTVQCKIDNC
5	BOPAM-SP1	*Homo sapiens*	6.9e-16	QGRDD----SIMRRRGLTSPCKDINTFIHGNKRSIKAICENKNG
6	BOPAM-SP2	*Felis catus*	5.1e-11	KGRND----SMMERRGLTTPCKDTNTFIHGNKGSIKAICGNKNG
7	BOPAM-INFA1	*Ciona savignyi*	0.015	TPTFI----PIPKQ
8	BOPAM-INFA 2	*Nematostella vectensis*	0.034	CPVIL----QVLPK
9	BOPAM-INFB1	*Callithrix jacchus*	0.0045	MDV----TFLRVPPQ
10	BOPAM-INFB2	*Callithrix jacchus*	0.0055	MDV----TFLMVPPQ
11	BOPAM-RSV1	*Brachyspira murdochii*	0.0057	IVSS----INLCKNKF
12	BOPAM-RSV2	*Brachyspira hyodysenteriae*	0.0074	EVSK----IDSNLSTV

LEGEND: S/N: Serial number; BOPAM-AB1-5: Putative Anti-*Acinetobacter baumannii* AMP; BOPAM-SP1-8: Putative Anti-*Streptococcus pneumoniae* AMP; BOPAM-KP1-2: Putative Anti-*Klebsiella pneumoniae* AMP; BOPAM-INFA1-13: Putative Anti-*Influenza A virus* AMP; BOPAM-INFB1-7: Putative Anti-*Influenza B virus* AMP; andBOPAM-RSV1-13: Putative Anti-*Respiratory Syncytial Virus* AMP.

2.7.6. Physicochemical properties of the AMPs

The physicochemical properties of the putative AMPs were determined using APD and BACTIBASE to ensure that the identified sequences conform to other known AMPs based on the characteristics measured. Physicochemical features such as molecular weight amino acid composition, hydrophobicity, Boman index, net charge, isoelectric potential and half life were used to evaluate the anti-bacterial and anti-viral AMPs.

Molecular weight and Common amino acids

The amino acid composition of the AMPs contributes to the molecular weight, since the AMPs are made up of amino acids and can be a distinguishing feature to discriminate

between two classes of protein/peptides.

The **antibacterial pneumonia AMPs** have some amino acids which are common to them for discrimination against one another. BOPAM-AB1-4 had common amino acid namely leucine while BOPAM-AB5 had asparagine, glycine, lysine and valine. BOPAM-KP2 had asparagine; BOPAM-KP3 had lysine and glycine; BOPAM-KP4-6 had lysine while BOPAM-KP7 had cysteine and serine. BOPAM-SP1 had arginine; BOPAM-SP2-6 had glycine while BOPAM-SP7 had glycine and valine. The presence of very specific amino acids sequence influences function and binding capacity of that particular sequence.

Apart from this, the **anti-viral pneumonia AMPs** also had common amino acids that could distinguish them from one another. BOPAM-INFA1, 6 and 8 had proline; BOPAM-INFA2 had proline, valine, iso-leucine and leucine; BOPAM-INFA3 had threonine; BOPAM-INFA4-5 had proline and glutamine; while BOPAM-INFA7 had proline, iso-leucine, serine and valine. BOPAM-INFB1 had valine and proline; BOPAM-INFB2-3 had asparagine and leucine; BOPAM-INFB4-5 had leucine; while BOPAM-INFB6 had asparagine and leucine. BOPAM-RSV1 had iso-leucine and lysine; BOPAM-RSV2 had serine; BOPAM-RSV3 and 12 had isoleucine and asparagine; BOPAM-RSV4, 5, 7, 8, 9, 10, 11 had asparagine; BOPAM-RSV6 had iso-leucine while BOPAM-RSV13 had aspartate.

In summary, the consequence of the presence of charged, polar and non-polar amino acids within the anti-bacterial and anti-viral putative AMP sequences is conferment of charge, reduced or increased hydrophobicity and reduced or increased binding potential on them.

Hydrophobicity

The anti-pneumonia AMPs such as BOPAM-KP4, BOPAM-SP1 and 2, BOPAM-INFA1, 3, 4, 5, 6 and 8, BOPAM-RSV9 had hydrophobicity less than 30% due to presence of more polar amino acid residues. Hydrophobicity result lower than 30% is not an ideal percentage (Chen *et al.*, 2007). The resultant effect of reduced hydrophobicity on the non-polar face of the amphipathic helix is poor peptide helicity, reduced self-associating ability in aqueous environments and poor antimicrobial activity. Peptides with higher hydrophobicity would penetrate deeper into the hydrophobic core of the bacterial membrane, causing stronger hemolysis by forming pores or channels (Tachi *et al.*, 2002). Thus, the anti-bacterial and the anti-viral putative AMPs with increased hydrophobicity could potentially penetrate the membrane core. Reduced hydrophobicity is a consequence of polar amino acids. Recently, AMPs from sugar-functionalized phosphonium polymers have been reported to require the hydrophilic domains in their molecular structure to exert anti-bacterial activities against Gram-negative Escherichia coli and Gram-positive Staphylococcus aureus (Cuthbert *et al.*, 2018).

Net charge

All **anti-bacterial pneumonia AMPs** were positively charged whereas the **anti-viral peptides** such as BOPAM-INFB3, 6, BOPAM-INFA1, 2, 3, 4, 6, 7, 8, BOPAM-RSV2, 3, 4, 6, 7, 8, 10, 11, 12, 13 were predominantly neutral or negative. Cationic AMPs are said to be positively correlated with increased anti-microbial activities. Therefore all the anti-bacterial AMPs which were positively charged indicated conformity with ideal AMPs with improved anti-microbial activities. However, the lack of positive charge in the net charge of anti-viral AMPs does not interpret lack of anti-microbial activities because some negatively charged AMPs have been reported such as surfactant associated anionic peptides in APD database (AP00528) with a net

charge of -5 showing anti-bacterial activity and maximin H5 with charge ranging between -1 to -7 which shows bacterial growth inhibition against *L. monocytogenes* (Jodoin and Hincke, 2018).

Isoelectric potential (pI)

Anti-bacterial pneumonia AMPs had pI values between 6.22 and 12.91 whereas anti-viral pneumonia AMPs pI values ranged from 3.85 to 12.50. This range of values shows the solubility properties of the AMPs despite the variability of charges in acid and alkaline media (Abraham *et al.*, 2014). Isoelectric potential (pI) of peptides is a function of individual amino acids in both backbone groups. At a pH below the pI, AMPs carry a net positive charge and *vice versa*.

Boman Index

The results of the Boman index showed negative values for BOPAM-AB1-4, BOPAM-INFA2, 7, and BOPAM-INFB4 and 5. A negative Boman index has been said to be positively correlated with a more hydrophobic peptide, indicating a high protein binding potential, while a more hydrophilic peptide tends to have a more positive index. However, the tendency of some peptides to have a positive Boman index has been reported with their ability to detect HIV in a lateral flow device (Tincho *et al.*, 2016).

Half-Life

The half-life results of the **anti-bacterial pneumonia AMPs** showed that BOPAM-AB1-4 has a 1.1 hour half-life; BOPAM-KP2-4 has a 30 hour half-life, BOPAM-KP5 has a 20 hour half-life while BOPAM-KP6 and 7 has half lifes of 5.5 hours. BOPAM-SP1 has a half-life of 0.8

hours, BOPAM-SP2 and 8 has a half-lifes of 1.3 hours, BOPAM-SP3-6 has a half-life of 1.1hours while BOPAM-SP7 has a half-life of 2.8 hours.

Anti-viral pneumonia AMPs had BOPAM-INFA1, 3, 4, 5, 6, 8 with a half-life of 7.2 hours, BOPAM-INFA2 and 7 having a half-life of 1.2 hours. BOPAM-INFB1, 2 has a half-lifes of 30 hours, BOPAM-INFB3-6 having a half-life of 5.5 hours. BOPAM-RSV1 having a half-life of 20 hours and BOPAM-RSV10 having a half-life of 100 hours, all other BOPAM-RSV had half-lives ranging between 1 to 4.5 hours. AMPs have been said to generally exhibit a short half-life because they are not stable. Half-life values as low as 1hour have been reported for AMP molecules used for HIV diagnosis (Tincho *et al.*, 2016).

The use of physicochemical parameters as indices to evaluate AMPs is in agreement with the work of Hollmann *et al.* (2018) where a re-evaluation of the physicochemical properties of antimicrobial peptides was investigated, resulting in a characteristic thermal transition profile in model vesicles which was used to categorize novel molecules with unknown biological activity.

2.7.7. Similarity Analysis using BLASTp

The result obtained from BLAST was based on the parameters or settings from the UniProt server network. The results were displayed with the information of similarities such as (a) mismatching, (b) the start and end point of similarity in the length of the sequence, and (c) the percentage similarity. Overall BLAST analysis showed that BOPAM-AB5 has 100% similarity to all Moricin from *Bombyx mori, Antheraea pernyi, Hyblaea puera,* and *Heliothis viresc*; BOPAM-KP1 was also similar to Drosomycin and were all discarded. All other predicted sequences showed no similarity to known AMPs sequences.

Table 2.6: Physicochemical parameter of the anti-bacterial and anti-viral pneumonia putative AMPs

S/N	Profile	Mass number(Da)	% Hydrophobic	Common amino acid	Net charge	PI	Boman Index (kcal/mol)	Half-life (Hours)
1	BOPAM-AB1	1 755.35	52	L	+2	10.81	-1.55	1.1
2	BOPAM-AB2	1 789.36	52	L	+2	10.81	-1.43	1.1
3	BOPAM-AB3	1 909.56	64	L	+2	10.81	-2.18	1.1
4	BOPAM-AB4	1 849.46	58	L	+2	10.81	-1.81	1.1
5	BOPAM-KP2	2 370.21	42	NKGV	+2	8.70	0.9	30
6	BOPAM-KP3	2 478.35	47	NKV	+3	8.82	1.18	30
7	BOPAM-KP4	2 532.14	28	K	+7	10.58	1.5	20
8	BOPAM-KP5	2 667.29	39	K	+5	9.66	2.3	20
9	BOPAM-KP6	2 300.00	38	K	+8	10.98	2.28	5.5
10	BOPAM-KP7	2 589.16	43	CS	+4	8.53	2.0	5.5
11	BOPAM-SP1	5 083.36	27	R	+6	10.06	3.62	0.8
12	BOPAM-SP2	4 903.88	25	G	+6	10.00	3.06	1.3
13	BOPAM-SP3	4 765.02	45	G	+1	6.22	0.47	1.1
14	BOPAM-SP4	4 779.05	45	G	+1	6.22	0.45	1.1
15	BOPAM-SP5	4 779.05	45	G	+1	6.22	0.43	1.1
16	BOPAM-SP6	4 765.02	45	G	+1	6.22	0.47	1.1
17	BOPAM-SP7	4 793.07	47	GV	+1	6.22	0.36	2.8
18	BOPAM-SP8	5 140.22	35	K	+7	9.82	3.27	1.3
19	BOPAM-INFA1	1 541.41	28	P	0	6.34	0.9	7.2

20	BOPAM-INFA2	1 496.20	57	PVIL	0	0.16	-0.54	1.2
21	BOPAM-INFA3	1 518.28	28	T	-2	3.49	1.37	7.2
22	BOPAM-INFA4	1 559.68	28	QP	-1	7.52	3.75	7.2
23	BOPAM-INFA5	1 541.64	28	PQ	-1	3.75	0.9	7.2
24	BOPAM-INFA6	1 569.42	28	P	-1	4.18	1.2	7.2
25	BOPAM-INFA7	1 546.21	50	PISV	0	6.16	-0.03	1.2
26	BOPAM-INFA8	1 542.35	28	P	-2	3.55	0.99	7.2
27	BOPAM-INFB1	1 869.55	40	VPR	+2	10.40	2.27	30
28	BOPAM-INFB2	1 844.56	46	NL	+1	7.55	1.12	30
29	BOPAM-INFB3	1 687.62	46	NL	0	5.76	0.94	5.5
30	BOPAM-INFB4	1 780.00	66	L	+2	10.81	-1.61	5.5
31	BOPAM-INFB5	1 752.93	66	L	+1	9.70	-1.81	5.5
32	BOPAM-INFB6	1 656.33	46	NL	0	5.76	0.57	5.5
33	BOPAM-RSV1	1 865.32	43	IK	+1	8.54	1.39	20
34	BOPAM-RSV2	1 776.07	31	S	-1	4.43	2.29	1
35	BOPAM-RSV3	1 729.06	43	IN	0	6.45	1.69	1.4
36	BOPAM-RSV4	1 784.11	43	N	-1	4.18	1.7	1.3
37	BOPAM-RSV5	1 687.21	37	N	+1	9.70	2	4.4
38	BOPAM-RSV6	1 813.76	50	I	-1	4.18	0.89	1.3
39	BOPAM-RSV7	1 827.37	31	N	0	6.41	2.8	1.4
40	BOPAM-RSV8	1 784.05	43	N	-1	3.85	1	1.9
41	BOPAM-RSV9	1 761.30	25	N	+1	9.7	2.5	1.4
42	BOPAM-RSV10	1 760.31	31	N	0	6.34	2.21	100
43	BOPAM-RSV11	1 716.96	37	N	-1	4.11	2.14	1.4

44	BOPAM-RSV12	1 834.16	37	IN	0	6.45	2.23	1.4
45	BOPAM-RSV13	1 942.32	31	D	0	6.60	4.12	1.3

LEGEND: S/N: Serial number; BOPAM-AB1-5: Putative Anti-*Acinetobacter baumannii* AMP; BOPAM-SP1-8: Putative Anti-*Streptococcus pneumoniae* AMP; BOPAM-KP1-2: Putative Anti-*Klebsiella pneumoniae* AMP; BOPAM-INFA1-13: Putative Anti-*Influenza A virus* AMP; BOPAM-INFB1-7: Putative Anti-*Influenza B virus* AMP; and BOPAM-RSV1-13: Putative Anti-*Respiratory Syncytial Virus* AMP.

2.8. Conclusion

This chapter identified novel AMPs for the potential differential diagnosis of bacterial and viral pneumonia using HMMER *in silico* technology where 18 anti-bacterial putative AMPs and 27 anti-viral peptides were generated. The putative anti-pneumonia AMPs showed conformity to other known AMPs in terms of their physicochemical characteristics as measured by APD and BACTIBASE. Also, the majority of the identified sequences showed no similarity to any other known AMP sequences. Further experimentation is required to confirm the discovery, through HMMER, of the anti-viral and anti-bacterial pneumonia putative AMPs.

Chapter 3

3.0. *In silico* Comparison of HMMER with CD-HIT for the Identification of Anti-Pneumonia Antimicrobial Peptides (AP-AMPs)

3.1. Introduction

Computational strategies are invaluable assets to provide insight into the activity of AMPs in order to exploit their potential as a new generation of diagnostic biomarkers. Identification of AMPs using a combination of supervised and unsupervised learning techniques is available for bioinformatics analysis. This allows a high performance machine learning model to be built for the classification of AMPs with domain features such as amino acid frequencies and composition extracted from known data using the most suitable performance measures (Manavalan *et al.*, 2018). HMMER and GLAM2 have been combined using sequences of experimentally validated biomolecules for the construction of probabilistic models to display their features based on the motifs of these input sequences with the desired activity against specific targets (Fjell *et al.*, 2009, Boman, 1995).

Combining different tools can ensure that the bio-molecule/biomarker discovery process is not hampered by the use of only one tool since all of these algorithms suffer one or more shortcomings. HMMER, for example, is very sensitive to the use of small data for analysis whilst GLAM2 produces false positive result with less specificity against targets.

Thus by combining more than one tool, it decreases the chance of missing out on the discovery of a bio-molecule due to inherent shortcomings of all tools. Tincho, 2014 (MSc thesis) compared HMMER and GLAM2 to identify novel AMPs against HIV. The study concluded that HMMER was far superior to GLAM2 since GLAM2 lacked strongly embedded statistical measures to produce AMPs with a high probability to be true positives. The AMPs identified by HMMER was experimentally proven to have high activity against HIV (Tincho *et al.*, 2016).

In the previous chapter, novel AMPs against pneumonia was discovered using HMMER with very little input sequences for the SP model which produced a profile with low sensitivity. In addition, all other profiles were also built with relatively low numbers of input sequences even though the calculated parameters to measure robustness of the profiles gave good values. For that reason, in this chapter, CD-HIT will be combined with HMMER for maximum AMP discovery.

3.1.1. CD-HIT Algorithm

The CD-HIT algorithm arranges the input sequences sequentially from long to short and processes them with the first sequence being automatically classified as the first cluster representative sequence. The query sequences of the remaining sequences are then compared to the representative sequences found before them, classified as either redundant or representative based on their similarity to the ones in the existing representative sequence cluster (Li, 2015). CD-HIT has the tendency to cluster non-redundant database (NR) at 90% identity in approximately 1 hour and at 75% identity in approximately 1 day on a 1 GHz Linux PC (Li *et al.*, 2001); Faster clustering speed is required because the size of protein databases are rapidly growing and many applications desire a lower attainable threshold. It is therefore not surprising that CD-HIT has been used by many groups, including Uniprot (Leinonen *et al.*, 2004), PDB (Finn *et al.*, 2004), and to generate non-redundant protein datasets to reduce the database search efforts and to improve homology detection sensitivity (Li *et al.*, 2002). For this reason, CD-HIT will be used in this chapter to check the novelty of the HMMER generated anti-pneumonia AMPs through comparison with the retrieved experimentally validated AMPs.

3.2 Aim

The aim of this chapter is to compare HMMER with CD-HIT algorithms for validity and proof of novelty of the identified AMPs.

3.3 Objectives

The specific objectives of this chapter include:

- Search for similarity between the experimentally validated AMPs list generated from various databases (section 2.2) and those generated using the CD-HIT and BLAST programs; and

- Finalise a list of potential AMPs for further validation that can serve as diagnostic and prognostic biomarkers for pneumonia.

3.4. Materials and Methods

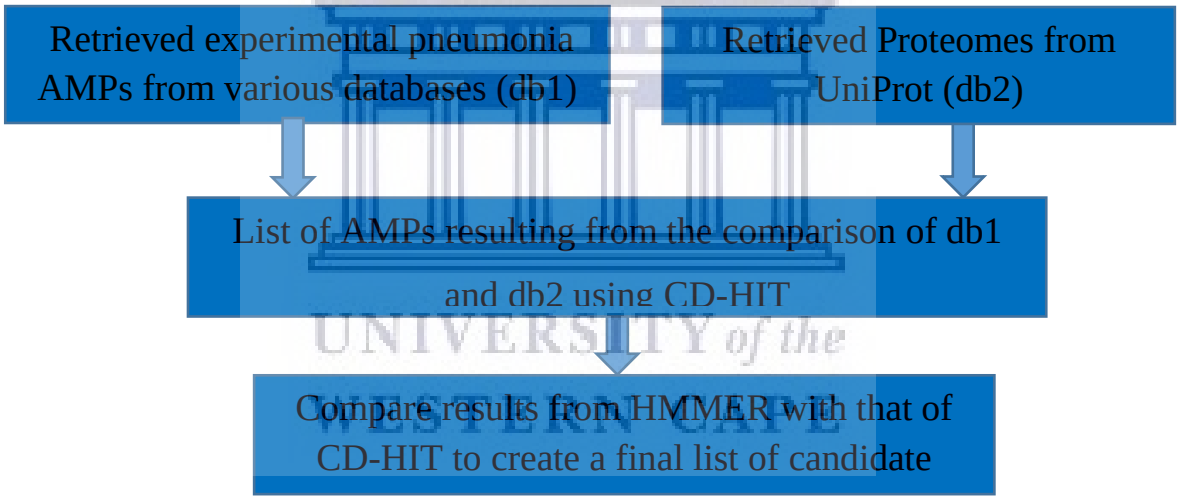


Figure 3. 1: Outline of the CD-HIT *in silico* methodology used for AMPs identification

3.4.1. Data retrieval

The experimentally validated anti-pneumonia AMPs for bacterial (*Streptococcus pneumoniae, Acinetobacter baumannii and Klebsiella pneumoniae*) and viral causative agents (*Respiratory syncytial virus, Influenza A and B* viruses) were retrieved from antimicrobial peptide databases such as namely: Antimicrobial Peptides Database (APD) (Wang and Wang, 2004, Wang *et al.*, 2008), Collection of Antimicrobial Peptides (CAMP) (Thomas *et al.*, 2009), UniProtKB (Wu

et al., 2006) and Antiviral peptides databases (AVPDB) (Sencanski *et al.*, 2015) and kept as database one (db1).

3.4.2. Identification of proteome sequences

Proteome sequences (in the fasta format) as used in chapter two, from the Ensembl (http:// www.ensembl.org/index.html) and the UniProt databases (http://www.uniprot.org/) were used as database two (db2).

3.4.3. AMPs Prediction using CD-HIT

CD-HIT algorithm (v 4.6.8.) was installed on Ubuntu and used to cluster for comparison of the experimentally validated anti-pneumonia sequences with various proteome sequences. This was achieved by using the database file (db1) and scanning it against the database file (db2) using the command line:

```
cd-hit 2d -i db1 -i2 db2 -o db2novel -c 0.9 -n 5 -d 0 -M 16000 -T 8  -S2 0.9 .....(i)
```


3.5. Results and Discussion

3.5.1. CD-HIT Analysis of db1 and db2

In this section, another *in silico* tool namely CD-HIT was employed to compare results obtained with HMMER in chapter 2. CD-HIT works on the basis of sequence similarity with the rationale that sequence is related to function i.e. if two sequences have similarity with the function of one sequence already proven, the function of the other sequence, without a known function, can be inferred. CD-HIT thus clusters a sequence of known function with a similar sequence without a known function.

Six db1 for each of the classes/pathogens SP, KP, AB, INFA, INFB and RSVM containing experimentally validated AMPs were created with db2 containing all proteome sequences

(1000 proteomes) as used in chapter 2. Table 3.1 tabulates the result of scanning the respective db1's against db2 using CD-HIT

Table 3.1: CD-HIT Analysis of the sequences within the respective db1' and db2

Classes	Db1	CD-HIT Match
AB	12	5
SP	16	8
KP	140	7
INFB	12	6
INFA	52	8
RSVM	112	13

Legend: AB: Anti-*Acinetobacter baumannii*; SP: Anti-*Streptococcus pneumoniae*; KP: Anti-*Klebsiella pneumoniae*; INFA: Anti-*Influenza A virus*; INFB: Anti-*Influenza B virus*; and RSVM: Anti-*Respiratory Syncytial Virus*

AB showed 5 of its AMP sequences to be similar to sequences within the entire proteome with SP showing 8; KP showing 7; INFB showing 6; INFA showing 8; and RSVM showing 13 anti-pneumonia AMPs across proteomes for the target organisms. The "hits' represented in table 3.1 are only sequences that met the threshold of 90 % similarity.

3.5.2. AMP Identification using CD-HIT

In this section, the CD-HIT identification of the AMPs was analysed by using the percentage similarity criteria from the algorithm to prioritize the novelty of the AMPs. The amino acid sequences within the proteome identified by the CD-HIT algorithm were compared with the experimentally validated AMPs of each class (6 x db1) as well as with the putative anti-pneumonia AMPs identified in chapter 2 using HIMMER.

The **anti-viral peptides** for each target class namely INFB, INFA and RSVM, showed less than 60% similarity across the proteomes with an uncharacterized sequence similarity to known anti-viral peptides. Thus none of the sequences within the 3 anti-viral classes matched a sequence within any of the proteomes scanned, at the 90 % threshold.

.**Table 3. 2**: Comparison of the Percentage Similarity of the anti-pneumonia AMPs with experimentally validated AMPs

S/N	AMPs Clusters	Percentage similarity (%)
1	BOPAM-SP1 Human beta-defensin	97.60
2	BOPAM-SP2 Uncharacterized	57.1
3	BOPAM-SP3 Beta-amyloid 1-42	52.1
4	BOPAM-SP4 Beta-amyloid 1-42	54.0
5	BOPAM-SP5 Beta-amyloid 1-40	53.0
6	BOPAM-SP6 Beta-amyloid 1-40	57.0
7	BOPAM-SP7 Angiogenin	56.8
8	BOPAM-SP8 Beta-amyloid 1-42	59.0
9	BOPAM-KP1 Mycin-7	59.1
10	BOPAM-KP2 Uncharacterized	57.1
11	BOPAM-KP3 Uncharacterized	59.0
12	BOPAM-KP4 Uncharacterized	58.0
13	BOPAM-KP5 DRS-11-3	56.1
14	BOPAM-KP6 Uncharacterized	55.0
15	BOPAM-KP7 Uncharacterized	54.5
16	BOPAM-AB1 Temporin Alf, Temporin Al	93.75
17	BOPAM-AB2 Uncharacterized	59.1
17	BOPAM-AB3 and 4 Temporin Ale, Temporin	100
18	BOPAM-AB5 Moricin 1	100

Legend: BOPAM-AB: Anti-*Acinetobacter baumannii*; BOPAM-SP: Anti-*Streptococcus pneumoniae*; BOPAM-KP: Anti-*Klebsiella pneumoniae.*

In table 3.2 above, BOPAM-SP1 and 7 showed 50.0% and 56.8% similarity to Angiogenin from *Homo sapiens* and *Monodelphis domestica* respectively; BOPAM-SP2 showed 57.1%

similarity to an uncharaterized protein in the proteome of *Felis catus* and no similarity to any known anti-*Streptococcus pneumoniae* AMP. BOPAM-SP3, 4, 5, 6 and 8 had less than 60% match to beta-amyloid peptides from *Xenopus tropicalis*.

Furthermore, the result from BOPAM-KPs showed less than 60% similarity to the experimentally validated anti-*Klebsiella pneumoniae* AMPs with BOPAM-KP1 having 59.1% similarity to the middle portion of mycin-7 from *Caenorhabditis latens* and BOPAM-KP5 having 56.1% similarity to DRS-11-3 from *Drosophila takahashii* with the rest of the BOPAM-KPs having no similarity to known anti-*Klebsiella pneumoniae* AMPs, thus, were considered novel.

The results for anti-*Acinetobacter baumannii* similarity comaprison using CD-HIT showed that the sequence *Amphibian Uniprot predicted* had a 93.75% degree of similarity to BOPAM-AB1. Comparison of the BOPAM-AB1 amino acid sequence to the experimentally validated AMPs such as Temporin Alf and Temporin Alh showed that there was a substitution of phenylalanine in Temporin Alf at position 2 with leucine in the predicted sequence. Furthermore, the analysis showed a serine substituted with leucine coupled with an insertion of lysine at position 10 which made the sequence of BOPAM-AB1 distinct from these matches. This distinction is based on the assumption that amino acid residues in a peptide influence its functionality and binding potential, thus a potentially novel role for BOPAM-AB1. Temporin Alf and Alh are experimentally validated anti-*Acinetobacter baumannii* AMPs from which db1 for the AB class was created. With this high similarity, it can be deduced that BOPAM-AB1 will have the same activity against *Acinetobacter baumannii* as Temporin Alf and Alh.

BOPAM-AB2 from *Amphibian Uniprot Predicted* showed similarity of 59.1% with a potentially uncharacterized AMP. BOPAM-AB 3 and 4 showed 100% similarity to Temporin Alg and Ale from *Amolops loloensis* respectively at the middle portion of the protein with an

insertion of lysine at position-10. BOPAM-AB5 from *Bombyx mori* had all constituent amino acid residues of Moricin-1 at position 7 – 23.

3.6. Conclusion

The aim of this chapter was to identify novel anti-pneumonia AMPs using an alternative algorithm as in chapter 2 namely, CD-HIT. Due to certain shortcoming of HMMER, as outlined in the introduction of this chapter, a second algorithm was deemed necessary to ensure that all potentially novel anti-pneumonia AMPs would be identified. Eighteen anti-bacterial pneumonia and twenty seven anti-viral pneumonia AMP domains were obtained from the CD-HIT analysis, which agreed with the HMMER results in chapter 2. It can be inferred from this chapter that functional similarity was observed between the HMMER and CD-HIT algorithms, which showed the same "hits" in the proteomes of the various organisms despite, differing scoring criteria. Therefore, combined *in silico* methods for the identification of AMP molecules is recommended as this would solve the inherent weaknesses of a particular algorithm.

Chapter 4

4.0. Identification of Receptors as Targets for Detection of Bacterial and Viral Pneumonia

4.1. Introduction

Proteins are biological macromolecules that maintain the structural and functional integrity of the cells, and many diseases are associated with protein malfunction (Jung *et al.*, 2017). Certain protective mechanisms have been developed by organisms, such as the adaptive immune response, that efficiently recognize and destroy specific pathogens (Druzd *et al.*, 2017), and thus restrict the spread of pathogens usually without causing significant non-specific inflammation. Bacterial and viral pneumonia pathogens, on the other hand, have evolved numerous bypass mechanisms to remain resistant to the host protective pathways (Kapoor *et al.*, 2017). However, the initiation, development, regulation, contraction, and re-activation of adaptive immune responses upon pathogen challenge remain questionable for future research. Due to this, sensitive identification of these receptor proteins may be an efficient way to target these pathogens from diagnostic and therapeutic perspectives.

4.2. Characteristics of Receptors to be used in diagnosis of pneumonia

The aim of this study was to identify AMPs that could be used as reporter molecules for the differential diagnosis of bacterial and viral pneumonia to ensure effective treatment regiments. The AMPs (ligands) have to engage with bacterial or viral receptors to facilitate diagnosis within a patient sample. These receptors should adhere to a strict set of criteria as discussed below:

- The receptors should be produced in appreciable high concentration for a particular strain within body fluids. This is important because it will aid their uniform distribution across the fluid for bioavailability during detection.

- The receptor should be specific to its ligand. This is important because bioreceptor is a crucial component as its biochemical properties assure high sensitivity and selectivity of the biomarker detection and permit to avoid interferences from other microorganisms or molecules present in the tested sample.
- The receptors should be easily accessible majorly cell surface receptors as to be easily bind to their respective ligands. The discriminate binding of a particular ligand to a specific receptor, will facilitate differential diagnosis of the causative agent.
- The receptors should be moderately stable to *in vitro* exposure. Their half-life within biological fluids should be long enough to ensure detection without degradation.

The section below will be discussing various types of receptors that can be targeted for differential diagnosis of bacterial and viral pneumonia.

4.3. Types of Receptors

4.3.1. Intracellular Receptors

Intracellular receptors are receptors found on the inside of the cell either in the cytoplasm or nucleus and are small and hydrophobic in nature. An example of this is the primary receptor for the steroid hormones such as estradiol and testosterone (Taoka *et al.*, 2011). When these hormones enter a cell and bind to their receptors, they cause the receptors to change shapes, allowing the receptor-hormone complexes to enter the nucleus and regulate gene activity. Hormone binding exposes regions of the receptor that have DNA-binding activity, meaning they can attach to specific sequences of DNA. These sequences are found next to certain genes in the DNA of the cell, and when the receptor binds next to these genes, it alters their level of transcription (Blasius and Beutler, 2010). Examples of intracellular receptor are NOD-like receptors (NLRs) and RIG-1-like receptors (RLRs) that are used for the detection of

inflammatory disorders caused by microbial components in the cytosol (Kanneganti *et al.*, 2007).

4.3.2. Cell Surface Receptors

Cell-surface receptors are membrane-anchored proteins that bind to ligands at the extracellular surface of the cell. In this type of signaling, the ligand does not need to cross the plasma membrane (Song *et al.*, 2005). So, many different kinds of molecules (including large, hydrophilic ones) may act as ligands (Christopoulos, 2002). A typical cell-surface receptor has three different domains, or protein regions: an extracellular ligand-binding domain, a hydrophobic domain extending through the membrane, and an intracellular domain, which often transmits a signal (Spear *et al.*, 2000). The size and structure of these regions can vary a lot depending on the type of receptor, and the hydrophobic region may consist of multiple stretches of amino acids that traverse the membrane (Anliker and Chun, 2004). Certain diseases, such as Alzheimer's disease, related to cell surface receptors, are due to deficiency or degradation of the receptor via changes in the genes that encode and regulate the receptor protein (Scheuer *et al.*, 1996).

4.3.3. Ligand Gated Ion Channel Receptors

It is a form of cell surface receptor and are ion channels that can open in response to the binding of a ligand (Sato *et al.*, 2008). This type of cell-surface receptor has a membrane-spanning region with a hydrophilic channel through its middle. The channel also allows ions to cross the membrane without touching the hydrophobic core of the phospholipid bilayer. A ligand binds to the extracellular region of the channel, allowing a change in the protein's structure in such a way that ions of a particular type, such as Ca^{2+}or Cl^{-}, can pass through (Bocquet *et al.*, 2009). In some cases, the channel is usually open, and ligand binding causes it to close. Changes in ion levels inside the cell can change the activity of other molecules, such as ion-binding

enzymes and voltage-sensitive channels, to produce a response (Collingridge *et al.*, 2009). Neurons, or nerve cells, have ligand-gated channels that are bound by neurotransmitters. Examples of ligand-gated ion channel receptors include Nicotinic cholinergic, GABA-A, and the 5-hydroxytryptamine$_3$ ($5HT_3$) used for neurodegenerative diseases detection (Li *et al.*, 2014).

4.3.4. G-Protein Coupled Receptors (GPCR)

These are a large family of cell surface receptors that share a common structure and method of signaling. The members of the GPCR family all have seven different protein segments that cross the membrane, and they transmit signals inside the cell through a type of protein called a G protein (Rosenbaum *et al.*, 2009). GPCRs are diverse and bind many different types of ligands. One particularly interesting class of GPCRs is the odorant (scent) receptors. When its ligand is not present, a G protein-coupled receptor waits at the plasma membrane in an inactive state. Some types of GPCRs, when docked to its signaling target in their inactive form, they become activated and cause the G protein to exchange GDP for GTP (Cherezov *et al.*, 2007). The now-active G protein separates into two pieces (one called the α subunit, the other consisting of the β and γ subunits). The subunits can interact with other proteins, triggering a signaling pathway that leads to a response (Venkatakrishnan *et al.*, 2013). Eventually, the α subunit will hydrolyze GTP back to GDP, at which point the G protein becomes inactive. Examples of G-Protein coupled receptors include arginine vasopressin receptor 1A, dopamine receptor D2 (D2R), metabotropic glutamatereceptor type 6 (mGluR6), histamine H4 receptor, G protein-coupled receptor 2 (GPR2), cannabinoid receptor type 1, and gamma-aminobutyric acid receptors (Zhao *et al.*, 2016).

4.3.5. Receptor Tyrosine Kinases

Enzyme-linked receptors are cell-surface receptors with intracellular domains that are associated with an enzyme (Tatsumi *et al.*, 2003). In some cases, the intracellular domain of the receptor actually *is* an enzyme that can catalyze a reaction. Other enzyme-linked receptors have an intracellular domain that interacts with an enzyme (Reichardt *et al.*, 2005). Receptor tyrosine kinases are a class of enzyme-linked receptors found in humans and many other species (Terracina *et al.*, 2003). Signaling molecules first bind to the extracellular domains of two nearby receptor tyrosine kinases. The two neighboring receptors then come together, or dimerize. The receptors then attach phosphates to tyrosines in eachothers' intracellular domains. The phosphorylated tyrosine can transmit the signal to other molecules in the cell (Bae-Harboe and Park, 2012). Example of this class is the class-III receptor tyrosine kinases used for acute leukemia detection (Berenstein, 2015).

4.4. Aim

The aim of this chapter is to identify specific bacterial and viral receptors of diagnostic relevance as targets for the anti-pneumonia AMPs/ligands.

4.5. Objectives

- Identify bacterial and viral pneumonia receptors as a target for the putative AMPs through literature mining;
- Compute the function and cellular location of these receptors; and
- Compute the amino acid sequences of the identified receptors.

4.6. Materials and Methods

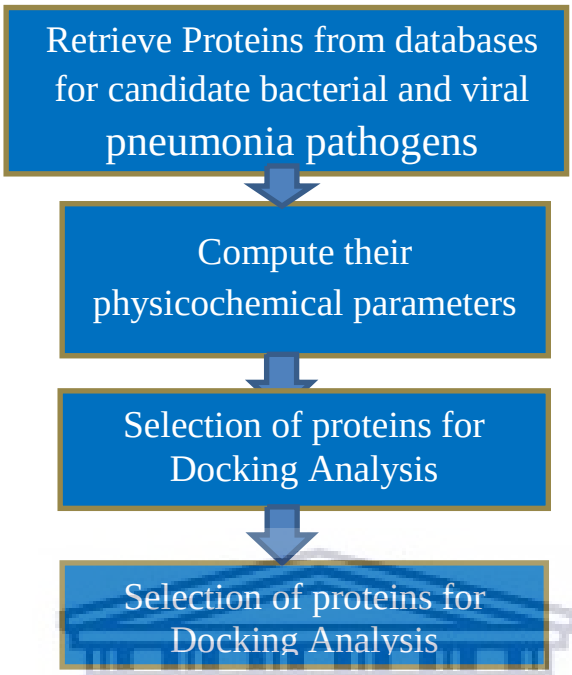


Figure 4. 1: Diagrammatic representation of the process for receptors identification

4.6.1. Identification of Receptors

Viral and bacterial receptors, such as cell surface receptors, were identified for the bacterial causative agents (*Streptococcus pneumoniae, Acinetobacter baumannii and Klebsiella pneumoniae*) and viral causative agents (*Respiratory syncytial virus, Influenza A and B)* implicated in pneumonia to serve as targets for the identified AMPs using several *in-silico* methods under the following subheading:

4.6.2. Protein Retrieval

Bacterial and viral pneumonia proteins were collected from various protein data banks (PDB) such as National Centre for Biotechnology Information (NCBI) (Coordinators, 2017), UniProt (Wu *et al.*, 2006), Google Scholar (Finn *et al.*, 2017), and Ensembl (Aken *et al.*, 2017) through literature mining. Thereafter, curation was performed to verify that all the retrieved pathogen

pneumonia proteins were complete or partial. Partial proteins were discarded and complete protein was kept for further analysis.

4.6.3. Determination of the physicochemical properties of the protein sequences

The physicochemical properties of the pneumonia pathogen proteins were calculated as carried out in section from the ProtParam tools (http://web.expasy.org/protparam/) using exPAsy server to reveal the protein structure and function (Gill and Von Hippel, 1989).

4.6.4. Selection of final list of pneumonia proteins

The proteins were sorted based on their roles and location in the bacterial and viral pneumonia pathogens for accessibility to be detected in a lateral flow device.

4.7. Results and Discussion

4.7.1. Retrieval of Protein Receptors of Pneumonia Pathogens

This section was carried out to determine the immunogenic proteins of bacterial and viral pneumonia of diagnostic relevance to serve as targets for the novel antimicrobial peptides for the diagnosis of the different pathogens. Several pneumonia proteins such as cell surface receptors and nucleoproteins were identified for the pneumonia pathogens *Acinetobacter baumanni*, *Influenza A* virus, *Influenza B* virus, *Respiratory Syncytial* virus, *Streptococococcus pneumoniae*, *Klebsiella pneumoniae*. These protein receptors of pneumonia pathogens were retrieved from the protein databank (PDB) in National Centre for Bioinformatics Institute (NCBI) database and are projected to be potentially relevant in diagnosis of bacterial and viral pneumonia pathogens.

4.7.2. Physicochemical properties of the pneumonia proteins

Table 4.1 below shows the summary of the physicochemical properties of the pneumonia protein receptors.

4.7.2.1. *Acinetobacter baumannii*

Acinetobacter baumannii is an emerging nosocomial pathogen that is resistant to many types of antibiotics, and hence, a fast, sensitive, specific, and economical test for its rapid diagnosis is needed. Literature mining and analysis of the pneumonia pathogen proteins using protparam tool of the exPAsy server showed that *Acinetobacter baumannii* has an iron regulated outer membrane receptor protein with molecular weight 85519.34Da, isoelectric point 7.55, hydrophobicity 31.22 %, charge +1, instability index 25.61 and Half-life of 30 hours in mammals.

Acinetobacter baumannii iron regulated outer membrane protein has a strong potential as a receptor for the diagnosis of pneumonia caused by this organism. The protein is well-conserved throughout evolution and stable *in vitro* as indicated by its instability index of 25.61 (see table 4.1). A protein whose instability index is smaller than 40 is predicted as stable *in vitro*. Studies have shown that the outer membrane protein (Omp) of *Acinetobacter baumannii* can be used as candidate bio-molecules in animal models for diagnostic tests (Islam *et al.*, 2011). This protein possesses attributes such as outer membrane localization, high adhesion probability (0.53), possession of a single transmembrane helix and absence of homology to human proteins ((Garg *et al.*, 2016). Taken together this protein presents itself as a very suitable candidate to be targeted for diagnostics of *Acinetobacter baumannii*.

4.7.2.2. *Streptococcus pneumoniae*

Streptococcus pneumoniae is presumed to be the primary bacterial cause of community-acquired lower respiratory infections and meningitis among children and elderly in many countries. The laboratory diagnosis of invasive pneumococcal disease continues to rely on culture-based methods from appropriate clinical samples such as blood, pleural fluid, or purulent sputum that have been used for many decades.

Pneumolysin was identified for *Streptococcus pneumoniae*, which has a molecular weight of 52896.42 Da, isoelectric point of 5.18, hydrophobicity of 33.97 %, charge -14, instability index 20.69 and Half-life of 30 hours in mammals.

The most significant recent developments in diagnosis of pneumonia has occurred with antigen detection assays using pneumolysin (Shima *et al.*, 2016). The use of pneumolysin is very essential for *S. pneumoniae* diagnosis because it is produced in high concentration and stable in different body fluid samples across virtually all clinical isolates. In addition, based on the protein's physicochemical properties using charge, instability index and half-life as measures, pneumolysin is an attractive candidate receptor for the diagnosis of *Streptococcus pneumoniae.*

4.7.2.3. *Klebsiella pneumoniae*

Evidence indicates that *Klebsiella pneumoniae* infections are characterized by a lack of an early inflammatory response, thus making detection difficult. However, it is unknown whether *Klebsiella pneumoniae* employs additional factors to modulate host inflammatory responses to escape detection.

Results indicated that *Klebsiella pneumoniae* has an iron regulated outer membrane protein with molecular weight 80401.89Da, isoelectric point 4.89, hydrophobicity 32.37%, charge -24, instability index 35.81 and Half-life of 30 hours in mammals. Iron regulated outer membrane protein is being used for the diagnosis of *K. pneumoniae* using antibody detection because it is well-conserved throughout evolution and stable across clinical samples (Gregorich and Ge, 2014).

4.7.2.4. Respiratory Syncytial Virus (RSV)

Respiratory syncytial virus has some immunogenic receptors which are believed to have potential diagnostic relevance such as membrane fusion core protein chains (Vabret *et al.*, 2001). Its RSV matrix protein chain A, molecular weight 28990.34Da, isoelectric point 8.49,

hydrophobicity 35.66%, charge +3, instability index 33.15 and half-life of 1 hour in mammals; it also has Human RSV fusion protein core chain X with molecular weight 4869.38Da, isoelectric point 4.38, hydrophobicity 39.53%, charge -4, instability index 49.41 and half-life of 1 hour in mammals.

Diagnosis of Respiratory syncytial virus (RSV) during acute infection is difficult because of the poor sensitivity of viral culture and antigen detection. Some protein chains such as fusion protein core A and matrix fusion protein X, which are integral RNA proteins of RSV, mediate entry into the transmembrane glycoproteins of the host cell to elicit apoptosis (Prendergast and Papenburg, 2013). They also play a crucial role in virus assembly, and interact with the RNA complex as well as with the viral membrane. Identification of these proteins in the body fluid has shown only slight antigenic variation, which is not progressive, an important factor for their use in the diagnosis of the virus (Falsey *et al*., 2002).

4.7.2.5. *Influenza* A virus

Influenza A virus has some protein receptors of potential relevance for its diagnosis. Matrix protein M1 has a molecular weight of 27780.10Da, isoelectric point 9.32, hydrophobicity 35.32%, charge +6, instability index 38.44 and half-life of 30 hours in mammals; and 416a monomeric nucleoprotein with molecular weight 56297.78Da, isoelectric point 9.45, hydrophobicity 29.52%, charge +12, instability index 46.35 and half-life of 30 hours in mammals.

Influenza A virus R416a monomeric nucleoprotein plays important structural and functional roles which could be explore for its diagnostics. On the other hand, M1 protein of Influenza A virus is a matrix protein that forms a coat at the viral envelop. It is a bi-functional membrane/RNA-binding protein that mediates the encapsulation of RNA-nucleoprotein cores into the membrane envelope. The use of influenza A matrix protein in diagnosis is justified

because it is the most abundant protein within the virus particle across strains and subtypes (Söderlin *et al.,* 2004). The nucleoprotein is also recently being used for diagnosis using antibodies (Voeten *et al.,* 1998).

4.7.2.6. *Influenza* B virus

Influenza B virus receptor proteins of diagnostic potential were identified and analyzed. *Influenza* B virus has nucleoprotein with molecular weight 61644.09Da, isoelectric point 9.43, hydrophobicity 31.61%, charge +18, instability index 43.78 and half-life of 30 hours in mammals.

Influenza B virus nucleoprotein has been used in the diagnosis of pneumonia (Voeten *et al.,* 1998). In both *Influenza* A and B viruses, factors associated with severe disease include older age, delay in hospitalization, lower-respiratory-tract involvement, and a low total peripheral white blood cell count or lymphopenia at admission (Yuen *et al.,* 1998).

Table 4.1: Physicochemical Properties of the Retrieved Pneumonia Receptor Proteins

S/N	Anti-Pneumonia proteins	Molecular weight (Da)	Hydrophobicity (%)	Net charge	Instability index	Half-life (hours)
1	AB Iron regulated OMP	85519.34	31.22	+1	25.61	30
2	INFA matrix protein M1	27780.10	35.32	+6	38.44	30
3	INFA R416a monomeric Nucleoprotein	56297.78	29.52	+12	36.35	30
4	INFB Nucleoprotein	61644.09	31.61	+18	43.78	30
5	RSV Chain A Matrix protein	28990.34	35.66	+3	33.15	1
6	RSV Chain X Fusion core protein	4869.38	39.53	-4	39.41	30
7	SP Pneumolysin	52896.42	33.97	-14	20.69	30
8	KP Iron regulated OMP	80401.89	32.37	-24	35.81	30

AB= *Acinetobacter baumanni*; INFA= *Influenza A* virus; INFB- *Influenza B* virus; RSV= *Respiratory Syncytial* virus; SP= *Streptococococcus pneumoniae*; KP= *Klebsiella pneumoniae*.

Additionally, viral pneumonia is very common in both children and elderly but more likely to occur in young children. This is because the bodies of young children have a harder time

fighting off the virus than people with a strong immune system. However, no unique identifying clinical characteristics are present that allow the physician to differentiate viral from bacterial pneumonia disease. This can be due to antigenic drift, antigenic shift and surface charge of the negative sense used by the viral pathogens to escape detection. The available methods of detection do not give accurate result, thus giving rise to a false positive result. Despite significant advancement in diagnosis and treatment research, viral pneumonia continues to be a major public health concern.

4.7.3. Final list of the pneumonia receptors

The use of receptor protein candidates such as *Klebsiella pneumoniae* iron regulated outer membrane protein (Paczosa and Mecsas, 2016), *Acinetobacter baumannii* iron regulated outer membrane protein (Singh *et al.*, 2016), *Streptococcus pneumoniae* pneumolysin (Andrade *et al.*, 2016), *Respiratory syncytial* virus fusion protein chains A and X (Fries *et al.*, 2017), *Influenza* A virus matrix protein M1 and nucleoprotein (Suarez, 2016, Vemula *et al.*, 2016), and *Influenza* B virus nucleoprotein (Hoffmann *et al.*, 2016) in diagnosis of pneumonia is justified because they are produced in relatively high amount within body fluids across all strains and subtypes of these pathogens; do not change with time; highly accessible either as cell surface receptor and relatively stable in a mild *in vitro* handling. These receptors were therefore retained for further analysis.

4.8. Conclusion

Prediction of protein candidates from the proteomes of pathogens could provide potential leads in the diagnosis of infections caused by these pathogens. Accessible localization in the outer membrane, stability and synthesis in high concentration are important criteria for receptors (biomarker) diagnostic potential. This chapter identified several viral and bacterial pneumonia proteins that can be targeted by the identifed AMPs (ligands) for differential diagnosis of this

disease. The idemtified proteins show high adherance to the criteria set out for their selection and use of Biomarkers. Although the literature indicates that some of these proteins are already employed in diagnosis of a singular pneumonial pathogen, this study aims at using AMPs rather than antibodies for the detection of these proteins within a biological sample. Furthermore, the AMPs will be engineered to target non-antigenic regions as not to compete with circulating antibodies prodoced by the infective person in response to the infection. Lastly, the study aim to design a singular test instead of multiple test that can sensitively and accurately differentiate bacterial from viral causative agents of pneumonia. Further studies would be required to establish the differential diagnostic potential of these viral and bacterial pneumonia receptor proteins to certain ligands through *in silico* structural prediction and binding interaction studies, an essential criterion for developing specificity-driven detection that would differentiate viral from bacterial pneumonia.

Chapter 5

5.0. Structure Prediction and Docking Analysis of the Putative Anti-pneumonia AMPs and Pneumonia Receptors

5.1. Introduction

Protein structure determination is crucial for accessing information about sequences and their physical properties. This relates to how the protein takes on its native conformation in solvent without the aid of a protein chaperone. Therefore, information on how to predict protein structure is imperative. The availability of feasible methods for protein structure prediction in the area of homology modelling and protein fold recognition are the two parameters currently being adopted; although, there are challenges in the accuracy of both methods. One disadvantage of these methods is that known structures have to be available otherwise these methods would fail when predicting the structure of a novel protein (Joyce *et al.*, 2014). Additionally, they tend to fail in predicting structures which are particularly sensitive to sequence differences, such as with random coils. However, not all protein structure prediction projects involve the use of these techniques (Zhang, 2008)

Several *in silico* tools are available for the purpose of structure prediction. Available tools include Psipred, I-Tasser. Lomets, Quark, just to mention a few. Only I-Tasser will be described in detail as it will be used wihin this study.

5.2. Structure Prediction using I-TASSER

The iterative threading assembly refinement (I-TASSER) server is designed with an integrated platform for automated protein structure and function prediction based on the sequence-to-structure-to-function paradigm (Roy *et al.*, 2010). Starting from an amino acid sequence, I-TASSER first generates three-dimensional (3D) atomic models from multiple threading alignments and iterative structural assembly simulations. The function of the protein is then

inferred by structurally matching the 3D models with other known proteins (Zhang, 2007). The output from a typical server run contains full-length secondary and tertiary structure predictions, and functional annotations on ligand-binding sites, Enzyme Commission numbers and Gene Ontology terms. An estimate of accuracy of the predictions is provided based on the confidence score of the modeling. This protocol provides new insights and guidelines for designing of online server systems for the state-of-the-art protein structure and function predictions (Roy *et al.*, 2010).

5.3. Docking Analysis using PATCHDOCK

Patchdock is an efficient algorithm for docking two protein molecules and it stands out among its equal because of its attractive running time and high quality results. It achieves this by dividing the two molecules involved into shape-based patches in order to address the efficiency of output. This enables the identification of hotspot, shape complementarity function and faster scoring due to multi-resolution surface data structure to allow for intermolecular penetration (Camacho *et al.*, 2000).

Two general approaches have been adopted for molecular docking. One approach is matching technique that uses a protein and the ligand as complementary surface (Halperin *et al.*, 2002). The second approach uses the ligand-protein pairwise interaction to simulate the calculation of the energies of the actual docking process (Wass *et al.*, 2011). The geometric matching or shape complementarity is fast and robust but it cannot be used to model dynamic changes in the ligand/protein conformation (Dutta *et al.*, 2017). Simulation approach, on the other hand, is more complicated because there is a physical distance in which the ligand only finds its space in the protein active site when there is a shift towards the conformation site. Such shift can be in the form of translation, rotation (such as torsion angle rotation) or internal changes to the ligand's structure (Norel *et al.*, 1995).

5.3.1. Visualization of the Structures

Graphical tools, called viewers, that enable protein structure visualization are parts of the standard software equipment of Bioinformaticians and Molecular Biologists. Among these is Rasmol which is a free, open-source and molecular graphics program. Three-dimensional structures can be constructed and displayed if the atomic-resolution coordinates of a biomolecule or its complex are available (Bernstein, 2000). However, RasMol is not suitable for homology modelling, to study the structural effect of mutations, to minimise energy generated during structures prediction or for molecular dynamics simulations (Goodsell, 2005). Pymol is another open-source molecular graphics program and it is probably the one most often used by Molecular Biologists to visualize three-dimensional structures of macromolecules (DeLano, 2002). The program is free for academic users and students, except for its molecular modelling modules and its advantage over RasMol is that it has superb publication-quality graphics outputs (Mottarella *et al.*, 2010).

5.4. Aim

The specific aim of this chapter was to predict *de novo,* the structures of the putative anti-pneumonia AMPs and pneumonia protein ligands for docking interaction analysis.

5.5. Objectives

The objectives of this chapter includes:

- Use I-TASSER server to predict the structures of the putative anti-pneumonia AMPs as ligands and the pneumonia proteins as receptors;
- Visualize the predicted structures using Pymol version 1.3;
- Compute the 3-D parameter scores from I-TASSER;
- Use PATCHDOCK server for the docking interaction analysis of the putative anti-pneumonia AMPs and the pneumonia proteins;

- Visualize the predicted structures using Rasmol version 2.7.5; and
- Compute the binding scores of the docking interactions.

5.6. Materials and Methods

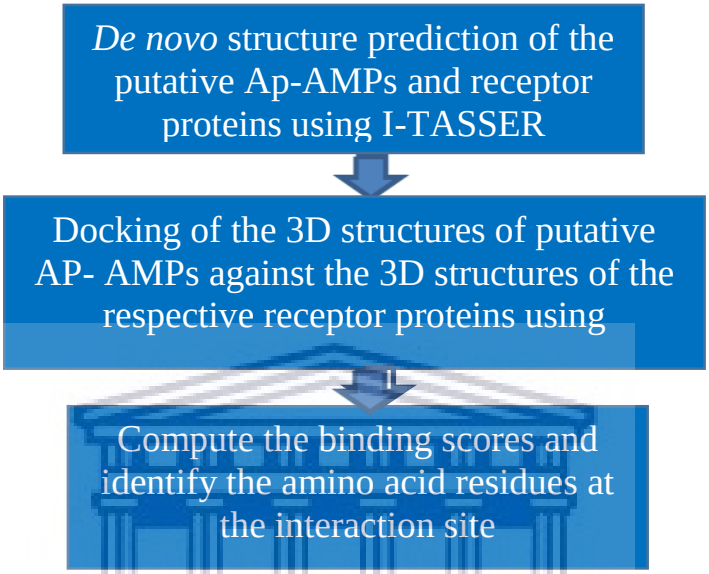


Figure 5. 1: Schematic representation of the methodology used within this chapter.

5.6.1. *De novo* structure predictions of the putative anti-Pneumonia AMPs and Pneumonia proteins (receptors) using I-TASSER

3-D structures of the anti-Pneumonia AMPs and the specific viral/bacterial pneumonia receptor protein was predicted by uploading each sequence onto the I-TASSER website. The user enters their email address to which the results link will be sent. After, naming the uploaded sequence, the menu "Run I-TASSER" was selected (Roy *et al.*, 2010).

5.6.2. PyMol Analysis of the I-TASSER 3-D Structures

The 3-D structures of the AMPs and their respective protein receptors were visualized using the PyMOL version 1.3. This was achieved by downloading the latest version of the PyMol on Ubuntu Linux, extracting and installing it using the terminal command line.

5.6.3. Docking Analysis of the Putative Anti-pneumonia AMPs and Pneumonia Proteins 3-D Structures using PATCHDOCK

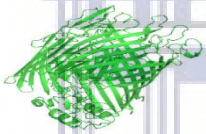
The 3-D structures of the anti-bacterial, anti-viral pneumonia putative AMPs and the bacterial and viral pneumonia protein receptors PDB files from I-TASSER were uploaded onto the PatchDock server, after which the user enters an email address. The cluster RMSD was set to 4.0Å and the complex type was selected as "protein-small ligand". The task was submitted by selecting "Submit Form". The docking results were sent via an email notification containing the web link to the docking results. Interaction analysis of the complex formation between the anti-bacterial with anti-viral pneumonia putative AMPs and their respective bacterial and viral pneumonia protein receptors was done using the RasMol 2.7.5. Software.

5.7. Results and Discussion

5.7.1. Structure Prediction of the putative anti-pneumonia AMPs and Pneumonia protein receptors

Representative output images from the I-TASSER server after predicting the 3-D structures of the anti-pneumonia AMPs (ligands) and the protein receptors are indicated in figure 5.2. The results indicates that all AMPs predicted exhibited various secondary structures including α-helices, parallel β-sheet, anti-parallel β-sheet, extended and loop conformational structures. The results observed are in line with the various structural conformations exhibited by known AMPs. Examples of known AMPs and their structures include Tachyplesin from horseshoe crabs and bovine Lactoferricin, which have a beta-sheet conformation (Ramamoorthy *et al.*, 2006); Magainin analog and Melittin having alpha helical conformations (Sengupta *et al.*, 2008).

Following the prediction of the putative anti-pneumonia peptides, it still had to be concluded whether these sequences can be considered *bona vide* AMPs. To classify these sequences as *bona vide* AMPs it had to conform to known AMPs in terms of characteristics as well as structure. As seen in Chapter 2, physicochemical characterization of the sequences was carried out on the predicted peptides with the results indicating that the peptides conform to known AMPs. In this Chapter the structures of the peptides was predicted and the results indicates that these peptides conform to known AMPs using structure as the measurement. Taken together, it can be concluded that based on their characteristics and structures, these peptides can be considered *bona vide* AMPs. However, the AMPs are still considered putative anti-pneumonia peptides due to lack of experimental evidence of their activity for these molecules currently.



(a) Receptor protein sample

(b1) Alpha helix AMP **(b2) Beta sheet AMP** **(b3) Extended AMP**

Figure 5. 2: 3D structures of the AMPs and pneumonia protein as determined by I-TASSER. 3D structure of (a) *Acinetobacter baumannii* outer membrane protein (b1) alpha helical AMP, (b2) beta sheet AMP, (b3) extended sheet AMP.

For structure prediction evaluation using I-TASSER (table 5.1 – 5.8) several parameters such as C-score, TM-score and RMSD was used for the prediction of the AMPs and pneumonia protein receptor 3-D structures.

5.7.1.1. C-Score

C-score is a confidence score for estimating the quality of predicted models by I-TASSER. Its calculation is based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations, which is typically in the range of -5 to 2, where a C-score within this range of values, signifies a model with a high confidence (Zhang, 2008).

The results indicates that the C-score of all the predicted 3-D structures for the anti-bacterial and anti-viral pneumonia AMPs and the pneumonia receptor proteins were between the values of -5 to 2 (see table) especially the C-score of the pneumonia receptor proteins for which the 3-D structures have already been solved indicating existing templates within protein databases for use by I TASSER for their structural prediction. Although the structures of these proteins exist, it was considered best to predict it once again with I TASSER to ensure uniformity (same parameters for structure prediction) for both the AMPs as well as their protein receptors.

The calculated C-scores of BOPAM-KP2, 3, 4, BOPAM-RSV11, BOPAM-SP3, 4, and 6 were lower than that of the other AMPs; and could indicate that these molecules did not have an available template for their prediction but were still not randomly predicted (Roy *et al.*, 2010). The lack of templates for the prediction of these molecules can indicate their novelty.

5.7.1.2. TM Score

TM-score, on the other hand, is a recently proposed scale for measuring the structural similarity between two structures (Zhang and Skolnick, 2004). A TM-score >0.5 indicates a model of correct topology and a TM-score <0.17 means a random similarity. These cut-offs do not depend on the protein length. From the results, the TM-score of the predicted structures of the AMPs were higher than the cut-off value of 0.5, except for BOPAM-KP2 with a TM-score of 0.49±0.15. BOPAM-KP2 also had a lower than expected C-score adding to the notion that

there is no templates for this particular molecule at the time of the study. AMPs with TM-scores higher than 0.5, signifies structural similarity with the templates that were used to predict their structures (Roy *et al.*, 2010, Zhang, 2008).

5.7.1.3. RMSD

Although there is not a defined RMSD value for 3-D structure prediction, a RMSD value of 2 - 4 Å is considered good and a RMSD ≤ 1 Å is considered ideal. Thus, all anti-bacteria and anti-viral pneumonia AMPs having RMSD within the accepted range (see table 5.1) had less distance and atomic deviation between the superimposed peptides and the templates, which were used for their 3-D structure prediction (Park *et al.*, 1997, Wei *et al.*, 1999).

Table 5. 1: Quality assessment scores of the predicted 3-D structures of the pneumonia receptors.

S/N	Anti-Pneumonia proteins	C-score	Exp. TM Score	Exp. RSMD	No of Decoys	Cluster Density
1	AB Iron regulated OMP	-0.18	0.69±0.12	8.8±4.6Å	2063	0.1476
2	INFA Membrane protein M1	-2.86	0.39±0.13	10.2±4.6Å	1540	0.0266
3	INFA Nucleoprotein	0.98	0.85±0.08	5.2±3.3Å	2104	0.4054
4	INFB Nucleoprotein	-0.96	0.59±0.14	9.8±4.6Å	1096	0.0697
5	RSV Chain A Matrix protein	1.99	0.99±0.04	2.0±1.6Å	10200	1.2500
6	RSV Chain X Fusion core protein	-0.21	0.70±0.12	2.6±1.9Å	7581	0.4296
7	SP Pneumolysin	1.77	0.96±0.05	3.5±2.4Å	2400	1.0000
8	KP Iron regulated OMP	-0.47	0.70±0.12	8.5±4.5Å	2035	0.1790

S/N= Serial number; AB= *Acinetobacter baumanni*; INFA= *Influenza A* virus; INFB- *Influenza B* virus; RSV= *Respiratory Syncytial* virus; SP= *Streptocococcus pneumoniae*; KP= *Klebsiella pneumoniae*.

The purpose of proposing TM-score is to solve the problem with RMSD, which is sensitive to local error since RMSD is an average distance of all residue pairs in two structures. For instance, a mis-orientation of the structure will give rise to a big RMSD value although the

global topology of the structure is correct. TM-score is not sensitive to mis-orientation in the distance of the residues, which makes the score insensitive to the local modeling error and thus a more reliable measure.

In summary, both TM and RMSD scores are known standards for measuring structural similarity between two structures for accuracy of structure modeling when the native structure is known (Roy *et al.*, 2010). The C-score is a metric developed for I-TASSER to estimate the confidence of the modeling, however, when the native structure is unknown, it becomes imperative to predict the quality of the modeling, that is, the distance between the predicted model and the native structures. This is the reason for calculating the TM and RMSD scores of the predicted models relative to the native structures, which is based on the C-score.

Table 5. 2: Quality assessment scores of the predicted 3-D structures of the putative anti-*Acinetobacter baumannii* AMPs.

S/N	Putative AMP	C-Score	Exp. TM Score	Exp. RSMD
1	BOPAM-AB1	-0.51	0.65±0.13	1.6±1.4Å
2	BOPAM-AB2	-0.49	0.65±0.13	1.6±1.4Å
3	BOPAM-AB3	-0.27	0.68±0.12	1.2±1.2Å
4	BOPAM-AB4	-0.32	0.67±0.13	1.3±1.3Å

Table 5. 3: Quality assessment scores of the predicted 3-D structures of the putative anti-*Influenza A* Virus AMPs.

S/N	Putative AMP	C-Score	Exp. TM Score	Exp. RSMD
1	BOPAM-INFA1	-1.18	0.57±0.15	2.5±1.9Å
2	BOPAM-INFA2	-0.21	0.69±0.12	0.8±0.8Å
3	BOPAM-INFA3	-1.61	0.57±0.15	2.4±1.8Å
4	BOPAM-INFA4	-1.40	0.54±0.15	2.9±2.1Å
5	BOPAM-INFA5	-1.21	0.56±0.15	2.5±1.9Å
6	BOPAM-INFA6	-1.19	0.57±0.15	2.5±1.9Å
7	BOPAM-INFA7	-1.11	0.58±0.14	2.3±1.8Å
8	BOPAM-INFA8	-1.14	0.57±0.15	2.4±1.8Å

Table 5. 4: Quality assessment scores of the predicted 3-D structures of the putative anti-*Influenza B* Virus AMPs.

S/N	Putative AMP	C-Score	Exp. TM Score	Exp. RSMD
1	BOPAM-INFB1	-0.56	0.64±0.13	1.5±1.4Å
2	BOPAM-INFB2	-0.57	0.64±0.13	1.5±1.4Å
3	BOPAM-INFB3	-1.09	0.58±0.14	2.4±1.8Å
4	BOPAM-INFB4	0.16	0.73±0.11	0.5±0.5Å
5	BOPAM-INFB5	-0.33	0.67±0.13	1.1±1.1Å
6	BOPAM-INFB6	-1.12	0.57±0.14	2.5±1.9Å

Table 5. 5: Quality assessment scores of the predicted 3-D structures of the putative anti-*Klebsiella pneumonia* AMPs.

S/N	Putative AMP	C-Score	Exp. TM Score	Exp. RSMD
1	BOPAM-KP2	-1.84	0.49±0.15	4.5±3.0Å
2	BOPAM-KP3	-1.55	0.52±0.15	3.9±2.7Å
3	BOPAM-KP4	-1.52	0.53±0.15	3.9±2.6Å
4	BOPAM-KP5	-0.18	0.69±0.12	1.6±1.4Å
5	BOPAM-KP6	-0.06	0.71±0.12	1.2±1.2Å

Table 5. 6: Quality assessment scores of the predicted 3-D structures of the putative anti-*Respiratory Syncytial* Virus AMPs.

S/N	**Putative AMP**	**C-Score**	**Exp. TM Score**	**Exp. RSMD**
1	BOPAM-RSV1	-0.03	0.71±0.12	0.7±0.7Å
2	BOPAM-RSV2	0.02	0.72±0.11	0.6±0.6Å
3	BOPAM-RSV3	-0.03	0.71±0.12	0.7±0.7Å
4	BOPAM-RSV4	-1.20	0.56±0.15	2.8±2.0Å
5	BOPAM-RSV5	-0.62	0.63±0.13	1.7±1.5Å
6	BOPAM-RSV6	-0.18	0.69±0.12	1.0±1.0Å
7	BOPAM-RSV7	-0.10	0.70±0.12	0.8±0.8Å
8	BOPAM-RSV8	-1.24	0.56±0.15	2.8±2.1Å
9	BOPAM-RSV9	-0.74	0.62±0.14	1.9±1.6Å
10	BOPAM-RSV10	-1.10	0.58±0.14	2.6±1.9Å
11	BOPAM-RSV11	-1.63	0.52±0.15	3.6±2.5Å
12	BOPAM-RSV12	-1.42	0.54±0.15	3.2±2.2Å
13	BOPAM-RSV13	-0.14	0.70±0.12	0.9±0.9Å

Table 5. 7: Quality assessment scores of the predicted 3-D structures of the putative anti-*Streptococcus pneumonia* AMPs.

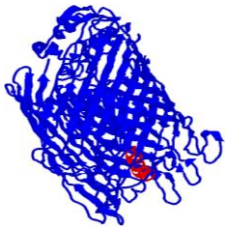
S/N	Putative AMP	C-Score	Exp. TM Score	Exp. RSMD
1	BOPAM-SP1	-0.05	0.71±0.12	2.5±1.9Å
2	BOPAM-SP2	-0.00	0.71±0.11	2.4±1.8Å
3	BOPAM-SP3	-1.74	0.50±0.15	5.8±3.6Å
4	BOPAM-SP4	-1.70	0.51±0.15	5.7±3.6Å
5	BOPAM-SP5	-1.71	0.51±0.15	5.7±3.6Å
6	BOPAM-SP6	-1.75	0.50±0.15	5.8±3.6Å
7	BOPAM-SP7	-0.08	0.70±0.12	2.6±1.9Å

In Table 5.2 – 5.7, BOPAM-AB = Anti-*Acinetobacter baumanni* AMPs; BOPAM-INFA = Anti-*Influenza A* virus AMPs; BOPAM-INFB = Anti-*Influenza B* virus AMPs; BOPAM-RSV = Anti-*Respiratory Syncytial* virus AMPs; BOPAM-SP= Anti-*Streptococococcus pneumoniae* AMPs; BOPAM-KP = Anti-*Klebsiella pneumoniae* AMPs.

The C-score has a strong correlation with the quality of the final models, which has been used to quantitatively estimate the RMSD and TM-score of the final models relative to the native structure. Unfortunately, such strong correlation only occurs for the first predicted model from the largest cluster. Thus, the C-scores of the lower-rank models (i.e., models 2-5 of the result output from the I-TASSER server) are listed only for reference and a comparison among them is not advised. In other word, even though the lower-rank models may have a higher C-score than the first model in some cases, the first model is on average the most reliable and should be considered if without special reasons (e.g., from biological sense or experimental data).

5.7.2. Docking Interaction Analysis of the putative anti-pneumonia AMPs with Bacterial and Viral Pneumonia Receptors

The output images from the PATCHDOCK server after predicting the docking interaction between the anti-pneumonia AMPs (ligands) and the protein receptors were analysed. The spatial docking interaction analysis showed that all the AMPs bound tightly to their respective proteins. Additionally, computational analysis was carried out to confirm the AMPs with the highest binding affinity, the amino acid residues which took part in the binding interaction and towards which terminal of the proteins the binding takes place.



(c) Ligand-receptor complex

Figure 5.3: Representative of the docking interaction of a protein receptor and putative anti-pneumonia AMPs produced by PATCHDOCK and visualized using RasMol.

Table 5. 8: Quality assessment scores of the docking analysis for the anti-pneumonia putative AMPs and the pneumonia receptors

S/N	Receptors	Ligands	Binding Scores
1	Membrane protein M1	BOPAM-INFA1	8986
2	Membrane protein M1	BOPAM-INFA2	8728
3	Matrix protein M1	BOPAM-INFA3	9056
4	Matrix protein M1	BOPAM-INFA4	9046
5	Matrix protein M1	BOPAM-INFA5	8836
6	Matrix protein M1	BOPAM-INFA6	9278
7	Matrix protein M1	BOPAM-INFA7	8806
8	Matrix protein M1	BOPAM-INFA8	10182
9	Nucleoprotein	BOPAM-INFA1	12134
10	Nucleoprotein	BOPAM-INFA2	10870
11	Nucleoprotein	BOPAM-INFA3	10572
12	Nucleoprotein	BOPAM-INFA4	12300
13	Nucleoprotein	BOPAM-INFA5	11778
14	Nucleoprotein	BOPAM-INFA6	12110
15	Nucleoprotein	BOPAM-INFA7	10982
16	Nucleoprotein	BOPAM-INFA8	12604
17	Nucleoprotein	BOPAM-INFB1	12816
18	Nucleoprotein	BOPAM-INFB2	13042
19	Nucleoprotein	BOPAM-INFB3	11606
20	Nucleoprotein	BOPAM-INFB4	12028
21	Nucleoprotein	BOPAM-INFB5	11704
22	Nucleoprotein	BOPAM-INFB6	11584
23	Iron regulated outer membrane protein	BOPAM-KP2	13036
24	Iron regulated outer membrane protein	BOPAM-KP3	12384
25	Iron regulated outer membrane protein	BOPAM-KP4	12960
26	Iron regulated outer membrane protein	BOPAM-KP5	10810

27	Iron regulated outer membrane protein	BOPAM-KP6	13208
28	Pneumolysin	BOPAM-SP1	12306
29	Pneumolysin	BOPAM-SP2	13606
30	Pneumolysin	BOPAM-SP3	12116
31	Pneumolysin	BOPAM-SP4	12384
32	Pneumolysin	BOPAM-SP5	12514
33	Pneumolysin	BOPAM-SP6	13306
34	Pneumolysin	BOPAM-SP7	11830
35	Chain X Fusion protein core	BOPAM-RSV1	5504
36	Chain A Protein	BOPAM-RSV1	8906
37	Chain X Fusion protein core	BOPAM-RSV2	5898
38	Chain A Protein	BOPAM-RSV2	8702
39	Chain X Fusion protein core	BOPAM-RSV3	5916
40	Chain A Protein	BOPAM-RSV3	9068
41	Chain A Protein	BOPAM-RSV4	9522
42	Chain X Fusion protein core	BOPAM-RSV4	6292
43	Chain A Protein	BOPAM-RSV5	8274
44	Chain X Fusion protein core	BOPAM-RSV5	5884
45	Chain X Fusion protein core	BOPAM-RSV6	5256
46	Chain A Protein	BOPAM-RSV6	8194
47	Chain X Fusion protein core	BOPAM-RSV7	5422
48	Chain A Protein	BOPAM-RSV7	8784
49	Chain X Fusion protein core	BOPAM-RSV8	5702
50	Chain A Protein	BOPAM-RSV8	8820
51	Chain X Fusion protein core	BOPAM-RSV9	5488
52	Chain A Protein	BOPAM-RSV9	8500
53	Chain X Fusion protein core	BOPAM-RSV10	5980
54	Chain A Protein	BOPAM-RSV10	9020
55	Chain X Fusion protein core	BOPAM-RSV11	5992
56	Chain A Protein	BOPAM-RSV11	8424
57	Chain X Fusion protein core	BOPAM-RSV12	6276
58	Chain A Protein	BOPAM-RSV12	8602
59	Chain X Fusion protein core	BOPAM-RSV13	5342
60	Chain A Protein	BOPAM-RSV13	8866
61	Iron Regulated OMP	BOPAM-AB1	10566
62	Iron Regulated OMP	BOPAM-AB2	11806
63	Iron Regulated OMP	BOPAM-AB3	12480
64	Iron Regulated OMP	BOPAM-AB4	11802

5.7.2.1. Anti-bacterial AMPs

Using this criteria of binding geometric score from table 5.8 above, BOPAM-AB3, 2 and 4 would be better diagnostic molecules for the detection of *Acinetobacter baumannii* since these peptides showed the highest binding geometry scores compared to any other putative anti-*Acinetobacter baumannii* AMP to the identified protein receptor of this pathogen.

In addition, BOPAM-SP2 and 6 would be the better diagnostic molecules against *Streptococcus pneumoniae* based on the peptides observed binding geometry scores to this organism's identified receptor. Furthermore, BOPAM-KP2 and 6 showed the highest binding geometry scores to the *Klebsiella pneumoniae* identified receptor than any other BOPAM-KPs

5.7.2.2. Anti-viral AMPs

BOPAM-RSVs bound more tightly to chain A protein than chain X fusion core protein with highest binding geometry score observed for BOPAM-RSV4. In other words, chain A protein of Respiratory Syncytial virus has more diagnostic relevance to detect this pathogen within a patient sample than chain X. In the same vein, the BOPAM-INFAs binds more tightly to nucleoprotein than membrane protein M1 of the *Influenza A* virus with highest binding geometry score observed for BOPAM-INFA4, whilst BOPAM-INFB2 and 1 are potential diagnostic molecules against the *Influenza B* virus based on the observed geometry scores for these docking interactions.

This result corresponds with the work of Tincho *et al.* (2016) where binding geometric scoring were used as the criteria in selection of candidate AMPs for HIV diagnostics. These observations were further confirmed using an "in house" lateral flow device in which the candidate AMPs was used to detect HIV in patient samples (Williams *et al.*, 2016).

The binding affinity of the putative anti-pneumonia AMPs corroborates the hypothesis that AMPs with a net positive charge less than +2 (Chapter 2) would have less electrostatic attraction towards the pathogenic organisms, thus a lesser binding affinity to these organisms and/or their proteins**.** BOPAM-INFA1, 2, 7, BOPAM-INFB3, 6, BOPAM-RSV3, 7, 12 and 13 which showed zero net charge through physiochemical analysis using APD and Bactibase, presented the lowest binding affinities with the pneumonia receptor proteins. These AMPs were also shown to have the lowest Boman indices as well as being predicted with the lowest E-values (table 2.4) using HMMER.

5.8. Conclusion

This chapter was carried out to predict the structures of the putative anti-pneumonia AMPs and pneumonia protein receptors for docking interaction analysis. From the I-TASSER results, the AMPs conformed structurally to known AMPs. Docking analysis of the 3D structures of the respective AMPs and their receptors showed high binding affinity of certain AMPs to their respective receptors. These AMPs also scored very well in terms of their net charge, Boman index as well as E-value. Thus several AMPs can already be shorlisted for use in a diagnostic test and will be further prioritized by anlysis within the subsequent chapter.

Chapter 6

6.0. *In silico* Site-Directed Mutagenesis (SDM) Analysis of Putative Anti-pneumonia Antimicrobial Peptides

6.1. Introduction

Site-directed mutagenesis (SDM) is a molecular biology method that is used to make specific and intentional changes to the DNA sequence of a gene and any gene products (Lockhart and Winzeler, 2000). It is used for investigating the structure and biological activity of DNA, RNA, and protein molecules, and for protein engineering. The application of site-directed mutagenesis to the study of protein function has been widely illustrated. For instance, it has become an important tool in biotechnology for the design of non-immunogenic antibodies for human therapeutics which underscores the practical benefits of mutagenesis and protein engineering; it is a powerful technique in the study of protein function which allows the assessment of the particular amino acid side-chains in a protein; it is most commonly used in the study of enzymes; and also very useful in identifying key residues in protein-protein interactions (Edelheit *et al.*, 2009). Site-directed mutagenesis is often beneficial in generating AMP derivatives since amino acids of AMPs play crucial roles in their functioning (Biro, 2006). Through SDM the functioning of AMPs can be augmented for instance increasing the interaction between the AMP and its target thus increasing the action of the AMP. However, site-directed mutagenesis is still an underexplored tool to determine the function of a gene or increase the function of a protein of interest because the principles guiding the use require a great deal of expertise (Goh *et al.*, 2012).

In chapter 2 and 3, novel AMPs were identified using *in silico* mathematical algorithms HMMER and CD-HIT. The binding of these AMPs to pneumonia pathogen receptors identified

in chapter 4 was carried out in chapter 5. In this chapter, the technique of site-directed mutagenesis will be used to increase the binding efficacy of the potential anti-pneumonia AMPs. Through increasing the binding affinity between the ligand and the receptor, the specificity of a potential test kit will be improved. Both sensitivity, the ability of a test kit to detect the pathogen at very low concentrations and specificity, the ability of a test kit to differentiate one pathogen from another and then binding to it with a high affinity, are paramount in creating a very good test kit.

Determination of the amino acid residues within the binding interface that contributes the most to the binding affinity known as "hotspots" residues are imperative to identify, as changes in these amino acids would greatly alter the structure and functioning of the identified putative AMPs. Thus non-"hotspot' residues will then be mutated to increase the efficacy of binding. Using a similar approach, Williams *et al.* (2016) demonstrated increased binding affinity of anti-HIV AMPs against p24 increasing both sensitivity and specificity of a Lateral flow device for detection of HIV 1 and 2 using AMPs instead of antibodies.

6.1.1. Knowledge-based FADE and Contacts (KFC)

The KFC server is a web-based implementation for the prediction of binding hotspots or subsets of residues that account for the protein's interface binding free energy (Darnell *et al.*, 2008). The server enhances the automated analysis of a user for a protein-protein or a protein-DNA interaction and the visualization of these "hotspot" predicted. It then characterizes its local structural environment and compares this with experimentally determined hotspots to predict if the interface residue is a hotspot (Guharoy *et al.*, 2011). Visualization is done using an interactive job viewer that is able to quickly highlight predicted hotspots and surrounding structural features within the protein structure.

The KFC server works on two predictive models principles namely; the basis of shape specificity properties and the other based on biochemical contacts (Darnell *et al.*, 2008). It has been investigated that KFC gives better predictive accuracy than computational alanine scanning (Robetta-Ala) for recognition of mutation sensitive residues (Hammami *et al.*, 2010).

6.2. Aim

The aim of this chapter was to use the parental anti-pneumonia AMPs as templates for the creation of derivative AMPs, which would bind the pneumonia protein receptors with greater affinity, demonstrated through *in silico* binding studies.

6.3. Objectives

The objectives of this chapter include:

- Analyse the amino acid residues at the interaction sites of the parental anti-bacterial and antiviral pneumonia AMPs with their respective receptors;
- Identify "hotspot" residues involved in the interaction using KFC;
- Perform *in silico* site-directed mutagenesis of non-"hotspot" residues;
- Perform physicochemical analysis of the mutated AMPs;
- Predict the 3-D structures of the mutated AMPs using I-TASSER; and
- Dock the mutated AMPs with the pneumonia receptors previously identified using PATCHDOCK.

6.4. Materials and Methods

6.4.1. Identification of hotspot residues

Before the commencement of *in silico* site-directed mutagenesis, essential amino acids have to be identified. Knowledge-based FADE and Contacts (KFC) server (http://kfc.mitchell-lab.org/upload.php) (Zhu and Mitchell, 2011) was employed for the precision identification of

"hotspot" residues or mutation sensitive residues to effectively perform *in silico* site-directed mutagenesis.

6.4.2. *In silico* site-directed mutagenesis:

An *in silico* approach was used to generate derivative AMPs that displayed increased predicted binding affinity to the pneumonia receptor proteins using the parental AMPs against viral and bacterial pneumonia as templates. All amino acids substituted to generate derivate AMPs were of similar characteristics to the amino acids present in the parental AMPs, as to maintain the predicted activity and functioning of the AMPs.

6.4.3. Physicochemical properties:

Submission of all derivative AMP sequences were carried out on the AMP characterization software Bactibase (physiochemical properties tab) (http://bactibase.pfba-lab- tun.org/ physicochem) (Fu *et al.*, 2010) and Antimicrobial Peptide Database (prediction tab) (http://aps.unmc.edu/AP/prediction/prediction_main.php) (Wang and Wang, 2004) to determine the characteristics of each derivative AMP.

6.4.4. *In silico* 3-D structure prediction:

Prediction of the derivative AMP's 3-D structures was carried out using the online software I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER) as previously described in section 5.6.1.

6.4.5. *In silico* protein-protein interaction study

The online protein-protein interaction server, PATCHDOCK, was employed in this study as carried out in section 5.6.3. Docking of each derivative AMP 3-D structure was carried out against its specific protein receptor's 3-D structure.

6.5. Results and Discussion

6.5.1 Knowledge-based FADE and Contacts (KFC) analysis

The Knowledge-based FADE and Contacts (KFC) server was used as a tool for the identification of “hotspots”, or the subset of residues that determined the receptor proteins interfaces' binding free energy. The results are tabulated in table 6.1 showing the “hotspot” residues for each interaction between a specific ligand and its receptor.

Table 6. 1: Identification of mutation sensitive amino acid residues "hotspots" using KFC.

S/N	Interacting Amino Acids of the Receptors	Interacting Amino Acids of the Putative AMPs
1	*Acinetobacter baumannii* Outer Membrane Protein: Arg188A, Arg358A, Asn597A, Arg600A, Val612A, Asp644A, Gln646A, Asp739A, Gln787A, Leu789A, Asn792A, Asn794A	BOPAM-AB1: Phe1a, Leu2a, Val5a, Leu9a, Ser11a, Ser14a, Gly15a, Leu16a, Leu17a
2	*Acinetobacter baumannii* Outer Membrane Protein: Glu180A, Arg188A, Ser189A, Arg203A, Ser206A, Arg207A, Arg600A, Val645A, Gln646A, Tyr736A, Arg737A, Gln787A, Leu789A, Arg790A	BOPAM-AB2: Phe2a, Gly6a, Lys7a, Leu9a, Ser11a, Gly15a, Leu16a, Leu17a
3	*Acinetobacter baumannii* Outer Membrane Protein: Arg176A, Arg188A, Ser189A, Arg203A, Asp204A, Asp431A, Ile433A, Leu473A, Leu595A, Asn597A, Ala598A, Arg600A, Asp644A, Leu789A, Arg790A, Asn793A	BOPAM-AB3: Phe1a, Phe2a, Pro3a, Leu8a, Leu9a, Lys10a, Leu13a, Phe14a, Leu17a
4	Acinetobacter baumannii Outer Membrane Protein: Arg188A, Ser189A, Gln587A, Asn597A, Ala598A, Val612A, Asp644A, Gln646A, Phe648A, Tyr736A, Asp739A, Gln787A, Leu789A	BOPAM-AB4: Phe1a, Phe2a, Ile4a, Val5a, Lys7a, Leu8a, Leu9a, Lys10a, Ser14a
5	Acinetobacter baumannii Outer Membrane Protein: Thr184A, Arg188A, Ser189A, Arg600A, Val612A, Asp644A, Val645Aa, Gln646A, Phe648A, Tyr736A, Asp739A, Gln787A, Leu789A	BOPAM-AB5: Thr4a, Gly6a, Lys7a, Ala8a, Lys11a, Arg14a, Ala15a, Asn17a
6	Matrix Protein M1: Pro54A, , Phe79A, Asn82A, Ala83A, Asn85A, Asn87A, Gly88A, Glu201A, Val205A, Gln208A, Tyr240A, Phe251A	BOPAM-INFA1: Thr1a, Pro2a, Thr3a, Phe4a, Gly7a, Pro10a, Ile11a

7	Matrix Protein M1: Pro54A, Arg78A, Phe79A, Asn82A, Glu201A, Val205A, Gln208A, Tyr240A, Met244A, Met248A, Phe251A	BOPAM-INFA2: Ile4a, Ala8a, Gln10a, Val11a, Leu12a, Pro13a, Lys14a
8	Matrix Protein M1: Pro54A, Phe79A, Asn82A, Ala83A, Asn87A, Glu201A, Gln208A, Tyr240A, Met244A, Met248A, Phe251A	BOPAM-INFA3: Thr1a, Thr3a, Phe4a, Ile5a, Val9a, Pro12a
9	Matrix Protein M1: Lys47A, Pro54A, Lys57A, Arg77A, Arg78A, Phe79A, Asn82A, Gln208A, Met212A, Tyr240A, Met248A	BOPAM-INFA4: Thr1a, Thr3a, Phe4a, Ile5a, Asp6a, Gln8a, Val9a, Gln14a
10	Matrix Protein M1: Pro54A, Phe79A, Asn82A, Asn87A, Glu201A, Val205A, Gln158A, Tyr240A, Met244A, Met248A, Phe251A	BOPAM-INFA5: Thr1a, Pro2a, Thr3a, Val9a, Pro10a, Gln13a
11	Matrix Protein M1: Ser13A, Ile14A, Val15A, Pro16A, Ser17A, Gln158A, His159A, Ser161A, His162A, Ala209A, Val213A, Leu229A, Lys230A, Asp232A	BOPAM-INFA6: Lys3a, Val9a, Pro10a, Ile11a, Pro12a, Glu13a, Gln14a
12	Matrix Protein M1: Ile14A, Pro16A, Ser17A, Ser157A, Gln158A, His162A, Met165A, Thr185A, Ala206A, Ala209A, Leu229A, Lys230A	BOPAM-INFA7: Cys1a, Val3a, Val11a, Phe12a, Pro13a, Lys14a
13	Matrix Protein M1: Lys47A, Pro54A, Arg78A, Phe79A, Asn82A, Ala83A, Asn87A, Glu201A, Val205A, Gln208A, Met248A	BOPAM-INFA8: Phe4a, Ile5a, Asp6a, Gly7a, Gln8a, Val9a, Pro10a
14	Nucleoprotein: Ser165A, Leu166A, Arg267A, Gly268A, Val270A, His272A, Phe338A, Glu339A. Asp340A, Arg342A, Val343A, Pro453A, Ser457A, Leu479A, Tyr487A	BOPAM-INFA1: Pro2a, Phe4a, Ile5a, Asp6a, Gly7a, Gln8a, Val9a
15	Nucleoprotein: Ser165A, Leu264A, Ile265A, Arg267A, Asp340A, Arg342A, Val343A, Pro453A, Val456A, Ser457A, Phe458A, Pro477A, Leu479A, Tyr487A	BOPAM-INFA2: Pro2a, Val3a, Leu5a, Ile9a, Gln10a, Lys14a
16	Nucleoprotein: Glu339A, Arg342A, Val343A, Ala387A, Thr390A, Asn395A, Pro453A, Glu454A, Asp455A, Val456A, Ser457A, Phe458A, Gly462A, Val463A, Leu479A	BOPAM-INFA4: Phe4a, Ile5a, Asp6a, Gly7a, Gln8a, Val9a, Pro10a, Gln14a

17	Nucleoprotein: Leu166A, Arg267A, Val270A, His272A, Glu339A, Asp340A, Val343A, Ser457A, Phe458A, Val463A, Pro477A, Leu479A	BOPAM-INFA5: Thr1a, Thr3a, Phe4a, Ile5a, Val9a, Ile11a, Gln13a, Gln14a
18	Nucleoprotein: Ser165A, Arg267A, Glu339A, Asp340A, Val343A, Glu454A, Asp455A, Ser457A, Phe458A, Val463A, Pro477A, Leu479A	BOPAM-INFA6: Pro2a, Phe4a, Ile5a, Asp6a, Gly7a, Val9a, Pro12a
19	Nucleoprotein: Ser165A, Leu166A, Leu264A, Ile265A, Arg267A, Phe338A, Asp340A, Val343A, Asn395A, Pro453A, Asp455A, Val456A, Ser457A, Phe458A, Tyr487A	BOPAM-INFA7: Pro2a, Val3a, Ile4a, Asp6a, Gln10a, Val11a, Phe12a
20	Nucleoprotein: Ser165A, Ile265A, Arg267A, Gly268A, Val270A, His272A, Ala336A, Glu339A, Arg342A, Val343A, Ser344A, Ile347A, Ala387A, Arg389A, Pro453A, Phe458A	BOPAM-INFA8: Thr1a, Pro2a, Thr3a, Ile5a, Val5a, Glu13a, Gln14a
21	Nucleoprotein: Ser289A, Ala290A, Val323A, Val324A, Arg325A, Arg398A, Arg447A, Glu452A, Met503A, Ser507A, Gly514A	BOPAM-INFB1: Met1a, Val3a, Ser4a, Arg6a, Trp7a, Thr8a, Phe9a, Leu10a, Val12a
22	Nucleoprotein: Arg116A, Lys125A, Gly151A, Arg170A, Arg235A, Glu253A, Arg256A, Phe257A	BOPAM-INFB2: Ser4a, Arg6a, Thr8a, Phe9a, Met11a, Val12a, Pro13a, Pro14a
23	Nucleoprotein: Arg116A, Leu119A, Ala120A, Asp123A, Lys125A, Asp137A, Glu140A, Lys142A, Glu143A, Thr150A, Gly151A, Pro234A	BOPAM-INFB3: Leu1a, Asn2a, Pro5a, Gln6a, Leu7a, Leu8a, Leu10A,
24	Nucleoprotein: Arg116A, Leu119A, Ala120A, Asp123A, Lys125A, Asp137A, Glu140A, Thr150A, Gly152A, Thr153A, Pro234A, Pro422A, Ala423A	BOPAM-INFB4: Leu4a, Gln5a, Leu7a, Leu8a, Leu10a, Lys1a, Val12a, Pro13a, Leu15a
25	Nucleoprotein: Aeg116A Leu119A, Ala120A, Lys125A, Phe129A, Asp137A, Glu140A, Glu143A, Gly151A, Gly152A, Thr153A, His201A, Thr232A, Pro234A, His420A	BOPAM-INFB5: Thr2a, Ile3a, Leu4a, Leu6a, Leu7a, Leu10a, Lys11a, Gln14a, Leu15a
26	Nucleoprotein: Gln203A, Ala229A, Ile244A, Val322A, Val323A, Arg325A, Val328A, Ser330A, Tyr394A, Glu395A, Asp396A, Gly449A, Met508A, Gly514A, Ala516A	BOPAM-INFB6: Leu1a, Asn4a, Leu7a, Cys9a, Asn11a, Asn13a, Pro14a, Gln15a

27	Chain A Protein: Trp37A, Cys93A, Asn95A, Val96A, Tyr199A, Ser200A, Ile225A, Asp227A, Ala230A	BOPAM-RSV1: Ile1a, Ile5a, Glu7a, Glu8a, Cys12a, Lys15a, Phe16a
28	Chain A Protein: Trp37A, Ile92A, Cys93A, Asn95A, Val96A, Tyr199A, Ser200A, Leu203A, Ile225A, Asp227A	BOPAM-RSV2: Asn6a, Glu7a, Ile9a, Asp10a, Leu13a, Ser14a
29	Chain A Protein: Trp37A, Met40A, Ile92A, Cys93A, Asn95A, Tyr199A, Ser200A, Leu203A, Ile225A, Asp227A	BOPAM-RSV3: Asn1a, Asp4a, Val5a, Lys8a, Ile9a, Asn12a, Thr13a
30	Chain A Protein: Leu27A, Ile30A, Arg31A, Asp34A	BOPAM-RSV4: Val2a, Glu4a, Ile5a, Ala7a, Asn11a
31	Chain A Protein: Trp37A, Ile92A, Cys93A, Leu128A, Thr130A, Tyr199A, Ser200A, Ile225A, Asp227A	BOPAM-RSV5: Ser6a, Ala7a, Gln8a, Asn10a, Lys11a, Asn15a
32	Chain A Protein: Trp37A, Pro39A, Met40A, Cys93A, Thr130A, Tyr199A, Ser200A, Leu203A, Val226A, Asp227A	BOPAM-RSV6: Lys1a, Gln4a, Ile5a, Ala8a, Ile9a, Gln12a
33	Chain A Protein: Trp37A, Cys93A, Asn95A, Thr130A, Tyr199A, Ser200A, Leu203A, Ile225A, Asp227A	BOPAM-RSV7: Lys6a, Asn10a, Gln11a, Ile13a, Asn14a
34	Chain A Protein: Trp37A, Pro39A, Met40A, Cys93A, Lys125A, Thr130A, Tyr199A, Ser200A, Leu203A, Ile225A, Asp227A	BOPAM-RSV8: Ile2a, Asn4a, Phe5a, Ser9a, Leu12a, Leu13a, Ser14a
35	Chain A Protein: Trp37A, Met40A, Cys93A, Asn95A, Lys125A, Leu128A, Thr130A, Tyr199A, Ser200A, Leu203A, Ile225A, Asp227A	BOPAM-RSV9: Ile2a, Ser3a, Lys4a, Thr6a, Asn7a, Asn10a, Thr11a
36	Chain A Protein: Trp37A, Met40A, Cys93A, Asn95A, Thr130A, Tyr199A, Ser200A, Leu203A, Ile225A, Asp227A	BOPAM-RSV10: Ser6a, Asn10a, Thr11a, Thr14a, Asn15a, Ile16a
37	Chain A Protein: Trp37A, Ser71A, Phe90A, Thr91A, Ile92A, Thr130A, Tyr199A, Ser200A, Leu203A, Ile225A, Val226A, Asp227A	BOPAM-RSV11: Asn1a, Val2a, Asp10a, Asn12a, Ala14a, Asp15a
38	Chain A Protein: Glu1A, Lys27A, Asn56A, Asn58A, Arg79A, Pro111A, Cys112A, Glu113A, Ile114A	BOPAM-RSV12: Asn4a, Ile5a, Asn8a, Lys12a, Phe13a, Ile16a
39	Chain A Protein: Trp37A, Tyr199A, Ser200A, Leu203A, Ile225A, Asp227A, Ala230A	BOPAM-RSV13: Leu5a, Glu7a, Lys8a, Asp11a, Arg12a
40	Chain X Fusion Protein Core: Leu27A, Asp34A, His38A	BOPAM-RSV1: Val2a, Lys6a, Ile9a, Lys13a

41	Chain X Fusion Protein Core: Ser15A, Gln18A, Lys22A, Gln25A	BOPAM-RSV2: Ile5a, Asn12a
42	Chain X Fusion Protein Core: Val19A, Lys22A, Ile23A, Ser26A	BOPAM-RSV3: Lys7a, Ala11a
43	Chain X Fusion Protein Core: Leu27A, Ile30A, Arg31A, Asp34A	BOPAM-RSV4: Val2a, Glu4a, Ile5a, Ala7a, Asn11a
44	Chain X Fusion Protein Core: Ile16A, Val19A	BOPAM-RSV5: Ser6a. Val9a
45	Chain X Fusion Protein Core: Asn24A, Arg31A	BOPAM-RSV6: Gln4a, Ala8a
46	Chain X Fusion Protein Core: Ser26A, Phe29A, Ser33A	BOPAM-RSV7: Val5a, Leu9a, Asn12a
47	Chain X Fusion Protein Core: Gln18A, Lys22A, Gln25A	BOPAM-RSV8: Leu12a, Ser14a
48	Chain X Fusion Protein Core: Val19A, Lys22A, Ser26A	BOPAM-RSV9: Asn1a, Val5a, Asn12a, Ile16a
49	Chain X Fusion Protein Core: Phe1A, Ser26A, Ile30A	BOPAM-RSV10: Ile2a, Ser6a, Thr13a
50	Chain X Fusion Protein Core: Phe12A, Ser15A, Gln18A, Val19A, Lys22A	BOPAM-RSV11: Val2a, Asn7a, Asp10a, Asn12a
51	Chain X Fusion Protein Core: Pro4A, Ser15A, Ile16A, Val19A	BOPAM-RSV12: Ile2a, Thr3a, Ala11a
52	Chain X Fusion Protein Core: Asp3A, Leu5A, Ser17A, Asn20A, Asn24A	BOPAM-RSV13: Glu7a, Asp10a, Asp11a, Thr14a
53	Pneumolysin: Thr57A, Ser58A, Asp59A, Met97A, Thr98A, Tyr99A, Ser100A, Lys196A, Ile198A, Thr201A, Ser203A, Asp205A, Ala206A, Asp212A, Ser239A, Ala241A, Ser330A, Thr332A, Phe335A, Val341A, Thr343A	BOPAM-SP1: Arg3a, Asp4a, Asp5a, Arg6a, Cys8a, Met12a, Ile24a, Thr26a, Phe27a, Ser34a, Ile38a, Cys39a, Asn43a, Gly44a
54	Pneumolysin: Ser58A, Asp59A, Thr98A, Tyr99A, Ser100A, Ile101A, Gln149A, Glu151A, Ile198A, Thr201A, Ser203A, Asp205A, Ser239A, Ala241A, Thr332A, Phe335A, Val341A, Thr343A	BOPAM-SP2: Arg3a, Asn4a, Cys8a, Met12a, Thr24a, Thr26a, Phe27a, His29a, Ser34a, Ile38a, Asn41a, Lys42a
55	Pneumolysin: Glu42A, Ser254A, Ser256A, Glu277A, Gln280A, Ile281A, Asn284A, Thr356A, Ala357A, Thr358A, Arg359A, Leu447A, Val448A, Asn470A, Asp471A	BOPAM-SP3: Phe1a, His13a, His14a, Gln15a, Lys16a, Leu17a, Val18a, Phe19a, Asp23a, Asn28a, Cys32a, Ala33a, Ile35a, Leu37a, Met38a

56	Pneumolysin: Val45A, Glu47A, Ser254A, Ser256A, Glu277A, Gln280A, Ile281A, Lys354A, Thr356A, Tyr358A, Arg359A, Arg419A, Pro446A, Leu447A, Val448A, Arg449A, Val468A, Asn470A, Asp471A	BOPAM-SP4: Arg2a, His3a, His13a, His14a, Gln15a, Lys16a, Leu17a, Val18a, Phe20a, Asp23a, Ser27a, Asn28a, Lys32a, Met38a, Ile44a
57	Pneumolysin: Lys19A, Leu20A, His23A, Glu26A, Val78A, Asp79A, Glu80A, Leu83A, Glu84A, Glu159A, Lys162A, Ser167A, Glu170A, Glu231A, Val351A	BOPAM-SP5: Phe1a, Arg2a, His3a, Glu4a, Phe20a, Val24a, Lys29a, Gly30a, Ile34a, Ile35a, Gly36a, Met38a
58	Pneumolysin: Glu42A, Thr253A, Ser254A, Lys255A, Ser256A, Glu277A, Ile281A, Asn284A, Ala357A, Tyr358A, Arg359A, Pro446A, Leu447A, Val448A, Asn470A	BOPAM-SP6: Glu11a, His14a, Gln15a, Lys16a, Leu17a, Val18a, Phe31a, Leu37a, Met38a
59	Pneumolysin: Thr55A, Thr57A, Asp59A, Met148A, Gln149A, Tyr150A, Glu151A, Lys164A, Phe165A, Asn194A, Lys196A, Glu264A, Lys268A, Val270A, Val272A, Ile314A, Glu315A, Phe344A	BOPAM-SP7: Arg3a, Cys8a, Asn25a, Phe27a, Asp34a, Ile38a, Lys40a, Asp41a, Lys42a, Asn43a, Gly44a
60	Iron Regulated Outer Membrane Protein: Ala94A, Arg108A, Thr109A, Ser111A, Arg112A, Tyr256A, Phe319A, Pro321A, Pro323A, Ser332A, Ser334A, Phe376A, Tyr501A, Tyr507A, Ser519A, Arg578A, Val620A, Thr621A, Gln696A	BOPAM-KP2: Lys3a, Tyr4a, Val5a, Lys7a, Gly9a, Leu10a, Asn11a, Gly13a, Gln15a, Lys17a, Ile18a, Asp19a, Asn20a
61	Iron Regulated Outer Membrane Protein: Arg93A, Lys500A, Tyr501A, Leu515A, Leu517A, Thr561A, Ile562A, Val563A, Val564A, Asp576A, Val609A, Trp618A, Val620A, Thr621A, Ala654A, Ala655A, Arg697A, Pro704A	BOPAM-KP3: Met1a, Lys3a, His4a, Val6a, Lys7a, Leu8a, Val14a, Gln15a, Cys16a
62	Iron Regulated Outer Membrane Protein: Gln81A, Asn85A, Thr109A, Tyr256A, Phe319A, Pro321A, Leu325A, Ser332A, Ser334A, Ser336A, Gln338A, Gln374A, Tyr507A, Ile518A, Ile693A, Gln696A, Arg697A, Ala698A, Leu700A, Leu711A	BOPAM-KP4: Ile1a, His2a, His3a, Glu4a, Ala5a, Lys7a, Gly8a, Tyr10a, Pro12a, Tyr13a, Leu14a, Trp16a, Leu18a
63	Iron Regulated Outer Membrane Protein: Ala22A, Gln23A, Gly52A, Gln53A, Glu56A, Glu123A, Lys149A, Asp158A, Glu160A, Phe590A, Asp595A, Ala635A, Ala672A, Lys677A, Thr728A	BOPAM-KP5: Arg7a, Cys12a, Ser13a, Ala14a, Ser15a, Leu16a, Lys17a, Cys18a, Trp19a, Phe20a

64	Iron Regulated Outer Membrane Protein: Ser79A, Gln81A, Asn85A, Met88A, Ala94A, Arg108A, Thr109A, Tyr256A, Phe376A, Tyr507A, Ile518A, Ser519A, Ile693A, Gln696A, Ala698A, Leu700A	BOPAM-KP6: Ala5a, Lys6a, Ala8a, Glu11a, Lys13a, Cys15a, Lys16a, Leu18a, Ala19a, Lys20a, Lys21a
65	Iron Regulated Outer Membrane Protein: Ser79A, Gln81A, Asn85A, Gly87A, Met88A, Arg93A, Ala94A, Arg108A, Thr109A, Ser111A, Arg112A, Tyr256A, Ser234A, Lys500A, Tyr501A, Tyr502A, Tyr507A, Ile518A, Ile562A, Ile693A, Gln696A, Arg697A, Tyr712A	BOPAM-KP7: Leu1a, Arg3a, Glu4a, Val8a, Ser9a, His11a, Cys12a, Leu16a, Cys18a, Arg20a, Met22a

Any amino acid "hotspot" residues identified within the interaction site cannot be substituted during SDM analysis in order to maintain their functionality as these residues contributes the most to the free energy involved in binding.

6.5.2. *In silico* site directed mutagenesis of the AMPs

Having identified the "hotspots" residues at the interaction sites between the anti-pneumonia AMPs and their receptors, site-directed mutagenesis was performed with these parental anti-pneumonia AMPs as input. All amino acids substituted to generate derivate AMPs were of similar characteristics to the amino acids present in the parental AMPs, as to maintain the predicted activity and functioning of the AMPs. BOPAM-AB1, 2 and 3 were mutated at position 10 where lysine was changed for arginine while BOPAM-AB4 and 5 had leucine and isoleucine substituted for methionine at position 16. Moreover, BOPAM-SP1, 2, 3, 4 and 5 had serine or asparagine substituted for Aspartate at position 25 while lysine was substituted for arginine in BOPAM-SP6 and 7 at position 36. Cysteine and tryptophan of BOPAM-KP1, 2, 3 and 7 were mutated at position 2 while glycine and serine of BOPAM-KP4, 5 and 6 were mutated at position 9, all with asparagine respectively.

The anti-viral AMPs were also subjected to site-directed mutagenesis with lysine of BOPAM-INFA1, 2, 4, , 6 and 7 at position 13 substituted for arginine while BOPAM-INFA3, 5 and 8 had a glycine at position 7, which was substituted for threonine. BOPAM-INFB1 and 2 had histidine at position 5 which were substituted for arginine, while BOPAM-INFB3 had cysteine at position 12 which was substituted for serine. BOPAM-INFB 5, 6 and 4 had leucine at position 8 and 15 which were substituted for methionine and phenylalanine respectively. Lastly, BOPAM-RSV1, 2, 3, 4, 6, 7, and 8 had serine, threonine, valine and tyrosine which were mutated at position 3 with asparagine while BOPAM-RSV5, 9, 10 and 13 had threonine and isoleucine which were substituted at position 13 with arginine.

6.5.3. Physicochemical properties of the derivative AMPs

Following substitution of amino acids distinct from the "hotspot" residues, mutated AMPs were again subjected to physicochemical analysis using APD and BACTIBASE servers as carried out in chapter 2. This was determined to ensure that the mutated AMPs retained the same characteristics as the parental ones.

6.5.3.1. Physicochemical properties of derived anti-bacterial AMPs

From the results observed in table 6.2, all the derivative anti-bacterial pneumonia AMPs had improved performance in terms of Boman Index and hydrophobicity compared to the parental AMPs. However, BOPAM-KP1,-3, and BOPAM-SP7 had slightly reduced hydrophobicity however, with a higher Boman Index. Similar result was generated for the site directed mutagenesis of anti-HIV AMP8 to yield AMP8.1 used for HIV diagnosis with reduced hydrophobicity however, with an improved binding potential to the HIV p24 protein (Williams *et al.*, 2016). Substituting an amino acid with one carrying a higher positive charge, the overall charge of the AMP and Boman Index will be increased. It is therefore expected that the derivative AMP would have a greater binding affinity to its receptor compared to its parental counterpart as a consequence of the specific amino acid substitution. However, the results indicated that BOPAM-SP3.1, 4.1, 5.1 and 6.1 had negative charges following SDM while BOPAM-KP1.1 was neutral. With the discovery of negatively charged AMPs such as maximin H5 which contradicts the believe that AMPs with high antimicrobial activities are cationic, there is an indication that charge might not affect an AMP's activity (Jodoin and Hincke, 2018).

6.5.3.2. Physicochemical properties of derived anti-viral AMPs

From the results observed, all derivative anti-viral pneumonia had improved performance in terms of Boman Index and hydrophobicity when compared to the parental AMPs. However, BOPAM-INFB3, BOPAM-RSV3 and 13 had slightly reduced hydrophobicity which does not

interfere with an AMP's binding potential as indicated above (Jodoin and Hincke, 2018). BOPAM-INFA5.1, BOPAM-RSV2.1, 4.1, 6.1, 8.1 and 11.1 of the anti-viral pneumonia AMPs had negative charges with neutral charge observed for BOPAM-INFA1.1, 4.1, BOPAM-INFB3.1, 6.1, BOPAM-RSV3.1 and 7.1 as indicated in table 6.3 below.

Table 6. 2: Physicochemical properties of the mutated anti-bacterial and anti-viral pneumonia AMPs

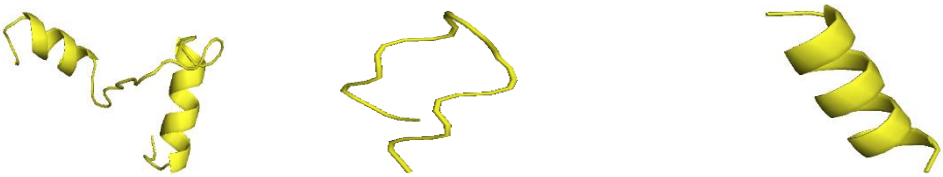
S/N	AMPs	Molecular Mass (Da)	% Hydrophobic	Common amino acid	Net charge	PI	Boman Index (kcal/mol)	Half-life in mammals (Hrs)
1	BOPAM-AB1.1	1783.33	52	L	+ 2	11.65	-1	1.1
2	BOPAM-AB2.1	1817.249	52	L	+2	11.65	-0.88	1.1
3	BOPAM-AB3.1	1937.58	64	L	+2	11.65	-1.63	1.1
4	BOPAM-AB4.1	1867.50	58	L	+2	10.81	-1.66	1.1
5	BOPAM-AB5.1	1714.1	47	G	+4	11.92	0.72	4.4
6	BOPAM-KP1.1	2496.90	30	SC	0	7.12	2.26	100
7	BOPAM-KP2.1	2297.701	38	N	+2	8.79	1.32	30
8	BOPAM-KP3.1	2406.25	42	N	+2	8.82	1.61	30
9	BOPAM-KP4.1	2589.19	28	K	+4	10.58	1.86	20
10	BOPAM-KP5.1	2694.32	39	K	+4	9.66	2.44	20
11	BOPAM-KP6.1	2357.05	38	K	+8	10.98	2.64	5.5
12	BOPAM-KP7.1	2600.13	39	S	+2	8.83	2.35	5.5
13	BOPAM-SP1.1	5084.34	27	R	+4	9.49	3.67	0.8
14	BOPAM-SP2.1	4904.86	25	G	+4	9,47	3.11	1.3
15	BOPAM-SP3.1	4793.03	45	G	-3	5.43	0.59	1.1
16	BOPAM-SP4.1	4 807.06	45	G	-3	5.43	0.57	1.1
17	BOPAM-SP5.1	4 807.06	45	G	-3	5.44	0.55	1.1
18	BOPAM-SP6.1	4 892.21	47	V	-1	6.78	0.72	2.8
19	BOPAM-SP7.1	5 168.24	25	K	+5	9.85	3.48	1.3
20	BOPAM-NFA1.1	1 569.43	28	P	0	6.34	1.57	7.2
21	BOPAM-NFA2.1	1 555.28	57	VIL	+1	8.55	0.52	1.2
22	BOPAM-NFA3.1	1 562.33	28	T	-2	3.49	1.63	7.2
23	BOPAM-NFA4.1	1 587.47	28	P	0	6.34	1.75	7.2
24	BOPAM-NFA5.1	1 585.69	28	TOP	-1	3.75	1.15	7.2
25	BOPAM-NFA6.1	1 596.50	28	P	+1	9.69	1.78	7.2

26	BOPAM-NFA7.1	1 605.29	50	ISV	+1	8.55	1.03	1.2
27	BOPAM-NFA8.1	1 586.40	28	PT	-2	3.55	1.24	7.2
28	BOPAM-NFB1.1	1 888.60	40	R	+2	12.20	2.95	30
29	BOPAM-NFB2.1	1 863.61	46	MRVP	+1	10.40	1.8	30
30	BOPAM-NFB3.1	1 671.56	40	NL	0	5.84	1.25	5.5
31	BOPAM-NFB4.1	1 798.04	66	L	+2	10.81	-1.44	5.5
32	BOPAM-NFB5.1	1 786.94	66	L	+1	9.70	-1.68	5.5
33	BOPAM-NFB6.1	1 690.34	46	N	0	5.76	0.7	5.5
34	BOPAM-RSV1.1	1 892.35	43	IKN	+1	8.54	1.6	20
35	BOPAM-RSV2.1	1 803.10	31	N	-1	4.43	2.49	1
36	BOPAM-RSV3.1	1 744.03	37	N	0	6.45	2.36	1.4
37	BOPAM-RSV4.1	1 797.11	43	N	-1	4.18	1.95	1.3
38	BOPAM-RSV5.1	1 742.30	37	N	+2	11.65	2.77	4.4
39	BOPAM-RSV6.1	1 840.79	50	I	-1	4.18	1.09	1.3
40	BOPAM-RSV7.1	1 854.40	31	N	0	6.41	3.01	1.4
41	BOPAM-RSV8.1	1 734.98	43	N	-1	3.85	1.41	1.9
42	BOPAM-RSV9.1	1 816.39	25	N	+2	11.65	3.27	1.4
43	BOPAM-RSV10.1	1 815.40	31	N	+1	9.69	2.99	100
44	BOPAM-RSV11.1	1 729.96	37	N	-1	4.11	2.39	1.4
45	BOPAM-RSV12.1	1 847.16	37	N	0	6.45	2.48	1.1
46	BOPAM-RSV13.1	1 985.35	25	DR	+1	9.53	5.36	1.3

BOPAM-AB = Anti-*Acinetobacter baumanni* AMPs; BOPAM-INFA = Anti-*Influenza A* virus AMPs; BOPAM-INFB = Anti-*Influenza B* virus AMPs; BOPAM-RSV = Anti-*Respiratory Syncytial* virus AMPs; BOPAM-SP= Anti-*Streptococococcus pneumoniae* AMPs; BOPAM-KP = Anti-*Klebsiella pneumoniae* AMPs.

6.5.4. Structure Prediction using I-TASSER

Figure 6.1 below showed the structures of derivative anti pneumonia AMPs with extended-partial α-helix and β-hairpin structures following SDM. With the slight changes in the structural characteristics of the AMPs post SDM. Derivative anti-pneumonia AMPs still contained structural features associated with this class of peptides



(c1) Beta sheet (c2) Extended sheet (c3) Alpha helix

Figure 6. 1: 3-D structures of the Derivative anti pneumonia AMPs using I-TASSER. The AMPs showed different secondary structures including alpha-helix, beta-sheets and extended conformations.

The table 6.3 below shows the quality output from I-TASSER server after predicting the derivative anti-pneumonia AMPs using C-score, TM score and RMSD as measure.

6.5.4.1. C score

The C-scores of all derivative AMPs were within the range -5 to 2 indicating accurate prediction from the templates used with correct topology. The C scores of the derivative AMPs did not deviate from the parental ones from which they were derived.

6.5.4.2. TM and RSMD scores

The TM scores were greater than 0.17 showing that the 3D structures of the derivative AMPs were not random predictions. The RMSD values were above 1Å further strengthening that the AMPs were correctly predicted.

Table 6. 3: Structure prediction scores of the anti-pneumonia AMPs

S/N	AMP Name	C-Score	TM-Score	RMSD
1	BOPAM-AB1.1	-0.45	0.66±0.13	1.5±1.4Å
2	BOPAM-AB2.1	-0.44	0.66±0.13	1.5±1.4Å
3	BOPAM-AB3.1	-0.16	0.69±0.12	1.0±1.0Å
4	BOPAM-AB4.1	-0.33	0.67±0.13	1.3±1.3Å
5	BOPAM-AB5.1	0.42	0.77±0.10	0.5±0.5Å
6	BOPAM-KP1.1	0.71	0.81±0.09	0.5±0.5Å
7	BOPAM-KP2.1	-1.45	0.54±0.15	3.7±2.6Å
8	BOPAM-KP3.1	-1.42	0.54±0.15	3.7±2.5Å
9	BOPAM-KP4.1	-1.56	0.52±0.15	3.9±2.7Å
10	BOPAM-KP5.1	0.03	0.72±0.11	1.2±1.2Å
11	BOPAM-KP6.1	-0.02	0.71±0.12	1.2±1.2Å
12	BOPAM-KP7.1	-0.57	0.64±0.13	2.3±1.8Å
13	BOPAM-SP1.1	0.04	0.72±0.11	2.3±1.8Å
14	BOPAM-SP2.1	-0.12	0.70±0.12	2.6±1.9Å
15	BOPAM-SP3.1	-1.88	0.49±0.15	6.1±3.8Å
16	BOPAM-SP4.1	-1.88	0.49±0.15	6.1±3.8Å
17	BOPAM-SP5.1	-1.97	0.48±0.15	6.3±3.8Å
18	BOPAM-SP6.1	-1.47	0.53±0.15	5.2±3.4Å
19	BOPAM-SP7.1	-0.14	0.70±0.12	2.7±2.0Å
20	BOPAM-INFA1.1	-1.03	0.58±0.14	2.2±1.7Å
21	BOPAM-INFA2.1	-0.14	0.70±0.12	0.7±0.7Å
22	BOPAM-INFA3.1	-1.50	0.53±0.15	3.1±2.2Å
23	BOPAM-INFA4.1	-1.05	0.58±0.14	2.2±1.7Å
24	BOPAM-INFA5.1	-0.97	0.59±0.14	2.1±1.7Å
25	BOPAM-INFA6.1	-1.00	0.59±0.14	2.1±1.7Å
26	BOPAM-INFA7.1	-0.71	0.62±0.14	1.6±1.4Å
27	BOPAM-INFA8.1	-1.13	0.57±0.14	2.4±1.8Å
28	BOPAM-INFB1.1	-0.77	0.62±0.14	1.9±1.5Å
29	BOPAM-INFB2.1	-0.78	0.61±0.14	1.9±1.6Å
30	BOPAM-INFB3.1	-1.19	0.57±0.15	2.6±1.9Å
31	BOPAM-INFB4.1	-0.03	0.71±0.12	0.6±0.6Å
32	BOPAM-INFB5.1	-0.35	0.67±0.13	1.1±1.1Å
33	BOPAM-INFB6.1	-0.91	0.60±0.14	2.1±1.7Å
34	BOPAM-RSV1.1	-0.00	0.71±0.11	0.7±0.7Å
35	BOPAM-RSV2.1	-0.01	0.71±0.11	0.7±0.7Å
36	BOPAM-RSV3.1	-0.23	0.68±0.12	1.1±1.1Å
37	BOPAM-RSV4.1	-1.53	0.53±0.15	3.4±2.3Å
38	BOPAM-RSV5.1	-0.37	0.67±0.13	1.3±1.3Å
39	BOPAM-RSV6.1	-0.85	0.61±0.14	2.1±1.7Å
40	BOPAM-RSV7.1	-0.13	0.70±0.12	0.9±0.9Å
41	BOPAM-RSV8.1	-1.42	0.54±0.15	3.2±2.2Å
42	BOPAM-RSV9.1	-0.11	0.70±0.12	0.9±0.9Å
43	BOPAM-RSV10.1	-0.78	0.61±0.14	2.0±1.6Å
44	BOPAM-RSV11.1	-1.26	0.56±0.15	2.9±2.1Å
45	BOPAM-RSV12.1	-1.46	0.53±0.15	3.2±2.3Å

46	BOPAM-RSV13.1	-0.10	0.70±0.12	0.8±0.8Å

BOPAM-AB = Anti-*Acinetobacter baumanni* AMPs; BOPAM-INFA = Anti-*Influenza A* virus AMPs; BOPAM-INFB = Anti-*Influenza B* virus AMPs; BOPAM-RSV = Anti-*Respiratory Syncytial* virus AMPs; BOPAM-SP= Anti-*Streptococcoccus pneumoniae* AMPs; BOPAM-KP = Anti-*Klebsiella pneumoniae* AMPs.

6.5.5. Docking interaction analysis of the derivative AMPs with Bacterial and Viral pneumonia Receptors using PATCHDOCK

The table 6.4 below showed the binding scores of the anti-bacterial and anti-viral pneumonia AMPs from the PATCHDOCK server with their respective receptors identified in chapter 4. All binding scores of the derivative AMPs were compared with parental AMPs for selection as candidate anti-pneumonia AMPs to be used in a lateral flow device (LFD).

6.5.5.1. Anti-bacterial mutated AMPs

BOPAM-AB1.1 and BOPAM-AB5.1 gave higher binding scores when compared to the parental AMPs against outer membrane receptor protein of *Acinetobacter baumannii*. Moreover, BOPAM-SP1.1, 3.1, 4.1 and 7.1 yielded improved binding scores against pneumolysin of *Streptococcus pneumoniae* while a similar result was observed for BOPAM-KP1.1, 3.1, 5.1 and 6.1 against iron-regulated outer membrane protein of *Klebsiella pneumoniae.*

6.5.5.2. Anti-viral mutated AMPs

Also, all mutated anti-Influenza A AMPs gave higher binding scores than the respective parental AMPs except BOPAM-INFA6.1 and 8.1 against the membrane protein M1 receptor, while similar improved binding efficacies was recorded for the AMPs against nucleoprotein of Influenza A virus except for BOPAM-INFA5.1. Only BOPAM-INFB5.1 had an increased binding score against *Influenza B* virus nucleoprotein, while BOPAM-

RSV3.1, BOPAM-RSV6.1, BOPAM-RSV9.1, BOPAM-RSV10.1, and BOPAM-RSV13.1 all had improved binding scores against *Respiratory Syncytial* virus Chain A, however, BOPAM-RSV2.1, BOPAM-RSV8.1, and 9.1 yielded relatively lower binding scores against chain X fusion protein core of *Respiratory Syncytial* Virus.

Table 6. 4: The docking analysis of only derivative viral and bacterial anti-pneumonia AMPs and their receptors which showed increased binding scores

S/N	Receptors	AMPs	Binding Score
1	Iron Regulated OMP	BOPAMAB1.1	11826
2	Iron Regulated OMP	BOPAMAB5.1	12952
3	Membrane protein M1	BOPAMINFA1.1	9198
4	Membrane protein M1	BOPAMINFA2.1	9392
5	Membrane protein M1	BOPAMINFA3.1	10708
6	Membrane protein M1	BOPAMINFA4.1	9728
7	Membrane protein M1	BOPAMINFA5.1	9236
8	Membrane protein M1	BOPAMINFA7.1	9250
9	Nucleoprotein	BOPAMINFA1.1	12346
10	Nucleoprotein	BOPAMINFA2.1	10888
11	Nucleoprotein	BOPAMINFA3.1	12578
12	Nucleoprotein	BOPAMINFA4.1	12458
13	Nucleoprotein	BOPAMINFA6.1	13256
14	Nucleoprotein	BOPAMINFA7.1	11120
15	Nucleoprotein	BOPAMINFA8.1	14170
16	Nucleoprotein	BOPAMINFB5.1	11932
17	Iron regulated outer membrane protein	BOPAMKP1.1	12268
18	Iron regulated outer membrane protein	BOPAMKP3.1	13216
19	Iron regulated outer membrane protein	BOPAMKP5.1	10984
20	Iron regulated outer membrane protein	BOPAMKP6.1	13870
21	Chain A Protein	BOPAMRSV3.1	9156
22	Chain A Protein	BOPAMRSV6.1	9236
23	Chain A Protein	BOPAMRSV9.1	9278
24	Chain A Protein	BOPAMRSV10.1	9158
25	Chain A Protein	BOPAMRSV13.1	9072
26	Chain X Fusion protein core	BOPAMRSV1.1	5506
27	Chain X Fusion protein core	BOPAMRSV3.1	5920
28	Chain X Fusion protein core	BOPAMRSV4.1	6566
29	Chain X Fusion protein core	BOPAMRSV5.1	5912
30	Chain X Fusion protein core	BOPAMRSV6.1	6610
31	Chain X Fusion protein core	BOPAMRSV7.1	6082
32	Chain X Fusion protein core	BOPAMRSV10.1	6068
33	Chain X Fusion protein core	BOPAMRSV11.1	6070
34	Chain X Fusion protein core	BOPAMRSV12.1	6422
35	Chain X Fusion protein core	BOPAMRSV13.1	5516
36	Pneumolysin	BOPAMSP1.1	13164
37	Pneumolysin	BOPAMSP2.1	12246
38	Pneumolysin	BOPAMSP3.1	14134
39	Pneumolysin	BOPAMSP4.1	12934
40	Pneumolysin	BOPAMSP5.1	11728

41	Pneumolysin	BOPAMSP6.1	11718
42	Pneumolysin	BOPAMSP7.1	12378

BOPAM-AB = Anti-*Acinetobacter baumanni* AMPs; BOPAM-INFA = Anti-*Influenza A* virus AMPs; BOPAM-INFB = Anti-*Influenza B* virus AMPs; BOPAM-RSV = Anti-*Respiratory Syncytial* virus AMPs; BOPAM-SP= Anti-*Streptococcus pneumoniae* AMPs; BOPAM-KP = Anti-*Klebsiella pneumoniae* AMPs.

All the derivative AMPs above (table 6.4) had increased binding affinity to their receptors when compared to their respective parental AMPs (Fig 6.2). In other words, they bound more tightly when compared to the respective parental AMPs. The mutation due to amino acids substitution in these derivative AMPs did not change their 3-D structures when compared to the parental ones and displayed a better-predicted fit at the binding pockets of bacterial and viral pathogens' receptors when compared to the parental AMPs**.**

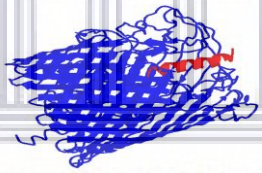


Figure 6. 2: Representative sample of the docking analysis of a mutated Ap-AMP with its receptor using PATCHDOCH and visualized using PyMOL. The blue colour represents the pneumonia receptor and the AMP is represented in red.

The improved binding affinity of the derived AMPs from the parental ones could be attributed to the amino acid residues which were used during the SDM analysis. However, there was no change in the binding region of the derivative AMPs with their receptors after the amino acids substitution of the parental AMPs. These physicochemical and binding characteristics of these AMPs would be very important for use in diagnostics. Also, these properties would be useful for the understanding of the mechanism of these peptides which allow them access to biological surfaces of membrane receptors. The

receptors used in this research have been carefully identified with their residues seating at the extracellular surface of the membrane for accessibility during binding.

The interaction studies indicates that the BOPAM-SPs bound pneumolysin at the extracellular site of *Streptococcus pneumoniae* at the N-terminal domain; BOPAM-KPs bound the extracellular outer membrane protein receptor *of Klebsiella pneumoniae* at the N-terminal domain; BOPAM-ABs bound the extracellular C-terminal of outer membrane protein receptors of *Acinetobacter baumannii*. In the same manner, binding affinity of the anti-viral AMPs occurred at the exposed regions of their receptors with BOPAM-INFAs bound matrix protein M1 at the C-terminal and Nucleoprotein at N-terminal of *Influenza A* virus; BOPAM-INFBs bound nucleoprotein of *Influenza B* virus at N-terminal; whilst BOPAM-RSVs bound both chain A and X proteins of *Respiratory Syncytial* virus at the N-terminal domains.

6.6. Conclusion

This chapter used the parental anti-pneumonia AMPs as templates for the identification of derivative AMPs for potential use in the differential diagnosis of pneumonia. The derivative AMPs have improved binding potential to the pneumonia receptors with greater affinity, demonstrated through *in silico* binding studies. Thus far, an additional 13 antibacterial and 20 anti-viral pneumonia AMPs with improved binding specificity and accuracy against pneumonia pathogens have been identified. These AMPs offer promising perspective to compete with current commercially available pneumonia diagnostic biomarkers to mitigate their shortcomings.

Chapter 7

7.0. General Discussion

7.1. Brief Overview

The serious concern from incessant mortality arising from pneumonia necessitates the quest for a standard diagnostic protocol to establish the status of patients before and after the onset of signs and symptoms. Mortality from this disease is a result of difficulty from treatment because there is no simple and reliable way to differentiate viral from bacterial cause of the disease. Several biomarkers, which hitherto have been used for the detection of pneumonia, are deemed unreliable as they are implicated in other diseases thereby giving false positive or false negative results. Blood culture results are valuable when positive for the detection of the disease, but negative results are more common even in severe pneumonia. Antibodies have been the gold standard for detection of samples due to its sensitivity to detection of targets within samples. However, it has recently been linked to various shortcomings, such as cross-reactivity and sequence coverage (Decuyper *et al.*, 2016).

As a result of inaccurate diagnosis, patients are treated with the wrong medication, which mostly include antibiotics. This leads to overuse of antibiotics and resistance by pathogens results from the overuse of antibiotics. This is worsened by the fact that there is no diagnostic protocol for the viral types of the disease. Hence the need for sensitive and specific detection of pneumonia is highly imperative.

The use of AMPs rather than antibodies to bind and recognize pneumonia biomarkers for diagnostics, offer several compensatory advantages over antibody shortcomings such as non-specificity and cost to mention only a few. With more accurate and sensitive diagnostics of pneumonia, the problem of indiscriminate overuse of antibiotics, toxicity due to the wrong

prescription, bacterial resistance, scarcity and the high cost of existing antibiotics can be reduced or even eliminated. The use of AMPs for HIV diagnostics on patient samples has been demonstrated by Williams *et al.* (2016), with a commercial kit currently in production.

The aim of this research work was to identify novel AMPs, using *in silico* techniques, to accurately differentiate between viral and bacterial pneumonia in patient samples.

7.1.1. Summary of Chapters

In chapter 2, experimentally validated AMPs showing activity towards *Streptococcus pneumoniae, Acinetobacter baumannii, Klebsiella pneumoniae, Respiratory Syncytial virus, Influenza A and B viruses* found in several databases such as APD (Wang and Wang, 2004; Wang *et al.*, 2008; Wang *et al.*, 2016), CAMP (Thomas *et al.*, 2009; Waghu *et al.*, 2014; Waghu *et al.*, 2015), AVPdb (Sencanski *et al.*, 2015) were retrieved. Literature and Data mining revealed after cross referencing between the databases, 80, 19, 12, 92, 50 and 12 experimentally validated AMPs as anti-pneumonia peptides for *Streptococcus pneumoniae, Acinetobacter baumannii, Klebsiella pneumoniae, Respiratory Syncytial virus, Influenza A and B viruses* respectively. Amino acid sequences were extracted to create a predictive model to potentially identify completely new AMPs against bacterial and viral pneumonia using HMMER. Performance measures of each model based on prediction of the positive and the negative testing datasets, tested the robustness of each HMMER profile, which was scanned against proteomes of various organisms to derive putative AMPs. Eighteen anti-bacterial and twenty seven anti-viral pneumonia AMPs were identified ranked according to their E-values, with the AMP predicted with the lowest E-value assumed to have the highest anti-pneumonia activity.

To ensure that the sequences identified were *bona vide* AMPs, their physicochemical properties were determined using the calculation and prediction section in APD and BACTIBASE. The result of this analysis showed that the predicted sequences had the same physicochemical properties characteristic of the AMP class of peptides. Sequence analysis of the predicted peptides showed no similarity to any experimentally validated AMPs indicating that all the identified AMPs can be considered novel.

In chapter 3, algorithm comparison was used to attest to the results obtained with HMMER using CD-HIT. Similarity was observed for the HMMER and CD-HIT algorithms, with their result showing the same "hits" for the same organisms. However, the HMMER algorithm showed result specific to the amino acid consensus, bit scores and E-values whereas CD-HIT gave result in percentage similarity. Apart from this, HMMER discriminated AMPs using insertion and rearrangement within the amino acid sequence whereas CD-HIT gave result in terms of genome loci.

In chapter 4, various PDB search engines such as Pubmed/NCBI, Google Scholar and ScienceDirect were accessed to identify the pneumonia receptors for the target organisms. The amino acid sequences of these proteins were extracted using NCBI and PDB that would serve as receptors to be identified by the AMPs (ligands).

In chapter 5, the 3D structure of the pneumonia receptors and AMPs were modeled using the I-TASSER server. The amino acid sequence of the receptors and the AMPs to be modeled were used individually as input in the FASTA format and submitted to the online server of I-TASSER and visualized using the PyMOL 1.3. Software. The AMPs showed secondary structures including α-helices, beta-sheets and extended shapes, seen in known AMPs (Schmidtchen *et al.*, 2014).

All anti-bacterial and anti-viral AMPs have similar structures, represented by extended partial α-helical structure or loop structure with partial α-helical secondary structure. Based on the C-score, TM-score and RMSD the 3D structures of the putative anti-pneumonia peptides as well of that of the receptor proteins, showed that of a good model. **In chapter 6**, the parental AMPs generated in chapter 2, were used as templates to generate derivative AMPs that would display increased predicted binding affinity for the pneumonia receptor proteins using site directed mutagenesis. Identification of mutation sensitive amino acids or "hotspot" amino acid residues of the AMPs using KFC server were performed and completed using site directed mutagenesis by changing 'non hot spot' amino acid residues within the binding interaction area to increase the binding affinity of the AMPs to their interacting proteins. The substitutions of amino acids ('non hot spots') introduced into the parental AMPs sequences were done based on the properties of the individual amino acid changed. All amino acids substituted to generate derivative AMPs were of similar characteristics to the amino acids present in the parental AMPs, so as to maintain the predicted structures and increase the binding affinities. After which the physicochemical characteristics of each mutated AMP were determined using BACTIBASE and APD. Amino acid sequences of the derivative AMPs were then extracted and 3D structures predicted using I-TASSER, in which the derived AMPS display similar α-helical secondary structures as their parental counterparts AMPs with slight variation in partial α-helical structure. Subsequently, *in-silico* protein-peptide interaction of the mutated anti-bacterial and anti-viral pneumonia AMPs with their receptor proteins respectively was accomplished using PatchDock. Visual output of the *in-silico* binding studies was done using RasMol and analyses of the geometric scores of the binding affinity of each anti-pneumonia AMP with each protein receptor was done.

The binding affinity scores were compared to the parental AMPs and those with lower binding potential were discarded. So far, 7 AMPs were generated for anti-*Streptococcus pneumoniae*, 4 for anti-*Klebsiella pneumoniae*, 2 for *Acinetobacter baumannii*, 11 for anti-*Respiratory Syncytial* virus, 1 for anti-*Influenza B*, and 8 for anti-*Influenza A*. Summarily, 13 anti-bacterial and 20 anti-viral pneumonia AMPs were generated as additional AMPs with improved binding potential to the respective receptors.

7.2. Conclusion

The main goal of this biomarker discovery using AMPs is to ease the search for early detection as well as differentiate bacterial and viral pneumonia. This would assist medical practitioners towards the correct treatment plan and subsequently, enable patients to develop accommodating lifestyles. AMPs have shown great promise in circumventing the drawbacks associated with the current diagnostic systems. 45 putative AMPs were identified from the use of HMMER and CD-HIT for both bacterial and viral pathogens while 33 AMPs were further identified with improved activities against the pathogens for differential diagnosis. This research offers new insights into the *in silico* modular architecture, evolution of host defense molecules containing core motifs and diagnostic helices associated with antimicrobial activity of putative AMPs functions against bacterial and viral pneumonia. This research work could be pursued for molecular validation through the binding of these AMPs with the viral and bacterial proteins respectively, using an “on/off” binding experiment in a LFD setting with a view to developing a prototype with these specific AMPs conjugated to gold nanoparticles (AuNPs) to accurately and sensitively detect the viral and bacterial pathogens within patient samples.

7.3. Future Work

7.3.1. Peptides synthesis

The forty-five novel AMPs from HMMER using CD-HIT as comparative tool and thirty-three AMPs from SDM analysis will be chemically synthesized by GL Biochem Ltd. using the solid-phase method and purified to >98% by reverse-phase High-Pressure Liquid Chromatography which will be subsequently shipped in a lyophilized form.

7.3.2. Molecular study

The DNA sequences of the identified bacterial and viral pneumonia receptors (GenScript, USA) will be cloned into a vector and transformed into respective cells and purified using the GST purification system. All samples will be analyzed by a 12% SDS-PAGE. The purified pneumonia receptor proteins will be lyophilized and stored at -80° C for further use in binding study.

7.3.3. Construction of a lateral flow device

A lateral flow device will be manufactured with potential to detect bacterial and viral pneumonia differentially and tested using patients samples as described by Williams *et al.* (2016). The AMPs will then be conjugated to gold nanoparticles (AuNPs) in a LFD to determine pneumonia disease using patient sera within 15 minutes. This will provide patients with accurate and sensitive diagnosis of pneumonia disease at various stages of infection to assess the consistency and detection power of the LFD prototype.

7.4. Future perspectives

In silico analysis can be used, to a large extent, to generate and construct putative AMPs and their binding affinity to the bacterial and viral pneumonia receptors, an important factor in the creation of an efficient and universal diagnostic point (POC) of care method

to detect pneumonia at an early stage to prevent serious complications. This will subsequently implement the best-suited AMPs in a POC device for detection of pneumonia**.**

Moreover, this work could bring about fast and reliable diagnostic test to detect pneumonia and improve its epidemiological surveillance. This *in silico* approach would lead to the identification of novel AMPs with diagnostic relevance against pneumonia pathogens. Moreover, the algorithms may be used to prioritize biomarkers in other pathogen species implicated in pneumonia disease. This research work is fairly novel and much, has yet to be explored and thus allowing more research to be done by future scientists, with possible training in new bioinformatics pipelines and model creations that could be used for any other disease.

A cheap POC diagnostic kit generated for diagnosis of pneumonia would be easy to access by medical personnel as well as the patients and would not require expensive training for personnel to be able to interpret result. The rational design of peptides using *in silico* methodologies will improve diseases diagnosis, promote health equity among citizenry and increase efficiency of health personnel, with decreased production costs of the detection devices.

REFERENCES

Abraham T., Prenner E. J., Lewis R. N., Mant C. T., Keller S., Hodges R. S. and McElhaney R. N. 2014. Structure–activity relationships of the antimicrobial peptide gramicidin S and its analogs: aqueous solubility, self-association, conformation, antimicrobial activity and interaction with model lipid membranes. *Biochimica et Biophysica Acta (BBA)-Biomembranes,* 1838**,** 1420-1429.

Aggarwal C. C. and Zhai C. 2012. *Mining text data*, Springer Science & Business Media.

Aguiar S. I., Frias M. J., Santos L., Melo-Cristino J., Ramirez M. and Pathogens P. S. G. F. T. S. O. R. 2006. Emergence of optochin resistance among Streptococcus pneumoniae in Portugal. *Microbial Drug Resistance,* 12**,** 239-245.

Ahn I. E., Jerussi T., Farooqui M., Tian X., Wiestner A. and Gea-Banacloche J. 2016. Atypical Pneumocystis jirovecii pneumonia in previously untreated patients with CLL on single-agent ibrutinib. *Blood,* 128**,** 1940-1943.

Aken B. L., Achuthan P., Akanni W., Amode M. R., Bernsdorff F., Bhai J., Billis K., Carvalho-Silva D., Cummins C. and Clapham P. 2017. Ensembl 2017. *Nucleic acids research,* 45**,** D635-D642.

Alan M., Grolimund E., Kutz A., Christ-Crain M., Thomann R., Falconnier C., Hoess C., Henzen C., Zimmerli W. and Mueller B. 2015. Clinical risk scores and blood biomarkers as predictors of long-term outcome in patients with community-acquired pneumonia: a 6-year prospective follow-up study. *Journal of internal medicine,* 278**,** 174-184.

Almeida A. and Boattini M. 2017. Community-acquired pneumonia in HIV-positive patients: an update on etiologies, epidemiology and management. *Current infectious disease reports,* 19**,** 2.

Almirall J., Bolibar I., Toran P., Pera G., Boquet X., Balanzó X. and Sauca G. 2004. Contribution of C-reactive protein to the diagnosis and assessment of severity of community-acquired pneumonia. *chest,* 125**,** 1335-1342.

Alsubaie H., Al-Shamrani A., Alharbi A. S. and Alhaider S. 2015. Clinical practice guidelines: Approach to cough in children: The official statement endorsed by the Saudi Pediatric Pulmonology Association (SPPA). *International Journal of Pediatrics and Adolescent Medicine,* 2**,** 38-43.

Andersson-Gäre B. and Neuhauser D. 2007. The health care quality journey of Jönköping County Council, Sweden. *Quality Management in Healthcare,* 16**,** 2-9.

Andrade D. C., Borges I. C., Ivaska L., Peltola V., Meinke A., Barral A., Käyhty H., Ruuskanen O. and Nascimento-Carvalho C. M. 2016. Serological diagnosis of pneumococcal infection in children with pneumonia using protein antigens: a study of cut-offs with positive and negative controls. *Journal of immunological methods,* 433**,** 31-37.

Anliker B. and Chun J. Cell surface receptors in lysophospholipid signaling. Seminars in cell & developmental biology, 2004. Elsevier, 457-465.

Bader M. W., Sanowar S., Daley M. E., Schneider A. R., Cho U., Xu W., Klevit R. E., Le Moual H. and Miller S. I. 2005. Recognition of antimicrobial peptides by a bacterial sensor kinase. *Cell,* 122**,** 461-472.

Bae-Harboe Y.-S. C. and Park H.-Y. 2012. Tyrosinase: a central regulatory protein for cutaneous pigmentation. *Journal of Investigative Dermatology,* 132**,** 2678-2680.

Bagheri M., Beyermann M. and Dathe M. 2009. Immobilization reduces the activity of surface-bound cationic antimicrobial peptides with no influence upon the activity spectrum. *Antimicrobial agents and chemotherapy,* 53**,** 1132-1141.

Benjamin E. J., Blaha M. J., Chiuve S. E., Cushman M., Das S. R., Deo R., Floyd J., Fornage M., Gillespie C. and Isasi C. 2017. Heart disease and stroke statistics-2017 update: a report from the American Heart Association. *Circulation,* 135**,** e146-e603.

Berenstein R. 2015. Class III Receptor Tyrosine Kinases in Acute Leukemia–Biological Functions and Modern Laboratory Analysis. *Biomarker insights,* 10**,** BMI. S22433.

Bhadra P., Yan J., Li J., Fong S. and Siu S. W. 2018. AmPEP: Sequence-based prediction of antimicrobial peptides using distribution patterns of amino acid properties and random forest. *Scientific reports,* 8**,** 1697.

Bhutta Z. A., Das J. K., Walker N., Rizvi A., Campbell H., Rudan I. and Black R. E. 2013. Interventions to address deaths from childhood pneumonia and diarrhoea equitably: what works and at what cost? *The Lancet,* 381**,** 1417-1429.

Biro J. 2006. Amino acid size, charge, hydropathy indices and matrices for protein structure analysis. *Theoretical Biology and Medical Modelling,* 3**,** 15.

Blanco-Covián L., Montes-García V., Girard A., Fernández-Abedul M. T., Pérez-Juste J., Pastoriza-Santos I., Faulds K., Graham D. and Blanco-López M. C. 2017. Au@ Ag SERRS tags coupled to a lateral flow immunoassay for the sensitive detection of pneumolysin. *Nanoscale,* 9**,** 2051-2058.

Blanco J. C., Boukhvalova M. S., Shirey K. A., Prince G. A. and Vogel S. N. 2010. New insights for development of a safe and protective RSV vaccine. *Human vaccines,* 6**,** 482-492.

Blasius A. L. and Beutler B. 2010. Intracellular toll-like receptors. *Immunity,* 32**,** 305-315.

Bocquet N., Nury H., Baaden M., Le Poupon C., Changeux J.-P., Delarue M. and Corringer P.-J. 2009. X-ray structure of a pentameric ligand-gated ion channel in an apparently open conformation. *Nature,* 457**,** 111.

Bogaert D., de Groot R. and Hermans P. 2004. Streptococcus pneumoniae colonisation: the key to pneumococcal disease. *The Lancet infectious diseases,* 4**,** 144-154.

Boman H. G. 1995. Peptide antibiotics and their role in innate immunity. *Annual review of immunology,* 13**,** 61-92.

Bommarius B., Jenssen H., Elliott M., Kindrachuk J., Pasupuleti M., Gieren H., Jaeger K.-E., Hancock R. and Kalman D. 2010. Cost-effective expression and purification of antimicrobial and host defense peptides in Escherichia coli. *Peptides,* 31**,** 1957-1965.

Bouchon A., Dietrich J. and Colonna M. 2000. Cutting edge: inflammatory responses can be triggered by TREM-1, a novel receptor expressed on neutrophils and monocytes. *The Journal of Immunology,* 164**,** 4991-4995.

Bouvier N. M. and Palese P. 2008. The biology of influenza viruses. *Vaccine,* 26**,** D49-D53.

Boven L. A., Van Meurs M., Van Zwam M., Wierenga-Wolf A., Hintzen R. Q., Boot R. G., Aerts J. M., Amor S., Nieuwenhuis E. E. and Laman J. D. 2005. Myelin-laden macrophages are anti-inflammatory, consistent with foam cells in multiple sclerosis. *Brain,* 129**,** 517-526.

Brackett D. J., Lerner M. R., Lacquement M. A., He R. and Pereira H. A. 1997. A synthetic lipopolysaccharide-binding peptide based on the neutrophil-derived protein CAP37 prevents endotoxin-induced responses in conscious rats. *Infection and immunity,* 65**,** 2803-2811.

Brahmachary M., Krishnan S., Koh J. L. Y., Khan A. M., Seah S. H., Tan T. W., Brusic V. and Bajic V. B. 2004. ANTIMIC: a database of antimicrobial sequences. *Nucleic acids research,* 32**,** D586-D589.

Breukink E., Wiedemann I., Van Kraaij C., Kuipers O., Sahl H.-G. and De Kruijff B. 1999. Use of the cell wall precursor lipid II by a pore-forming peptide antibiotic. *Science,* 286**,** 2361-2364.

Brogden K. A. 2005. Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? *Nature reviews microbiology,* 3**,** 238.

Brumfitt W., Salton M. R. and Hamilton-Miller J. M. 2002. Nisin, alone and combined with peptidoglycan-modulating antibiotics: activity against methicillin-resistant Staphylococcus aureus and vancomycin-resistant enterococci. *Journal of Antimicrobial Chemotherapy,* 50**,** 731-734.

Camarena L., Bruno V., Euskirchen G., Poggio S. and Snyder M. 2010. Molecular mechanisms of ethanol-induced pathogenesis revealed by RNA-sequencing. *PLoS pathogens,* 6**,** e1000834.

Caruana R., Lou Y., Gehrke J., Koch P., Sturm M. and Elhadad N. Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2015. ACM, 1721-1730.

Chakraborty K., Raundhal M., Chen B. B., Morse C., Tyurina Y. Y., Khare A., Oriss T. B., Huff R., Lee J. S. and Croix C. M. S. 2017. The mito-DAMP cardiolipin blocks IL-10 production causing persistent inflammation during bacterial pneumonia. *Nature communications,* 8**,** 13944.

Chang K. Y. and Yang J.-R. 2013. Analysis and prediction of highly effective antiviral peptides based on random forests. *PloS one,* 8**,** e70166.

Chastre J., Luyt C.-E., Trouillet J.-L. and Combes A. 2006. New diagnostic and prognostic markers of ventilator-associated pneumonia. *Current opinion in critical care,* 12**,** 446-451.

Chen Y., Guarnieri M. T., Vasil A. I., Vasil M. L., Mant C. T. and Hodges R. S. 2007. Role of peptide hydrophobicity in the mechanism of action of α-helical antimicrobial peptides. *Antimicrobial agents and chemotherapy,* 51**,** 1398-1406.

Cherazard R., Epstein M., Doan T.-L., Salim T., Bharti S. and Smith M. A. 2017. Antimicrobial resistant Streptococcus pneumoniae: prevalence, mechanisms, and clinical implications. *American journal of therapeutics,* 24**,** e361-e369.

Cherezov V., Rosenbaum D. M., Hanson M. A., Rasmussen S. G., Thian F. S., Kobilka T. S., Choi H.-J., Kuhn P., Weis W. I. and Kobilka B. K. 2007. High-resolution crystal structure of an engineered human β2-adrenergic G protein–coupled receptor. *science,* 318**,** 1258-1265.

Chinnapaiyan S., Parira T., Dutta R., Agudelo M., Morris A., Nair M. and Unwalla H. 2017. HIV infects bronchial epithelium and suppresses components of the mucociliary clearance apparatus. *PloS one,* 12**,** e0169161.

Cho C. H., Woo M. K., Kim J. Y., Cheong S., Lee C.-K., An S. A., Lim C. S. and Kim W. J. 2013. Evaluation of five rapid diagnostic kits for influenza A/B virus. *Journal of virological methods,* 187**,** 51-56.

Choi C. H., Lee E. Y., Lee Y. C., Park T. I., Kim H. J., Hyun S. H., Kim S., Lee S. K. and Lee J. C. 2005. Outer membrane protein 38 of Acinetobacter baumannii localizes to the mitochondria and induces apoptosis of epithelial cells. *Cellular microbiology,* 7**,** 1127-1138.

Chopra S., Ramkissoon K. and Anderson D. 2013. A systematic quantitative proteomic examination of multidrug resistance in Acinetobacter baumannii. *Journal of proteomics,* 84**,** 17-39.

Christ-Crain M., Breidthardt T., Stolz D., Zobrist K., Bingisser R., Miedinger D., Leuppi J., Tamm M., Mueller B. and Mueller C. 2008. Use of B-type natriuretic peptide in the risk stratification of community-acquired pneumonia. *Journal of internal medicine,* 264**,** 166-176.

Christopoulos A. 2002. Allosteric binding sites on cell-surface receptors: novel targets for drug discovery. *Nature reviews Drug discovery,* 1**,** 198.

Cilloniz C., Martin-Loeches I., Garcia-Vidal C., San Jose A. and Torres A. 2016. Microbial etiology of pneumonia: Epidemiology, diagnosis and resistance patterns. *International journal of molecular sciences,* 17**,** 2120.

Coelho L., Póvoa P., Almeida E., Fernandes A., Mealha R., Moreira P. and Sabino H. 2007. Usefulness of C-reactive protein in monitoring the severe community-acquired pneumonia clinical course. *Critical care,* 11**,** R92.

Collingridge G. L., Olsen R. W., Peters J. and Spedding M. 2009. A nomenclature for ligand-gated ion channels. *Neuropharmacology,* 56**,** 2-5.

Collins P. L. and Graham B. S. 2008. Viral and host factors in human respiratory syncytial virus pathogenesis. *Journal of virology,* 82**,** 2040-2055.

Conlon J. M., Kolodziejek J. and Nowotny N. 2004. Antimicrobial peptides from ranid frogs: taxonomic and phylogenetic markers and a potential source of new therapeutic agents. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics,* 1696**,** 1-14.

Coordinators N. R. 2017. Database resources of the national center for biotechnology information. *Nucleic acids research,* 45**,** D12.

Cortés G., Borrell N., de Astorza B., Gómez C., Sauleda J. and Albertí S. 2002. Molecular analysis of the contribution of the capsular polysaccharide and the lipopolysaccharide O side chain to the virulence of Klebsiella pneumoniae in a murine model of pneumonia. *Infection and immunity,* 70**,** 2583-2590.

Costa F., Carvalho I. F., Montelaro R. C., Gomes P. and Martins M. C. L. 2011. Covalent immobilization of antimicrobial peptides (AMPs) onto biomaterial surfaces. *Acta biomaterialia,* 7**,** 1431-1440.

Curbelo J., Bueno S. L., Galván-Román J. M., Ortega-Gómez M., Rajas O., Fernández-Jiménez G., Vega-Piris L., Rodríguez-Salvanes F., Arnalich B. and Díaz A. 2017. Inflammation biomarkers in blood as mortality predictors in community-acquired pneumonia admitted patients: Importance of comparison with neutrophil count percentage or neutrophil-lymphocyte ratio. *PloS one,* 12**,** e0173947.

Cushnie T. T. and Lamb A. J. 2005. Detection of galangin-induced cytoplasmic membrane damage in Staphylococcus aureus by measuring potassium loss. *Journal of Ethnopharmacology,* 101**,** 243-248.

Cuthbert T. J., Hisey B., Harrison T. D., Trant J. F., Gillies E. R. and Ragogna P. J. 2018. Surprising antibacterial activity and selectivity of hydrophilic polyphosphoniums featuring sugar and hydroxy substituents. *Angewandte Chemie,* 130**,** 12889-12892.

Cutts F., Zaman S., Enwere G. y., Jaffar S., Levine O., Okoko J., Oluwalana C., Vaughan A., Obaro S. and Leach A. 2005. Efficacy of nine-valent pneumococcal conjugate vaccine against pneumonia and invasive pneumococcal disease in The Gambia: randomised, double-blind, placebo-controlled trial. *The Lancet,* 365**,** 1139-1146.

Darnell S. J., LeGault L. and Mitchell J. C. 2008. KFC Server: interactive forecasting of protein interaction hot spots. *Nucleic acids research,* 36**,** W265-W269.

Davi M. V., Cosaro E., Piacentini S., Reimondo G., Albiger N., Arnaldi G., Faggiano A., Mantovani G., Fazio N. and Piovesan A. 2017. Prognostic factors in ectopic Cushing's syndrome due to neuroendocrine tumors: a multicenter study. *European journal of endocrinology,* 176**,** 453-461.

Davis T. R., Evans H. R., Murtas J., Weisman A., Francis J. L. and Khan A. 2017. Utility of blood cultures in children admitted to hospital with community-acquired pneumonia. *Journal of paediatrics and child health,* 53**,** 232-236.

Decuyper I., Ebo D., Uyttebroek A., Hagendorens M., Faber M., Bridts C., De Clerck L. and Sabato V. 2016. Quantification of specific IgE antibodies in immediate drug hypersensitivity: More shortcomings than potentials? *Clinica Chimica Acta,* 460**,** 184-189.

Demaria M., O'Leary M. N., Chang J., Shao L., Liu S., Alimirah F., Koenig K., Le C., Mitin N. and Deal A. M. 2017. Cellular senescence promotes adverse effects of chemotherapy and cancer relapse. *Cancer discovery,* 7**,** 165-176.

Dhople V., Krukemeyer A. and Ramamoorthy A. 2006. The human beta-defensin-3, an antibacterial peptide with multiple biological functions. *Biochimica et Biophysica Acta (BBA)-Biomembranes,* 1758**,** 1499-1512.

Druzd D., Matveeva O., Ince L., Harrison U., He W., Schmal C., Herzel H., Tsang A. H., Kawakami N. and Leliavski A. 2017. Lymphocyte circadian clocks control lymph node trafficking and adaptive immune responses. *Immunity,* 46**,** 120-132.

Dutta S. R., Gauri S. S., Ghosh T., Halder S. K., DasMohapatra P. K., Mondal K. C. and Ghosh A. K. 2017. Elucidation of structural and functional integration of a novel antimicrobial peptide from Antheraea mylitta. *Bioorganic & medicinal chemistry letters,* 27**,** 1686-1692.

Eckert R., Qi F., Yarbrough D. K., He J., Anderson M. H. and Shi W. 2006. Adding selectivity to antimicrobial peptides: rational design of a multidomain peptide against Pseudomonas spp. *Antimicrobial agents and chemotherapy,* 50**,** 1480-1488.

Eddy S. R. 1998. Profile hidden Markov models. *Bioinformatics (Oxford, England),* 14**,** 755-763.

Edelheit O., Hanukoglu A. and Hanukoglu I. 2009. Simple and efficient site-directed mutagenesis using two single-primer reactions in parallel to generate mutants for protein structure-function studies. *BMC biotechnology,* 9**,** 61.

Eggerbauer E., de Benedictis P., Hoffmann B., Mettenleiter T. C., Schlottau K., Ngoepe E. C., Sabeta C. T., Freuling C. M. and Müller T. 2016. Evaluation of six commercially available rapid immunochromatographic tests for the diagnosis of rabies in brain material. *PLoS neglected tropical diseases,* 10.

Eierhoff T., Ludwig S. and Ehrhardt C. 2009. The influenza A virus matrix protein as a marker to monitor initial virus internalisation. *Biological chemistry,* 390**,** 509-515.

Falsey A. R., Formica M. A. and Walsh E. E. 2002. Diagnosis of respiratory syncytial virus infection: comparison of reverse transcription-PCR to viral culture and

serology in adults with respiratory illness. *Journal of clinical microbiology,* 40**,** 817-820.

Farnaaz N. and Jabbar M. 2016. Random forest modeling for network intrusion detection system. *Procedia Computer Science,* 89**,** 213-217.

Feldman S. A., Audet S. and Beeler J. A. 2000. The fusion glycoprotein of human respiratory syncytial virus facilitates virus attachment and infectivity via an interaction with cellular heparan sulfate. *Journal of virology,* 74**,** 6442-6447.

Fernandes F. C., Rigden D. J. and Franco O. L. 2012. Prediction of antimicrobial peptides based on the adaptive neuro-fuzzy inference system application. *Peptide Science,* 98**,** 280-287.

Fernández-Barat L., Ferrer M., De Rosa F., Gabarrús A., Esperatti M., Terraneo S., Rinaudo M., Bassi G. L. and Torres A. 2017. Intensive care unit-acquired pneumonia due to Pseudomonas aeruginosa with and without multidrug resistance. *Journal of Infection,* 74**,** 142-152.

Finn R. D., Attwood T. K., Babbitt P. C., Bateman A., Bork P., Bridge A. J., Chang H.-Y., Dosztányi Z., El-Gebali S. and Fraser M. 2017. InterPro in 2017—beyond protein family and domain annotations. *Nucleic acids research,* 45**,** D190-D199.

Finn R. D., Marshall M. and Bateman A. 2004. i Pfam: visualization of protein–protein interactions in PDB at domain and amino acid resolutions. *Bioinformatics,* 21**,** 410-412.

Fischer S., Gill V. J., Kovacs J., Miele P., Keary J., Silcott V., Huang S., Borio L., Stock F. and Fahle G. 2001. The use of oral washes to diagnose Pneumocystis carinii pneumonia: a blinded prospective study using a polymerase chain reaction–based detection system. *The Journal of infectious diseases,* 184**,** 1485-1488.

Fjell C. D., Hancock R. E. and Cherkasov A. 2007. AMPer: a database and an automated discovery tool for antimicrobial peptides. *Bioinformatics,* 23**,** 1148-1155.

Fjell C. D., Hiss J. A., Hancock R. E. and Schneider G. 2012. Designing antimicrobial peptides: form follows function. *Nature reviews Drug discovery,* 11**,** 37.

Fjell C. D., Jenssen H., Hilpert K., Cheung W. A., Pante N., Hancock R. E. and Cherkasov A. 2009. Identification of novel antibacterial peptides by chemoinformatics and machine learning. *Journal of medicinal chemistry,* 52**,** 2006-2015.

Flanders S. A., Stein J., Shochat G., Sellers K., Holland M., Maselli J., Drew W. L., Reingold A. L. and Gonzales R. 2004. Performance of a bedside C-reactive protein test in the diagnosis of community-acquired pneumonia in adults with acute cough. *The American journal of medicine,* 116**,** 529-535.

Franzin L. and Cabodi D. 2005. Legionella pneumonia and serum procalcitonin. *Current microbiology,* 50**,** 43-46.

Fries L., Shinde V., Stoddard J. J., Thomas D. N., Kpamegan E., Lu H., Smith G., Hickman S. P., Piedra P. and Glenn G. M. 2017. Immunogenicity and safety of a respiratory syncytial virus fusion protein (RSV F) nanoparticle vaccine in older adults. *Immunity & Ageing,* 14**,** 8.

Fu E., Lutz B., Kauffman P. and Yager P. 2010. Controlled reagent transport in disposable 2D paper networks. *Lab on a Chip,* 10**,** 918-920.

Gan S. W., Ng L., Lin X., Gong X. and Torres J. 2008. Structure and ion channel activity of the human respiratory syncytial virus (hRSV) small hydrophobic protein transmembrane domain. *Protein Science,* 17**,** 813-820.

Ganz T. 2003. Defensins: antimicrobial peptides of innate immunity. *Nature reviews immunology,* 3**,** 710.

Garg N., Singh R., Shukla G., Capalash N. and Sharma P. 2016. Immunoprotective potential of in silico predicted Acinetobacter baumannii outer membrane nuclease, NucAb. *International Journal of Medical Microbiology,* 306**,** 1-9.

Gennaro R. and Zanetti M. 2000. Structural features and biological activities of the cathelicidin-derived antimicrobial peptides. *Peptide Science,* 55**,** 31-49.

Gill S. C. and Von Hippel P. H. 1989. Calculation of protein extinction coefficients from amino acid sequence data. *Analytical biochemistry,* 182**,** 319-326.

Ginsburg A. S., Delarosa J., Brunette W., Levari S., Sundt M., Larson C., Agyemang C. T., Newton S., Borriello G. and Anderson R. 2015. mPneumonia: development of an innovative mHealth application for diagnosing and treating childhood pneumonia and other childhood illnesses in low-resource settings. *PloS one,* 10.

Goel V. K. and Kapil A. 2001. Monoclonal antibodies against the iron regulated outer membrane proteins of Acinetobacter baumannii are bactericidal. *BMC microbiology,* 1**,** 16.

Goh K., Honga G. P., Chyi N. H., Piawa C. and Rahman R. 2012. Trends and tips in protein engineering, a review. *Jurnal Teknologi (Sci. Eng.),* 59**,** 21-31.

Gregorich Z. R. and Ge Y. 2014. Top-down proteomics in health and disease: Challenges and opportunities. *Proteomics,* 14**,** 1195-1210.

Guharoy M., Pal A., Dasgupta M. and Chakrabarti P. 2011. PRICE (PRotein Interface Conservation and Energetics): a server for the analysis of protein–protein interfaces. *Journal of structural and functional genomics,* 12**,** 33-41.

Haeuptle J., Zaborsky R., Fiumefreddo R., Trampuz A., Steffen I., Frei R., Christ-Crain M., Müller B. and Schuetz P. 2009. Prognostic value of procalcitonin in Legionella pneumonia. *European journal of clinical microbiology & infectious diseases,* 28**,** 55.

Hajisharifi Z., Piryaiee M., Beigi M. M., Behbahani M. and Mohabatkar H. 2014. Predicting anticancer peptides with Chou′ s pseudo amino acid composition and investigating their mutagenicity via Ames test. *Journal of Theoretical Biology,* 341**,** 34-40.

Hallak L. K., Spillmann D., Collins P. L. and Peeples M. E. 2000. Glycosaminoglycan sulfation requirements for respiratory syncytial virus infection. *Journal of virology,* 74**,** 10508-10513.

Hallock K. J., Lee D.-K. and Ramamoorthy A. 2003. MSI-78, an analogue of the magainin antimicrobial peptides, disrupts lipid bilayer structure via positive curvature strain. *Biophysical Journal,* 84**,** 3052-3060.

Hammami R. and Fliss I. 2010. Current trends in antimicrobial agent research: chemo- and bioinformatics approaches. *Drug discovery today,* 15**,** 540-546.

Hammami R., Zouhir A., Hamida J. B. and Fliss I. 2007. BACTIBASE: a new web-accessible database for bacteriocin characterization. *Bmc Microbiology,* 7**,** 89.

Hammami R., Zouhir A., Le Lay C., Hamida J. B. and Fliss I. 2010. BACTIBASE second release: a database and tool platform for bacteriocin characterization. *Bmc Microbiology,* 10**,** 22.

Hammer T. J., Janzen D. H., Hallwachs W., Jaffe S. P. and Fierer N. 2017. Caterpillars lack a resident gut microbiome. *Proceedings of the National Academy of Sciences,* 114**,** 9641-9646.

Hancock R. E. and Chapple D. S. 1999. Peptide antibiotics. *Antimicrobial agents and chemotherapy,* 43**,** 1317-1323.

Hancock R. E. and Rozek A. 2002. Role of membranes in the activities of antimicrobial cationic peptides. *FEMS microbiology letters,* 206**,** 143-149.

Hancock R. E. and Scott M. G. 2000. The role of antimicrobial peptides in animal defenses. *Proceedings of the national Academy of Sciences,* 97**,** 8856-8861.

Hartmann M., Berditsch M., Hawecker J., Ardakani M. F., Gerthsen D. and Ulrich A. S. 2010. Damage of the bacterial cell envelope by antimicrobial peptides gramicidin S and PGLa as revealed by transmission and scanning electron microscopy. *Antimicrobial agents and chemotherapy,* 54**,** 3132-3142.

Hein K. Z., Takahashi H., Tsumori T., Yasui Y., Nanjoh Y., Toga T., Wu Z., Grötzinger J., Jung S. and Wehkamp J. 2015. Disulphide-reduced psoriasin is a human apoptosis-inducing broad-spectrum fungicide. *Proceedings of the National Academy of Sciences,* 112**,** 13039-13044.

Henriques-Normark B. and Tuomanen E. I. 2013. The pneumococcus: epidemiology, microbiology, and pathogenesis. *Cold Spring Harbor perspectives in medicine,* 3**,** a010215.

Hirche T. O., Gaut J. P., Heinecke J. W. and Belaaouaj A. 2005. Myeloperoxidase plays critical roles in killing Klebsiella pneumoniae and inactivating neutrophil elastase: effects on host defense. *The Journal of Immunology,* 174**,** 1557-1565.

Hirst D. J., Lee T.-H., Swann M. J., Unabia S., Park Y., Hahm K.-S. and Aguilar M. I. 2011. Effect of acyl chain structure and bilayer phase state on binding and penetration of a supported lipid bilayer by HPA3. *European Biophysics Journal,* 40**,** 503-514.

Hoffmann J., Machado D., Terrier O., Pouzol S., Messaoudi M., Basualdo W., Espínola E. E., Guillen R. M., Rosa-Calatrava M. and Picot V. 2016. Viral and bacterial co-infection in severe pneumonia triggers innate immune responses and specifically enhances IP-10: a translational study. *Scientific reports,* 6**,** 38532.

Hollmann A., Martinez M., Maturana P., Semorile L. C. and Maffia P. C. 2018. Antimicrobial peptides: interaction with model and biological membranes and synergism with chemical antibiotics. *Frontiers in chemistry,* 6.

Hong Y. C., Choi E. J. and Park S. 2017. Risk Factors of readmission to hospital for pneumonia in children. *Pediatric Infection & Vaccine,* 24**,** 146-151.

Howard A., O'Donoghue M., Feeney A. and Sleator R. D. 2012. Acinetobacter baumannii: an emerging opportunistic pathogen. *Virulence,* 3**,** 243-250.

Hsu C.-H., Chen C., Jou M.-L., Lee A. Y.-L., Lin Y.-C., Yu Y.-P., Huang W.-T. and Wu S.-H. 2005. Structural and DNA-binding studies on the bovine antimicrobial peptide, indolicidin: evidence for multiple conformations involved in binding to membranes and DNA. *Nucleic acids research,* 33**,** 4053-4064.

Huang L. and Crothers K. 2009. HIV-associated opportunistic pneumonias. *Respirology,* 14**,** 474-485.

Iorio M. V. and Croce C. M. 2012. MicroRNA dysregulation in cancer: diagnostics, monitoring and therapeutics. A comprehensive review. *EMBO molecular medicine,* 4**,** 143-159.

Islam A. H. M. S., Singh K.-K. B. and Ismail A. 2011. Demonstration of an outer membrane protein that is antigenically specific for Acinetobacter baumannii. *Diagnostic microbiology and infectious disease,* 69**,** 38-44.

Ito A., Ishida T., Tokumasu H., Washio Y., Yamazaki A., Ito Y. and Tachibana H. 2017. Prognostic factors in hospitalized community-acquired pneumonia: a retrospective study of a prospective observational cohort. *BMC pulmonary medicine,* 17**,** 78.

Izoré T., Contreras-Martel C., El Mortaji L., Manzano C., Terrasse R., Vernet T., Di Guilmi A. M. and Dessen A. 2010. Structural basis of host cell recognition by the pilus adhesin from Streptococcus pneumoniae. *Structure,* 18**,** 106-115.

Jacobs A. C., Hood I., Boyd K. L., Olson P. D., Morrison J. M., Carson S., Sayood K., Iwen P. C., Skaar E. P. and Dunman P. M. 2010. Inactivation of phospholipase D diminishes Acinetobacter baumannii pathogenesis. *Infection and immunity,* 78**,** 1952-1962.

Jain S., Self W. H., Wunderink R. G., Fakhran S., Balk R., Bramley A. M., Reed C., Grijalva C. G., Anderson E. J. and Courtney D. M. 2015. Community-acquired pneumonia requiring hospitalization among US adults. *New England Journal of Medicine,* 373**,** 415-427.

Järvå M., Lay F. T., Phan T. K., Humble C., Poon I. K., Bleackley M. R., Anderson M. A., Hulett M. D. and Kvansakul M. 2018. X-ray structure of a carpet-like antimicrobial defensin–phospholipid membrane disruption complex. *Nature communications,* 9**,** 1962.

Jodoin J. and Hincke M. T. 2018. Histone H5 is a potent Antimicrobial Agent and a template for novel Antimicrobial Peptides. *Scientific reports,* 8**,** 2411.

Johnson S. B., Lissauer M., Bochicchio G. V., Moore R., Cross A. S. and Scalea T. M. 2007. Gene expression profiles differentiate between sterile SIRS and early sepsis. *Annals of surgery,* 245**,** 611.

Jorgensen J. H., Barry A. L., Traczewski M., Sahm D. F., McElmeel M. L. and Crawford S. A. 2000. Rapid automated antimicrobial susceptibility testing of Streptococcus pneumoniae by use of the bioMerieux VITEK 2. *Journal of clinical microbiology,* 38**,** 2814-2818.

Joyce A. P., Zhang C., Bradley P. and Havranek J. J. 2014. Structure-based modeling of protein: DNA specificity. *Briefings in functional genomics,* 14**,** 39-49.

Julián-Jiménez A., del Castillo J. G. and Candel F. J. 2017. Usefulness and prognostic value of biomarkers in patients with community-acquired pneumonia in the emergency department. *Medicina Clínica (English Edition),* 148**,** 501-510.

Jung D.-Y., Park J. B., Lee E.-N., Lee H.-A., Joh J.-W., Kwon C. H., Ki C.-S., Lee S.-Y. and Kim S.-J. 2008. Combined use of myeloid-related protein 8/14 and procalcitonin as diagnostic markers for acute allograft rejection in kidney transplantation recipients. *Transplant immunology,* 18**,** 338-343.

Jung J., Michalak M. and Agellon L. B. 2017. Endoplasmic reticulum malfunction in the nervous system. *Frontiers in neuroscience,* 11**,** 220.

Juretić D., Vukičević D., Petrov D., Novković M., Bojović V., Lučić B., Ilić N. and Tossi A. 2011. Knowledge-based computational methods for identifying or designing novel, non-homologous antimicrobial peptides. *European biophysics journal,* 40**,** 371-385.

Kadioglu A., Weiser J. N., Paton J. C. and Andrew P. W. 2008. The role of Streptococcus pneumoniae virulence factors in host respiratory colonization and disease. *Nature Reviews Microbiology,* 6**,** 288.

Kalil A. C., Metersky M. L., Klompas M., Muscedere J., Sweeney D. A., Palmer L. B., Napolitano L. M., O'grady N. P., Bartlett J. G. and Carratalà J. 2016. Management

of adults with hospital-acquired and ventilator-associated pneumonia: 2016 clinical practice guidelines by the Infectious Diseases Society of America and the American Thoracic Society. *Clinical Infectious Diseases,* 63**,** e61-e111.

Kanneganti T.-D., Lamkanfi M. and Núñez G. 2007. Intracellular NOD-like receptors in host defense and disease. *Immunity,* 27**,** 549-559.

Kapoor G., Saigal S. and Elongavan A. 2017. Action and resistance mechanisms of antibiotics: A guide for clinicians. *Journal of anaesthesiology, clinical pharmacology,* 33**,** 300.

Kerlin B. A., Yan S. B., Isermann B. H., Brandt J. T., Sood R., Basson B. R., Joyce D. E., Weiler H. and Dhainaut J.-F. 2003. Survival advantage associated with heterozygous factor V Leiden mutation in patients with severe sepsis and in mouse endotoxemia. *Blood,* 102**,** 3085-3092.

Khor C. C., Chapman S. J., Vannberg F. O., Dunne A., Murphy C., Ling E. Y., Frodsham A. J., Walley A. J., Kyrieleis O. and Khan A. 2007. A Mal functional variant is associated with protection against invasive pneumococcal disease, bacteremia, malaria and tuberculosis. *Nature genetics,* 39**,** 523.

Kim I.-W., Lee J.-H., Park H.-Y., Kwon Y.-N., Yun E.-Y., Nam S.-H., Kim S.-R., Ahn M.-Y. and Hwang J. S. 2012. Characterization and cDNA cloning of a defensin-like peptide, harmoniasin, from Harmonia axyridis. *J Microbiol Biotechnol,* 22**,** 1588-90.

Kim M., Lee S. M., Song J.-W., Do K.-H., Lee H. J., Lim S., Choe J., Park K. J., Park H. J. and Kim H. J. 2017. Added value of prone CT in the assessment of honeycombing and classification of usual interstitial pneumonia pattern. *European journal of radiology,* 91**,** 66-70.

Kojima Y., Okudela K., Matsumura M., Omori T., Baba T., Sekine A., Woo T., Umeda S., Takemura T. and Mitsui H. 2017. The pathological features of idiopathic interstitial pneumonia-associated pulmonary adenocarcinomas. *Histopathology,* 70**,** 568-578.

Kolokoltsov A. A., Deniger D., Fleming E. H., Roberts N. J., Karpilow J. M. and Davey R. A. 2007. Small interfering RNA profiling reveals key role of clathrin-mediated endocytosis and early endosome formation for infection by respiratory syncytial virus. *Journal of virology,* 81**,** 7786-7800.

Kriegsmann M., Casadonte R., Kriegsmann J., Dienemann H., Schirmacher P., Kobarg J. H., Schwamborn K., Stenzinger A., Warth A. and Weichert W. 2016. Reliable entity subtyping in non-small cell lung cancer by matrix-assisted laser desorption/ionization imaging mass spectrometry on formalin-fixed paraffin-embedded tissue specimens. *Molecular & Cellular Proteomics,* 15**,** 3081-3089.

Krüger S., Papassotiriou J., Marre R., Richter K., Schumann C., von Baum H., Morgenthaler N. G., Suttorp N., Welte T. and Group C. S. 2007. Pro-atrial natriuretic peptide and pro-vasopressin to predict severity and prognosis in community-acquired pneumonia. *Intensive care medicine,* 33**,** 2069-2078.

Ku C.-L., Von Bernuth H., Picard C., Zhang S.-Y., Chang H.-H., Yang K., Chrabieh M., Issekutz A. C., Cunningham C. K. and Gallin J. 2007. Selective predisposition to bacterial infections in IRAK-4–deficient children: IRAK-4–dependent TLRs are otherwise redundant in protective immunity. *Journal of Experimental Medicine,* 204**,** 2407-2422.

Kurt-Jones E. A., Popova L., Kwinn L., Haynes L. M., Jones L. P., Tripp R. A., Walsh E. E., Freeman M. W., Golenbock D. T. and Anderson L. J. 2000. Pattern

recognition receptors TLR4 and CD14 mediate response to respiratory syncytial virus. *Nature immunology,* 1**,** 398.
Lai Y. and Gallo R. L. 2009. AMPed up immunity: how antimicrobial peptides have multiple roles in immune defense. *Trends in immunology,* 30**,** 131-141.
Larrick J. W., Hirata M., Balint R. F., Lee J., Zhong J. and Wright S. C. 1995. Human CAP18: a novel antimicrobial lipopolysaccharide-binding protein. *Infection and immunity,* 63**,** 1291-1297.
Lata S., Sharma B. and Raghava G. 2007. Analysis and prediction of antibacterial peptides. *BMC bioinformatics,* 8**,** 263.
Lázár V., Martins A., Spohn R., Daruka L., Grézal G., Fekete G., Számel M., Jangir P. K., Kintses B. and Csörgő B. 2018. Antibiotic-resistant bacteria show widespread collateral sensitivity to antimicrobial peptides. *Nature microbiology,* 3**,** 718-731.
Lee C.-R., Lee J. H., Park M., Park K. S., Bae I. K., Kim Y. B., Cha C.-J., Jeong B. C. and Lee S. H. 2017. Biology of Acinetobacter baumannii: pathogenesis, antibiotic resistance mechanisms, and prospective treatment options. *Frontiers in cellular and infection microbiology,* 7**,** 55.
Lee K.-Y. 2017. Pneumonia, acute respiratory distress syndrome, and early immune-modulator therapy. *International journal of molecular sciences,* 18**,** 388.
Lee T.-H., Heng C., Swann M. J., Gehman J. D., Separovic F. and Aguilar M.-I. 2010. Real-time quantitative analysis of lipid disordering by aurein 1.2 during membrane adsorption, destabilisation and lysis. *Biochimica et Biophysica Acta (BBA)-Biomembranes,* 1798**,** 1977-1986.
Leinonen R., Diez F. G., Binns D., Fleischmann W., Lopez R. and Apweiler R. 2004. UniProt archive. *Bioinformatics,* 20**,** 3236-3237.
LeVine A. M. and Whitsett J. A. 2001. Pulmonary collectins and innate host defense of the lung. *Microbes and infection,* 3**,** 161-166.
Li B., Han S., Liu F., Kang L. and Xv C. 2017a. Budesonide nebulization in the treatment of neonatal ventilator associated pneumonia. *Pakistan journal of medical sciences,* 33**,** 997.
Li J., Koh J.-J., Liu S., Lakshminarayanan R., Verma C. S. and Beuerman R. W. 2017b. Membrane active antimicrobial peptides: translating mechanistic insights to design. *Frontiers in neuroscience,* 11**,** 73.
Li S., Wong A. H. and Liu F. 2014. Ligand-gated ion channel interacting proteins and their role in neuroprotection. *Frontiers in cellular neuroscience,* 8**,** 125.
Li W. 2015. Fast program for clustering and comparing large sets of protein or nucleotide sequences. *Encyclopedia of Metagenomics.* Springer.
Li W. and Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics,* 22**,** 1658-1659.
Li W., Jaroszewski L. and Godzik A. 2001. Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics,* 17**,** 282-283.
Li W., Jaroszewski L. and Godzik A. 2002. Sequence clustering strategies improve remote homology recognitions while reducing search times. *Protein engineering,* 15**,** 643-649.
Liu H., Grantham M. L. and Pekosz A. 2018. Mutations in the Influenza A virus M1 protein enhance virus budding to complement lethal mutations in the M2 cytoplasmic tail. *Journal of virology,* 92**,** e00858-17.
Liu L., Oza S., Hogan D., Perin J., Rudan I., Lawn J. E., Cousens S., Mathers C. and Black R. E. 2015. Global, regional, and national causes of child mortality in 2000–

13, with projections to inform post-2015 priorities: an updated systematic analysis. *The Lancet,* 385**,** 430-440.

Liu S., Fan L., Sun J., Lao X. and Zheng H. 2017a. Computational resources and tools for antimicrobial peptides. *Journal of Peptide Science,* 23**,** 4-12.

Liu X., Gallay C., Kjos M., Domenech A., Slager J., van Kessel S. P., Knoops K., Sorg R. A., Zhang J. R. and Veening J. W. 2017b. High-throughput CRISPRi phenotyping identifies new essential genes in Streptococcus pneumoniae. *Molecular systems biology,* 13.

Liu Y., Su L., Jiang S.-J. and Qu H. 2017c. Risk factors for mortality from pneumocystis carinii pneumonia (PCP) in non-HIV patients: a meta-analysis. *Oncotarget,* 8**,** 59729.

Lo M. S., Brazas R. M. and Holtzman M. J. 2005. Respiratory syncytial virus nonstructural proteins NS1 and NS2 mediate inhibition of Stat2 expression and alpha/beta interferon responsiveness. *Journal of virology,* 79**,** 9315-9319.

Lockhart D. J. and Winzeler E. A. 2000. Genomics, gene expression and DNA arrays. *Nature,* 405**,** 827.

Loeffler J. M., Nelson D. and Fischetti V. A. 2001. Rapid killing of Streptococcus pneumoniae with a bacteriophage cell wall hydrolase. *Science,* 294**,** 2170-2172.

Loppnow H., Libby P., Freudenberg M., Krauss J., Weckesser J. and Mayer H. 1990. Cytokine induction by lipopolysaccharide (LPS) corresponds to lethal toxicity and is inhibited by nontoxic Rhodobacter capsulatus LPS. *Infection and immunity,* 58**,** 3743-3750.

Łuksza M., Riaz N., Makarov V., Balachandran V. P., Hellmann M. D., Solovyov A., Rizvi N. A., Merghoub T., Levine A. J. and Chan T. A. 2017. A neoantigen fitness model predicts tumour response to checkpoint blockade immunotherapy. *Nature,* 551**,** 517-520.

Madera M. and Gough J. 2002. A comparison of profile hidden Markov model procedures for remote homology detection. *Nucleic acids research,* 30**,** 4321-4328.

Malhotra-Kumar S., Abrahantes J. C., Sabiiti W., Lammens C., Vercauteren G., Ieven M., Molenberghs G., Aerts M. and Goossens H. 2010. Evaluation of chromogenic media for detection of methicillin-resistant Staphylococcus aureus. *Journal of clinical microbiology,* 48**,** 1040-1046.

Manavalan B., Subramaniyam S., Shin T. H., Kim M. O. and Lee G. 2018. Machine-learning-based prediction of cell-penetrating peptides and their uptake efficiency with improved accuracy. *Journal of proteome research,* 17**,** 2715-2726.

Marik P. E. 2010. Aspiration syndromes: aspiration pneumonia and pneumonitis. *Hospital Practice,* 38**,** 35-42.

Marik P. E. and Kaplan D. 2003. Aspiration pneumonia and dysphagia in the elderly. *Chest,* 124**,** 328-336.

Martin P. and Leibovich S. J. 2005. Inflammatory cells during wound repair: the good, the bad and the ugly. *Trends in cell biology,* 15**,** 599-607.

Matsuzaki K. 2009. Control of cell selectivity of antimicrobial peptides. *Biochimica et Biophysica Acta (BBA)-Biomembranes,* 1788**,** 1687-1692.

McManus A. M., Dawson N. F., Wade J. D., Carrington L. E., Winzor D. J. and Craik D. J. 2000. Three-dimensional structure of RK-1: a novel α-defensin peptide. *Biochemistry,* 39**,** 15757-15764.

Meher P. K., Sahu T. K., Saini V. and Rao A. R. 2017. Predicting antimicrobial peptides with improved accuracy by incorporating the compositional, physico-chemical and structural features into Chou's general PseAAC. *Scientific reports,* 7**,** 1-12.

Mehl M. 2018. Laryngeal Paralysis. *Textbook of Small Animal Emergency Medicine***,** 193-195.

Melbye H. and Stocks N. 2006. Point of care testing for C-reactive protein: a new path for Australian GPs? *Australian family physician,* 35**,** 513.

Menendez R., Torres A., Zalacain R., Aspa J., Villasclaras J. M., Borderías L., Moya J. B., Ruiz-Manzano J., de Castro F. R. and Blanquer J. 2004. Risk factors of treatment failure in community acquired pneumonia: implications for disease outcome. *Thorax,* 59**,** 960-965.

Messiaen P. E., Cuyx S., Dejagere T. and van der Hilst J. C. 2017. The role of CD 4 cell count as discriminatory measure to guide chemoprophylaxis against P neumocystis jirovecii pneumonia in human immunodeficiency virus-negative immunocompromised patients: A systematic review. *Transplant Infectious Disease,* 19**,** e12651.

Mitchell A. and Mitchell T. 2010. Streptococcus pneumoniae: virulence factors and variation. *Clinical Microbiology and Infection,* 16**,** 411-418.

Mok J. H., Eom J. S., Jo E. J., Kim M. H., Lee K., Kim K. U., Park H. K., Yi J. and Lee M. K. 2016. Clinical utility of rapid pathogen identification using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry in ventilated patients with pneumonia: A pilot study. *Respirology,* 21**,** 321-328.

Momosaki R., Yasunaga H., Matsui H., Horiguchi H., Fushimi K. and Abo M. 2016. Predictive factors for oral intake after aspiration pneumonia in older adults. *Geriatrics & gerontology international,* 16**,** 556-560.

Morgenthaler N. G., Struck J., Alonso C. and Bergmann A. 2006. Assay for the measurement of copeptin, a stable peptide derived from the precursor of vasopressin. *Clinical chemistry,* 52**,** 112-119.

Mulvenna J. P., Wang C. and Craik D. J. 2006. CyBase: a database of cyclic protein sequence and structure. *Nucleic acids research,* 34**,** D192-D194.

Munoz-Price L. S., Poirel L., Bonomo R. A., Schwaber M. J., Daikos G. L., Cormican M., Cornaglia G., Garau J., Gniadkowski M. and Hayden M. K. 2013. Clinical epidemiology of the global expansion of Klebsiella pneumoniae carbapenemases. *The Lancet infectious diseases,* 13**,** 785-796.

Murdoch D. R., O'Brien K. L., Scott J. A. G., Karron R. A., Bhat N., Driscoll A. J., Knoll M. D. and Levine O. S. 2009. Breathing new life into pneumonia diagnostics. *Journal of clinical microbiology,* 47**,** 3405-3408.

Nerenberg K. A., Zarnke K. B., Leung A. A., Dasgupta K., Butalia S., McBrien K., Harris K. C., Nakhla M., Cloutier L. and Gelfer M. 2018. Hypertension Canada's 2018 guidelines for diagnosis, risk assessment, prevention, and treatment of hypertension in adults and children. *Canadian Journal of Cardiology,* 34**,** 506-525.

Ngari C. G., Malonza D. M. and Muthuri G. G. 2014. A model for childhood Pneumonia Dynamics.

Nguyen L. T., Haney E. F. and Vogel H. J. 2011. The expanding scope of antimicrobial peptide structures and their modes of action. *Trends in biotechnology,* 29**,** 464-472.

Nguyen T., Tran T., Roberts C., Fox G., Graham S. and Marais B. 2017. Risk factors for child pneumonia-focus on the Western Pacific Region. *Paediatric respiratory reviews,* 21**,** 95-101.

Nickler M., Ottiger M., Steuer C., Kutz A., Christ-Crain M., Zimmerli W., Thomann R., Hoess C., Henzen C. and Bernasconi L. 2017. Time-dependent association of glucocorticoids with adverse outcome in community-acquired pneumonia: a 6-year prospective cohort study. *Critical Care,* 21**,** 72.

Niyonsaba F., Iwabuchi K., Matsuda H., Ogawa H. and Nagaoka I. 2002. Epithelial cell-derived human β-defensin-2 acts as a chemotaxin for mast cells through a pertussis toxin-sensitive and phospholipase C-dependent pathway. *International immunology,* 14**,** 421-426.

O'Brien K. L., Wolfson L. J., Watt J. P., Henkle E., Deloria-Knoll M., McCall N., Lee E., Mulholland K., Levine O. S. and Cherian T. 2009. Burden of disease caused by Streptococcus pneumoniae in children younger than 5 years: global estimates. *The Lancet,* 374**,** 893-902.

O'Driscoll N. H., Labovitiadi O., Cushnie T. T., Matthews K. H., Mercer D. K. and Lamb A. J. 2013. Production and evaluation of an antimicrobial peptide-containing wafer formulation for topical application. *Current microbiology,* 66**,** 271-278.

Opal S. 2003. The link between coagulation and innate immunity in severe sepsis. *Scand J Infect Dis,* 35**,** 535-544.

Oppenheim J., Biragyn A., Kwak L. and Yang D. 2003. Roles of antimicrobial peptides such as defensins in innate and adaptive immunity. *Annals of the rheumatic diseases,* 62**,** ii17-ii21.

Otvos L., O I., Rogers M. E., Consolvo P. J., Condie B. A., Lovas S., Bulet P. and Blaszczyk-Thurin M. 2000. Interaction between heat shock proteins and antimicrobial peptides. *Biochemistry,* 39**,** 14150-14159.

Paczosa M. K. and Mecsas J. 2016. Klebsiella pneumoniae: going on the offense with a strong defense. *Microbiol. Mol. Biol. Rev.,* 80**,** 629-661.

Palacios G., Hornig M., Cisterna D., Savji N., Bussetti A. V., Kapoor V., Hui J., Tokarz R., Briese T. and Baumeister E. 2009. Streptococcus pneumoniae coinfection is correlated with the severity of H1N1 pandemic influenza. *PloS one,* 4**,** e8540.

Papo N., Oren Z., Pag U., Sahl H.-G. and Shai Y. 2002. The consequence of sequence alteration of an amphipathic α-helical antimicrobial peptide and its diastereomers. *Journal of Biological Chemistry,* 277**,** 33913-33921.

Park C. B., Kim H. S. and Kim S. C. 1998. Mechanism of action of the antimicrobial peptide buforin II: buforin II kills microorganisms by penetrating the cell membrane and inhibiting cellular functions. *Biochemical and biophysical research communications,* 244**,** 253-257.

Phua J., Koay E., Zhang D., Tai L., Boo X., Lim K. and Lim T. 2006. Soluble triggering receptor expressed on myeloid cells-1 in acute respiratory infections. *European Respiratory Journal,* 28**,** 695-702.

Piers K. L., Brown M. H. and Hancock R. E. 1993. Recombinant DNA procedures for producing small antimicrobial cationic peptides in bacteria. *Gene,* 134**,** 7-13.

Polanco C., Samaniego J. L., Castañón-González J. A. and Buhse T. 2014. Polar profile of antiviral peptides from AVPpred Database. *Cell biochemistry and biophysics,* 70**,** 1469-1477.

Poon I. K., Baxter A. A., Lay F. T., Mills G. D., Adda C. G., Payne J. A., Phan T. K., Ryan G. F., White J. A. and Veneer P. K. 2014. Phosphoinositide-mediated oligomerization of a defensin induces cell lysis. *Elife,* 3**,** e01808.

Porto W., Pires A. and Franco O. 2017. Computational tools for exploring sequence databases as a resource for antimicrobial peptides. *Biotechnology Advances,* 35**,** 337-349.

Porto W. F., Silva O. N. and Franco O. L. 2012. Prediction and rational design of antimicrobial peptides. *Protein Structure, InTech***,** 377-396.

Potter S. C., Luciani A., Eddy S. R., Park Y., Lopez R. and Finn R. D. 2018. HMMER web server: 2018 update. *Nucleic acids research,* 46**,** W200-W204.

Prendergast C. and Papenburg J. 2013. Rapid antigen-based testing for respiratory syncytial virus: moving diagnostics from bench to bedside? *Future microbiology,* 8**,** 435-444.

Qureshi A., Thakur N., Tandon H. and Kumar M. 2013. AVPdb: a database of experimentally validated antiviral peptides targeting medically important viruses. *Nucleic acids research,* 42**,** D1147-D1153.

Rajpurkar P., Irvin J., Zhu K., Yang B., Mehta H., Duan T., Ding D., Bagul A., Langlotz C. and Shpanskaya K. 2017. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225*.

Ramamoorthy A., Thennarasu S., Tan A., Gottipati K., Sreekumar S., Heyl D. L., An F. Y. and Shelburne C. E. 2006. Deletion of all cysteines in tachyplesin I abolishes hemolytic activity and retains antimicrobial activity and lipopolysaccharide selective binding. *Biochemistry,* 45**,** 6529-6540.

Ramirez J. A., Wiemken T. L., Peyrani P., Arnold F. W., Kelley R., Mattingly W. A., Nakamatsu R., Pena S., Guinn B. E. and Furmanek S. P. 2017. Adults hospitalized with pneumonia in the United States: incidence, epidemiology, and mortality. *Clinical Infectious Diseases,* 65**,** 1806-1812.

Randou E. G., Veltri D. and Shehu A. Binary response models for recognition of antimicrobial peptides. Proceedings of the International Conference on Bioinformatics, Computational Biology and Biomedical Informatics, 2013. ACM, 76.

Reichardt W., Hu-Lowe D., Torres D., Weissleder R. and Bogdanov Jr A. 2005. Imaging of VEGF receptor kinase inhibitor-induced antiangiogenic effects in drug-resistant human adenocarcinoma model. *Neoplasia,* 7**,** 847-853.

Rosenbaum D. M., Rasmussen S. G. and Kobilka B. K. 2009. The structure and function of G-protein-coupled receptors. *Nature,* 459**,** 356.

Roy A., Kucukural A. and Zhang Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature protocols,* 5**,** 725.

Roy S., Ninkovic J., Banerjee S., Charboneau R. G., Das S., Dutta R., Kirchner V. A., Koodie L., Ma J. and Meng J. 2011. Opioid drug abuse and modulation of immune function: consequences in the susceptibility to opportunistic infections. *Journal of Neuroimmune Pharmacology,* 6**,** 442.

Rozek A., Friedrich C. L. and Hancock R. E. 2000. Structure of the bovine antimicrobial peptide indolicidin bound to dodecylphosphocholine and sodium dodecyl sulfate micelles. *Biochemistry,* 39**,** 15765-15774.

Rudan I., Lawn J., Cousens S., Rowe A. K., Boschi-Pinto C., Tomašković L., Mendoza W., Lanata C. F., Roca-Feltrer A. and Carneiro I. 2005. Gaps in policy-relevant

information on burden of disease in children: a systematic review. *The Lancet*, 365**,** 2031-2040.

Rutter W. C., Burgess D. R. and Burgess D. S. 2017. Increasing incidence of multidrug resistance among cystic fibrosis respiratory bacterial isolates. *Microbial Drug Resistance*, 23**,** 51-55.

Ruuskanen O., Lahti E., Jennings L. C. and Murdoch D. R. 2011. Viral pneumonia. *The Lancet*, 377**,** 1264-1275.

Sager R., Kutz A., Mueller B. and Schuetz P. 2017. Procalcitonin-guided diagnosis and antibiotic stewardship revisited. *BMC medicine*, 15**,** 15.

Salditt T., Li C. and Spaar A. 2006. Structure of antimicrobial peptides and lipid membranes probed by interface-sensitive X-ray scattering. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1758**,** 1483-1498.

Sato K., Pellegrino M., Nakagawa T., Nakagawa T., Vosshall L. B. and Touhara K. 2008. Insect olfactory receptors are heteromeric ligand-gated ion channels. *Nature*, 452**,** 1002.

Scheuer K., Maras A., Gattaz W., Cairns N., Förstl H. and Müller W. 1996. Cortical NMDA receptor properties and membrane fluidity are altered in Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders*, 7**,** 210-214.

Schiffner-Rohe J., Witt A., Hemmerling J., von Eiff C. and Leverkus F.-W. 2016. Efficacy of PPV23 in preventing pneumococcal pneumonia in adults at increased Risk–A systematic review and meta-analysis. *PloS one*, 11**,** e0146338.

Schmidtchen A., Pasupuleti M. and Malmsten M. 2014. Effect of hydrophobic modifications in antimicrobial peptides. *Advances in colloid and interface science*, 205**,** 265-274.

Schooley R. T., Biswas B., Gill J. J., Hernandez-Morales A., Lancaster J., Lessor L., Barr J. J., Reed S. L., Rohwer F. and Benler S. 2017. Development and use of personalized bacteriophage-based therapeutic cocktails to treat a patient with a disseminated resistant Acinetobacter baumannii infection. *Antimicrobial agents and chemotherapy*, 61**,** e00954-17.

Scott M. G., Rosenberger C. M., Gold M. R., Finlay B. B. and Hancock R. E. 2000. An α-helical cationic antimicrobial peptide selectively modulates macrophage responses to lipopolysaccharide and directly alters macrophage gene expression. *The Journal of Immunology*, 165**,** 3358-3365.

Seligman R., Papassotiriou J., Morgenthaler N. G., Meisner M. and Teixeira P. J. 2008. Copeptin, a novel prognostic biomarker in ventilator-associated pneumonia. *Critical Care*, 12**,** R11.

Sencanski M., Radosevic D., Perovic V., Gemovic B., Stanojevic M., Veljkovic N. and Glisic S. 2015. Natural products as promising therapeutics for treatment of influenza disease. *Current pharmaceutical design*, 21**,** 5573-5588.

Sengupta D., Leontiadou H., Mark A. E. and Marrink S.-J. 2008. Toroidal pores formed by antimicrobial peptides show significant disorder. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1778**,** 2308-2317.

Senthilkumar B. and Rajasekaran R. 2016. Computational resources for designing peptide based drugs preferred in the field of nanomedicine. *Journal of Bionanoscience*, 10**,** 1-14.

Sharma A., Kapoor P., Gautam A., Chaudhary K., Kumar R., Chauhan J. S., Tyagi A. and Raghava G. P. 2013. Computational approach for designing tumor homing peptides. *Scientific reports*, 3**,** 1607.

Shehabi Y. and Seppelt I. 2008. Pro/Con debate: is procalcitonin useful for guiding antibiotic decision making in critically ill patients? *Critical Care,* 12**,** 211.

Shima K., Coopmeiners J., Graspeuntner S., Dalhoff K. and Rupp J. 2016. Impact of micro-environmental changes on respiratory tract infections with intracellular bacteria. *FEBS letters,* 590**,** 3887-3904.

Sieprawska-Lupa M., Mydel P., Krawczyk K., Wójcik K., Puklo M., Lupa B., Suder P., Silberring J., Reed M. and Pohl J. 2004. Degradation of human antimicrobial peptide LL-37 by Staphylococcus aureus-derived proteinases. *Antimicrobial agents and chemotherapy,* 48**,** 4673-4679.

Singh R., Garg N., Shukla G., Capalash N. and Sharma P. 2016. Immunoprotective efficacy of Acinetobacter baumannii outer membrane protein, FilF, predicted in silico as a potential vaccine candidate. *Frontiers in microbiology,* 7**,** 158.

Sinha R. and Shukla P. 2019. Antimicrobial peptides: Recent insights on biotechnological interventions and future perspectives. *Protein and peptide letters,* 26**,** 79-87.

Siu L. K., Yeh K.-M., Lin J.-C., Fung C.-P. and Chang F.-Y. 2012. Klebsiella pneumoniae liver abscess: a new invasive syndrome. *The Lancet infectious diseases,* 12**,** 881-887.

Slotved H.-C., Facklam R. R. and Fuursted K. 2017. Assessment of a novel bile solubility test and MALDI-TOF for the differentiation of Streptococcus pneumoniae from other mitis group streptococci. *Scientific reports,* 7**,** 1-8.

Society A. T. and America I. D. S. o. 2005. Guidelines for the management of adults with hospital-acquired, ventilator-associated, and healthcare-associated pneumonia. *American journal of respiratory and critical care medicine,* 171**,** 388.

Söderlin M., Kastbom A., Kautiainen H., Leirisalo-Repo M., Strandberg G. and Skogh T. 2004. Antibodies against cyclic citrullinated peptide (CCP) and levels of cartilage oligomeric matrix protein (COMP) in very early arthritis: relation to diagnosis and disease activity. *Scandinavian journal of rheumatology,* 33**,** 185-188.

Solé-Violán J., de Castro F. R., García-Laorden M. I., Blanquer J., Aspa J., Borderías L., Briones M. L., Rajas O., Carrondo I. M.-L. and Marcos-Ramos J. A. 2010. Genetic variability in the severity and outcome of community-acquired pneumonia. *Respiratory medicine,* 104**,** 440-447.

Son Y. G., Shin J. and Ryu H. G. 2017. Pneumonitis and pneumonia after aspiration. *Journal of dental anesthesia and pain medicine,* 17**,** 1-12.

Song E., Zhu P., Lee S.-K., Chowdhury D., Kussman S., Dykxhoorn D. M., Feng Y., Palliser D., Weiner D. B. and Shankar P. 2005. Antibody mediated in vivo delivery of small interfering RNAs via cell-surface receptors. *Nature biotechnology,* 23**,** 709.

Southey B. R., Amare A., Zimmerman T. A., Rodriguez-Zas S. L. and Sweedler J. V. 2006. NeuroPred: a tool to predict cleavage sites in neuropeptide precursors and provide the masses of the resulting peptides. *Nucleic acids research,* 34**,** W267-W272.

Spear P. G., Eisenberg R. J. and Cohen G. H. 2000. Three classes of cell surface receptors for alphaherpesvirus entry. *Virology,* 275**,** 1-8.

Strieter R. M., Belperio J. A. and Keane M. P. 2003. Host innate defenses in the lung: the role of cytokines. *Current opinion in infectious diseases,* 16**,** 193-198.

Suarez D. L. 2016. Influenza A virus. *Animal Influenza***,** 1-30.

Suntrup-Krueger S., Kemmling A., Warnecke T., Hamacher C., Oelenberg S., Niederstadt T., Heindel W., Wiendl H. and Dziewas R. 2017. The impact of lesion location on dysphagia incidence, pattern and complications in acute stroke. Part 2: Oropharyngeal residue, swallow and cough response, and pneumonia. *European journal of neurology,* 24**,** 867-874.

Swart E. C. 2012. *The information-dense genomes of Oxytricha trifallax.* Princeton University.

Sweet D. G., Carnielli V., Greisen G., Hallman M., Ozek E., Plavka R., Saugstad O. D., Simeoni U., Speer C. P. and Vento M. 2017. European consensus guidelines on the management of respiratory distress syndrome-2016 update. *Neonatology,* 111**,** 107-125.

Sychantha D. and Clarke A. J. 2018. Peptidoglycan modification by the catalytic domain of Streptococcus pneumoniae OatA follows a ping-pong bi-bi mechanism of action. *Biochemistry,* 57**,** 2394-2401.

Tachi T., Epand R. F., Epand R. M. and Matsuzaki K. 2002. Position-dependent hydrophobicity of the antimicrobial magainin peptide affects the mode of peptide− lipid interactions and selective toxicity. *Biochemistry,* 41**,** 10723-10731.

Tagami H., Ray-Gallet D., Almouzni G. and Nakatani Y. 2004. Histone H3. 1 and H3. 3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. *Cell,* 116**,** 51-61.

Tahar R., Sayang C., Foumane V. N., Soula G., Moyou-Somo R., Delmont J. and Basco L. K. 2013. Field evaluation of rapid diagnostic tests for malaria in Yaounde, Cameroon. *Acta tropica,* 125**,** 214-219.

Tamposis I. A., Tsirigos K. D., Theodoropoulou M. C., Kontou P. I. and Bagos P. G. 2019. Semi-supervised learning of Hidden Markov Models for biological sequence analysis. *Bioinformatics,* 35**,** 2208-2215.

Tang B. M., McLean A. S., Dawes I. W., Huang S. J., Cowley M. J. and Lin R. C. 2008. Gene-expression profiling of gram-positive and gram-negative sepsis in critically ill patients. *Critical care medicine,* 36**,** 1125-1128.

Tang Q. Y. and Zhang C. X. 2013. Data Processing System (DPS) software with experimental design, statistical analysis and data mining developed for use in entomological research. *Insect Science,* 20**,** 254-260.

Taoka K.-i., Ohki I., Tsuji H., Furuita K., Hayashi K., Yanase T., Yamaguchi M., Nakashima C., Purwestri Y. A. and Tamaki S. 2011. 14-3-3 proteins act as intracellular receptors for rice Hd3a florigen. *Nature,* 476**,** 332.

Tatsumi T., Herrem C. J., Olson W. C., Finke J. H., Bukowski R. M., Kinch M. S., Ranieri E. and Storkus W. J. 2003. Disease stage variation in CD4+ and CD8+ T-cell reactivity to the receptor tyrosine kinase EphA2 in patients with renal cell carcinoma. *Cancer research,* 63**,** 4481-4489.

Taubenberger J. K., Reid A. H., Lourens R. M., Wang R., Jin G. and Fanning T. G. 2005. Characterization of the 1918 influenza virus polymerase genes. *Nature,* 437**,** 889.

Techaarpornkul S., Collins P. L. and Peeples M. E. 2002. Respiratory syncytial virus with the fusion protein as its only viral glycoprotein is less dependent on cellular glycosaminoglycans for attachment than complete virus. *Virology,* 294**,** 296-304.

Tejera A., Santolaria F., Diez M.-L., Alemán-Valls M.-R., González-Reimers E., Martínez-Riera A. and Milena-Abril A. 2007. Prognosis of community acquired pneumonia (CAP): value of triggering receptor expressed on myeloid cells-1

(TREM-1) and other mediators of the inflammatory response. *Cytokine,* 38**,** 117-123.

Terracina M., Franchi J., Bonté F., Romagnoli G., Maurelli R., Failla C., Dumas M., Marconi A., Fila C. and Pincelli C. 2003. Expression and function of neurotrophins and their receptors in cultured human keratinocytes. *Journal of investigative dermatology,* 121**,** 1515-1521.

Thomas Jr C. F. and Limper A. H. 2004. Pneumocystis pneumonia. *New England Journal of Medicine,* 350**,** 2487-2498.

Thomas S., Karnik S., Barai R. S., Jayaraman V. K. and Idicula-Thomas S. 2009. CAMP: a useful resource for research on antimicrobial peptides. *Nucleic acids research,* 38**,** D774-D780.

Thompson J. D., Higgins D. G. and Gibson T. J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research,* 22**,** 4673-4680.

Thornton C. R. 2008. Development of an immunochromatographic lateral-flow device for rapid serodiagnosis of invasive aspergillosis. *Clin. Vaccine Immunol.,* 15**,** 1095-1105.

Tincho M., Gabere M. and Pretorius A. 2016. In silico identification and molecular validation of putative antimicrobial peptides for HIV therapy. *Journal of AIDS and Clinical Research,* 7.

Todhunter D. A., Smith K. and Hogan J. 1991. Antibodies to iron-regulated outer membrane proteins of coliform bacteria isolated from bovine intramammary infections. *Veterinary immunology and immunopathology,* 28**,** 107-115.

Torrent M., Di Tommaso P., Pulido D., Nogués M. V., Notredame C., Boix E. and Andreu D. 2011. AMPA: an automated web server for prediction of protein antimicrobial regions. *Bioinformatics,* 28**,** 130-131.

Torres M. D., Sothiselvam S., Lu T. K. and de la Fuente-Nunez C. 2019. Peptide design principles for antimicrobial applications. *Journal of molecular biology,* 431**,** 3547-3567.

Tyrer P., Reed G. M. and Crawford M. J. 2015. Classification, assessment, prevalence, and effect of personality disorder. *The Lancet,* 385**,** 717-726.

Uteng M., Hauge H. H., Markwick P. R., Fimland G., Mantzilas D., Nissen-Meyer J. and Muhle-Goll C. 2003. Three-dimensional structure in lipid micelles of the pediocin-like antimicrobial peptide sakacin P and a sakacin P variant that is structurally stabilized by an inserted C-terminal disulfide bridge. *Biochemistry,* 42**,** 11417-11426.

Vabret A., Mouthon F., Mourez T., Gouarin S., Petitjean J. and Freymuth F. 2001. Direct diagnosis of human respiratory coronaviruses 229E and OC43 by the polymerase chain reaction. *Journal of virological methods,* 97**,** 59-66.

van der Linden M., Falkenhorst G., Perniciaro S. and Imöhl M. 2015. Effects of infant pneumococcal conjugate vaccination on serotype distribution in invasive pneumococcal disease among children and adults in Germany. *PloS one,* 10**,** e0131494.

van der Meer V., Neven A. K., van den Broek P. J. and Assendelft W. J. 2005. Diagnostic value of C reactive protein in infections of the lower respiratory tract: systematic review. *Bmj,* 331**,** 26.

Veltri D., Kamath U. and Shehu A. 2018. Deep learning improves antimicrobial peptide recognition. *Bioinformatics,* 34**,** 2740-2747.

Vemula S., Zhao J., Liu J., Wang X., Biswas S. and Hewlett I. 2016. Current approaches for diagnosis of influenza virus infections in humans. *Viruses,* 8**,** 96.

Venkatakrishnan A., Deupi X., Lebon G., Tate C. G., Schertler G. F. and Babu M. M. 2013. Molecular signatures of G-protein-coupled receptors. *Nature,* 494**,** 185.

Vizioli J. and Salzet M. 2002. Antimicrobial peptides from animals: focus on invertebrates. *Trends in Pharmacological Sciences,* 23**,** 494-496.

Voeten J., Groen J., Van Alphen D., Claas E., De Groot R., Osterhaus A. and Rimmelzwaan G. 1998. Use of recombinant nucleoproteins in enzyme-linked immunosorbent assays for detection of virus-specific immunoglobulin A (IgA) and IgG antibodies in influenza virus A-or B-infected patients. *Journal of clinical microbiology,* 36**,** 3527-3531.

Von Bernuth H., Picard C., Jin Z., Pankla R., Xiao H., Ku C.-L., Chrabieh M., Mustapha I. B., Ghandil P. and Camcioglu Y. 2008. Pyogenic bacterial infections in humans with MyD88 deficiency. *Science,* 321**,** 691-696.

Wade J. D., Lin F., Hossain M. A. and Dawson R. M. 2012. Chemical synthesis and biological evaluation of an antimicrobial peptide gonococcal growth inhibitor. *Amino Acids,* 43**,** 2279-2283.

Waghu F. H., Barai R. S., Gurung P. and Idicula-Thomas S. 2015. CAMPR3: a database on sequences, structures and signatures of antimicrobial peptides. *Nucleic acids research,* 44**,** D1094-D1097.

Walport M. J. 2001. Complement. *New England Journal of Medicine,* 344**,** 1058-1066.

Wang C.-K., Shih L.-Y. and Chang K. 2017a. Large-scale analysis of antimicrobial activities in relation to amphipathicity and charge reveals novel characterization of antimicrobial peptides. *Molecules,* 22**,** 2037.

Wang G., Li X. and Wang Z. 2008. APD2: the updated antimicrobial peptide database and its application in peptide design. *Nucleic acids research,* 37**,** D933-D937.

Wang J. T., Zaki M. J., Toivonen H. T. and Shasha D. 2005. Introduction to data mining in bioinformatics. *Data Mining in Bioinformatics.* Springer.

Wang L., Wang Y. and Chang Q. 2016. Feature selection methods for big data bioinformatics: A survey from the search perspective. *Methods,* 111**,** 21-31.

Wang P., Hu L., Liu G., Jiang N., Chen X., Xu J., Zheng W., Li L., Tan M. and Chen Z. 2011. Prediction of antimicrobial peptides based on sequence alignment and feature selection methods. *PloS one,* 6**,** e18476.

Wang Y., Chen H., Guo Z., Sun L., Fu Y., Li T., Lin W. and Lin X. 2017b. Quantitative proteomic analysis of iron-regulated outer membrane proteins in Aeromonas hydrophila as potential vaccine candidates. *Fish & shellfish immunology,* 68**,** 1-9.

Wang Z. and Wang G. 2004. APD: the antimicrobial peptide database. *Nucleic acids research,* 32**,** D590-D592.

Wattanathum A., Manocha S., Groshaus H., Russell J. A. and Walley K. R. 2005. Interleukin-10 haplotype associated with increased mortality in critically ill patients with sepsis from pneumonia but not in patients with extrapulmonary sepsis. *Chest,* 128**,** 1690-1698.

Wazny K., Zipursky A., Black R., Curtis V., Duggan C., Guerrant R., Levine M., Petri Jr W. A., Santosham M. and Scharf R. 2013. Setting research priorities to reduce mortality and morbidity of childhood diarrhoeal disease in the next 15 years. *PLoS medicine,* 10**,** e1001446.

Weinberg E. D. 2009. Iron availability and infection. *Biochimica et Biophysica Acta (BBA)-General Subjects,* 1790**,** 600-605.

Weinberger D. M., Trzciński K., Lu Y.-J., Bogaert D., Brandes A., Galagan J., Anderson P. W., Malley R. and Lipsitch M. 2009. Pneumococcal capsular polysaccharide structure predicts serotype prevalence. *PLoS pathogens,* 5**,** e1000476.

Wenzler E., Goff D. A., Mangino J. E., Reed E. E., Wehr A. and Bauer K. A. 2016. Impact of rapid identification of Acinetobacter baumannii via matrix-assisted laser desorption ionization time-of-flight mass spectrometry combined with antimicrobial stewardship in patients with pneumonia and/or bacteremia. *Diagnostic microbiology and infectious disease,* 84**,** 63-68.

Williams M., Tincho M., Gabere M., Uys A., Meyer M. and Pretorius A. 2016. Molecular validation of putative antimicrobial peptides for improved Human Immunodeficiency Virus diagnostics via HIV protein p24. *J. AIDS Clin. Res,* 7**,** 571.

Wrammert J., Koutsonanos D., Li G.-M., Edupuganti S., Sui J., Morrissey M., McCausland M., Skountzou I., Hornig M. and Lipkin W. I. 2011. Broadly cross-reactive antibodies dominate the human B cell response against 2009 pandemic H1N1 influenza virus infection. *Journal of Experimental Medicine,* 208**,** 181-193.

Wright P. F., Karron R. A., Belshe R. B., Shi J. R., Randolph V. B., Collins P. L., O'Shea A. F., Gruber W. C. and Murphy B. R. 2007. The absence of enhanced disease with wild type respiratory syncytial virus infection occurring after receipt of live, attenuated, respiratory syncytial virus vaccines. *Vaccine,* 25**,** 7372-7378.

Wu C. H., Apweiler R., Bairoch A., Natale D. A., Barker W. C., Boeckmann B., Ferro S., Gasteiger E., Huang H. and Lopez R. 2006. The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic acids research,* 34**,** D187-D191.

Wu W., Wieckowski S., Pastorin G., Benincasa M., Klumpp C., Briand J. P., Gennaro R., Prato M. and Bianco A. 2005. Targeted delivery of amphotericin B to cells by using functionalized carbon nanotubes. *Angewandte Chemie International Edition,* 44**,** 6358-6362.

Xiao X., Wang P., Lin W.-Z., Jia J.-H. and Chou K.-C. 2013. iAMP-2L: a two-level multi-label classifier for identifying antimicrobial peptides and their functional types. *Analytical biochemistry,* 436**,** 168-177.

Xuan V.-T. 2018. Rice production, agricultural research, and the environment. *Vietnam's rural transformation.* Routledge.

Yang H.-B., Hou W.-T., Cheng M.-T., Jiang Y.-L., Chen Y. and Zhou C.-Z. 2018. Structure of a MacAB-like efflux pump from Streptococcus pneumoniae. *Nature communications,* 9**,** 196.

Yeaman M. R. and Yount N. Y. 2003. Mechanisms of antimicrobial peptide action and resistance. *Pharmacological reviews,* 55**,** 27-55.

Yu S.-L., Chen H.-W., Yang P.-C., Peck K., Tsai M.-H., Chen J. J. and Lin F.-Y. 2004. Differential gene expression in gram-negative and gram-positive sepsis. *American journal of respiratory and critical care medicine,* 169**,** 1135-1143.

Yuan C. 2018. Hidden Markov Models.

Yuen K., Chan P., Peiris M., Tsang D., Que T., Shortridge K., Cheung P., To W., Ho E. and Sung R. 1998. Clinical features and rapid viral diagnosis of human disease associated with avian influenza A H5N1 virus. *The Lancet,* 351**,** 467-471.

Zhang G.-H., Mann D. M. and Tsai C.-M. 1999. Neutralization of endotoxin in vitro and in vivo by a human lactoferrin-derived peptide. *Infection and immunity,* 67**,** 1353-1358.

Zhao J., Deng Y., Jiang Z. and Qing H. 2016. G protein-coupled receptors (GPCRs) in Alzheimer's disease: a focus on BACE1 related GPCRs. *Frontiers in aging neuroscience,* 8**,** 58.

Zhu X. and Mitchell J. C. 2011. KFC2: a knowledge-based hot spot prediction method based on interface solvation, atomic density, and plasticity features. *Proteins: Structure, Function, and Bioinformatics,* 79**,** 2671-2683.

Zilberberg M. D., Nathanson B. H., Sulham K., Fan W. and Shorr A. F. 2017. Carbapenem resistance, inappropriate empiric treatment and outcomes among patients hospitalized with Enterobacteriaceae urinary tract infection, pneumonia and sepsis. *BMC infectious diseases,* 17**,** 279.

APPENDICES

Appendix A: Supplementary Tables for Chapter 2

Table A1: Anti-*Acinetobacter baumannii* AMPs

ID	AMP	*SOURCE*	SEQUENCE
CAMPSQ3632	Temporin-ALd	*Amolops loloensis*	FLPIAGKLLSGLSGLL
CAMPSQ3633	Temporin-Ale	*Amolops loloensis*	FFPIVGKLLFGLSGLL
CAMPSQ3634	Temporin-Alf	*Amolops loloensis*	FFPIVGKLLSGLSGLL
CAMPSQ3635	Temporin-ALg	*Amolops loloensis*	FFPIVGKLLFGLFGLL
CAMPSQ3636	Temporin-ALh	*Amolops loloensis*	FLPIVGKLLSGLSGLS
CAMPSQ3424	CPF-C1	*Xenopus clivii*	GFGSLLGKALRLGANVL
CAMPSQ3327	B2RP-Era	*Hylarana erythraea*	GVIKSVLKGVAKTVALGML
CAMPSQ1090	Moricin-1, Moricin-2	*Bombyx mori*	AKIPIKAIKTVGKAVGKGLRAI NIASTANDVFNFLKPKKRKH
CAMPSQ2591	Hydramacin-1	*Hydra vulgaris [Hydra]*	IVDCWETWSRCTKWSQGGTGT LWKSCNDRCKELGRKRGQCEE KPSRCPLSKKAWTCICY
CAMPSQ3331	Hydramacin-1	*Hydra magnipapillata*	QIVDCWETWSRCTKWSQGGT GTLWKSCNDRCKELGRKRGQC EEKPSRCPLSKKAWTCICY
AP00094	Temporin A	*Rana temporaria*	FLPLIGRVLSGIL
AP00524	Human beta defensin 2 (hBD-2)	*Homo sapiens*	GIGDPVTCLKSGAICHPVFCPR RYKQIGTCGLPGTKCCKKP

Table A2: *Anti-Influenza A* virus AMPs

ID	AMP	SOURCE	SEQUENCE
CAMPSQ350	Alloferon-1	Calliphora vicina [Blue blowfly]	HGVSGHGQHGVHG
AVP0554	EB	FGF-4 signal sequence, Orthomyxoviridae	RRKKAAVALLPAVLLALLAP
AVP0961	18-c01	Phage display	GWWYKGRARPVSAVA
AVP0962	C18-c03	Phage display	RAVWRHSVATPSHSV
AVP0963	C18-c01W2A	Phage display	GAWYKGRARPVSAVA
AVP0964	C18-c01P10A	Phage display	GWWYKGRARAVSAVA

AVP0965	C18-c01r	Phage display	AVASVPRARGKYWWG
AVP0966	C18-p1b	Phage display	DFRRLPGAFWQLRQP
AVP0967	C18-cp3c	Phage display	AETVESCLAKPHTEN
AVP0968	c01	Phage display	GWWYKGRARPVSAVA
AVP0969	c03	Phage display	RAVWRHSVATPSHSV
AVP0977	45658	INFV A polymerase (PB1)	MDVNPTLLFLKVPAQNAISTTFPYT
AVP0978	45717	INFV A polymerase (PB1)	VNPTLLFLKVPAQNAISTTFPYT
AVP0979	45778	INFV A polymerase (PB1)	PTLLFLKVPAQNAISTTFPYT
AVP0980	45839	INFV A polymerase (PB1)	LLFLKVPAQNAISTTFPYT
AVP0981	45901	INFV A polymerase (PB1)	FLKVPAQNAISTTFPYT
AVP0982	45962	INFV A polymerase (PB1)	KVPAQNAISTTFPYT
AVP0983	43831	INFV A polymerase (PB1)	MDVNPTLLFLKVPAQNAIST
AVP0984	43101	INFV A polymerase (PB1)	MDVNPTLLFLKVPAQNAI
AVP0985	42370	INFV A polymerase (PB1)	MDVNPTLLFLKVPAQN
AVP0986	42005	INFV A polymerase (PB1)	MDVNPTLLFLKVPAQ
AVP0987	41640	INFV A polymerase (PB1)	MDVNPTLLFLKVPA
AVP0988	41275	INFV A polymerase (PB1)	MDVNPTLLFLKVP
AVP0989	41609	INFV A polymerase (PB1)	MDVNPTLLFLKV
AVP0990	41579	INFV A polymerase (PB1)	MDVNPTLLFLK
AVP0991	41548	INFV A polymerase (PB1)	MDVNPTLLFL
AVP0992	41518	INFV A polymerase (PB1)	MDVNPTLLF
AVP0993	41487	INFV A polymerase (PB1)	MDVNPTLL
AVP0994	41456	INFV A polymerase (PB1)	MDVNPTL
AVP0995	41426	INFV A polymerase (PB1)	MDVNPT
AVP0996	PB1-1-15 B	INFV A polymerase (PB1)	MNINPYPLFIDVPIQ
AVP0997	PB1-1-15 A D2N,V3I,L14I	INFV A polymerase (PB1)	MNINPTLLFLKVPIQ
AVP0998	PB1-1-15 A L10I,K11D	INFV A polymerase (PB1)	MDVNPTLLFIDVPAQ

AVP0999	PB1-1-15 A D2N,V3I	INFV A polymerase (PB1)	MNINPTLLFLKVPAQ
AVP1000	PB1-1-15 A T6Y,L7F	INFV A polymerase (PB1)	MDVNPYFLFLKVPAQ
AVP1001	PB1-1-15 A L7F	INFV A polymerase (PB1)	MDVNPTFLFLKVPAQ
AVP1002	PB1-1-15 A T6Y	INFV A polymerase (PB1)	MDVNPYLLFLKVPAQ
AVP1003	PB1-1-15 A T6F	INFV A polymerase (PB1)	MDVNPFLLFLKVPAQ
AVP1004	PB1-1-15 A T6W	INFV A polymerase (PB1)	MDVNPWLLFLKVPAQ
AVP1005	PB1-1-15 A T6H	INFV A polymerase (PB1)	MDVNPHLLFLKVPAQ
AVP1006	PB1-1-15 A T6C	INFV A polymerase (PB1)	MDVNPCLLFLKVPAQ
AVP1753	FP3	INFV A polymerase (PB1)	WLVFFVIFYFFRRRKK
AVP1754	FP4	INFV A polymerase (PB1)	RRKKWLVFFVIFYFFR
AVP1755	FP2	INFV A polymerase (PB1)	WLVFFVIAYFAR
AVP1756	FP1 (Tkip)	Mimetic for the SOCS protein	WLVFFVIFYFFR
AVP1759	FP7	Mimetic for the SOCS protein	RRKKIFYFFR
AVP1806	FP8	Mimetic for the SOCS protein	WLVFFVRRKK
AVP1808	FP9	Mimetic for the SOCS protein	FFVIFYRRKK
AVP2028	Peptide 6	INFV A matrix protein 1	CATCEQIADSQHRSHRQMV
AVP2057	Mucroporin-S2	Scorpion venom	LFGLIPSLIGGLVSAFK

Table A3: *Anti-Influenza B* virus AMPs

ID	AMP	SOURCE	SEQUENCE
AVP1007	PB1-1-15 A	INFV A polymerase (PB1)	MDVNPTLLFLKVPAQ
AVP1008	PB1-1-15 B	INFV A polymerase (PB1)U	MNINPYPLFIDVPIQ
AVP1009	PB1-1-15 A D2N,V3I,L14I	INFV A polymerase (PB1)	MNINPTLLFLKVPIQ
AVP1010	PB1-1-15 A L10I,K11D	INFV A polymerase (PB1)	MDVNPTLLFIDVPAQ
AVP1011	PB1-1-15 A D2N,V3I	INFV A polymerase (PB1)	MNINPTLLFLKVPAQ
AVP1012	PB1-1-15 A T6Y,L7F	INFV A polymerase (PB1)	MDVNPYFLFLKVPAQ

AVP1013	PB1-1-15 A L7F	INFV A polymerase (PB1)	MDVNPTFLFL KVPAQ
AVP1014	PB1-1-15 A T6Y	INFV A polymerase (PB1)	MDVNPYLLFL KVPAQ
AVP1015	PB1-1-15 A T6F	INFV A polymerase (PB1)	MDVNPFLLFL KVPAQ
AVP1016	PB1-1-15 A T6W	INFV A polymerase (PB1)	MDVNPWLLF LKVPAQ
AVP1017	PB1-1-15 A T6H	INFV A polymerase (PB1)	MDVNPHLLFL KVPAQ
AVP1018	PB1-1-15 A T6C	INFV A polymerase (PB1)	MDVNPCLLFL KVPAQ

Table A4: Anti-*Klebsiella pneumoniae* AMPs

ID	AMP	SOURCE	SEQUENCE
CAMPSQ2791	Anionic peptide SAAP	*Pasteurella haemolytica*	DDDDDD
CAMPSQ3121	Microcin 7	*E. coli*	MRTGNAD
CAMPSQ834	Microcin C7	*E. coli*	MRTGNAN
CAMPSQ3110	Jelleine-1	*Apis mellifera*	PFKLSLHL
CAMPSQ665	Jelleine-1	*Apis mellifera*	PFKLSLHL
CAMPSQ3111	Jelleine-2	*Apis mellifera*	TPFKLSLHL
CAMPSQ666	Jellein-2	*Honey bee*	TPFKISIHL
CAMPSQ2985	Serracin-P 43 kDa subunit	***Erwinia amylovora-bacterium***	DYHHGVRVL
CAMPSQ3841	HaA4	*Harmonia axyridis*	IGGYCSWLRL
CAMPSQ3842	Temporin-ECa	*Euphlyctis cyanophlyctis*	FLPGLLAGLL
CAMPSQ736	Decoralin	*Oreumenes decoratus*	SLLSLIRKLIT
CAMPSQ4322	HLP6	*Homo sapiens*	FQWQRNPRKVR
CAMPSQ170	Tritrpticin	*Sus scrofa*	VRRFPWWWPFLRR
CAMPSQ8186	Mastoparan-S	*Sphodromantis viridis*	LRLKSIVSYAKKVL
CAMPSQ3843	Buforin-EC	*Euphlyctis cyanophlyctis*	RAGLKFPVGRVHRLLR

CAMPSQ215	Ponericin-G6	*Pachycondyla goeldii*	GLVDVLGKVGGLIKKLLP
CAMPSQ3118	Ascaphin-8	*Ascaphus truei*	GFKDLLKGAAKALVKTVLF
CAMPSQ608	Pilosulin-1	*Myrmecia pilosula*	GLGSVFGRLARILGRVIPKV
CAMPSQ3844	Cyanophlyctin	*Euphlyctis cyanophlyctis*	FLNALKNFAKTAGKRLKSLLN
CAMPSQ271	Thanatin	*Podisus maculiventris [Spined soldier bug]*	GSKKPVPIIYCNRRTGKCQRM
CAMPSQ573	Kassinatuerin-1	*Kassina senegalensis*	GFMKYIGPLIPHAVKAISDLI
CAMPSQ575	Nigrocin-2	*Rana nigromaculata*	GLLSKVLGVGKKVLCGVSGLC
CAMPSQ609	PYLa/PGLa precursor	*Xenopus laevis*	GMASKAGAIAGKIAKVALKAL
CAMPSQ2711	LP1A	*Manduca sexta*	QRFSQPTFKLPQGRLTLSRKF
CAMPSQ1009	Brevinin-1TSa	*Rana tsushimensis*	FLGSIVGALASALPSLISKIRN
CAMPSQ109	Magainin-1	*Xenopus laevis*	GIGKFLHSAGKFGKAFVGEIMKS
CAMPSQ110	Magainin 2	*Xenopus laevis*	GIGKFLHSAKKFGKAFVGEIMNS
CAMPSQ4332	Ascaphin-1	*Ascaphus truei*	GFRDVLKGAAKAFVKTVAGHIAN
CAMPSQ283	Gaegurin-5	*Rana rugosa*	FLGALFKVASKVLPSVFCAITKKC
CAMPSQ223	Ponericin-W5	*Pachycondyla goeldii*	FWGALIKGAAKLIPSVVGLFKKKQ
CAMPSQ4333	Ascaphin-5	*Ascaphus truei*	GIKDWIKGAAKKLIKTVASHIANQ
CAMPSQ263	Maximin 5	*Bombina maxima*	SIGAKILGGVKTFFKGALKELASTYLQ
CAMPSQ3666	Hymenochirin-3B	*Hymenochirus boettgeri*	IKIPAVVKDTLKKVAKGVLSAVAGALTQ
CAMPSQ3667	Hymenochirin-4B	*Hymenochirus boettgeri*	IKIPAFVKDTLKKVAKGVISAVAGALTQ
CAMPSQ3668	Hymenochirin-5B	*Hymenochirus boettgeri*	IKIPPIVKDTLKKVAKGVLSTIAGALST
CAMPSQ3664	Hymenochirin-1B	*Hymenochirus boettgeri*	IKLSPETKDNLKKVLKGAIKGAIAVAKMV
CAMPSQ3665	Hymenochirin-2B	*Hymenochirus boettgeri*	LKIPGFVKDTLKKVAKGIFSAVAGAMTPS

CAMPSQ877	Cathelin-related peptide SC5	*Cathelin-related peptide*	RGLRRLGRKIAHGVKKYGPTVLRII RIAG
CAMPSQ213	Ponericin-G4	*Pachycondyla goeldii*	DFKDWMKTAGEWLKKKGPGILKA AMAAAT
CAMPSQ2590	L-amino-acid oxidase	*Bothropoides matogrossensis*	KFEPPLPPKKAHKKFWEDDGIYYP PNHNFP
CAMPSQ3497	Cliotide T4, Cyclotide cter-P	*Clitoria ternatea [Butterfly pea]*	GIPCGESCVFIPCITAAIGCSCKSKV CYRN
CAMPSQ1054	Fowlicidin-2 , Myeloid antimicrobial peptide 27	*Gallus gallus [Chicken]*	LVQRGRFGRFLRKIRRFRPKVTITIQ GSARFG
CAMPSQ574	Nigrocin-1	*Rana nigromaculata (dark-spotted frog) [Black-spotted frog]*	GLLDSIKGMAISAGKGALQNLLKV ASCKLDKTC
CAMPSQ69	Gaegurin-1	*Rana rugosa [Japanese wrinkled frog]*	SLFSLIKAGAKFLGKNLLKQGACY AACKASKQC
CAMPSQ70	Gaegurin-2	*Rana rugosa [Japanese wrinkled frog]*	GIMSIVKDVAKNAAKEAAKGALST LSCKLAKTC
CAMPSQ71	Gaegurin-3	*Rana rugosa [Japanese wrinkled frog]*	GIMSIVKDVAKTAAKEAAKGALST LSCKLAKTC
CAMPSQ3845	Sarcotoxin Pd	*Paederus dermatitis-insect*	GWLKKIGKKIERVGQHTRGLGIAQI AANVAATAR
CAMPSQ3865	Papillosin	*Halocynthia papillosa [Red sea-squirt]-fish*	GFWKKVGSAAWGGVKAAAKGAA VGGLNALAKHIQ
CAMPSQ284	Sapecin-B	*Sarcophaga peregrina [Flesh fly]*	LTCEIDRSLCLLHCRLKGYLRAYCS QQKVCRCVQ
CAMPSQ571	Canine beta-defensin	*Canis lupus familiaris [Dog]*	KCWNLRGSCREKCIKNEKLYIFCTS GKLCCLKPK
CAMPSQ1107	Cecropin-A	*Aedes aegypti [Yellowfever mosquito]*	GGLKKLGKKLEGAGKRVFNAAEK ALPVVAGAKAL

CAMPSQ7512	Paneth cell-specific alpha-defensin 1	*Equus caballus [horse]*	SCTCRRAWICRWGERHSGKCIDQKGSTYRLCCRR
CAMPSQ4340	Chicken AvBD2	*Gallus gallus [Chicken]*	LFCKGGSCHFGGCPSHLIKVGSCFRSCCKWPWNA
CAMPSQ3132	Brevinin-2PRa	*Rana pirica-frog*	GLMSLFKGVLKTAGKHIFKNVGGSLLDQAKCKITGEC
CAMPSQ3133	Brevinin-2PRb	*Rana pirica*	GLMSLFRGVLKTAGKHIFKNVGGSLLDQAKCKITGEC
CAMPSQ3134	Brevinin-2PRc	*Rana pirica*	GLMSVTKGVLKTAGKHIFKNVGGSLLDQAKCKISGQC
CAMPSQ3135	Brevinin-2PRd	*Rana pirica*	GLMSVLKGVLKTAGKHIFKNVGGSLLDQAKCKITGQC
CAMPSQ3136	Brevinin-2Pre	*Rana pirica*	GLLSVLKGVLKTTGKHIFKNVGGSLLDQAKCKISGQC
CAMPSQ3767	Esculentin-2Cha	*Lithobates chiricahuensis-frog*	GFSSIFRGVAKFASKGLGKDLAKLGVDLVACKISKQC
CAMPSQ324	Tracheal antimicrobial peptide	*Bos taurus [Bovine]-cattle*	NPVSCVRNKGICVPIRCPGNMKQIGTCVGRAVKCCRKK
CAMPSQ135	Tracheal antimicrobial peptide precursor	*Bos taurus [Bovine]*	NPVSCVRNKGICVPIRCPGSMKQIGTCVGRAVKCCRKK
CAMPSQ734	Cathelicidin	*Canis lupus familiaris [Dog]*	RLKELITTGGQKIGEKIRRIGQRIKDFFKNLQPREEKS
CAMPSQ4339	Chicken AvBD1	*Gallus gallus [Chicken]*	GRKSDCFRKSGFCAFLKPSLTLISGKCSRFYLCCKRIW
CAMPSQ397	Sapecin-C	*Sarcophaga peregrina [Flesh fly]*	ATCDLLSGIGVQHSACALHCVFRGNRGGYCTGKGICVCRN
CAMPSQ3854	Bacteriocin	*Lactococcus sp.*	TSYGNGVHCNKSKCWIDVSELETYKAGTVSNPKDILWSLKE
CAMPSQ106	Hadrurin	*Hadrurus aztecus [Mexican scorpion]*	GILDTIKSIASKVWNSKTVQDLKRKGINWVANKLGVSPQAA
CAMPSQ3	Bactenecin 5 – bovine	*Bos taurus [Bovine]*	RFRPPIRRPPIRPPFYPPFRPPIRPPIFPPIRPPFRPPLRFP
CAMPSQ3356	CBD-1	*Canis lupus [Gray wolf]*	KCWNLRGSCREKCIKNEKLYIFCTSGKLCCLKPKFQPNMLQR
CAMPSQ4323	Drosomycin	*Drosophila melanogaster [Fruit fly]*	CLSGRYKGPCAVWDNETCRRVCKEEGRSSGHCSPSLKCWCEGC

CAMPSQ313	Opistoporin-1	*Opistophthalmus carinatus [African yellow leg scorpion]*	GKVWDWIKSTAKKLWNSEPVKELKNTALNAAKNLVAEKIGATPS
CAMPSQ631	Antimicrobial peptide eNAP-2	*Equus caballus [Horse]*	EVERKHPLGGSRPGRCPTVPPGTFGHCACLCTGDASEPKGQKCCSN
CAMPSQ7576	Vejovine	*Vaejovis mexicanus [Mexican scorpion]*	GIWSSIKNLASKAWNSDIGQSLRNKAAGAINKFVADKIGVTPSQAAS
CAMPSQ1166	Pg-AMP	*Psidium guajava [Guava]*	RESPSSRMECYEQAERYGYGGYGGGRYGGGYGSGRGQPVGQGVERSHDDNRNQPR
CAMPSQ3830	CCL27	*Homo sapiens [Human]*	PPSTACCTQLYRKPLSDKLLRKVIQVELQEADGDCHLQAFVLHLAQRSICIHPQNP
CAMPSQ3476	Vejovine	*Vaejovis mexicanus(Arachnid)*	GIWSSIKNLASKAWNSDIGQSLRNKAAGAINKFVADKIGVTPSQAASMTLDEIVDAMYYD
CAMPSQ3078	Tachycitin	*Tachypleus tridentatus [Japanese horseshoe crab]*	YLAFRCGRYSPCLDDGPNVNLYSCCSFYNCHKCLARLENCPKGLHYNAYLKVCDWPSKAGCT
CAMPSQ881	ABF-2	*Caenorhabditis elegans-nematoda*	DIDFSTCARMDVPILKKAAQGLCITSCSMQNCGTGSCKKRSGRPTCVCYRCANGGGDIPLGALIKRG
CAMPSQ3831	MCCL28	*Mus musculus [Mouse]*	FQTSEAILPMASSCCTEVSHHVSGRLLERVSSCSIQRADGDCDLAAVILHVKRRRICISPHNRTLKQWMRASEVKKNGRENVCSGKKQPSRKDRKGHTTRKHRTRGTHRHEASR
CAMPSQ3726	Human angiogenin	*Homo sapiens [Human]*	QDNSRYTHFLTQHYDAKPQGRDDRYCESIMRRRGPTSPCKDINTFIHGNKRSIKAICENKNGNPHRENLRISKSSFQVTTCKLHGGSPWPPCQYRATAGFRNVVVACENGLPVHLDQSIFRRPRP
CAMPSQ3829	CCL28	*Homo sapiens [Human]*	MQQRGLAIVALAVCAALHASEAILPIASSCCTEVSHHISRRLLERVNMCRIQRADGDCDLAAVILHVKRRRICVSPHNHTVKQWMKVQAAKKNGKGNVCHRKKHHG

			KRNSNRAHQGKHET YGHKTPY
AP00026	Lactoferricin B (LfcinB, UCSS1a; 1S=S, cattle, ruminant, animals, ZZHa; BBL; Derivatives: lactoferrin peptide 2; LTX-302; LfcinB6; MPLfcinB6; JJsn)	*Bos taurus*-*cattle*	FKCRRWQWRMKKLGAPSITCVRR AF
AP00058	Maximin 1 (UCLL1c; XXA; toad, amphibians, animals; ZZS)	*Chinese red belly toad, Bombina maxima , Yunnan, China, Asia*	GIGTKILGGVKTALKGALKELAST YAN
AP00060	Maximin 3 (UCLL1; toad, amphibians, animals; ZZHa; ZZS)	*Bombina maxima*	GIGGKILSGLKTALKGAAKELASTY LH
AP00061	Maximin 4 (UCLL1c; XXA; Maximin-4; toad, amphibians, animals)	*Bombina maxima*	GIGGVLLSAGKAALKGLAKVLAEK YAN
AP00062	Maximin 5 (UCLL1; toad, amphibians, animals)	*Bombina maxima*	SIGAKILGGVKTFFKGALKELASTY LQ
AP00474	Piscidin 3 (Pis-3; fish, animals)	*hybrid striped bass (Morone saxatilis x Morone chrysops)-fish*	FIHHIFRGIVHAGRSIGRFLTG
AP02146	ALFpm3 (ALF-Pm3; P.	*the black tiger shrimp,*	QGWEAVAAAVASKIVGLWRNEKT ELLGHECKFTVKPYLKRFQVYYKG

	monodon anti-lipopolysaccharide factor 3; UCSS1a; Crustaceans, arthropods, invertebrates, animals; 1S=S; BBS)	*Penaeus monodon*	RMWCPGWTAIRGEASTRSQSGVA GKTAKDFVRKAFQKGLISQQEANQ WLSS
AP00009	BACTENECIN 5 (BtBac5; bac5, bac 5, bac-5; Pro-rich; bovine cathelicidin, cattle, ruminant, animals)	*Bovine neutrophils, Bos Taurus*	RFRPPIRRPPIRPPFYPPFRPPIRPPIFP PIRPPFRPPLGPFP
BAC098	Subtilosin	*Bacillus subtilis*	MKLPVQQVYS VYGGKDLPKG HSHSTMPFLS KLQFLTKIYL LDIHTQPFFI =50
BAC143	**Enterocin AS-48 (BACTERIOCIN AS-48)**	*Streptococcus pneumoniae*	MVKENKFSKI FILMALSFLG LALFSASLQF LPIAHMAKEF GIPAAVAGTV LNVVEAGGWV TTIVSILTAV GSGGLSLLAA AGRESIKAYL KKEIKKKGKR AVIAW
BAC174	Enterocin E-760	*Streptococcus pneumoniae*	NRWYCNSAAG GVGGAAVCGL AGYVGEAKEN IAGEVRKGWG MAGGFTHNKA CKSFPGSGWA SG

Table A5: Anti-*Respiratory Syncytial Virus* AMPs

ID	AMP	SOURCE	SEQUENCE
AP00281	mCRAMP (mouse cathelin-related antimicrobial peptide	*Mouse, Mus musculus*	GLLRKGGEKIGEKLKKIGQKIKNF FQKLVPQPEQ

AP00310	LL-37 [LL37; FALL-39; cathelicidin; UCLL1; human; chimpanzee; primates, mammals, animals; XXX; XXY; XXZ; BBBh2o, BBBm; BBMm, BBPP, BBN, BBL, BBR, JJsn; Derivatives: GF-17, IG-25, KR-12, etc.	*neutrophils, moncytes; lymphocytes, Mesenchymal Stem Cells; islets; skin, sweat; airway surface liquid, saliva; Homo sapiens; Also Pan troglodytes*	LLGDFFRKSKEKIGKEFKRIVQRIK DFLRNLVPRTES
AP02337	RNase 2 (eosinophil-derived neurotoxin, EDN; EPX; Ribonuclease superfamily; UCSS1a; 4S=S, humans; primates, mammals, animals)	*liver, lung, spleen, eosinophilic leukocytes; neutrophils, and monocytes, Homo sapiens*	KPPQFTWAQWFETQHINMTSQQC TNAMQVINNYQRRCKNQNTFLLT TFANVVNVCGNPNMTCPSNKTRK NCHHSGSQVPLIHCNLTTPSPQNIS NCRYAQTPANMFYIVACDNRDQ RRDPPQYPVVPVHLDRII
AP02338	mEar2 (mouse eosinophil-associated ribonuclease 2, mRNase 2, Ribonuclease superfamily; mice; mammals, animals)	*Mus musculus*	LGQTPSQWFAIQHINNNANLQCN VEMQRINRFRRTCKGLNTFLHTSF ANAVGVCGNPSGLCSDNISRNCH NSSSRVRITVCNITSRRRTPYTQCR YQPRRSLEYYTVACNPRTPQDSP MYPVVPVHLDGTF
AVP0127	T-104	*RSV fusion (F) protein*	IINFYDPLVFPSDEFDASISQVNEKI NQSLAFIRK
AVP0128	T-105	*RSV fusion (F) protein*	INFYDPLVFPSDEFDASISQVNEKI NQSLAFIRKS

AVP0129	T-106	*RSV fusion (F) protein*	NFYDPLVFPSDEFDASISQVNEKINQSLAFIRKSD
AVP0130	T-107	*RSV fusion (F) protein*	FYDPLVFPSDEFDASISQVNEKINQSLAFIRKSDE
AVP0131	T-108	*RSV fusion (F) protein*	YDPLVFPSDEFDASISQVNEKINQSLAFIRKSDEL
AVP0132	T-109	*RSV fusion (F) protein*	DPLVFPSDEFDASISQVNEKINQSLAFIRKSDELL
AVP0133	T-110	*RSV fusion (F) protein*	PLVFPSDEFDASISQVNEKINQSLAFIRKSDELLH
AVP0134	T-111	*RSV fusion (F) protein*	LVFPSDEFDASISQVNEKINQSLAFIRKSDELLHN
AVP0135	T-112	*RSV fusion (F) protein*	VFPSDEFDASISQVNEKINQSLAFIRKSDELLHNV
AVP0136	T-113	*RSV fusion (F) protein*	FPSDEFDASISQVNEKINQSLAFIRKSDELLHNVN
AVP0137	T-114	*RSV fusion (F) protein*	PSDEFDASISQVNEKINQSLAFIRKSDELLHNVNA
AVP0138	T-115	*RSV fusion (F) protein*	SDEFDASISQVNEKINQSLAFIRKSDELLHNVNAG
AVP0139	T-116	*RSV fusion (F) protein*	DEFDASISQVNEKINQSLAFIRKSDELLHNVNAGK
AVP0140	T-117	*RSV fusion (F) protein*	EFDASISQVNEKINQSLAFIRKSDELLHNVNAGKS
AVP0141	T-118	*RSV fusion (F) protein*	FDASISQVNEKINQSLAFIRKSDELLHNVNAGKST
AVP0142	T-119	*RSV fusion (F) protein*	DASISQVNEKINQSLAFIRKSDELLHNVNAGKSTT
AVP0143	T-118	*RSV fusion (F) protein*	FDASISQVNEKINQSLAFIRKSDELLHNVNAGKST
AVP0427	Peptide 80 to 94	*RSV Rho-A protein*	ILMCFSIDSPDSLEN
AVP0428	Peptide 83A	*RSV Rho-A protein*	ILMAFSIDSPDSLEN
AVP0429	Peptide 80 to 94-N	*RSV Rho-A protein*	ILMCFSINSPNSLQN
AVP1310	T-142	*RSV fusion (F) protein*	YTSVITIELSNIKENKCNGTDAKVKLIKQELDKYK
AVP1311	T-143	*RSV fusion (F) protein*	TSVITIELSNIKENKCNGTDAKVKLIKQELDKYKN
AVP1312	T-144	*RSV fusion (F) protein*	SVITIELSNIKENKCNGTDAKVKLIKQELDKYKNA
AVP1313	T-145	*RSV fusion (F) protein*	VITIELSNIKENKCNGTDAKVKLIKQELDKYKNAV
AVP1314	T-146	*RSV fusion (F) protein*	ITIELSNIKENKCNGTDAKVKLIKQELDKYKNAVT
AVP1315	T-147	*RSV fusion (F) protein*	TIELSNIKENKCNGTDAKVKLIKQELDKYKNAVTE

AVP1316	T-148	*RSV fusion (F) protein*	IELSNIKENKCNGTDAKVKLIKQELDKYKNAVTEL
AVP1317	T-149	*RSV fusion (F) protein*	ELSNIKENKCNGTDAKVKLIKQELDKYKNAVTELQ
AVP1318	T-150	*RSV fusion (F) protein*	LSNIKENKCNGTDAKVKLIKQELDKYKNAVTELQL
AVP1319	T-151	*RSV fusion (F) protein*	SNIKENKCNGTDAKVKLIKQELDKYKNAVTELQLL
AVP1320	T-152	*RSV fusion (F) protein*	NIKENKCNGTDAKVKLIKQELDKYKNAVTELQLLM
AVP1321	T-153	*RSV fusion (F) protein*	IKENKCNGTDAKVKLIKQELDKYKNAVTELQLLMQ
AVP1322	T-154	*RSV fusion (F) protein*	KENKCNGTDAKVKLIKQELDKYKNAVTELQLLMQS
AVP1323	T-155	*RSV fusion (F) protein*	ENKCNGTDAKVKLIKQELDKYKNAVTELQLLMQST
AVP1324	T-104	*RSV fusion (F) protein*	IINFYDPLVFPSDEFDASISQVNEKINQSLAFIRK
AVP1325	T-105	*RSV fusion (F) protein*	INFYDPLVFPSDEFDASISQVNEKINQSLAFIRKS
AVP1326	T-106	*RSV fusion (F) protein*	NFYDPLVFPSDEFDASISQVNEKINQSLAFIRKSD
AVP1327	T-107	*RSV fusion (F) protein*	FYDPLVFPSDEFDASISQVNEKINQSLAFIRKSDE
AVP1328	T-108	*RSV fusion (F) protein*	YDPLVFPSDEFDASISQVNEKINQSLAFIRKSDEL
AVP1329	T-109	*RSV fusion (F) protein*	DPLVFPSDEFDASISQVNEKINQSLAFIRKSDELL
AVP1330	T-110	*RSV fusion (F) protein*	PLVFPSDEFDASISQVNEKINQSLAFIRKSDELLH
AVP1331	T-111	*RSV fusion (F) protein*	LVFPSDEFDASISQVNEKINQSLAFIRKSDELLHN
AVP1332	T-112	*RSV fusion (F) protein*	VFPSDEFDASISQVNEKINQSLAFIRKSDELLHNV
AVP1333	T-113	*RSV fusion (F) protein*	FPSDEFDASISQVNEKINQSLAFIRKSDELLHNVN
AVP1334	T-114	*RSV fusion (F) protein*	PSDEFDASISQVNEKINQSLAFIRKSDELLHNVNA
AVP1335	T-115	*RSV fusion (F) protein*	SDEFDASISQVNEKINQSLAFIRKSDELLHNVNAG
AVP1336	T-116	*RSV fusion (F) protein*	DEFDASISQVNEKINQSLAFIRKSDELLHNVNAGK
AVP1337	T-117	*RSV fusion (F) protein*	EFDASISQVNEKINQSLAFIRKSDELLHNVNAGKS
AVP1338	T-118	*RSV fusion (F) protein*	FDASISQVNEKINQSLAFIRKSDELLHNVNAGKSI
AVP1924	77-95	*Rho-A protein*	TDVILMCFSIDSPDSLENI
AVP1925	77-95	*Rho-A protein*	CSIELSDIPLSVDFNTMID

AVP1926	77-95-77A	*Rho-A protein*	ADVILMCFSIDSPDSLENI
AVP1927	77-95-78A	*Rho-A protein*	TAVILMCFSIDSPDSLENI
AVP1928	77-95-79A	*Rho-A protein*	TDAILMCFSIDSPDSLENI
AVP1929	77-95-80A	*Rho-A protein*	TDVALMCFSIDSPDSLENI
AVP1930	77-95-81A	*Rho-A protein*	TDVIAMCFSIDSPDSLENI
AVP1931	77-95-82A	*Rho-A protein*	TDVILACFSIDSPDSLENI
AVP1932	77-95-83A	*Rho-A protein*	TDVILMAFSIDSPDSLENI
AVP1933	77-95-84A	*Rho-A protein*	TDVILMCASIDSPDSLENI
AVP1934	77-95-85A	*Rho-A protein*	TDVILMCFAIDSPDSLENI
AVP1935	77-95-86A	*Rho-A protein*	TDVILMCFSADSPDSLENI
AVP1936	77-95-87A	*Rho-A protein*	TDVILMCFSIASPDSLENI
AVP1937	77-95-88A	*Rho-A protein*	TDVILMCFSIDAPDSLENI
AVP1938	77-95-89A	*Rho-A protein*	TDVILMCFSIDSADSLENI
AVP1939	77-95-90A	*Rho-A protein*	TDVILMCFSIDSPASLENI
AVP1940	77-95-91A	*Rho-A protein*	TDVILMCFSIDSPDALENI
AVP1941	77-95-92A	*Rho-A protein*	TDVILMCFSIDSPDSAENI
AVP1942	77-95-93A	*Rho-A protein*	TDVILMCFSIDSPDSLANI
AVP1943	77-95-94A	*Rho-A protein*	TDVILMCFSIDSPDSLEAI
AVP1944	77-95-95A	*Rho-A protein*	TDVILMCFSIDSPDSLENA
AVP1945	77-86	*Rho-A protein*	TDVILMCFSI
AVP1946	77-89	*Rho-A protein*	TDVILMCFSIDSP
AVP1947	77-92	*Rho-A protein*	TDVILMCFSIDSPDSL
AVP1948	78-95	*Rho-A protein*	DVILMCFSIDSPDSLENI
AVP1949	79-95	*Rho-A protein*	VILMCFSIDSPDSLENI
AVP1950	80-95	*Rho-A protein*	ILMCFSIDSPDSLENI
AVP1951	83-95	*Rho-A protein*	CFSIDSPDSLENI
AVP1952	80-94	*Rho-A protein*	ILMCFSIDSPDSLEN
AVP1953	80-93	*Rho-A protein*	ILMCFSIDSPDSLE
AVP1954	80-92	*Rho-A protein*	ILMCFSIDSPDSL
AVP1955	80-91	*Rho-A protein*	ILMCFSIDSPDS
AVP1956	80-90	*Rho-A protein*	ILMCFSIDSPD
AVP1957	80-89	*Rho-A protein*	ILMCFSIDSP
AVP1958	80-88	*Rho-A protein*	ILMCFSIDS
AVP1959	80-87	*Rho-A protein*	ILMCFSID
AVP1960	80-86	*Rho-A protein*	ILMCFSI
AVP1961	80-85	*Rho-A protein*	ILMCFS
AVP1971	HR1-30a	*RSV fusion (F) protein*	AVSKVLHLEGEVNKISALLSTNK AVVSLSNGVSVLTSKVLDLDNYI DKQLLPIVNK
AVP1972	HR2-30a	*RSV fusion (F) protein*	NFYDPLVFPSDEFDASISQVNEKIN QSLASIRKSDELLHNVNAGK
AVP2000	G149-197	*RSV attachment glycoprotein*	KQRQNKPPSKPNNDFHFEVFNFV PCSICSNNPTCWAICKRIPNKKPG KK
AVP2001	G163-197	*RSV attachment glycoprotein*	FHFEVFNFVPCSICSNNPTCWAIC KRIPNKKPGKK

AVP2002	G171-197	*RSV attachment glycoprotein*	VPCSICSNNPTCWAICKRIPNKKPGKK
AVP2003	G173-197	*RSV attachment glycoprotein*	CSICSNNPTCWAICKRIPNKKPGKK
AVP2004	G149-189	*RSV attachment glycoprotein*	KQRQNKPPSKPNNDFHFEVFNFVPCSICSNNPTCWAICKRI
AVP2005	G154-189	*RSV attachment glycoprotein*	KPPSKPNNDFHFEVFNFVPCSICSNNPTCWAICKRI
AVP2006	G158-189	*RSV attachment glycoprotein*	KPNNDFHFEVFNFVPCSICSNNPTCWAICKRI
AVP2007	G149-169	*RSV attachment glycoprotein*	KQRQNKPPSKPNNDFHFEVFN
AVP2008	G149-165	*RSV attachment glycoprotein*	KQRQNKPPSKPNNDFHF
AVP2009	G154-172	*RSV attachment glycoprotein*	KPPSKPNNDFHFEVFNFVP
AVP2010	G154-171	*RSV attachment glycoprotein*	KPPSKPNNDFHFEVFNFV
AVP2011	G154-170	*RSV attachment glycoprotein*	KPPSKPNNDFHFEVFNF
AVP2012	G154-169	*RSV attachment glycoprotein*	KPPSKPNNDFHFEVFN

Table A6: Anti-*Streptococcus pneumoniae* AMPs

ID	AMP	SOURCE	SEQUENCE
CAMPSQ3440	Jcpep7	*Jatropha curcas*	KVFLGLK
CAMPSQ2785	Gramicidin A	*Bacillus brevis*	VGALAVVVWLWLWLW
CAMPSQ278	Buforin-2	*Bufo gargarizans [Asian toad]*	TRSSRAGLQFPVGRVHRLLRK
CAMPSQ393	Misgurin	*Misgurnus anguillicaudatus [Oriental weatherfish]*	RQRVEELSKFSKKGAAARRRK

CAMPSQ3373	Gallidermin	*Staphylococcus gallinarum (F16/P57) Tu3928, Staphylococcus cohnii T*	IASKFLCTPGCAKTGSFNSYCC
CAMPSQ3300	Lichenicidin	*Bacillus licheniformis ATCC 14580 or VK 21*	TITLSTCAILSKPLGNNGYLCTVTKECMPSCN
CAMPSQ3300	Lichenicidin	*Bacillus licheniformis ATCC 14580 or VK 21*	TITLSTCAILSKPLGNNGYLCTVTKECMPSCN
CAMPSQ693	Beta defensin 1	*Chinchilla lanigera*	QRYFCRVRGGRCAALTCLPRETQIGRCSVKGRKCCR
CAMPSQ277	Buforin-1	*Bufo gargarizans*	AGRGKQGGKVRAKAKTRSSRAGLQFPVGRVHRLLRKGNY
CAMPSQ3418	Beta-amyloid peptide(1-40)	*Homo sapiens*	DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVV
CAMPSQ942	Plectasin	*Pseudoplectania nigrella*	GFGCNGPWDEDDMQCHNHCKSIKGYKGGYCAKGGFVCKCY
CAMPSQ3419	Beta-amyloid peptide(1-42)	*Homo sapiens*	DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIA
CAMPSQ193	Royalisin precursor	*Apis mellifera*	VTCDLLSFKGQVNDSACAANCLSLGKAGGHCEKVGCICRKTSFKDLWDKYF
CAMPSQ3091	Brochocin C	*Brochothrix campestris ATCC 43754*	YSSKDCLKDIGKGIGAGTVAGAAGGGLAAGLGAIPGAFVGAHFGVIGGSAACIGGLLGN
CAMPSQ3726	Human angiogenin	*Homo sapiens*	QDNSRYTHFLTQHYDAKPQGRDDRYCESIMRRRGPTSPCKDINTFIHGNKRSIKAICENK NGNPHRENLRISKSSFQVTTCKLHGGSPWPPCQYRATAGFRNVVVACENGLPVHLDQSIF RRPRP
AP00549	Plectasin (fungi, fungal defensin; UCSS1a; 3S=S, fungii; BBW; Derivatives: NZ2114)	*Pseudoplectania nigrella*	GFGCNGPWDEDDMQCHNHCKSIKGYKGGYCAKGGFVCKCY
BAC171	Thiocillin	*Bacillus cereus*	MSEIKKALNT LEIEDFDAIE MVDVDAMPEN EALEIMGASC TTCVCTCSCC

TT

Appendix B

Supplementary Tables for chapter 4: Pneumonia Receptor Proteins from NCBI

Table B1: >SCY38300.1 Iron-regulated outer membrane proteins [*Acinetobacter baumannii*]

MHFSLRSRHLRPSLLASSLLLAVSAQGQEKLDVDLPAAPLGQAINALAQQSSVQI
LFAGDLGAGRQAPALKGRFTPEEALRQLLRDSGLKAQARDEHTFIVVPASEAAV
PATQARSEPLDMEQMEITASRTSSDLVSATRQSTVIEHAQLEELRQGSDSLATVL
AKAVPGMSDSSRTITEYGQTLRGRSMLVMVDGVPLNTNRDSSRNLANIDPALIE
RIEVIRGSSAIYGSGATGGIISITTRPAGGENRAETRLSATSSLTRLGSDGLGGQFQ
QYFAGSLGALDYSFDFGTRHVGASYDAHGDRIAPEPSQGDLFDSNVYNIGGKL
GLRIDENQRVQLALSHYDARQDTDYATDPRVARLPPGSVPANAIKGLELDEQNRI
RNTLANLEYENLDILGSRLSAQLYYRDYFTRFTPFDARAVSTRGGNVDQIMQNS
EVFGSRLTLRTPLGESGNTELVWGGDYNQERSDMPLDVFDPAAYDASGGLVFD
KIGKLTYMPPLRTRSAGAFAQLQHRFDEHWSIDGGLRYEYSTAEFDDFVPLSESK
AASPVTVKGGDLDYDAVLSNLGIVYSPVAGQEIYASFSQGFQLPDVGIQLRNAR
RGFDIGSSNLEPVKTNNYELGWRGAIGGNTLGSLALFYTTSKLGDVQSFNNGLI
LTRTKERIYGVEASADWLSDDEVWGAGGSATWMRGREKPDGKDWQDMTGYR
VPPLKLTAYLQYKPDADWNNRLQATFFDSKDYRLDGVESFGRRQVSTYTTVDL
VSQYRITPDDQLSLGIQNLFNRDYYPLYSQLLRNNNNTSHLPAPGTVLTASYTHN
W

Table B2: >AIT02889.1 Iron-regulated outer membrane protein [*Klebsiella pneumoniae]*

MKKRLWVLHPLLLASTLPALAAQSDEDSIIVSANRTHRTVAEMAQTTWVIEGQE
IEQQVQGGKEFKDVLAQLIPGIDVSSQGRTNYGMNMRGRAIVVLIDGVRLNSSR
TDSRQLDAIDPFNIEHIEVISGATSLYGGGSTGGLINIVTKKGQQDRQVDIEVGSK
SGFANSNDHDERVAAAVSGGTDHASGRLSVAYQRFGGWYDGNNDALILDNTQT
GLQHSDRLDVMGTGTIEIDNNRQLQLVTQYYKSQGDDDYGLWLGKNMSAVTS
GGKAYTTDGLNSDRIPGTERHLISLQYSDADFFGQNLVSQVYYRDESLTFYPFPT
LTKGQVSSFSSSQQDTDQYGAKLTLNSQPLAGWDLTWGLDADHETFNANQMF
FDLPQSMASGGLHNESIYTTGRYPGYSISNVAPFLQSSYDLNDIFTVSGGVRYQW
TENRVDDFVGYAQQQDIANGKARSADAIKGGKTDYDNFLFNAGIVAHLTERQQ
TWFNFSQGVELPDPGKYYGIGKYGAAVNGHLPLISSVNVDDSPLQGIKVNSYEL
GWRYTGDNLRTQLAAYYSTSDKTIVVNRTDMTIDVQSDKRRIYGVEGAVDYFIP
DSDWSVGGNFNVLKSQVQTDGRWQKWDVTLASPSKATAWVGWAPDPWSLRV
QSQQVFDLSDAAGNKLEGYNTVDFIGSYALPVGKLTFSIENLLNEDYVTIWGQR
APLLYSPTYGSSSLYEYKGRGRTFGLNYALTF

Table B3: >AJS15225.1 pneumolysin [*Streptococcus pneumoniae*]
MANKAVNDFILAMNYDKKKLLTHQGESIENRFIKEGNQLPDEFVVIERKKRSLS
TNTSDISVTATNDSRLYPGALLVVDETLLENNPTLLAVDRAPMTYSIDLPGLASS
DSFLQVEDPSNSSVRGAVNDLLAKWHQDYGQVNNVPARMQYEKITAHSMEQL
KVKFGSDFEKTGNSLDIDFNSVHSGEKQIQIVNFKQIYYTVSVDAVKNPGDVFQ
DTVTVEDLKQRGISAERPLVYISSVAYGRQVYLKLETTSKSDEVEAAFEALIKGV
KVAPQTEWKQILDNTEVKAVILGGDPSSGARVVTGKVDMVEDLIQEGSRFTAD
HPGLPISYTTSFLRDNVVATFQNSTDYVETKVTAYRNGDLLLDHSGAYVAQYYIT
WNELSYDHQGKEVLTPKAWDRNGQDLTAHFTTSIPLKGNVRNLSVKIRECTGL
AWEWWRTVYEKTDLPLVRKRTISIWGTTLYPQVEDKVEND

Table B4: >YP_308854.1 membrane protein M1 [*Influenza A* virus (A/Korea/426/1968(H2N2))]
MSLLTEVETYVLSIVPSGPLKAEIAQRLEDVFAGKNTDLEALMEWLKTRPILSPL
TKGILGFVFTLTVPSERGLQRRRFVQNALNGNGDPNNMDRAVKLYRKLKREITF
HGAKEVALSYSAGALASCMGLIYNRMGAVTTEVAFAVVCATCEQIADSQHRSH
RQMVTTTNPLIRHENRMVLASTTAKAMEQMAGSSEQAAEAMEVASQARQMV
QAMRAIGTPPSSSAGLKDDLLENLQAYQKRMGVQMQRFK

Table B5: >pdb|3ZDP|C Chain C, R416a Monomeric Nucleoprotein of *Influenza A* Virus
MATKGTKRSYEQMETDGERQNATEIRASVGKMIDGIGRFYIQMCTELKLSDYE
GRLIQNSLTIERMVLSAFDERRNKYLEEHPSAGKDPKKTGGPIYRRVDGKWRRE
LILYDKEEIRRIWRQANNGDDATAGLTHMMIWHSNLNDATYQRTRALVRTGMD
PRMCSLMQGSTLPRRSGAAGAAVKGVGTMVMELIRMIKRGINDRNFWRGENG
RRTRIAYERMCNILKGKFQTAAQRTMVDQVRESRNPGNAEFEDLIFLARSALILR
GSVAHKSCLPACVYGSAVASGYDFEREGYSLVGIDPFRLLQNSQVYSLIRPNENP
AHKSQLVWMACHSAAFEDLRVSSFIRGTKVVPRGKLSTRGVQIASNENMETME
SSTLELRSRYWAIRTRSGGNTNQQRASSGQISIQPTFSVQANLPFDRPTIMAAFTG
NTEGRTSDMRTEIIRLMESARPEDVSFQGRGVFELSDEKATSPIVPSFDMSNEGS
YFFGDNAEEYDN

Table B6: >AAB72046.1 nucleoprotein [*Influenza B* virus]
MSNMDIDGINTGTIDKTPEEITSGTSGTTRPIIRPATLAPPSNKRTRNPSPERATTSS
EADVGRKTQKKQTPTEIKKSVYNMVVKLGEFYNQMMVKAGLNDDMERNLIQ
NAHAVERILLAATDDKKTEFQRKKNARDVKEGKEEIDHNKTGGTFYKMVRDD
KTIYFSPIRITFLKEEVKTMYKTTMGSDGFSGLNHIMIGHSQMNDVCFQRSKAL
KRVGLDPSLISTFAGSTLPRRSGATGVAIKGGGTLVAEAIRFIGRAMADRGLLRDI
KAKTAYEKILLNLKNKCSAPQQKALVDQVIGSRNPGIADIEDLTLLARSMVVVR
PSVASKVVLPISIYAKIPQLGFNVEEYSMVGYEAMALYNMATPVSILRMGDDAK
DKSQLFFMSCFGAAYEDLRVLSALTGIEFKPRSALKCKGFHVPAKEQVEGMGAA
LMSIKLQFWAPMTRSGGNEVGGDGGSGQISCSPVFAVERPIALSKQAVRRMLSM
NIEGRDADVKGNLLKMMNDSMAKKTNGNAFIGKKMFQISDKNKTNPVEIPIKQ
TIPNFFFGRDTAEDYDDLDY

Table B7: >pdb|4D4T|A Chain A, Respiratory syncytial virus Matrix Protein
EFMETYVNKLHEGSTYTAAVQYNVLEKDDDPASLTIWVPMFQSSMPADLLIKEL
ANVNILVKQISTPKGPSLRVMINSRSAVLAQMPSKFTICANVSLDERSKLAYDVT

TPCEIKACSLTCLKSKNMLTTVKDLTMKTLNPTHDIIALCEFENIVTSKKVIIPTYL
RSISVRNKDLNTLENITTTEFKNAITNAKIIPYSGLLLVITVTDNKGAFKYIKPQSQ
FIVDLGAYLEKESIYYVTTNWKHTATRFAIKPMED

Table B8: >pdb|1G2C|X Chain X, Human *Respiratory Syncytial* Virus Fusion Protein Core
FYDPLVFPSDEFDASISQVNEKINQSLAFIRKSDELLHNVNAG