The discovery and characterisation of novel antimicrobial peptides against bacterial pathogens with porcine host specificity.

Britany S. Clarke

Student ID: 4303978

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

Supervisors: Professor Kevin Gough, Dr Rob Atterbury, Dr Ben Maddison

School of Veterinary Medicine and Science, 2022

1.1 Abstract

Natural antimicrobial peptides (AMPs) are crucial host defense peptides with multifactorial biological activity and AMPs reinforce the first-line innate immune response to safeguard hosts from invading microbes and viruses. Focal aims of this study concerned constructing a discovery pipeline to identify novel AMPs against bacterial pathogens with porcine host specificity. Outlined herein the AMP discovery pipeline was structured to screen large and diverse peptide libraries via two parallel screening approaches; *i)* Screening libraries $\geq 10^9$ in size via phage display technology (pIII system, 3+3) to identify enriched binder ligands. Alternatively, explored was the exploitation of *ii)* an *Escherichia coli* recombinant peptide expression soft agar-based assay; for direct antimicrobial characterisation of peptide sub-libraries 10^6 to 10^7 in size. *E. coli* JE5505 (*F− 1po his proA argE thi gal lac xyl mtl tsx*) is a lipoprotein *Δlpp*-254 mutant and the overlay centres on recombinant peptides leaking from the periplasm into surrounding media generating visible zones of inhibition around colonies. A 0.7% Mueller-Hinton *E. coli* JE5505 soft agar overlay assay was optimised for both the scanning of known AMPs and screening randomised peptide library. The Mueller-Hinton leaky *E. coli* JE5505 overlay assay approach offered high-throughput ($\leq 10^4$ library colonies per assay) and cost-effective screening. However, the overlays were incompatible with screening randomised libraries as no antimicrobial activity was observed when screening AMPs which were not class-II bacteriocin Plantaricin-423 (Pln-423) or respective improved Pln-423 mutants obtained from literature.

Screening methods were optimised alongside designing a degenerate primer strategy to bias amino acid distribution in a manner which increased the prevalence of peptide library candidates with antibacterial phenotypes. A training dataset of 187

natural AMPs against pig bacterial pathogens were collected from the ADP3 database and used as the baseline for amino acid distribution analysis (APD3_ AMP_PigPathogenDataset). The degenerate primer strategies utilised to construct two 16 mer-peptide phage libraries herein were; *i)* the NNK randomisation scheme; "N": A/T/G/C and "K": G/T, a highly exploited conventional peptide degenerate mutagenesis strategy. Compared to *ii)* the $VNN_{15}+(TTT)_1$ randomisation, "V" = A/G/C, and "TTT" reintroduces VNN scheme excluded phenylalanine residue in one out of sixteen peptide region amino acid residues. $VNN_{15}+(TTT)_1$ randomisation "AMP-biased approach" was designed with the rationale of biasing the amino acid distribution towards that observed in the APD3_ AMP_PigPathogenDataset. Whole-construct inverse PCR was exploited for library construction wherein degenerate primers introduced the $VNN_{15}+(TTT)_1$ or NNK randomised peptide region diversity into *pSD3* phagemids. Additionally, with ease peptide phage libraries were *SpeI* digested into sub-libraries (absent of phage -*gIII*) for screening in the 0.7% Mueller-Hinton *E. coli* JE5505 soft agar overlay assay. Three categorised AMP amino acids groupings were outlined herein; (a) hydrophobic, (b) basic (cationic), (c) acidic (anionic), or neutrally charged and/or hydrophilic and/or stop codons. Ion Torrent Next Generation Sequencing (NGS) analysis quality control suggested $VNN_{15}+(TTT)_1$ surpassed NNK equivalent libraries, and more closely resembled the overall AMP amino acid groupings of the APD3_ AMP_PigPathogenDataset.

A whole-cell phage display approach was then used to screen the $VNN_{15}+(TTT)_1$ and NNK phage propagations against bacterial pathogens with porcine host specificity; *Streptococcus suis* P1/7 and *Salmonella enterica subsp. enterica serovar Typhimurium 4/74*. The panning strategy entailed three rounds, and in the final panning step all round two output phage were screened against both test species

generating 8 panning datasets with 10 replicates. Round three panning outputs were rescued and then sequenced via Ion Torrent NGS to elucidate $>10^6$ library sequences and identify enriched peptide ligands against bacteria targets. TopN and Z score bioinformatic analysis were utilised (Z score: ≥2 or Top50 frequency analysis) and peptides were shortlisted on the presence across ten panning sample replicates, with positive dataset reproducibility of ≥90%, 9/10 panning replicates preferred. Eighty-five peptides were crude synthesised (C-terminal amidation), with the majority being derived from $VNN_{15}(TTT)_1$ library. As in most panning conditions $VNN_{15}(TTT)_1$ library scheme generated a greater number of enriched binders against panning test species. Unfortunately, *in vitro* microdilution minimum inhibitory concentration (MIC) assays screening revealed <6% of shortlisted candidates demonstrated inhibitory antimicrobial activity against bacteria pig pathogen strains; *S. suis* P1/7, *S. Typhimurium* 4/74 or *E. coli* P433. Antimicrobial peptides identified were weakly active 50 - 200 µM and the most potent AMPs herein were NNK derived. This unfortunately countered the inference of superiority gained from NGS analysis of the $VNN_{15}(TTT)_1$ amino acid distribution and enriched panning ligand pools.

Nonetheless, $VNN_{15}(TTT)_1$ phage display derived peptides; Pep_VNN/43 (MIC =100µM) and Pep_VNN/55 (MIC = >100µM) represent novelty in two-aspects; (*i)* novel application of the degenerate VNN scheme to identify phage display peptide ligands with antibacterial phenotype, and *ii)* both peptides inhibit *S. suis P1/7 in vitro.* Pep_NNK/23 additionally inhibited *S. suis PI/7,* and in both instance the discovery of $VNN_{15}+(TTT)_1$ and NNK phage display derived peptides extends the current knowledge of novel AMP sequences against *S. suis*. Enriched peptides panned against both *S. Typhimurium* 4/74 and *S. suis* P1/7 were able to demonstrate more potent MICs against non-panned test species; *E. coli* P433. For example, broad-spectrum inhibiting

Pep_NNK/17 and *E. coli* P433 biased Pep_NNK/21 demonstrated the most potent inhibitory activity (MIC = 50µM). Pep_NNK/21 was unable to demonstrate antibacterial activity against the two panning species at 200µM and 800µM, *x4* and *x*16 MIC of *E. coli* P433. Essentially, both Pep_NNK/17 and NNK/21 further demonstrates that in some instances; phage display technologies likely select for peptide ligands with weak antibacterial action but sufficient binding affinity to be enriched against the test species panned.

Pep_VNN/43, Pep_NNK/17 and Pep_NNK/21 exemplify novel antimicrobial peptides with activity against key bacterial pathogens with porcine host specificity. Pig livestock production is highly reliant on the use of antimicrobials to control and prevent the dissemination of disease. *S. typhimurium, S. suis and E. coli* are amongst the most prevalent bacterial pathogens associated with economically impactful porcine diseases and certain strains can be multidrug resistant. Antimicrobial resistance continues to challenge both animal and human health prospects, and the ever-increasing exhaustion of treatment options is ominously encroaching further towards last-resort antibiotics. Consequently, novel antimicrobial therapeutics with divergent antibacterial mechanisms of action, membrane or intracellular biological targets are required to supplant antibiotic reliance. Presently, AMPs are an underutilised class of antimicrobials however, due to their therapeutic potential AMPs are increasingly garnering the spotlight for substituting antibiotics.

1.2 Acknowledgements

Firstly, I would like to thank Professor Kevin Gough, Dr Rob Atterbury, Dr Ben Maddison, the broader Gough research and ADAS team for their guidance.

There are various loved ones I would like to dedicate this thesis to;

Firstly, my mother; ***Faith,*** you are my inspiration and the reason I fight. As you say, education is power. Thank you for fostering my learning heart.

Countless friends of new and old who I am equally indebted to;

Maria Tsoumpeli – Your intellect, passion and dedication motivated me beyond belief. We will cycle into the sunset together my friend.

Katie East – A beautiful soul. My guiding moon throughout and may we always continue to foster each other's growth and success.

Akshay Kumar – Thank you for your endless kindness and support, my closet childhood friend. I will forever cherish you.

Isabel Foo, Parichehr Zoharbi-Redwood & Claudia Leong – Thank you for your endless wisdom, love and reminding me family is a word for those who cultivate and see the best in each other.

Kimberley Slinger, Jayasree Dey, Rachel Bates, Anthony Chiu, Louis Burrows, Jess Kyle, Jared Porter, Dimitar Ivanov, Christopher Waugh, dearly-loved friends. I've finished the story the place ours began. Good *ol'* Sutton Bonington, a place that has forever changed my life.

Finally, in dedication to Mr Campbell and Mr Sumner two teachers who saw my iconoclast nature and cultivated it into an unquenchable passion for science. Guess what Sir! I did it and you were absolutely right, science was my path……

Margaret Cavendish – 17th century English philosopher poet and scientist captures the ceaseless ways our reality is underpinned by mindboggling beauty, even down to the most mundane. The opportunity to understand the world beyond what we see is the greatest gift science bestows.

Of Many Worlds in This World

Just like as in a nest of boxes round,
Degrees of sizes in each box are found:
So, in this world, may many others be
Thinner and less, and less still by degree:
Although they are not subject to our sense,
A world may be no bigger than two-pence.
Nature is curious, and such works may shape,
Which our dull senses easily escape:
For creatures, small as atoms, may there be,
If every one a creature's figure bear.
If atoms four, a world can make, then see
What several worlds might in an ear-ring be:
For, millions of those atoms may be in
The head of one small, little, single pin.
And if thus small, then ladies may well wear
A world of worlds, as pendents in each ear.

Margaret Cavendish

TABLE OF CONTENTS

1.1 Abstract **1**

1.2 Acknowledgements **5**

1.3 List of Tables **20**

1.4 List of Figures **22**

1.5 Key Abbreviations **25**

Chapter 1: Introduction **26**

1.1 GLOBAL PIG PRODUCTION, FOOD SECURITY AND ANTIMICROBIALS **27**

1.1.1 One Health: food security and pig production 27

1.1.2 Pig farming and antimicrobials 28

1.1.3 Pig production and antibiotic usage 29

1.1.4 Antibiotics utilised in pig farming. 31

1.1.5 Antibiotic resistance mechanisms 33

1.2 DISEASES IN PIGS **35**

1.2.1 Global pig markets and disease trends 35

1.2.2 Gastrointestinal and respiratory bacterial pathogens in pigs 36

1.3 BACTERIA PATHOGEN WITH PORCINE HOST SPECIFICITY: *SALMONELLA* **42**

1.3.1 Salmonella classification 42

1.3.2 Salmonella: Zoonotic potential 43

1.3.3 Salmonella in pigs and the epidemiology, prevalence and impact of pig-related serovars 44

1.3.4 Clinical signs of Salmonella infection in pigs 45

1.3.5 Salmonella Typhimurium antimicrobial resistance profiles in pigs 47

1.3.6 Salmonella virulence factors 48

1.4 BACTERIA PATHOGEN WITH PORCINE HOST SPECIFICITY: *STREPTOCOCCUS SUIS* 50

1.4.1 S. suis epidemiology, prevalence and impact of pathogenic strains in pigs. 50

1.4.2 S. suis in pig farm environment and clinical signs of disease in pigs 52

1.4.3 S. suis pathogenicity and virulence factors 53

1.4.4 S. suis; treatment, emerging resistance and zoonotic potential, ... 55

1.5 NON-ANTIBIOTIC ALTERNATIVES IN PIG FARMING: CONTROLLING *SALMONELLA* AND *S. SUIS* 56

1.6 ANTIMICROBIAL PEPTIDES 58

1.6.1 Introduction to AMPs: generalised characteristics, source origins and classification 58

1.6.2 Classification of AMPS by structure 60

1.6.3 Biological functions of AMPs: antimicrobial activity immune modulating properties 61

1.6.4 AMP mechanisms of action 62

1.6.5 Clinical applications of AMPs 64

1.7 PEPTIDE LIBRARY CONSTRUCTION 65

1.7.1 Established peptide library mutagenesis strategies 65

1.7.2 Degenerate primers and codon schemes 66

1.7.3 Trinucleotide phosphoramidite mutagenesis (TRIM) technology 69

1.8 SCREENING STRATERGIES FOR NOVEL ANTIMICROBIAL PEPTIDE DISCOVERY 70

1.8.1 Phage display technology: Structure, genetics and biology of filamentous phage 70

1.8.2 Phage display technology: filamentous phage systems 73

1.8.3 Peptide phage display libraries 74

1.9 PHAGE DISPLAY DISCOVERY OF ANTIMICROBIAL PEPTIDES AGAINST BACTERIAL PATHOGENS 75

1.9.1 Whole-cell phage display platforms for AMP discovery 75

1.9.2 STAMPs & Phage display: Species specific AMP tethering systems 80

1.9.3 AMP Screening: phage display and magnetic bead technology 81

1.9.4 Soft-agar overlay assays for AMP scanning studies 83

1.9.5 SPOT-synthesis AMP peptide assays 85

1.10 HYPOTHESIS AND AIMS 87

1.11 Hypothesis: 87

1.12 Aims 88

Chapter 2: Materials and Methods 89

2.1 BACTERIOLOGICAL MEDIA USED AND MATERIALS THEREOF USED FOR THE CULTIVATION OF BACTERIA TEST STRAINS AND LIBRARY SCREENING APPROACHES 90

2.1.1 Nutrient Broth media 90

2.1.2 MacConkey Agar (MAC) 90

2.1.3 Brilliant Green Agar (BGA) 90

2.1.4 Mueller Hinton Broth (MHB) and Mueller Hinton Agar (MHA) 90

2.1.5 2× Yeast Extract Tryptone (2xYT) and Luria-Bertani Broth media..... 91

2.1.6 Tryptone Soya Broth (TSB) and Tryptone Soya Broth Agar (TSA) 91

2.1.7 de Man, Rogosa and Sharpe (MRS) medium 91

2.1.8 Selective media for Listeria, Salmonella and Streptococcus suis. 92

2.1.9 Soft Agar Media..... 92

2.1.10 Antibiotics for supplemented media 93

2.1.11 25% Glucose 94

2.1.12 Phage Display: M9 minimal growth media reagents 94

2.1.13 Phage Display: M9 minimal growth media preparation 95

2.2 BACTERIAL TEST STRAINS AND PHAGEMID CLONING VECTOR MATERIALS 95

2.2.1 Bacterial test strains 95

2.2.2 Glycerol Stocks 98

2.2.3 pSD3 Phagemid Construct 98

2.3 PEPTIDES, BUFFERS, REAGENTS, KITS AND MATERIALS THEREOF USED FOR ANTIMICROBIAL ASSAYS, LIBRARY CONSTRUCTION AND SCREENING APPROACHES 100

2.3.1 Crude peptides 100

2.3.2 Tetrazolium red 100

2.3.3 Isopropyl β-D-1-thiogalactopyranoside (IPTG) 100

2.3.4 Tris-Acetate-EDTA buffer 100

2.3.5 Phosphate buffered saline (PBS) x 1 100

2.3.6 Phage Display: Polyethylene glycol (PEG) precipitation buffer 100

2.3.7 Phage Display: Glycine-HCL 100

2.3.8 Phage Display: Tris-HCL 100

2.3.9 Macherey-Nagel^{Tm} NucleoSpinTm Gel and PCR Clean-up Kit 100

2.3.10 E.Z.N.A.® M13 DNA Mini Kit 100

2.3.11 The QIAprep Spin Miniprep Kit 100

2.3.12 Qubit™ 100

2.3.13 Agencourt AMPure XP Bead Clean-up kit 100

2.4 METHODS FOR THE ENUMERATION OF BACTERIA TEST STRAINS 100

2.4.1 Bacteriological media preparation 100

2.4.2 Antibiotic stocks and antibiotic supplemented media 100

2.4.3 Enumeration of bacteria 100

2.4.4 Growth curve methodology 101

2.5 METHODS FOR ANTIMICROBIAL CHARACTERISATION AND SCREENING ASSAYS 102

2.5.1 Preparing crude peptides 102

2.5.2 Broth microdilution minimum inhibitory concentration (MIC) assay 104

2.5.3 Qualitative analysis of broth minimum inhibitory concentrations 105

2.5.4 Tetrazolium red colorimetric broth assay 106

2.5.5 Verifying recombinant product from IPTG inductions 107

2.5.6 Agar diffusion assay: Spotting serial dilutions of antimicrobials 107

2.5.7 Plug assay 108

2.5.8 Overlay plug assays 109

2.5.9 Cross-streaking and Co-streaking methods 109

2.5.10 Leaky E. coli JE5505 overlay assay. 110

2.5.11 Periplasm Extraction 112

2.6 GENERALISED METHOD FOR WHOLE-CONSTRUCT INVERSE PCR, BACTERIAL TRANSFORMATION AND SANGER SEQUENCING 114

2.6.1 Inverse PCR and reaction formulation 114

2.6.2 Preparation of agarose gels for electrophoresis 116

2.6.3 Preparation of DNA sample for agarose gel electrophoresis 117

2.6.4 PCR and Gel Extraction Clean-up 117

2.6.5 Quantification of DNA 119

2.6.6 DpnI Digestion 119

2.6.7 T4 Ligase Ligation 119

2.6.8 SpeI restriction enzyme digestion 120

2.6.9 Electrocompetent cells 121

2.6.10 Electroporation 122

2.6.11 Minipreps 122

2.6.12 Sanger sequencing of plasmid constructs 123

2.7 Inverse PCR cloning AMPs into pSD3 vector construct 124

2.7.1 Literature and repository database AMPS cloned into pSD3 phagemid for the optimisation of screening approaches. 124

2.8 METHODS FOR GENERATING A TRAINING SET OF ANTIMICROBIAL PEPTIDES AND IMPLEMENTING *IN-SILICO* ANALYSIS TO EVALUATE AMINO ACID DISTRUBUTION 128

2.8.1 Manually extracting repository AMPs from The Antimicrobial Peptide Database 3, (APD3). 128

2.8.2 Generating control AMP and Non-AMP training datasets 130

2.8.3 In silico tools for analysing antimicrobial peptides; data mining and characterisation of antimicrobial peptides 130

2.8.4 In silico pipeline for analysing antimicrobial peptides; amino acid composition and hydrophobic stretches. 133

2.9 METHODS FOR THE DESIGN OF AN "AMP-BIASED" NOVEL DEGENERATE CODON SCHEME 136

2.10 METHODS FOR 16 MER PEPTIDE LIBRARY CONSTRUCTION. 139

2.10.1 Modifying pSD3 phagemid vector library template 139

2.10.2 $VNN_{15}(TTT)_1$ and NNK Library Construction 141

2.10.3 Library cloning strategy 143

2.10.4 Library Inverse PCR 145

2.10.5 Library electroporation and quality control 146

2.10.6 E. coli JE5505 sub-library production 147

2.10.7 Schematic overview of the library and sub-library restriction enzyme pathway 148

2.11 WHOLE-CELL PHAGE DISPLAY METHOD FOR SCREENING PEPTIDE PHAGE DISPLAY LIBRARIES AGAINST BACTERIAL TARGETS 149

2.11.1 Phage display panning strategy: 149

2.11.2 Phage production 151

2.11.3 Preparation of key reagents for panning 152

2.11.4 Whole-cell phage display panning 152

2.11.5 M13 DNA extraction 153

2.12 NEXT GENERATION ION TORRENT BARCONDING AND IN SILCO ANALYSIS TOOLS 154

2.12.1 Ion torrent NGS primers 154

2.12.2 NGS Sample Preparation: PCR & Clean-ups 156

2.12.3 Extracting datafiles, identifying barcoded sequences and analysing the amino acid composition of 16mer peptides. 158

2.12.4 Summarised overview of panning sampleset analysis in TopN and Z score pipelines. 159

2.12.5 Z score analysis pipeline 161

2.12.6 TopN analysis pipeline 164

Chapter 3: Whole-construct inverse PCR method for the generation of naïve 16mer peptide libraries randomised using degenerate primer $VNN_{15}+(TTT)_1$ "AMP-biased" and "conventional" NNK approaches 166

3.1 Introduction 167

3.2 Aims 1 170

3.2.1 Summary of key research aims for results chapter 1; 170

3.3 Generating and analysing a dataset of natural AMPs with activity against target bacteria pathogens 171

3.3.1 In silico analysis of AMP and non-AMP training datasets 171

3.4 Constructing a degenerate codon scheme mutagenesis strategy to generate a novel "AMP-biased" peptide library. 177

3.5 Whole-plasmid inverse-PCR method for the construction of naïve 16mer peptide libraries randomised using degenerate primer strategies $VNN_{15}+(TTT)_1$ and NNK 181

3.5.1 Library construction: pSD3 phagemid vector library template construction 181

3.5.2 Library construction: Phage display peptide libraries 183

3.6 Sanger and Next Generation Sequencing analysis of naïve $VNN_{15}+(TTT)_1$ and NNK 16mer peptide libraries. 185

3.6.1 E. coli TG1 library sanger sequencing 185

3.6.2 E. coli JE5505 Sub-library sanger sequencing 187

3.6.3 Next generation sequencing of naïve peptide phage libraries 189

3.6.4 Deep sequencing and quality control of naïve peptide libraries 190

3.6.5 Comparing the expected frequency of amino acids to the observed distribution in naïve libraries 193

3.6.6 Positional analysis of amino acid distribution across peptide insert region in naïve peptide libraries compared to natural AMPs..... 196

3.7 Discussion..... 202

3.7.1 In silico analysis of known AMPs for knowledge-based design of AMP biased peptide library. 202

3.7.2 Semi-rational design of an AMP biased peptide library: 205

3.7.3 Selection of degeneration primer randomisation schemes..... 206

3.7.4 Cloning strategy 208

3.7.5 Quality control of naïve peptide libraries: Sanger sequencing 210

3.7.6 Deep sequencing naïve peptide libraries: next generation sequencing library quality control 211

3.7.7 Quality control via next-generation sequence analysis of naïve peptide libraries: amino acid expected frequency vs observed frequency 213

Chapter 4: A leaky periplasm *E. coli* JE5505 expression system and agar-based overlay assay method for screening AMPs 219

4.1 Introduction..... 220

4.2 Result Chapter 2 Aims; 222

4.3 – Optimising a novel leaky periplasm *E. coli* JE5505 expression system within soft agar overlay assays for screening AMPs 223

4.3.1 Microbiological cultivation and phagemid transformation of E. coli JE5505 223

4.3.2 Characterisation of Pln-423wt and mutant variants using a Luria-Bertani (LB) and Tryptone -Soya Broth soft agar E. coli JE5505 overlay assay method 224

4.3.3 Optimisation of a leaky E. coli JE5505 pSD3 expression system in a Mueller-Hinton soft agar overlay assay. 229

4.3.4 Growth kinetics of E. coli JE5505 carrying pSD3 vectors cultured with glucose or IPTG. 233

4.3.5 Evaluating the use of tetrazolium red in broth and agar AMP characterisation assays 240

4.3.6 Overlay assay limitation; E. coli expression vs peptide diffusivity; 242

4.4 Screening degenerately randomised 16-mer peptide libraries in soft agar overlay assay approach for the identification of library AMP candidates ... 247

4.4.1 Screening VNN_{15}+$(TTT)_1$ and NNK randomised 16-mer peptide libraries in 0.7% Mueller-Hinton soft agar overlay assay 247

4.5 Discussion 250

4.5.1 Key considerations made for the optimisation of the Mueller-Hinton soft agar E. coli JE5505 overlay assay; media characteristics and assay format 250

4.5.2 Reproducibility of zones of inhibition in leaky E. coli JE5505 overlay assays; semi-quantitative or purely qualitative? 252

4.5.3 Characterisation of antimicrobial activity of several literature and repository database AMPs in agar-based approaches 255

4.5.4 Class II bacteriocins; case study for their quintessential applicability with E. coli JE5505 expression systems 258

4.5.5 The pitfalls and prizes of recombinant production in E. coli JE5505 expression systems 261

Chapter 5: Discovering and characterising novel AMPs identified from the Next-Generation Sequencing Phage Display screening of $VNN_{15}(TTT)_1$ and NNK 16 mer phage peptide libraries. 264

5.1 Introduction 265

5.2 Aims: 267

5.3 ISOLATION OF NOVEL PHAGE-DISPLAY DERIVED ANTIMICROBIAL PEPTIDES AGAINST BACTERIAL PATHOGENS OF PIGS 268

5.3.1 Whole-cell phage display screening of degenerately randomised phage peptide libraries 268

5.3.2 E. coli JE5505 overlay assay screening of phage display panning outputs 269

5.3.3 Next generation sequencing preparation for panning output samples 269

5.3.4 Phage display next generation sequencing analysis pipeline 1; Frequency analysis TopN 270

5.3.5 Phage display next generation sequencing analysis pipeline 2; Z scores 277

5.3.6 Characterisation for shortlisted enriched binders against key bacteria pathogens of pigs 283

5.4 Discussion..... 293

Chapter 6: Final Discussion..... 298
Chapter 7: Conclusion and future perspectives..... 303
Chapter 8: References..... 307
Chapter 9: Appendix..... 328
9.1 ADP3 database "PigDataset"..... 328

9.2 Next Generation Sequencing Barcode Primers..... 330

9.3 Panning strategy, enlarged schematic diagram..... 334

9.4 Whole-construct PCR cloning method, enlarged schematic..... 335

9.5 Molecular cloning strategy, restriction enzyme digestion; enlarged image 336

1.3 List of Tables

Table Title:	Page
Table 1.1 Phenotypic classification of the primary bacterial pathogens of pigs by genera…….	**37**
Table 1.2 Gram-negative bacterial pathogens of pig digestive and respiratory systems and their propensity of occurrence at varying pig developmental stages…..	**38**
Table 1.3 Gram-positive bacterial pathogens of pig digestive and respiratory systems and their propensity of occurrence at varying pig developmental stages…..	**38**
Table 1.4 *S. enterica* serotypes most frequently associated with disease in key livestock animals and humans…….	**44**
Table 1.5 Classification of AMPs based on structural conformations……	**61**
Table 1.6 Degenerate codon schemes exploited in random mutagenesis and peptide library design……...	**68**
Table 1.7 Antimicrobial peptides discovered by phage display……	**77**
Table 2.1 Formulation of 2xYT and Luria-Bertani (LB) broth culture media…….	**91**
Table 2.2. Gram-Negative and Gram-Positive bacteria strains utilised herein……..	**97**
Table 2.3. *E. coli* strains used for periplasm extractions, overlay assays and phage display library…….	**97**
Table 2.4 Physiochemical properties and sources of AMPs screened in the 0.7% Mueller-Hinton leaky *E. coli* JE5505 overlay assay or minimum inhibition assay (broth microdilution)……..	**103**
Table 2.5 Sanger sequencing primers utilised to map pSD3 phagemid vector…….	**124**
Table 2.6 Inverse PCR primers for Pln-423 and Guralp *et al.*, (2013) improved mutants……	**126**
Table 2.7 Inverse PCR primers for literature or repository database derived AMPs……..	**127**
Table 2.8 Key pathogenic bacteria in pigs prioritised for natural AMP selection from the ADP3 database……..	**129**
Table 2.9 Summarised characteristics of natural AMPs in the "ADP3_AMP_PigPathogenDataset"……...	**129**
Table 2.10. AMP online databases and a selection of computational tools utilised for AMP in silico analysis……..	**132**
Table 2.11 Primers designed for the insertion of SpeI(opal stop "TGA") downstream of *gIII*……	**140**
Table 2.12 $VNN_{15}(TTT)_1$ and NNK degenerate primer design and Inverse PCR thermocycling……	**142**
Table 2.13 PCR Round 1 "Linker" NGS primers for pSD3 _16mer library (sequences reverse strand)……..	**155**
Table 2.14 PCR Round 2 "Barcoding & Adapter Linker " NGS primers for pSD3 _16mer library……	**156**
Table 3.1. Summary of the hydrophobic stretches identified in AMPs of varying peptide length ranges from the APD3_AMP_PigPathogensDataset…….	**174**

Table 3.2 Sequential stretches of ≥2 hydrophobic amino acids observed in ≥3 peptide sequences in the APD3_AMP_PigPathogensDataset…… **176**
Table 3.3 Sanger sequencing of randomly handpicked pIII peptide phage VNN_{15}+$(TTT)_1$ and NNK library *E. coli* TG1 clones…… **186**
Table 3.4 Sanger sequencing of randomly handpicked -pIII (No gIII) peptide VNN_{15}+$(TTT)_1$ and NNK sub library *E. coli* JE5505 clones……. **188**
Table 3.5. Summary of pIII 16mer peptide phage library (*E. coli* TG1 and M13 phage) Next Generation Sequencing analysis……. **192**
Table 3.6. Summary of -gIII absent 16mer peptide sub library (*E. coli* JE5505) Next Generation Sequencing analysis…….. **192**
Table 4.1 Sanger sequencing of handpicked *E. coli* JE5505 carrying pSD3_Pln423 and improved mutants…… **225**
Table 4.2 2xYT and Mueller-Hinton broth cultivation of *E. coli* JE5505 carrying a pSD3 vector……. **230**
Table 4.3 The characterisation of antimicrobial activity of natural or phage-display derived AMPs; Peptide RLL, Pln-423, Bacteriocin E50-52, Cathelicidin-BF and Plectasin defensin…….. **234**
Table 4.4 Summary of the spotting peptide serial dilution agar diffusion assays of several "AMPs" against *L. innocua, E. coli JE5505, S. Typhimurium* and *S. suis*….. **245**
Table 5.1. Top50 enriched binders: ≥40% replicate reproducibility identified when panning VNN_{15}+$(TTT)_1$ and NNK respectively against *S. Typhimurium 4/74* and *S. suis P1/7*……. **273**
Table 5.2 Summary of the physiochemical and charged residue composition of Top50 enriched VNN_{15}+$(TTT)_1$ and NNK library binders…… **274**
Table 5.3 VNN_{15}+$(TTT)_1$ and NNK derived library peptides enriched over three rounds of panning and possess Z scores ≥ 2 and ≥40% positive dataset sample reproducibility……. **275**
Table 5.4. Summary of the physiochemical and charged residue composition of Top50 enriched VNN_{15}+$(TTT)_1$ and NNK library binders continued…… **275**
Table 5.5 Quantitative and qualitative presentation of ranked peptides with Z scores ≥2 identified in VNN_{15}+$(TTT)_1$ and NNK panning sample sets against *S. suis P1/7 and S. Typhimurium 4/74*…….. **279**
Table 5.6. Physiochemical properties and relevant amino acid compositional characteristics of Z score identified peptide binders……. **280**
Table 5.7. Quantitative and qualitative presentation of ranked peptides with Z scores ≥2 identified in VNN_{15}+$(TTT)_1$ and NNK panning sample sets against *S. suis P1/7 and S. Typhimurium 4/74* continued…….. **281**
Table 5.8 .Physiochemical properties and relevant amino acid compositional characteristics of Z score identified peptide binders continued…….. **281**
Table 5.9. Minimum inhibitory concentrations for VNN_{15}+$(TTT)_1$ and NNK library phage display derived peptides against *S.Typhimurium 4/74, S. suis P1/7* and *E. coli P433 (ETEC)*……. **289**
Table 5.10 The physiochemical, hydrophobic and structural estimates of NNK and VNN_{15}+$(TTT)_1$ phage display derived peptides……. **291**

1.4 List of Figures

Figure Title:	Page
Figure 1.1 AMP structural conformations and membrane modes of action……	**63**
Figure 1.2 Filamentous phage M13 mature virion structure and non-lytic lifecycle..	**72**
Figure 2.1 pSD3 16mer peptide phagemid vector map……..	**99**
Figure 2.2 Mechanism of whole-construct site-directed mutagenesis using inverse PCR primers……	**115**
Figure 2.3 Amino acid codon wheel for NNN (Unaltered genetic code: "N" = A/T/G/C). NNN contains 61 codons for twenty amino acids and three stops; TAA (opal), TAG (amber), TGA (ochre)…….	**136**
Figure 2.4 Amino acid codon wheel for NNK ("K" = G/T). The NNK scheme restricts the codon subset of the unaltered genetic code to 48. Amber stops (TAG) are still retained in this scheme……..	**137**
Figure 2.5 Amino acid codon wheel for the VNN degenerate scheme, where "V" = A/G/C (No T) and N: A/T/G/C…….	**138**
Figure 2.6 Schematic overview of the inverse-PCR method used for generating randomised $VNN_{15}+(TTT)_1$ and NNK 16mer peptide libraries……..	**143**
Figure 2.7 Schematic overview of ligases and restriction endonucleases utilised for peptide and library cloning strategies……	**148**
Figure 2.8 Overview of the whole-cell panning strategy utilised to screen $VNN_{15}+(TTT)_1$ and NNK degenerately randomised phage peptide libraries against *S. Typhimurium 4/74* and *S. suis P1/7*…….	**150**
Figure 2.9 Next Generation Sequencing coverage and PCR schematic for pSD3_16mer vectors…….	**155**
Figure 2.10 *in silico* peptide phage display analysis pipeline; TopN strategy……	**160**
Figure 2.11 *in silico* peptide phage display analysis pipeline; Z score strategy……..	**161**
Figure 3.1 Key physiochemical properties of small datasets of AMPs and Non-AMPs peptide sequences, (A) Peptide sequence length (B) Overall net charge, (C) Isoelectric point……..	**172**
Figure 3.2. Amino acid composition of the ADP3_AMP_PigPathogensDataset…	**173**
Figure 3.3 Cumulative percentage frequency of AMP-amino acid groups in AMP and Non-AMP datasets……	**173**
Figure 3.4. The percentage frequency of amino acids in the sequential hydrophobic stretches of peptides in APD3_AMPDataset_PigPathogens……..	**175**
Figure 3.5. Expected frequency of respective encoded amino acids and stop codons across a 16mer peptide region randomised using NNN or NNK schemes……	**177**
Figure 3.6. Cumulative percentage expected frequency difference of AMP amino acid groups with a peptide region randomised using NNN or NNK compared against the APD3_AMP_PigPathogensDataset…….	**178**
Figure 3.7 Expected frequency of respective encoded amino acids and stop codons across a 16 mer peptide region randomised using the VNN scheme……...	**179**
Figure 3.8 Cumulative percentage expected frequency difference of AMP amino acid grouping across a 16mer peptide region randomised using; VNN, $VNN_{15}+(TTT)_1$, NNK and $NNK+(TTT)_1$ compared against the APD3_AMP_PigPathogensDataset……..	**180**
Figure 3.9 Verifying PCR modifications used to generate library template; pSD3_RLL_-BspQI_+SpeI_opal stop visualised by agarose gel electrophoresis…..	**182**

Figure 3.10 Exemplar image of the agarose gel electrophoresis used to confirm amplification after library construction inverse PCR……. **184**
Figure 3.11. Agarose gel electrophoresis visualisation of SpeI digests of VNN_{15}+$(TTT)_1$ and NNK pIII peptide phage constructs…… **187**
Figure 3.12 Representative agarose gel electrophoresis images post-PCR NGS sample preparation for VNN_{15}+$(TTT)_1$ and NNK naïve peptide libraries in *E.coli* TG1, *E. coli* JE5505 and M13 phage…….. **190**
Figure 3.13. Peptide insert lengths in barcoded sequences identified with the correct flanking motifs and >1Aa in insert across all six naive peptide libraries…… **192**
Figure 3.14. VNN_{15}+$(TTT)_1$ and NNK pIII 16mer peptide phage libraries transformed into *E. coli* TG1; Heat map of the percentage difference between the expected and observed frequency for the twenty natural amino acids and three stops. **193**
Figure 3.15. VNN_{15}+$(TTT)_1$ and NNK pIII 16mer peptide phage libraries transformed into *E. coli* JE5505 and M13 phage; Heat map of the percentage difference between the expected and observed frequency for the twenty natural amino acids and three stops…… **196**
Figure 3.16. Observed frequency of phenylalanine across the 16mer peptide region in VNN_{15}+$(TTT)_1$ randomised peptide libraries……. **197**
Figure 3.17. Percentage frequency of amino acids and stop codons across the 16mer peptide region in VNN_{15}+$(TTT)_1$ and NNK libraries in *E.coli* TG1, JE5505 and M13 phage…….. **198**
Figure 3.18 Overview of the AMP amino acid grouping distribution in naive $VNN_{15}(TTT)_1$ and NNK naïve 16mer peptide libraries…… **201**
Figure 4.1 Growth curve of wildtype *E. coli* JE5505 (lpo-) mutant grown in 2xYT broth, 200rpm at 37^{o}C…….. **223**
Figure 4.2. Verifying PCR modifications used to generate pSD3_Pln423_-BspQI_-gIII and respective improved mutants by agarose gel electrophoresis visualisation… **224**
Figure 4.3 *E. coli* JE5505 colonies carrying pSD3_Pln423-BspQI,-gIII vectors tested against *L. innucoa* strain in a modified Gurlap et al., (2013) LB and TSB soft agar overlay assay…… **227**
Figure 4.4 Optimised Mueller-Hinton soft agar leaky *E. coli* JE5505 overlay assay method for the screening of known or uncharacterised AMPs……. **231**
Figure 4.5. *E. coli* JE5505 (pSD3_3-BspQI,-gIII) encoding Pln-423 and improved mutants against *L. innocua* in a 0.7% Mueller-Hinton soft agar overlay assay screen.. **232**
Figure 4.6 Growth curves of *E. coli* JE5505 wt and pSD3 carrying variants in Mueller-Hinton broth supplemented with 1% Glucose or 1mM IPTG…….. **235**
Figure 4.7. Modified Mueller-Hinton leaky *E. coli* JE5505 overlay assay library screening approach and recovery of zone producing JE5505 colonies for antimicrobial peptide variant identification…… **239**
Figure 4.8 *S. Typhimurium* broth cultures of varying colony forming units in a Mueller-Hinton broth tetrazolium red colorimetric assay……. **241**
Figure 4.9 *E. coli* JE5505 broth cultures of varying colony forming units in a Mueller-Hinton/tetrazolium red colorimetric assay…… **241**
Figure 4.10 Spotting peptide serial dilution agar diffusion assay for Cathelicidin-BF and Peptide RLL tested against *L. innocua* and *E. coli* JE5505…… **244**
Figure 4.11 Spotting peptide serial dilution agar diffusion assay for Cathelicidin-BF and Peptide RLL tested against *E. coli* JE50505……. **245**
Figure 4.12. 0.7% soft agar Mueller-Hinton *E. coli* JE5505 overlay assay screening of 16mer peptide libraries degenerately randomised via VNN_{15}+$(TTT)_1$ or NNK scheme……. **249**

Figure 5.1 Next Generation Sequencing (Ion Torrent) round 3 panning sample preparation…….. **270**
Figure 5.2 Selection of Top50 enriched binders originating from screening VNN_{15}+$(TTT)_1$ phage library vs S. suis round 2 output phage against both test species; *S. suis* P1/7 and *S. Typhimurium* 4.74 in the third panning round…… **271**
Figure 5.3 VNN_{15}+$(TTT)_1$ and NNK derived library peptides enriched over three rounds of panning and possess Z scores ≥ 2 and ≥40% positive dataset sample reproducibility…….. **278**
Figure 5.4. Broth microdilution 200µM screening of shortlisted phage display derived peptides against key Gram-negative bacterial pathogens of pigs; *S. Typhimurium* 4/74……. **285**
Figure 5.5 Broth microdilution 200µM screening of shortlisted phage display derived peptides against key Gram-negative bacterial pathogens of pigs; *E.coli* P433 (porcine ETEC)……. **285**
Figure 5.6 Broth microdilution 200µM screening of shortlisted phage display derived peptides against key Gram-positive bacterial pathogens of pigs; *S. suis* P1/7 Screen A……. **286**
Figure 5.7 Broth microdilution 200µM screening of shortlisted phage display derived peptides against key Gram-positive bacterial pathogens of pigs; *S. suis* P1/7 Screen B…….. **288**

1.5 Key Abbreviations

Antimicrobial peptide(s)	AMP/AMPs
2× Yeast Extract Tryptone	2xYT
Mueller-Hinton Agar	MHA
Mueller-Hinton Broth	MHB
Polymerase Chain Reaction	PCR
Next Generation Sequencing	NGS
NNK	N:"A/T/G/C" and K: "G/T"
VNN	V: "A/G/C" and N: "A/T/G/C"
LIB_SS2_SS3	Library panned against *S. suis* P1/7 for rounds 2 and 3.
LIB_TY2_TY3	Library panned against *S. Typhimurium 4/74* for rounds 2 and 3.
LIB_SS2_TY3	Library panned against S. suis *P1/7* in round 2, and this output phage was panned against *S. Typhimurium 4/74*
LIB_TY2_SS3	Library panned against S. suis *P1/7* in round 2, and this output phage was panned against *S. Typhimurium 4/74*

Amino acid abbreviations

Ala	**A**	Alanine
Arg	**R**	Arginine
Asn	**N**	Asparagine
Asp	**D**	Aspartic acid
Cys	**C**	Cysteine
Gln	**Q**	Glutamine
Glu	**E**	Glutamic acid
Gly	**G**	Glycine
His	**H**	Histidine
Ile	**I**	Isoleucine
Leu	**L**	Leucine
Lys	**K**	Lysine
Met	**M**	Methionine
Phe	**F**	Phenylalanine
Pro	**P**	Proline
Ser	**S**	Serine
Thr	**T**	Threonine
Trp	**W**	Tryptophan
Tyr	**Y**	Tyrosine
Val	**V**	Valine

TAG	Amber
TAA	ochre
TGA	opal or umber

Chapter 1: Introduction

1.1 GLOBAL PIG PRODUCTION, FOOD SECURITY AND ANTIMICROBIALS

1.1.1 One Health: food security and pig production

The "One World - One Health" principle is a contemporary iteration of the humanistic desire for the conscious custodianism of agri-food and health systems, by the means of interdisciplinary sustainable solutions (Destoumieux-Garzón *et al.*, 2018). However, the strategic stewardship of livestock production to alleviate global food insecurity and animal-human interface health risks, whilst refraining from further depleting the environment and agri-resource base, still remains a significant challenge (Barrett, 2010; Garcia, Osburn and Jay-Russell, 2020). Human and animal health are persistently threatened by several factors of growing concern. For instance, infectious disease dissemination; many of which are zoonotic in origin (Destoumieux-Garzón *et al.*, 2018), antimicrobial resistance (AMR) development; which leads to ~700, 000 human deaths globally per annum (Pokharel, Shrestha and Adhikari, 2020) and in absence of substantially effective climate change mitigations, by 2050 the risk of hunger and malnutrition is projected to rise by ~20% globally (WFP, 2021).

Malnutrition is a key anthropometric index used to quantify global food insecurity and involves imbalances, deficiencies or excesses, in an individual's energy and nutrient intake. Consequently, malnutrition exhibits an intrinsic dualism; manifesting both as high rates of undernutrition, which more broadly includes wasting, stunting and micronutrient deficiencies, co-existing in tandem with overweight, obesity and various diet associated non-communicable diseases (Perez-Escamilla *et al.*, 2018; Hoffman, 2019). Despite global efforts to curtail this double burden, the prevalence of malnutrition has increased in children (≤ five years old), and undernutrition in this key developmental stage contributes to ~45% of children deaths

globally (Fanzo *et al.,* 2018). Malnutrition is also a significant risk factor associated with increased mortality in older adults irrespective of the cause of death, and this will be of particular concern while global populations continue to include higher proportions of adults ≥ sixty years old (Stratton *et al.,* 2006; Söderström *et al.,* 2017)

The world's population will reach ~9.7 billion by 2050, which is a ~24% increase than at present (United Nations, Department of Economic and Social Affairs, Population Division, 2019). Food insecurity is inextricably linked to the upward trajectory of human population dynamics, rapid global urbanisation and land scarcity (Prosekov and Ivanova, 2018). As regions globally shift from rural to urban economically prosperous majorities, various diet transitions will ensue and nutritious diets will increasingly demand livestock products; with meat and dairy demand forecasted to increase ~60% by 2050 (Guyomard *et al.,* 2012; Hatab, Cavinato and Lagerkvist, 2019).

1.1.2 Pig farming and antimicrobials

Pigmeat will play a vital role in future agri-food supply chains. Global pig production will observe an ~8.6% increase by 2030 and in the same year, global consumption of antimicrobials will increase by 67% compared to levels in 2010 (Monger et al., 2021; Tsekouras et al., 2022). The intensification of global animal production will be the source of two-thirds of this predicted increase (Monger et al., 2021; Tsekouras et al., 2022). Pig farming typically involves vertically integrated high-intensive production systems albeit, these highly cost-efficient farming approaches often necessitate antimicrobial usage for increased productivity and disease reduction (Niemi et al., 2020). Van Boeckel *et al.,* (2015) combined antimicrobial consumption data with high-resolution geographical mapping of livestock populations; to estimate the average annual antimicrobial consumption per

kilogram animal meat produced. Pig production ranked higher than any other livestock, with antimicrobial consumption(mg) per kilogram of animal meat produced ranging from; 45 $mg \cdot kg^{-1}$, 148 $mg \cdot kg^{-1}$, and 172 $mg \cdot kg^{-1}$ for cattle, chicken and pigs respectively (Van Boeckel et al., 2015). Antimicrobial usage for livestock often outweighs use for human medical needs, with food-producing livestock consuming >70% of antimicrobials produced globally (Zeineldin et al., 2019). In pig farming antimicrobials typically include antibiotics and mineral feed-additives such as zinc oxide (ZnO) and copper sulfate ($CuSO_4$) (Højberg et al., 2005). Antimicrobial usage is governed by complex interconnecting factors of animal health, productivity, regulatory laws and farm economics (Wittum et al., 1996; Højberg et al., 2005; Zimmerman et al., 2017).

1.1.3 Pig production and antibiotic usage

Traditionally, the key pillars of antibiotic utility in pig production pertains to controlling the dissemination of bacterial infections (metaphylaxis) or preventing disease onset (prophylaxis) in high-risk groups or animals under stress (Monger et al., 2021; Tsekouras et al., 2022). Specifically, this is achieved by improving feed intake, growth performance and productivity, particularly during weaning, post-vaccination and post-farrowing periods (Monger et al., 2021; Tsekouras et al., 2022). Lekagul *et al.,* (2019) systematically reviewed thirty-six global studies conducted between 2000-2017 and reported 93% of total antibiotics administered to pigs were for prophylaxis whole-group treatment strategies; via medicated feed or drinking water containing a top dressing of oral antibiotic powder.

Pig farms routinely use antibiotic treatments for entire pig production batches at strategic timepoints of likely increased infection susceptibility such as; birth and castration (week 1), weaning (~ 2-5 weeks) and the commencement of the finishing

period (~week 9) (Sjölund et al. 2016). The prevalence of this practice is due to the vulnerability of underdeveloped pig immune systems. For example, enterotoxin producing *Escherichia coli* in pig herds are particularly prevalent in suckling pigs (neonatal: <7 days old) due to lacking passive lacteal immunity from sows with insufficient antibodies and vaccination profiles (Wittum et al., 1996; Katsuda et al., 2006). Weaning piglets (~8-35 days) equally possess poorly developed immune systems as maternal acquired immunity dissipates within three to four weeks from birth (Monger et al., 2021).

Over-reliance on antimicrobial usage at younger ages has a positive association with higher usage in older pigs, leading often to higher disease pressures and susceptibility in pig herds (Sarrazin et al., 2019). Additionally, when animal stress coincides with antibiotic or environmental pressures, this drastically alters normal flora population dynamics and intestinal motility thus, reducing the replication required for pathogenic bacteria to manifest disease (Zimmerman et al., 2017;Monger et al., 2021). Therefore, high antibiotic usage perpetuates gastrointestinal bacteria pathogens, for instance antibiotic usage in pigs induces a five times higher risk of susceptibility to seropositive *Salmonella* (De Lucia and Ostanello, 2020).

The European Union placed limitations on the prophylactic use of antibiotics however, antimicrobial classes are still frequently administered orally for pig group treatments or applied parentally, and this includes antibiotics such as; aminoglycosides, amphenicols, cephalosporins, fluoroquinolones, lincosamides, macrolides, penicillins (β-lactamase-sensitive and extended-spectrum), polymyxins, sulphonamides and tetracyclines (Sarrazin et al., 2019). A portion of these antimicrobials are favoured for their growth promotion benefits; for instance, Che *et al,* (2019) demonstrated the low antibiotic dosage of a premix of chlortetracycline and

synergistic virginiamycin led to an enrichment in the intestinal microbiota. Contributing to short-chain fatty acid production, thus improving carbohydrate metabolism and energy utilisation in piglets. However, Gao *et al.*, (2018) demonstrated the fine balance required to observe antibiotic-induced growth promoting effects; as antibiotic cocktails of ampicillin, gentamicin and metronidazole can alter pig microflora metabolism, in fact resulting in lower concentrations of faecal branched short-chain fatty acids (acetic, propionic, and butyric acids). These fatty acids can supplement pig nutritional uptake in the gut and are key catabolites resulting from microbiota fermentation of undigested dietary carbohydrates and proteins (Zeineldin et a., 2019). Gao *et al.*, (2018) additionally reported an upsurge of *E. coli* replacing the complex network of normal microflora bacterial genera populations such as *Bifidobacterium*, *Lactobacillus* and *Ruminococcus.*

1.1.4 Antibiotics utilised in pig farming.

A significant portion of the common antibiotics used in pig farming are listed on the World Health Organisation (WHO) list of Critically Important Antimicrobials for Human Medicine (Scott et al., 2019). Subcategorisation into "Highest priority" or "High priority" critically important antimicrobials denotes the reliance and impact of said antimicrobial in several instances such as; treating serious human bacterial infections (*i.e.* last resort antimicrobials), various settings of bacteria resistance transmission (zoonotic, community, among others) and consideration of risk populations (*i.e.* patients in health care settings susceptible to further opportunistic infections) and whether non-human transmission is extensively evidenced as a source of emerging resistant bacteria. Notwithstanding the paucity of validated pathogen epidemiological data, various studies have evidenced the zoonotic transmission of resistant bacteria mainly, non-typhodial *Salmonella*, *E. coli*, *Campylobacter spp,*

Enterococcus spp, and methicillin-resistant *Staphylococcus aureus* (Nhung et al., 2016). Highest priority critically important antimicrobials include; cephalosporins (3^{rd} /4^{th}/5^{th} generation), glycopeptides, macrolides and ketolides, polymyxins (*i.e.* colistin), quinolones whereas, penicillin and analogue derivatives, aminoglycosides, penems (*i.e.* Carbapenems), ansamycins are all deemed high priority antimicrobials (Scott et al., 2019).

In the extant of literature, β-lactams such as penicillins; benzylpenicillins, aminopenicillins and tetracyclines such as doxycycline, chlortetracycline are reported as the most frequently used antibiotic classes in pig farming (Nhung *et al.,* 2016; Lekagul et al., 2019; Monger et al., 2021). The synthetic sulfonamides and *Streptomyce*s spp derived lincosamides, were reported as relatively common, and colistin was especially exploited in farrow-to-finish and fattening farms across Europe and Asia (Lekagul et al., 2019). Chlortetracycline is a common antibiotic used for growth promotion in the US and Southern Asia (Apley et al, 2012; Nhung *et al.,* 2016;). Broad-spectrum antibiotic amoxicillin is the most common prophylactic antibiotic used in Thailand (Lekagul et al., 2020) and in nursery pig farms in Canada (Bosman et al, 2022). Highest priority CIA for human use such as macrolides, quinolones, third and fourth generation cephalosporins were underutilised in comparison to aforementioned antibiotic classes typically constituting <20% of the total use in studies reviewed (Lekagul et al., 2019)

Sub-therapeutic doses of an antimicrobial cultivates selection pressure, not only to the antimicrobial of treatment, but potentially numerous antimicrobial classes which can highly differ in their properties and structures (Pokharel et al, 2020). Livestock animals are rich reservoirs of antimicrobial resistance genes (AGR), some of which can be horizontally transferred, and this contributes to continued epizootic

or zoonotic outbreaks of AMR bacterial pathogens (Woolhouse et al., 2015; Pokharel et al, 2020). AGRs in pigs often flux with changing populations of intestinal microbiota especially with antibiotic usage and exogenous bacteria colonisation of the gut (Græsbøll et al., 2019). AMR development in the microbiome of animal reservoir populations is positively correlated to livestock antimicrobial consumption for instance, Chantziaras *et al.*, (2013) reported that commensal *E. coli* isolates from pigs were rich sources of antimicrobial resistance to several classes of antibiotics including; ampicillin, sulphonamides, streptomycin and tetracycline, and this was specifically the case in evaluated countries with the highest antimicrobial consumption.

1.1.5 Antibiotic resistance mechanisms

Bacteria can acquire resistance to antibiotics classes either by the *i)* effluxion of antibiotics from the cell often coupled with the overexpression of efflux pumps, *ii)* enzymatically modifying and metabolising antibiotics or *iii)* altering the antibiotic target. For instance, in Gram-negative and Gram-positive bacteria, tetracycline resistance is commanded by ~36 known *Tet* genes involved in efflux pump mechanisms (*TetA-E, G, H, J, K, Y, Z* and *Tet30*), production of ribosomal protection proteins (*TetO, M, Q, S, P*) which are highly homologous to elongation factors and destabilise tetracyclines ribosomal interactions, and enzymatic alteration mechanisms (*TetX*) (Blair et al., 2014). Numerous efflux systems can confer resistance to tetracycline in different bacteria; *E. coli* (AcrB-TolC, AcrAB-TolC, AcrEF-Tol-C), *S. aureus* (MepA), *Psudomanas aeruginosa (*MexXY-OprM, MexCD-OprJ, MexAB-OprM) and *Campylobacter jejuni* (CmeABC) (Blair et al., 2014).

AcrAB-TolC once assembled can export and thus confer resistance to a diverse array of substrates with minimal compositional and structure similarity including and not limited to Triton X-100, bile salts, β-lactams, erythromycin,

chloramphenicol, tetracycline, fluoroquinolones, among others (Blair et al., 2014). Efflux-mediated resistance has been widely evidenced for fluoroquinolones. In comparison to susceptible strains, efflux containing bacteria typically only exhibit ~2-8 fold higher minimum inhibitory concentrations (MICs) for the same antimicrobial agent whereas, >100-fold increase is associated with resistance derived from enzyme antimicrobial inactivation or antimicrobial target alterations (Piddock et al, 2006).

Polymyxin resistance in Gram-positive bacteria is conferred by the *D*-Alanylation of cell wall teichoic acids (WTA/LTA) mediated by *dlt* operon and/or *L*-lysine incorporation into phosphatidylglycerol membrane lipids via the *mprF* gene products (Weidenmaier and Peschel, 2008; Sarr-Dover et al., 2012). S. *aureus, Streptococcus* (Group A, Group D and *S. pneumoniae*) and *Enterococcus faecalis dlt* operon deficient mutants lack *D*-alanyl esters in teichoic acids and exhibit increased susceptibility to defensins, protegrin and other cationic AMPs (Sarr-Dover et al., 2012). Polymxyin resistance is conferred in Gram-negative bacteria by the reduction of anionic charges in lipopolysaccharide (LPS) by the incorporation of phosphoethanolamine or 4-amino-4-deoxy-L-arabinoase (Ara4N) into lipid A (Yin et al., 2020).

Plasmid-meditated colistin resistant gene (*mcr-1*), was identified in commensal *E. coli* isolated from humans, pigs and pork products in China (Li et al., 2020). *mcr-1* encodes an LPS-modifying phosphoethanolamine transferase that catalyses the modification of bacteria outer membrane lipid A, and this is hypothesised to lower permeability and protect against hydrophobic antibiotics which are heavily reliant on lipid interactions for membrane traversing (Khondker and Rheinstädter, 2020; Li et al., 2020). *E. coli MCR-1* carriage within global pig production is emblematic of growing concerns for AMR. As colistin is a Highest priority CIA relied upon for the

last resort treatment of human infections caused by multidrug resistant (MDR) bacterial pathogens such as *Pseudomonas aeruginosa*, *Acinetobacter baumannii* and various pathogenic *Enterobacteriaceae* (Lekagul et al., 2019).

Bacteria can produce β-lactamases to inactivate β-lactam antibiotics via hydrolysation of the amide bond. Members of the *Enterobacteriaceae* family, and Gram-positive bacteria such as *S. aureus*, can express extended spectrum β-lactamases that hydrolyse specific antibiotics such as penicillins, $1^{st}/2^{nd}/3^{rd}$ generation cephalosporin including oxyimino-cephalosporins such as cefotaxime, ceftazidime, ceftriaxone, cefuroxime, cefepime) and monobactams (aztreonam) (Kapoor et al., 2017). Numerous studies have reported extended spectrum β-lactamases in pig farms specifically where animal production is supported by antibiotic usage (Ada et al., 2021). Hammerum *et al.*, (2014) found that of 195 Danish pig farms, extended spectrum β-lactamases-producing *E. coli* were detected in 79% of pig farms linked with the high consumption of cephalosporins, 59% higher than farms with no cephalosporin consumption. *E. coli* encoding extended spectrum β-lactamases have been isolated from pig carcasses in the EU and are readily emerging in both humans and other animal reservoirs (Bergspica et al., 2020).

1.2 DISEASES IN PIGS

1.2.1 Global pig markets and disease trends

The demand for pork increasingly contributes to the intensification of global pig production, and reliance on more intensive forms of pig husbandry (>high density pig herds and housing). This provides the quintessential conditions necessary for the rapid dissemination of pathogens, and the consequences of infectious diseases in pig

production manifests as; livestock mortality, farm productivity losses and reduced meat market value (VanderWaal and Deen, 2018). Broadly summarising, the most concerning pig diseases for pig health and production are viral or bacterial pathogens. Common viruses of concern in pig production include; rotavirus, coronaviruses including Transmissible gastroenteritis virus, TGE and porcine epidemic diarrhea virus, Porcine reproductive and respiratory syndrome virus, PRRSV, Classical swine fever virus, and Swine influenza virus, among others (Katsuda et al., 2006; Saade et al., 2020).

1.2.2 Gastrointestinal and respiratory bacterial pathogens in pigs

In pigs, viral and bacterial pathogens assimilate their pathogenic niche within one or more (polysystemic) systems and in particular this includes diseases of the skin, gastrointestinal, respiratory, urinary and reproductive systems (Zimmerman et al, 2017). Bacterial pig pathogens such as *E. coli* strains, *Clostridium perfringens* and *Actinobacillus pleuropneumoniae* infect one system contrastingly *Streptococcus*, *Salmonella* and PRRSV are polysystemic pathogens (Monger et al., 2021). **Table 1.1** summarises the principal bacterial pathogens of swine by phenotypic variance gram stain, cellular shape, oxygen tolerance and spore formation. Bacterial pathogens of pigs typically exhibit niches in the gastrointestinal or respiratory systems. Gram-negative and Gram-positive bacteria which cause gastrointestinal and/or respiratory infections in pigs are summarised in **Table 1.2** and **Table 1.3** respectively.

Table 1.1 **Phenotypic classification of the primary bacterial pathogens of pigs by genera.**

Summary of key bacteria genera with porcine host specificity	
Phenotypic classifiers	**Genus/Genera**
Cell wall containing, Gram-positive bacteria	
Cocci, aerobic to facultatively anaerobic	*Enterococcus* *Staphylococcus* *Streptococcus*
Bacilli (rod), (non-spore forming) a) Aerobic to microaerophilic anaerobic b) Anaerobic	 *Arcanobacterium* *Eysipelothrix* *Listeria* *Mycobacterium* *Rhodococcus* *Actinobaculum*
Bacilli (rod), (spore forming) a) Aerobic b) Anaerobic	 *Bacillus* *Clostridium*
Cell wall containing, Gram- negative bacteria	
Straight bacilli (rod), a) Aerobic to facultatively anaerobic	 *Actinobacillus* *Bordetella* *Brucella* *Burkholderi*a *Escherichia* *Salmonella* *Haemophilus* *Pasteurella* *Yersinia*
Curved to spiral-shaped bacilli (rod), a) Aerobic to facultatively anaerobic	 *Brachyspira* *Campylobacter* *Lawsonia* *Leptospira* *Treponema*
Cell wall absent bacteria (intracellular obligates)	
Pleomorphic cellular shape **a)** Aerobic to facultative anaerobic	 *Mycoplasma*
Dimorphic, cocci or bacilli (rod) cellular shape a) Anaerobic	 *Chlamydophila*
Adapted from Zimmerman *et al*, 2017	

Table 1.2 Gram-negative bacterial pathogens of pig digestive and respiratory systems and their propensity of occurrence at varying pig developmental stages.

	Diseases and/or clinical signs of disease	Gram-negative bacteria	Occurrence across development			
			Pre weaning	Post weaning nursey	Post weaning grows to finish	Adult mature
Digestive system	Enteritis[1] Colibacillosis[2] Swine Dysentery[3] Colitis[4] Colonic spirochetosis [5] Porcine proliferative enteropathy [6]	*Brachyspira hyodysenteriae* [3]	++	++	+++	++
		Brachyspira pilosicoli [5]	+	+	+	
		Bordetella bronchiseptica [7]	+	+	+	+
		Campylobacter jejuni [1]	+	+		
		Escherichia coli [2]	++	++	+	+
		Lawsonia intracellularis [6]		+	+++	+
		Salmonella enterica Typhimurium, Dublin, Enteritidis [1]	+	++	++	
		Yersinia enterocolitica [1]				
		Yersinia pseudotuberculosis [4]	++	+	++	++
Respiratory system	Pneumonia [7] Pleuropneumonia [8] Pneumonic pasteurellosis Pasteurellosis [9] Glässer's disease [10]	*Actinobacillus pleuropneumoniae* [8]	++	++	++	++
		Actinobacillus suis [7]		+	+++	+
		Glaesserella (Haemophilus) parasuis [10]	+	+++	+++	
		Pasteurella multocida [9]	+	++		
		Salmonella enterica choleraesuis [7]		+	++	+
				+	+++	

Bacteria spp [*n*], where *n* = 1 to 13 signifies associated diseases or clinical signs of disease.

+ (occasional), ++ (common), and +++ (routine).

Adapted using Helke et al., 2015; Zimmerman et al., 2017; Robbins et al., 2014; CABI [date accessed; 14/04/2022]

CABI, current year. Invasive Species Compendium. Wallingford, UK: CAB International. www.cabi.org/isc.

Table 1.3 Gram-positive bacterial pathogens of pig digestive and respiratory systems and their propensity of occurrence at varying pig developmental stages.

	Diseases and/or clinical signs of disease	Gram-positive bacteria	Occurrence across development			
			Pre weaning	Post weaning nursey	Post weaning grows to finish	Adult mature
Digestive system	Enteritis[1] Listeriosis [11] Necrotic enteritis [14]	*Clostridium dificile* [1]	+++			
		Clostridium perfringens Type A and Type c [1]/[14]	+++	++		
		Enterococcus spp (E. durans, E hirae) [1]	+	+		
		Listeria monocytogenes [11]	+	+		
		Staphylococcus aureus [14]	+	+		
Respiratory system	Pneumonia [7] Meningitis [12] Tuberculosis [13]	*Mycobacterium spp (M.avium, M.bovis, M. tuberculosis)* [13]	*	*	*	*
		Streptococcus dysgalactiae [7]	+++	++	++	
		Streptococcus porcinus [7]/[12]		+	+	
		Streptococcus suis [7]/[12]	++	+++	+++	
		Mycoplasma hyopneumoniae [7]		+	+++	+

Bacteria spp [*n*], where *n* = 1 to 13 signifies associated diseases or clinical signs of disease

*rare, + (occasional), ++ (common), and +++ (routine).

Adapted using Helke et al., 2015; Zimmerman et al., 2017; Robbins et al., 2014; CABI [date accessed; 14/04/2022]

CABI, current year. Invasive Species Compendium. Wallingford, UK: CAB International. www.cabi.org/isc.

In the case of the gastrointestinal system, digesta flow and bacterial growth rates both dictate bacterial adherence to gastrointestinal mucosal and epithelial cell surface layers (Canibe et al., 2001). The stomach and small intestine of pigs contains 10^3–10^5 culturable bacteria per gram of content contrastingly, the high diversity (>400 bacterial species) of culturable bacteria in the large intestines exceeds 10^{10} bacteria per gram content (Canibe et al., 2001). Predominant bacteria cultivated from pig intestines include but is not limited to; *E. coli*, *Streptococcus*, *Lactobacillus*, *Clostridia*, *Selenomonas*, *Mitsuokella*, and *Enterobacteria* (Canibe et al., 2001). *Enterobacteriaceae* are Gram-negative rods which are oxidase-negative and catalase-positive, and this family of bacteria colonise the small and large intestines with some pathogenic strains causing enteric diseases in pigs (Schierack et al., 2007). *Enterobacteriaceae* such as *E. coli* can be either commensal autochthonous microbiota (non-pathogenic) and/or non-indigenous allochthonous strains which pass through microhabitats, via faecal-oral route or ingestion of contaminated foodstuff (Pluske et al., 2002).

Enteric infections are attributed to significant economic losses, reduced pig productivity and regardless of infectious agent(s) these infections result in the loss of water, electrolytes and nutrients in pigs (Mesonero-Escuredo et al., 2018). Diarrhea (faecal dry mater with excess water) is often observed as a clinical sign of enteric infections (Mesonero-Escuredo et al., 2018). The manifestation of diarrhea is complexly determined by interactions between various pathogen types, gut dysfunction, host immunity and inflammation responses (enteritis) (Zimmerman et al., 2017). Numerous distinct genera and species of enteric bacteria pathogens inhabit different regions of the pig gastrointestinal tract and prevalent infections are often associated with varying pig developmental age as demonstrated **in Table 1.2 and 1.3.**

Lesions from *L. intracellularis* and *Brachyspira spp* are typically observed in the small and large intestines respectively, with *B. hyodysenteriae* and B. *pilosicoli* showing differing abilities to cause mucohaemorrhagic intestinal mucosa necrosis (Casas et al, 2017).

Pre- and post-weaning piglets are particularly disposed to diarrhea inducing infections and cases of endemic gastrointestinal diseases can threaten finishing pigs (Katsuda et al., 2006). *E. coli* strains can cause post-weaning colibacillosis colonising the small intestine (~3-10 days post weaning), pathogenic *E. coli* strains (ETEC, EPEC, EHEC, STEC) and *C. perfringens* type C (α- and β-toxin producing) and *C. perfringens* type A (α-toxin and ß2-toxin producing) infections are frequent sources of piglet diarrhea cases and outbreaks often carry high morbidity and mortality (Katsuda et al., 2006; Helke et al., 2015; Mesonero-Escuredo et al., 2018). Diseases of the respiratory system cause significant mortality in US nursery farms and fattening-finishing farms, contributing to 47.3% and 75.1% of the farm mortality rate respectively (Monger et al., 2021).

Weaning pigs are susceptible to early colonising agents which are potentially pathogenic under the presence of external stressors; *H. parasuis, S. suis and Actinobacillus spp* are considered the most common and economically impactful respiratory pathogens of pigs (Olvera et al, 2007; Zimmerman et al, 2017). Additionally, respiratory disease in pigs is often complex and involves co-infection with viral and bacterial pathogens. The Porcine Respiratory Disease Complex, PRDC consists of co-infections or superinfections involving viruses such as Swine influenza A virus, Porcine Circovirus type 2, PRRSV and highly virulent bacterial pathogens such as *A. pleuropneumoniae, M. hyopneumoniae, P. multocida* and *B. bronchiseptica* (Opriessnig et al., 2011).

The GB Pig Disease Surveillance dashboard stores >10 years' worth of data derived from submissions to the UK-wide veterinary diagnostic network, and the database includes > 9, 255 diagnosed cases of mono- or polysystemic pathogens. In the UK, common bacterial pathogens diagnosed occupied gastrointestinal and respiratory system niches; *Streptococcal spp* (11.3%) with *Streptococcus suis* contributing 7.1%, *Salmonella* spp (10.2%) with serovar *Typhimurium* attributing 8.5% of this and the monophasic ST-like variants (3.3%), *E. coli* (5.3%) , *Lawsonia spp* (4.0%), *A. pleuropneumoniae* (2.5%), *Pasteurella multocida (3.9%), Brachyspira spp* (4.9%) with both *B. pilosicoli* and *B. hydodysenteriae* attributed to (3.2%) and (1.7%) respectively, *Glaesserella* (*Haemophilus)* spp (2.5%), (Animal and Plant Health Agency: The Pig Disease Surveillance Dashboard, 2022).

VanderWaal and Deen, (2018) systematically reviewed ~57,500 global articles, published through 1966- 2016, focusing on infectious agents in pigs. Machine assisted annotations identified the top forty cited pig pathogens globally and regionally and this included; sixteen viruses, fifteen bacteria, eight parasites and one protozoan. VanderWaal and Deen, (2018) reported the most common problematic bacterial pathogens in pigs were; *Salmonella spp* (Rank 1: 6,466 citations), *E. coli* (Rank 2: 4,985 citations), *A. pleuropneumoniae* (Rank 8: 2, 552 citations), *P. multocida* (Rank 11: 2, 099 citations), *S. suis* (Rank 16: 1,776 citations) *M. hyopneumoniae* (Rank 17: 1, 665 citations). VanderWaal and Deen, (2018) emphasised there was an increasing threat of multi-AMR *S. suis* globally, which is additionally reflected in the frequent isolation of *S. suis* serotypes 1, 2, 7 and 14 in APHA, UK: GB Pig Disease Surveillance data (Animal and Plant Health Agency: The Pig Disease Surveillance Dashboard, 2022) Other notable, but lower ranked (>20) bacterial pathogens included *B.*

hyodysenteriae, H. parasuis and *L. intracellularis* with *Campylobacter* spp (Rank 22: 974) emerging as a regional threat in Latin America.

1.3 BACTERIA PATHOGEN WITH PORCINE HOST SPECIFICITY: *SALMONELLA*

1.3.1 Salmonella classification

Salmonella are motile facultative anaerobic Gram-negative rod shaped (2-5μm) bacteria belonging to the *Enterobacteriaceae* family (Andino and Hanning, 2015). *Salmonella* classification is a triplet hierarchy of a) species; *S. enterica* or *S. bongori* (V), b) subspecies; *S. enterica* subsp. *enterica (I), salamae (II), arizonae (IIIa), diarizonae (IIIb), houtenae (IV), indica (VI)* and c) ~2600 serotypes of *Salmonella* have been classified by the Kauffman–White scheme with *S. enterica* subsp. *enterica* (I) containing the largest number of serovars (Brenner et al. 2000). The Kauffman–White scheme centres upon three major antigenic determinants; somatic (O), capsular (K) and flagellar (H) (Brenner et al. 2000). *Salmonella* motility is derived from the diphasic heat liable peritrichous flagella, which can activate host immune responses.

Salmonella spp often possess two distinct flagellar genes phase IH antigens for immunological identity and phase II antigens are non-specific and highly conserved across serotypes. *Salmonella* monophasic variants can exist, wherein one flagellar gene is expressed at a time, and isolates from chickens and pigs have been identified with deletions to the flagella antigen *loci* (Clark et al., 2020). Observed alterations to the *loci* consist of combinations gene deletions in; *fliC* (1^{st}-phase flagellar antigen), *fljA* (1^{st}-phase transcription repressor), *fljB* (2^{nd}-phase flagellar antigen), and *hin* (DNA invertase for a DNA sequence containing a promoter for *fljA* and *fljB* gene transcription) (Clark et al., 2020). For bacterial motility, the commonly referenced monophasic *Salmonella enterica* serovar I Typhimurium 4,[5],12:i:- pig isolate

expresses *fliC* but does not express the second-phase flagellar antigen via the *fliB* gene (Shippy et al., 2018).

1.3.2 Salmonella: Zoonotic potential

Human salmonellosis can lead to mortality and consistently incurs economic losses additionally, severe infections are increasingly involving multi-drug resistant *Salmonella* (Niemi et al 2019). Ao *et al.,* (2015) estimated ~2.1-6.5 million human Salmonellosis (non-typhoidal) cases occur globally per year, and the general nature of underreported foodborne infections means this figure is likely an underestimation (Ao et al., 2015). Majowicz *et al,* (2010) application of the Monte Carlo simulation to surveillance data and literature studies estimated ~93.8m cases nontyphoidal Salmonella gastroenteritis occur globally per annum with ~155,000 related deaths, and 85.7% of cases are foodborne. Evangelopoulou *et al.,* (2015) estimated that salmonellosis cases in humans and pigs incurs ~€600m loss per annum for the EU trading area due to the removal of *Salmonell*a contaminated pork and treatment management for human infection cases(~€600-800 each). In 2009, The European Food Safety Agency, EFSA (2011) reported that 7.8% of foodborne outbreaks were verified as salmonellosis, and 9.8% of outbreaks were caused by the ingestion of *Salmonella* contaminated pigmeat and derived products thereof. *S. Enteritidis*, *S. Typhimurium* and respective monophasic *S. Typhimurium* (mST) variants such as *S. Typhimurium* 1,4,[5],12:i:- [4,5,14,15,16,17] are primarily responsible for human infections across industrialised regions (Campos et al., 2019; Cevallos-Almeidaac et al., 2019). As summarised in **Table 1.4**, the epidemiology of *Salmonella* serotypes in humans is often a coalition of *Salmonella* prevalence in poultry flocks, pig herds, cattle and respective animal derived products thereof such as eggs.

Table 1.4 *S. enterica* serotypes most frequently associated with disease in key livestock animals and humans.

***S. enterica* serotypes <u>most frequently</u> associated with disease in key livestock animals and humans.**		
Reservoir	***S. enterica* serotypes**	
Swine	*S. Typhimurium* *S. Enteritidis* *S. Dublin*	*S. Choleraesuis, S. Typhisuis*
Chicken		*S. Gallinarum, S. Pullorum*
Cattle		
Sheep		*S. Abortusovis*
Humans	*S. Typhimurium, S. Enteritidis*	*S. Typhi* *S. Paratyphi* A, B, C
Adapted from Bäumler et al., 1998		

1.3.3 Salmonella in pigs and the epidemiology, prevalence and impact of pig-related serovars

Salmonella was initially discovered in the intestines of pigs in the 19th century by American microbiologist Theobald Smith, and *S. enterica* subsp. *enterica* (I) serotypes remain intertwined with this warm-blooded animal host. Porcine host specific *S. Choleraesuis* exclusively causes disease in swine with clinical signs manifesting as lesions and sepsis (Bäumler et al., 1998; Helke et al., 2015). Whereas *S. Typhimurium* strains possess broad-host specificity and persists as the most frequently isolated serotype in clinically ill pigs exhibiting enterocolitis (Zimmerman et al., 2017). Pigs are susceptible to infection from most *Salmonella* serotypes all of which can potentially contaminate pigmeat and pork-derived product. *Salmonella* colonisation can occur across all stages of pig production chain by horizontal transfer (*e.g.*, environment, wild animal/pest vectors and feed) or vertical transfer (*e.g.*, sow to piglet or within pig herd to slaughter).

Often a circular transmission pattern occurs, with a permanent cycle of horizontal and vertical transfer between the herd and farm environment wherein colonisation is perpetuated by the faecal shedding of *Salmonella* and oral-faecal transmission between pigs (Campos et al., 2019). Pig pathogens such as PRRSV can be significant risk factors for *Salmonella* shedding. Co-infection and synergistic effects of pig pathogens inducing immunosuppression likely play in the epidemiology of *Salmonella.* Likewise, Salmonella pervasive infection of pigs is associated with an increased susceptibility to other farm productivity decreasing diseases such as postweaning multisystemic wasting syndrome. Consequently, high health status pig herds are often linked to low risk of *Salmonella* infection.

1.3.4 Clinical signs of Salmonella infection in pigs

During the initial stages of *Salmonella* infection, pigs are often lethargic, febrile with temperatures of 40.5 - 41.6^{o}C (onset ~4 hr) and exhibit reduced movement (Moura et al., 2021). *Salmonella* induced proinflammation, high blood cytokine concentrations, fever with prostration, yellowish diarrhoea and neutrophil influx are key indicators of *Salmonella* infection (Côté, et al., 2004). *Salmonella* infections in pigs carries a case fatality rate of <10% (Zimmerman et al., 2017). High concentrations of *Salmonella* are typically seen in faeces in the initial few days following infection, followed by a gradual decrease after 2-3 weeks, (Bellido-Carreras et al., 2019). Typically, >10^3 CFU of *Salmonella* is required for the induction of acute infection and high doses of *S. Typhimurium* (i.e. 10^9 CFU) exacerbates the severity of the aforementioned clinical signs (Loynachan and Harris 2005; Boyen et al., 2009).

Most pigs are healthy carriers of *Salmonella* and seldom exhibit clinical signs of salmonellosis nonetheless shedding and the infiltration of intestinal tissues, ileum, lymph nodes and tonsils can still persist (García-Feliz et al., 2009; Pala et al., 2019).

Sub-clinical *Salmonella* infections in pigs causes reductions in projected productivity, feed conversion rates and lower body weights at the point of slaughter (Evangelopoulou et al., 2015). Asymptomatic carriage of *Salmonella* and intermittent faecal shedding are vital steps in the pathogenesis life cycle, and both are considerate threats to disease control in in pigs. Ivanek *et al.,* (2012) demonstrated faecal shedding and immune response to *Salmonella* are profoundly determined by the serotype and dosage of exposure; pigs challenged with *Salmonell*a at 10^9 CFU transitioned 1.5 days faster into *Salmonella* shedding, with the continuous shedding state extending by 10-26 days compared to a lower dose of 10^6 CFU. *S. Typhimurium* and *S. Derby* undergo rapid seroconversion and exhibit longer-lasting host infection (Ivanek et al., 2012).

D'Incau *et al.,* (2021) analysed 25,215 faecal, rectal swabs, gut content, samples among others from pig farms in northern Italy, 15.8% of samples were *Salmonella*-positive and over time monophasic *S. Typhimurium* (mST) became increasingly more dominant. *Salmonella spp* were identified across production phases from weaning to fattening period, however the manifestation of clinical signs was significantly lower for mST isolates; ~45% weaners and ~26% of growers-fatteners contrasted to non-mST *Salmonella Typhimurium* where clinical signs were observed in ~53% weaners and ~43% growers-fatteners respectively (D'Incau et al., 2021). Cevallos-Almeidaac et al., 2019 additionally reported piglets inoculated with mST continuously shed *Salmonella* and exhibit no fever symptoms. The absence clinical symptoms in pigs likely promotes mST circular dissemination in pig farms and increases the risk of mST contamination in slaughterhouse and abattoirs.

1.3.5 Salmonella Typhimurium antimicrobial resistance profiles in pigs

S. Typhimurium types DT9, DT204, DT193, DT104 and, monophasic *S. Typhimurium* 4,[5],12;i:- were consecutively the most epidemic isolates identified in pigs and humans (EFSA, 2017;Tassinari et al., 2020). Consumption of *S. Typhimurium* 4,[5],12;i:- contaminated pork has been linked to numerous food-borne outbreaks in humans (Bearson et al., 2020). *S. Typhimurium* 4,[5],12;i:- first emerged as an epidemic concern in European pig populations in ~2005, replacing the dominant DT104 clone, the zoonotic potential of this variant was rapidly seen in clinical human cases, with the consumption of contaminated pork being the main linked source (Tassinari et al., 2020). Epidemic isolates of *S. Typhimurium* have shared similar MDR profiles; typically including resistance to ampicillin, chloramphenicol, sulphonamides, streptomycin and tetracycline conferred via chromosomally insertion of an R- type *ASSuT "RR3"* and plasmid resistance acquisition (Tassinari et al., 2020). The *fjiAB* operon possesses a higher-than-average GC content (52.2%) comparative to the remainder of the *Salmonella* genome, +7.2 difference, this is hypothesised to increase the susceptibility to integration loss making this genetic region a hotspot for foreign DNA (Garcia et al., 2016). The RR3 complex and the novel integrative *Salmonella genomic island 4* (*SGI-4*), which harbours genes for an enhanced metal tolerance (copper, silver, arsenic resistance among others) has been jointly identified in *S. Typhimurium* 4,[5],12;i:- sequence type ST34.

Heavy metals are widely utilised in agriculture primarily due to their physiological benefit (*i.e.* enzymatic co-factors) however, these metals can promote the persistence of bacterial antibiotic resistance via co-selection mechanisms. Bearson et al, (2020) demonstrated that weaned crossbred pigs (barrows/gilts 19-21 day age) administered pre- and post-feed regimes of elevated ZnO and $CuSO_4$

additives, 2000mg/kg and 200mg/kg feed respectively, following intranasal inoculation of 8 x 10^7 CFU *S. Typhimurium* 4,[5],12;i:- ST34 SX 240 strain exhibited a slower decline in the rate of *Salmonella* shedding. This was compared to the control groups with increased no Zn+Cu supplementation but provided a basal diet with traces of Cu (16.5mg/kg feed), Zn (165mg/kg feed) and other metals including Fe, I, Mn (Bearson et al, 2020). Suggesting metal tolerance genes existing in-tandem with MDR modules possibly extend benefits to this *Salmonella* variant, promoting prolonged pig colonisation and environmental survival in metal-containing feed and farm environments.

1.3.6 Salmonella virulence factors

Horizontal gene transfer via temperate phages likely accounts for the majority of *S. Typhimurium* lineage-specific accessory genome elements rather than plasmids or non-phage chromosomal genes. Phage mediated transfer of virulence factors is well documented in *Salmonella spp,* and several virulence genes, such as *Sop* have been identified in *Salmonalla* specific prophages (Rabsch et al, 2002). *S. Typhimurium* 4,[5],12;i:- acquired the atypical *Salmonella* virulence factor, *SopE* via horizontal gene transfer from lysogenic conversion by mTmV bacteriophage (Tassinari et al., 2020). *Salmonella pathogenicity islands (SPIs)* are necessary genetic virulence elements for *in viv*o infection, *SPI-1* concerns cell invasion and post-invasion processes. The *Inv/Spa* system found in *SPI-1* supports *Salmonella* entry into nonphagocytic epithelial cells, influences host cell cytoskeleton rearrangements and promotes the production of proinflammatory signals (Bakshi et al., 2000). >40 *SPI* encoded effectors interact with the type III secretion system (T3SS) apparatus contained within both *SPI-1* and *SPI-2* (Yang et al., 2021). *SPI-1* T3SS is associated with early infection and bacterial invasion whereas *SPI-2* T3SS translocates

Salmonella effector proteins into host cytosol eliciting various responses post translocation, including modulating host inflammation and immunity as well as promoting bacterial survival (Kim et al., 2018).

Vital *Salmonella* related SPI-T3SS effectors include *SopE/SopE2 homolog*; an effector for *Salmonella* invasion, activator of Rho-GTPase which facilitates cytoskeleton rearrangements, entry into epithelium cells and activation of pro-inflammatory transcriptional responses, *SopB*; signal transduction kinase of Akt, a protein kinase B that promotes *Salmonella* survival, *SopD*; promotes membrane fission and non-specific uptake of extracelluar fluids, *SptP*; transcriptional inhibitor of NF-κB dependent genes and *AvrA* inhibits NF-κB NF activation (Ioannidis et al., 2013). TTSS translocation *SopE*/*SopE2* and *SopB* effector subset leads to pro-inflammatory effects, which allow *Salmonella* to hijack host inflammatory responses in order to outcompete resident microbiota populations for localised nutrient access, an essential strategy for fast replication common to enteric pathogens (Sun *et al.,* 2020).

SPI-2 orchestrates intracellular survival mechanisms for instance, *SpvB*, a plasmid encoded cytotoxic protein is secreted by SPI-2 T3SS to enhance *Salmonella* intracellular replication (Garcia-Gil et al., 2018; Yang et al., 2021). Similar to *SptP* and *AvrA, Salmonella secreted factor L*, encoded in the *SPI-2* can suppress NF-κB activity via degradation and the production of kinase inhibitors such as IκBα. Interrupting signalling pathways with several effectors targeting NF-κB functionality among other kinases advantageously attenuates host inflammation and promotes *Salmonella* survival (Yang et al, 2021)

1.4 BACTERIA PATHOGEN WITH PORCINE HOST SPECIFICITY: *STREPTOCOCCUS SUIS*

1.4.1 S. suis epidemiology, prevalence and impact of pathogenic strains in pigs.

Streptococcus bacteria astonishingly consists of some of the most invasive bacterium of host mucosal membranes. The *Streptococcus* genus is segmented into >100 species and various sub-species, and gene clustering phylogenetics typically designates eight distinct species groups; *sanguinis, mitis, anginosus, downei, bovis pyogenic, mutans, and salivalius* and various ungrouped species which *Streptococcus suis* is classified under (Yumoto et al., 2019). >60% of *Streptococcus* species have been associated with invasive infections of human and animal hosts alike (Krzyściak et al., 2013). In ode to invasive capacity, many *Streptococcus* species become indigenous and early colonisers of host physiological flora with the confines of oral cavities, skin, throat upper respiratory and intestinal tracts of host. For this reason, *Streptococcus* species are often recognised as possible pathobionts, as weakened host immunity systems can lead to the manifestation of opportunistic infections.

Streptococcus suis is a heterogenous grouping of facultatively anaerobic encapsulated, spherical or ovid Gram-positive cocci-shaped bacterium (1.0–1.5 μm) which exist as pairs or short chains (Goyette-Desjardins et al., 2014; Dutkiewicz et al., 2017). *S. suis* can be classified under the Lancefield R, S and T groupings and generally these bacteria demonstrate α-hemolysis when growing on selective media plates supplemented with preferentially sheep, bovine or horse blood. A totality of 35 *S. suis* serotypes are recognised, specifically serotypes 1-34 and ½ are described, and distinction primarily relies upon capsular typing via assessing antigenicity of capsular polysaccharides (CPS) (King et al., 2002). PCR-based molecular serotyping is often exploited for *S. suis* isolate identification, nonetheless due to evolutionary divergence

PCR strategies are limited in the capacity to distinguish between clinically important serotypes. For instance, shared CPS gene loss can result in difficulty discerning between serotypes 1, 2, 14 and ½ (King et al., 2002; Goyette-Desjardins et al., 2014). Many of the aforementioned *S. suis* serotypes are commonly isolated from swine and thus the inability of most approaches to distinguish between these serotypes is particularly concerning (Goyette-Desjardins et al., 2014; Zimmerman et al., 2017).

Consequently, capsular serotyping is often coupled with multilocus sequence typing (MLST) for *S. suis* characterisation. *S. suis* MLST entails sequencing seven housekeeping genes primarily; *cpn60* (chaperonin), *dpr* (supposed peroxide resistance protein), *recA* (homologous recombination factor), key enzyme encoding genes; *aroA* (5-enolpyruvylshikimate 3-phosphate synthase), *thrA* (aspartokinase/homoserine dehydrogenase), *gki* (glucose kinase) and *mutS* (DNA repair enzyme) (King et al., 2002). MLST offers enhanced typing discrimination and has resulted in over >3,000 registered *S. suis* sequence types (ST) at present and numerous clonal complexes (CC), nonetheless invasive strains are limited to certain STs/CCs (Dutkiewicz et al., 2017)

Diseased pigs can be subject to infection from *S. suis* serotype 1 (ST1) through to serotype 9 (ST9), 1/2, 14, 16, 21 24 and 31 (Vötsch et al., 2018; Lunha 2022). In the extant of literature, serotypes 1 (ST1) and 2 (ST2) are described as the most virulent and epidemiologically prevalent serotype in both diseased humans and pigs globally (Vötsch et al., 2018). *S. suis* serotype 9 systematic carriage was identified in both healthy and diseased pigs across several European countries (Segura et al., 2017). Whereas *S. suis* serotypes ST25, ST28, ST7 and ST101-104 have been underlying sources of major outbreaks in pigs across the Asia-pacific (Vötsch et al., 2018). Nonetheless, geographical discrepancies do affect the prevalence of certain serotypes.

Even isolates of certain serotypes can exhibit phenotypic and genotypic disparities additionally, numerous un-typable *S. suis* exist (Goyette-Desjardins et al., 2014; Vötsch et al., 2018).

1.4.2 S. suis in pig farm environment and clinical signs of disease in pigs

S. suis was initially reported by veterinarians in 1954, due to outbreaks of meningitis, septicaemia arthritis amongst piglets (Hughes et al., 2009) and ever since *S. suis* has been recognised as one of the major and economically impactful porcine pathogens. Pigs with underdeveloped immune facilities are particularly susceptible to *S. suis* infection, and this can occur to pigs of any age however, mortality post-weaning is most common (Segura et al., 2016). Nonetheless, morbidity associated with *S. suis* infections of swine is generally ≤5% but within farms with poor biosecurity and hygiene this can exceed >50%, and prompt administration and availability of appropriate treatment can reduce mortality by 15% compared to untreated pig herd cohorts (Dutkiewicz et al., 2017)

As with most bacterial pathogens of pigs, the dissemination of *S. suis* can occur by both horizontal; (transmission through carrier members of pig herds) and vertical mechanism (sow and piglet bacteria flora exchanges). In the farm environments horizontal transfer is most common route; *S. suis* pathobionts initiate carrier states in hosts, with some pig farms reporting carriage rates as high as ~80% (Hughes et al., 2009). Animal stress during periods such as weaning or co-infections with other viral/bacteria pathogens and highly intensive farming systems all can predispose pigs to develop clinical *S. suis* infections (Hughes et al., 2009; Obradovic et al., 2021). Poly systemic infection caused by *S. suis* is characterised by several disease symptoms including; pneumonia, meningitis, arthritis, endocarditis and septicaemia which can lead to streptococci septic shock (Helke et al., 2015; Zimmerman et al., 2017). Head

tilting, ataxia and convulsing in pigs are clinical indications of *S. suis* neurological infections or meningitis, whilst septicaemia is associated with signs of malaise and anorexia and arthritis if often detected after the emergence of lameness. Whereas, persistent coughing can be a clinical sign of *S. suis* respiratory disease (Helke et al., 2015; Gottschalk and Segura, 2019).

The complicated mechanisms which dictate respiratory disease in pigs often disrupts concluding *S. suis* as the primary cause of disease manifestation, especially as invasive bacteria strains are classified within the expansive respiratory disease complex of pig pathogens (Obradovic et al., 2021; Gottschalk and Segura, 2019). *S. suis* remains as the only Gram-positive bacteria defined within the porcine respiratory disease complex and is the secondary most infectious agent detected in pigs (Petrocchi-Rilo et al., 2021). Co-infection of pigs with *S. suis* and bacteria pathogens; *P. multocida, B. bronchiseptica A. pleuropneumoniae* and/or *M. hyopneumoniae* can lead to various manifestations of suppurative or fibrinous bronchopneumonia (Obradovic et al., 2021). Additionally, the co-infection of highly virulent *S. suis* serotypes and viruses such as PRRSV has been identified in pigs (Obradovic et al., 2021).

1.4.3 S. suis pathogenicity and virulence factors

>60 virulence associated, and host-specificity factors play vital roles in the clinical manifestation of *S. suis* infection and diseases in pigs and other animals. The knowledge-base underpinning the pathogenicity of *S. suis* is segmented into four groups of virulence-associated factor genes which encode or facilitate; *i)* membrane surface, biofilm formation or secreted elements; *SspA*, *cps*, *epf* and *fbp* among others; *ii)* enzymes and peptidase; *gdh*, *gapdh* iii) transcription factors (*ccpA*, catabolite control factor) or *iii*) regulatory, transporter or secretion systems (Baums and

Valentin-Weigand, 2009; Fittipaldi et al., 2012; Krzyściak et al., 2013). Capsular polysaccharide (*cps), e*xtra-cellular protein factor (*epf), f*ibronectin binding protein (*fbp)* and glutamine dehydrogenase-encoding (*gdh)* genes, are most well-characterised *S. suis* virulence factors.

Host-cell adhesion requires receptor-ligand interactions between the host and bacterial cell components. Fibronectin binding protein, extra-cellular protein factors (including muramidase released protein) are proteinaceous virulence associated factors, which facilitate *S. suis* colonisation, and adhesion to host cells or the complex extracellular matrix rich in collagen, fibrogen and fibronectin (Fittipaldi et al., 2012). Surface Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and suilysin are virulence-associated factors of *S. suis* serotype 2 and support the adhesion to epithelial cells (Fittipaldi et al., 2012).

Aforementioned virulence factors are well-studied due to the *S. suis* serotype 2 research bias which exists. *S. suis* P1/7 and other serotype 2 stains are highly virulent and geographical divergencies occur within this group of microorganisms. For instance; Surface-associated subtilisin-like serine protease (*SspA*) and zinc-binding lipoprotein virulence factors have all been isolated from serotype 2. *S. suis* P1/7 pig isolates in Northern America countries such as Canada whereas, *S. suis* P1/7 carrying suilysin thio-activated hemolysin (*sly*) were initially identified in the Netherlands and remain prevalent in pigs across Europe (Baums and Valentin-Weigand, 2009).

1.4.4 S. suis; treatment, emerging resistance and zoonotic potential,

Conventionally, β-lactams, tetracyclines, sulfonamides, ceftriaxone fluoroquinolones (enrofloxacin) and macrolides are exploited for the treatment of pathogenic *Streptococci* infections in both pigs and humans. Petrocchi-Rilo *et al.,* (2021) reviewed *S. suis* antimicrobial resistance profiles from 147 intensive Spanish pig farms during 2019-2020. *S. suis* isolates demonstrated a broad range of antimicrobial resistance profiles against ribosomal targeting and folic acid disrupting; Tulathromycin (<64 μg/mL) and Sulphadimethoxine (>256 μg/mL) respectively. Further examination verified the current arsenal of antibiotics used for *S. suis* infection are enrofloxacin (2 μg/mL), penicillin (<2 μg/mL) ampicillin (0.5 μg/mL) or ceftriaxone (8 μg/mL) (Petrocchi-Rilo *et al.,* 2021) *S. suis* biofilm encapsulation supports evading antimicrobials and enzymes which modify key membrane components; lipoteichoic acid (LTA)-d-alanylation and peptidoglycan N-deacetylation are exploited by S. suis to circumvent antibiotic activity (Fittipaldi et al., 2012).

Similar to *S. Typhimurium*, the zoonotic transmission of *S. suis* relies primarily on the consumption of contaminated pork meat, and increasingly this threatens both animal and human health. Especially as, zoonotic transmission of *S. suis* carrying resistance profiles against tetracyclines and macrolides have been widely reported globally for numerous years (Lunha et al., 2022). The prevalence of *S. suis* human infections and subsequent outbreaks are associated with occupational-based and gender factors; *i.e.* male farmers, veterinarians, butchers, abattoir and food processing workers have a greater propensity of infection risk (Hughes, 2009). This is especially the case in global regions such as Asia, where approximately ~90% of the global *S. suis* cases through 2002-2013 were registered (Hughes, 2009; Dutkiewicz et al., 2017).

1.5 NON-ANTIBIOTIC ALTERNATIVES IN PIG FARMING: CONTROLLING *SALMONELLA* AND *S. SUIS*

Salmonella prevalence in pigs can be curtailed via long-term surveillance programmes and alternative therapeutic, farm and slaughtering management strategies. Controlling *Salmonella* prevalence at farm-level, the risk posed to productivity and human health, fundamentally relies upon both individual and herd-specific health, most critically preventing *Salmonella* dissemination in latter production stages (*e.g.*, rearing to fattening pigs), as well as high standards of farm, transport and slaughter hygiene management (De Lucia and Ostanello, 2020). Focus has additionally been placed on optimising pig feeding practices, as feed is a well-recognised introductory source of *Salmonella* in farm settings. Feeding *Salmonella*-free feedstuffs to pigs was predicted to reduce *Salmonella* prevalence by 10-20% and 60-70% in EU member states with high and low *Salmonella*-positive prevalence respectively (EFSA, 2016). Pig herds reliant on dry feeds are associated with high levels of *Salmonella* infection, the addition of water to feed lessens this risk however, fermented or acid feed by-products are typically the most beneficial for lowering *Salmonella* seroprevalence in pigs (Bahnson et al. 2006; Gavin *et al.*, 2018). However, conflicting evidence exists on the economic viability of feed-related interventions for *Salmonella* control as regulating *Salmonella* status during pig processing further squeezes cost-profit ratios for individual farms (Niemi et al 2019; Gavin et al., 2018).

Phage therapy is an alternative means of nonantibiotic bacterial disease control that has been shown to reduce enteropathogenic colonisation and shedding in experimental pig studies. Saez *et al.*, (2011) explored administrating a phage cocktail via feed to young pigs at 2hr and 4hrs post *S. Typhimurium* challenge. Faecal shedding and ileal concentrations of *Salmonella* in the phage-treated group at necropsy were

significantly reduced. Desiree *et al.,* (2021) conducted a meta-analysis of 19 pig studies (576 observations) where bacteriophages isolated against *S. Typhimurium* or *E. coli* were delivered to pigs before or after bacterial challenge. Despite considerable data heterogeneity the general trend saw concentrations of both pathogens significantly reduced in phage-treated pigs challenged with *Salmonella* or *E. coli* (Desiree et al., 2021).

Probiotic strategy for pigs can putatively promote stable protective gut endogenous flora and reduce *Salmonella* shedding. Prior to orally challenging weaning piglets with *S. Typhimurium* PT12 (1×10^8 CFU, 3 days) Casey *et al.*, (2007) implemented the pre-administration of milk probiotic fermentate or suspensions containing a cocktail of five lactic acid bacteria strains delivered at ~ $4x10^9$ - $4x10^{10}$ CFU/day; *Lactobacillus murinus* (2) *Lactobacillus* salivarius subsp. *Salivarius* (1), *Lactobacillus pentosus* (1), and *Pediococcus pentosaceous* (1). Probiotic treatment reduced the elicitation, severity and duration of diarrhea symptoms and the mean prevalence of *Salmonella* in faecal samples significantly reduced by ~3-fold in the probiotic treatment group (+15 days post infection) compared to the control (Casey et al., 2007). Lactic acid bacteria can persist within pig intestinal tracts, preventing epithelial invasion and these bacteria behave as humoral immune adjuvant, stimulating targeted *Salmonella* antibody production. Naqid *et al.,* 2015 demonstrated the provision of either prebiotic lactulose (1% w/w feed) and water based probiotic feed mix containing *Lactobacillus plantarum* B2984 ($\sim 1 \times 10^{10}$cfu/pig/day) or inclusion of both in the diets of weaning piglets significantly enhanced antibody (IgM, IgA and IgG) responses to *S. Typhimurium* SL1344 and ameliorated diarrhoea symptoms.

Gonçalves de Oliveira Moura, er al., (2021) demonstrated *S. Choleraesuis* and *S. Typhimurium* bacterins (inactivated) vaccines administered to forty weaning piglet

was partially protective against *Salmonella* colonisation and reduced shedding. Farzan and Friendship, (2010) observed similar results in nine cohorts of ≤350 pigs when testing *S .Typhimurium* bacterin and commercially live *S. Choleraesuis* vaccinations. *Salmonella* shedding reduced twice as much compared to the control group over the entire farrow-to-finish production operation (Farzan and Friendship, 2010). Cross-protective vaccine approaches for *S. suis* and other *Streptococci* are largely impaired by the heterogeneity of this species and serotype representation. Nevertheless, field trial vaccines targeting with singular serotype targets 1, 2, 1/2, 3, 9 and 14 have been explored however *S. suis* vaccines remain commercially unavailable thus, further necessitating the demand for novel antimicrobials for *S. suis* control (Goyette-Desjardins et al., 2016). Nevertheless, Wisner *et al.,* (2021) reported from a meta-literature analysis of 441 clinical vaccination trials on nursery pigs that one-third of vaccine trails failed to report clinically important outcomes.

1.6 ANTIMICROBIAL PEPTIDES

1.6.1 Introduction to AMPs: generalised characteristics, source origins and classification

Antimicrobial peptides (AMPs) have emerged as putative alternatives for treating bacterial infections in light of the ever-growing antibiotic resistance crisis caused by the widespread usage of antibiotics in human, veterinary medicine and animal agriculture. AMPs are evolutionary conserved oligopeptides ubiquitously produced in all life forms. Generally, AMPs are considered as short (≤100 amino acids), cationic (+2 to +11 overall positive net charge), hydrophobic (≥ 50% hydrophobic amino acids) peptides that fold into a variety of diverse structures exhibiting amphipathic conformations (Shai, 2006; Bahar and Ren, 2013; Mahlapuu *et al.,* 2016). However, several anionic AMPs (AAMPs) rich in negatively charged amino acids, such as

aspartic acid and glutamic acid, can be encoded genetically or form as proteolysis fragments (Huan et al., 2020 and Zhang et al., 2022). AAMPs are theorised to facilitate their interaction with targets by utilising metal ions and the negatively charged components of microbial membranes to form cationic salt bridges (Zhang et al., 2022).

The diversity of natural AMPs stifles classification especially, as minor peptide sequence variations often lead to significant differences in AMP structure and activities (Khamis et al., 2018). AMP quadripartite classification encapsulates; source, mechanisms of action, structural characteristics and amino acid rich species. The Antimicrobial Peptide Database(ADP3) is one of the largest AMP repository databases. ADP3 exemplifies the ubiquity of AMPs, with other 3283 natural AMP repository sequences originating from across the six life kingdoms; Bacteria (11.30%), Archaea (0.15%), Protists (0.24%), Fungi (0.67%), Plants (11.00%), Animals (74.05%, synthetic peptides accounts for 2%) (https://aps.unmc.edu/, figures updated July 2022).

Early discoveries of animal-originating AMPs include: rabbit leukocyte defensins (Hirsch, 1956), amphibian epithelia derived bombinin (Kiss and Michl, 1962), lactoferrin isolated from cow milk (Groves, Peterson and Kiddy, 1965) and human leukocyte AMPs (Zeya and Spitznagel, 1963). Tissues and organs that are frequently exposed to airborne pathogens are cultivars for large numbers of AMPs, for example; over 300 different AMPs have been sourced from frog skin (Bahar and Ren, 2013). AMPs produced by bacteria function to kill other bacteria allowing AMP producing bacteria to compete and usurp ecological niches (Mahlapuu *et al.*, 2016). Variations within AMP families consequently lead to functional divergence of isoforms to extend specificity, antimicrobial spectra of activity or to acquire novel immune functionality (Schmitt 2016).

Cathelicidins and defensins are the main families of mammalian (vertebrates) AMPs. Depending on disulfide bond position defensins can be classified into a-, β- and θ-defensins. Human host defense peptides (HDPs) are innate immunity effector molecules exhibiting broad-spectrum antimicrobial activity exerting exclusively lytic effects on microbes (Blyth et al., 2020). HDPs such as cathelicidin LL-37 fluctuate in their expression during different stages of growth whereas Casein201, derived from breast milk, exhibits different levels in preterm and term human colostrum (Huan et al., 2020). Colostrum across mammals are important sources of AMPs, formed primarily through the enzymatic hydrolysis of milk components, to which several well-characterised AMPs have been derived such as lactoferrin, α-lactalbumin and β-lactoglobulin AMPs. Magainin, equally well-characterised, is an example of an AMP derived from the skin secretions of frog genera such as Xenopus (Mangoni and Casciaro, 2020). Wang, (2020) identified that amino acids leucine (L), alanine (A), glycine (G) and lysine (K) were the most abundant in amphibian AMPs and variation of L, A, G and K modulated peptide activity in different HDPs found in frog spp. across Asia, Europe and North America.

1.6.2 Classification of AMPS by structure

AMPs can adopt a variety of structural conformations depending on peptide sequence, inherent physiochemical properties and physiological conditions. As demonstrated in **Table 1.5.** AMP classification predominately centres upon structure. AMPs tend to be α-helices or complexes of 2-4 β-sheets however, other structures: loop and extended have been reported (Jenssen 2006; Bahar and Ren, 2013). To retain amphipathicity a minimum length of ≈22 mers is required for α-helical AMPs to transverse lipid bilayers whilst, for AMP adopting a β-sheeted conformation ≈8 mers are required (Shai, 2002).

Table 1.5 Classification of AMPs based on structural conformations.

Classification of 3D structures of antimicrobial peptides in APD3 AMP database		
AMP Structure conformation	**Peptide Count**	**Exemplar AMPs**
α- helical	**456**	Cathelicidin LL-37, Cecropin, dermcidin, magainin,
β – sheet	**123**	Humanalpha defensins (HNP-1, HNP- 4, and HD-5), plant kalata B1,
α β	**64**	Human beta defensins (HBD-1, HBD-4), Drosomycin
Non-αβ (loop or extended structures)	**12**	Indolicidin, tritrpticin, drosocin, nisin A
Data retrieved from ADP3 (https://aps.unmc.edu/, updated 2022). *Note <50% of AMPs on APD3 are annotated with structural information.		

1.6.3 Biological functions of AMPs: antimicrobial activity immune modulating properties

In higher organisms AMPs are often denoted as "HDPs" that form an integral part of the innate immunity. Moreover, AMPs are present in the epithelial barriers of multicellular eukaryotic organisms and they exhibit broad spectrum activity against Gram-negative and Gram positive bacteria, fungi, unicellular protozoa and viruses (Mahlapuu *et al.*, 2016). HDPs have demonstrated *in vitro* the ability to recruit immune cells, support wound healing and the initiation of cytokine production (Blyth et al., 2020). Defensins are a highly convergent group of small, disulphide-rich, cationic HDPs with highly diverse sequences and structures expressed in essentially all eukaryotes. Defensins display a broad range of antimicrobial activity at relatively low micromolar concentrations, and they possess diverse roles including ion channel perturbation, enzyme inhibition, toxicity functions and cell signalling facets for example immune cell recruitment and self-nonself recognition systems (Makarova et al., 2018), Rumio *et al.*, (2004) additionally exemplified that *in vivo* stimulation of

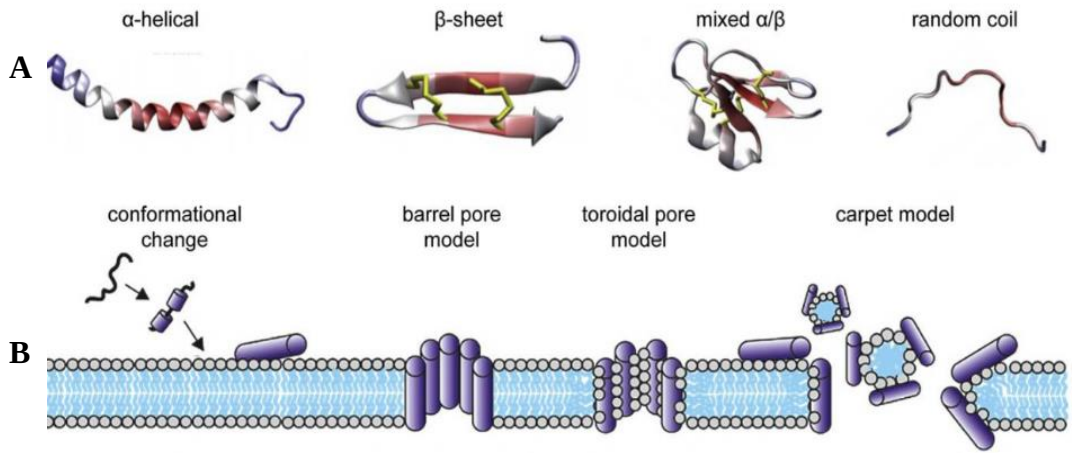
Toll-Like Receptor (TLR) 9 and 3 triggered Paneth cell degranulation and the excessive release of antimicrobials including high quantifies of AMPs, which inferred protection against *S. Typhimurium* infection in mice. Additionally, the absence of paneth cell α-defensin cryptdins in mice promoted high susceptibility to orally administered pathogens (Lazzaro et al., 2020).

1.6.4 AMP mechanisms of action

The aforementioned AMP properties supplants these peptides as therapeutic alternatives for treating microbial infections in human and animal medicine. Especially considering eukaryotes have zwitterionic lipids in their cell membranes that weakly interact with AMPs (Glukhov et al., 2005). However, bacterial membranes are negative due to phosphate groups of lipopolysaccharides (Gram-negative) and lipoteichoic acids (Gram-positive). Antibacterial AMPs mode of action centres on the affinity to interact and disrupt the integrity of anionic bacterial membranes (Harris and Pierpoint, 2012). Overall net charge and hydrophobicity vitally underpin AMP bacterial membrane binding affinity and interactions; cationic AMPs (CAMPs) exploit electrostatic attraction to interact with anionic bacteria membrane components such as LPS, LTA (Zhang and Yang, 2022). This attachment leads to membrane destabilisation and permeabilisation whereupon hydrophobic amino acids residues and polar side chains enable bilayer penetration. Well established AMPs such as defensins, magainins, melittin and LL-37 are examples of transmembrane pore forming AMPs (Pushpanathan, Gunasekaran and Rajendhran, 2013). Contrastingly, AMPs such as buforin II, indolicidin, dermaseptin, HNP-1 and pluerocidin lead to the death of cells by disrupting cell membranes as they translocate across to inhibit intracellular processes such as DNA and protein synthesis (Pushpanathan, Gunasekaran and Rajendhran, 2013).

Several models exist to explain the complexities of AMP-bacteria membrane interaction (**Figure 1.1**). The leading scientific propositions include barrel-stave, carpet (detergent like), toroidal pore, membrane thinning and aggregate models (Jenssen et al., 2006; Liming Lin et al., 2021). Under aforementioned AMP models, membrane dysfunction is initiated by varying intermediaries including; transient channel formation, micellarisation, membrane dissolution or translocation across the membrane.

Figure 1.1 AMP structural conformations and membrane modes of action.
(Liming Lin et al., 2021);
(A) Four main structural conformations for AMPs, **(B)** Four main AMP modes of action



The "carpet-like" model entails an initial electrostatic interaction where a micelle α-helical AMP clusters align parallel to lipid bilayers and engross the membrane. Following penetration of the lipid bilayers, α-helical AMPs conform to "wormhole" pore forming secondary structures which lead to the prompt lysis of susceptible bacteria (Jenssen et al., 2006; Bahar and Ren, 2013;Liming Lin et al., 2021). Under the barrel-stave model transmembrane pore AMPs, α-helical amphipathic AMPs position as touching pillars or "staves" which penetrate across the transmembrane due to hydrophobic non-polar side chains grappling with phospholipids orientating hydrophilic side chain regions inwards forming aqueous pores (Jenssen, Hamill and Hancock, 2006 ;Lin et al., 2021). According to the toroidal

model, AMPs operate under a similar mechanism of transmembrane pore formation except AMPs align perpendicularly unlike the adjacent staves positioned within close proximity theorised by the barrel-stave. Rather under the toroidal model, AMPs operate using an multipore state which relies on integrating within the membrane and flipping inward by attaching to hydrophobic lipids (Lin et al., 2021). The aggregate channel model similarly proposes AMPs possess and utilise their affinity for hydrophobic lipid heads, and upon attachment to bacteria membranes AMPs adopt clusters forming unorganised aggregates which across the membrane lead to ion leakage and subsequent lysis (Jenssen, Hamill and Hancock, 2006 ;Lin et al., 2021).

1.6.5 Clinical applications of AMPs

The most clinically relevant examples of CAMPs include polymyxins, nisin and gramicidin among others. Several AMPs have transitioned through clinical trial development and some of the most prominent examples include Pexiganan which exhibited promising *in vitro* broad-spectrum activity. However FDA approval was denied as the efficacy of Pexiganan was ineffective in comparison with existing antibiotics used for foot ulcer treatment (Goldstein et al., 2017). Emphasising novel AMPs must surpass antibiotic alternatives in activity to be clinically and commercially viable. Omiganan was originally developed to limit catheter contamination and represents an example of exploiting synthetic analogues of characterised natural AMPs (*i.e.* bovine neutrophil indolicidin) (Niemeyer-van der Kolk et al., 2020). AMP therapy clinical trials have been halted at different stages NVB-302 (phase I), POL7080 (phase II), Iseganan, Omiganan, Surotomycin at phase III (Kosikowska and Lesner, 2016; Greber and Dawgul, 2017; Sierra et al., 2017).

At present there are no FDA approved synthetically derived AMPs in the market, especially for use in veterinary medicine. Fundamentally, synthetic peptides are often

*x*5 - 20 times more expensive to develop than conventional antibiotics (Zambrowicz et al., 2013) and peptide-based drugs can suffer from proteolytic instability and degradation *in vivo*, among other shortcomings such as being toxic to mammalian cells (Otvos and Wade, 2014). Additionally, the commercialisation of peptide drugs are limited due to the cost effectiveness of manufacturing, especially as synthetic peptide production costs typically range between USD $300–500 per gram (Bray, 2003; da Costa et al., 2015;Molchanova et al., 2017).

1.7 PEPTIDE LIBRARY CONSTRUCTION

1.7.1 Established peptide library mutagenesis strategies

Researchers continue to use both rational and directed evolution strategies to accelerate the discovery of high-value peptide or protein variants. Despite differences in the desired peptide length most modern combinatorial peptide libraries produce a complexity 10^{8}-10^{11} (Larsson *et al.*, 2002; Castel *et al.*, 2011). Theoretically, it is possible to achieve a library diversity of $10^{12\text{-}13}$. Nonetheless, high library diversity imposes higher screening workloads and does not necessarily improve the chances of isolating desired peptide variants (Larsson *et al.*, 2002; Castel *et al.*, 2011). The fortuitous discovery of desired peptide variants depends upon library complexity but additionally concerns starting library quality. The pursuit of maximum peptide library quality, defined in terms of frequency of unique peptide variants with desired functionality, fundamentally relies on the efficiency of a variety of mutagenesis techniques. The most frequently utilised mutagenesis approaches for peptide libraries are; polymerase chain reaction(PCR) amplification with degenerate primers, error-prone PCR, saturation mutagenesis and DNA shuffling (Reetz, Prasad and Carballeira, 2010).

In spite of numerous successes in the application of these mutagenesis methods to create fully randomised libraries three disadvantages are sustained; (i) truncated peptides due to unwanted integration of stop codons, (ii) high screening workload and, (iii) conflicts between the theoretical and achieved library complexity (Larsson *et al.*, 2002). To reduce these drawbacks directed evolution of peptides is becoming increasingly popular, and numerous efficient semi-rational approaches systemised under the branch of iterative saturation mutagenesis(ISM) have been developed (Kille *et al.*, 2013). Unlike fully randomised mutagenesis, ISM requires to some extent knowledge of the structural, chemical or mechanistic characteristics of the peptide of interest or the assimilation of this information through *in silico* analysis (Tang *et al.*, 2018). Archetypally, small numbers of appropriate amino acid sites in peptides are initially targeted for mutagenesis. These preliminary libraries often serve as gene templates for further iterations of saturation mutagenesis targeting other sites (Lane and Seelig, 2014; Currin *et al.*, 2015). Common ISM approaches include; Combinatorial Active-site Saturation Test (CAST), Structure-based Combinatorial Protein Engineering (SCOPE) and Gene Site Saturation Mutagenesis (GSSM) (Chen *et al.*, 2012; Kille *et al.*, 2013).

1.7.2 Degenerate primers and codon schemes

Generating high-quality libraries using saturation mutagenesis is still experimentally challenging. As factors such as; DNA sequence, G+C content, primer quality and randomisation scheme all affect the quality of the resulting peptide library (Acevedo-Rocha, Reetz and Nov, 2015). Whilst some aforementioned factors are inherent to the DNA sequence, and thus not straightforwardly modified, other factors such as the randomisation scheme can be revised. Consequently, when implementing mutagenesis techniques to design peptide libraries that are less complex but of higher

quality, the selection of an appropriate randomisation scheme is imperative (Acevedo-Rocha, Reetz and Nov, 2015). The unaltered genetic code (NNN, where N = A/T/G/C bases) is redundant and 18/20 canonical amino acids are encoded by multiple synonymous codons to which 61 sense codons exist (Spencer and Barral, 2012). For example, arginine, serine and leucine amino acids are encoded by ≤6 different sense codons however, methionine and tryptophan are encoded by a singular sense codon (Spencer and Barral, 2012; Kille *et al.*, 2013). The contrasting disproportionality in synonymous codons for different amino acids facilitates the under and over-representation of certain amino acid residues in peptide libraries and this disproportionality exaggerates as the number of randomisation target sites increases.

To eliminate the unequivocal bias of the unaltered genetic code (NNN) libraries are largely constructed via using degenerate primer codon schemes which restrict subsets of the genetic code, alleviating the effects of redundancy on library quality **see Table 1.6** (Kille *et al.*, 2013). Unique codons within these schemes typically, reflect the codon bias of expression hosts such as *E. coli* (Quax *et al.*, 2015). Popular NNK/NNS degenerate codon schemes encode the 20 canonical acids using 32 codons (Tonikian et al., 2007; Sieber et al., 2015). When randomising for example four amino acid sites using NNK/NNS the theoretical ratio between the most overrepresented and underrepresented amino acids declines from 1296:1(NNN) to 81:1 (Acevedo-Rocha, Reetz and Nov, 2015). Additionally, the frequency of prematurely truncated peptide variants falls from 17.5% (NNN) to 11.9% (NNK/S) (Acevedo-Rocha, Reetz and Nov, 2015). Nonetheless, NNK/NNS degenerate codon schemes still contain; 3 codons for arginine, leucine, and serine and, 2 codons for glycine, alanine, threonine, proline and valine (Kille et al., 2013).

Table 1.6 Degenerate codon schemes exploited in random mutagenesis and peptide library design

Degenerate codon	Total number of unique codons	No. of stop codons	Stop Codons	No. of amino acids encoded	Amino acids	Characteristics of amino acids encoded
NNN	64	3	TAG, TAA, TGA	20	All 20	All characteristics
NNK	32	1	TAG	20	All 20	All characteristics
NNC	16	None	-	15	A, C, D, F, G, H, I, L, N, P, R, S, T, V, Y	Mixed
NNS	32	1	TAG	20	All 20	All characteristics
NNB	48	1	TAG	20	All 20	All characteristics
NDT	12	None	-	12	R, N, D, C, H, G, I, L, F, S, Y, V	Mixed
NWW	16	1	TAA	11	D, E, F, H, I, K, L, N, Q, V, Y	Charged or hydrophobic
NVT	12	None	-	10	C, D, G, H, N, P, R, S, T, Y	Charged or/and hydrophilic (except C)
NNT	16	None	-	12	A, D, G, H, I, L, N, P, R, S, T, V	Mixed
NRT	8	None	-	8	R, N, D, C, G, H, S, Y	Mixed
NTT	4	None	-	4	F, I, L, V	Hydrophobic
DBK	18	None	-	12	A, R, C, G, I, L, M, F, S, T, W, V	Mixed
DVT	9	None	-	8	A, C, D, G, N, S, T, Y	Hydrophilic (except A and C)
RVK	12	None	-	10	A, D, E, G, H, K, N, R, S, T	Charged or/and hydrophilic (except A)
RST	4	None	-	4	A, G, S, T	Small side chains
VVC	9	None	-		A, D, G, H, N, P, R, S, T	Hydrophilic (except A)
TDK	6	1	TAG	5	C, F, L, W, Y	Hydrophobic

[a] DNA degeneracy are schematised using the International Union of Biochemistry (IUB) one letter code, where each letter represents a selection of degenerate nucleotide bases (A: Adenine, C: Cytosine, T: Thymine and, G: Guanine). Singular letter abbreviations include *(Note: / = or); N:* A/C/G/T; *S:* G/C; *K:* G/T; *D:* A/G/T; *R:* A/G; *V:* A/C/G; *W*, A/T, *B*: C/G/T *T*: T only

[b] Single letter amino acid abbreviations; G: Glycine, P: Proline, A: Alanine, V: Valine, L: Leucine, I: Isoleucine, M: Methionine, C:Cysteine, F: Phenylalanine, Y: Tyrosine, W: Tryptophan, H: Histidine, K: Lysine, R: Arginine, Q: Glutamine, N: Asparagine, E:Glutamic Acid, D: Aspartic Acid, S: Serine, T: Threonine.

[c] Some of these degenerate codon schemes are still redundant (unique codons exceeds the number of unique amino acids encoded) however, the redundancy is significantly less compared to the of NNN (unaltered genetic code) degenerate codon.

[e] Characteristics of amino acids are; hydrophobic, hydrophilic, polar, non-polar, aromatic, aliphatic, charged (acids or basic)/neutral

(Tonikian *et al.*, 2007; Reetz, Kahakeaw and Lohmer, 2008; Sieber *et al.*, 2015; Quax *et al.*, 2015)

1.7.3 Trinucleotide phosphoramidite mutagenesis (TRIM) technology

Inarguably, compared to rudimental mutagenesis techniques make advancements on controlling amino acid distribution at saturated codons. Nonetheless, various shortcomings still exist as; randomised mutation of several contiguous amino acid residues and achievement of the theoretical levels of amino acid representation are challenging to achieve. These drawbacks can be circumvented by trinucleotide mutagenesis technology, an emerging technique which relies on phosphoramidite chemistry (Gaytán *et al.*, 2009). DNA synthesis in TRIM technology requires trinucleotide phosphoramidites which are the synthon equivalents of codons. Since the early 1990's the challenge has been to discovery suitable orthogonal protecting groups that enable; synthetic preparation of trinucleotide phosphoramidites, and conversion into coupling competent building blocks for use in chemical synthesis of DNA (Virnekas *et al.*, 1994; Kayushin, Korosteleva and Miroshnikov, 2000)

Approaches to chemically synthesise trinucleotides in solution (Arunachalam *et al.*, 2012), solid phase (Jabgunde *et al.*, 2015) and in soluble polymers (Kayushin, Korosteleva and Miroshnikov, 2000) have been developed. Furthermore, a variety of phosphorylating reagents have been identified which reincorporate a phosphate group to the 5' end of trinucleotides to ensure these molecules are suitable for DNA synthesis (Suchsland, Appel and Müller, 2018). The most commercially widespread synthesis strategy for trinucleotides centres on solid supports and 4,4′-dimethoxytrityl (DMTr) as orthogonal protecting groups (Suchsland, Appel and Müller, 2018). Trinucleotide technology gene mutation can be conducted at the codon level rather than targeting individual bases (Yagodkin *et al.*, 2007) additionally, in TRIM technology sense codons can be synthesised for or, excluded from the mutagenesis reaction mixture. Consequently, codon redundancy can be perfectly optimised to allow uniform amino

acid distribution and eliminate undesirable stop codons (Gaytán *et al.*, 2009). At present the complexity of trinucleotide synthesis results in high-costs for commercial TRIM technology such as: Thermo Scientific GeneArt (Gaytán *et al.*, 2009). Commercial trinucleotide mixes are becoming readily available at lower costs however, implementing trinucleotide mutagenesis in standard laboratories does not suffice as, specialised apparatus such as DNA synthesisers are required (Gaytán *et al.*, 2009).

1.8 SCREENING STRATERGIES FOR NOVEL ANTIMICROBIAL PEPTIDE DISCOVERY

1.8.1 Phage display technology: Structure, genetics and biology of filamentous phage

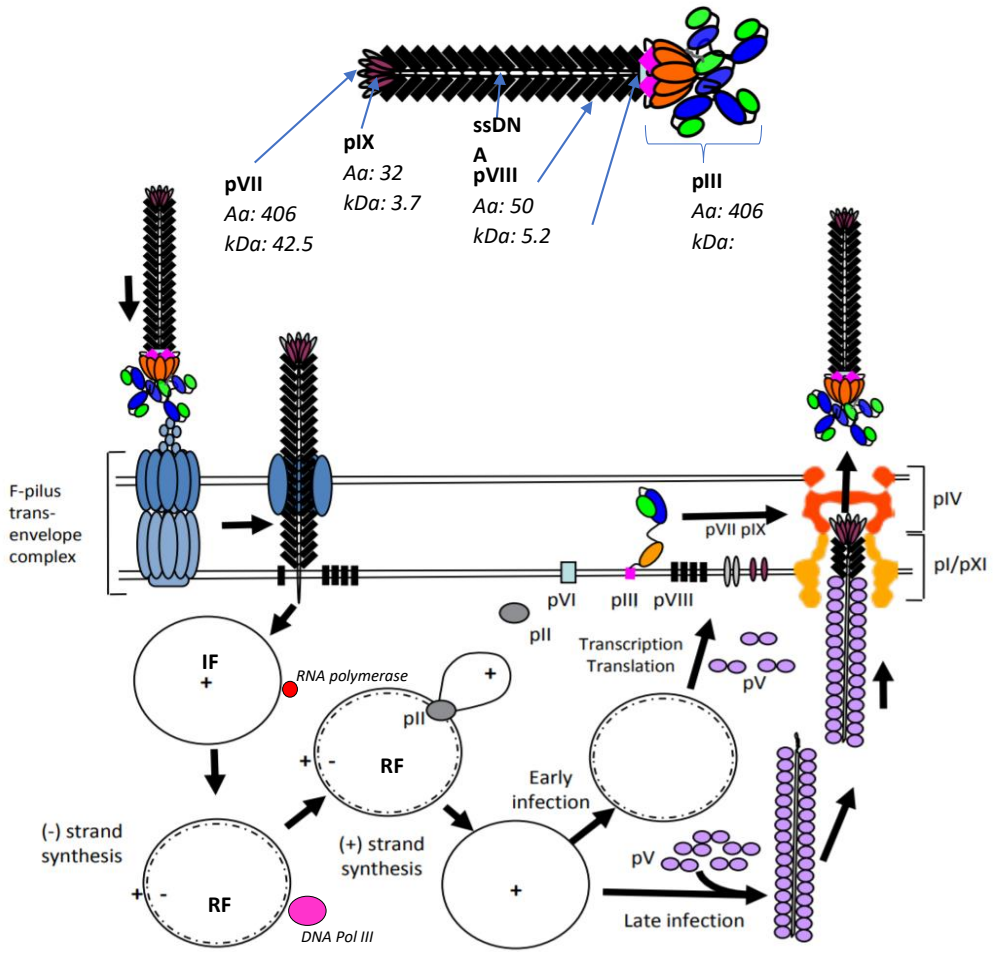
Phage display is highly reliant on filamentous Ff phages; M13, f1 and fd (Rakonjac and Bennett, 2009). Filamentous phage belong to the *Inoviridae* family and, are long rod shaped (900nm long and 7nm narrow) viruses with a single-stranded DNA (ss-DNA) genomes. (Ackermann, 2003; Rakonjac and Bennett, 2009; Ebrahimizadeh and Rajabibazl, 2014). Predominately, filamentous phage infect a multitude of Gram-negative bacteria however, infection of Gram-positive organisms has also been documented (Chopin *et al.*, 2002). Unlike lytic phages such as T7, that utilise cytoplasmic assembly and release of progeny phage via cell lysis, filamentous phage assemble in the periplasm whereupon, phage progeny are continuously extruded through the cell envelope in a process which pairs assembly with export (Ebrahimizadeh and Rajabibazl, 2014). This non-lytic lifecycle does not kill bacteria hosts but, due to the manipulative strain on bacterium machinery caused by phage, the growth rates of hosts are significantly reduced (Wilson and Finlay, 1998).

The sequenced genomes of Ff phages share 98.5% homology and, this group of phage owe their name to their dependence on the F-pilus for infection of F-conjugative plasmid carrying *Escherichia coli* strains (Rakonjac and Bennett, 2009; Tikunova and Morozova, 2009). This selective infection route is manipulated in screening methodologies like phage display. As following infection the F-pilus is depolymerised therefore, each bacterial clone can only permit infection by one phage that specifies a unique library variant (Schmitz *et al.*, 2000). M13 bacteriophage is the most widely utilised display platform due to the well characterised genetic and biological features of M13 phage. Defined simply, bacteriophages are proteomic complexes and the structural, chemical and biological functionality of these viral entities are specified by their encapsulated genomes. The M13 6407bp ss-DNA genome encodes 11 genes which can be subcategorised according to protein product function; (i) phage DNA replication protein genes *gII, gV, gX* (ii) virion particle protein genes g*III, gVI, gVII, gVIII, gIX* and finally (ii) proteins required for virion assembly are encoded by genes g*I, gIV and gXI* (Tikunova and Morozova, 2009). Furthermore, MI3 phage ss-DNA genome has an intergenic region (~570bp) which contains an origin of replication (*ori*) for synthesising both (+) and (-) DNA strands and also contained is a packaging signal which initiates phage virion assembly (Lu, Weers and Stellwagen, 2003; Tikunova and Morozova, 2009).

M13 infection firstly requires interaction between pIII C-terminal N1N2 domains with *E. coli* F-pilus and, TolA domains of the TolQRA complex (Nilsson, Malmborg and Borrebaeck, 2000; Bennett and Rakonjac, 2006). Following this, phage ss-DNA is inserted into host's cytoplasm and the events that follow are illustrated in **Figure 1.2.** Mature M13 virion structure is also summarised **in Figure 1.2**, and consists of; five molecules of each minor virion protein (pII, pVI, pVII and pIX), and 2, 700

molecules of the major virion protein pVIII (Mai-Prochnow *et al.*, 2015). Protein copy number disparities occur as the M13 genome possesses two coding regions; one of which is controlled by a strong promoter that encodes high copy number proteins and, the second coding region contains a weaker promoter that encodes for low copy number proteins

Figure 1.2 Filamentous phage M13 mature virion structure and non-lytic lifecycle
Mature M13 filamentous phage virion structure initiates infection by pIII attaching to host F-pilus, inserting phage ssDNA into cytoplasm, and dissolving coat proteins into cell envelope (Ebrahimizadeh & Rajabibazl, 2014). Host machinery synthesises negative strand using IF positive ss-DNA and RNA polymerase (red circle) at negative strand ori, generating an RNA primer for host DNA polymerase (pink circle) (Zenkin et al., 2006). Replication cycle produces replicative form (RF), used to create more RNA primers, phage proteins, progeny RF's, and ssDNA (positive strand) when pII binds to RF (Rakonjac & Bennett, 2009; Ebrahimizadeh & Rajabibazl, 2014). Proteins pII, pV, and pX mediate genome replication and phage packaging in cytoplasm. pI/pXI (yellow) and pIV (orange) form an assembly/export transport complex spanning hosts' inner and outer membrane. Virion particle proteins are inserted into the inner membrane prior to assembly. pV phage dimers encapsulate positive ss-DNA strands and are directed to pI/pXI-pIV via a packaging signal, forming progeny phage particles subsequently exported. – image adapted from Ebrahimizadeh & Rajabibazl, 2014



1.8.2 Phage display technology: filamentous phage systems

Several phage display systems have been founded on the five virion proteins however, pIII (type 3) and pVIII (type 8) are mainly exploited (Tikunova and Morozova, 2009). In type 3 and 8 systems the exogenous inserts are theoretically displayed on all the five pIII and 2, 700 pVIII molecules on a virion respectively. However, in reality recombinant phage protein fusions do not represent the totality of the available molecules displayed on respective phage. Unequivocally it is recognised that type 3/8 systems peptides ranging within the length of 6 -45 mers can be functionally displayed on the surface of bacteriophage (Rasooly and Herold, 2009). Larger exogenous inserts can lead to abnormal phage assembly or weakened cell infectivity. Host proteolytic enzymes can remove the disruptive foreign peptides encoded on progeny phage (Tikunova and Morozova, 2009). In order to negate this limitation a plethora of specialised vector systems can introduce recombinant and wild-type pIII and pVIII proteins thus, restoring recombinant phage assembly and infectivity.

Type 33 and 88 systems carry the recombinant gene and wild-type phage gIII and gVIII genes respectively (Enshell-Seijffers, Smelyanski and Gershoni, 2001; Hamzeh-Mivehroud *et al.*, 2013). Whereas type 3+3 and 8+8 utilise the coinfection of a phagemid vectors and "helper" phages. In these systems wild type gIII and gVIII are introduced to infected *E. coli* via helper phages (Tikunova and Morozova, 2009). Whilst, phagemids (specialised Ff phage derived vectors) introduce the recombinant gene and contain; *ori* of a bacterial plasmid, selection markers (*i.e.* antibiotics resistance genes), Ff phage intergenic genome region, restriction enzyme sites and a signal peptide DNA segment proceeded by a promoter (Qi *et al.*, 2012).

The type 88/8+8 vector systems are multivalent as both wild-type and recombinant pVIII proteins are visible on phage particle surfaces. As pVIII is a major virion protein typically around ≤700 exogenous molecules are present (Sidhu *et al.,* 2005). Phage display researchers manipulate this high valence to identify ligands with low affinity (Sidhu *et al.,* 2005; Tikunova and Morozova, 2009). Contrastingly, type 33/3+3 vector systems are effectively monovalent as 0-5 copies of the recombinant proteins are present on phage particle surfaces (Tikunova and Morozova, 2009). Unsurprisingly, due to the vital role of pIII during infection initiation, typically ~10% of the phage population carry one recombinant protein and, the percentage with 2 or more recombinant protein copies is considerably lower (Tikunova and Morozova, 2009). This low level of valence contributes to the restrictive avidity of the pIII systems nonetheless, this in turn enables the selection of high affinity ligands (Sidhu *et al.,* 2005; Tikunova and Morozova, 2009).

1.8.3 Peptide phage display libraries

Traditionally peptide library technology exploits peptide-encoding nucleic acids. More specifically, DNA-encoded peptide libraries are frequently employed due to the lowered cost in oligonucleotide synthesis and, the compatibility with screening technologies such as phage display (Larsson *et al.,* 2002). The diversity of peptide phage libraries available commercially derives from the choice of; phage vector, desired peptide length, peptide structure (linear or cyclised) and, techniques utilised to randomise peptide sequences. Researchers can also elect to generate custom peptide phage libraries using commercial cloning kits *e.g.* M13 phage Ph.D. cloning system (NEB) and, T7phage-Select systems (Merck Millipore) (Fukunaga and Taki, 2012). Routinely, small peptides are used in peptide phage display and, researchers predominately exploit pre-made commercial libraries; Ph.D.-12 a

dodecapeptide(12mer), Ph.D.-7, linear heptapeptide(7mer) and, Ph.D.-C7C loop constrained heptapeptide(7mer) library systems (New England Biolabs, NEB) (Fukunaga and Taki, 2012; Castel *et al.*, 2011; Sieber *et al.*, 2015Rami *et al.*, 2017).

1.9 PHAGE DISPLAY DISCOVERY OF ANTIMICROBIAL PEPTIDES AGAINST BACTERIAL PATHOGENS

1.9.1 Whole-cell phage display platforms for AMP discovery

A variety of display technologies have been utilised for the discovery of novel AMPs. **Table 1.7** summarises examples of literature antibacterial AMPs identified from phage peptide libraries and display technologies. The vast majority of strategies focus on whole cell-phage display platforms. For instance; Bishop-Hurley et al., (2005) designed a subtractive whole-cell phage display strategy by repurposing the 15mer f88-4/15-mer phage peptide library (Smith, University of Missouri-Columbia). Non-typeable *Haemophilus influenzae,* is an obligate parasite which can cause mucosal infections in humans for instance; meningitis, sinusitis and bronchitis. Subtractive panning of the f88-4/15-mer phage peptide library was conducted against non-virulent H. influenzae Rd KW20 strain. Sixteen round-three panning candidates were randomly selected, propagated as phage and tested in vitro against a panel of *H. influenzae* strains.

Clone hi3/17; KQRTSIRATEGCLPS; demonstrated strong bacteria species selectivity and lacked bactericidal activity against several Gram-Negative (*Proteus mirabilis, P. aeruginosa, Salmonella enterica*) and Gram-Positive bacteria (*E. faecalis, S. aureus*). However, the hi3/17 clone exhibited potent activity against *H. influenzae* R2866 at MICs 0.8 µM to 9.4 µM and MBCs 1.2 µM to 9.4 µM (MBC = lowest concentration representative of a ~ 99.5% bactericidal killing effect on the test bacteria). hi3/17 clone was within range or surpassing the antimicrobial activity of

well-detailed AMPs including polymyxin (2 µM), LL-37 (20 µM), Magainin II (5 µM) and Human β- defensin-2 (8 µM). *C. jejuni* gastrointestinal tract carriage in meat-producing chickens is a growing zoonotic concern and Bishop-Hurley et al., (2010) utilised a similar subtractive panning strategy to identify novel *C. jejuni* specific AMP from the f88-4/15-mer phage peptide library. In both panning schemes PBS-G was utilised in panning round one, whereas sub-libraries of output phage were split into testing with washing steps with PBS-G or TBST (TBS buffer containing 0.5% (v/v) Tween-20; TBST, pH 7.5) at respective rounds remaining. A larger portion of the candidate library clones with the most potent antimicrobial activity were identified from PBS-G washing throughout the panning rounds (Bishop-Hurley et al., 2010).

Table 1.7 Antimicrobial peptides discovered by phage display

	Characteristics		Micro bicidal effect level	Antimicrobial activity		Peptide library and phage display system	
	Length	**Peptide Sequence**	**%**	**MIC/MBC**	**Target species**		
M6	10 mers	QKKIRVRLSA	90-99.5	4-8 µg/mL 4-16 µg/mL 8 µg/mL >128 µg/mL	All multidrug resistant clinical isolates Gram-negatives; *Pseudomonas aeruginosa* 88514, 89, VA463/9 *Klebsiella pneumoniae* W99FI0057, W03NO0078, W03BG0019 *E. coli*, W99FI0077, W03BG002, W03NO0013 Gram positives; *Staphylococcus aureus* MIU-68A	The Ph.D.-10 mer decapeptide library and Filamentous phage – M13 – pIII display	(Pini *et al.*, 2005)
hi3/17	15 mers	KQRTSIRATEGCLPS	99.5	1.2 µM	Activity specifically against; *Haemophilus influenza* (no activity seen against other Gram negatives)	The f88-4/15-mer peptide library phage (pVIII) display	(Bishop-hurley *et al.*, 2005)
CP3/1 CP3/3	15 mers 15 mers	GRFLIRVTSSPLGPD FLIDSPLASIGPTSM	96 91-97	1.5 µM 3.1 µM	Activity specifically against; *Campylobacter jejuni* ACM 3393 and *C. jejuni* C338 (poultry isolate)	The f88-4/15-mer peptide library phage (pVIII) display	(Bishop-hurley, Rea and Mcsweeney, 2010)
EC5	12 mers	RLLFRKIRRLKR	-	 8 µg/mL 8 - 16 µg/mL 32 - 64 µg/mL 64 µg/mL 64 µg/mL 128-256 µg/mL	Gram-negatives; *E. coli* ATCC 700928, ATCC 25922 *P. aeruginosa* ATCC 27853, ATCC 12121 *K. pneumoniae* ATCC 10031, ATCC 13885 Gram-positives; *Staphylococcus epidermidis* ATCC 35983 *Bacillus cereus* ATCC 11778 *S. aureus* ATCC 25923	Ph.D 12-mer peptide and Filamentous phage pIII display	(Sainath Rao, Mohan and Atreya, 2013)
L2 L3	12 mers 12 mers	DQFVHDVKGTKH NSWIQAPDTKSI	-	30 µM 30 µM	Activity specifically against; *L. monocytogenes* (Multidrug resistant isolated from the cerebrospinal fluid)	Ph.D 12-mer peptide and Filamentous phage pIII display	(Flachbartova *et al.*, 2016)

The most potent or relevant phage display peptides identified in each strategy is summarised in the present table.

[a] MIC, minimal inhibitory concentration of the phage-displayed or synthetic peptides resulting in no detectable growth.
[b] MBC, minimum bactericidal concentration is the lowest concentration of required to kill.
[c] Underlined in black are the ideal targets or Gram-spectrum for aforementioned AMPs

Sainath Rao *et al.,* (2013) screened the 12-mer random peptide phage display library, pIII fusion (New England Biolabs) using a whole-cell subtractive strategy to identify novel AMPs against *E. coli.* The Gram-Positive model bacteria selected for the initial round of panning was *S. aureus* ATCC 25923. *S. aureus* depleted output phage sub-library was successively screened against *E. coli* ATCC 700928 in six rounds of panning. Similar to aforementioned approaches, within all contexts Sainath Rao e*t al.,* (2013) suspended whole cells in PBS (pH 7.4) and exposed target bacteria to pIII fusion displayed library (2×10^{10} per 100 µl/well). The phage library was diluted in TSBT, which was additionally utilised to conduct ten successive washes to remove unbound phage. HCl (100 mM) pH shifting followed by 1 M Tris (pH 8.0) was used to elute and neutralise bound output phage. During this process five clones were identified carrying the same library peptide sequence "EC5"; RLLFRKIRRLKR (+7 charge, 41% hydrophobicity). EC5 will hereafter be referred to as Peptide RLL.

Peptide RLL outer membrane permeabilising activity against *E. coli* and *P. aeruginosa;* with 6.25 µg/mL and 12.5 µg/mL concentrations of Peptide RLL being representative of the lowest concentrations of detected membrane permeability (Sainath Rao et al., 2013). Peptide RLL secondary structure analysis identified propensity for α-helix conformations. Similar to previous assays, MBCs revealed from the panel of Gram-Negative and Gram-positive bacteria Peptide RLL only displayed activity against only *E. coli* and *P. aeruginosa.* 12.5 – 50 µg/mL concentrations of Peptide RLL were tested in cation-supplemented Mueller-Hinton broth a 5 $\log_{10}$ CFU/mL reduction in was observed *E. coli* and *P. aeruginosa* (Sainath Rao et al., 2013). Later work demonstrated the low cytotoxicity in eukaryotic cells and non-hemolytic activity against chicken red blood cells.

Ph.D. Peptide Library Kits Ph.D.-7, Ph.D.-C7C, and Ph.D.-12 (New England BioLabs) are highly exploited libraries repurposed for AMP identification (Silva Jr et al., 2002). Flachbartova *et al.,* (2016) identified novel AMPs by screening a 2×10^{10} PFU combinatorial Ph.D-12 phage display library (New England Biolabs) against the cell surface of multidrug resistant *L. monocytogenes* (4×10^5 CFU/mL, LB medium). Succeeding incubation, samples were centrifuged (x 2000 g, 15 minutes RTP) and unbound phage relinquished by >3 washes with 0.1% PBS-T, whereas 0.2 M Glycin-HCl (pH 2.2) and 1 M Tris–HCl (pH 9.0) was utilised for eluting bound phage and neutralisation.

Flachbartova *et al.,* (2016) conducted four rounds of panning in the absence of a subtractive step, and twenty phage clones were rescued randomly from a 10^{-8} dilution of output phage were analysed for antimicrobial activity. Phage display peptides L1, L2 and L3 were later identified as present in ≤ 50% of final panning sample replicates; 10/20, 4/20, 6/20 respectively (Flachbartova *et al.,* 2016). All three peptides were within the bounds of conventional AMP descriptions; L1 adopted a β-sheets conformation, whilst L2 and L3 were α-helical peptides, and physicochemical property ranges included; overall charge; 0 - +1, pI = 6.34 – 7.55, hydrophobicity 25-33%. L2 and L3 were inhibitory to bacterial growth at 30μM in microdilution susceptibility testing assays, whereas L1 was inactive. Nonetheless, all three peptides were found to be non-toxic to eukaryotic cells.

Contrasting whole-cell panning, Xiong *et al.,* 2019 targeted panning a Ph.D.-7mer phage display peptide library (New England Biolabs) specifically against the gastric acid receptor ArsRS, which plays a vital role in *Helicobacter pylori* sensing and adaption in highly acidic gastric environments. Several phage display derived peptide candidates were isolated with ELISA determined biding affinity for ArsRS

and activity against *H. pylori* (<16 µM). Xie et al., (2006) similarly explored non-whole-cell approaches to the identification of AMPs. More specifically, a cassette randomised phage peptide library was screened against bacterial and eukaryotic bacterial membrane models. The liposome compositions of both models contained phosphatidylcholine, phosphatidyl glycerol, 1.4% biotinylated phosphatidyl ethanolamine at varying rations and the mammalian model contained cholesterol. Xie *et al.,* (2006) finalised selection from the bacterial model included 95 different sequences, and interesting motif biases and combinations of R/V/G/K/A/L amino acids were identified *i.e.* "ALR", "KVL" "RVG" across the peptide region of several library peptide candidates. The top 20% most potent AMPs from the shortlisted 95 peptides exhibited MIC ranges; B. subtilis (3.1-200 µM), S. aureus (1.56 – 200 µM), *E. coli* and S. enteritidis strains (3.1 -200 µM) (Xie et al., 2006).

1.9.2 STAMPs & Phage display: Species specific AMP tethering systems

Binders with species specific binding affinity can be dimerised with an additional AMP with characterised antimicrobial activity and, this technique is often referred to as STAMPs. The STAMP strategy typically involves constructing a hybrid mutant peptide of two independently functional and characterised AMPs. Primarily, one of two tethered peptides facilitate either a) affinity for target-bacteria binding or b) demonstrates broad-spectrum activity (~ < 50 µM). For instance, Kim et al., (2020) tethered de novo generated; GNU7 (RLLRPLLQLLKQKLR) to several phage display candidates enriched against *P. aeruginosa* (whole-cell phage display approach). The Ph.D-12 phage display peptide library (New England Biolabs) was screened against *P. aeruginosa* in four rounds of panning. Twenty phage clones were randomly selected from the eluted and diluted output phage population from *P. aeruginosa* panning. PA1 (SCSSLTTLRPCG), PA2 (SQRKLAAKLTSK) and PA3

(VILTGPEAEYFW) bound to *P. aeruginosa* and possessed limited cross-reactivity with *E. coli, S. Typhimurium, S. aureus, and C. albicans.* Individually PA2 and GNU7 MICs against *P. aeruginosa* were ~ 32μM each. However the STAMP; PA2-GNU7 demonstrated an MIC of 2 μM, a 16-fold increase, against *P. aeruginosa*. PA2-GNU7 STAMP compared to individualised MICs, exhibited a 2- to 8-fold difference in susceptibility for *E. coli, S. Typhimurium, S. aureus, and C. albicans.*

Peng Tan *et al.,* (2020) utilised the framework of selecting *E. coli* active STAMP (WKKIWK^{D}PGIKKWI) and rationally improving characteristic AMP features; *e.g.* increased charge or altered amphipathicity. Primarily, N- and C- terminal ends of the branching STAMP were modified with amino acid combinations or stretches of cationic, hydrophobic and/or hydrophilic; K/R/G/V/L/A/P/Q/S. The twenty STAMPs identified by Peng Tan *et al.,* (2020) were active at 1-128μM activity against *E. coli* and several other pathogenic and beneficial lactic acid producers. The modification N terminal-WKKI and C-terminal KWIK was the most potent clone candidate against *E. coli* (0.5-2μM). However, candidate peptides demonstrated larger 4-fold variance amongst the six *E. coli* strains exemplifying AMPs can have varying MICs amongst strains of the same susceptible bacterial species.

1.9.3 AMP Screening: phage display and magnetic bead technology

Tanaka, Kokuryu and Matsunaga, (2008) conducted six biopanning (selection affinity) rounds against a Ph.D.-12 mer (NEB) peptide phage displayed library with a complexity of 2.7×10^{9}. Selection affinity was facilitated through the use of nanosized (50-100nm) BacMPs surrounded by the strongly ferrimagnetic lipid bilayer of *Magnetospirillum magneticum* AMB-1 (Yoshino *et al.,* 2010). The BacMP-AMB-1s were incubated with 4×10^{10} PFU(Plaque forming units) of the phage library whereupon, lipid membranes were extracted and phage eluted using mild

phospholipase D treatment (Tanaka, Kokuryu and Matsunaga, 2008). Generally, after each successive round of biopanning, the percentage occurrence of phage clones carrying recombinant proteins with cationic characteristics increased (Round 1: 13.6% and Round 6: 65%). Subtractive panning with *B. subtilis* coated BacMP identified six peptides (5-24, 6-2, 6-7, 6-8, 6-17, and 6- 45) exhibiting stereotypical AMP characteristics (Tanaka, Kokuryu and Matsunaga, 2008).

Peptide 6-7 (KPQQHNRPLRHK), possessed a -33 hydropathy index, and unsurprisingly showed the highest binding capacity (Tanaka, Kokuryu and Matsunaga, 2008). Consequently, Tanaka, Kokuryu and Matsunaga, (2008) opted to use KPQ as a template for site-specific mutations; proline(P)/glutamine(Q) to phenylalanine(F), asparagine(N) to valine(V) and, histidine(H) to tryptophan (W). These specific substitutions were chosen to increase the hydrophobicity of KPQ variants as F, V and W amino acids have highly hydrophobic side chains (Moon and Fleming, 2011). KPQ peptide exhibited binding to *B. subtilis* but, weak antimicrobial activity against *B. subtilis* and little to no activity against *E.coli* K-12 and *Saccharomyces cerevisiae* (NRBC 0224). Whereas mutant peptide 6-7/7 (KFVVWVRFLRWK) generated a 10^3 -fold decrease of *S. cerevisiae* CFU/mL (colony forming units) +8hrs incubation at 100µM (Tanaka, Kokuryu and Matsunaga, 2008). This enhancement of antimicrobial activity was also mirrored against *B. subtilis* as most notably, modified peptides 6-7/5 (KFVVHVRFLRHK) and 6-7/7 at 100µM exhibited activity 1.1 x 10^4 greater than KPQ.

Furthermore, *B.subtilis* growth rate kinetics analysis showed these two peptides (1 or 10µM) were able to induce antimicrobial activity within 1 hour, whilst KPQ possessed no significant effect when incubated for 4 hours at 100µM (Tanaka, Kokuryu and Matsunaga, 2008). This demonstrated that amino acid substitutions of

P/Q and N to F and V respectively generated peptides with higher hydrophobicity and antimicrobial activity. Furthermore, substituting histidine into tryptophan, in peptide 6-7/7 resulted in slightly greater hydrophobicity, and this variant possessed greater antimicrobial activity at low concentrations (1µM) that were 2.6 x 10^2 times greater than KPQ (Tanaka, Kokuryu and Matsunaga, 2008).

1.9.4 Soft-agar overlay assays for AMP scanning studies

Guralp *et al.*, (2013) described the application of an *E. coli* JE5505 (lpp-254, lipoprotein deficient mutant) secretory expression host system. Lpp is a major outer membrane protein in *E. coli* that anchors the outer membrane to cell wall via a C-terminal lysine and N-terminal N-acyl-S-diacylglycerylcysteine (Chang *et al.*, 2012; Teh *et al.*, 2014). This was a beneficial characteristic as Guralp *et al.*, (2013) used *E. coli* JE5505 to facilitate the leaching of periplasmic-expressed recombinant peptides, with putative antimicrobial activity, into the soft agar of overlay assays.

Succeeding selecting an appropriate screening methodology Guralp *et al.*, (2013) began the antimicrobial discovery pipeline by making a peptide library with 12,000 plantaricin-423 (Pln-423) mutants (≤50Aa) created by standard phosphoramidite chemistry and maskless photolithography on glass slides. Plantaricin-423 is a candidate biopreservative as this 3.5kDa Class II-bacteriocin isolated from *Lactobacillus plantarum 423* exhibits bactericidal activity against a range of Gram-positive bacteria including foodborne pathogens such as; *Listeria innocua and L. monocytogenes* (van Reenen, Dicks and Chikindas, 1998). The *Pln-423* variant ss-DNA library was amplified using emulsion-PCR, which entails emulsifying the PCR reaction mixture in oil so that amplification of independent ss-oligonucleotides occurs in individual microdroplets (Guralp *et al.*, 2013). Emulsion PCR; (i) created ds-DNA for the Pln-423 library variants, (ii) introduced restriction sites for subsequent cloning

and, (iii) retained library diversity as competition between DNA molecules is reduced. Following this, PCR products were digested using *Hind*III and Eco*RI* and, ligated into pFLAG-CTS expression vector to transform *E. coli* JE5505. The FLAG epitope pertains to a small hydrophilic tag (DYKDDDDK) that allows sensitive detection and purification through the application of anti-FLAG molecules (Tatsumi *et al.*, 2017).

Transformed *E. coli* JE5505 were then diluted to 3000–5000 CFU/mL in 25 mL Luria-Bertani (LB) soft agar (0.8% at 45^{o}C). This precise CFU/mL value was determined from growth curve graphs (*y axis:* OD and CFU/mL plotted against *x axis:* time) and, aimed to produce <50 colonies per screening petri dish plate (150mm). Following media solidification, another layer of LB soft agar was overlaid (acts as agar layer barrier between *E. coli* JE5505 and test species) and, plates were incubated at 37^{o}C for 24hrs to allow *E. coli* JE5505 colonies to form. Following this, plates were overlaid with 10mL of Trypticase Soy soft agar containing; the test species *L. innocua* ATCC 33090 diluted to OD_{600nm}= 0.03 along with 1mM IPTG to induce production of recombinant Pln-423 peptides via *lac* promoter. Following overnight incubation (30^{o}C) zones of inhibition (ZoI) indicated by clear growth were used to isolate *E. coli* JE5505 clones encoding peptides with antimicrobial activity. Whereupon. Pln-423 variants from these clones were characterised by sanger sequencing and, desired peptides were synthesised for MIC assays (Guralp *et al.*, 2013).

Guralp *et al.*, (2013), identified a Pln-423 variant: Pln-S23E (Serine residue 23 to Arginine) that exhibited a 13-fold higher MIC compared to unmodified Pln-423. This demonstrated the importance of C-terminal serine residues in Pln-423 antimicrobial potency. However, the Pln-S23E variant highlighted a negative of this overlay approach as, it appeared in two overlay assay selections. Suggesting that variability of recombinant peptide expression can lead to false positives. Nonetheless, Guralp *et al.*,

(2013) identified two key *E. coli* JE5505 mutants carrying Pln-423 variants; H28R/H34E (Histidine residue 28 to Arginine/Histidine residue 38 to Arginine) and S27D/K26N (Serine residue 28 to Aspartic Acid/Lysine residue 38 to Asparagine). Compared to the wild-type Pln-423 that possessed an MIC: 0.075µM and ZoI(2.5µM): 14.3mm, the H28R/H34E and S27D/K26N Pln-423 variants possessed higher activities MIC: 0.037µM (2-fold lower than wild type Pln-423) and ZoI(2.5µM): 18.6 and 19.1mm respectively. This, consequently, identified other site-specific residues (Lys36 and His28), that were critical to the antimicrobial activity of Pln-423 against *L. innocua*.

1.9.5 SPOT-synthesis AMP peptide assays

SPOT-synthesis potentially can curtail the often-expensive cost associated with the crude synthesis of screening output panels of improved mutant AMPs. SPOT-synthesis is an innovative technique fundamentally based on dispensing small droplets of reaction mixtures containing in-situ activated or pre-activated Fmoc-protected amino acid specifically onto a porous planar surface membrane (López-Pérez et al., 2015). Hilpert and Hancock, (2007) demonstrated the antimicrobial screening of SPOT-synthesised and single codon-substitution matrix of Bac2A (RLARIVVIRVAR) a linear mutant of bactenecin (natural AMP in bovine neutrophils). The substitution Bac2A matrix generated 228 unique mutant variants for SPOT synthesis, which were subsequently screened against *P. aeruginosa*. Single substitutions were identified which improved activity by 6-fold compared to wildtype bactenecin MIC against *P. aeruginosa* (50 µg/mL) whereas multiple substitution *i.e.* substitutions of Ala (pos 11) to C, W, R, K and H conferred improved antibacterial activity (Hilpert and Hancock, 2007; López-Pérez et al., 2015).

Soluble peptide-cellulose complexes can be spotted onto glass slides forming CelluSpots microarrays for biological testing (López-Pérez et al., 2015. Knappe et al., (2016) later went on to demonstrate Oncocin (proline-rich natural AMP) mutants could be tethered to membranes and demonstrate antibacterial activity against *P. aeruginosa* and *S. aureus*. However, attempting singular mutations/per mutant produced limited improvements in activity, reflecting the relatively poor pre-characterised activity of oncocin against bacteria test species. Nonetheless, perseverance was rewarded as the combination of two favourable substitutions were combined within one peptide mutant, improving MIC to 0.5 – 8 µg/mL for *P. aeruginosa* and *S. aureus*.

1.10 HYPOTHESIS AND AIMS

1.11 Hypothesis:

The hypothesis of this study posits that the constituent amino acid distribution within natural AMPs correlates with their exhibition of antimicrobial phenotype(s); wherein certain amino acid biases, motifs or patterns, both in terms of physiochemical properties and amino acid positional incorporation, confer the enhanced propensity of antimicrobial activity.

By primarily analysing datasets of natural AMPs, this complex relationship between amino acid distribution and antimicrobial activity can be further elucidated. Peptide library mutagenesis can be exploited to systematically construct libraries with diverse peptide variants aligned to the revealed amino acid biases underpinning natural AMPs. Peptide library variants with antimicrobial properties can be identified and characterised via numerous screening approaches and potentially be utilised as early hits for the development of novel antimicrobial therapies.

Working hypothesis: Puts forth strategically replicating the amino acid distribution patterns present in natural AMPs in a peptide library mutagenesis strategy increases the number of library candidates with the desired antimicrobial phenotype.

1.12 Aims

The principal aim was to generate a novel antimicrobial peptide discovery pipeline;

i. **"AMP-biased" library strategy:** Analyse the amino acid distribution of natural AMPs to generate a semi-rational degenerate codon scheme approach to restrict encoded amino acids in peptide library variants towards the bias identified in natural AMPs.

ii. **Site-saturated mutagenesis:** Devise a mutagenesis strategy which exploits the use of degenerate primers to introduce randomised 16mer peptides into phagemid vectors.

iii. **Peptide library design and screening:** Devise a library cloning and vector strategy which supports more than one downstream screening approach to identify library variants with antimicrobial activity. Specifically, high-throughput Next Generation Phage Display (pIII : 3+3 system) and the *leaky* periplasm *E. coli* JE5505 recombinant expression peptide overlay assay.

iv. **Benchmark the technical effect of the "AMP" biased strategy:** Evaluate randomised 16mer "AMP-biased" peptide libraries, generated under the provisions of aims *i – iii,* and compare the identification of library variants with antibacterial phenotypes against that observed in a conventional NNK degenerate peptide mutagenesis library.

v. **Industrial applicability:** The constant thread between *i* and *iv.* is the desire to discover and characterise novel AMPs with antimicrobial phenotypes against bacterial pathogens with porcine host specificity; *S. Typhimurium 4/74*, *S. suis P1/7* and *E. coli P433*.

Chapter 2: Materials and Methods

2.1 BACTERIOLOGICAL MEDIA USED AND MATERIALS THEREOF USED FOR THE CULTIVATION OF BACTERIA TEST STRAINS AND LIBRARY SCREENING APPROACHES

2.1.1 Nutrient Broth media

Nutrient Broth ("NB"; CM0001, Oxoid) was prepared according to manufacturer's instructions. 15 g/L of agar powder (VWR) was added to NB to make Nutrient Agar (NA) plates. NB cultures and agar plates were prepared according to **(Method 2.4.1).** NB and NA plates were utilised for the cultivation of non-fastidious bacterial strains.

2.1.2 MacConkey Agar (MAC)

The differential medium MacConkey Agar ("MAC"; CM0007, Oxoid) was prepared according to manufacturer's instructions and **(Method 2.4.1)**. MAC agar plates were used for the cultivation of *Enterobacteriaceae* specifically, *E. coli* strains and to a lesser extent *Salmonella spp*.

2.1.3 Brilliant Green Agar (BGA)

Selective medium Brilliant Green Agar ("BGA"; CM0263, Oxoid) was prepared according to manufacturer's instructions and **(Method 2.4.1).** BGA was utilised for the selective isolation of *Salmonella* spp., specifically *Salmonella Typhimurium.*

2.1.4 Mueller Hinton Broth (MHB) and Mueller Hinton Agar (MHA)

Non-selective Mueller-Hinton Broth ("MHB"; CM0337, Oxoid) was prepared according to manufacturer's instructions. 15 g/L of agar powder (VWR) was added to MHB to generate Mueller-Hinton Agar ("MHA") plates **(Method 2.4.1)**. Mueller-Hinton media was primarily used for antimicrobial testing assays and the cultivation of leaky *E. coli* JE5505 *Δlpp-254* strain.

2.1.5 2× Yeast Extract Tryptone (2xYT) and Luria-Bertani Broth media

2× Yeast Extract Tryptone ("2xYT") and Luria-Bertani Broth ("LB") was prepared using the formulation shown in **Table 2.1** and sterilised (**Method 2.4.1**). 15 g/L and 12 g/L of agar powder (VWR) were used to prepare 2xYT or LB agar respectively. 2xYT and LB media were utilised for the cultivation of *E. coli* strains. 2xYT and LB ampicillin and/or carbenicillin antibiotic supplementation **(Method 2.4.2)** were used to isolate *E.coli* strains carrying phagemids.

Table 2.1 Formulation of 2xYT and Luria-Bertani (LB) broth culture media.

Media	Mass of component (g/L)			
	Yeast Extract	**Sodium Chloride**	**Tryptone**	**Peptone**
2xYT	10.0	5.0	16.0	
LB	5.0	5.0		10.0

2.1.6 Tryptone Soya Broth (TSB) and Tryptone Soya Broth Agar (TSA)

Tryptone Soya Broth ("TSB"; CM0129, Oxoid) was prepared according to manufacturer's instructions. 15 g/L of agar powder (VWR) was added to TSB to make Tryptone Soya Broth Agar plates **(Method 2.4.1).** TSB and Tryptone Soya Agar ("TSA") were utilised for cultivating *Listeria spp* more specifically, *L. innocua* and *L. monocyotogenes* strains used in this study**.**

2.1.7 de Man, Rogosa and Sharpe (MRS) medium

de Man, Rogosa and Sharpe ("MRS", CM0359 Oxoid) medium formulation was developed and optimised to be selective for *Lactobacilli* and other lactic acid bacteria genera. Small opaque white colonies (<0.5 mm) with slightly feathery undulate margins are distinctive of *Lactobacilli* growth on selective MRS media. *Lactobacillus acidiophillus* ATCC 43561 was cultivated under microaerophilic conditions, often under extended incubations for 24- 48 hours (37^{o}C). *Lactobacilli* microaerophiles and as such require overlaying in aerobic cultivation on MRS agar medium. MRS was

made according to manufacturer's instructions, and 15 g/L of agar was utilised to make MRS agar **(Method 2.4.1).**

2.1.8 Selective media for Listeria, Salmonella and Streptococcus suis.

The following media were high quality standardised formulations purchased as pre-prepared agar plates. *Listeria* isolation Oxford Medium in 90mm plates (EO Labs) was utilised to verify *Listeria* species herein. Specifically, *L. innocua* and *L monocytogenes. Listeria spp.* hydrolyse aesculin, forming black halo zones around colonies (2-3 mm). Certain coagulase negative staphylococci might appear as aesculin negative colonies and enterococci grow poorly exhibiting weak aesculin reaction +40hrs of incubation. Pre-prepared Columbia CAP Selective Agar plates with 5% Sheep Blood (Thermo Scientific) were utilised for the selective cultivation of *Streptococcus suis*. Translucent colonies with green halos of clearing are characteristic of *S. suis* haemolysis activity on Columbia CAP Selective Agar supplemented with Sheep Blood. *Salmonella Typhimurium* was grown on selectively differential Oxoid Prepared XLD Agar (Thermo Scientific). On XLD, *S Typhimurium* are differentiated as transparent, black-centred colonies on red coloured agar due to xylose fermentation, lysine decarboxylation sugar metabolism and hydrogen sulphide production. All aforementioned plates were stored as recommended by manufacturer (typically <8 weeks at 4^{o}C).

2.1.9 Soft Agar Media

Soft agar was utilised for agar-based antimicrobial screening specifically the overlay assay method **(Method 2.5.10).** Broadly, 0.35 – 1.5% agar percentages were prepared for bacteriological media; MHB, 2xYT, LB and TSB **(Method 2.1.4 - 2.1.6).** 0.35% to 0.8% soft agar percentages were utilised for overlay assays. A "0.35%" soft

agar represents a 75% reduction in the agar within media compositions previously stated **(Method 2.1.4 - 2.1.6).**

2.1.10 Antibiotics for supplemented media

Ampicillin sodium salt (Fisher BioReagents), Carbenicillin disodium salt (Fisher BioReagents), Kanamycin antibiotic (Alfa-Aesar), Nisin (Sigma) and Polymyxin B (Calbiochem) were antibiotics utilised herein. Antibiotic stock solutions were generated using molecular grade ddH_2O (GE Healthcare Hyclone) passed through a prewashed 0.22 µm sterile filter (Merck Millipore). 1 mL aliquots filter sterilised antibiotic solutions were transferred into autoclaved sterile 1.5 mL eppendorfs. Antibiotic aliquots were stored at -20 °C for 1 year or at 4 °C for 3 months.

Positive controls for antimicrobial characterisation assays: The cationic and hydrophobic nature of nisin and polymyxin B lends to antimicrobial peptide categorisation (Zavascki et al., 2007; Abee and Delves-Broughton, 2003). Nisin is a low molecular weight polypeptide, consisting of 34 amino acids (Abee and Delves-Broughton, 2003). Originally isolated from the diary starter culture bacteria species *Lactococcus lactis*, nisin is considered a potent bacteriocin exhibiting activity against a wide variety of Gram-positive bacteria and sporeformers (Abee and Delves-Broughton, 2003). Contrastingly, Polymyxin B sulfate is a polycationic peptide ring, with a tripeptide side chain containing fatty acid tail (Zavascki et al., 2007) and exhibits activity against Gram- negative bacteria. In antimicrobial characterisation assays, Nisin, MW: 3354.07 g/mol (Sigma-Aldrich, N5764) was utilised as the positive control for Gram-positive bacteria, whereas Polymyxin B sulfate MW: 1,301.56 g/mol (Calbiochem, 5291) was used for Gram-negative bacteria in this study **(Method 2.2.1).** Both controls were suspended in ultrapure molecular grade ddH_2O.

Nisin and Polymyxin B sulfate stocks (5 mg/mL and 2.5 mg/mL respectively) were used to make 200 µM stocks, stored at -20 ^{o}C <6 months.

2.1.11 25% Glucose

25 g glucose (Fisher Scientific) was added to 100mL of autoclaved ddH_2O (<25 ^{o}C). The glucose solution when fully dissolved was filter sterilised via 0.22 µm sterile filter decanting bottle unit (Corning). Solution was stored at RTP <6 months.

2.1.12 Phage Display: M9 minimal growth media reagents

M9 Minimal Salts are formulated for generating deficient M9 growth medium namely utilised for phage display panning protocols. Supplementation herein of M9 salts was adjusted with the appropriate carbon and amino acids for *E. coli* TG1 cultivation. M9/Thiamine Hydrochloride minimal media maintains the selection of the *F' pillus* episome which is vital for manipulation in phage propagation techniques and reliant display technologies for M13 phage.

M9 salts solution x5 Solution (500 mL): 32 g Sodium phosphate (dibasic) heptahydrate (Na_2HPO_4**.**$7H_2O$), 7.5 g Monopotassium phosphate (KH_2PO_4), 1.25 g Sodium chloride (NaCl) in 500 mL of DEPC-treated ddH_2O (certified nuclease/RNAse free ddH_2O). All purchased from Fisher Scientific except for Monopotassium phosphate which was obtained from Sigma-Aldrich.

Magnesium chloride solution (20%): 10 g of $MgCl_2$ (Fisher Scientific) was added to 50 mL DEPC-treated water and further sterilised via autoclaving. Magnesium chloride supports bacterial growth and viability. Additionally, divalent metal ions are necessary for phage adsorption and putatively supports higher phage yields.

Thiamine Hydrochloride (10 mg/mL): Full nomenclature, 3-((4-Amino-2-methylpyrimidin-5-yl)methyl)-5-(2-hydroxyethyl)-4-methylthiazolium iodide monohydroiodide. 100 mg Thiamine HCl (Sigma-Aldrich) was added to 10 mL DEPC-treated ddH_2O and filter sterilised using a of 0.22 µm membrane filter (Merck Millipore). Thiamine HCl was stored foil protected at 4 ^oC for <6 months. Thiamine Hydrochloride is an essential vitamin nutrient component.

2.1.13 Phage Display: M9 minimal growth media preparation

Optimised M9 minimal growth medium is necessary for the selection of F' (+) pillus episome of *E. coli* TG1, which facilitates the M13 bacteriophage infection process. Minimal M9 growth medium was prepared as follows; 25 mL M9 salts solution (5x), 100 µl of $MgCl_2$, 20% solution **,** 50 µl Thiamine Hydrochloride (10 mg/mL) **(Method 2.1.12**), 1.6 mL Glucose, 25% solution **(Method 2.1.11)** were poured into molten agar; 1.5 g agar in 75 mL of DEPC-treated ddH_2O. 90mm M9 agar platers were used immediately or stored at $4^oC \leq 2$ days.

2.2 BACTERIAL TEST STRAINS AND PHAGEMID CLONING VECTOR MATERIALS

2.2.1 Bacterial test strains

Table 2.2 and Table 2.3 lists several bacterial species and respective strains collected for this study. For clarification, the optimisation of the overlay assay exploited *E. coli* JE5505 **(Table 2.3)** and *Listeria* test species **(Table 2.2).** Library construction and phage display utilised *E. coli* TG1 strain and sub-libraries were transformed into *E. coli* JE5505 **(Table 2.3).** As this study focuses on characterisation of AMPs against bacterial pathogens with porcine host specificity, *S. Typhimuium* 4/74 was the "model Gram- negative test strain ", *S. suis* P1/7 "model Gram- positive test strain" **(Table 2.1)** and *E. coli* P433 "model enteric pathogen" **(Table 2.3) .** The remaining strains listed in **Tables 2.2/2.3** are control strains for overlays, periplasm

extractions or were considered as candidates for subtractive panning strategies (*Lactobacillus acidiophillus* ATCC 4356) or strains susceptible to literature AMPs herein **see Table 2.4** *(i.e., B. subtilis* was the target of Tanaka, Kokuryu and Matsunaga, 2008 isolated AMP: "KPQQHNRPLRHK).

E. coli strains are highly utilised for the production of recombinant proteins. *E. coli* JE5505 cell factories can be biotechnologically manipulated to drive recombinant production of vector encoded peptides (Tominaga and Hatakeyama, 2006; Guralp *et al.,* 2013). Consequently, the initial aim focused on assessing the *E. coli* JE5505 expression system; specifically, with Pln-423 wt and improved mutants under assay structuring in parity with Guralp *et al.,* (2013). *E. coli* JE5505 originates from the K12 strain lineage, and was formerly characterised by Hirota et al., (1977) who identified the deletion of *Δlpp-254* as one of the thirteen mutations present in this strain. As an *Δlpp-254* mutant, *E. coli* JE5505 uniquely exhibits considerable leakage from the periplasmic space. Nevertheless, it was reported *E. coli* JE5505 maintains relatively undisturbed physiological operations, growth, division and synthesis pathways barring the lipoprotein *Δlpp-254* pathway (Hirota et al., 1977).

Hirota *et al.,* (1977) *E. coli* JE5505 strain was obtained from the *E. coli* Genetic Resources at The Coli Genetic Stock Center (Yale). On non-selective agars such as Luria-Bertani (LB), 2xYT medium and Mueller-Hinton agar (MHA); *E.coli* JE5505 was characterised herein as a catalase (+ive) and oxidase (-ive) Gram-negative rod, which produces butyrous slightly cream-coloured circular colonies (~1-2 mm), raised in elevation with entire margins. As conventional with *E.coli* strains, on selective MacConkey agar *E. coli* JE5505 produces flat non-mucoid pink colonies surrounded by pink agar due to the precipitation of bile salts.

Table 2.2. Gram-Negative and Gram-Positive bacteria strains utilised herein

Bacteria	Gram stain	Growth condition	Control Antibiotic used	Reference
S. Typhimurium 4/74	Gram- Negative	37°C , Aerobic	Polymyxin B	Rankin and Taylor, 1966.
Bacillus subtilis (2 strains) *B. subtilis ATCC 6633*[1]	Gram- Positive	37°C , Aerobic	Nisin	Shibata et al., 1976
Lactobacillus acidiophillus ATCC 4356	Gram- Positive	37°C, Microaerophilic or 5% CO_2	Nisin	Boot et al., 1995
Listeria innocua (2 strains)	Gram- Positive	37°C , Aerobic	Nisin	Obtained from University of Nottingham Food Science Department.
Listeria monocytogenes (3 strains)	Gram- Positive	37°C , Aerobic	Nisin	
Streptococcus suis P1/7	Gram- Positive	37°C , Aerobic	Nisin	Nicholson et al., 2020

Key:
Reference strain; American Type Culture Collection, ATCC https://www.atcc.org/

Aerobic growth conditions = ≥21% Oxygen, Microaerophilic conditions = 5% Oxygen, CO_2 incubator conditions = 15% Oxygen

Table 2.3. *E. coli* strains used for periplasm extractions, overlay assays and phage display.

	Lineage	Control Antibiotic used	Amber stop suppression	Reference
E. coli JE5505	K-12 strain	Polymyxin B Ampicillin Carbenicillin	Amber suppressor	Hirota *et al.*, 1977
E. coli TG1	K-12 strain		Amber suppressor	Obtained from University of Nottingham, School of Veterinary Medicine and Science.
E. coli HB2151	B strain		Non- Amber suppressor	
E. coli BL21	B strain		Non- Amber suppressor	
Enterotoxigenic *E. coli* (ETEC) P433	K-12	Kanamycin, Polymyxin B	Amber suppressor	Smith and Huggins, 1983

Key:
E. coli strains were grown at 37°C under aerobic growth conditions = ≥21% Oxygen
Specified genotype;
E. coli JE5505 *(F− 1po his proA argE thi gal lac xyl mtl tsx)* , The Coli Genetic Stock Center (Yale). *E. coli* Genetic Resources, CGSC[1]

E. coli TG1 *[F' traD36 proAB lacIqZ ΔM15] supE thi-1 Δ(lac-proAB) Δ(mcrB-hsdSM)5(rK – mK -)* Lucigen TG1 Electrocompetent, Cells Lucigen[2]

2.2.2 Glycerol Stocks

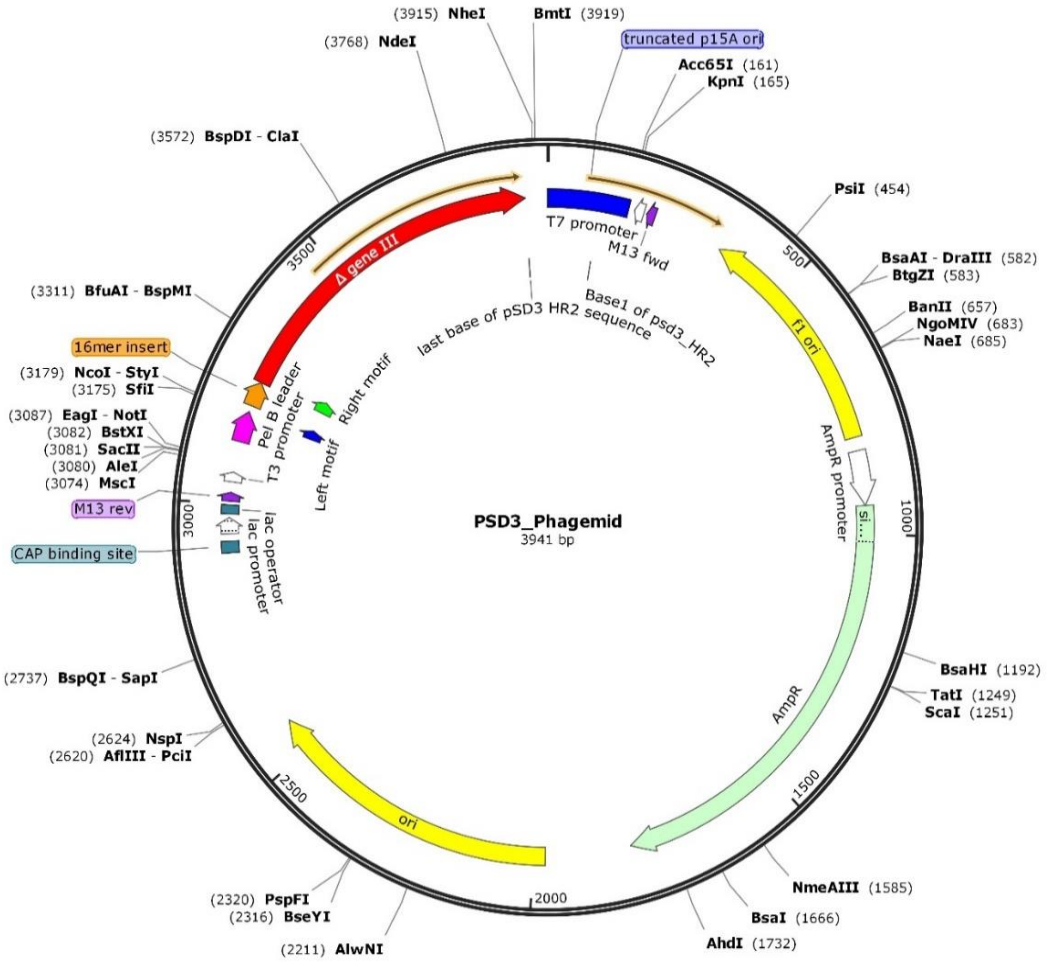
Glycerol storage was utilised for all bacteria test strains; 20% glycerol stocks using +99% Glycerol, ("Fisher Scientific, 10579570") were prepared then stored at -20 ^{o}C (short-term, ≤6 months) and -80 ^{o}C (long-term, ≥1yr). Mid-log (OD_{600nm} = ~0.4-0.6) glycerol stocks or overnight stocks were utilised. 50% glycerol stocks were used for *E. coli* JE5505 due to membrane fragility and/or high vulnerability to freeze-thaw damage.

2.2.3 pSD3 Phagemid Construct

pSD3 is an M13 bacteriophage derived phagemid containing key nucleic elements required for phage display. The pSD3 vector (<4kb), **Figure 2.1**, was designed herein for recombinant peptide production in *E. coli* and was the template vector for phage display library construction. The pSD3 vector contains PelB leader sequence "MKYLLPTAAAGLLLLAAQPAMA" an alkaline phosphatase signal peptide which directs fusions for export to bacterial periplasmic spaces. The PelB signal peptide lies at the N-terminus of the peptide insert region. Several other important pSD3 element features include but is not limited to; *gIII* phage gene (encodes pIII protein), bacteriophage *f1* origin of replication, plasmid origin of replication for *E. coli* propagation, *AmpR* ampicillin resistance gene used as the antibiotic selection marker, *lac* promoter for controlling recombinant peptide production and various regions perfectly apt for sanger sequencing (*T3* promoter, *M13* fwd and rev). The specific utility of the pSD3 phagemid was the applicability both to phage display and recombinant production in *E. coli* strains, including the *E. coli* JE5505 strain exploited for the overlay assay.

Figure 2. 1 *pSD3* 16mer peptide phagemid vector map

pSD3 phagemid map was obtained from SnapGene viewer. The complete *pSD3* phagemid including a 16 mer peptide insert (orange) equates to 3941 bp. The left and right motifs flanking the peptide insert site are utilised for *in silico* analysis to positionally locate the peptide encoding region. The vector map additionally depicts the positions of the *f1* origin of replication and plasmid origin of replication for *E. coli* propagation (yellow), *AmpR* ampicillin resistance gene (green), lac operon (turquoise), PelB periplasm leader sequence (pink) and *gIII* phage gene (red). *bspQI / SapI* restriction cleavage site is located at 2737 bp in the naïve phagemid.



2.3 PEPTIDES, BUFFERS, REAGENTS, KITS AND MATERIALS THEREOF USED FOR ANTIMICROBIAL ASSAYS, LIBRARY CONSTRUCTION AND SCREENING APPROACHES

2.4 METHODS FOR THE ENUMERATION OF BACTERIA TEST STRAINS

2.4.1 Bacteriological media preparation

Unless otherwise acknowledged, bacteriological growth broth and agar media were prepared to manufacturer recommendations in Reverse Osmosis (RO) water, autoclaved sterilised (121 ^{o}C for ~20 minutes at 15 psi) and stored at room temperature and pressure (RTP, ~20 ^{o}C) for < 6 months. Agar media were tempered to 45-50 ^{o}C then poured into sterile petri dishes using aseptic technique. The agar volumes typically used depended on petri dishes diameter; 90 mm, ~25 mL; 150 mm, 40 mL; 100 mm, 27.5 mL). Agars plates were dried in laminar flow cabinets until solidified and no condensation present on petri dish lids. Plates were stored in the dark at 4 ^{o}C for up to one month. Where necessary broth and agar media were preheated to 37 ^{o}C for fastidious bacteria test strains.

2.4.2 Antibiotic stocks and antibiotic supplemented media

Broths and agars were prepared **(Method 2.4.1)** and cooled to 45 – 50 ^{o}C for antibiotic supplementation. Ampicillin and carbenicillin containing media were supplemented at a final working concentration of 100 µg/mL. Ampicillin and Carbenicillin containing plates were immediately used or stored foil-protected in the dark at 4 ^{o}C for ≤1 month. Kanamycin working concentration was 150 µg/mL, and kanamycin supplemented media was used herein for phage production.

2.4.3 Enumeration of bacteria

Enumeration method used herein aligns to the technique described by Miles and Misra (1938). Bacterial culture suspensions were serially diluted in tenfold steps down

to a dilution of 10^{-10} *i.e.* 20 µl culture suspension from 96-well plate (polypropylene plate) diluted into 180 µl of broth diluent create 10^{-1} dilution and repeated per triplicate specimen. Pipette tips were discarded between dilution steps and whilst spotting either; six 10 µl (**2.4.3A**) or 20 µl (**2.4.3B**) volumes. A smaller spotting volume (*i.e.* 10 µl) is used for bacteria that produce medium-large colonies. Spotting was placed onto dried and pre-warmed (37 ^{o}C) agar media with or with antibiotic supplement in 100 mm square petri dishes. Plates were allowed to dry ~20 - 30 minutes, then inverted and incubated at 37 ^{o}C for 16-18 hours or alternatively for bacteria that produce larger colonies 25 ^{o}C for 24 hours. Dilution giving rise to reproducible 3-30 colonies (**2.4.3A**) or 6-60 colonies (**2.4.3B**) were used for enumeration of colony forming units (Hall et al., 2013).

2.4.4 Growth curve methodology

Chin-Yi Chen *et al.,* (2003) drop plate method and the standardised enumeration calculations as outlined by Hall *et al.,* (2013) were utilised to conduct bacteria growth curves and/or time-kills assays. Glycerol stocks were streaked onto agar media with no selective supplementation. A minimum of two biological specimens of bacteria were analysed for the growth curve from colony plates. 10 mL test species inoculum broth cultures were incubated <18 hours at 37 ^{o}C, 200 rpm under appropriate oxygen exposure. OD_{600nm} ~0.025 culture was generated for overnight bacterial cultures $T_{0\ sample}$ (Time = 0 minutes). Every + 30 minutes succeeding T_0 (for >5 hours), triplicate cell suspensions (200 µl) were serially diluted and later enumerated in accordance with **(Method 2.4.3).** Graph Pad Prism was utilised to plot Log_{10}CFU/mL OD_{600nm} [y axis] against Time [x-axis]. Growth curves for *E. coli* JE5505 and other bacteria test species were conducted to empirically determine colony forming units for the overlay assay optimisation and evaluate growth kinetics under different conditions.

2.5 METHODS FOR ANTIMICROBIAL CHARACTERISATION AND SCREENING ASSAYS

2.5.1 Preparing crude peptides

Crude peptides for literature-based or repository AMPs **(Table 2.4)** and library peptide variants of interest were obtained from Biomatik (≤5mg crude peptide vials) or GeneCust (2mg crude peptide vials). Both companies implemented solid-phase peptide synthesis and high-performance liquid chromatography for quality control of lyophilised peptides. The re-suspension of crude peptides was executed with the solvent recommended by manufacturer(s). Due to the hydrophobic nature of AMPs, crude peptide equivalents were suspended typically in 500 µl of dimethyl sulfoxide, DMSO solvent (Fisher Scientific) containing either; 100% DMSO or ddH_2O/DMSO (1:1). Crude peptide serial molar concentrations were generated, typically via half-fold serial dilutions initiating from 200 µM. The solvent used to resuspend lyophilised crude peptides was used as the diluent to generate molar concentrations. Additionally, solvent diluents were utilised as solvent controls in antimicrobial characterisation assay (agar and broth-based approaches).

Table 2.4 Physiochemical properties and sources of AMPs screened in the 0.7% Mueller-Hinton leaky *E. coli* JE5505 overlay assay or minimum inhibition assay (broth microdilution)

Name/Abbreviation(s)	Peptide sequence	Length (mers)	Charge	Hydrophobicity (%)	Isoelectric point
Peptide RLL [(2)L]	RLLFRKIRRLKR	12	7.00	41.67	13.11
Peptide KPQ [(2)L]	KPQQHNRPLRHK	12	5.00	25.00	12.53
Peptide KFV [(2)L]	KFVVWVRFLRWK	12	4.00	66.67	12.53
Peptide WRE [(2)IR]	WRERLSAPGCFLAYLL	16	1.00	75.00	7.99
Peptide SLN [(2)IR]	SLNSVLSLLDNEYHPN	16	-1.50	43.75	4.18
Peptide AFV [(2)IR]	AFVPPRPAFLHKLFCL	16	2.50	81.25	9.67
Pln-423 wt* [(1)L]	KYYGNGVTCGKHSCSVNWGQAFSCSVSHLANFGHGKC	37	4.50	56.76	8.52
Pln-H28R/H34E [(3)]	KYYGNGVTCGKHSCSVNWGQAFSCSVSRLANFGEGKC	37	3.50	56.76	8.51
Pln-S27D/K36N [(3)]	KYYGNGVTCGKHSCSVNWGQAFSCSVDHLANFGHGNC	37	2.50	56.76	7.70
Bacteriocin E50-52 [(1)D]	TTKNYGNGVCNSVNWCQCGNVWASCNLATGCAAWLCKLA	39	2.00	64.10	7.93
Cathelicidin-BF, Peptide CATHELICIDIN- [(1)D/L]	KFFRKLKKSVKKRAKEFFKKPRVIGVSIPF	30	11.00	50.00	12.35
Cathelicidin 4F buCATHL4 [(1)D]	AIPWSIWWRLLFKG	14	2.00	78.57	11.65
Caerin-2.6 [(1)D]	GLVSSIGKVLGGLLADVVKSKGQPA	25	2.00	68.00	10.50
Palustrin-Ca [(1)D]	GFLDIIKDTGKEFAVKILNNLKCKLAGGCPP	31	2.00	64.52	8.55
Lactoferricin-B1-15 [(1)D]	FKCRRWQWRMKKLGA	15	6.00	53.33	12.26
Plectasin defensin* [(1)D]	GFGCNGPWDEDDMKCHNHCKSIKGYKGGYCASAGFVCKCY	40	2.00	62.50	7.56
MP1102 [(3)L]	GFGCNGPWNEDDLRCHNHCKSIKGYKGGYCAKGGFVCKCY	40	4.00	62.50	8.24
NZZ114 [(3)L]	GFGCNGPWQEDDVKCHNHCKSIKGYKGGYCAKGGFVCKCY	40	4.00	62.50	8.24

Key: '*' refers to peptides with multiple mutated sequences, and 'X' denotes the amino acid substitutions which differ between the wildtype peptide sequence and respective mutant.

[(1)] Natural AMP and/or segment of known AMP [L] = Literature search derived or [D] = DBAASP AMP database derived.

DBAASP database derived antibacterial natural AMP sources include; Bacteria source: Bacteriocin E50-52, *Enterococcus faecium/* Pln-423 wt, *L. plantarum 423*, Reptilian source: Cathelicidin-BF (Peptide CATHELICIDIN-), *Bungarus fasciatus*, Fungi source: Plectasin defensin, *Pyronema omphalodes CBS 100304*, Animal source: Cathelicidin 4F buCATHL, *Bubalus bubalis* (water buffalo)

DBAASP ID; Bacteriocin E50-52 (5134), Peptide CATHELICIDIN- (1737), Cathelicidin 4F buCATHL4 (8618), Caerin-2.6 (1993), Palustrin-Ca (5841); Lactoferricin-B1-15 (4699); Plectasin (13059)

[(2)] Phage display derived peptide ([L]= Literature search derived); Sainath Rao *et al.*, 2013; Tanaka *et al.*, 2008 or ([IR] internal research derived peptides isolated from phage display[data not shown])

[(3)] Synthetically improved mutant of a natural AMP* Sources of literature peptides; Guralp *et al.*, 2013; Yang *et al.*, 2019

2.5.2 Broth microdilution minimum inhibitory concentration (MIC) assay

The antimicrobial susceptibility testing assay format selected for characterising activity against Gram-positive and Gram-Negative bacteria was adopted from Wiegand, Hilpret and Hancock *et al.,* (2008) and standardised guidelines for microdilution MIC assays suggested by (CLSI, 2019). As recommended Mueller-Hinton broth **(2.14)** was used for all MIC assays. Bacterial test species culture suspensions at 5 x 10^5 CFU/mL is the recommended standardised starting CFU/mL inoculum for wells in MIC assays (Andrews, 2001;Wiegand, Hilpret and Hancock et al., 2008).

Single colony material was inoculated into 5mL Mueller-Hinton broth. The inoculum was then incubated at 37ºC shaking at 225 rpm under optimum oxygen levels for ~3-5hrs until mid-log phase ($OD_{600nm.}$ 0.4-0.6). Mueller-Hinton broth bacteria cultures at mid-log phase were diluted to 0.5 McFarland (1.5 x 10^8 CFU/mL) which equates to optical densities between ~0.04-0.08 at OD_{600nm} (Hudzicki, 2009; Kralik et al., 2012). 1:200 dilution of the 0.5 McFarland suspension was used to generate bacterial cultures at ~5 x 10^5 CFU/mL. The following steps were completed within the obliged 25–30 minutes window allowed to sustain assay validity (Wiegand, Hilpret and Hancock et al., 2008). 100 µl of 5 x 10^5 CFU/mL suspensions were transferred into wells in a polypropylene 96-well plate. Optical density of seed bacteria per well at OD_{600nm} was assessed using a spectrophotometer plate reader.

At RTP, triplicates of a crude peptide molar concentrations were spiked into designated wells. Initial screening involved assessing peptide molar concentrations; 200 µM, 100 µM, 50 µM, 25 µM, 12.5 µM, 6.25 µM and 3.125 µM. Each test plate contained ≥3 wells per assay of positive controls Nisin or Polymyxin B (spiked at a concentration higher than validated MIC), solvent controls (3%/well DMSO or

H_2O/DMSO), negative controls (ddH_2O or previously characterised non-active crude peptides). In accordance with **Method 2.4.3**; growth and neat broth control wells were enumerated to assess starting inoculum cell density and contamination respectively. The 96-well plate assay was secured with a sterile PCR-grade plate lid, incubated at 37^{o}C overnight (18 hours maximum incubation), under optimum oxygen conditions for the test bacterial specimen (**Table 2.2 and Table 2.3**). 96-well plate growth control well enumeration was utilised to verify whether wells contained ~5 x 10^5 CFU/mL starting inoculum.

2.5.3 Qualitative analysis of broth minimum inhibitory concentrations

Visual analysis of the 96-well plate can be used to identify the lowest molar concentration of a peptide/positive control that exhibited inhibitory activity. Alternatively, the potency of spiked peptides across the concentration series can be indirectly accumulated from assay optical density absorbance data. Following the necessary <18hr incubation at 37^{o}C, MIC assay plates were gently agitated on 96-well plate shakers for ~5 minutes to reconstitute any clumped bacterial material in the well whereupon a spectrophotometer was used to analyse wells at OD_{600nm}.

OD_{600nm} absorbance change across triplicate wells can be an indirect estimate of changes to bacterial growth in liquid cultures. By subtracting the pre-spiking absorbance (*i.e.* seed bacteria) per well from the absorbance observed post incubation, this assimilated the bacterial growth per well which occurred post to spiking test materials (crude peptides, positive, negative, solvent controls etc). Average percentage uninhibited estimations can be garnered by simply dividing the average OD_{600nm} absorbance change of triplicate test wells by the average negative control absorbance change and multiplying by *x*100. GraphPad Prism can be utilised to visualise average

percentage uninhibited(%) [y-axis] value plotted against each molar concentration tested (3.125- 200µM, unless otherwise stated) [x-axis].

2.5.4 Tetrazolium red colorimetric broth assay

McLaughlin and Balaa, (2005) utilised 2,3,5-triphenyltetrazolium chloride (tetrazolium red) specifically to improve the visualisation of phage plaque assays against *Salmonella enterica subsp. enterica.* Tetrazolium red (white compound) is a redox indicator for cellular respiration and when reduced forms the red precipitate formazan red (McLaughlin and Balaa, 2005). For optimisation experiments, tetrazolium red was assessed with two contrasting starting inoculums; 0.5 McFarland (1.5×10^8 CFU/mL) and 5×10^5 CFU/mL suspensions generated from mid-log cultures of *S. Typhimurium* 4/74, *L. innocua* and *E. coli* JE5505. Inoculums were suspended into the same polypropylene 96-well plates (100µl per well). 0.004% to 1.0% w/v of tetrazolium red solutions were prepared (diluent; ddH_2O) and 15 µl of respective w/v tetrazolium red (company) solutions were spiked per well along with necessary controls (i.e., positive control AMPs nisin & polymyxin B). Plates were left to incubate at RTP for < 1hr to develop colorimetric shifts.

To assess the toxicity, triplicate plate wells for each tetrazolium red w/v concentration were enumerated. <0.0039% (w/v) tetrazolium red concentrations were determined as sufficiently non-toxic. 0.0019% tetrazolium red (w/v) was the lowest concentration which produced observable formazan red precipitate for both starting inoculums. Consequently, to improve the observation of growth in MIC assay wells and thus the visual determination of MICs, <0.0039% (w/v) tetrazolium red can be spiked into broth microdilution MIC assays post incubation following the completion of **Method 2.5.2** and **Method 2.5.3.** Plates should be left at RTP for 1hr to develop

colorimetric shifts, whereupon formazan red can be spectrophotometrically measured at wavelengths $OD_{490\text{-}570nm}$.

2.5.5 Verifying recombinant product from IPTG inductions

This study sought to assess recombinant peptide production from *E. coli* JE5505 growing in liquid cultures, and whether periplasm tagged peptides were leached into media following IPTG induction. Therefore, *E. coli* JE5505 cultures (50 mL MHB) were grown to mid-log (OD_{600nm} 0.6 - 0.65) and then spiked with 0.5 mM IPTG or 1 mM IPTG (T_0 sample) and returned to the 37°C shaking incubator. At the following time stamps 2 mL volumes were taken from IPTG induced broths; T_0, 1 hour, 2 hours, 4 hours, 6 hours, 8 hours and 18 hours. 1 mL / 2 mL sample was utilised to assess optical density (OD_{600nm}), whereas the remaining sample was passed through pre-washed (ddH_2O) 0.22 µm sterile filters (Merck Millipore) to generate cell-free supernatant for each time point. Filtered supernatants were stored in 1.5 mL polypropylene eppendorfs < 4 days at 4°C. Embedded lawn plates of desired indicator species were generated from 1 mL 0.5 McFarland test train culture spiked into ~27.5 mL MHA in 100mm petri dish, <4 mm agar thickness. Triplicate 5 µl, 15 µl, 50 µl and 100 µl volumes for each time point filtered supernatant was spotted onto lawn agar plates. Following this, plates were incubated at the recommended conditions for bacteria species (**Table 2.2/2.3**). A positive result was assumed to be any regions or zones of inhibition surrounding spots.

2.5.6 Agar diffusion assay: Spotting serial dilutions of antimicrobials

The peptide serial dilution agar assay optimised herein mirrors the Kirby–Bauer disk diffusion method in design and the guidelines suggested by Hudzicki, (2009); Ahman *et al.,* (2020). This assay was particularly deployed to assess peptide diffusion through agar and verify agar-based antimicrobial activity. Bacteria test species were

grown overnight from a single colony in 10 mL of Mueller-Hinton (37ºC, aerobic conditions shaking 200rpm). Overnight cultures were diluted to OD_{600nm} ~0.025 and then incubated at 37ºC, aerobic conditions shaking 200rpm to generate mid-log cultures. 0.5 McFarland standard (~1.5 x 10^8 CFU/mL) suspensions were generated from mid-log cultures. 1 mL of 0.5 McFarland broths were spiked into ~27.5 mL of Mueller-Hinton agar (45 ºC) and decanted into 100x100 mm petri dishes to achieve an agar thickness <4 mm. Crude peptide serial dilutions starting from 200 µM were prepared and 5 µl triplicate spots were plated per molar concentration. As guidelines suggest only <9 spots were allowed per agar assay to ensure no disruption or merging on zones of inhibition. Plates were inverted and placed into a static incubator at 37ºC for <18 hrs (aerobic conditions). Zones of inhibition were analysed both quantitively and qualitatively; whereby the absence/presence of zones was noted along with zone diameters.

2.5.7 Plug assay

Balouiri *et al.,* (2016) plug assay methodology was exploited herein. Bacteria test species 0.5 McFarland cultures were generated. 27.5ml molten agar (45 ºC) was spiked with 1 mL of test species 0.5 McFarland suspension and subsequently poured into 90mm or 150mm petri dishes, maintain assay <4 mm agar thickness. 5mm to 7mm agar plugs were excised and triplicate ~5-50 µl volumes of crude peptides or positive controls (nisin or polymyxin B) were transferred into plugs at molar concentrations; 200 µM, 100 µM, 50 µM, 25 µM, 12.5 µM, 6.25 µM and 3.125 µM unless stated otherwise. Plugs were additionally designated for equivalent volumes of the negative and solvent controls. Once crude peptide and controls were absorbed into the agar media, plates were then inverted and incubated aerobically at 37 ºC for 16-18 hours. The emergence of zones of inhibition surrounding a plug well was considered a

positive result. In this study, plug assays were conducted for peptides RLL, KPQ and WRE in Mueller-Hinton and Brilliant Green Agar against *S. Typhimurium 4/74* and *B. subtilis ATCC 66331*.

2.5.8 Overlay plug assays

Contrary to **Method 2.5.7**, for the overlay plug assay the agar layer spiked with the test species was soft agar and was poured to overlay the plugged agar with absorbed crude peptides. Specifically, 50 µl volumes of crude peptides at different molar concentrations were allowed to diffuse into agar from plug wells (2-3hrs at RT). Under aseptic conditions, 10mL aliquots of BGA and MHA soft agars stored at 45 ^{o}C were spiked with 1mL of *S. Typhimurium* broth culture at 0.5 McFarland. The bacteria spiked soft agar media was overlaid and petri dish lids were left slightly ajar to allow agar to cool and solidify. Plates were then incubated in aerobic conditions at 37 ^{o}C. Post overnight incubation (≤20 hrs) plates were analysed for antimicrobial activity via the visualisation of zones of inhibition in the top layer of soft agar.

2.5.9 Cross-streaking and Co-streaking methods

Cross-streaking methodology utilised in this study was modified from Carvajal, (1947). A "master streak" of bacteria test species was streaked onto 42.5 mL 2xYT/1 mM IPTG or MHA/1 mM IPTG agar in 100 mm x 120 mm petri dishes. The master streak was created by vertically streaking a mid-log broth suspension of the *pSD3* carrying *E. coli* test species four times down the middle of the agar plate. *pSD3* carrying *E.coli* strains used are summarised in Method 2.7. The negative control was *E. coli* JE5505 carrying no *pSD3* vector. Whereas positive control master streaks were created by streaking 5 µl loop of Nisin (200 µM) or Polymyxin B (200 µM). The plate containing the master streak is incubated at 30 ^{o}C for < 24 hrs, with the presence of IPTG in the agar initiating recombinant protein production in the test species. Post

24hr incubation, *S. typhimurium* or *B. subtilis* the principal indicator species were streaked at a right angle (90^{o}) in relation to the test species master streak. Plates were then incubated at 30 ^{o}C for 24 hrs. When the streaked indicator species was unable to grow towards the master streak, this was deemed a positive result. A negative result would be characterised as the indicator bacteria colony growth being in close proximity/merging with the master streak.

Co-streaking is a simplified version of the aforementioned technique except test and indicator inoculum were seeded onto the agar on the same day. This involves firstly streaking the indicator inoculum onto ~25 mL 2xYT/1mM IPTG or MHA/1mM IPTG agar plates (90 mm petri dishes). Streaks are allowed to dry, whereupon the test species inoculum is then streaked perpendicular to the initial indicator streaks. Antimicrobial activity was verified as inhibition of growth arising at the several junctions where indicator and test species co-streaks overlapped. Similar incubations, positive and negative controls were deployed as mentioned for cross-streaking were utilised.

2.5.10 Leaky E. coli JE5505 overlay assay.

The overlay assay methodology herein was adapted from Tominaga and Hatakeyama, (2006); Guralp *et al.,* (2013), and exploits *E. coli* JE5505 lpp mutant for the periplasm leaky expression of pSD3 encoded AMPs. The present iteration of the overlay assay uses 0.7% soft agar Mueller-Hinton media however various non-selective soft agar compositions (LB, 2xYT, TSB) can be exploited for this approach. The method commences with the recovery of *pSD3* carrying *E. coli* JE5505 (MHA/100µg/mL carbenicillin) and *E. coli* JE5505 wt, negative control (MHA, no carbenicillin). MHA platers were then incubated at 37 ^{o}C for ≤18 hrs. JE5505 *E. coli* strains carrying *PSD3*_peptide or JE5505 *E. coli (*negative control) colonies were

transferred into respective 10 mL MHB/1% glucose broths (pre-warmed to 37^{o}C) and incubated at 37^{o}C shaking, 200 rpm for ≤18 hrs. Overnight MHB cultures of indicator strains were prepared in a similar manner to the negative control except neat MHB was used for broth cultivation. The three main indicator strains utilised in this study were *L. innocua*, *S. typhimurium* and *S. suis* **(Table 2.2).**

Overnight cultures for *E. coli* JE5505 and the indicator strains were diluted to <0.025 (OD_{600nm}) and then grown to mid-log phase. Fundamentally, the soft agar *E. coli* JE5505 overlay assay can be summarised into five key methodological steps;

i) A mid-log broth culture of *E. coli* JE5505 is spun gently at 3500 rpm for 10 minutes (RTP), and the MHB/1% glucose broth is removed. The colony pellet is resuspended into neat pre-warmed MHB, and optical density re-assessed. This neat MHB suspension is diluted to OD_{600nm} = 0.238 (1.92 x 10^7 CFU/mL). 350 µl volumes of 1:10 serial dilutions (10^{-1} to 10^{-6}) of the OD_{600nm} = 0.238 can be plated onto the solidified 25 mL bottom soft MHA base (150mm petri dish). 350 µl volume of the 10^{-6} dilution OD_{600nm} = 0.238 generates ~50 colonies per assay, whereas for high-throughput library screening 350 µl volume of the 10^{-3} dilution suffices to produce 1.4 x 10^4 colonies per assay. 350 µl suspensions are spread plated to evenly distribute colony formation and triplicate assay plates are generated per test strain and negative control.

ii) Once plates have sufficiently dried, a 10 mL intermediate soft agar is then overlaid. Overlay assays are incubated at 37^{o}C for <18 hrs.

iii) Following *E. coli* JE5505 host colony formation another 10mL soft agar layer is overlaid spiked with 1mM IPTG and 1 mL of the indicator species at 0.5 McFarland (1.5 x 10^8 CFU/mL). Plates are once again incubated, however at 30 ^{o}C for <18hrs

(aerobic conditions). ChemiDoc MP Imaging System (Bio-RAD) was set to colorimetric (white epi light, auto exposure= 2-6s), with light exposure and contrast adjusted to ensure zones of inhibition surrounding colonies were clearly distinguished in images.

iv) Delicate aseptic excision of plugs containing desired *E. coli* JE5505 colony were utilised to isolate zone-forming colonies. Plugs (diameter <5 mm) were subsequently transferred into 10 mL MHB/1% glucose (37 ^{o}C) and incubated at 37^{o}C, 200 rpm for ≤18 hrs. Following incubation and removal of MHB/1%glucose, 50% glycerol stocks were generated by transferring 500 µl from each plug broth into a 1.5mL eppendorf containing 500 µl glycerol. Secondary verification of antimicrobial activity was achieved by repeating the aforementioned overlay assay methodology using the remaining plug overnight broth (~9.5 mL) as the starting culture for a mid-log broth, and plating 10^{-6} dilutions at OD_{600nm} =~0.2385 on the bottom base layer.

v) Colonies that produced zones of inhibition in this secondary verification overlays were picked using a pipette tip and transferred into 10mL MHB for incubation at 37^{o}C, 200 rpm for ≤18hrs. Overnight cultures were used to make 50% glycerol stocks as previously described. The remaining broth culture was minipreped and sent off for sanger sequencing. *T3* promoter sanger sequencing primers were utilised to identify the peptide sequence present in *E. coli* JE5505 isolated *pSD3* constructs.

2.5.11 Periplasm Extraction

A osmotic shock treatment, modified from Kontermann and Dübel, (2001) methodology, was used for extracting periplasmic proteins. Under aseptic conditions, 20% glycerol stocks HB2151 *E. coli* carrying *pSD3*_peptide RLL constructs and HB2151 no *pSD3*_construct (negative control) were streaked onto 25mL 2xYT/100

µg/mL carbenicillin plates. Following incubation at 37ºC for 24hrs, a single colony was transferred into 5mL 2xYT/100µg/mL carbenicillin /1% glucose broth and incubated for 18hrs at 37 ºC (200 rpm). 1mL of overnight culture was transferred into a 1.5 mL eppendorfs an spun 3000 *xg*, 10 min, RTP. Broth supernatant was removed and overnight uninduced pellet stored at -20 ºC. 500 µl of overnight broth culture was transferred into 50 mL 2xYT/100 µg/mL carbenicillin /1% glucose broth.

50 mL cultures were incubated at 37°C (200 rpm) until mid-log phase (OD_{600nm} 0.4 to 0.6). Mid-log phase 2xYT/100 µg/mL carbenicillin /1% glucose broth cultures were centrifuged at 3000 *xg*, 10 min, 4 °C. Supernatant was discarded and cell pellet was resuspended in 50 mL 2xYT/100 µg/mL carbenicillin/1 mM IPTG broth. Cultures were incubated for 18 hrs at 30ºC (200 rpm).

1mL of IPTG induced overnight culture was transferred into a 1.5mL eppendorfs, microfuge (3000 *xg*, 10min, 4ºC). Broth supernatant was removed and overnight uninduced pellet stored at RT. Remaining overnight culture was centrifuged at 3000 *xg*, 10 min, 4°C. Supernatant was discarded and pellet resuspended in 5 mL 1x PBS, then re-centrifuged at 3000 *xg*, 15 min, 4°C. Supernatant was discarded and pellet resuspended in 2 mL PPB Buffer (20% sucrose, 1 mM EDTA, 20 mM Tris-HCL, pH= 8). Resuspended pellet in PPB Buffer was transferred into 1.5mL eppendorf tubes and kept on ice for 20 mins. Cells were spun down in microcentrifuge at 6000 rpm for 5 mins. Post PPB buffer supernatant was transferred into a 1.5mL eppendorf tube and stored at RT. For osmotic shock preparation, pellets were resuspended in 1mL 5 mM $MgSO_4$ per eppendorf (1/50 of total growth volume) and then kept on ice for 20 min. Cells were spun down in microcentrifuge at 13,000 rpm for 5 min. Whereupon supernatant (periplasmic fraction) was transferred into a 1.5 mL eppendorf tube. Periplasmic fraction and respective pellet were stored at RT.

25 mL 0.7% soft MHA agar was poured into 150mm petri dishes and allowed to solidify. 5 µl triplicate volumes from the; overnight uninduced pellets (total cell), overnight IPTG induced pellets (total Cell), supernatant post PPB Buffer, supernatant and pellet post $MgSO_4$ were spotted onto 0.7% soft MHA agar. Petri dish lids were left slightly ajar for ≤ 2hrs at RT to allow pellets and supernatant to dry and absorb. *S. typhimurium* was cultured and diluted to create a 0.5 McFarland standard broth culture**.** Under aseptic conditions, 10mL aliquots of 0.35% soft MHA agar stored at 45^{o}C were spiked with 1mL of respective 0.5 McFarland *S. typhimurium* broth culture. Overlaid soft agar was allowed to cool and solidify. Plates were then incubated in aerobic conditions at 37^{o}C for 24hrs. Post incubation plates were analysed for antimicrobial activity via the visualisation of zones of inhibition in the lawn of *S. typhimurium* surrounding colony pellets or periplasm extraction spots.

2.6 GENERALISED METHOD FOR WHOLE-CONSTRUCT INVERSE PCR, BACTERIAL TRANSFORMATION AND SANGER SEQUENCING

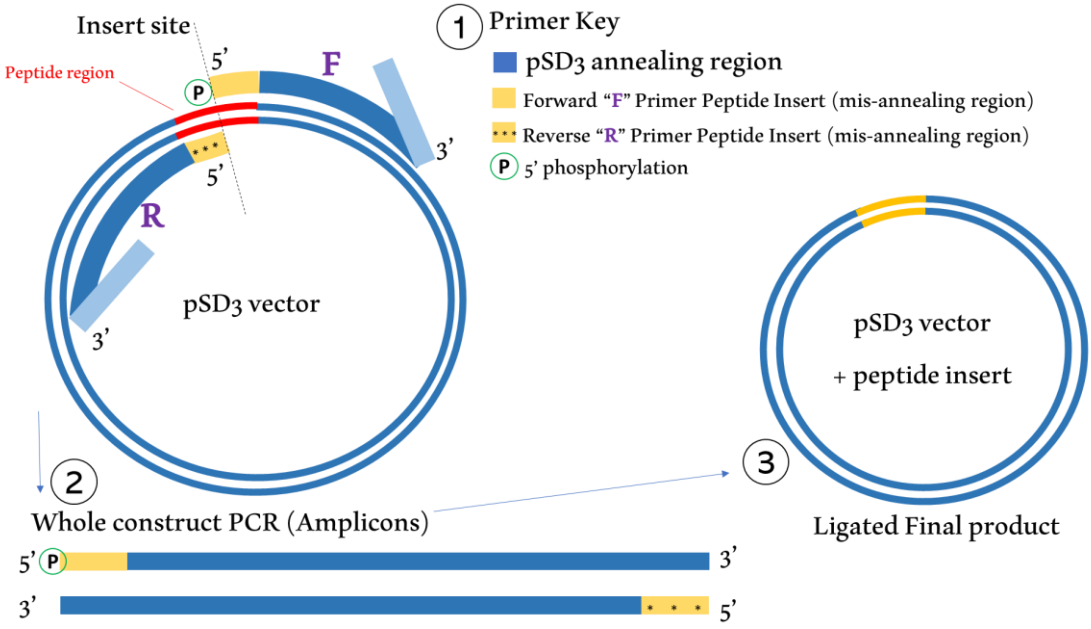
2.6.1 Inverse PCR and reaction formulation

Conventional PCR is a highly exploited molecular biology technique that involves the *in vitro* enzymatic amplification of a specific region in a nucleic acid. Unlike traditional PCR's, the oligonucleotide primers employed in inverse PCR procedures are designed to be orientated in opposite directions, as summarised in **Figure 2.2**. Inverse PCR facilitates the *in vitro* amplification of flanking sequences upstream and downstream a specific known internal region (Innis *et al.*, 1990). Thus, inverse PCR has been traditionally utilised to identify flanking regions of known

genomic sites, and facilitate insertions, mutations or delete regions in nucleic acids (Innis *et al.*, 1990;). In this study, inverse PCR was exploited with low error prone polymerase (i.e. *Q5 DNA* Polymerase) for whole-construct pSD3 phagemid amplification, wherein the desired sequence present in oligonucleotide primers (*i.e.* peptides/AMPs) now becomes inserted into PCR amplicons **(Figure 2.2).**

Figure 2.2 Mechanism of whole-construct site-directed mutagenesis using inverse PCR primers

The diagram illustrates the working principle of inverse PCR primers**. (1)** The target pSD3 vector is represented with blue circles, with the peptide region to be replaced by the desired insert depicted in red. Inverse primers (indicated by blue arrows) anneal to opposite strands and face outward from the internal region targeted for replacement (peptide region, red). The mis-annealing primer regions carry the peptide to be inserted, both the forward and reverse primers are designed to carry ~50% of peptide insert. Phosphorylated 5' end of inverse PCR primers are indicated with a circle labelled "P". **(2)** As a consequence of the primers binding to their complementary sequences, enabling whole construct amplification of both strands, linear PCR amplicons are generated. **(3)** Phosphorylation at the 5' end enhances ligation efficiency, the ligated circular final product has the full peptide sequence desired in the insert region.



All PCR reactions were performed in 50 µl volumes, except for library PCR reactions wherein in some instances 100 µl reactions were utilised. The generalised method for whole-construct inverse PCR herein utilised reaction mixtures consisting of; molecular grade nuclease free dH_2O water (GE Healthcare Hyclone, 20 µl); 5X *Q5* Reaction buffer (New England Biolabs, 10 µl); 5X *Q5* High GC Enhancer (New

England Biolabs, 10 µl); dNTP mixture [10 mM] (New England Biolabs, 2 µl); Forward primer [10 µM] (Sigma-Aldrich, 2.5 µl); Reverse primer [10µM] (Sigma-Aldrich, 2.5 µl); 2.5 µl Template DNA [10 ng/ µl] and *Q5* High-Fidelity DNA Polymerase [0.02 U/ µl] (New England Biolabs, 0.5 µl). Unless otherwise stated, the general PCR thermocycling programme used began with PCR reactions being hot started at 105^{o}C, then followed by one heating cycle at 94^{o}C or 95^{o}C for 300s to denature the template DNA, subsequent to this is 20 cycles of denaturation, 94^{o}C or 95^{o}C for 30s; primer annealing, 52 ^{o}C for 60s; and elongation, 72 ^{o}C for 330s; 10 cycles of denaturation, 94 ^{o}C for 30 s; primer annealing, 55 ^{o}C for 60 s; and elongation, 72 ^{o}C for 330s. Followed by another 10 more cycles of denaturation, 94^{o}C for 30 s; primer annealing, 58 ^{o}C for 60 s; and elongation, 72^{o}C for 330 s. Final elongation at 72 ^{o}C for 300s and PCR reactions are then held at 4 ^{o}C. < 40 cycles of amplification were utilised for whole construct inverse PCRs, and irrespective of peptide cloning the specifics of the denaturation and elongation steps remained constant nonetheless, annealing temperatures ranged between 60 - 69.5^{o}C. Inverse PCR optimisation procedure for different oligonucleotide primer sets required incremental changes of the annealing temperature through the cycles to increase stringency *i.e.,* temperature gradient PCRs (touchdown approaches etc).

2.6.2 Preparation of agarose gels for electrophoresis

Agarose gel percentages 1-3% w/v were prepared by mixing agarose powder in 1 x TAE Buffer [40 mM Tris, 20mM Acetate and 1mM EDTA, pH ~8.6] followed by heating the agarose/TAE to aid agarose dissolving and dispersing. Agarose/TAE was cooled (~50 ^{o}C), and Nancy 520 (Sigma-Aldrich, 01494) added to give a final concentration of 1 µg mL^{-1}. Succeeding the gel setting, well combs are removed and the agarose gel is transferred to an electrophoresis chamber with 1 x TAE buffer.

2.6.3 Preparation of DNA sample for agarose gel electrophoresis

DNA samples were mixed with dilutions of 6x purple loading dye [10mM EDTA, 3.3mM Tris-HCl, 0.02% w/v Dye 1, 0.001% w/v Dye 2 at pH 8] (New England Biolabs, B7024S). The ratio of sample to loading dye depended on the sample volume capacity of the wells present in the agarose gel. **2.6.3A:** Agarose gels cast with 8 well combs; well maximum volume capacities ~30-35 µl. Typically, electrophoresis for these agarose gels were performed at 70-100V for 45-60 minutes. **2.6.3B:** Agarose gels cast with 12, 16 or 20 well combs (well maximum volume capacities ≥20 µl) were used. **2.6.3A**. 8 µl of 1kb DNA ladders were typically added to each well. Contrastingly, in **2.6.3B** 3-4 µl of 1kb DNA ladder is added to each well. Once electrophoresis is complete, agarose gels were removed from the 1x TAE buffer flooded chamber and then gel images were captured under the UV, Ethidium bromide exposure settings.

2.6.4 PCR and Gel Extraction Clean-up

The Macherey-Nagel^{Tm} NucleoSpin^{Tm} Gel and PCR Clean-up Kit (12303368, Fisher Scientific) exploits silica membrane technology and chaotropic salt binding; DNA binds to the silica columns (high salt environment), contaminates are removed (high salt environment/ethanol washes) and then DNA is eluted from the column (low salt environment). Manufacturer instructions were utilised with only two alterations to reduce ethanol carryover which can limit downstream enzymatic reactions. Firstly, following the second wash of the silica membrane with NT3, an ethanol-based wash buffer, the silica column is spun for 3 minutes instead of the recommended 1 minute. Succeeding this, silica membranes are placed on a heat block at 70^{o}C ≤5 minutes to evaporate residual ethanol. Manufacturer's optional method recommendations were followed; NE, elution fresh buffer was heated to 70^{o}C for 5 mins to increase DNA

recovery of large fragments. Total volume of 30 µl of NE buffer (70ºC) was split equally across two elution steps. Macherey-Nagel^{Tm} NucleoSpin^{Tm} Gel and PCR Clean-up Kit was used for recovering DNA from enzymatic digestion reactions (*i.e. DpnI* etc). NT1 binding buffer is diluted with nuclease free dH_2O water at a ratio of 1:5 (NT1:dH_2O) to remove smaller undesired fragments whilst maintaining high levels of extraction for larger fragments (i.e. *pSD3* digestion products). The Macherey-Nagel^{Tm} NucleoSpin^{Tm} Gel and PCR Clean-up Kit (12303368, Fisher Scientific) was utilised to perform agarose gel extract clean-up. 1-3% agarose gels were visualised using a UV transilluminator, and bands of desired size excised. The mass of gel extracts were determined, and the volume of NTI buffer required was calculated as follows; 100mg of agarose requires 200 µl of NTI ∴ [1mg agarose = 2 µl NTI]. The volume of NTI required doubles for agarose gels with >2% (w/v).

Bijous or 50mL containers containing gel extracts were incubated in a water bath set to 50 ºC for a total of 10 mins, samples were removed from the water bath and vortexed briefly at the 3 min, 6 min and 8 min incubation intervals. NTI buffer + gel extract sample was transferred into columns then, centrifuged for 30 s at 11, 000 *xg* and the flow-through was discarded. For each centrifugation step ≤650 µl of NTI buffer + gel extract sample was run through columns. Following this 650 µl of NT3 (ethanol wash buffer) was pipetted into each column. Columns were centrifuged for 30s, 11, 000 *xg*. Once complete the flow-through was discarded. Columns were centrifuged for 3 minutes at 11, 000 *g* to remove any residual NT3 buffer. 30 µl of NE buffer (70ºC) was transferred into each column and left for 5 minutes at RT. Columns were then centrifuged for 60s at 11, 000 *g*. Repeat elution steps were conducted by heating the column to 70ºC for 3 minutes, run-through DNA elution from previous

steps was transferred back into the column and centrifuged once again (60s at 11, 000 *g).*

2.6.5 Quantification of DNA

Qubit™ dsDNA (High-sensitivity, HS) and (Broad-range, BR) Assay Kits and Qubit 2.0 Fluorometer were used to quantify DNA in samples. This assay relies on the use of fluorescent dyes to identify the concentration of DNA present by comparing the level of fluorescence observed in a sample to the fluorescence signals of standards with known DNA concentrations. Per sample to be quantified, 199 µL of Qubit dsDNA HS/BR Assay Buffer was mixed with 1 µL fluorophore dye reagent (provided in the kit). 2 µL DNA sample was transferred to a kit provided tube containing 198 µL of the Qubit dsDNA HS/BR Assay Buffer and dye working solution. Tubes were mixed gently (flicking) and incubated at RTP for 5 mins. Whereupon the Qubit 2.0 Fluorometer 2.0 (Invitrogen) was used to read samples and produce a standard curve for DNA concentration quantification based on fluorescence.

2.6.6 DpnI Digestion

DpnI is a TypeII restriction enzyme that recognises and cleaves DNA containing methylated adenines in the sequence: 5'- GATC- 3', 3'-CTAC -5'. Therefore, PCR reactions herein were *DpnI* (New England Biolabs, R0176S) restriction digested to remove methylated DNA templates. *DpnI* digestions were implemented as recommended by supplier; 10 units *DpnI* to 1µg DNA in a total reaction mixture of 50 µl and incubating reactions at 37^{o}C for 15 minutes whereupon *DpnI* was heat inactivated at 80^{o}C for 20 minutes.

2.6.7 T4 Ligase Ligation

The DNA prepared from Method **2.6.5-2.6.6 , were** utilised in a ligation reactions facilitated by *T4* Ligase (New England Biolabs, M0202). Manufacture's

reaction mixture structure was utilised; total volume of ligation equalling 20 µl (*T4* DNA Ligase 2 µl/20) and the DNA concentration being 2.5ng/ µl. To commence the ligation, ***2.6.7A:*** samples were then incubated ≤48hrs at RTP (~20^{o}C) or ***2.6.7B:*** samples were incubated at RTP for 2hrs. *T4* ligation reactions are then heat inactivated at 65^{o}C for 10 minutes. A singular 0.025µM Millipore membrane filter (Fisher Scientific, VSWp02500) was used per ligation sample to help remove undesired sample components such as *T4* ligase buffer salts. 0.025µM Millipore membrane filter were placed onto a 90mm petri dish filled with ~25mL of molecular grade nuclease free dH_2O water for 30 minutes at RTP. ~20 µl ligation end product were collected from the filter and then stored at -20^{o}C (in 0.5 mL eppendorf) or used immediately for bacterial transformation.

2.6.8 SpeI restriction enzyme digestion

SpeI is a restriction enzyme that recognises 5'-AC^TAGT-3' cutting between the C and T nucleotides, producing overhangs useful for sticky end ligation. Herein restriction *SpeI* digestion of dual *SpeI* RE sites in modified *pSD3* constructs was utilised to cleave the phage gene -*gIII* where appropriate (e.g., for non-phage display library screening approaches). *SpeI* digestion reactions were conducted on ~1 µg of DNA, New England Biolabs Time-Saver qualified *SpeI* enzyme (NEB, R0133), 5 µl of 10X rCutSmart Buffer (final concentration of 1X. 1.0 µl of *SpeI* enzyme, 10 units) and nuclease-free water added to bring the total reaction volume to 50 µl. Reaction mixtures were incubated at 37°C for 5-15 minutes or an extended 1-hour digestion. If needed, overnight digestions can be safely performed using Time-Saver qualified enzymes. To avoid star activity (non-specific cleavage) caused by excess glycerol, the enzyme volume did not exceed 10% of the total reaction volume. Heat inactivation (80^{o}C) was performed before proceeding to clean-up and DNA purification step

(Method 2.6.4). *SpeI* digest DNA clean-ups were then quantified **(Method 2.6.5)** and digested DNA was then ligated **(Method 2.6.**7) for downstream applications.

2.6.9 Electrocompetent cells

Electrocompetent cells were created for *E.coli* TG1, HB2151, BL21 and JE5505 strains. Glycerol stocks for each *E. coli* strain were streaked onto agars and incubated under optimum conditions for 24hrs **(Table 2.3).** Under aseptic conditions a single colony was collected then transferred into 10 mL of 2xYT Broth [pre-warmed 37^{o}C] **(2.1.15).** Broths were incubated aerobically at 37^{o}C, shaking 250 rpm for 18hrs. 250 µl of this overnight broth culture was transferred into 250 mL of 2xYT Broth [pre-warmed 37^{o}C] contained in a 500 mL conical flask.

Flasks were then incubated aerobically at 37^{o}C, shaking 250 rpm for ~3-4hrs. 1mL of broth culture samples were assessed periodically to identify when cultures reached mid-log phase (0.5-0.6 at OD_{600nm}). Mid-log phase cultures were then placed on ice for 15 minutes and under aseptic conditions in a laminar flow cabinet, 125 mL of the chilled flask cultures were transferred into four 250 mL centrifuge bottles (chilled to 4^{o}C). Samples were chilled on ice then centrifuged in a rotor chilled to 4 ^{o}C spun at 5000 rpm, for 10 minutes. Succeeding each chilled centrifugation broth or 10% glycerol supernatant separated from the cell pellet was removed and aspirated. Before the second and third centrifuge steps (5000 rpm for 10 minutes), 125 mL of 10% glycerol solution was added to each centrifuge bottle and cell pellets were reconstituted via gentle swirling. ~10mL of 10% glycerol solution was retained per centrifuge bottle to re-suspend pellets. 100 µl volumes of the electrocompetent cells were transferred into autoclaved 0.5 mL eppendorfs. Samples are then stored at -80 ^{o}C.

2.6.10 Electroporation

Competent cells of the desired *E. coli* strain **(Table 2.3)** and phagemids (≥100 ng/ µl) were thawed and defrosted on ice along with the 2mm electrocuvttes (SLS, FBR-202). 2mm electrocuvettes were used for *E. coli* JE5505 whereas 1mm cuvettes were used for other *E. coli* strains (TG1 *etc*). Under aseptic conditions 2 µl of plasmid was transferred and then mixed into 50-85 µl of competent cells contained in a 0.5 mL eppendorf. Mixing was achieved by gently flicking the eppendorf. The competent/plasmid sample was then left on ice for 2 minutes and then transferred into a pre-chilled 2mm electrocuvette. The sample was pulsed using the electroporation machine, voltages of 2.5 kV or 2.0 kV were chosen for the bacterial strains utilised herein. Following this 1 mL of SOC outgrowth medium (New England Biolabs, B9020S) was added to revive pulsed cells. Electrocuvettes were inverted gently ~6 times and depending on the size of petri dishes used, 50-100 µl of the pulsed competent cells were then spread plated onto petri dishes containing the 25-40 mL of Carbenicillin (100 µg/mL) agar; 2xYT: HB2151, TG1, BL21 *E.coli*; MHA: JE5505 *E.coli* **(Table 2.3)**. Agar spread plates were allowed dry and then incubated aerobically at 37 ^{o}C for 24hrs to form single colonies carrying the desired plasmid construct. For titrations 100 µl volume of the recovered transformations were 1:10 diluted in SOC medium then plated on 2YT/150 µg/mL ampicillin/1% glucose agar plates and incubated at 37 °C static overnight to enumeration of transformation efficiency.

2.6.11 Minipreps

A single pSD3 transformant colony **(2.6.10)** was transferred into 5-10 mL of broth and incubated aerobically overnight (≤18 hrs) at 37 ^{o}C, shaking at 200-220 rpm. Transformant overnight broth cultures were briefly vortexed (~10s) whereupon ~5 mL of overnight cultures were transferred into sterile 50 mL centrifuge tubes. The

remaining 5 mL suspension was utilised to generate 20 or 50% overnight glycerol stocks of transformants**.** Minipreps were then conducted using The QIAprep Spin Miniprep Kit (QIAGEN, 27104), and the manufacture's recommended starting volume of 5 mL. The protocol utilised is identical to the kit supplier's instructions apart from the fact another centrifugation step (top speed ~17,000 *xg*) for 3 minutes was completed to remove any residual ethanol-based wash buffer. The DNA concentration present in the 50 µl minipreps were then quantified **(2.6.5).**

2.6.12 Sanger sequencing of plasmid constructs

Sanger sequencing is reliant on the chain termination method; amplified complementary DNA (cDNA) is derived from DNA polymerase random incorporation of deoxynucleotide triphosphates, dNTPs and/or dideoxynucleotide (ddNTPs) (fluorescently labelled) (Grada and Weinbrecht, 2013; Crossley et al., 2020). cDNA amplicons vary in length and are capped sequentially at each successive base because rate-limiting ddNTP incorporation terminates nucleotide synthesis (Grada and Weinbrecht, 2013; Crossley et al., 2020). Chromatogram analysis facilitates the identification of nucleotide bases, and currently upper limits of <800-1000bp length sequence reads are supported by Sanger sequencing (Crossley et al., 2020).

To prepare samples for sanger sequencing method **2.6.12A** was utilised: 5 µl of DNA samples **(2.6.11)** at 100 ng/ µl were sent off for Sanger sequencing (Source Bioscience, UK). Sanger sequencing quality control was utilised to analyse inverse PCR peptide inserts, RE site inserts or to check the carriage of the correct construct in a transformant. Due to the location of the peptide region in pSD3 **(Figure 2.1)**, the *T3* promoter primer 5'-AATTAACCCTCACTAAAGG-3' [3.2pmol/µl] (Source Bioscience, UK) was selected as the primer site to initiate sanger sequencing for

analysis peptide inserts. **2.6.12B:** Contrastingly, if all of the pSD3 phagemid required sanger sequencing then 30 µl of the DNA sample at 100 ng/ µl was sent off for sequencing. The primers utilised to Sanger sequence the entire pSD3 construct are summarised in **Table 2.5.**

Table 2.5 Sanger sequencing primers utilised to map pSD3 phagemid vector

Primer name	**Primer Sequence 5' to 3'**	**Source**
T7 Promoter Forward	5'– TAATACGACTCACTATAGG- 3'	(Source Bioscience, UK)
M13 Promoter Forward	5'– GTTTTCCCAGTCACGAC- 3'	(Source Bioscience, UK)
F1 ori Reverse	5'– GACGTTGGAGTCCACGTTC- 3'	(Sigma-Aldrich, UK)
VHH_Check 1F	5' – CGTATTACAATTCACTGGCCG- 3'	(Sigma-Aldrich, UK)
Amp273F sequencing/qPCR primer	5'– TATGCAGTGCTGCCATAACCA-3'	(Sigma-Aldrich, UK)
Amp273R sequencing/qPCR primer	5'– AACTTTATCCGCCTCCATCC-3'	(Sigma-Aldrich, UK)
VHH check 3F primer	5'– CCCTTAACGTGAGTTTTCGTTCC - 3'	(Sigma-Aldrich, UK)
BSPQ1 deleted Forward Primer	5'– CGTAGCTGCCCAATACGCAAAC-3'	(Sigma-Aldrich, UK)
BSPQ1 sequencer Forward Primer	5'–TTTCGCCACCTCTGACTTGAG-3'	(Sigma-Aldrich, UK)
VHH check 5F primer	5' – GGACGCAGGTGACAGTTT - 3'	(Sigma-Aldrich, UK)
M13 Promoter Reverse	5' –AGCGGATAACAATTTCACACAGGA- 3'	(Source Bioscience, UK)
T3 Promoter Reverse	5' – AATTAACCCTCACTAAAGG- 3'	(Source Bioscience, UK)
* 5ul (3.2pmol/ul) volumes of in-house primers were sent off for Sanger sequencing (Source Bioscience, UK)		

2.7 Inverse PCR cloning AMPs into pSD3 vector construct

2.7.1 Literature and repository database AMPS cloned into pSD3 phagemid for the optimisation of screening approaches.

Guralp *et al.*, 2013 demonstrated the ability to visualise the antimicrobial activity of the Pln-423 wt and library mutants' peptide in the JE5505 leaky *E.coli* overlay

assay. To investigate the efficacy of Guralp *et al.,* 2013 overlay methodology this study transformed **(2.6.10)** JE5505 leaky *E.coli* with *pSD3* phagemid constructs encoding recombinant peptides including Pln-423 and two selected improved mutants **(Table 2.6 / Table 2.7)**.

Additionally, included were 12 mer peptides identified by phage display approaches Sainath Rao, Mohan and Atreya, (2013) and Tanaka, Kokuryu and Matsunaga, (2008). MIC assays were conducted to verify antimicrobial activity against target test strains using purified, lyophilized Peptide RLL, KPQ and KFV among other crude peptides dissolved in H_2O/DMSO (1:1) **(2.5.2-2.5.3)**. 16mers peptides identified from previous phage display experiments were additionally included [*data not shown*]. Furthermore, these three peptides demonstrated activity against *S. Typhimurium* in minimum bactericidal concentration assays [*data not shown*]. The full set of peptides characterised via MIC assays and subsequently designed for inverse PCR cloning into the pSD3 vector are summarised in **Table 2.6.**

The inverse PCR primers for Pln-423 and Peptides KFV, WRE, SLN and AFV among others were designed as previously mentioned in **(Method 2.5/Figure 2.2).** PCR amplicons were run on 1% agarose gels and desired bands (~3.9kb) were excised. Gel extraction clean-ups were performed to elute PCR product into 30 µl of elution buffer (**2.6.4**). The eluted DNA was then subjected to *DpnI* digestion **(2.6.6)**. Following this the PCR-clean up methodology was used to remove smaller digest fragments and heat-inactivated *DpnI* enzyme. *T4* ligation was then conducted on *DpnI* digested samples to circularise the linear constructs (**2.6.7**). ~20 µl *T4* heat-inactivated ligations of pSD3_ Pln-423 or Peptide KFV/WRE/SLN/AFV samples were utilised to transform 50 µl volumes of *E. coli* JE5505 electrocompetent cells (**2.6.10**). ~100 µl of SOC-media recovered *E. coli* JE5505 pulsed (3.0kV) cells were spread plate onto

~25mL MHA (Ampicillin 100µg/mL) plates. Once dry, plates were incubated aerobically for 24hrs at 37°C. Several transformant colonies for each pSD3_peptide construct were transferred into 50 mL falcon tubes containing 10mL MHB. Overnight MHB broth cultures were used to create 50% glycerol stocks and then minipreped for Sanger sequencing (**2.6.11 to 2.6.12**). Translational analysis of sanger sequence data was used to verify clones with the correct peptide inserts.

Table 2.6 Inverse PCR primers for Pln-423 and *Guralp* et al., (2013) improved mutants.

	Inverse PCR: Forward primer	**Inverse PCR: Reverse primer**
Pln-423 wild type	**[PHOS]5'**GGCCAGGCGTTTAGCTGCAG CGTTAGCCATCTGGCGAACTTTGGCCA TGGCAAATGCTAATCATGCCAGTTCTT TTGGC **3'** *(79bp, Tm = 78 °C, GC% = 53)* 5'3' Frame 1- *Amino acid seq* GQAFSCSVSHLANFGHGKC-SCQFFW	**5'**CCAGTTAACGCTGCAGCTATGTTTGCC GCAGGTAACGCCGTTGCCATAATATTTG GCCATCGCCGGCTGGGCCGC **3'** *(75bp, Tm = 80 °C, GC% = 59)* 3'5' Frame 1 - *Amino acid seq* AAQPAMAKYYGNGVTCGKHSCSVNW
Pln-H28R/H34E	**[PHOS]5'**GGCCAGGCGTTTAGCTGCAG CGTTAGCCGTCTGGCGAACTTTGGCGA AGGCAAATGCTAATCATGCCAGTTCTT TTGGC **3'** *(79bp, Tm = 79 °C, GC% = 54)* 5'3' Frame 1- *Amino acid seq* GQAFSCSVSRLANFGEGKC-SCQFFW	**5'**CCAGTTAACGCTGCAGCTATGTTTGCC GCAGGTAACGCCGTTGCCATAATATTTG GCCATCGCCGGCTGGGCCGC **3'** *(75bp, Tm = 80 °C, GC% = 59)* 3'5' Frame 1 - *Amino acid seq* AAQPAMAKYYGNGVTCGKHSCSVNW
Pln-S27D/K36N	**[PHOS]5'**GGCCAGGCGTTTAGCTGCAG CGTTGATCATCTGGCGAACTTTGGCCA TGGCAACTGCTAATCATGCCAGTTCTT TTGGC **3'** *(79 bp, Tm = 78 °C, GC% = 53)* 5'3' Frame 1 – *Amino acid seq* GQAFSCSVDHLANFGHGNC-SCQFFW	**5'**CCAGTTAACGCTGCAGCTATGTTTGCC GCAGGTAACGCCGTTGCCATAATATTTG GCCATCGCCGGCTGGGCCGC **3'** *(75bp, Tm = 80 °C, GC% =59)* 3'5' Frame 1 - *Amino acid seq* AAQPAMAKYYGNGVTCGKHSCSVNW

Key:
XXX = *PelB* C-terminus annealing region (for the Reverse primer)
XXX = Peptide encoding region
XXX = *pSD3* region downstream peptide insert region. (for the Forward primer)
[PHOS] = Phosphorylating the 5' end of inverse PCR primers improves ligation efficiency, prevents self-ligation, and enables directional cloning.

Primer design: - refer to Figure 2.2, *"G" refers to the mid-point insert amino acid in each Pln-423wt/mutant, this was utilised to segment the peptide region to be carried by either the forward or reverse primer.*
Pln-423: KYYGNGVTCGKHSCSVNWGQAFSCSVSHLANFGHGKC
Pln-H28R/H34E: KYYGNGVTCGKHSCSVNWGQAFSCSVSRLANFGEGKC
Pln-S27D/K36N: KYYGNGVTCGKHSCSVNWGQAFSCSVDHLANFGHGNC

Table 2.7 Inverse PCR primers for literature or repository database derived AMPs

	Inverse PCR: Forward primer	**Inverse PCR: Reverse primer**
E50-52	**[PHOS]5'**GTGTGGGCGAGCTGCAAC CTGGCGACCGGCTGCGCGGCGTGG CTGTGCAAACTGGCGTAATCATGC CAGTTCTTTTGGCTAGCTAA **3'** *(86bp, Tm = 82ºC , GC% = 60)*	**5'**GTTGCCGCACTGGCACCAGTTC ACGCTGTTGCACACGCCGTTGCC ATAGTTTTTGGTGGTGGCCATCGC CGGCTGGGCCGC **3'** *(81bp, Tm = 83 ºC, GC% = 64)*
Cathelicidin-BF	**[PHOS]5'**GAATTTTTTAAAAAACCG CGCGTGATTGGCGTGAGCATTCCG TTTTAATCATGCCAGTTCTTTTGGC TAGCTAA **3'** *(74bp Tm = 72 ºC , GC% = 41)*	**5'**TTTCGCGCGTTTTTTCACGCTTT TTTTCAGTTTGCGAAAAAATTTGG CCATCGCCGGCTGGGCCGC **3'** *(66bp, Tm = 75 ºC, GC% = 50)*
Cathelicidin-4F buCATHL4F	**[PHOS]5'**TGGCGCCTGCTGTTTAAA GGCTAATCATGCCAGTTCTTTTGGC TAGCTAA **3'** *(50bp, Tm = 70 ºC , GC% = 46)*	**5'**CCAAATGCTCCACGGAATCGCG GCCATCGCCGGCTGGGCCGC **3'** *(42bp, Tm = 78 ºC, GC% = 71)*
Caerin-2.6	**[PHOS]5'**CTGGCGGATGTGGTGAAA AGCAAAGGCCAGCCGGCGTAATCA TGCCAGTTCTTTTGGCTAGCTAA **3'** *(65bp, Tm = 76 ºC , GC% = 52)*	**5'**CAGGCCGCCCAGCACTTTGCCA ATGCTGCTCACCAGGCCGGCCAT CGCCGGCTGGGCCGC **3'** *(60bp, Tm = 84 ºC, GC% = 73)*
Palustrin-Ca	[**PHOS]5'**ATTCTGAACAACCTGAAA TGCAAACTGGCGGGCGGCTGCCCG CCGTAATCATGCCAGTTCTTTTGGC TAGCTAA **3'** *(74bp, Tm = 77 ºC , GC% = 51)*	**5'**TTTCACCGCAAATTCTTTGCCG GTATCTTTAATAATATCCAGAAA GCCGGCCATCGCCGGCTGGGCCG C **3'** *(69bp, Tm = 77 ºC, GC% = 52)*
Lactoferricin-B1-15	**[PHOS]5'**CGCATGAAAAAACTGGGC GCGTAATCATGCCAGTTCTTTTGGC TAGCTAA **3'** *(50bp, Tm = 70 ºC, GC% = 46)*	**5'**CCACTGCCAGCGGCGGCATTTA AAGGCCATCGCCGGCTGGGCCGC **3'** *(45bp, Tm = 79 ºC, GC% = 71)*
NZZ114	**[PHOS]5'**AGCATTAAAGGCTATAAA GGCGGCTATTGCGCGAAAGGCGGC TTTGTGTGCAAATGCTATTAATCAT GCCAGTTCTTTTGGCTAGCTAA **3'** *(89bp, Tm = 75 ºC, GC% = 44)*	**5'**TTTGCAATGGTTATGGCAGCGC AGATCATCTTCGTTCCACGGGCC GTTGCAGCCAAAGCCGGCCATCG CCGGCTGGGCCGC **3'** *(81bp, Tm = 82 ºC, GC% =63)*
MP1102	**[PHOS]5'**AGCATTAAAGGCTATAAA GGCGGCTATTGCGCGAAAGGCGGC TTTGTGTGCAAATGCTATTAATCAT GCCAGTTCTTTTGGCTAGCTAA **3'** *(89bp, Tm = 75 ºC, GC% =44)*	**5'**TTTGCAATGGTTATGGCATTTC ACATCATCTTCCTGCCACGGGCC GTTGCAGCCAAAGCCGGCCATCG CCGGCTGGGCCGC **3'** *(81bp, Tm = 81 ºC, GC% =60)*
Plectasin, defensin	**[PHOS]5'**AGCATTAAAGGCTATAAA GGCGGCTATTGCGCGAGCGCGGGC TTTGTGTGCAAATGCTATTAATCAT GCCAGTTCTTTTGGCTAGCTAA **3'** *(89bp, Tm = 76 ºC, GC% =46)*	**5'**TTTGCAATGGTTATGGCATTTC ATATCATCTTCATCCCACGGGCC GTTGCAGCCAAAGCCGGCCATCG CCGGCTGGGCCGC **3'** *(81bp, Tm = 80 ºC, GC% = 58)*

Key:
XXX= *PelB* C-terminus annealing region (for the Reverse primer)
XXX= Peptide encoding region
XXX= *pSD3* region downstream peptide insert region (for the Forward primer) "TAATCATGCCAGTTCTTTTGGCTAGCTAA", Aa: -SCQFFWLA this was slightly longer than outlined in **Table 2.6** increasing the length targeted for annealing was to compensate the slightly larger peptides required for insertion.
[PHOS] = Phosphorylating the 5' end of inverse PCR primers improves ligation efficiency, prevents self-ligation, and enables directional cloning.

Similar strategy for primer design as outlined in **Table 2.6,** the peptide region was separated equally where possible across both the forward and reverse primer. See Table 2.4 for the peptide encoding regions for respective peptides.

2.8 METHODS FOR GENERATING A TRAINING SET OF ANTIMICROBIAL PEPTIDES AND IMPLEMENTING *IN-SILICO* ANALYSIS TO EVALUATE AMINO ACID DISTRUBUTION

2.8.1 Manually extracting repository AMPs from The Antimicrobial Peptide Database 3, (APD3).

Several online free-access AMP databases exist, and all vary in curation, annotation and repository criteria for AMP collection (Kang et al., 2019). "The Antimicrobial Peptide Database 3, APD3" (http://aps.unmc.edu/AP/, accessed from 1st May 2018) at present contains ≥3435 natural AMPs, with 84.49% of peptides annotated with antibacterial activity. Source organisms range from all six life kingdoms further exemplifying the ubiquity of these host defense peptides in nature, in the ADP3 database sources include; Bacteria (~11.30%), Archaea (~0.15%), Protists (~0.24%), Fungi (~0.67%), Plants (~11.00%), Animals (~74.05%, synthetic peptides accounting for 2%). APD3 is a regularly manually curated online database, and well cited repository of natural AMPs with defined amino acid sequences and biological activities denoted via antimicrobial testing assays such as MICs (Wang *et al.*, 2016; Kang et al., 2019). A relatively unrestricted rudimental search (*peptide length:* ≤50 mers, *Net charge:* any value, *Hydrophobicity:* any value, *D- and unnatural amino acids:* none) was conducted in APD3 to identify AMPs with known activity (MIC: ≤100μM or ≤100μg/mL) against pig pathogens in **Table 2.8.** Under the search parameters, 187 AMPs were selected from the APD3 **(Appendix Figure 9.1)** and collectively defined in a dataset which will be referred to as "APD3_ AMP_PigPathogenDataset", see **Table 2.9**.

Table 2.8 Key pathogenic bacteria in pigs prioritised for natural AMP selection from the ADP3 database.

Gram reaction	Gram-Negative bacteria	Gram-Positive bacteria
Pathogenic bacteria of pigs	*E.coli* *S.typhimurium* *S. Choleraesuis* *B.pilosicoli* *B.hyodisintariae* *C. jejuni* *C. coli*	*C. difficile* *C. perfringens (types A and C).* *S. suis*

Table 2.9 Summarised characteristics of natural AMPs in the "ADP3_AMP_PigPathogenDataset"

Feature	**Value**	**Additional information**
Total No. peptides identified from naïve search of ADP3	187	~5.46% of the total AMP sequences in APD3. Peptide length: ≤50 mers, no D- or unnatural amino acids,
Antimicrobial activity* of peptides	*MIC: 0.25-100µM* *MIC: 0.2-62µg/mL*	Annotated activity* search; MIC/MBC ≤100µM or ≤100µg/mL)
AMPs with activity* stated against pig bacterial pathogen	*Salmonella spp* (2.5%) *Clostridium spp* (1%) *E. coli* (97%) *S. Typhimurium* (~36%) *C. jejuni* (3.5%) *B. pilosicoli* (0%) *B. hyodisintariae* (0%) *C. coli* (0%) *C. difficle* (0.5%) *C. perfringens* (1.5%) *S. suis* (0%)	*i.e. S. enetrica (I); Enteritidis, Choleraesuis* Species level identification not provided stated '*clostridia*'
Spectrum of Anti-bacteria activity	Gram-negative only (60 peptides) Gram-positive only (1 peptide) Broad spectrum (141 peptides)	
* "Annotated Activity" relates to MIC values from antimicrobial assays conducted in the literature article source of each AMP in APD3.		

2.8.2 Generating control AMP and Non-AMP training datasets

An excel tool was developed to randomly extract a defined number of samples from a defined list (*i.e*. peptide sequence dataset). This randomisation selector tool was applied to ~3435 AMP sequences downloaded from ADP3 (Wang et al., 2016) and 9777 curated Non-AMPs database (Bhadra *et al.,* 2018). Bhadra *et al.,* (2018) collected 9777 non-AMPs from Uniport to train an AMPprediction tool (AMPep). 187 sequences were randomly extracted to form the "APD3_RandomSelection187" and "NonAMPep_RandomSelection187" datasets. These two datasets were generated to; *a)* identify whether the sampling selection of AMPs from APD3 *i.e.* selecting on the condition the AMP displayed activity against specific bacteria pig pathogen(s) resulted in any observable differences and *b)* further delineate amino acid and positional distribution differences between AMPs and non-AMPs.

2.8.3 In silico tools for analysing antimicrobial peptides; data mining and characterisation of antimicrobial peptides

Computational analysis of AMPs is vital for classification, characterisation and design. Numerous computational tools were utilised to analyse AMP amino acid composition and physiochemical properties including; web-based peptide analysis, spreadsheet programmes, coding programming such as Python and R. As summarised **in Table 2.10,** a variety of useful AMP specific databases and free-access tools; web, naïve applications or programmes exist for the depository or analysis of AMPs (Osorio et al., 2015; Waghu et al., 2016). All databases and tools listed **in Table 2.10** were utilised to some extent during this study. DBAASP v3, R "Peptides" package outlined by Osorio *et al.,* (2015), and some Perl scripts mentioned in **Method 2.11** were primarily utilised for AMP physiochemical property characterisation; sequence length (size), amino acid composition, net charge, isoelectric point, boman index, hydrophobicity index, hydrophobic moment index *etc*.

DBAASP v3 property calculation via the Eisenberg and Weiss scale was implemented for physiochemical analysis (Vishnepolsky and Pirtskhalava, 2014). Osorio *et al.,* (2015) provides the ability to change parameters for equations such as hydrophobic moment unlike DBAASP v3. Osorio *et al.,* (2015) AMP tool requires R Statistics 4.2.1 programme: language and environment for statistical computing. To install the "Peptides" package, in a new R Statistics 4.2.1 workspace a) enter the following command into the R console > install.packages ("name of package"), wherein desired package is "Peptides" or b) in the R menu tab go to Packages and select 'Install package(s)' in the menu drop-down. Following (b) step will be the selection of a secure CRAN mirror downalod server (UK, London 1 :https selected) and confirmation that "Peptides" is the desired package for install. Once downloaded to utilise this package > require (Peptides) or > library (Peptides) can be used to initiate the package in the R workspace.

Net charge: Amino acid side chains confer electrical charges under certain pH environments, the sum of charges provides the 'Overall net charge'. Isoelectric point: pH at which a particular molecular is electrically neutral, which is reliant on dissociation constants (pK_a) of charged amino acids and peptide termini (NH+ and COO-). Bjellqvist optimised pK_a scale was selected in the Osorio et al., (2015) "Peptides" package due to the relative prediction accuracy of this scale as demonstrated by Kozlowski, (2021) and similarity to Eisenberg *et al.,* 1982 adopted method in DBAASP v3. Instability index/Boman Index: Estimation of *in vivo* stability and due to their short nature AMPs with instability index values <40 . The Boman index reflects protein interactions, with values <0 signifying a weak protein-protein interaction, which is untypical for AMPs as mechanism of action centre upon protein-membrane interactions (Osorio et al., 2015). Hydrophobicity index/Hydrophobic

moment index: hydrophobicity is vital for protein folding stabilisation and the moment index is a quantitative measure of the amphiphilicity of a peptide structures, such as the a-helix or b-sheet. (Osorio et al., 2015). The moment index Eisenberg scale with $\delta = 100^{o}$ angle, as utilised by Giangaspero *et al.*, (2001) for the analysis of α-helical AMPs.

XXX = Peptide package script input.

> charge(seq = "One-letter AMP amino acid sequence", pKscale = “Bjellqvist”)

> pI(seq = "One-letter AMP amino acid sequence", pKscale = "Bjellqvist")

> instaindex(seq = "One-letter AMP amino acid sequence")

> boman(seq = "One-letter AMP amino acid sequence")

>hydrophobicity(seq = "One-letter AMP amino acid sequence", scale = "Eisenberg")

>hmoment(seq = "One-letter AMP amino acid sequence", angle = 100, window = 11)

Table 2.10. AMP online databases and a selection of computational tools utilised for AMP *in silico* analysis

In silico tool or AMP database	**Function(s)**	**Reference**
ExPASy	*ExPASy Protoparam:* Analysis of multiple physical and chemical parameters for proteins; molecular weight, MW, theoretical pI, amino acid composition, grand average of hydrophobicity, GRAVY) *ExPASy Translate:* DNA sequence translation into peptide amino acid sequence considering Frame 5’3’ (1, 2, 3) and Frame 3’5’(1, 2, 3)	(Gasteiger *et al.*, 2005) rp https://web.expasy.org/translate/
EMBOSS Pepstats	Web based analysis tool for protein characterisation (MW, isoelectric point, probability of expression in inclusion bodies). Number of amino acids with certain properties (Tiny, polar, basic, acidic *etc* as seen in Osorio *et al.*, 2015)	(Rice et al., 2000) https://www.ebi.ac.uk/Tools/seqstats/emboss_pepstats/
R statistics package	“*Peptides” package;* computational analysis of) 23 different possible variables including; MW, amino acid composition, net charge, isoelectric point, membrane position, indices: aliphatic, instability, Boman, hydrophobicity, hydrophobic moment. Biophysical chemical simulations (GROMACS)	(Osorio *et al.*, 2015)
Peptide Synthesis and Proteotypic Peptide Analyzing Tool	Assess the difficulty of peptide synthesis and key peptide characteristics (hydrophobicity percentage).	https://www.thermofisher.com/uk/en/home/life-science/protein-biology/peptides-proteins/custom-peptide-synthesis-services/peptide-analyzing-tool.htmL
PepDraw	Tool to visualise peptide primary structure and calculate various theoretical properties (isoelectric point, hydrophobicity, net charge)	http://www2.tulane.edu/~biochem/WW/PepDraw/

Uniport	Free access database resource for protein sequences, >5.5 x 10^5 (AMP and non-AMPs) and related functional information. BLAST/align peptides to sequences within the database.	https://www.uniprot.org/
ADP3	Natural AMP online database (>3000 sequences) with amino acid sequences, antimicrobial activity and structural characterisation. AMP designer and predictor tool.	(Wang et al., 2016) https://aps.unmc.edu/home
CAMP/CAMPR3	AMP database containing AMP amino acid sequences structures and signatures. CAMPR3 bolsters AMP machine learning prediction tools; Random Forest; RF, Artificial neural network; ANN, Support Vector Machine; SVM and Discriminant analysis; DA	(Thomas *et al.*, 2010; Waghu *et al.*, 2016) http://www.camp3.bicnirrh.res.in/index.php
DBAASP/DBAASP v3	DBAASP is an AMP database (>19,000) with analysis tools including; AMP properties (hydrophobic moment, net charge, isoelectric point etc), structural information (chemical, 3D). Prediction tools for; general antibacterial activity (Machine learning programme based on Moon-Fleming scale), linear AMP activity against specific bacteria species or strain-level.	(Gogoladze et al., 2014; Pirtskhalava et al., 2021) https://dbaasp.org/

2.8.4 In silico pipeline for analysing antimicrobial peptides; amino acid composition and hydrophobic stretches.

The classification criteria of amino acids can vary on size (A^3), chemical structure, hydropathy and electronic charges (Katchalski-Katzir et al, 2006). In this study the twenty standard amino acids are classified primarily on hydropathy and charge into three main groups, collectively termed "AMP- amino acid grouping" *a)* hydrophobic (polar or non-polar) amino acids (Aa's): G, A, V, L, I, P, F, M, W, C, and Y; *b)* basic (positively-charged side chains) polar amino acids: K, R, and H, *c)* polar amino acids which are either acidic (negatively-charged side chains) or neutrally charged and/or hydrophilic; S, N, Q, T, E and D. In some instances, the three stop codons are aligned into the third AMP-amino acid grouping, as they are desired to a lesser extent and/or in *E. coli* stop suppressor strains utilised herein generate readthrough contributions to this group. A python-based script was constructed to conduct amino acid compositional analysis of *n* number of peptides with varying lengths. The script was specifically implemented on AMP datasets specifically to enumerate hydrophobic and

basic amino acids. Additionally, a key focus of the script was to analyse and identify sequential stretches of two or more hydrophobic amino acids.

```
with open("output01.txt") as b:
   for amp, ref in zip(b, e):
      d.write(ref)
      for ind, char in enumerate(amp[:-1]):
         next_char = amp[ind + 1]
         if char in hydrophobic and next_char in hydrophobic:
            temp = temp + char
         else:
            if len(temp) > 0:
               temp = temp + char + "\n"
               d.write(temp)
               temp = ''
            else:
               pass
 d.close()
e.close()
```

'with open ("filename")' Python syntax instruction to open file for amp, ref in zip(b, e)'. An iterable is an object which contains continuous variable *i.e.* it contains either; 0, 1 or many (>2) elements. Zip function 'zip(b, e)' turns iterables into arguments, this combines b/e datasets and identifies pairs of items. 'for' can loop over an iterable. "Ind, char in enumerate", is utilised to loop across strings within characters "char". This aids scanning over each amino acid residue in a peptide sequence and returns an index and value. In this context, value equates to hydrophobicity classification. 'd.write(ref)' instructs Python to use write mode 'w' to transcribe ref to denote the stretches observed for each peptide identify if present or "pass".

```
f = open("tally_of_aa_in_amps.txt", 'w')
 def aa_count(input_file, aa_group):
    with open(input_file) as b:
        content = b.read()
        return dict([(i, content.count(i)) for i in aa_group])
 hyd = aa_count("output01.txt", hydrophobic)
pol = aa_count("output01.txt", polar)
bas = aa_count("output01.txt", basic)
aci = aa_count("output01.txt", acidic)
 f.write("Hydrophobic Amino Acids: \n")
 for key, value in hyd.items():
    text = str(key) + " : " + str(value) + "\n"
    f.write(text)
 f.write("Polar Amino Acids: \n")
 for key, value in pol.items():
    text = str(key) + " : " + str(value) + "\n"
    f.write(text)
 f.write("Basic Amino Acids: \n")
   for key, value in bas.items():
    text = str(key) + " : " + str(value) + "\n"
    f.write(text)
 f.write("Acidic Amino Acids: \n")
 for key, value in aci.items():
    text = str(key) + " : " + str(value) + "\n"
    f.write(text)
```

The frequency of amino acids “Aa” in AMPs were derived as a tally and this file is instructed to be opened. (input file) The input file is compared to “aa_group” file containing the list of amino acids with each physiochemical property “hyd” – hydrophobic , “pol” – polar, “bas” – basic, “acid” – acidic. Aa_count function is utilised to enumerate all amino acid groupings. Directed counts are then formulated into the designated output files “output01.txt”

2.9 METHODS FOR THE DESIGN OF AN "AMP-BIASED" NOVEL DEGENERATE CODON SCHEME

2.9.1.1 *Site-directed mutagenesis and degenerate codon scheme selection*

The NNN scheme summarised in **Figure 2.3** (unaltered genetic code, 20 amino acids via 61 codons), where "N" represents any of the four nucleotide bases A/T/G/C, is inherently redundant and various schemes exist to improve this facet of gene diversification (Sieber et al., 2015;Lindenburg et al., 2020). Site-saturation schemes which are frequently explored in peptide library synthesis include NNB, NNK **(summarised in Figure 2.4)** or NNS (B: G/T/C; K: G/T; S: G/C), all encode 20 amino acids with reduced codon numbers; 48 (B) and 32 (K/S) respectively (Sieber et al., 2015). Nonetheless, withstanding redundancy (≥8 amino acid redundancies exist within the codon mixture and the presence of one stop codon (amber stop, TAG) in NNK/NNS schemes persists as bottlenecks for both schemes (Sieber et al., 2015)

Figure 2.3 Amino acid codon wheel for NNN (Unaltered genetic code: "N" = A/T/G/C). NNN contains 61 codons for twenty amino acids and three stops; TAA (opal), TAG (amber), TGA (ochre).

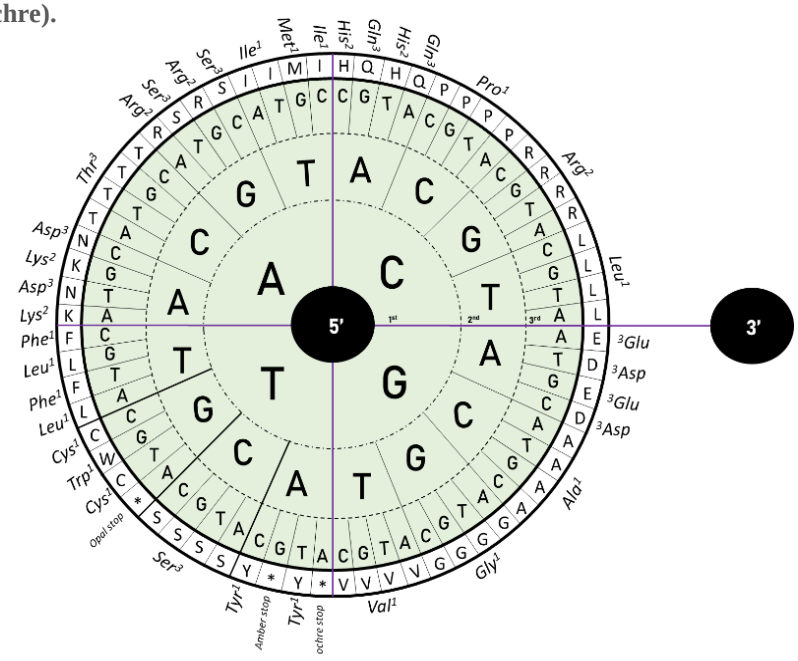
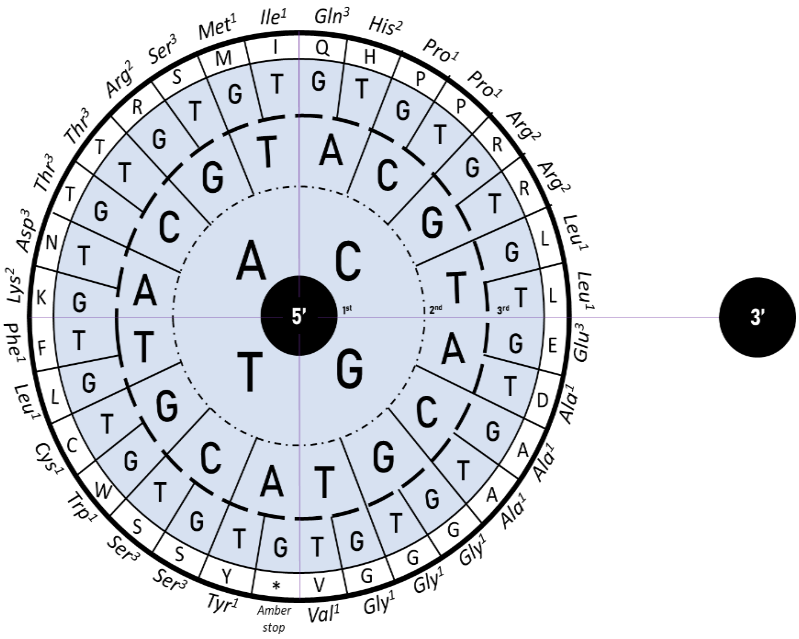


Figure 2.4 Amino acid codon wheel for NNK ("K" = G/T). The NNK scheme restricts the codon subset of the unaltered genetic code to 48. Amber stops (TAG) are still retained in this scheme.

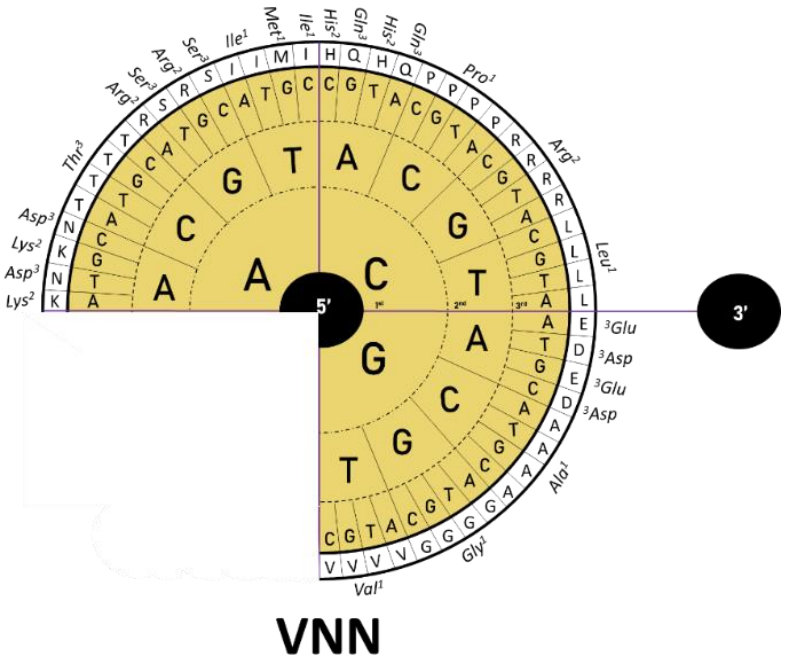


To eliminate or reduce bottlenecks of the NNK degenerate codon scheme, in this study we pursued the application of the VNN scheme (48 codons :16Aa), where "V" = A/G/C, N:A/T/G/C to design a simple AMP-biased library via a novel degenerate oligonucleotide primer approach. The VNN scheme was designed to match the amino acid distribution identified in the natural "AMP_PigPathogenDataset". The VNN degenerate scheme is summarised in **Figure 2.5.** Essentially, VNN degeneracy restricts the occurrence of the thymine oligonucleotide base in the first codon position. This removes twelve amino acid encoding codons; four codons encoding serine, two codons each encoding phenylalanine, leucine, tyrosine and cysteine respectively and one codon for tryptophan. Resultingly, the VNN scheme reduces the genetic code to sixteen amino acids and forty-eight synonymous codons **Figure 2.5**. Additionally,

unlike NNK/S approaches, the VNN scheme beneficially excludes all three stop codons TAG (amber), TAA (ochre) and TGA (opal/umber).

Figure 2.5 Amino acid codon wheel for the VNN degenerate scheme, where "V" = A/G/C (No T) and N: A/T/G/C.

Exclusion of nucleotide "T" in the first position in a codon triplet eliminates the representation of all three stop codons and four amino acids (C, F, W, Y), whilst reducing redundancy by only retaining two serine codons (AGT, AGC). The number of codons for the 16 encoded amino acids is equivalent to NNN and NNK schemes (**Figures 2.3 and 2.4**)



In the VNN scheme, retains six Arg encoding codons within the VNN scheme, twice as many as NNK/S schemes respectively however, this redundancy might be beneficial due to the desired basic characteristics and prevalence of R in the APD3_AMP_PigPathogensDataset. VNB where "B" = T/G/C was considered due to lower redundancy rations (36:16). However, only 1/2 Lys codons were retained by the VNB scheme. This was putatively disadvantageous as the basic charged lysine was the most frequently observed amino acid residue in the APD3_AMP_PigPathogensDataset. The VNN/VNB schemes both encode 9/10 of the top ranked amino acids observed in the APD3_AMP_PigPathogensDataset . The four amino acids not represented by the VNN scheme and their ranked occurrence in

APD3_AMP_PigPathogensDataset are as follows; phenylalanine (F; ranked 8th), cysteines (C; ranked 11th), tryptophan (W, ranked 16th) and tyrosine (Y, ranked 19th).

2.10 METHODS FOR 16 MER PEPTIDE LIBRARY CONSTRUCTION

2.10.1 Modifying pSD3 phagemid vector library template

The phage display libraries in this study were randomised 16mer peptides fused to the minor coat protein (pIII) of *M13* phage. pIII library construction required a distinct phagemid vector template, which was a modified version of the pSD3 vector shown in **(Figure 2.1).** Initial modifications of the pSD3 vector entailed removing a pre-existing *BSPQI* site (5'GCTCTTCC 3', 3'CGAGAAGGCGA 5') located at 2738-2748bp. Aforementioned inverse PCR reaction mixture utilising Q5 DNA polymerase kit (NEB) and pSD3_RLL as the DNA template in 50 µl total reaction was implemented **(Method 2.6.1).** Primers utilised for this were; *BSPQI removal*_Forward primer: 5' CGTAGCTGCCCAATACGCAAAC 3' and *BSPQ1 removal* Reverse primer: 5' AGCTACGCGCTTCCTCGCTCAC 3'. PCR reactions were hot started at 105°C, then followed by one heating cycle at 95°C for 180s to denature the template DNA, subsequent to this is 20 cycles of denaturation, 95 °C for 30 s; primer annealing, 58 °C for 30 s; and elongation, 72 °C for 140s; Final elongation at 72 °C for 300 s and PCR reactions are then held at 4 °C. PCR reaction mixtures were run on a 1% agarose gel. Bands of desired size were excised, and subsequent to this a PCR gel clean-up was conducted following manufacturer instructions. DNA was eluted into 30 µl NE buffer, and DNA concentrations were quantified **(2.6.4/2.4.6).** *DpnI* digestion was then completed **(2.6.6),** reactions mixtures were then cleaned up and DNA concentration quantified **(2.6.4/2.4.6).** 50ng of DNA with 2 µl of *T4* DNA ligase

were ligated in 20 µl total reaction for 18hs at 16ºC followed by incubation 6h at RT **(2.6.7).** Millipore filters were used to clean-up ligation reaction mixtures **(2.6.7)**. DNA was then electroporated into JE5505 *E.coli* **(2.6.9),** transformants were picked and cultured in broth in preparation for minipreps **(2.6.11).** Removal of undesired *BSPQI* site was determined by sanger sequencing with the following primer 5'–CGTAGCTGCCCAATACGCAAAC-3' (BSPQI deleted primer, **2.6.11**).

The overlay assay screening approach does not require the random 16mer peptides to be fused to phage protein pIII, therefore further modifications to the *pSD3_RLL_-BspQI_gIII* entailed placing a secondary *SpeI* restriction enzyme followed by an opal(umber) stop codon TGA downstream of *gIII*. Consequently, *SpeI* restriction enzyme digestion of this modified pSD3 vector would remove of the *gIII,* and upon re-ligation the constructs contain the 16-mer peptide region succeeded by TAG (amber) and TGA stop codons to terminate transcription. Primers outlined in **Table 2.11** were utilised to insert the *SpeI* site-opal stop codon into pSD3 vector were;

Table 2.11 Primers designed for the insertion of *SpeI*(opal stop "TGA") downstream of *gIII*

Primer	**Oligonucleotide primer sequence [1]**	**(bp)**	**Features**
***SpeI*TGA Forward primer:**	**5'**AGTTGATCATGCCAGTTCTTTTGGCTAG CTAATAATTATGCCTAGAGGTG **3'**	50	T_m = 67.9 ºC GC%= 40%
***SpeI*TGA Reverse primer:**	**5'**AGTTTAAGACTCCTTATTACGCAGTATG TTAGCAAACGTAGAAAATACATACAT **3'**	54	T_m = 65 ºC GC%= 31%
[1]All primers were manufactured by Sigma-Aldrich			

Inverse PCR reaction mixture utilising Q5 DNA polymerase and pSD3_RLL(-BSPQI) as the DNA template in 50 µl total reaction was implemented **(2.6.1)**. PCR thermocycling programme; PCR reactions were hot started at 105ºC, then followed by one heating cycle at 94ºC for 300s to denature the template DNA, subsequent to this is 20 cycles of denaturation, 94ºC for 30s; primer annealing, 55ºC for 60s; and

elongation, 72°C for 330s; 10 cycles of denaturation, 94°C for 30s; primer annealing, 58°C for 60s; and elongation, 72°C for 330s. Followed by another 10 more cycles of denaturation, 94°C for 30s; primer annealing, 60°C for 60s; and elongation, 72°C for 330s. Final elongation at 72°C for 300s and PCR reactions are then held at 4°C. Cloning strategy described above for the generation of the *pSD3_RLL_-BspQI_gIII* construct were repeated. Sanger sequencing via the *SpeI* forward primer was utilised to verify the correct *SpeI*-opal stop codon insertion in the *pSD3_RLL_-BspQI_gIII* construct.

2.10.2 $VNN_{15}(TTT)_1$ and NNK Library Construction

The strategy for library construction distinctly utilises principles from both Kong *et al*, (2020); and Tsoumpeli *et al.*, (2022), with the specific aim of generating large diverse naïve 16 mer peptide libraries. Wherein the randomised peptide diversity was introduced to whole construct amplicons of the library template *pSD3_RLL(-BSPQI, + SpeITGA)*, modified to support a simplified molecular cloning strategy which not only facilitates phage peptide library construction (Kong et al., 2020; Tsoumpeli et al., 2022) but additionally alternative downstream screening methodologies such as Guralp et al., (2013) *E. coli* JE5505 overlay assay. **Figure 2.6** summarises the library construction process, and the primer design and PCR protocols employed are summarised in **Table 2.12**. Both the $VNN_{15}(TTT)_1$ and NNK libraries were constructed with an identical reverse primer and distinct randomised forward generate primers as shown in **Table 2.12**. Additionally, forward degenerate primers were phosphorylated on the 5' end to facilitate downstream enzymatic activity summarised in **Figure 2.2**. The sixteen $VNN1_{5(}TTT)_1$ forward primers were distinguished from one another by the position of the encoded phenylalanine (TTT) across the 16mer peptide region as shown in **Table 2.12.**

Table 2.12 $VNN_{15}(TTT)_1$ and NNK degenerate primer design and Inverse PCR thermocycling

Degenerately randomised primers and respective annealing temperatures for successful amplification of whole pSD3 constructs				
Primer	**Oligonucleotide primer sequence [1]**	**(bp)**	**Features**	**PCR thermocycling**
Universal library pSD3 reverse primer	**5'**GGTAGCTCTTCGGGCGGCCATCGCCGGCTGGGCCGCCAGTAACAG **3'**	45	T_m = 84ºC GC%= 71.1%	**1 cycle** 95ºC, 3 min **25 cycles** ***Denature:*** 95ºC, 30s ***Anneal:*** 59.2ºC OR 62.2ºC, 60s NNK – (59.2ºC) $VNN_{15}(TTT)_1$ – (62.2ºC) ***Extension:*** 72ºC, 5.5min **1 cycle** Final extension: 72ºC. 5 min Hold at 4oC ∞
Forward NNK randomised degenerate primers	**5'**GATTGCTCTTCGGCC$(NNK)_{16}$ACTAGTTCTAGTGGTGGC **3'**	81		
Forward $VNN_{15}(TTT)_1$ randomised degenerate primers	**5'**GATTGCTCTTCGGCC$[(VNN)_{15}TTT_1]$ACTAGTTCTAGTGGTGGC **3'**	81		
[1] = All primers were manufactured by Sigma-Aldrich				

Universal library pSD3 reverse primer	5'GGTAGCTCTTCGGGCGGCCATCGCCGGCTGGGCCGCCAGTAACAG 3'
Forward NNK randomised degenerate primers	*P5'GATTGCTCTTCGGCCNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKACTAGTTCTAGTGGTGGC 3'
Forward $VNN_{15}(TTT)_1$ randomised degenerate primers	*P5'GATTGCTCTTCGGCCTTTVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNACTAGTTCTAGTGGTGGC 3' *P5'GATTGCTCTTCGGCCVNNTTTVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNACTAGTTCTAGTGGTGGC 3' *P5'GATTGCTCTTCGGCCVNNVNNTTTVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNACTAGTTCTAGTGGTGGC 3' *P5'GATTGCTCTTCGGCCVNNVNNVNNTTTVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNACTAGTTCTAGTGGTGGC 3' *P5'GATTGCTCTTCGGCCVNNVNNVNNVNNTTTVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNACTAGTTCTAGTGGTGGC 3' *P5'GATTGCTCTTCGGCCVNNVNNVNNVNNVNNTTTVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNACTAGTTCTAGTGGTGGC 3' *P5'GATTGCTCTTCGGCCVNNVNNVNNVNNVNNVNNTTTVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNACTAGTTCTAGTGGTGGC 3' *P5'GATTGCTCTTCGGCCVNNVNNVNNVNNVNNVNNVNNTTTVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNACTAGTTCTAGTGGTGGC 3' *P5'GATTGCTCTTCGGCCVNNVNNVNNVNNVNNVNNVNNVNNTTTVNNVNNVNNVNNVNNVNNVNNVNNVNNACTAGTTCTAGTGGTGGC 3' *P5'GATTGCTCTTCGGCCVNNVNNVNNVNNVNNVNNVNNVNNVNNTTTVNNVNNVNNVNNVNNVNNVNNVNNACTAGTTCTAGTGGTGGC 3' *P5'GATTGCTCTTCGGCCVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNTTTVNNVNNVNNVNNVNNVNNVNNACTAGTTCTAGTGGTGGC 3' *P5'GATTGCTCTTCGGCCVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNTTTVNNVNNVNNVNNVNNVNNACTAGTTCTAGTGGTGGC 3' *P5'GATTGCTCTTCGGCCVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNTTTVNNVNNVNNVNNACTAGTTCTAGTGGTGGC 3' *P5'GATTGCTCTTCGGCCVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNTTTVNNVNNACTAGTTCTAGTGGTGGC 3' *P5'GATTGCTCTTCGGCCVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNTTTVNNACTAGTTCTAGTGGTGGC 3'

Key; XXX (Nucleotide bases/codon restriction): Annealing region and/or specific purpose

GATT = Nucleotides required for *BspQI/SapI* (Isoschizomers) to attach to DNA sequence.
A/T/G/C = *BspQI/SapI* restriction enzyme (RE) site.
GCC = Encodes Alanine (A) amino acid, replicating the last amino acid present on the 3'of the *pelB* leader sequence
A/T/G/C[1] = Anneals to *PelB* leader sequence
NNK/VNN = Degenerate randomised peptide region, where N = A/T/G/C, K = G/T, V = A/G/C.
TTT = Phenylalaine amino acid residue introduced at 1-16 codons in the 16 mer peptide region
ACTAGT [1] = *SpeI* RE site
A/T/G/C[1] = 5' start of the *gIII* phage gene

[1] = denotes a primer region 'X, X, X' which specifically anneals to the pSD3 vector template. The remainder of the primer mis anneals to the template and contains our desired peptide diversity.

2.10.3 Library cloning strategy

pSD3_RLL_-BspQI_gIII_+SpeI_opal stop construct was generated **(Method 2.10.1)** and used as the library cloning template. **Figure 2.6** summarises the key steps in the library cloning strategy.

Figure 2.6 Schematic overview of the inverse-PCR method used for generating randomised $VNN_{15}+(TTT)_1$ and NNK 16mer peptide libraries.
(A)"Phage vector library template": Simplified construct map provided for the *pSD3_RLL_-BspQI +SpeI_opal stop* phagemid library template with *SpeI* RE sites flanking the 5' and 3' end-terminus of the *gIII* gene **("pSD3 Phagemid Library Template ~3.94kb").** The vector with PelB highlighted in blue **"pSD3_RLL Phagemid library template",** has the same vector components but for ease of explanation only the PelB to opal stop are represented. **"Degenerate primers"** visualises the key aspects of the degenerate forward primer design; 5' starts with the *SapI site+CCG* codon followed by randomised 16mer peptide region and the 3' primer end complements the *pSD3* template at the 5' start of *gIII*. $VNN_{15}+(TTT)_1$ forward primers have one phenylalanine encoded at position 1-16, the first primer encodes phenylalanine on the first codon of the peptide region, whereas the sixteenth primer encodes phenylalanine on the last codon.

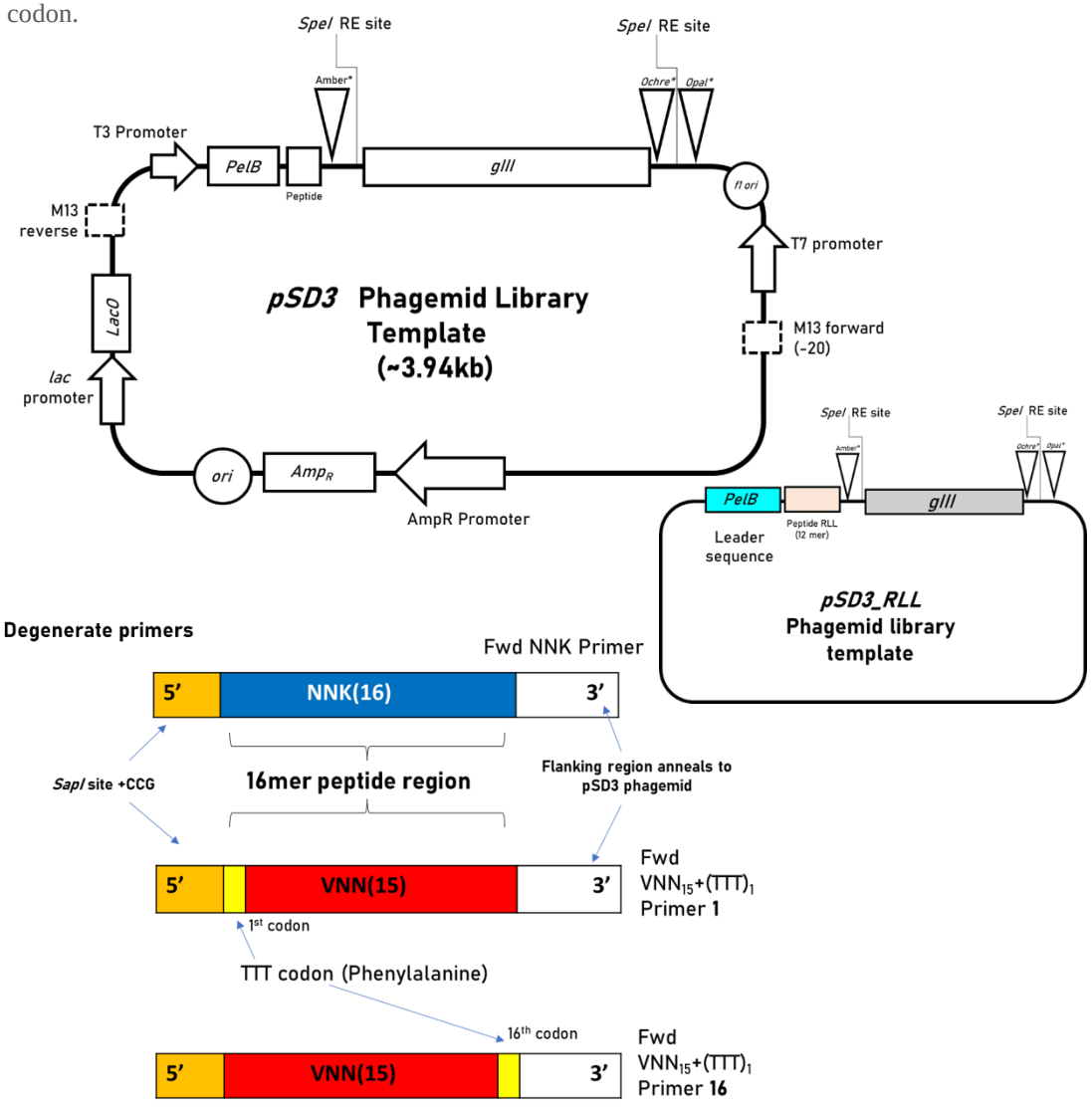


Figure 2.6 (B) "Inverse PCR" illustrates regions of the reverse and degenerate forward primers which anneal (White sections) or misanneal (coloured regions of primer) to library templates. **"Linear PCR library amplicons"** *SapI+(CCG or GGC codons)* are at the 3' end of respective reverse or forward primer thus becoming the terminal residues at opposing ends of the linear PCR amplicon.

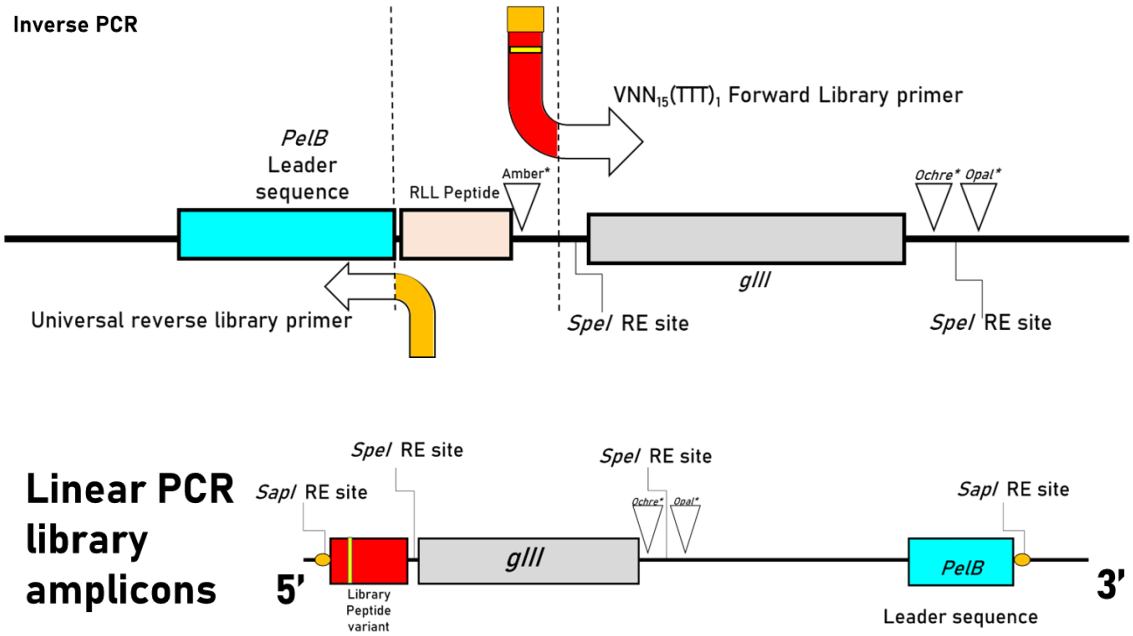
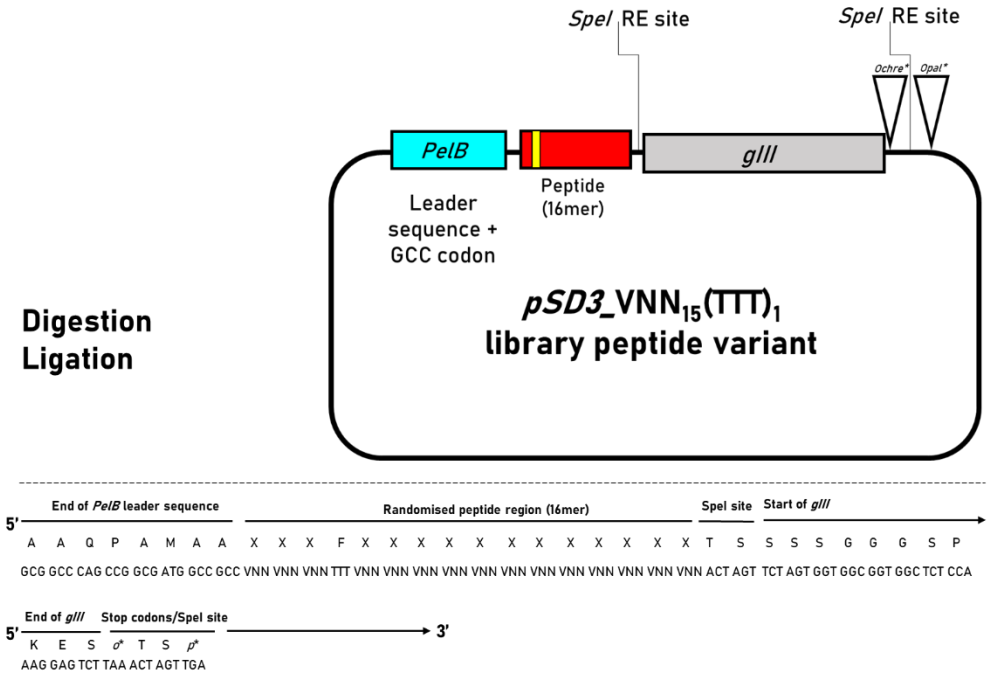


Figure 2.6(C) "Digestion/Ligation" – depicts the circular library amplicons generated following *DpnI* and *SapI* digestion and subsequent ligation. Importantly, peptide diversity in degenerate primers are introduced into whole-plasmid amplicons of the *pSD3_-BspQI_+SpeI_opal stop* construct.



2.10.4 Library Inverse PCR

Library inverse PCR reactions mixtures were formulated using the Hot Start High-Fidelity DNA Polymerase kit (NEB) as summarised in **Method 2.6**. Annealing gradient PCRs were conducted from 59-69.5^{o}C to determine the highest annealing temperature with sufficient amplification. The primers and finalised PCR protocol for $VNN_{15}(TTT)_1$ and NNK libraries are summarised in **Table 2.12**. The inverse PCR programme utilised consisted of; 95°C for 3 mins, succeeded by 25 cycles of 95°C for 30s, 59.2°C or 62.2°C for 60s and 72°C for 5 min and 30s, followed by an incubation at 72°C for 5 mins. Library PCR products were run on 1% agarose gels **(2.6.2)** and cleaned up using the The Macherey-Nagel^{Tm} NucleoSpin^{Tm} Gel and PCR Clean-up Kit (Fisher Scientific); protocol was implemented according to manufacturer's instructions, recommended method steps to improve elution were adopted **(2.6.4).** DNA concentrations were quantified using the Qubit dsDNA HS Assay Kit (ThermoFisher) **(2.6.5).** Double restriction digestion with the restriction enzymes *DpnI* and *SapI* (NEB) was then conducted. 1µg of DNA in total reaction mixture of 50 µl was digested according to the NEB recommendations, reactions were incubated for 1hr at 37°C then 20 mins at 80^{o}C to inactivate the *DpnI* and *SapI* restriction enzymes. DNA was then cleaned up using the The Macherey-Nagel^{Tm} NucleoSpin^{Tm} Gel and PCR Clean-up Kit and quantified via the Qubit dsDNA HS Assay Kit (ThermoFisher)**.**

Ligations were formulated to contain 1-2µg of DNA along with 4 µl of *T4* DNA ligase (NEB) in a 200 µl total reaction volume. Reactions were initially incubated for ≤6hrs at 4^{o}C and then incubated at 16^{o}C for 18hrs, followed by 6hr incubation at RT. DNA was cleaned up using Nucleospin kit (Macherey Nagel), manufacturer's instruction was followed **(2.6.4)** however for the final elution 20 µl Ultrapure

molecular biology grade H_2O was used instead of the NE buffer. Repeat elution steps were conducted to improve the elution of DNA from the column, subsequent to this the concentration of DNA was measured with the Qubit dsDNA HS Assay Kit **(2.6.5).**

2.10.5 Library electroporation and quality control

1µg of library DNA was transformed into 50 µl of highly competent TG1 *E. coli* electrocompetent cells (Lucigen,60502-1). 1mL of pre-warmed (37ºC) SOC Outgrowth Medium (New England Biolabs) was immediately added to pulsed TG1 *E. coli* cells. Recovery was achieved by incubating cells for 45 minutes at 37ºC, shaking at 200rpm. Post recovery for library titration 1 µl volume of transformants was transferred into 99 µl of Ultrapure molecular biology grade H_2O, serial dilution down to 10^{-6} was generated. 100 µl volume of serial dilutions were plated onto ~25mL 2xYT/150µg/mL ampicillin/1% glucose agar plates for transformation efficiency calculations.

The remaining recovered transformation volume was spread plated onto a large bioassay dish containing ~125mL 2xYT/150µg/mL ampicillin/1% glucose agar. All 2xYT/150µg/mL ampicillin/1% glucose agar plates were inverted and incubated at 37°C for 16-18hrs. Transformant colonies were scraped off plates, whereupon 25% glycerol stocks were generated and stored at -80ºC. Transformant colonies on the library titration plates were individually picked and transferred to 5mL 2xYT/1% glucose broth, and were incubated at 37°C for 18hrs, 200-220rpm. Library pSD3 plasmids were extracted from overnight broth cultures via the Qiagen Miniprep kit using the according to manufacturer's instructions **(2.6.11).** Plasmids were then Sanger sequenced (Source Bioscience) to verify pSD3 plasmid conformation and randomised 16mer peptide region.

2.10.6 E. coli JE5505 sub-library production

pSD3_-BspQI +*SpeI*_opal stop VNN15+(TTT)1 and NNK *E. coli* TG1 glycerol stocks were midiprepped **(Method 2.6.11)** using E.Z.N.A.® M13 DNA Mini Kit. DNA concentrations were then quantified via Qubit **(Method 2.6.5).** >30µg library DNA obtained from the midiprep of 1000 µl *E. coli* TG1 VNN_{15}+$(TTT)_1$ and NNK library glycerol stocks were then subsequently *SpeI* digested to remove the *gIII* in library pSD3 constructs **(Method 2.6.8).** Gel electrophoresis was performed to visualise the ~679bp *gIII* restriction digestion fragment of library DNA. Nonetheless, gel visualisation of *SpeI* library digests intended for ligations was omitted and the verification of complete and correct digestion was conducted under brief UV exposure during gel excision of the desired ~3.37kb band. *SpeI*-digested sub-library DNA was then ligated and <10µg of respective circular DNA for each sub-library was transformed into electrocompetent *E. coli* JE5505 **(Method 2.6.8).** Titrations of *E. coli* JE5505 sub-libraries estimated library sizes for both libraries were; 1×10^7 for VNN_{15}+$(TTT)_1$ and 5×10^6 for NNK. *E. coli* JE5505 sub-libraries. *E. coli* JE5505 sub-libraries generated by *SpeI* digestion were then applicable for downstream utilisation in the modified Muller-Hinton overlay assay screening approach, wherein library variant peptides would be produced as recombinant peptides without the phage pIII (due to *gIII* removal).

2.10.7 Schematic overview of the library and sub-library restriction enzyme pathway

Figure 2.7 summarises where and why *T4* Ligase and restriction endonucleases, including *DpnI* (digests methylated DNA to remove parental DNA), *SpeI* and *SapI/BspQI*, were utilised in the peptide and library cloning process.

Figure 2.7 Schematic overview of ligases and restriction endonucleases utilised for peptide and library cloning strategies.
(A) Schematic depiction of the *SapI/BspQI* digestion of pSD3 $VNN_{15}(TTT)_1$ PCR amplicon. *SapI* and *BspQI* are isoschizomers both recognising 5' GCTCTTC(N)1/N4▼ 3' (where N represents any nucleotide). For both restriction endonucleases there is a prerequisite for an additional small region called the "cutting anchoring site" adjacent to the recognition sequence for efficient cleavage, herein the cutting anchoring site was 5' TACC 3'. *SapI* and *BspQI* produces a 3 base long overhang 5' GCC 3' by cleaving DNA in a staggered manner creating single-stranded ends with protruding nucleotides "sticky ends" useful for improved ligation necessary for peptide library production.

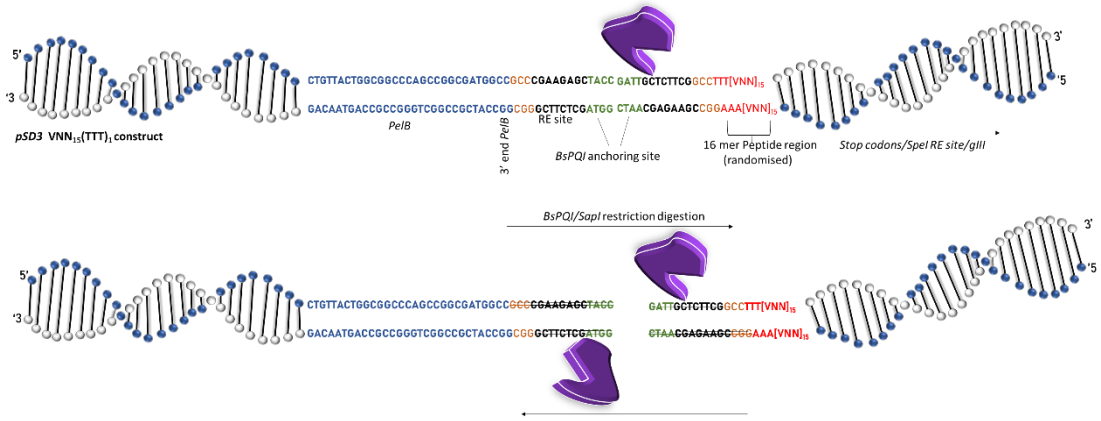


Figure 2.7 (B) *T4* ligase catalyses the formation of phosphodiester bonds between DNA fragments and requires ATP activation to carry out sticky-cohesive or blunt ends for ligation. 3bp overhangs promotes *T4* ligase efficiency by promoting alignment, reducing hindrance, and increasing stability during DNA fragment ligation.

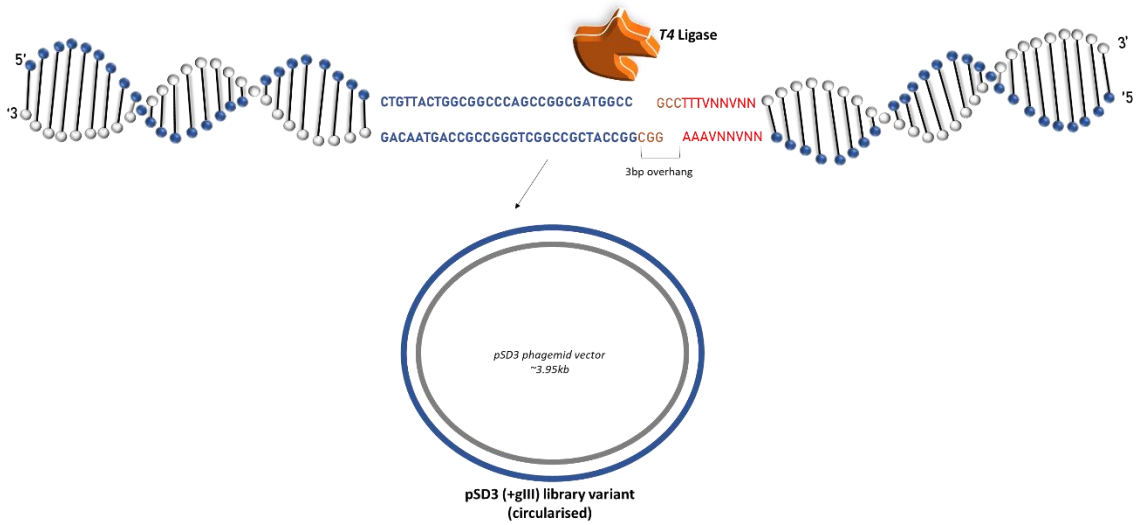
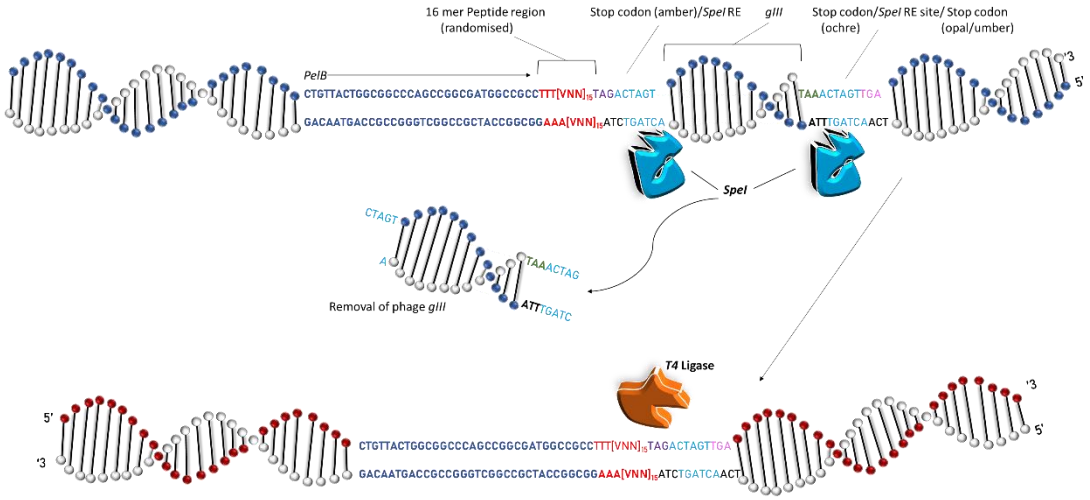


Figure 2.7 (C) Schematic representation of the use of *SpeI*. For the generation of the peptide sub-libraries for *E. coli* JE5505 overlay assay screening, it was a preference to have recombinant library peptides expressed without pIII (*-gIII*). Therefore, dual *SpeI* restriction cloning site 5' A▼CTAGT 3'sites located up and downstream of *gIII* were used for gIII removal. *T4* Ligase joins the *SpeI* digested DNA, which now has an amber stop – *SpeI* – opal stop following the $VNN_{15}(TTT)_1$ randomised peptide region.



2.11 WHOLE-CELL PHAGE DISPLAY METHOD FOR SCREENING PEPTIDE PHAGE DISPLAY LIBRARIES AGAINST BACTERIAL TARGETS

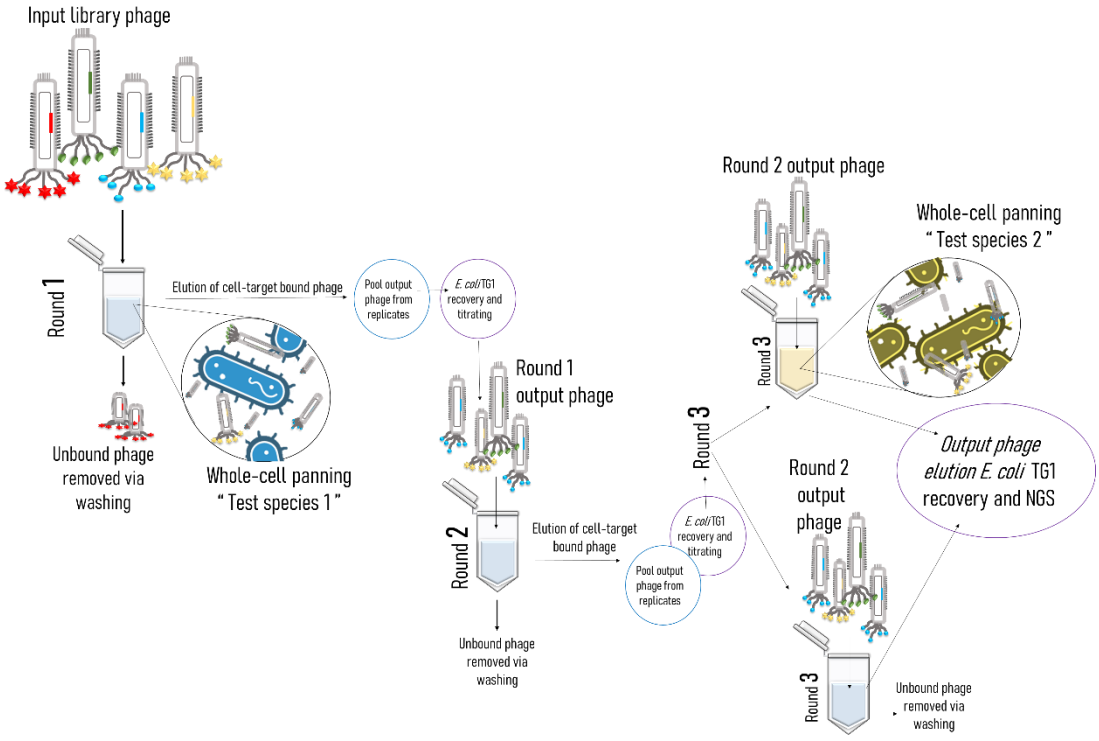
2.11.1 Phage display panning strategy:

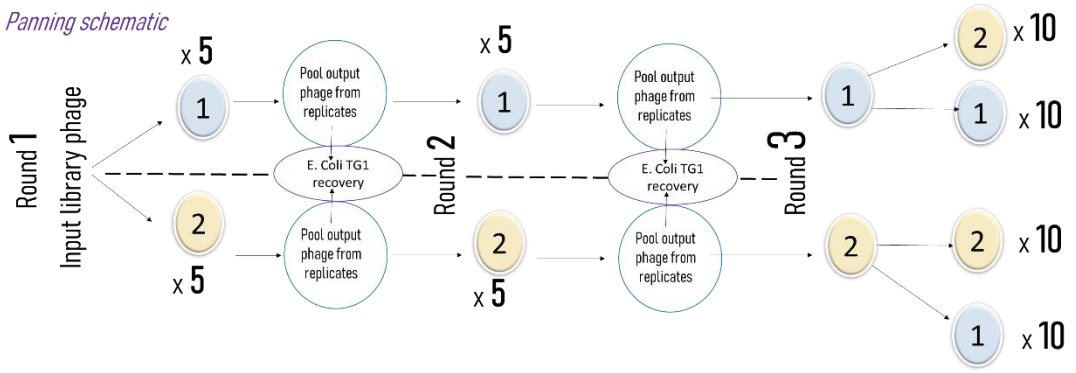
For round one of panning; input titre of 10^{11} to 10^{12} PFU/mL bacteriophage for both libraries were panned in individual replications against whole bacteria cells in phosphate buffer saline with 0.1% Gelatin (0.1% PBS-G) suspended; *S. Typhimurium* 4/74 and *S. suis* P1/7 at ≤ 2 x 10^8 CFU/mL (**Figure 2.8**) Panning round 1 and 2 both included five replicate 1 mL panning samples, wherein eluted target cell-bound phage was collectively pooled, titrated and recovered in *E. coli* TG1 to generate input phage for successive panning rounds (**Figure 2.8**). For round 1 and 2, output phage observed

ranged from 1 x 10^6 to 5 x 10^7 PFU/mL for both $VNN_{15}(TTT)_1$ and NNK libraries against *S. Typhimurium* 4/74 and *S. suis* P1/7 at 10^8 CFU/mL. Each library therefore contained output screening from 3*x* rounds of iterative panning against each of the two test species, and two sample sets wherein the round two and three test species differ in preference for the opposite Gram-model test species (**Figure 2.8**). Round three panning sample arrangement was structured in accordance with the desire to explore whether species-specific or broad-spectrum binders could be *in silico* identified by alternating the test species.

Figure 2.8 Overview of the whole-cell panning strategy utilised to screen $VNN_{15}(TTT)_1$ and NNK degenerately randomised phage peptide libraries against *S. Typhimurium* 4/74 and *S. suis*P1/7.

For both the $VNN_{15}(TTT)_1$ and NNK libraries input phage were screened against the two test species; Gram-positive model (*S. suis* P1/7) *e.g.*, test species 1, and Gram-negative model (*S. Typhimurium* 4/74) *e.g.*, test species 2. A whole-cell phage display approach in 0.1% PBS-G was selected. All rounds of biopanning aimed to screen 10^{11} to 10^{12} PFU/mL of library phage against empirically enumerated ≤ 2 x 10^8 CFU/mL of either test species. The eluted phage from five replicates for R1/R2 in (0.2M glycine-HCl, pH 2.2 neutralised with 1M Tris-HCl, pH 9) were collectively pooled. *E. coli* TG1 recovery of phage outputs was implemented. The final round of panning herein generated 8 sample sets with 10 replicates each.





2.11.2 Phage production

A singular colony of *E. coli* TG1 (Lucigen, 60502-1) grown on M9 minimal agar plates was transferred into 10 mL 2xYT broth and grown for <18 hrs 37°C, shaking at 220 rpm. Post-overnight incubation, the 10 mL *E. coli* TG1 broth was combined with 500 mL of freshly prepared and pre-warmed 2xYT broth. To support aeration, large 1L flasks were utilised and cultures were left shaking at 220rpm (37^{o}C aerobic) until mid-log phase was achieved (OD $_{600nm}$ 0.4 – 0. 6). PEG precipitated helper ex-phage stocks (~10^{12} PFU/mL) were introduced to mid-log phase *E. coli* TG1 and left to incubate at 37^{o}C for 45 mins (static). Ex-phage and *E. coli* TG1 cultures were then spun at 5000 rpm for 20 mins and bacterial pellets were resuspended in 1L 2xYT (pre-warmed 37^{o}C) containing a working concentration of 150 µg/mL kanamycin. Cultures were incubated at 220 rpm at 30^{o}C for 16hrs. Following incubation, samples were centrifuged at 5000 rpm for 20 minutes. To precipitate the phage, 2~50 mL of PEG buffer was combined with supernatant and placed on ice for <1hr. Phage were then pelleted by centrifuging at 8000 g for 20 mins and resuspended in ~10mL of 1 x PBS. PEG precipitated phage were titrated by generating serial dilution and transferring 20 µl from each phage dilution into 180 µl *E. coli* TG1 (mid-log, OD600nm 0.4 – 0.6)

in 2xYT. 500 µl glycerol stocks instead of singular *E. coli* TG1 colonies were recovered (500mL 2xYT) to mid-log phase <0.5 before inoculation of helper phage.

2.11.3 Preparation of key reagents for panning

Nutrient broth media (**2.1.1**) was prepared for cultivating panning test species whilst 2xYT media and minimal M9 agar plates (**2.1.12 and 2.1.13**) were exploited for *E. coli* TG1. Where necessary plates were supplemented with ampicillin at a final working concentration of 150µg/mL. PBS with 0.1% Gelatin "PBS-G" (Fisher); ~50mL per test species was prepared for use as the solution to suspend whole-bacteria cells. Tris-HCl and Glycine-HCL were prepared to pH 9 and pH 2.2 respectively **(2.3.7 to 2.3.8).**

2.11.4 Whole-cell phage display panning

~ Two colonies of the target test bacteria isolated from nutrient agar plates were transferred to 10mL of nutrient broth, and grown at 37^{o}C, shaking at 200rpm. Mid-log nutrient broths containing the test species were spun down (5000 g, 10 mins, 4^{o}C), broth removed, and bacterial pellets resuspended in 10 mL 1x PBS w/ 0.1% gelatin (PBS-G, chilled to 4^{o}C). The mid-log phase PBS-G resuspended test species was then diluted to ~ 1 x 10^{7} CFU/mL. Test species mid-log phase was empirically determined, as well as the necessary optical density (OD_{600nm}) equating to ~ 1 x 10^{7} CFU/mL however, for each panning round test species input CFU/mL was verified. 100 µl of the ~ 1 x 10^{7} CFU/mL test species PBS-G suspension was aliquoted into five individual replicates of polypropylene eppendorfs and challenged with peptide phage libraries 10^{11} and 10^{12} PFU/mL. Panning samples were then rotated gently for 30 mins at RTP, whereupon samples were spun at 5000 g for 5 mins to pellet test species bacteria. Seven washes with 1x PBS-G were conducted, with spin steps equating to 5000 *xg* for 4 mins. Bound phages were eluted by the addition of ~200 µl of 0.2M

Glycine-HCl (pH 2.2) and slow rotation at RTP for 10 mins. Panning samples were then spun at 12, 000 *xg* for 2 min. Phage containing supernatant from the five replicates were pooled and neutralised with the addition of 145 µl 1M Tris-HCl, pH 9 (on ice). 2-3 colonies of *E. coli* TG1 isolated from M9 minimal plates and utilised to inoculate 15mL 2xYT. Output phage serial dilutions were titrated with mid-log phase *E. coli* TG1 recovery on 1 % glucose ampicillin 2xYT plates. ~500 µl of the neutralised eluted phages were placed into 4.5mL of *E. coli* TG1 cells at mid-log phase and incubated at 37^{o}C for ~30 mins (static). Output -phage *E. coli* mixture was spun down and reconstituted into 1mL of 2xYT. Similar to titration plates, this 1mL reconstituted sample was spread onto 1% glucose, ampicillin 2xYT plates. Following incubation (37^{o}C, aerobic and static), large bioassay dishes were scrapped to recover *E. coli* TG1 and create 30% glycerol stocks for each panning output sample. Phage propagation as explored in **Method 2.11.2** was utilised to generate panning sub-library phage for subsequent rounds of panning. Should be noted, round 3 panning involved increasing the sample replicates to ten, and samples were not pooled but individually *E. coli* TG1 recovered to maintain sample integrity.

2.11.5 M13 DNA extraction

E.Z.N.A.® M13 DNA Mini Kit (Omega Bio-tek) was utilised for M13-phage DNA extraction, specifically from round three TG1 panning outputs pooled for analysis in the leaky *E. coli* overlay assay. E.Z.N.A.® M13 DNA Mini Kit manufacturer's instructions were applied with the inclusion of using 2xYT broth diluent to bring samples to desired starting volume **(Method 2.3.10)**. The DNA from round three panning experiments were subjected to sub-library preparations *i.e., SpeI* digestion **(Method 2.6.8),** ligated **(Method 2.6.7),** and then transformed into *E. coli* JE5505 to screen round three phage output peptides in the overlay assay.

2.12 NEXT GENERATION ION TORRENT BARCONDING AND IN SILCO ANALYSIS TOOLS

2.12.1 Ion torrent NGS primers

$VNN_{15}(TTT)_1$ or NNK library *E. coli* TG1 glycerol stocks or final panning round *E. coli* TG1 recovery stocks were miniprepped **(Method 2.6.11)** to isolate phagemid dsDNA to be prepared for next generation sequencing. The two-round of PCR strategy involves standardly orientated forward and reverse primers **(Table 2.13)**. PCR amplification strategy aimed to introduce linker and barcoding sequences necessary for Ion torrent Next generation sequencing analysis, see **Figure 2.9**. NGS round one amplicons contain linker 1 (via reverse primer) and linker 2 (via forward primer) sequences which serve as annealing regions for Round two PCR primers. The second round introduces A-key barcodes and P1prim linker via Akey-BCXX-lnk1 reverse primers (anneals to linker 1 sequence) and P1prim-linker2 forward primer (anneals to linker 2 sequence) respectively. P1prim is the forward sequencing primer in NGS, initiating the sequencing reaction by binding to DNA sequencing adapters. A-key primer allows for the identification and demultiplexing of the sequencing data as DNA fragments can be distinguished via unique barcodes.

Figure 2.9 Next Generation Sequencing coverage and PCR schematic for pSD3_16mer vectors

(NGS PCR Round 1) The NGS left motif (located between *M13 reverse* and *T3 promoter*) and right motif (downstream peptide sequence, dependent on the pSD3 inclusion of *gIII*) provide the guide for the annealing locations of the Round 1 forward "**F**" and reverse "**R**" primers respectively shown as blue arrows.

(NGS PCR Round 2) The linkers from the previous round of PCR amplification serves as annealing targets for Round 2 PCR primers containing either A-key barcodes (forward primer) P1prim-linker (reverse primer). The final barcoded NGS Round 2 PCR product is a pSD3 fragment with the coverage localised to capture the peptide encoding region.

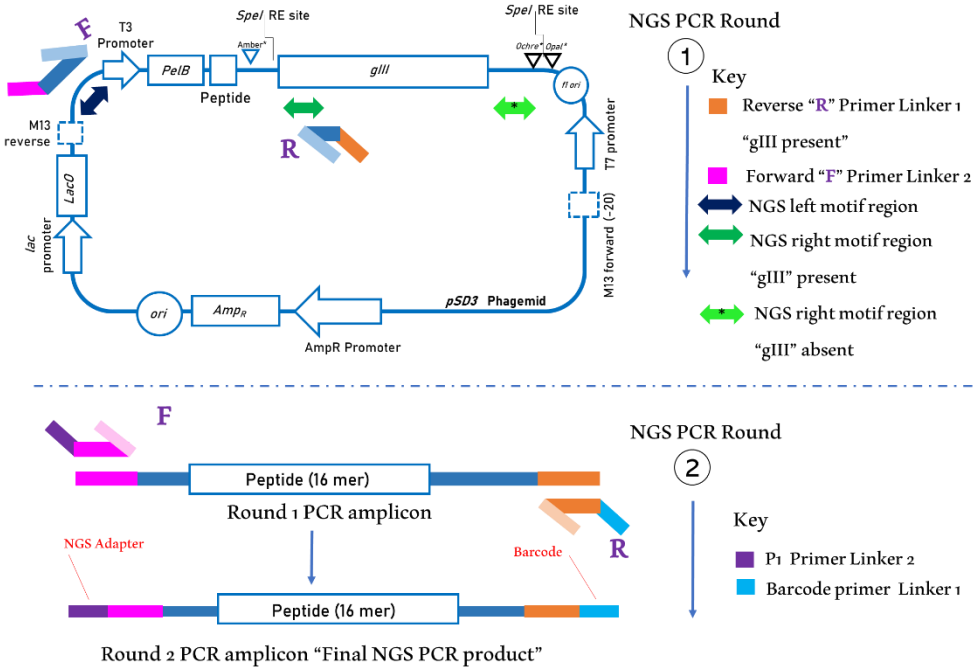


Table 2.13 PCR Round 1 "Linker" NGS primers for pSD3 _16mer library (sequences reverse strand)

Primer	**Oligonucleotide primer sequence** [1]	**(bp)**	**Features**
pSD3_16_R_lnk1 Reverse Primer	**5'**GTAATCCTTGTGGTATCGCGAA TGGAGAGCCACCGC **3'**	36	T_m = 69 °C GC%= 56%
pSD3_(gIII deleted)_16_R_lnk1 Reverse Primer	**5'**GTAATCCTTGTGGTATCGAGTA CGGTCAAGAAAACC **3'**	36	T_m = 64 °C GC%= 44%
pSD3_16_F_lnk2 Forward Primer	**5'**CTAGAACATTTCACTTACATGA TTACGCCAAGCTCG **3'**	36	T_m = 63 °C GC%= 42 %
[1]All primers were manufactured by Sigma-Aldrich XXX or XXX= Linker sequence			

Table 2.14 PCR Round 2 "Barcoding & Adapter Linker " NGS primers for pSD3 _16mer library

Primer	Oligonucleotide primer sequence [1] 5' → 3'	(bp)	Features
P1prim-linker2 "Forward"	CCTCTCTATGGGCAGTCGGTGAT**CTAGAA CATTTCACTTAC**	46	T_m = 68 °C GC%= 46%
"Reverse 1" Akey-BC1-lnk1[2]	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTAAGGTAACCGTAATCCTTGTGGTATCG		
"Reverse 2" Akey-BC96-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTAAGCGGTCCGTAATCCTTGTGGTATCG		

[1]Manufactured by Sigma-Aldrich.
P1prim-linker2
XXX = refers to the Round 1 Forward primer introduced linker region (Table 2.12) annealed to by P1prim-linker2 primer.
XXX = refers to P1prim-linker2 adapter region.

Akey-BCXX-link1
"XXX" relates to the A-key sequence, whereas "XXX" is the unique barcode and "XXX" refers to the Round 1 Reverse primer introduced linker region (Table 2.12) annealed to by **Akey-BCXX-lnk1.** The barcoding primers were at least 58bp in length. [2]See the full list of the barcode primers **in Appendix Section 9.2.**

2.12.2 NGS Sample Preparation: PCR & Clean-ups

50 µl total volume PCR reaction mixtures were generated by combing; 1 µl (10ng) template and master mix consisting of; 25.5 µl Nuclease-free water, 10 µl Q5 buffer (x1), 10 µl GC enhancer (x1), 1 µl dNTPs (200 µM), 1 µl forward primer "pSD3_16_F_lnk2" (10 µM), 1 µl reverse primer "…._16_R_lnk1" (10 µM) and 0.5 µl of Hot-Start Q5 enzyme (units: 0.02 U/µl, High-fidelity DNA polymerase, M0493). First round NGS PCR conditions were: 95°C for 3 mins, succeeded by 30 cycles of 95°C for 30s, 60°C for 30s, 65°C (pSD3_-*gIII*) or 68°C (pSD3_*gIII*) for 30s, and 72°C for 5 mins and 30s. 3% (w/v) agarose/TAE gel electrophoresis with <1kb DNA ladders were utilised to visualise amplifications, preferred where possible were gels with singular desired size bands **(Method 2.6.2/2.6.3).** The first round amplifies the peptide encoding region of the pSD3 construct generating 287bp (pSD3_*gIII*) or 215bp (pSD3_-*gIII*) PCR amplicons containing desired linkers. DNA was purified using the Nucleospin kit (Macherey Nagel) according to the manufacturer's

recommendations for high weight volume agarose gel clean-ups **(Methods 2.6.4).** The final concentration of the DNA in clean-ups were quantified utilising Qubit dsDNA High Sensitivity & Broad-range Assay Kit (ThermoFisher Scientific) / Qubit 2.0 Fluorometer 2.0 (Invitrogen) **(Method 2.6.5).**

Aforementioned PCR reaction mixtures were replicated for round two NGS PCR, using 10ng of the quantified round one clean-up as the template. Round two NGS PCR utilised the universal forward primer P1prim-linker2 and reverse barcoded primers **(Appendix 9.2)** under the following thermocycling conditions 95°C for 3 mins, succeeded by 15 cycles of 95°C for 30s, 62°C for 30s, 67°C (pSD3_-*gIII*) or 69°C (pSD3_*gIII*) for 30s, and 72°C for 5 mins and 30s. NGS Round two PCR amplification final product applicable for sequencing and confirmed by agarose gel electrophoresis; wherein 347bp (pSD3_*gIII*) or 275bp (pSD3_-*gIII*) PCR amplicons containing barcodes and P1prim linker were desired. Following agarose gel clean-ups, pools were generated by ensuring concentrations of each NGS sampleset DNA were quantified via Qubit and equal represented in the final pool.

Agencourt AMPure XP Bead Clean-up kit (Beckman coulter) was utilised to ultra-clean pooled samples, all steps conducted were in accordance with the manufacturer's instructions; 1.8 μL of AMPure XP per 1.0 μL of sample (DNA fragment and paramagnetic bead binding). Two wash steps with 70% Ethanol are performed to eliminate impurities and contaminants. Purified DNA fragments are eluted from the beads using nuclease-free elution buffer provided. 200ng pool containing AMPure eluted barcoded DNA were then sent to the University of Pennsylvania for sequencing via the Ion Proton platform with an SS 540 Chip.

2.12.3 Extracting datafiles, identifying barcoded sequences and analysing the amino acid composition of 16mer peptides.

The following instructions were run in Linux, XXX.pl denotes the perl script or commands, XXX relates to script input files whereas XXX relate to the name of outfiles of importance.

1. seqtk seq -a **NGSFILENAME**.fastq > **NGSFILENAME**.fastq.gz
2. perl PIPELINE1.06.pl --infile NGSFILENAME_.fastq.gz –outfile [filename:LIB_outfile1.06pl] --barcodes BCXX.txt --left PAMAA.txt --right TSSSGG.txt --minimum 1 --fulloutput –revcomp
3. perl aa_distribution_all_peps.pl --input.fasta outfile1.06pl_BCXX_LR.fasta --length 16.

The first step pertained to decompressing the raw NGS Ion Torrent data retrieved from University of Pennsylvania via seqtk seq -a NGSFILENAME.fastq > **NGSFILENAME.fastq.gz.** Pipeline 1.06.pl was used to demultiplex raw NGS data, and the script processes decompressed single line FASTAQ OR.gz compressed FASTAQ files "**i.e. --infile NGSFILENAME_.fastq.gz".** Certain NGS samples were pools of up to 80 barcoded samplesets. Pipeline 1.06 script extracts one barcode dataset per run, therefore text files containing each barcode sequences " --barcodes **BC01.txt** to **BC95.txt**" were referenced in the script where appropriate for each sample. The left (N-terminal) and right(C-terminal) motifs flank the peptide region in pSD3 constructs **(Figure 2.1)**, and in addition to the peptide encoding region the motifs are retained in the final product of the two-round NGS preparation (**Figure 2.9).**

For the pSD3 constructs with *gIII* present the left and right motifs of NGS prepared amplicons were amino acid sequences **PAMAA** and **TSSSGG** respectively, which were saved as txt files **---left PAMAA.txt** and **--right TSSSGG.txt** for Pipeline 1.06 script to reference when running. Whereas for pSD3 constructs with *gIII* absent NGS

samples (e.g., *E. coli* JE5505 $VNN_{15}(TTT)_1$ and NNK sub-libraries) the **--right motif** file was **"SCQFFWL".txt.** Ion Torrent sequencing reads one strand of DNA molecules in the reverse direction; therefore it is essential to convert the raw NGS data to the opposite strand by reverse complementation "--rev comp."

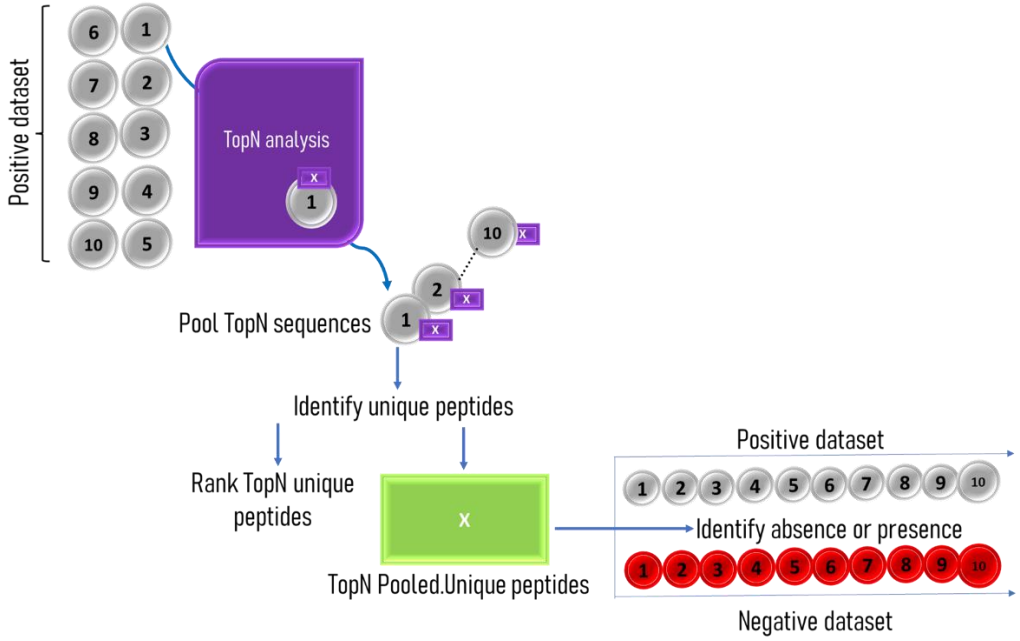
PIPELINE1.06.pl produces several outfiles including the reverse complementation read out (outfile1.06pl_BCXXrevcomp.fasta), files with barcoded sequence translation in each of the three reading frames (*i.e.*, frame 1: translation initiates from the first nucleotide.); "outfile1.06pl_BCXXframe1.fasta, outfile1.06pl_BCXXframe2.fasta and outfile1.06pl_BCXXframe1.fasta" and "outfile1.06pl_BCXXframe1.fasta," and finally the outfile1.06pl_BCXX_LR.fasta, contains the frame shifted sequences where the left and right motif sequences were correctly identified. Perl aa_distribution_all_peps.pl script was implemented to analyse the amino acid distribution of the peptide encoding regions, wherein the peptide was specifically 16 mers long (--length 16)

2.12.4 Summarised overview of panning sampleset analysis in TopN and Z score pipelines.

Figure 2.10 and Figure 2.11 succinctly summarise the two parallel forms of *in silico* analysis applied to round 3 panning output NGS data; namely Top*N* and Z score analysis pipelines. Compared to Z scores, the Top*N* pipeline represents a simpler implementation of frequency analysis. The Top*N* pipeline entailed pooling the most frequent sequences across a selected positive dataset ($p_{dataset}$), and returning the absence or presence of unique ranked pooled sequences within each replicate of the $p_{dataset}$ and negative dataset ($n_{dataset}$). Implementing the Top*N* pipeline coupled two groups of $p/n_{datasets}$.

Figure 2.10 ***in silico peptide*** **phage display analysis pipeline; TopN strategy**

The Top*N* strategy, where N = 50, 100, 200, 500 or 1000 ranks peptides by relative frequency and representation across the 10 replicates of the positive and negative dataset . The use of Top*N* becomes more stringent by selecting a cut-off for the minimum number of times a peptide occurs *i.e.* X/10 panning replicates. Herein X/10 threshold for the positive dataset ($p_{dataset}$) is ≥4/10; equates to ≥40% of samples in the positive dataset containing said peptide at TopN. Applying $p_{dataset}$≥40% to both the positive and negative datasets aided the identification of binders with reproducibility in panning samples against two different test species. Alternatively, similar to the Z score approach this can be altered to $p_{dataset} \geq 40\%$ and $n_{dataset} = \leq 20\%$) for a more species-specific approach for TopN use.

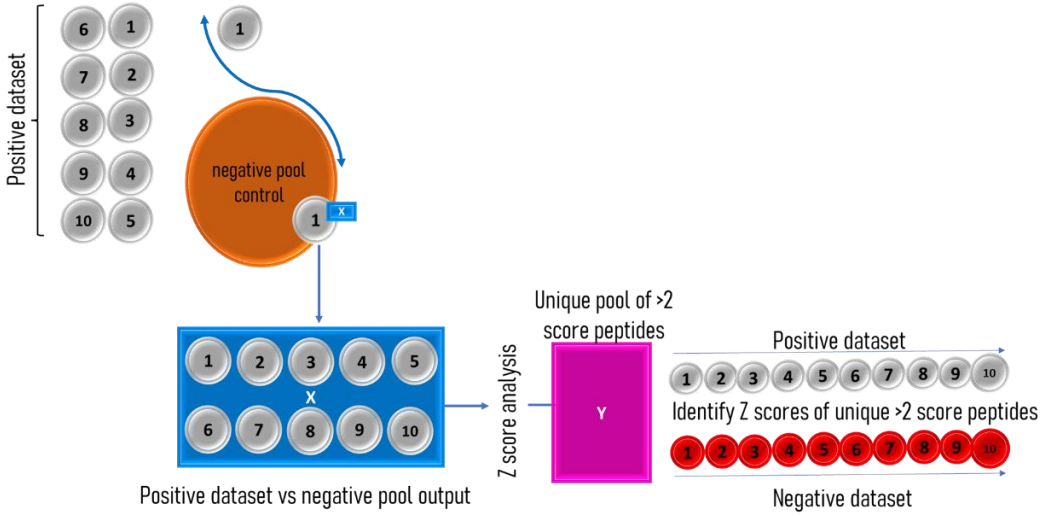


In comparison to the Top*50* approach, the Z scores pipeline begins by considering a designated pool of negative panning replicates (negative dataset) (**Figure 2.11**). The coupling of panning sample sets at this stage was performed to conduct "*in silico* subtraction" for the identification of species-specific enriched binders. A Z score cut-off value of >2 has been broadly utilised to analyse peptide libraries for enriched peptide ligands, share parity with other recent peptide phage display applications (Rajan et al., 2021). Z score ≥ 2 cut-off was applied to the analysis herein except for the dataset VNN_SS2_SS3 where Z score cut off was increased to ≥ 2.5 due to the number of peptide candidates (Z score ~2.0) present in this dataset. Across the eight panning datasets 721 peptides were identified with a Z

score higher than the 2 or 2.5 cut-off parameters combined with presentation in 4/10 positive dataset panning replicates, i.e., ≥40% of $p_{datasets}$.

Figure 2.11 *in silico* peptide phage display analysis pipeline; Z score strategy

Positive dataset vs negative pool output is a crucial step which reflects a form of "in silico subtraction". The Z score of retained sequences in the negative pool comparison output was determined. Z score: ≥2 cut-off was applied for selection. The unique pool of Z score >2 peptides were then mined for Z scores and frequency across replicates and the alternative negative dataset to the pool which shares the same round 2 panning test species but differ in round 3.



$$Z\ Score = \frac{(\rho 1 - \rho 2) - 0}{\sqrt{\mathrm{P}(1 - \mathrm{P})\left(\frac{1}{n1} + \frac{1}{n2}\right)}}$$

2.12.5 Z score analysis pipeline

Maximising next generation sequencing of antibody phage display libraries relies on manipulating the phenotype-information-cycle of target vs phage and *E. coli* gene recovery transformation (Zhang *et al.*, 2011). The same exploitation of phenotype-information can be applied to peptide phage libraries. Panning enriches multiple phenotypes, which arise both as noise and desired signals within the phenotypic pool. Therefore, interests mainly concern discerning the frequency distribution of each individual panning peptide within the dataset relative to a designated control. Z scores are values thereby created from implementing a two-

proportion Z test to identify the frequency ratio of proportions across two independent but experimentally related test sets. The application of Z scores in this manner exploits the assumption in large phage display libraries (> 10^9 size) a given sequence should not be seen multiple times unless it has been selectively enriched.

Z Score Script Pipeline: XXX.pl denotes the perl script or command, XXX relates to script input files whereas XXX relate to the name of outfiles of importance. Script instructions below, wherein BC01-10 equates to the positive barcoded panning dataset and BC11-20 is the negative.

1. perl Nmerfastafilter.pl --length 16 --infile BC01.LR.fasta --outfile BC01.LR.fasta_16mers
2. cat BC01.LR.fasta_16mers BC02.LR.fasta_16mers BC03.LR.fasta_16mers…….> BC01-10_positive_pool.LR.fasta.
3. cat BC11.LR.fasta_16mers BC12.LR.fasta_16mers BC13.LR.fasta_16mers……..> BC11-20_negative_pool.LR.fasta.
4. perl compare2.1.pl BC01.LR.fasta_16mers BC11-20_negative_pool.LR.fasta > 01vsnegative_pool.table.txt.
5. perl topZscores.pl --seqsonly --zscore 2 --infile 01vsnegative_pool.table.txt --outfile 01vsnegative_pool_Z2_sequences.
6. Cat 01vsnegative_pool_Z2_sequences 02vsnegative_pool_Z2_sequences 03vsnegative_pool_Z2_sequences…….> all_10_positve_Z2_sequences.
7. perl uniq.pl all_10_positve_Z2_sequences > all_10_positve_Z2_sequences.unique.
8. ls -1 {01..10}vsnegative_pool_Z2_sequences > 01-10_top_Z2_file_list.
9. perl seenhownmany1.1.pl --posmin 0 --peplist all_10_positve_Z2_sequences.unique --posfiles 01-10_top_Z2_file_list > topZ2_peptides_seen_across_all_Z2_positives.txt.
10. **ls -1** {01..10}vsnegative_pool.table.txt > Zscore_table_file_list
11. perl Zcountpep1.0.pl --position 4 --peplist topZ5_peptides_seen_across_all_Z5_positives.sequences.txt --posfiles Zscore_table_file_list > topZ5_Zscore_table.tsv
12. **ls -1** BC{01..20}.LR.fasta_16mers > positive_and_negative_LR.fasta_file_list
13. **perl countpep1.3.pl** topZ5_peptides_seen_across_all_Z5_positives.sequences.txt positive_and_negative_LR.fasta_file_list

NGS raw data is decompressed, barcoded sequences for each panning sample are demultiplexed and frame-shifted with reference to left/right motifs (**Method 2.12.3).** perl Nmerfastafilter.pl generates "**outfile1.06pl_BCXX_LR.fasta_16mers**" filtered outfiles with barcoded samples frame shift translated with the correct left/right motifs flanking a 16mer peptide region only. The cat command was utilised where applicable

to then combine Outfile1.06pl_BCXX_LR16.fasta to create negative or positive poolsets i.e., "**BCXX-XX_negative_pool.LR.fasta.**" Important to emphasise herein negatives and positives were aligned from the panning strategy to identify species specific sequences. For example, if the 10 replicates derived from panning the NNK library against *S. suis* P1/7 for three rounds "NKK_SS2_SS3" was deemed the positive dataset, this would necessitate the Z score pipeline utilising the NNK library screened for three rounds against *S. typhimurium* 4/74 "NKK_TY2_TY3" as the negative pool.

Compare2.1.pl script compares two FASTA files (singular positive sample **outfile1.06pl_BCXX_LR.fasta vs** negative pool **BCXX-XX_negative_pool.LR.fasta)**, to calculate sequence counts, normalised counts, Z-scores, and percentages which are captured in a sorted tabulated outfile **"BCXXvsnegative_pool.table.txt"** with sequence identifiers and their associated statistics. perl topZscores script reads tabulated files, filters lines based on a specified Z-score cutoff (i.e. Z score ≥ 2), generating an outfile with sequences and respective Z-scores **"BCXXvsnegative_pool_Z2_sequences".** The cat | sort | uniq -c | sort -nr > command rank sorts BCXX to BCXXvsnegative_pool_Z2_sequences outfiles generating a file with the ≥ 2 Z score sequences across each panning barcoded sample **"all_10_positve_Z2_sequences".** perl uniq.pl was the utilised to identify unique sequences, denote their relative counts, frequency and Z scores **"all_10_positve_Z2_sequences.unique". ls -1** command was utilised to combine files into top Z-score files lists, BCXX…BCXXvsnegative_pool_Z2_sequences > **"barcoderange_top_Z2_file_list".**

Seenhownmany1.1.pl script was exploited to determine how many times each unique ≥ 2 Z score sequence appears in the top Z-score file list of the positive samples. The seenhowmay1.1 script is repeated with the optional command --seqsonly thereby

generating a list of sequence readouts. **ls -1** is used to combine files from both the positive and negative datasets BC{XX..XX}vsnegative_pool.table.txt > "**Zscore_table_file_list".** Perl script **Zcountpep1.0.pl** analyses the presence of top Z sequences in the Zscore_table_file_list collating information on file names, Z scores and percentages which can be analysed in Excel to highlight cells based on specific criteria. **Countpep1.3.pl** identified topZ score sequences across the outfile1.06pl_BCXX_LR.fasta_16mers outfiles in the positive and negative datasets.

2.12.6 TopN analysis pipeline

Compared to Z_scores, the TopN analysis pipeline centres on a simple frequency analysis. **TopN.pl.** reads input files containing peptide sequences, identifies counts of each unique peptide using a hash data structures then sorts the peptides in descending order based on frequency counts. N is specified according to the --top option, when N = 50, this generates a Top50 list of 50 top-ranked peptides with the highest frequency count in the input file. Herein Top50, Top100, Top200 and Top1000 peptides were analysed.

TopN pipeline: XXX.pl denotes the perl script or command, XXX relates to script input files whereas XXX relate to the name of outfiles of importance. Script instructions below, N = 50 (Top50), wherein BC01-10 equates to the positive barcoded panning dataset and BC11-20 is the negative.

1. **perl topN.pl** --top 50 --infile **BC01.LR.fasta_16mers** --outfile **BC01_top50.sequences.** The **--top** value and the output file adjusted accordingly. Repeated for negative and positive files.
2. **cat** BC01_top50.sequences BC02_top50.sequences ... > **positive_top50_sequences.pooled.**
3. Option A: **cat** positive_top50_sequences.pooled | sort | uniq -c | sort -nr > positive_top50_sequences.pooled.ranked.

   Option B: perl uniq.pl positive_top50_sequences.pooled > positive_top50_sequences.pooled.unique.

4. **ls -1 BC{01..10}_top50.sequences > positive_top50_sequences_file_list**.
5. perl seenhownmany1.1.pl --posmin 0 --peplist positive_top50_sequences.pooled.unique --posfiles positive_top50_sequences_file_list > positive_top50_sequences.pooled.ranked.
6. Negative filelist generated **ls -1 BC{11..20}_top50.sequences > negative_top50_sequences_file_list**.
7. perl seenhownmany1.1.pl --posmin 4 --negmax 2 --peplist positive_top50_sequences.pooled.unique --posfiles positive_top50_sequences_file_list --negfiles negative_top50_sequences_file_list > positive_top50_sequences.seenmin4pos.max2neg.txt. –seqsonly can be additionally inserted into the script instructions for sequence read outs.

TopN.pl. was applied to both positive and negative datasets**. "TopN_sequences.pooled"** were created by concatenating **(cat command)** "BC01_topN.sequences, BC02_topN.sequences….." The pooled file can then be piped to sort | uniq -c | sort -nr to obtain a ranked list. However, opted for was making a unique list of peptides **"positive_topN_sequences.pooled.unique"** from the pooled file using the Perl script **uniq.pl** . A list of positive and negative files containing the top N sequences was generated by **ls-** command (ls -1 BC{01..10}_top50.sequences). **Seenhownmany1.1.pl** was utilised to analyse the occurrences of each non-duplicated unique peptide in the positive and negative files via filters based on minimum occurrences in the positive files (--posmin) and maximum occurrences in the negative files (--negmax). --seqsonly option was included in the previous seenhownmany1.1.pl command to extract sequences. The output ranked peptide sequences based on frequency were further analysed via Excel or directly used in AMP predictor software (**Table 2.10**).

Chapter 3: Whole-construct inverse PCR method for the generation of naïve 16mer peptide libraries randomised using degenerate primer $VNN_{15}+(TTT)_1$ "AMP-biased" and "conventional" NNK approaches

3.1 Introduction

Anti-bacterial AMPs vary in length, structural conformities, bacteria membrane targets and spectrum of activity (Wang et al., 2016). Nonetheless, this class of peptides often share certain properties such as; amphipathicity, and the tendency to be highly abundant in cationic and/or hydrophobic amino acid residues (Pane et al., 2017). To accelerate AMP-drug design and discovery; peptide collection databases (ADP3, CAMPR3, DBAASP v3), numerous computational methods and bioinformatic prediction tools exist to further rationalise AMP-candidate identification (Waghu et al., 2016; Wang et al., 2016; Bhadra et al., 2018; Pirtskhalava et al., 2021). Most AMP *in silico* tools operate via the principle of parallel properties, often analysing similarities in amino acid distribution between peptides (Wu et al., 2014). AMP activity inherently relies upon amino acid composition, which in itself is complex and further dictated by positional incorporation and contribution towards physiological properties favourable for peptide structural integrity, binding and/or antimicrobial activity (Vishnepolsky et al., 2018; Pirtskhalava et al., 2021). Consequently, to semi-rationally design novel AMPs against several bacterial targets requires deep amino acid-level analysis of known AMPs with desired bacteria target affinity.

Site saturation mutagenesis is a powerful tool for the directed evolution of peptides and degenerate oligonucleotide primers are widely utilised for amino acid residue-specific randomisation (Acevedo-Rocha et al., 2015). The NNK/NNS degenerate codon schemes, NNK/S, where "N" signifies A/T/G/C bases, "K" = G/T and "S" = C/G bases, remain popular and have been frequently exploited to generate peptide libraries (Kille et al., 2013). More broadly, application of degenerate schemes wherein redundancy is eliminated with equimolar ratios of degenerate codons or primers, *e.g.* Tang, MAX and 22c-Trick methods, at present cannot generate large

peptide libraries where variant sequences are respectively saturated by >five contiguous amino acids (Bozovičar et al., 2020). However, ceaseless methods exist to generate libraries with >5mers randomised peptides; for instance, Kretz *et al., (*2018) and Tsoumpeli *et al.,* (2022) exemplified the generation of NNK-randomised phage peptide libraries (10^8 to 10^9 library size) via contrasting molecular PCR cloning methods.

Generally, identification of novel AMPs orientates towards two main approaches. *i*) The construction of randomised-residue specific peptide libraries wherein, library variants of a known AMP are characterised and scanned for; improved antimicrobial functionality and critical position specific amino acid residues which influence antibacterial effects against bacterial targets (Tominaga and Hatakeyama, 2006; Guralp *et al.*, 2013). For instance, Tominaga and Hatakeyama, (2006) utilised NNK-scanning and peptide library variants to identify amino acids responsible for the antimicrobial nature of natural AMP pediocin PA-1. Secondly, *ii*) literature exists wherein novel AMPs are identified from screening repurposed commercially available combinatorial small peptide phage libraries (Pini et al., 2005; Bishop-hurley et al., 2005; Sainath Rao, Mohan and Atreya, 2013;Flachbartova et al., 2016). For example, Sainath Rao, Mohan and Atreya, (2013) identified RLLFRKIRRLKR (EC5), an arginine and lysine-rich phage-display derived candidate AMP with activity against *E. coli*.

Needless to say, novel AMP discovery has been exemplified from a plethora of approaches. However, the application of aforementioned molecular cloning methods to generate NNK-randomised and/or novel degenerate scheme randomised $\geq 10^9$ peptide libraries for AMP discovery which facilitates both phage display and other screening methods remains limited. Consequently, this chapter aims to exploit the

interdisciplinary progress made in *in silico* tools, advanced peptide library mutagenesis techniques and various screening strategies or display platforms to construct a novel AMP discovery pipeline with broad reaching industrial applicability.

3.2 Aims 1

3.2.1 Summary of key research aims for results chapter 1;

i. Delineate amino acid compositional and positional biases in a dataset of natural "AMPs" with activity against targets of interest; several key bacteria gastrointestinal and respiratory pathogens of pigs.

ii. Identify a novel AMP-biased mutagenesis strategy for library construction, which theoretically biases peptide library amino acid distribution towards the observed composition of an AMP dataset *(i).*

iii. Design a phagemid vector, degenerate primer and library cloning strategy for the construction of peptide libraries compatible with high-throughput peptide library screening strategies including phage display technology.

iv. Use a semi-rational directed mutagenesis strategy (*ii*) and *(iii*), to construct a randomised 16mer "AMP-biased" peptide library using a novel degenerate codon scheme approach *(ii).*

v. Construct a randomised 16mer peptide library with a directed mutagenesis strategy (*iii*) utilising the conventionally exploited NNK degenerate codon scheme.

vi. Deep sequence peptide libraries generated via (*iv*) and (*v*), evaluate library quality, diversity and amino acid composition via quality control analysis of sanger and next-generation sequencing data.

3.3 Generating and analysing a dataset of natural AMPs with activity against target bacteria pathogens

3.3.1 In silico analysis of AMP and non-AMP training datasets

Analysis of peptide length, charge and isoelectric point of AMPs and Non-AMPs in the respective three training datasets was conducted and summarised in **Figure 3.1** The APD3_ AMP_PigPathogenDataset peptides were 7 - 47 mers long (average ~26mers) with +1 to +15 overall net charges (average ~+4.35) and hydrophobic residues constituted <80% of amino acids present in sequences, with the lowest hydrophobicity being ~30%. Fundamentally AMP datasets aligned towards literature-based descriptions of natural AMPs (Tan et al., 2021). Computational tools **(Method 2.9)** were utilised to conduct amino acid composition and positional analysis of the 187 peptide sequences in APD3_AMP_PigPathogensDataset, APD3_RandomSelection187, and NonAMPep_RandomSelection187 datasets. The *in-silico* analysis strategy focused on numerating the frequency of all natural 20 amino acids and analysing the incidence and composition of hydrophobic stretches of amino acids along each peptide sequence. Amino acid composition analysis for our dataset of interest, ADP3_AMP_PigPathogensDataset, is summarised in **Figure 3.2.**

Figure 3.1 Key physiochemical properties of small datasets of AMPs and Non-AMPs peptide sequences, (A) Peptide sequence length (B) Overall net charge, (C) Isoelectric point

(A), (B): Selection of ADP3 database derived AMPs on the basis of activity against specified bacteria targets (ADP3_AMPDataset_PigPathogens) identified a set of peptide which were slightly more hydrophobic and cationic in nature in comparison to peptides randomly selected from the same database (APD3_RandomSelection187).

The ADP3_RandomSelection187 presents a profile which demonstrates the diversity of AMPs and contains peptides with varying targets e.g.; anti-fungal and anti-viral. Naturally, randomly selected AMPs are typically shorter than non-AMP equivalents (NonAMPep_RandomSelection187). Both AMP datasets exhibit high medians for isoelectric points ≥9 **(C),** overall net charge (high tendency to be cationic) and hydrophobicity profiles which are discreetly distinguishable from non-AMPs. This discreetness necessitates the exploration of amino acid distribution wherein compositional frequency is placed in context of positionally incorporation.

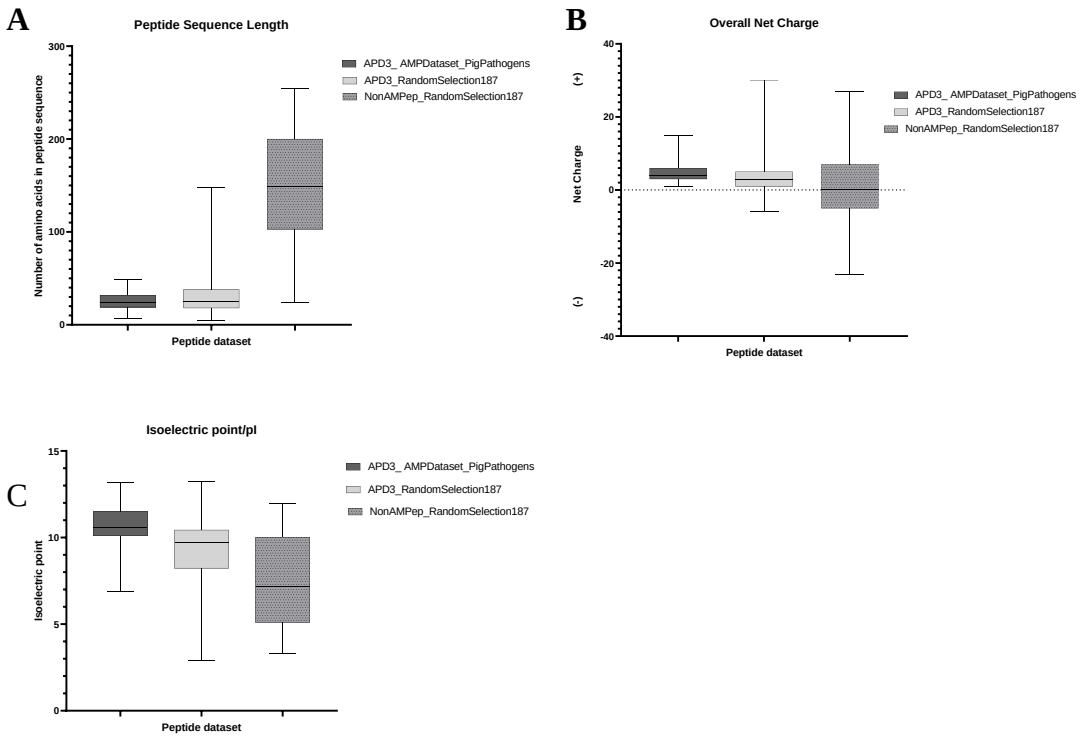


Figure 3.2. Amino acid composition of the ADP3_AMP_PigPathogensDataset. The observed percentage frequency of the twenty naturally encoded amino acids in 187 antibacterial natural AMPs. The top 10 ranked amino acids accounted for 76.8% of total amino acid frequency, and 9/10 were hydrophobic or basic residues.

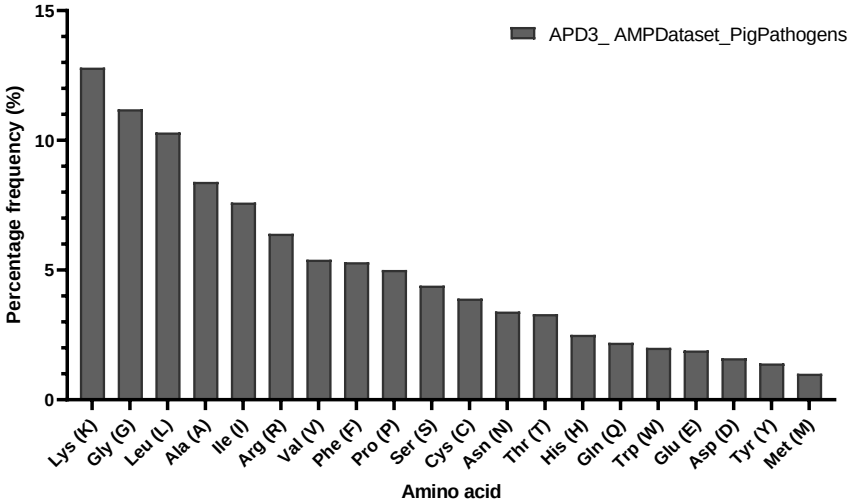
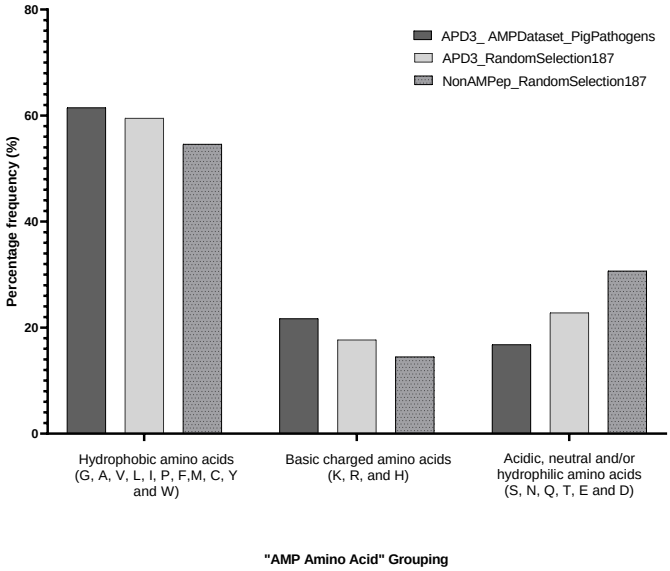


Figure 3.3 Cumulative percentage frequency of AMP-amino acid groups in AMP and Non-AMP datasets.

Non-AMPs and AMPs are discretely distinguishable, and this places certain challenges on design and prediction. Primarily, non-AMPs tend to contain more acidic, neutral and/or hydrophilic amino acids, and the comparative underrepresentation of this AMP-Amino acid group in AMPs is favoured largely for basic and hydrophobic residues. The ADP3_AMP_PigPathogensDataset contained zero anionic AMPs unlike the APD3_RandomSelection187, which likely makes this basic/hydrophobic > hydrophilic bias more prominent.



As shown in **Figure 3.3.**, there are clear biases in the representation of amino acids in the three groupings; basic charged amino acids; Lys is represented twice and four-times as much as Arg and His respectively. Biases exist within the hydrophobic amino acids with Gly, Lue, Ile and Ala being particularly prominent within peptides in the APD3_AMP_PigPathogensDataset. Stretches of hydrophobic amino acids are common in ultra-short AMPs and play a vital role in contributing to the overall amphathicity (Anunthawan et al., 2013). Python script **(Method 2.9)** was implemented to identify sequential stretches of hydrophobic amino acids (G, A, V, L, I, P, F, M, C, Y and W) across the peptide length of the 187 AMPs in the APD3_AMP_PigPathogensDataset. As demonstrated in **Table 3.1** increased peptide length in AMPs typically denotes more stretches of hydrophobic amino acids and peptides 11-20mers typically contain 2-5 stretches across the peptide region.

Table 3.1. Summary of the hydrophobic stretches identified in AMPs of varying peptide length ranges from the APD3_AMP_PigPathogensDataset

Hydrophobic stretches in the APD3_AMP_PigPathogensDataset against peptide length					
Peptide length (mers)	1-10	11-20	21-30	31-40	41-50
No. peptides within peptide length range	3	54	77	39	15
No. of hydrophobic stretches identified	1 to 3 stretches	2 to 5 stretches	2 to 8 stretches	3 to 8 stretches	5 to 10 stretches

374 combinations of ≥2 hydrophobic amino acids stretches were observed, 36.63% were only observed once whereas 63.36% were in more than one peptide in the APD3_AMP_PigPathogensDataset. Segmenting peptides by 25%, 50% and 75% residue is common way to evaluate composition across AMPs with a thread of commonality in spite of the varying peptide lengths of analysed AMPs (Bhadra et al., 2018). 123 (65.78%) peptides in the APD3_AMP_PigPathogensDataset begin with a

hydrophobic stretch (≥2 more hydrophobic amino acids), and of the 64 peptides remaining ~61% of these peptides possess a hydrophobic stretch by the 25% residue from the N-terminal. In terms of commencing the peptide sequence (N-terminal); Gly, Phe, Ile, Arg and Ala account for 78.07% of the first amino acids in the peptide chain. Hydrophobic stretches were less common at the C-terminus of peptides (29.41%), C-terminal end of peptides mainly consisted of stretches of basic amino acids (Lys, His, Arg) in combination with singular hydrophobic and/or hydrophilic amino acid residues. Cys, Leu, Ser, and Gly were the most common last amino acid residue at the C-terminus and accounted for ~56% of these residues. As shown in **Table 3.2,** twin (IG, VG, VL, LL), triplet (GLL, FLP, FLG, AAL) and quadruplet (AGLG, GIAA, CFGP, GLLL) sequential hydrophobic amino acids were most commonly observed within each parameter of stretch size. The longest hydrophobic stretches was 11 sequential amino acids were observed. Generally, longer stretches (≥5) were derived from unique peptide motifs rather than highly common in the majority of peptides in the dataset. The observed percentage of hydrophobic amino acids (G, A, V, L, I, P, F,M, C, Y and W) in these stretches is conveyed in **Figure 3.4.**

Figure 3.4. The percentage frequency of amino acids in the sequential hydrophobic stretches of peptides in APD3_AMPDataset_PigPathogens.

Clear bias for Gly (G), Lue (L) and Ala (A) amino acids in hydrophobic stretches (≥2 Aa's) of AMPs in the APD3_AMP_PigPathogensDataset. Gly (G) is the most common and observed ~ *x*10 more than the least represented hydrophobic amino acid Met (M).

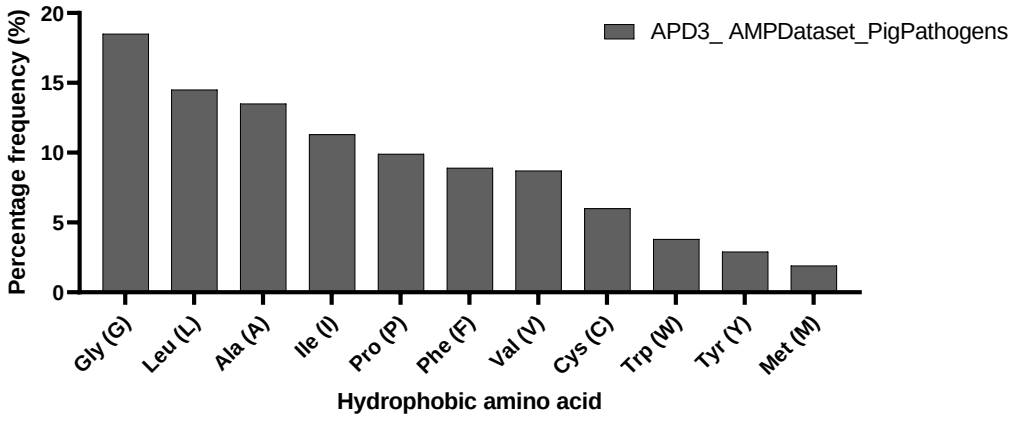


Table 3.2 Sequential stretches of ≥2 hydrophobic amino acids observed in ≥3 peptide sequences in the APD3_AMP_PigPathogensDataset

APD3_AMP_PigPathogensDataset: Sequential hydrophobic stretches (≥2 Aa's) observed in ≥3 peptides.		
Hydrophobic amino acid stretch	**No. of hydrophobic amino acid residues**	**No. of observations**
IG	2	21
VG	2	17
VL	2	16
LL	2	14
AL	2	14
AA	2	14
GLL	3	13
IF	2	13
LA	2	13
GV	2	13
GL	2	12
AG	2	11
IL	2	11
VA	2	10
FF	2	10
WL	2	10
GG	2	9
CC	2	9
LV	2	9
PP	2	8
FA	2	8
LF	2	8
LG	2	8
VF	2	8
FG	2	7
GC	2	6
AI	2	6
GI	2	6
IA	2	6
LI	2	5
FLP	3	5
FI	2	5
FLG	3	5
AAL	3	5
LP	2	5
II	2	5
GW	2	5
GIG	3	5
GLA	3	5
VAA	3	4
GA	2	4
GLW	3	4
AAG	3	4
AV	2	4
GF	2	4
AW	2	4
PI	2	4
PFP	3	3
PVYIP	5	3
GVA	3	3
ML	2	3
IFCAI	5	3
CI	2	3
IGA	3	3
GLI	3	3
IPF	3	3
GLV	3	3
IW	2	3
GFW	3	3
LAG	3	3
YG	2	3
VV	2	3
WF	2	3
AGLG	4	3
GIAA	4	3
PPI	3	3
IP	2	3
LLF	3	3
CPG	3	3
GAL	3	3
FV	2	3
LLG	3	3
IM	2	3
YL	2	3

3.4 Constructing a degenerate codon scheme mutagenesis strategy to generate a novel "AMP-biased" peptide library.

The amino acid distribution of the APD3_AMP_PigPathogensDataset summarised in **Figure 3.2** was utilised to devise the AMP biased degenerate codon strategy. Expected amino acid frequency can be assimilated as a ratio of the number of encoding codons / total codons present in the scheme. Initially explored was the impact of NNN and NNK schemes on expected frequency of amino acids across a 16 mer peptide region **(Figure 3.5).** Combining the expected frequency for each of the amino acid grouping herein demonstrated, hydrophobic and cationic charged amino acids are underrepresented compared to the representation identified in the APD3_AMP_PigPathogensDataset **(Figure 3.6).**

Figure 3.5. Expected frequency of respective encoded amino acids and stop codons across a 16mer peptide region randomised using NNN or NNK schemes.

The expected frequency of the twenty naturally occurring amino acids and stop codons across a 16mer peptide region is summarised for NNN(unaltered, unrestricted nucleotide randomisation) or NNK degenerate codon scheme, which offers reduced redundancy and similar amino acid representation as seen in NNN. Except hydrophobic amino acids W/M/I and stop codons. The duality of reduced redundancy means although NNK encodes two fewer stops, the retained amber stop codon is slightly overrepresented in comparison.

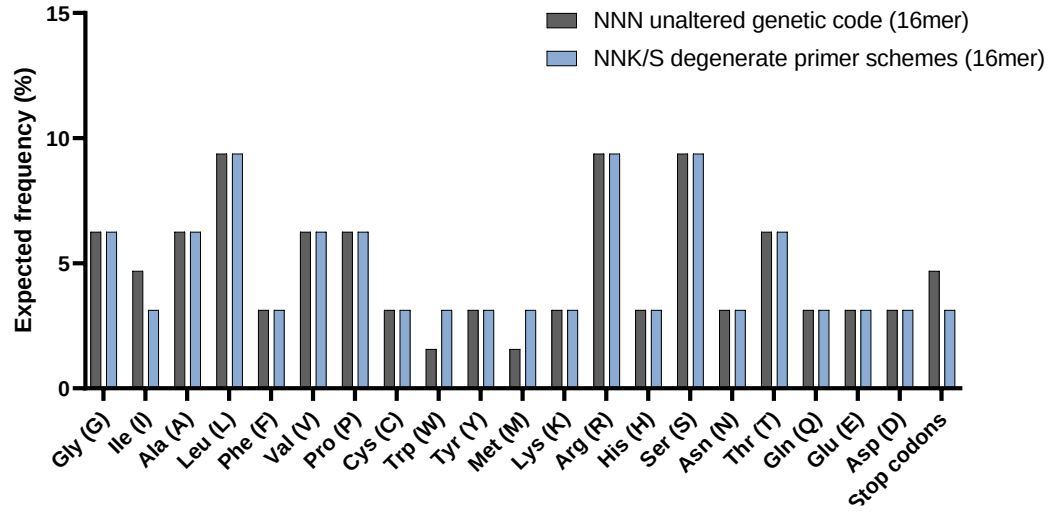
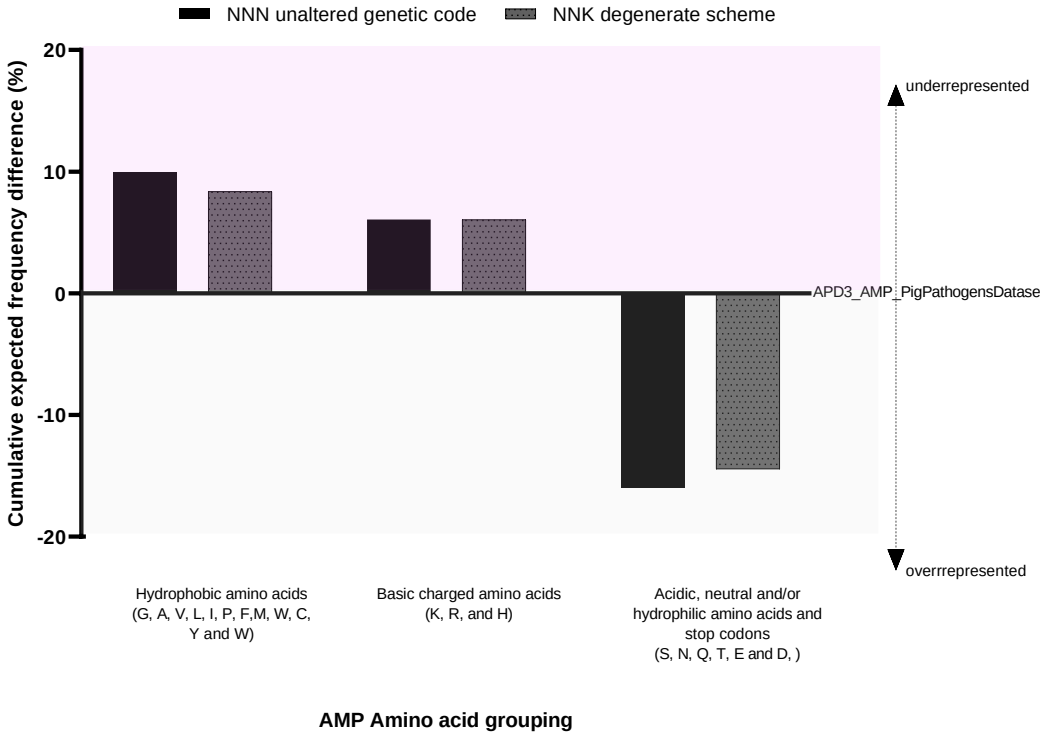


Figure 3.6. Cumulative percentage expected frequency difference of AMP amino acid groups with a peptide region randomised using NNN or NNK compared against the APD3_AMP_PigPathogensDataset.

The NNK degenerate scheme generates slightly more hydrophobic residues than NNN thus resulting in a marginal closer representation of this AMP-amino acid group. NNK overrepresentation of amber stops provides the difference for the acidic/neutral/hydrophilic and/or stop codon group. However, both hydrophobic and basic residues are underrepresented when compared to the AMP dataset (baseline).



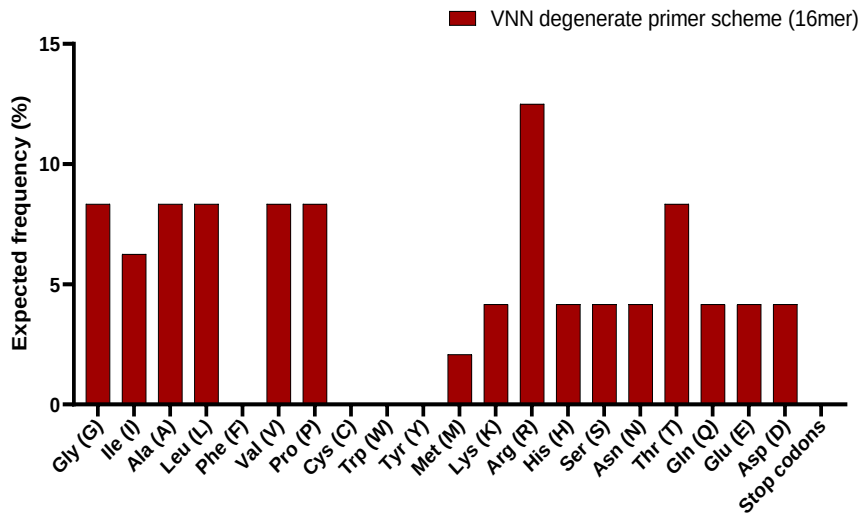
To overcome the drawbacks of NNN/NNK, the VNN was devised and **Figure 3.7** demonstrates VNN scheme expected frequency. The VNN scheme excludes C, W, Y hydrophobic amino acids which are poorly represented in the APD3_AMP_PigPathogensDataset, equating to 3.90%, 2.00% and 1.40% respectively **(Figure 3.2).** Hence the representation of C, W, Y in hydrophobic stretches is limited and in the APD3_AMP_PigPathogensDataset. <34% of AMPs contained ≥1 cysteine residues. Only <14% of the cysteine residues represented in AMPs within the APD3_AMP_PigPathogensDataset were evenly coupled cysteine residues for instance, hydrophobic stretches “CC”, “LCCL”, “GCGYCC”, “YCC”, “LWAFCC” thus indicating the role of cysteines extends much further than simply

structural contributions. Upon oxidation these twin cysteines can cross-link via disulfide bond formation which likely plays vital role in protein-structure stabilisation for certain AMPs (Katzir et al., 2006).

Nonetheless, the VNN scheme retains serine which shares similar chemical structure and properties to cysteine. Both amino acids differ by a single atom in the R-group; wherein hydroxyl (-OH) functional group of serine is replaced by the sulphur thiol (-SH) group in cysteine (Catalano et al., 2021). Additionally, as the majority of cysteine residues in the APD3_AMP_PigPathogensDataset were odd and uncoupled, if displayed these peptides could corrupt display in filamentous bacteriophage screening systems (Sieber et al., 2015). Consequently, it can be argued the VNN scheme cysteine omission is partially negated by other amino acids (*i.e.* serine and other hydrophobic/polar Aa's), and exclusion would putatively be advantageous for downstream phage display system applications.

Figure 3.7 Expected frequency of respective encoded amino acids and stop codons across a 16mer peptide region randomised using the VNN scheme.

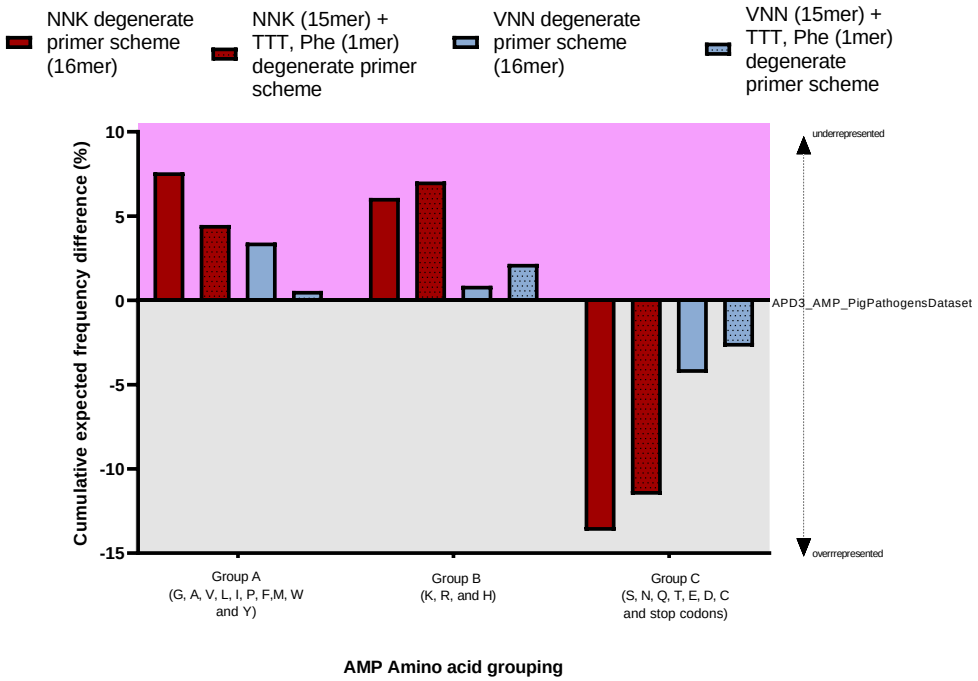
The VNN scheme contains 16 amino acids with varying expected frequencies; six hydrophobic Aa's; G/A/L/P/V (8.33% each), I (6.25%) M (2.08%) and all three basic Aa's; R (~12.50%), H/K (~4.17%) as well as hydrophilic Aa's; S/N/Q/E/D (~4.17% each) and T (8.33%). All three key AMP amino acid groupings are retained. Advantageously, compared to NNK the scheme facilitates a higher expected frequency of arginine and 45% of all hydrophobic Aa's.



However, phenylalanine reincorporation was prioritised due to the relatively high representation of phenylalanine in AMPs and respective sequential stretches of hydrophobic Aa's, especially at the N-terminus of AMP sequences in the APD3_AMP_PigPathogensDataset. The "VNN_{+}(TTT)" strategy describes the reincorporation of phenylalanine into the degenerate scheme and respective primer design. This resulted in the finalised AMP-biased degenerate strategy and **Figure 3.8** depicts the expected frequency of the AMP amino acid groupings across a 16-mer peptide region using VNN, VNN+$(TTT)_1$, NNK and NNK+$(TTT)_1$ degenerate codon randomisation strategies.

Figure 3.8 Cumulative percentage expected frequency difference of AMP amino acid grouping across a 16mer peptide region randomised using; VNN, VNN_{15}+$(TTT)_1$, NNK and NNK+ $(TTT)_1$ compared against the APD3_AMP_PigPathogensDataset.

"TTT" encodes the amino acid phenylalanine and the reintroduction of this hydrophobic residue to the VNN scheme equates to the VNN_{15}+$(TTT)_1$, strategy. Both VNN/NNK schemes overrepresent hydrophilic Aa's and amber stop codons in the case of NNK. Phenylalanine occupancy of one out of sixteen codon positions results in an overall greater representation of the hydrophobic Aa's grouping for both schemes. Expected representation of VNN encoded hydrophobic Aa's observes a -3.13% reduction but this is negated by the +6.25% expected frequency contribution of phenylalanine. This brings the representation of hydrophobic residues closer to that seen in the APD3_AMP_PigPathogensDataset. The representation of basic amino acids in the VNN_{15}+(TTT) scheme observes a <2.16% difference compared to the APD3_AMP_PigPathogensDataset and for both strategies (with or without "TTT") basic Aa's are represented ≥ three times more than the expected frequency of NNK strategies.



3.5 Whole-plasmid inverse-PCR method for the construction of naïve 16mer peptide libraries randomised using degenerate primer strategies VNN_{15}+$(TTT)_1$ and NNK.

3.5.1 Library construction: pSD3 phagemid vector library template construction

The bacteriophage display vector *pSD3* phagemid was originally constructed to maintain library diversity with optimised PCR and restriction enzyme digestion requirements (Li et al., 1999). The pIII phagemid *pSD3* construct utilised herein to generate a template for library production was constructed as outlined in **Method 2.10.1**. This *pSD3* vector contained 5' to 3'; The *pelB* leader sequence encoding pectate lyase B in *Erwinia carotovora* "MKYLLPTAAAGLLLLAAQPAMA" (Jestin et al., 2001; Mirzadeh et al., 2020), an *E. coli* codon optimised 12mer peptide insert "RLLFRKIRRLKR", (Sainath Rao, Mohan and Atreya, 2013) followed by an amber stop codon (TAG) and the *SpeI* restriction cloning site 5' A▼CTAGT 3' all of which proceed *gIII* gene with an ochre stop codon (TAA). The selected strategy herein maximises the restriction digestion of *SapI* (isoschizomer of *BspQI*) to generate overhanging sticky ends necessary to improve ligation efficiency and to allow cloning without introduction of further vector residues (from a non-type II restriction enzyme site). Therefore, the first modification required the deletion of a *BspQI* restriction site (RE site) 5' GCTCTTC$(N)_1/N_4$▼ 3' (2738-2748bp) present in this construct. *BspQI* RE site removal was achieved by using inverse PCR primers designed to introduce two amino acids (RS) in replacement of the two glutamate (E) residues in the *BspQI* site.

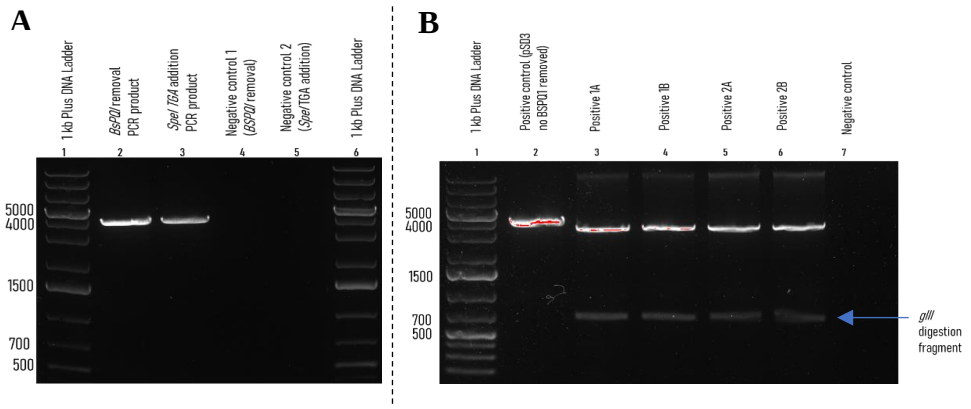
BspQI site deletion was verified via sanger sequencing with a primer which annealed ~189bp upstream of the removed restriction enzyme site in the *pSD3* vector template. Secondly, to accommodate alternative non-phage display techniques used to screen peptide libraries and ensure the sequence read-through was compatible with *E.*

coli amber suppressors and alike, it was necessary to insert an additional *SpeI* restriction site with an opal stop codon (TGA) after the *gIII*/ochre stop in the original construct. Consequently, *SpeI* restriction digestion of library DNA extractions would digest *pSD3* library variants, snipping out the phage display gene *gIII*, whereupon constructs can be ligated and utilised for downstream screening applications which do not require peptide pIII fusions. Importantly, TGA stop follows the peptide region in *SpeI* digested *pSD3* constructs to facilitate the termination of translation in *E. coli* strains irrespective of stop suppression (*SupE* etc). Consequently, a inverse PCR protocol centred on designing primers wherein the misannealing region contained the desired *SpeI*/*opal stop* replacement and the annealing region sat at the 3' end (C-terminus) of *gIII* in the *pSD3_RLL_-BsPQI* construct **Method 2.10.** PCR products were then run-on agarose gels as seen in **Figure 3.9,** with bands visualised at ~4000bp excised for *DpnI* digestion.

Figure 3.9 Verifying PCR modifications used to generate library template; *pSD3_RLL_-BspQI_+SpeI_opal stop* visualised by agarose gel electrophoresis.

(A) Lane 2 = PCR Amplicon (~3.94kb) generated for *BsPQI* RE site removal from the *pSD3_RLL* construct to generate *pSD3_RLL_-BspQI*. Lane 3 = PCR Amplicon (~3.95kb) generated from the addition of *SpeI*/TGA(opal stop) at the 3' end of *gIII* gene. Generating the finalised library construction template. Negative controls contained ddH_2O instead of DNA templates.

(B) Representative *SpeI* digestion of library template: To verify the addition of the *SpeI/opal stop* in the library template *pSD3_RLL-BspQI_+SpeI_opal stop* a *SpeI* digestion of this template and the original *pSD3*_RLL construct (only contains one *SpeI* stop). ~679bp *gIII* digestion product is generated only for the library template positives. The negative control was ddH_2O to ensure no nucleic acid contaminants were present in digestion reactions. (1kb Plus DNA ladder, Thermo Fisher)



DpnI is a methylation sensitive (dam, methylated adenine, $_{m}$A) restriction enzyme which only cleaves 5' G$_{m}$A$^{\blacktriangledown}$TC 3' found in bacterially derived constructs. *DpnI* digestion of purified PCR gel product facilitated the preservation of newly synthesised library template vector; *pSD3_RLL_-BspQI_+SpeI_*opal stop. Linear *DpnI* digested template vectors were then ligated using *T4* Ligase and subsequently transformed into *E. coli* TG1 bacteria. Whole construct sanger sequencing using the primers outlined in **Method 2.10** were employed for construct mapping and verifying *E. coli* TG1 clones with the correctly modified template *pSD3_RLL_-BspQI_+SpeI_opal stop*. *SpeI* digestion verification samples were analysed to ensure gIII gene excision could be performed on the library template **(Figure 3.9).** Functionality of this vector was verified by utilising this template for phage biopanning (Results Chapter 3) and the generation of constructs containing AMPs (Results Chapter 2) used in alternative screening strategy explored in this study.

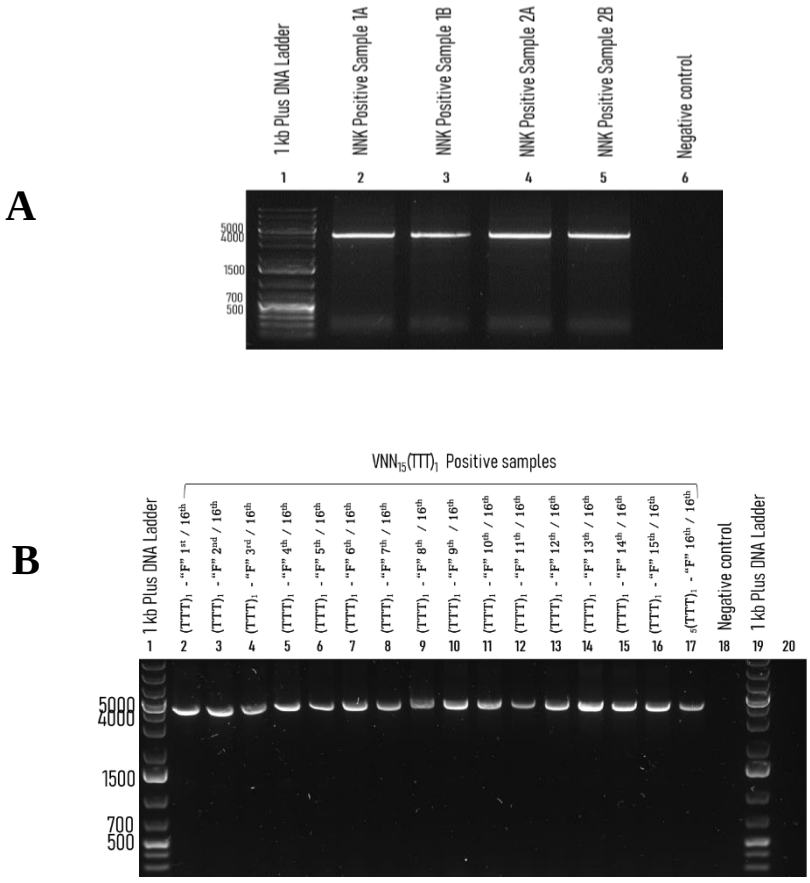
3.5.2 Library construction: Phage display peptide libraries

Exploiting a singular phagemid vector template *pSD3_RLL_-BspQI +SpeI_opal stop*, high diversity 16mer peptide libraries were generated via site directed mutagenesis strategies respectively utilising VNN$_{15}$+(TTT)$_{1}$ or NNK degenerate primers **Method 2.10.2**. The randomised peptide region diversity was introduced to the *pSD3_RLL_-BspQI_+SpeI_opal stop* vector via an inverse PCR approach, which maintains introduced diversity from misannealing degenerate forward primers within whole-plasmid construct amplicons of the library template vector. For both libraries the degenerate forward primers anneal to a constant region flanking the N-terminal region of *gIII* to ensure peptides are displayed as pIII fusions. Primer annealing additionally omitted the residing amber stop in the library template. Inverse PCR reactions were individually carried out using VNN$_{15}$+(TTT)$_{1}$_1st to 16th/16th primers.

This individualised sample preparation of the $VNN_{15}+(TTT)_1$_1^{st} to $16^{th}/16^{th}$ primers remained the case until pooling each for ligation. Gel electrophoresis confirmed amplification of correctly sized library amplicons as shown in **Figure 3.10.**

Figure 3.10 Exemplar image of the agarose gel electrophoresis used to confirm amplification after library construction inverse PCR.

Inverse PCR amplification using NNK degenerate forward prime**(A)** or $VNN_{15}+(TTT)_1$_1^{st} to $16^{th}/16^{th}$ forward primers **(B)**, desired amplicon ~4.05kb. Degenerate $VNN_{15}+(TTT)_1$_1^{st} to $16^{th}/16^{th}$ primers were carried out individually all at the same annealing temperature.. Negative control is no library DNA template, replaced with ddH_2O. (1kb Plus DNA ladder, Thermo Fisher)



DpnI and *SapI* (*BspQI* isoschizomer) were respectively utilised to cleave methylated library templates of bacterial origin or cleave the *BspQI* RE site which resides within the degenerate primer introduced region of amplicons. The appending

BspQI RE site within degenerate primers when digested results in an 3bp overhang, wherein this codon specifically encodes the last amino acid (A) represented in the *PelB* leader sequence. Intramolecular interactions between these slight overhangs were used to promote the ligation of the two ends of PCR amplicon products. Both libraries were ligated with an input of >40µg *DpnI/SapI* digested library DNA into the totality of *T4* ligase reactions. Ligation of VNN_{15}+$(TTT)_1$_F_1^{st} to 16^{th}/16^{th} *DpnI/SapI* digested library DNA was performed as an equimolar pool; wherein library DNA derived from individual primers within this degenerate set were pooled ~3µg of each, except VNN_{15}+$(TTT)_1$_F4^{th}/16^{th} and VNN_{15}+$(TTT)_1$_F12^{th}/16^{th} which were added at 2.72µg and 2.45µg respectively. Ligated library DNA was transformed into highly electrocompetent *E. coli* TG1 cells (≥× 10^{10} cfu/µg). The estimated library pIII peptide phage library size for VNN_{15}+$(TTT)_1$ and NNK were 2.0 x 10^9 and 2.32 x 10^9 respectively when titrated. *E. coli* TG1 titration of unligated NNK and VNN_{15}+$(TTT)_1$ pIII library constructs estimated the *pSD3_RLL_-BspQI_+SpeI_opal stop* library template equates to 1:1478 and 1:23,809 respectively for each library.

3.6 Sanger and Next Generation Sequencing analysis of naïve VNN_{15}+$(TTT)_1$ and NNK 16mer peptide libraries.

3.6.1 E. coli TG1 library sanger sequencing

≤26 *pSD3_-BspQI +SpeI_opal stop* VNN_{15}+$(TTT)_1$ and NNK library *E. coli* TG1 colony clones were miniprepped then sent for sanger sequencing. For the NNK pIII peptide phage library all of clones with sufficient sequence reads (11/20 clones) contain a 16mer randomised peptide region, and none of the clones recovered were library templates as shown in **Table 3.3.** The vast majority of the sufficiently sanger sequenced NNK phage library clones contained the necessary stop codon, *SpeI* restriction site and *gIII* arrangement designated downstream of the peptide region from

the library template construction. Similarly, sanger sequencing analysis of *pSD3_-BspQI +SpeI_opal stop* $VNN_{15}+(TTT)_1$ library clones revealed no library templates and correct *pSD3* modifications (post *gIII SpeI* RE site/TGA *etc*). Sufficient sequence reads of $VNN_{15}+(TTT)_1$ library clones (11/26 clones) contained varying randomised peptide region lengths; 13, 15, 16 or 17mers as summarised in **Table 3.3.**

Table 3.3 Sanger sequencing of randomly handpicked pIII peptide phage $VNN_{15}+(TTT)_1$ and NNK library *E. coli* TG1 clones.

Table 3.3A; : $VNN_{15}+(TTT)_1$ sanger sequencing of handpicked pIII peptide phage library clones

Clone	read length (bp)	Peptide region amino acid sequence
1	369	XXVDVGQVQVIDGVFL
2	408	VIVFTQPEQTGVRTG [2]
3	540	RRGPRGFIRGLVNERD
4	660	TDGTSFDSSPKIVLEG
5	537	LXNXXKRGMAQDSGDRE [3]
6	403	PLLKIRGNTXSGVLXT
7	595	QDRDVLKGPDQFARVG
8	369	IFLRVMASGLTAAVRG
9	885	VDGATHEQKVRDDAFD
10	561	TVHMSSPTRGITDFVR
11	285	EKSLPPSFIPTDF [1]

Denotation of peptide region lengths which vary from desired 16 mers; [Peptide amino acid sequence] [1] = 13 mers , [2] = 15 mers, [3] = 17 mers

Single letter amino acid code utilised, 'X' refers to an unquantified amino acid codon.

Table 3.3B: NNK sanger sequencing handpicked pIII peptide phage library clones

Clone	Read length (bp)	16 mer peptide region amino acid sequence
1	875	AM*VKNFVSPWMNDS
2	872	WKSRLHSAARERSTGL
3	873	GVLLVSGLEVEHXI*G
4	1146	HXDXWTLXTELM*WPS
5	872	ATPASGVVWGFWNSAT
6	872	LEYGNCG*CLAARPLA
7	554	LEYGNCG*CLAARPLA
8	841	ISVAVLIYCGSTCGLG
9	872	GQTSLGTLWYLVRFLA
10	528	GRTHPALMKCFSDAFE
11	873	PNCWSLKRSTKXRDAW

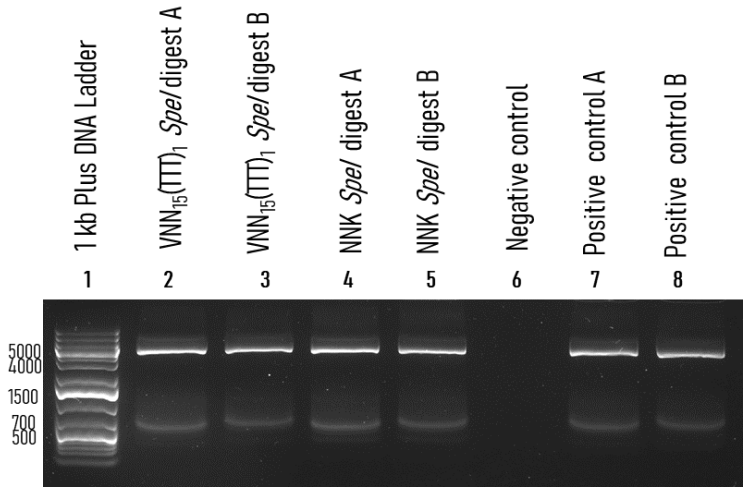
'*' in this context refers to an amber stop codon, specifically TAG sequence read.

Single letter amino acid code utilised, 'X' refers to an unquantified amino acid codon.

pSD3_-BspQI +SpeI_opal stop $VNN_{15}+(TTT)_1$ and NNK *E. coli* TG1 glycerol stocks were midiprepped and quantified via Qubit. >30µg library DNA obtained from the midiprep of 1000 µl *E. coli* TG1 $VNN_{15}+(TTT)_1$ and NNK library glycerol stocks were subsequently *SpeI* digested to remove the *gIII* in library *pSD3* constructs. Gel electrophoresis was performed, as shown in **Figure 3.11,** to visualise the ~679bp *gIII* restriction digestion fragment of library DNA. Nonetheless, gel visualisation of *SpeI*

library digests intended for ligations was omitted and the verification of complete and correct digestion was conducted under brief UV exposure during gel excision of the desired ~3.37kb band. *SpeI*-digested sub-library DNA was then ligated and <10µg of respective circular DNA for each sub-library was transformed into electrocompetent *E. coli* JE5505. Titrations of *E. coli* JE5505 sub-libraries estimated library sizes for both libraries were; 1 x 10^7 for VNN_{15}+$(TTT)_1$ and 5 x 10^6 for NNK. *E. coli* JE5505 sub-libraries.

Figure 3.11. Agarose gel electrophoresis visualisation of *SpeI* digests of VNN_{15}+$(TTT)_1$ and NNK pIII peptide phage constructs.
Lane 2-5: One *SpeI* digestion reaction (Method 2.6.8) was split into two lanes for both libraries (A or B). ~679bp *gIII* digest fragment can be seen for these positive library samples. The positive control was the *SpeI* digest of library template; *pSD3_RLL_-BspQI +SpeI_opal stop* (lane 7-8). Negative control = No DNA digest template (Lane 6). (1kb Plus DNA ladder, Thermo Fisher).



3.6.2 E. coli JE5505 Sub-library sanger sequencing

Titrations of *E. coli* JE5505 sub-libraries estimated library sizes for both libraries were; 1 x 10^7 for VNN_{15}+$(TTT)_1$ and 5 x 10^6 for NNK. *E. coli* JE5505 sub-libraries. Ten *E. coli* JE5505 transformants per sub-library were analysed via sanger sequencing to verify the removal of *gIII* and once again review the introduced peptide diversity in constructs **(Table 3.4)**. Sanger sequencing of sufficient quality (70-90% of clones)

highlighted all peptide regions of transformant clones were 16mers and correctly followed by a *SpeI* RE site and opal stop (TGA). No library template containing the RLL peptide was identified. However, the vast majority of clones contained amber stop codons in the peptide region; equating to 57.1% for the NNK sub-library clones and 77.7% of the $VNN_{15}+(TTT)_1$ sub-library clones. Additionally, one $VNN_{15}+(TTT)_1$ sub-library clone possessed an opal stop. 2/9(~22.2%) $VNN_{15}+(TTT)_1$ sub-library clones did not contain a phenylalanine present in the peptide region. Exclusion of phenylalanine and inclusion of stop codons in the peptide region might indicate some minor inefficiency in degenerate oligonucleotide primer synthesis or sanger sequencing accuracy.

Table 3.4 Sanger sequencing of randomly handpicked -pIII (No *gIII*) peptide VNN15+(TTT)1 and NNK sub library *E. coli* JE5505 clones.

Table 3.4A $VNN_{15}+(TTT)_1$ sanger sequencing of randomly handpicked handpicked *E. coli* JE5505 sub-library clones

Clone	read length (bp)	Peptide region amino acid sequence
1	219	VLINL*VLRRSIRRRQ
2	214	DKSLRKWV*IQFCEMW
3	216	SVLDRWGPLCRLGLGG
4	213	IRAVCILDSTMDY*GW
5	215	FLPRYGCGVNMGWG*K
6	217	SYGFIGRPISG*FYLW
7	217	RLNARFT*GFEKFVGC
8	213	YIYGCTA*GLLLFRAT
9	212	FFITLQW*EYMCFARA

'*' and '*' in this context refers to an amber stop codon (TAG) or opal stop codon (TGA) respectively.

Single letter amino acid format used.

Figure 3.4B: NNK sanger sequencing of randomly handpicked *E. coli* JE5505 sub-library clones

Clone	Read length (bp)	16 mer peptide region amino acid sequence
1	208	LRL**E*GPA*TAWRN
2	215	TRSFCPWFAWSWTHKP
3	214	VW*SL*KSWHTWCLLS
4	214	STFPVSARWMHMLRCK
5	214	ALVWMGMDGHKVM*SW
6	215	RLWITVPGFRQGNGNP
7	211	EGNIHCYWTLDRYR*E

'*' in this context refers to an amber stop codon, specifically TAG sequence read.

Single letter amino acid format used.

3.6.3 Next generation sequencing of naïve peptide phage libraries

For both VNN_{15}+$(TTT)_1$ and NNK pIII phage display libraries; Ex-phage superinfection of mid-log phase *E. coli* TG1 was utilised to propagate libraries as phage (**Method 2.11.2**). This phagemid-helper phage system fundamentally relies upon recombinant displayed peptide pIII fusion (phagemid derived) and additional vital phage proteins supplied (helper phage derived) in order to generate peptide displaying phage particles (Ayriss and Bradbury, 2006). VNN_{15}+$(TTT)_1$ and NNK pIII phage libraries in *E. coli* TG1 and respective -*gIII* absent sub-libraries in *E. coli* JE5505 were miniprepped. Aforementioned library DNA extraction samples were then subjected to Next Generation Sequencing (NGS) preparation, for library quality control and peptide region compositional analysis. Barcoded NGS primers would anneal and amplify a region of *pSD3_-BspQI +SpeI_opal stop* VNN_{15}+$(TTT)_1$ and NNK peptide library constructs from transformants; *E. coli* TG1, *E. coli* JE5505 stocks and *M13* phage propagations (**Method 2.12**). More specifically, NGS barcoded primers designed and utilised herein amplify a short sequence including the peptide region, and amplicon size differed depending on the presence of *gIII* within the library DNA construct.

NGS round 1 PCR amplification as shown in **Figure 3.12**, *pSD3* gIII+ VNN_{15}+$(TTT)_1$ and NNK libraries produced an ~287bp amplicon whereas *gIII*- sub-libraries generate a slightly shorter amplicon of ~215bp. The second sequential PCR step for NGS preparation increases round 1 amplicon by ~60bp; resulting in a ~347bp amplicon for *pSD3 gIII*+ VNN_{15}+$(TTT)_1$ and NNK libraries and ~275bp for *gIII*- sub-libraries (**Figure 3.12**). The Ion Torrent sequencing data returned for both libraries

contained within *E. coli* TG1, *E. coli* JE5505 and *M13* phage were *in silico* analysed to determine any cloning or library biases (**Method 2.12**).

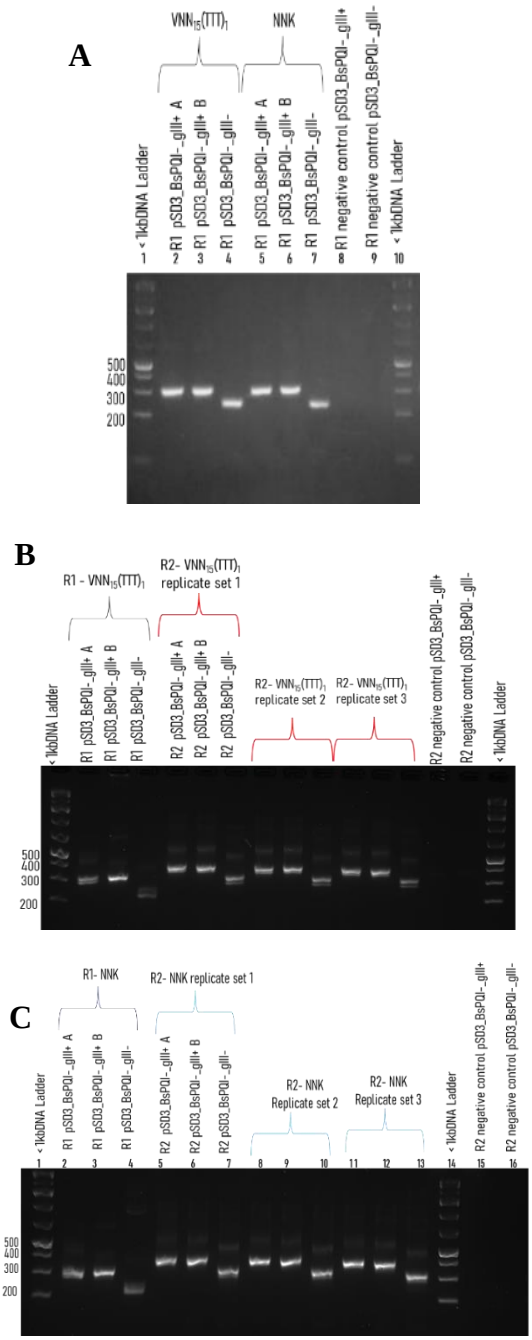


Figure 3.12 Representative agarose gel electrophoresis images post-PCR NGS sample preparation for VNN_{15}+$(TTT)_1$ and NNK naïve peptide libraries in *E.coli* TG1, *E. coli* JE5505 and *M13* phage.

(A) Round 1 amplification for both VNN_{15}+$(TTT)_1$ and NNK; *gIII (*present) peptide phage libraries in *E. coli* TG1 and *M13* phage produce ~287bp desired PCR amplicon Sub-libraries(gIII absent) in *E. coli* JE5505 produce a ~215bp desired PCR amplicon. Negative controls contained no DNA template to ensure no unspecific amplification occurred. <1kb DNA Ladder (NEB). Round 2 amplification for both VNN_{15}+$(TTT)_1$**(B)** and NNK **(C)** increases the desired PCR amplicon size by ~60bp in comparison to Round 1. Negative controls contained no DNA template, replaced with ddH_2O. <1kb DNA Ladder (NEB)

3.6.4 Deep sequencing and quality control of naïve peptide libraries

NGS data was analysed using several perl scripts (**Method 2.12**). For the pIII 16mer peptide phage libraries contained in *E. coli* TG1 and *M13* phage > 1.3 x 10^6 barcoded sequences with ≥1Aa insert were identified **(Table 3.5).** The insert

represents the amino acid sequence between our designated flanking left and right constant motifs and corresponds to the length of peptides displayed as pIII fusions. Irrespective of degenerate primer randomisation strategy; for the pIII phage libraries the vast majority of displayed peptides were the desired length, as shown in **Figure 3.13**. For instance, 16mer peptide insert occurrence for the $VNN_{15}+(TTT)_1$ pIII phage libraries were; *E. coli* TG1, 92.27% and *M13* phage, 89.84% with a slight increase in inserts ≥17Aa in the M13 propagated library. NNK pIII phage peptide libraries observed similar prevalence however, with no obvious deviations between library TG1 transformants and *M13* propagation. 16mer peptides insert occurrence observed were; *E. coli* TG1 NNK library, 91.53% and *M13* phage NNK library, 91.27%. In this study, 'unique sequences' denotes the number of sequences with the desired 16mer insert size and correct constant flanking motifs. For $VNN_{15}+(TTT)_1$ and NNK pIII phage libraries described herein, a significant majority of unique sequences identified were singletons (**Table 3.5**), with <8% of sequences identified as repeats.

As depicted in **Tale 3.6**, for the peptide libraries (absent of pIII) transformed into *E. coli* JE5505, $\sim 10^5$ barcoded unique sequences with 16Aa insert were analysed for both $VNN_{15}+(TTT)_1$ and NNK sub-libraries, which were titrated at library size: $\sim 10^{6-7}$. Sequencing demonstrated the vast majority of peptide inserts were 16mers in length specifically; 92.11% and 92.01% for $VNN_{15}+(TTT)_1$ and NNK *E. coli* JE5505 sub-libraries respectively. However, in comparison to the pIII phage libraries, the *E. coli* JE5505 sub-libraries possessed a greater proportion of repeated unique sequences with only 50.4-51.8% of these sequences identified as singletons (**Figure 3.13/Table 3.6**).

Table 3.5. Summary of pIII 16mer peptide phage library (*E. coli* TG1 and M13 phage) Next Generation Sequencing analysis

Unique sequences analysed and prevalence of singletons for pIII peptide phage libraries				
Library	**$VNN_{15}+(TTT)_1$ *E. coli* TG1**	**NNK *E. coli* TG1**	**$VNN_{15}+(TTT)_1$ *M13* phage**	**NNK *M13* phage**
Unique sequences[3]	1.399601×10^6	1.218179×10^6	1.247828×10^6	1.402110×10^6
Total number of singleton sequences (%)[4]	1.294692×10^6 (92.5%)	1.116541×10^6 (91.7%)	1.183182×10^6 (94.8%)	1.312380×10^6 (93.6%)
Key [1] = Sequences identified with the specified barcode present in the NGS analysis dataset which contained in totality ~8.86×10^8 reads. [2] = Number of translated barcoded sequences (Frames 1-3) with≥1 Amino acid in the insert site, percentage analysed from the total number of specified barcoded sequences. [3] = The number of sequences with the correct insert size (16 mer) and correct left/right motifs. [4] = Singleton sequences refers to barcoded sequences which were identified as unique and only seen once in the analysed dataset. This is expressed as a percentage of the unique sequences.				

Figure 3.13. Peptide insert lengths in barcoded sequences identified with the correct flanking motifs and >1Aa in insert across all six naive peptide libraries.

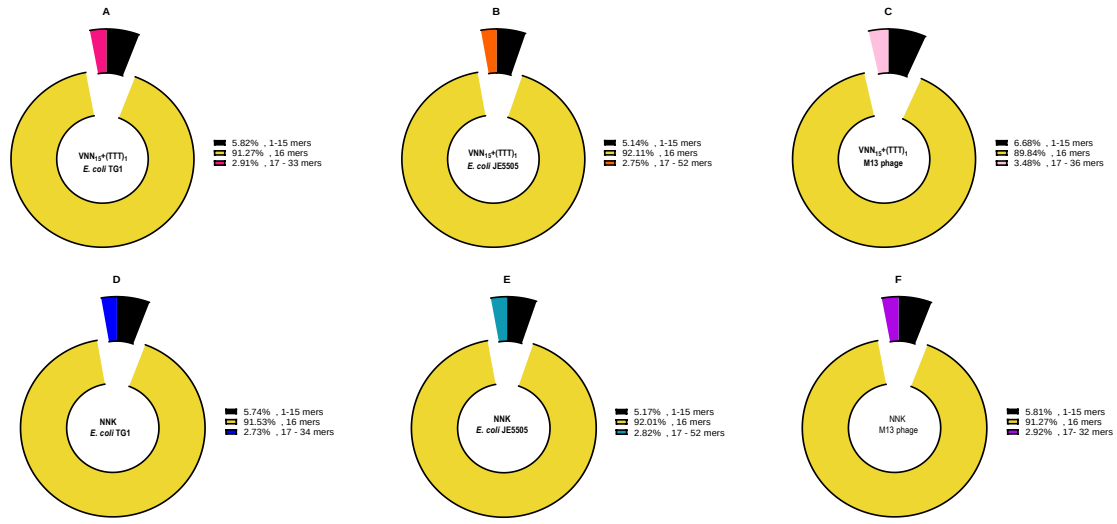


Table 3.6. Summary of *-gIII* absent 16mer peptide sub library (*E. coli* JE5505) Next Generation Sequencing analysis

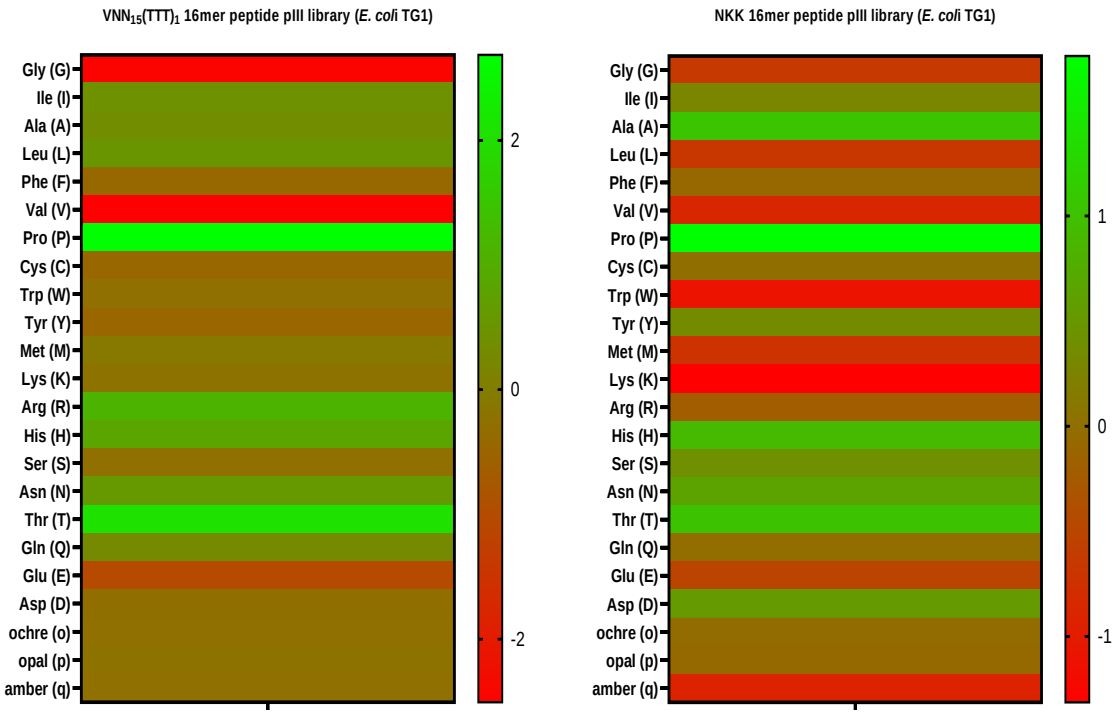
Table 3.6 : Summary of -gIII absent 16mer peptide sub library (E. coli JE5505) Next Generation Sequencing analysis		
Library	**$VNN_{15}+(TTT)_1$ *E. coli* JE5505**	**NNK *E. coli* JE5505**
Unique sequences[3]	3.12034×10^5	2.90224×10^5
Total number of singleton sequences (%)[4]	1.61702×10^5 (51.8%)	1.46343×10^5 (50.4%)
Key [3] = Selected number of barcoded sequences with two correct flanking regions either side of a distinct peptide insert site, ≥1Aa. [4] = Singleton sequences refers to barcoded sequences which were identified as unique and only seen once in the analysed dataset. This is expressed as a percentage of the unique sequences.		

3.6.5 Comparing the expected frequency of amino acids to the observed distribution in naïve libraries

Fundamentally in comparison to NNK-randomisation, the $VNN_{15}+(TTT)_1$ scheme should advantageously; *i)* eliminate cysteines and thus any effect of unpaired cysteine on phage display screening *ii)* reduce hydrophobic amino acids which tend to be poorly represented in AMPs (*i.e.* tryptophan, tyrosine) and *iii*) eliminate the occurrence of stop codons in the peptide inset region. **Figure 3.14** summarises the percentage difference between the expected frequency predictions for the $VNN_{15}+(TTT)_1$ and NNK schemes compared to the amino acid distribution observed from the NGS analysis of naïve pIII phage display libraries transformed into *E. coli* TG1.

Figure 3.14. $VNN_{15}+(TTT)_1$ and NNK pIII 16mer peptide phage libraries transformed into *E. coli* TG1; Heat map of the percentage difference between the expected and observed frequency for the twenty natural amino acids and three stops.

Two colour scale heat map wherein Red denotes "overrepresentation" and Green denotes "underrepresentation" of a specific codon compared to expected frequency.



VNN_{15}+$(TTT)_1$ (*E. coli* TG1) naïve library exhibited an amino acids and stop codons observed occurrence which deviated by ~ ± 3% than theoretical expected frequency (**Figure 3.14)**. In the VNN_{15}+$(TTT)_1$ pIII peptide phage library (*E. coli* TG1) the most underrepresented amino acid (Pro, P) possessed a +2.68 percentage difference compared to projections, signifying proline residues were identified 2.68% less than expected (**Figure 3.14).** The most overrepresented amino acid (Valine, V) registered an -2.51 percentage difference compared to expected projections (**Figure 3.14**). This signifies valine composition in library peptides was observed at 2.51% higher than expected. In addition to valine, other overrepresented amino acids included; Glycine (G), Glutamate (E), F, C and Y. Nonetheless, then observed glycine and valine overrepresentation by percentage difference from expected is five times higher than cysteine and tyrosine.

However, as mentioned previously the VNN_{15}+$(TTT)_1$ strategy sought the elimination of cysteine and tyrosine and NGS analysis revealed the residual inclusion of both amino acids. This accounts for ~ <0.5% of the total amino acids analysed across VNN_{15}+$(TTT)_1$ libraries. This arguably negligible inclusion was additionally observed for all three stop codons (<0.63%). The occurrence of stops codons in VNN_{15}+$(TTT)_1$ (*E. coli* TG1) library was approximately ~five to six times lower than the expected frequency (3.13%) and observed frequency (<4.03%) of amber stop codons for NNK library in *E. coli* TG1. VNN_{15}+$(TTT)_1$ degenerate scheme design dictated library clones possess phenylalanine as 1/16 amino acids (6.25%) in respective peptide region sequences. Phenylalanine was slightly more overrepresented than expected, with a 0.41% percentage higher representation than expected (**Figure 3.14**). This contrasts sanger sequence elucidation, wherein this reintroduced amino

acid was absent from the peptide region of the majority of $VNN_{15}+(TTT)_1$ handpicked library clones.

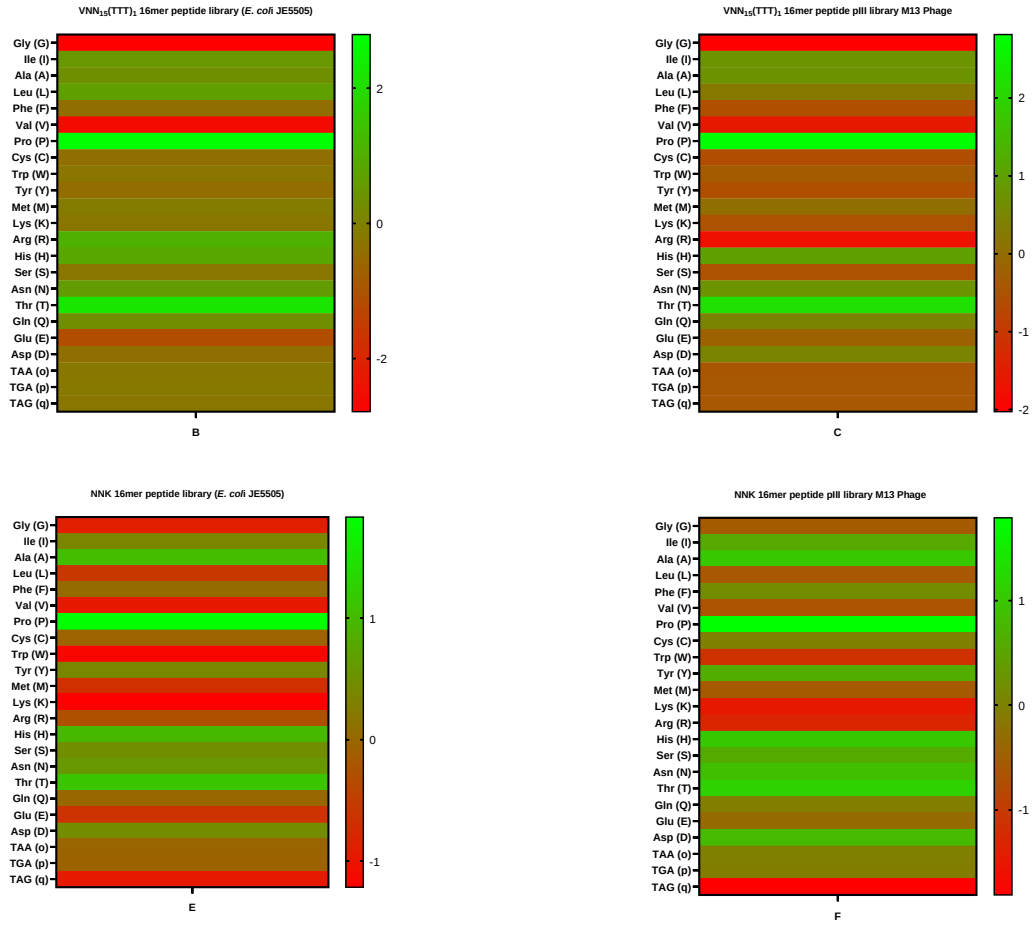
The pIII phage library randomised using the NNK degenerate scheme and transformed into *E. coli* TG1 demonstrated a differing pattern of over- and under-representation to the $VNN_{15}+(TTT)_1$ equivalent library (**Figure 3.14**). However, the most underrepresented amino acid in the NNK pIII phage library (TG1) was proline with a +1.76% difference compared to expected projections. Histidine was slightly underrepresented across both randomisation schemes however, unlike the $VNN_{15}+(TTT)_1$ equivalent library, basic amino acids Arg and Lys were slightly overrepresented collectively than expected, with Lys being the most over-represented amino acid (1.32% overrepresentation) in the NNK pIII phage library. Other overrepresented residues included; Val, Trp(W), amber stops (TAG), and Met (M) with amber stops consisting of ~4% of codons in peptide insert regions.

The libraries generated in *E. coli* TG1 transformants were applied downstream for sub-library construction and phage propagation. For both the $VNN_{15}+(TTT)_1$ and NNK randomised libraries the occurrence of amino acids in naïve *E. coli* JE5505 sub-libraries weakly deviated from the patterns of over- and under-representations seen in *E. coli* TG1, as demonstrated in **Figure 3.15**. Interestingly for both randomisation schemes, phage propagation resulted in censorship shifts of over- and underrepresentation in favour of more hydrophobic and basic amino acids residues **Figure 3.15.** For instance, both $VNN_{15}+(TTT)_1$ and NNK pIII phage library propagated in *M13* observed a sharp increase in the overrepresentation of arginine (R) in peptides. Specifically, for NNK pIII phage libraries arginine was overrepresented approximately six times more in M13 phage than *E. coli* TG1. This difference was even more drastic in $VNN_{15}+(TTT)_1$ pIII libraries where arginine underrepresentation

in *E. coli* TG1 (1.13% less than expected) was reversed in M13 (1.74% more than to expected) **Figure 3.14 and Figure 3.15.**

Figure 3.15. VNN_{15}+$(TTT)_1$ and NNK pIII 16mer peptide phage libraries transformed into *E. coli* JE5505 and *M13* phage; Heat map of the percentage difference between the expected and observed frequency for the twenty natural amino acids and three stops.

Two colour scale heat map wherein Red denotes "overrepresentation" and Green denotes "underrepresentation" of a specific codon compared to expected frequency.



3.6.6 Positional analysis of amino acid distribution across peptide insert region in naïve peptide libraries compared to natural AMPs

Phenylalanine was the seventh most common amino acid identified in the APD3_AMP_PigPathogensDataset. To generate the VNN_{15}+$(TTT)_1$ library a relatively equimolar mixture (<3ug each) of library DNA generated from sixteen degenerate primers with alternating phenylalanines were pooled and ligated. Despite

almost equimolar insertion into the ligation mixture phenylalanine representation across the 16mer peptide region in *E. coli TG1* varied (**Figure 3.16**). This could reflect the quantity of $VNN_{15}+(TTT)_1$_F 1st to 16th library variants in the finalised ligation clean-up and/or *E. coli* biases during transformation. Nonetheless, in both *E. coli* strains and *M13* phage $VNN_{15}+(TTT)_1$ peptide library variant tends to be composed of phenylalanine at terminal positions 1, 2, 13, 14 and 15 rather than the interior of the peptide region. In *M13* phage and *E. coli* JE5505, a significant reduction in phenylalanine is observed at the 16th codon C-terminus of peptides and largely in favour for phenylalanine at the 13th codon position (**Figure 3.16).** Interestingly, this censorship of phenylalanine at the 16th codon position was additionally observed for NNK 16mer peptide libraries in *E. coli* TG1, JE5505 and M13 (**Figure 3.17**). The saturation of phenylalanine at terminal ends in $VNN_{15}+(TTT)_1$ peptide library variant in some capacity coincidentally reflects the common positional incorporation pattern of phenylalanine in some natural AMPs analysed within the APD3_AMP_PigPathogensDataset.

Figure 3.16. Observed frequency of phenylalanine across the 16mer peptide region in $VNN_{15}(TTT)_1$ randomised peptide libraries

Three scale heat map utilised to visualise percentage frequency <lighter shades to yellow represent greater representation of phenylalanine.

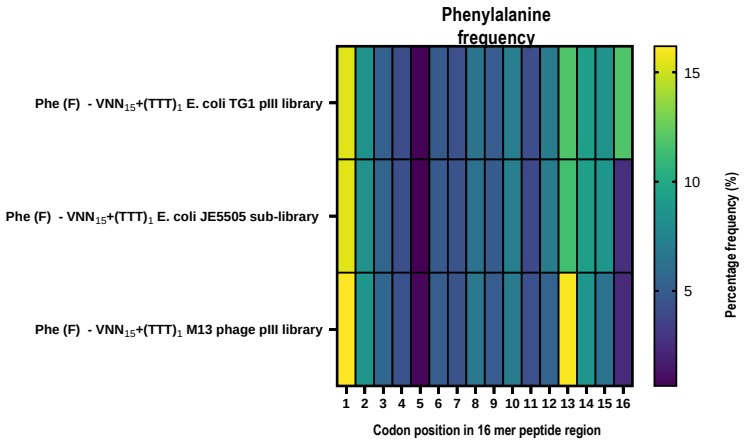
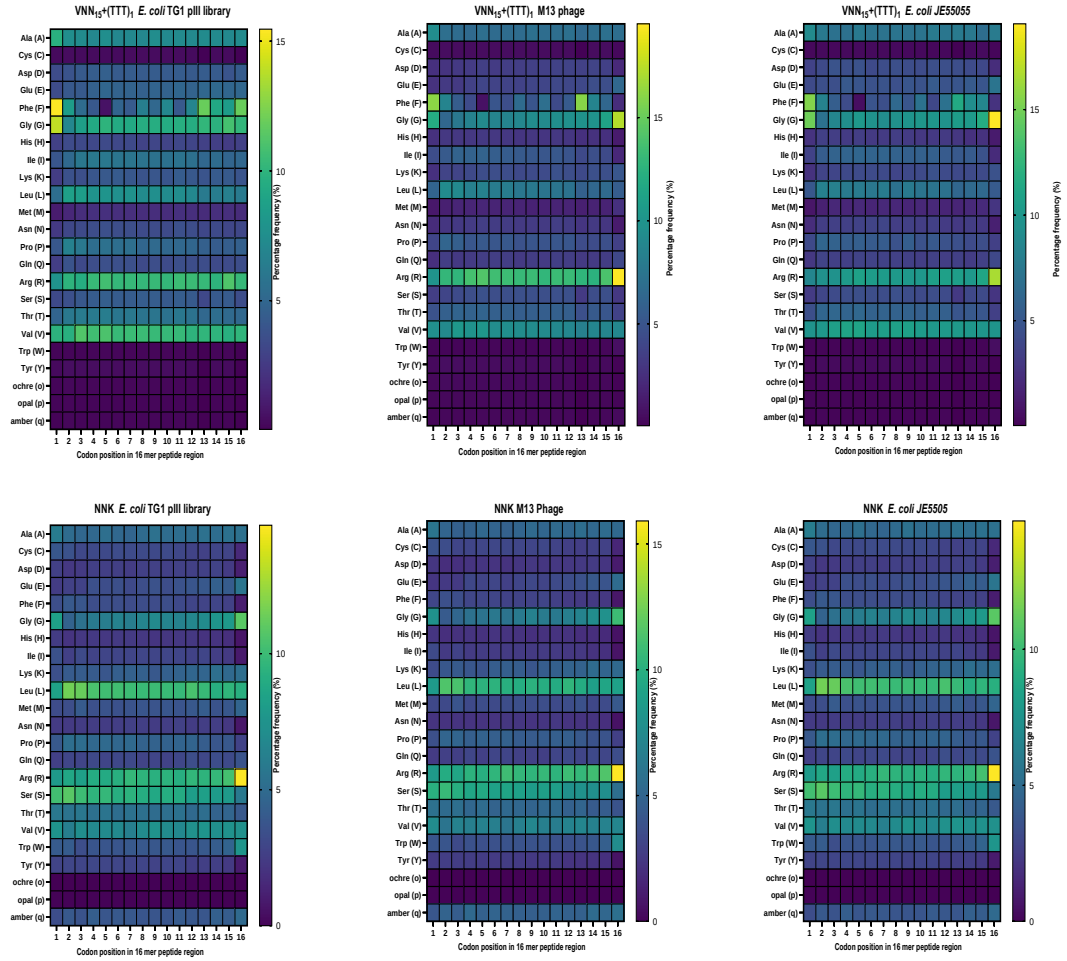


Figure 3.17. Percentage frequency of amino acids and stop codons across the 16mer peptide region in $VNN_{15}(TTT)_1$ and NNK libraries in *E.coli* TG1, JE5505 and *M13* phage

Three scale heat map utilised to visualise percentage frequency <lighter shades to yellow represent greater representation of a specific codon residue (amino acid or stops).



Herein this study, three principal groupings were architected in context of the key contributing characteristics of the twenty natural amino acids and stop codons in AMPs. Principally this includes two highly desired groups; hydrophobic Aa's G, A, V, L, I, P, F, M, W, C and Y, basic (cationic) Aa's; K, R and to a lesser extent desired however crucial for contributing to peptide amphiphilicity are hydrophilic and/or neutral Aa's; S, N, Q, T, E and D. Amber stops and the remaining two stops are placed in the third grouping of codons desired to a lesser extent. Focusing on these three

groupings complements analysis of the true randomisation achieved in VNN_{15}+$(TTT)_1$ and NNK libraries with amino acid distribution, both compositionally and positionally, to assess whether peptides emulate previously analysed natural AMPs.

Additionally, of interest for this process of analysis is to draw conclusions on whether any alterations to amino acid censorship or library diversity are present when libraries are transformed into *E. coli* or propagated as *M13* phage. In context of the VNN_{15}+$(TTT)_1$ libraries, hydrophobic Aa's; such as A, F, I and V as well as hydrophilic Aa's; D, E across the 16mer region are favoured in *E. coli* TG1 however, these Aa's are censored somewhat in observed frequency in *M13* phage as shown in (**Figure 3.17**). Whereas the overrepresentation of certain amino acids *i.e.* glycine in both *E. coli* TG1 and *M13* phage are retained. Glycine overrepresentation might be a beneficial trait as one of the most common amino acids in hydrophobic stretches of AMPs analysed in the APD3_ AMP_PigPathogenDataset. VNN_{15}+$(TTT)_1$ 16mer peptide sequences analysed for both *E. coli* TG1 and *M13* phage retained positional bias of glycine, with this amino acid saturating representation at the terminal ends, especially the C-terminus, of the peptide region. Contrastingly, the overrepresented valine residues in *E. coli* TG1 are slightly reduced in representation from N- to C-terminal across the 16mer peptide region in the VNN_{15}+$(TTT)_1$ *M13* propagation (**Figure 3.17**).

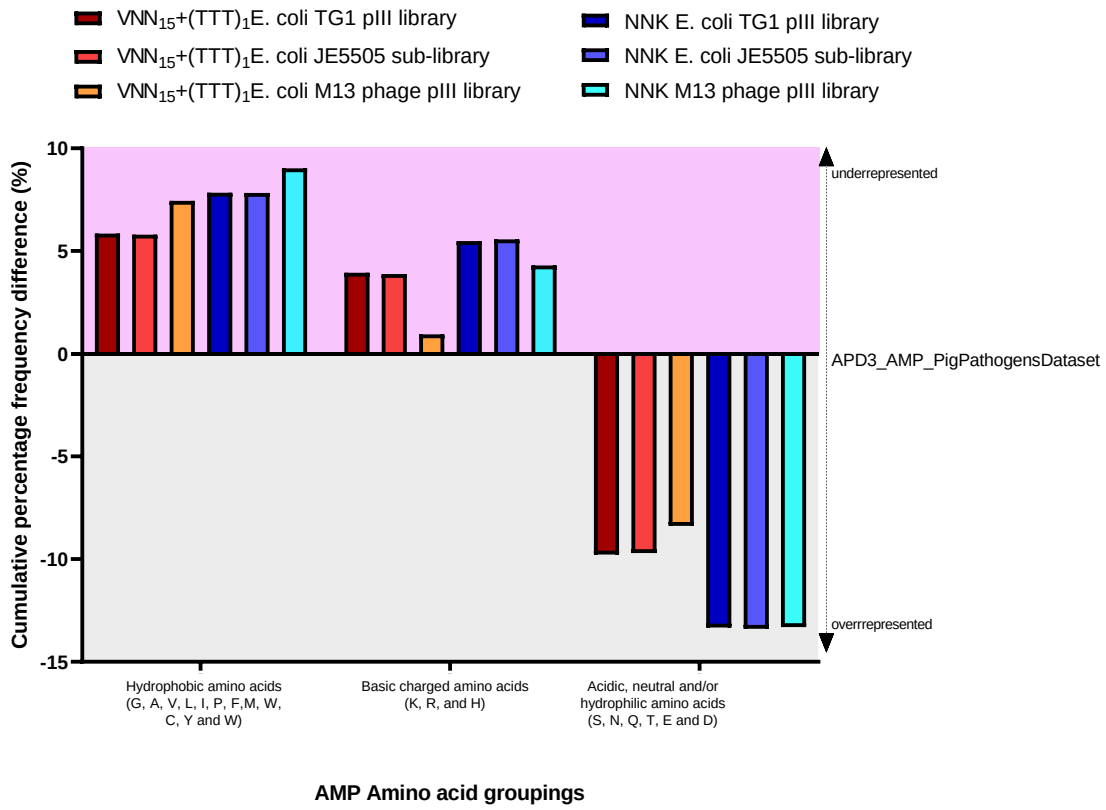
The compositional reduction of A, F, I, V, D, E representation in *M13* phage propagation of VNN_{15}+$(TTT)_1$ peptide library seemed to be distributed primarily to; basic amino acids arginine (R) which positionally seems to saturate at the C-terminus of the peptide region. Notably, arginine representation in *M13* phage was biased in comparison to the retention of peptides consisting of lysine (K) or histidine (H). A similar trend of R > K/H bias was observed in the VNN_{15}+$(TTT)_1$ peptide *E. coli*

library. The pattern of R > K/H basis for the basic amino acids and arginine saturation at the C-terminus as seen in $VNN_{15}+(TTT)_1$ library was additionally replicated in NNK 16mer peptide libraries.

Similarly, in comparison to *E. coli* strains a series of hydrophobic and hydrophilic amino acids (D, N, F, I, V, Y) saw minute reductions in representation in NNK derived peptides propagated in *M13* phage. Whilst arginine and amber stop codons largely benefited in higher representation across the peptide region. NGS analysis highlighted stop codons opal, amber and ochre are retained at minute levels *e.g.* <0.4% of codons represented in the $VNN_{15}+(TTT)_1$ 16mer peptide libraries. The overrepresentation of amber stops across all three NNK 16mer peptide libraries, means < 4-5% increase in the occurrence of hydrophilic glutamine (Q) across the peptide region. As *E. coli* amber suppressors convert amber stop codons into glutamine (Q), and this readthrough efficiency results in the translation of ~70% of an untruncated peptides (Normanly et al., 1986; Belin and Puigbò, 2022). Consequently, this would result a slight increase in overall overrepresentation of hydrophilic in NNK 16mer peptide library variants.

Across both VNN/NNK libraries censorship correction in *M13* phage seems to lead generally to reduced observed frequency of certain hydrophobic or hydrophilic Aa's (*i.e.* phenylalanine, valine and aspartate) but these shifts were more obvious in the $VNN_{15}+(TTT)_1$ library. Nonetheless, noteworthy to state these slight fluctuations in observed frequency rarely negated the over- or underrepresentation of a specific amino acid against the expected frequency each degenerate scheme was theorised to contain. The only significant exception was the overrepresentation of arginine residues in $VNN_{15}+(TTT)_1$ library in M13 phage compared to *E. coli* strains. Nonetheless, fluctuations in amino acid observed frequency did result in positive shifts in the overall hydrophobicity of peptides identified from *M13* phage for both libraries (**Figure 3.17).**

Figure 3.18. Overview of the AMP amino acid grouping distribution in naive $VNN_{15}(TTT)_1$ and NNK naïve 16mer peptide libraries



In context of the AMP amino acid groupings "basic, hydrophobic, hydrophilic Aa's, the $VNN_{15}+(TTT)_1$ library from *E. coli* TG1, JE5505 to *M13* phage all show closer representation than their NNK equivalent to the distribution pattern observed in APD3_ AMP_PigPathogenDataset (**Figure 3.18**). Nonetheless, for all three amino acid groupings $VNN_{15}+(TTT)_1$ and NNK libraries deviate from the desired pattern (-13.4 to +9.0). AMPs in the APD3_ AMP_PigPathogenDataset were more hydrophobic and contained less hydrophilic/neutral amino acids than NGS observed peptides fron both degenerate schemes. However, propagation in *M13* phage facilitated a censorship correction which inclined library peptides towards being more cationic and thus a closer representation of this overall physiochemical property as seen in the APD3_ AMP_PigPathogenDataset (**Figure 3.18**). This was largely due to the

overrepresentation of arginine in M13 peptide library variants, which was particularly drastic in the in $VNN_{15}+(TTT)_1$ library.

Our analysis of the APD3_ AMP_PigPathogenDataset highlighted specific biases exist within the three AMP amino acid groupings. For basic Aa's, the APD3_ AMP_PigPathogenDataset highlighted a clear bias for K > R/H however, randomisation via $VNN_{15}+(TTT)_1$ and NNK schemes in fact resulted in an R > K/H bias in library peptides. $VNN_{15}+(TTT)_1$ libraries possess an increased cumulative frequency of hydrophobic amino acids compared to NNK libraries (**Figure 3.18**). The restrictions of the VNN scheme specifically on amino acids; W, Y and C thus inclines the hydrophobicity of peptides to be more likely derived from Aa's highly represented in the APD3_ AMP_PigPathogenDataset; such as glycine (G), alanine (A), isoleucine (I).

3.7 Discussion

3.7.1 In silico analysis of known AMPs for knowledge-based design of AMP biased peptide library.

The initial phase of this study led to the generation of a manually curated dataset of 187 characterised natural AMPs against several bacterial pathogens of pigs (APD3_ AMP_PigPathogenDataset). AMPs were obtained from a well-established database; specifically, APD3: the antimicrobial peptide database (Wang et al., 2016). The only restrictions placed upon the selection of AMPs for the APD3_ AMP_PigPathogenDataset concerned; amino acids (size; ≤50mers and no unnatural amino acids) and characterised antibacterial activity against pig pathogens (MIC: ≤100μM or ≤100μg/mL). Manual selection of APD3_ AMP_PigPathogenDataset in comparison to randomly selecting from the entire ADP3 database with no restrictions

i.e. APD3_RandomSelection187 dataset still resulted in highly akin peptides. Nonetheless, the challenges for in silico design of AMPs is reflected in the fact both datasets remain discreetly distinguishable from a set of non-AMPs with certain physiochemical properties; *e.g.* median charge and isoelectric points (Tan et al., 2021; Bhadra et al., 2018).

The complexity of AMPs as a class of peptides often results in their characteristics being loosely generalised. Rationalisation of these generalisations solely via amino acid composition alone cannot facilitate the prediction nor designed assimilation of membrane acting AMPs. Rather residue composition of AMPs must be placed within the context of physiochemical and structural features (Fjell et al., 2011). The antimicrobial activity of AMPs is affected by several factors related to structural integrity such as; charge, hydrophobicity, amphiphilicity, and polar/non-polar angle (Kim and Cha, 2010; Shiqi He et al., 2020). Clusterisation of amino acids in accordance with physicochemical features has been previously exploited for identifying AMPs against microbes (Kim and Cha, 2010;Vishnepolsky et al., 2018). However, the discreteness of the physicochemical profiles of AMPs can be hampered by insufficient datasets of verified AMPs and the variation of AMP-target cell envelopes and modes of action (Vishnepolsky et al., 2018; Pirtskhalava et al., 2021).

Nonetheless, in this study we constructed a simplified *in silico* analysis strategy to identify putative determinants of antimicrobial functionality via patterns of amino acid distribution. Specifically, the composition of naturally occurring amino acids were assessed by respective contributions to physiochemical properties; *i*) hydrophobicity, hydrophilicity, charge, polarity and contribution to amphiphilicity; *ii*) distribution within or proximal to hydrophobic stretches or positional biases across the peptide region. APD3_ AMP_PigPathogenDataset demonstrated characteristic AMP-

like features and were all cationic (+1 - +15) and highly hydrophobic (Tan et al., 2021). Peptides tend to retain a positive charge at pH below their isoelectric points, whereas a pH higher than the isoelectric point results in a negative charge (Tokmakov et al., 2021). Cationic AMPs in both datasets typically retained ≥7 isoelectric points, which is putatively advantageous for the maintenance of overall cationic charge in physiological pH conditions. Random selection from the ADP3 database did identify natural anionic AMPs which typically rely heavily on amphiphilicity and obliquely slanted α-helices conformations for antibacterial activity (Dennison *et al.,* 2006).

In context of the three AMP amino acid groupings designated in this study, the following highly abundant amino acids were identified and equated to 47.1% of the total amino acid occurrence in the APD3_ AMP_PigPathogenDataset; hydrophobic Aa's: Gly (11.2%), Leu(10.3%), Ala (8.4%), basic; Lys (12.8%) and hydrophilic; Ser (4.4%). Mishra and Wang (2012) and Pirtskhalava et al., (2021) similarly shortlisted Gly, Leu, Ala, Lys and Ser amino acids as the most preferred Aa's in bacterial AMPs. Glycine is a hydrophobic amino acid with an R group which is typically classified as non-polar. Gly-rich regions are common in natural AMPs and Gly repeats provide structural flexibly thus, exerting effects on tertiary structure formation *(e.g.* loop conformations) and lipid attachment (Stephan Tavares et al., 2012;Huan et al., 2020). Hydrophobic amino acids such as; Gly, Leu, Ile and Ala improve hydrophobic parameters of antimicrobial peptides, thus contributing to amphiphilic structures which partition bacterial membranes (Saint Jean et al., 2017). Ser amino acid residues can facilitate complex ring formations in AMPs, however, serine is particularly susceptible to protease action (Pirtskhalava et al., 2021). Lys was the most common basic amino acid residue followed by Arg then His, and all three are key for maintaining cationic charges. Vorobyov *et al.,* (2019) demonstrated charged Arg/Lys

residues deformed bilayers of DPPC lecithin via extraction of water molecules and lipid head groups, with Arg pulling in more lipids. Charged Arg forms more electrostatic interactions nonetheless these residues weakly interact with polar molecules in comparison to Lys (Li, Vorobyov et al., 2019). So, perhaps natural AMPs have favoured Lys>Arg inclusion on the merit of strength quality of bonding interactions rather than their mere quantity.

In the AMP_PigPathogenDataset, Gly, Leu and Ala were the most common Aa's in hydrophobic stretches of peptides with lengths between 11-20mers, and these peptides typically contained; 2-5 sequential stretches of ≥2 hydrophobic residues. Stretches were immediate or within the first 25% residues from the N-terminal, and AMPs analysed shared the propensity to be highly saturated with hydrophobic or basic residues at terminal ends. Cationic and hydrophobic rich-regions can dictate the adoption of amphipathic α-helix or β-sheet conformations, and thus vital for transmembrane interactions and pore formation in AMPs (Chen et al., 2020). Terminal saturation of hydrophobicity and charge is common in natural AMPs <20 amino acids (Zelezetsky and Tossi 2006; Tossi et al., 2000). This has fuelled the design of chimeric peptides constituting adjoined or bridged terminal ends of natural AMPs such as cecropin A and melittin (Shengyue Ji et al., 2014).

3.7.2 Semi-rational design of an AMP biased peptide library:

Analysis of the AMP_PigPathogenDataset corroborated certain trends in amino acid composition reported in the extant of literature. However, most importantly analysis of this dataset supported the downstream knowledge-based design of an "AMP" biased peptide library strategy. This study aimed to select an approach to construct naïve peptide libraries of DNA variant constructs with randomised 16-mer peptide regions (encoded in phagemids), and screen library peptide functionality

either; binding affinity for bacterial targets but primarily antimicrobial activity. Directed molecular evolution and combinatorial mutagenesis strategies, both chemical and biological among others, exist for protein engineering and the generation of randomised peptide libraries (Neylon, 2004).

TRIM (trimer/trinucleotide) technology is a redundancy-free mutagenesis strategy for the generation of completely randomised 20/20 libraries (Krumpe et al., 2007 Sieber et al., 2015) which primarily exploits equimolar (20/20 mutagenesis) or precise concentrations (biased mutagenesis) of triplet phosphoramidites codon mixtures, and removal of unwanted stop codons (Ono et al., 1995; Almagro et al., 2019). Fundamentally, TRIM technology provides unparalleled levels of control for both complete randomisation and selective amino acid codon saturation. Utilising TRIM technology in this study was explored, however this technology remains proprietary and an inaccessible strategy in research settings, and this underutilisation is derived from; *i*) the complexity of phosphoramidite production, and utilisation requires expensive DNA synthesiser equipment, or *ii*) poor triphosphoramidite yields from solid supports and/or *iii*) commercially offered TRIM-combinatorial libraries remain extremely expensive (Virnekäs et al., 1994; Lindenburg et al., 2020).

3.7.3 Selection of degeneration primer randomisation schemes

Degenerate primer schemes exist, with varying redundancy ratios (codon:amino acids), which restrict the genetic code in peptide library construction dependent on desired amino acid properties hydrophilicity; VRK (12:8), hydrophobicity; NYC (8:8), charge; RRK (8:7) (Kille et al., 2013). AMPs can contain all 20 naturally occurring amino acid and a singular primer scheme cannot completely eradicate genetic redundancy whilst concurrently encoding all 20 amino acids (Bozoviˇcar and Bratkoviˇc et al., 2019). Additionally, selection of singular or combinations degenerate

codon schemes such as VRK, NYC, RRK among others states in literature either; *i*) reduces or excludes highly represented amino acids in AMPs *i*) complicates oligonucleotide synthesis requirements and/or *iii*) necessities several or more primers for the randomised peptide library construction.

The nature of the NNK generate scheme still begets redundancy issues nonetheless, redundancy-free libraries can be constructed via strategies (MAX, ProMAX 22c-Trick) however, these techniques are restricted to mutagenesis of <5 contiguous amino acid residues. Therefore, aforementioned strategies are more apt for site-directed mutagenesis of specific amino acid sites of characterised AMP to improve AMP potency, reduce cytotoxic effects and/or conduct amino acid scanning studies. Depending on synthesis and molecular biology approach, degenerate primer mutagenesis can be a relatively inexpensive (Lindenburg et al., 2020), and conventionally NNK/NNS degenerate primers are highly relied upon to generate peptide libraries (Tonikian *et al.*, 2007; Sieber *et al.*, 2015). *E. coli* codon bias, specifically the asymmetric distribution of high G/T representation in the third base of codons increases the commercial preference for NNK rather than NNS in peptide library construction (Wan et al., 2004;Lindenburg et al., 2020).

Consequently, the degenerate primer strategy selected herein aimed to generate two libraries; *i*) NNK randomisation, and *ii*) $VNN_{15}+(TTT)_1$ randomisation with sixteen degenerate primers with successive 1/16 peptide region amino acid designated as phenylalanine (F). VNN scheme where "V" = A/G/C, "N" = A/T/G/C encodes 16 amino acids via 48 codons and by its nature therefore excludes; cysteine (C), phenylalanine (F), tryptophan (W), tyrosine (Y) and all three stop codons. Contrastingly, the NNK scheme, where K: G/T is less redundant and encodes all 20 natural amino acids and one stop codon (amber, TAG expected frequency: ~3.13%).

The retention of phenylalanine was prioritised due to its's 7th placed ranking in the most commonly identified amino acids in the AMP_PigPathogenDataset. Phenylalanine is a hydrophobic amino acid with aromatic side chains and phenylalanine residues in potent short sequence AMPs has been shown to facilitate membrane anchoring, as demonstrated in natural AMP aurein 1.2 (13 mers) (Saint Jean et al., 2017; Migon et al., 2019). Additionally, when present as the sole hydrophobic residue phenylalanine containing AMPs can bind with high affinity to bacterial lipid bilayers often leading to bacteriostatic consequence (Saint Jean et al., 2017; Migon et al., 2019). Analysis of the expected frequency for NNK and $VNN_{15}+(TTT)_1$ randomisation against the amino acid distribution of the AMP_PigPathogenDataset revealed both schemes were likely to underrepresent hydrophobic and basic Aa's, whereas the opposite would be observed for hydrophilic Aa's. In all three AMP amino acid groupings, the expected frequency theorised for the $VNN_{15}+(TTT)_1$ randomisation scheme shared a closer resemblance to the AMP_PigPathogenDataset, and this was especially apparent in the expected basic amino acid residue representation.

3.7.4 Cloning strategy

This study applied a modified cloning method generated large and diverse 16mer phage peptide libraries via a the whole-plasmid inverse PCR method. The cloning strategy was adapted from two approaches *i*) Tsoumpeli *et al.,* (2022) specifically the use of degenerate primers and their respective design, *ii)* and the simple whole-construct amplification strategy outlined by Kong *et al.,* (2020). DNA coding the peptide library diversity was introduced to the phagemid vector via degenerately randomised (NNK and $VNN_{15}+(TTT)_1$) forward primers. A 48/81bp region of forward primers was selected for randomisation to generate 16mer peptides adjoined to the N-

terminus region of the pIII protein to display peptides as pIII fusions. Inverse PCR amplification of phagemid vectors necessitates high specificity and sensitive DNA polymerases with efficient proof-reading to ensure low error rates.

As such a *Taq* DNA polymerase was selected for PCRs; *Taq* DNA polymerase possesses 5′→3′ polymerising and 5′-exonuclease activity with fidelity range over-ten fold; 3×10^{-4} to 3×10^{-6} errors per nucleotide polymerised (Abramson, 1995). Using a standard PCR master mix, a reaction constituted a volume of 50µL and >1µg linear PCR vectors per reaction was the typical yield observed herein similar to Kong *et al.,* (2020). Sixty 50 µl PCR replicates were utilised to generate sufficient starting DNA yields for NNK library production, whereas the higher primer load for the VNN_{15}+$(TTT)_1$ library construction required ≤ 10 PCR replicates per primer. >100µg of PCR product were extracted and flowed downstream to the *SapI/DpnI* double digestion. For both libraries, the reverse and forward library construction primers contained *SapI* RE sites, wherein *SapI* digestion facilitates ligating opposing ends of an engineered whole-construct PCR amplicon via intramolecular interactions (Tsoumpeli et al., 2022; Kong *et al.,* 2020).

Significant losses of PCR or double digestion library DNA was a bottleneck observed during clean-ups, wherein extraction often-underperformed manufacturers advertised ~60% retention values, for instance some of the VNN_{15}+$(TTT)_1$ clean-ups recovered 5-44% of PCR products. Nonetheless, to generate libraries $>10^9$ in size; ~10µg of engineered ligated library DNA is required for *E. coli* transformations as <1% of DNA is typically retained (Lindner et al., 2011) In this study, 18-20µg each of ligated DNA required to generate NNK and VNN_{15}+$(TTT)_1$ libraries, <20 electroporation of highly electrocompetent *E. coli* TG1 were conducted per library wherein ~1 x 10^7 and 1 x 10^8 cfu/µg of ligated library DNA were achieved

respectively. Large peptide libraries were generated, with titrations estimating library sizes were >1 x 10^9 and comparable to previous literature where NNK was utilised for randomisation (Tsoumpeli et al., 2022). Additionally, by the nature of the study aims both libraries are comparatively larger than AMP-designed mutant peptide libraries used by Tucker *et al.*, (2018) and Tominaga *et al.* (2006). VNN_{15}+$(TTT)_1$ and NNK pIII phage display libraries were propagated in *M13* phage for downstream-phage display application. *E. coli* TG1 glycerol stocks for both libraries were miniprepped and extracted library DNA subjected to *SpeI* digestion and subsequent ligation to form respective sub-libraries in *E. coli* JE5505.

3.7.5 *Quality control of naïve peptide libraries: Sanger sequencing*

Sanger sequencing utility in this study encapsulated *i*) prompt verification of library construction PCR template modifications and *ii*) selection of handpicked clones for the evaluation of library cloning success. A deep insight into the library quality and peptide diversity of randomised libraries cannot be garnered from sanger sequencing; especially as this sequencing technology suffers from low throughput and readout numbers (Linars et al., 2022). For instance, ~50% of handpicked library clones sent for sanger sequencing produced unusable poor sequence reads which limited the pool of clones analysed. Nonetheless, sufficient sanger sequencing data did loosely indicate; Phenylalanine was absent from the peptide region in ~22% of sufficient sanger sequenced VNN_{15}+$(TTT)_1$ library clones, and some scheme-excluded amino acids (C, W and Y) were present in the randomised peptide region. Additionally, across libraries/sub-libraries sanger sequencing suggested non-16mer peptide inserts and stop codons in the encoding peptide regions were present.

3.7.6 Deep sequencing naïve peptide libraries: next generation sequencing library quality control

NGS platforms are relatively low-cost and high-throughput parallel sequencing technologies and NGS remains heavily exploited to assess library quality before screening for ligands (Grada and Weinbrecht, 2013; Shave et al., 2018). NGS deep sequencing naïve peptide libraries was useful for assessing the success of the modified cloning strategy and identifying amino acid distribution of peptide variants. The parameters of the NGS perl script analysis pipeline omits the evaluation of low sequence quality reads and/or sequences with incomplete/unpaired flanking constant motifs. Random sampling of $>10^6$ NGS sequences from a single sample is a fraction of peptide libraries however, it is assumed representative for the entirety of the library (Shave et al., 2018). $>1 \times 10^6$ NGS sequences were analysed for $VNN_{15}+(TTT)_1$ and NNK libraries/sub-libraries in *E. coli* TG1, *E. coli* JE5505 and *M13* phage. For instance, $1.2\text{-}1.3 \times 10^6$ unique sequences were analysed for the $VNN_{15}+(TTT)_1$ pIII phagemid libraries.

For both $VNN_{15}+(TTT)_1$ and NNK libraries the number of barcoded sequences with the desired 16mer randomised peptide region was; $VNN_{15}+(TTT)_1$ pIII phagemid libraries ~89 - 92%, *gIII- E. coli* JE5505 sub-library 92.11% and NNK; pIII phagemid libraries ~91%, *gIII- E. coli* JE5505 sub-library ~92%. Across these pIII phagemid datasets 91.7- 94.8% of $> 1 \times 10^6$ sequences were singletons, which is similar or slightly less than previously reported for peptide libraries of a similar ethos (Kong et al., 2020; Tsoumplei et al., 2022). Consequently, <8.3% of sequences were identified as repeats across phagemid libraries, which is comparable to previous reports on phage display libraries (Kong et al., 2020; Tsoumplei et al., 2022). More specifically; $VNN_{15}+(TTT)_1$ pIII phagemid libraries observed 5.2% and 7.5% unique sequences as

repeats for *E. coli* TG1 and M13 libraries respectively. This was incrementally higher for NNK pIII phagemid libraries; 8.3% (*E. coli* TG1 library) and 6.4% (M13 library) attributed to repeat sequences.

As the entire 15-16 mer peptide region will be randomised via both schemes, this means with limited restriction of encoding amino acid a complex matrix of positional combinations of amino acids are possible. Lindner *et al.,* (2011) outlined when 15mer or 16mer peptide regions are randomised this correlates to 3×10^{19} and 6×10^{20} maximal peptide library diversity. Rebollo *et al.,* (2014) methodology of extrapolating library diversity was applied, wherein the absolute number of different peptide sequences are derived as the proportion of the titrated library with peptides of the correct insert size. Consequently, the library diversity observed was 1.83×10^9 and 2.12×10^9 respectively for $VNN_{15}+(TTT)_1$ and NNK libraries transformed into *E. coli* TG1. Considering the titres achieved for both *E. coli* JE5505 sub-libraries, library diversity was estimated as 9.21×10^6 and 4.60×10^6 for $VNN_{15}+(TTT)_1$ and NNK sub-libraries respectively.

Library construction in this study utilised high-performance liquid chromatography (HPLC) purified degenerate oligonucleotide primers, with the forward primers for both schemes equating to 81bp (27 mers). The application of advanced techniques of anion-exchange or reversed phase HPLC are common for oligonucleotide purification, and each technique is suitable for separating <30mer and <20 oligonucleotides respectively (Gilar and Bouvier, 2000; Qiulong Zhang et al., 2016). HPLC purification varies between 95-98% and the high purity of primers used herein is demonstrated across the *E. coli* TG1 libraries as by <9% of barcoded sequences analysed, with correct flanking regions/>1Aa insert were deemed non-16mer peptide inserts (Gilar and Bouvier, 2000).

Nonetheless, as library forward degenerate oligonucleotide primers exceed the upper limits of suitable separation for HPLC methods, co-elution of similarly size oligonucleotides with undesired peptide lengths could occur. Additionally, in PCR reactions oligonucleotide primers can yield various PCR amplicon sizes such as; self-ligated vectors or large multimeric amplicons consisting of several peptide inserts, vectors adjoined or an amalgamation of both inefficiencies (Potapov and Ong, 2017). To limit non-specific amplification, we conducted annealing gradient PCR to identify the highest annealing temperature with sufficient PCR amplicon DNA yields however, the degeneracy of randomised primer regions does influence GC% content, primer T_m which can influence annealing temperature ranges. Varying library peptide inserts lengths were replicated in the VNN_{15}+$(TTT)_1$ and NNK *E. coli* JE5505 sub-libraries however the range extended to 1-52 amino acids in the peptide region. Possibly, indicating the recombination of *SpeI* digested library DNA possessed a greater propensity to facilitate intramolecular interactions of sticky ends of varying vectors.

3.7.7 Quality control via next-generation sequence analysis of naïve peptide libraries: amino acid expected frequency vs observed frequency

The results from deep sequencing naïve peptide libraries via the NGS analysis contrasted sanger sequencing data in two-fronts; *i*) putative phenylalanine representation across peptide region of the VNN_{15}+$(TTT)_1$ randomised library and *ii*) stop codon representation across both schemes. The high propensity for stop codons in sufficiently sanger sequenced clones likely reflect poor chromatogram baseline distinguishment of A/T/G/C nucleotides which is typical as the sequence progress further away from the annealing primer initiation site (Grada and Weinbrecht, 2013; Crossley et al., 2020). Sanger sequencing of random handpicked clones identified

phenylalanine as absent in a significant portion of peptide regions analysed for $VNN_{15}+(TTT)_1$ *E. coli* TG1 and JE5505. NGS data revealed phenylalanine was slightly overrepresented in the *E. coli* TG1 library.

$VNN_{15}+(TTT)_1$ *E. coli* TG1 library phenylalanine (F) slight overrepresentation is in line with the inclusion of cysteine, tyrosine and stop codons. All amino acids/stops encoded via codons consisting with a Thymine 'T' inclusion within the first nucleotide position were slightly overrepresented (0.20-0.44%). Thymine 'T' inclusion within the first nucleotide position of a codon (except for the reintroduction of phenylalanine) was not completely omitted in degenerate primers as theoretical designed for the VNN15+(TTT)1 scheme. The overrepresentation of cysteine and tyrosine is almost double that of stop codons (TAA, TGA, TAG) and tryptophan (TGG). This might indicate the incorrect Thymine 'T' inclusion on the first codon position was additionally biased towards being coupled with C/T nucleotides in the third codon position thus giving rise to slightly more cysteine (TGC) and tryptophan (TAC, TAT) residues and phenylalanine (TTT).

In concern of the $VNN_{15}+(TTT)_1$ naïve peptide libraries (*E. coli* TG1, *E. coli* JE5505, *M13* phage), the observed frequency of the twenty natural amino acids and stop codons (TAG, TGA, TAA) deviated by ~ ± 3% from the theoretical expected frequency. The amino acids overrepresented in the $VNN_{15}+(TTT)_1$ *E. coli* TG1 library were; V, G, E and those underrepresented were P, R, H, T, similar representations for these amino acids were reflected in the *E. coli* JE5505 equivalent sub-library except the underrepresentation of R and H. Contrastingly, the NNK naïve peptide libraries similarly in *E. coli* TG1, JE5505 and M13 phage, the observed frequency of amino acids and stops deviated by ~ ± 2% from the theoretical expected frequency. This

range of deviation from expected is only >1% out from Tsoumpeli *et al.,* (2022) NNK-16mer pVIII peptide library.

Again, similar to the VNN*15*(TTT)*1 E. coli* peptide libraries, vestige within two different *E. coli* strains resulted in limited wavering from the theorised levels of expected frequency for NNK libraries, nonetheless patterns of representation for certain amino acids did differ across both. The amino acids overrepresented in the NNK *E. coli* TG1 library were; K, W, M and amber stops and those underrepresented included A, P, H, and T. Across libraries generated via either degenerate codon scheme, amino acids P, T and H remained underrepresented across *E. coli* TG1 and JE5505 library. The only exception being histidine observant frequency being similar to expected in VNN_{15}+$(TTT)_1$ *E. coli* JE5505 sub-library. Nonetheless, the consistent underrepresentation of P/T/H across libraries indicates to specific codon usage and/or *E.coli* derived biological censorship, which has been extensively explored for peptide libraries (Zanconato et al., 2011).

In this study, R and amber stops were overrepresented, and P/H/A/T were both underrepresented across both VNN_{15}+$(TTT)_1$ and NNK *M13* phage libraries. This resulted in a significant shift in the amino acid distribution observed in *M13* phage equivalent VNN_{15}+$(TTT)_1$ and NNK libraries, specifically the overall representation of Aa's under the AMP amino acid basic and hydrophobic groups outlined herein. Arguably propagation in phage seems to favour the selection of peptides which more closely resemble the necessary charge and hydrophobicity to facilitate efficient phage pIII production and bacterial binding. The censorship of *M13* phage on short peptide libraries randomised utilising the NNK scheme has been identified in literature (Krumpe et al., 2007). The observed amino acid distribution across *M13* phage propagated libraries, with no other forms of selection pressure (*i.e.* biopanning affinity

selection), likely indicates *E. coli* amino acid codon usage bias as well as biased production selecting phage bearing recombinant pIII/library peptide with particular amino acid content.

Stop codons in the encoding peptide region were observed across libraries and sub-libraries; NNK (TAG) and $VNN_{15}+(TTT)_1$ (TAG/TGA) from sanger sequencing handpicked clones. NGS data of $VNN_{15}+(TTT)_1$ scheme derived libraries revealed all three stop codons were prevalent in the *E. coli* TG1 and M13 phage accounting for 0.66% and 1.12% of codons in the peptide region respectively. Contrastingly, opal and ochre stops were limited as expected in the NNK libraries, with the NGS data revealing amber stops constituted 4.11% and 5.02% of codons in the peptide regions of the NNK scheme randomised libraries in *E. coli* TG1 and M13 phage.

Kretz *et al.,* (2018) constructed NNK randomised 6-mer peptide phage display libraries (*M13* display, library size: 2.3×10^8) to conduct active site scanning studies on coagulation serine protease thrombin and metalloprotease ADAMTS13 (proteolytically regulates platelet-binding of von Willebrand factor). Kretz *et al.,* (2018) observed against expected frequency a -3 to +2.5 deviation for all twenty amino acids and stop codons, slightly higher deviations than reported herein, with the most overrepresented Aa's including; G, V, E, W, D, R and amber stops whereas underrepresented Aa's were; I, F, K, N, P, Y (Kretz *et al.*, 2018). 0.04% of Kretz *et al.,* (2018) sequencing reads (10^6) contained stop codons, wherein; amber stops were overrepresented compared to expected, TAG (91.8% of stops), however residual retention of ochre and opal stops were observed, TAA (1.7% of stops) TGA (6.5% of stops) respectively.

Screening strategies utilised herein such as phage display or Guralp et al., (2013) overlay assay rely in some capacity of *E. coli* strains. Phage-display systems overcome the representation of amber stops via the use of amber supressing *E. coli* strains (Oh et al., 2007). *E. coli* amber suppressors insert an amino acid upon recognition of the nonsense stop codon mutation (Normanly et al., 1986) for instance, the highly electrocompetent (≥1 x 107 cfu/μg efficiency) *E. coli* TG1 utilised herein contains the *SupE* gene and thus replaces TAG stops with glutamine (Q) (Normanly et al., 1986;Lessard and Perham, 1993; Oh et al., 2007). The inclusion of stop codons can limit display in phage systems, however in pVIII (type 8) systems this is counter-balanced by pVIII high copy number leading to hundreds of recombinant molecules on phage particles (Tikunova and Morozova, 2009). In comparison, pIII (type 3) systems are highly monovalent as wildtype pIII molecules are retained on phage particles however, type 3 systems are preferred as it can carry peptide <50mers without reduced infectivity (Lindner et al., 2011). Nonetheless, during infection a small fraction of phage carry 1-5 copies are recombinant pIII molecules on the particle surfaces (Tikunova and Morozova, 2009; Mi-young Oh et al., 2007). Therefore, non-amber stops (TGA/TAA) within the peptide region could produce truncated peptide variants and reduce recombinant molecule display in the pIII system utilised herein (Kretz et al., 2018). Strategies exist to improve pIII wt to recombinant molecule ratios, for instance Baek et al., (2002) describes the use of gIII_amberstop Ex-phage mutants wherein wt pIII inclusion is inhibited by the presence of amber stops in the encoding gIII however, this can lead to issues with phage propagation (*i.e.* generates no-pIII phage) and low titres achieved (Mi-young Oh et al., 2007). Nonetheless, the pIII system low valence advantageously offers the selection of high affinity ligands, which

is highly desirable for this study's aim to isolate library peptide variants which bind to bacterial pathogens of pigs (Sidhu *et al.,* 2005; Tikunova and Morozova, 2009).

The NGS data revealed the naïve *M13* pIII peptide phage library wherein the peptide region was randomised via $VNN_{15}+(TTT)_1$ scheme resulted in the closet representation of the three key AMP amino acid groups herein. However, as aforementioned amino acid composition cannot solely determine antimicrobial functionality of peptides. At this stage of the study implementation, the analysis of library peptide antimicrobial functionality cannot be derived although, it can be modelled and assimilated via largely under evaluated AMP-prediction tools with varying models (Tucker *et al.,* 2018). Nonetheless, to evaluate both $VNN_{15}+(TTT)_1$ and NNK libraries to identify peptide variants with desired functionality, high-throughput screening is necessary and the endeavours to achieve this are explored further in the remainder of this study.

Chapter 4: A leaky periplasm *E. coli* JE5505 expression system and agar-based overlay assay method for screening AMPs

4.1 Introduction

Small AMPs are increasingly becoming highly sought-after therapeutics, and wet-lab techniques persist as fundamental pillars of AMP novel drug discovery. Several distinct approaches have exemplified novel AMP identification; SPOT-synthesis and cellulose tethered AMP arrays (López-Pérez et al., 2015), AMP display libraries; phage and/or bacterial (Sainath Rao et al., 2013; Tucker et al., 2018), colorimetric broth assays (Allen et al., 2022) and/or agar assays. The utilisation of solid medium screening strategies have identified numerous AMPs. For example, the well-characterised LL-37 human neutrophil cathelin-associated AMP was originally isolated via agarose radial diffusion assays (Turner et al., 1998). Miller *et al.,* (1998) were initial innovators of analysing a characterised AMP via an overlay soft agar assay approach. Miller *et al.,* (1998) demonstrated efficient recombinant production of chimeras of bacteriocin AMP Pediocin AcH (identical sequence to pediocin PA-1) from *E. coli* E609L (*lpp::Tn10;* periplasmic leaky; *Tcr)* colonies. The formation of zones of inhibition in 0.8% soft tyrptone glucose extract overlays were observed against the *L. innocua* test species (Miller et al., 1998).

More recently, Tominaga and Hatakeyama, (2006) and Guralp *et al.,* (2013) overlay assay approaches exemplified high-throughput AMP screening using the *E. coli* JE5505 (F− *1po his proA argE thi gal lac xyl mtl tsx*) *Δlpp-254* mutant expression system. Guralp *et al.,* (2013) designed a plantaricin-423 (Pln-423, 37mers) mutant peptide library wherein phosphoramidite synthesised library oligonucleotides were amplified into the *pFLAG-CTS* plasmid vector via an emulsion PCR approach. Wild-type Pln-423 (37mers) is a class IIa natural AMP bacteriocin isolated from *Lactobacillus plantarum* 423, and this small heat-stable AMP is bactericidal against several Gram-positive bacteria including *Listeria spp* (Van Reenen et al., 2003). A

leaky *E. coli* JE5505 expression system within an agar assay was used to screen the antimicrobial activity of the constructed *pFLAG-CTS_Pln-423* mutant library. ~1.0 x 10^5 colonies were screened within five successive attempts which achieved ~8-fold library coverage (Guralp *et al.,* 2013). Overlay assay screening with *L. innocua* 33090 as the test species, identified improved mutant Pln-423 variants. Crude peptides of Pln-H28R/H34E and Pln-S27D/K36N library mutants observed a 0.037μM MIC against *L. innocua* 33090, almost halving the MIC value and increasing observed zone diameters by +4.3 to 4.8mm wider than respective Pln-423 wt expression (Guralp *et al.,* 2013).

The direct evaluation of library variants antimicrobial phenotypes against a test species, albeit within agar conditions, is the primary benefit and appeal of Miller et al., (1998) and later adopted overlay assay approaches. Leaky *E. coli* JE5055 expression system overlays putatively offer the potential to generate quantitative data, *e.g.* the continuous variable of zone of inhibition diameters, and qualitative information on the presence or absence of antimicrobial activity within this specific expression system. Guralp *et al.,* (2013) approach demonstrates a succinct AMP discovery pipeline however, similar to Tominaga and Hatakeyama, (2006), the leaky *E. coli* JE5505 overlay assay employed in both contexts focused on specific characterised bacteriocin AMPs Pln-423 or pediocin PA-1 respectively. Inarguably, overlay assays have proved useful in high-throughput scanning studies for improved AMP mutant identification. Nevertheless, with the growing need for novel antibacterial therapeutics, the present chapter seeks to evaluate whether antibacterial AMPs can be isolated from randomised peptide libraries wherein characterised antimicrobial activity was not a prerequisite of deploying overlay assay screening.

4.2 Result Chapter 2 Aims;

i. Utilise modified overlay assay method (*ii*) and other broth or agar characterisation assays to demonstrate the antimicrobial activity against bacterial pathogens of pigs using phage display isolated AMPs or natural AMPs.

ii. Evaluate the efficacy of Guralp *et al.*, (2013) leaky *E. coli* JE5505 overlay methodology with the novel application of a *pSD3* phagemid (*-gIII* absent) vector system guiding the recombinant peptide expression.

iii. Optimise a modified leaky *E. coli JE5505* overlay assay agar methodology for screening large degenerately randomised peptide libraries against both Gram-negative and Gram-positive bacteria test species.

iv. Utilise optimised overlay assay strategy (*iii*) to screen peptide sub-libraries randomised via "conventional" NNK or "AMP biased" $VNN_{15}(TTT)_1$ schemes against a Gram-negative and Gram-positive model bacterial pathogen of pigs; *S. Typhimurium 4/74* and *S. suis P1/7* respectively.

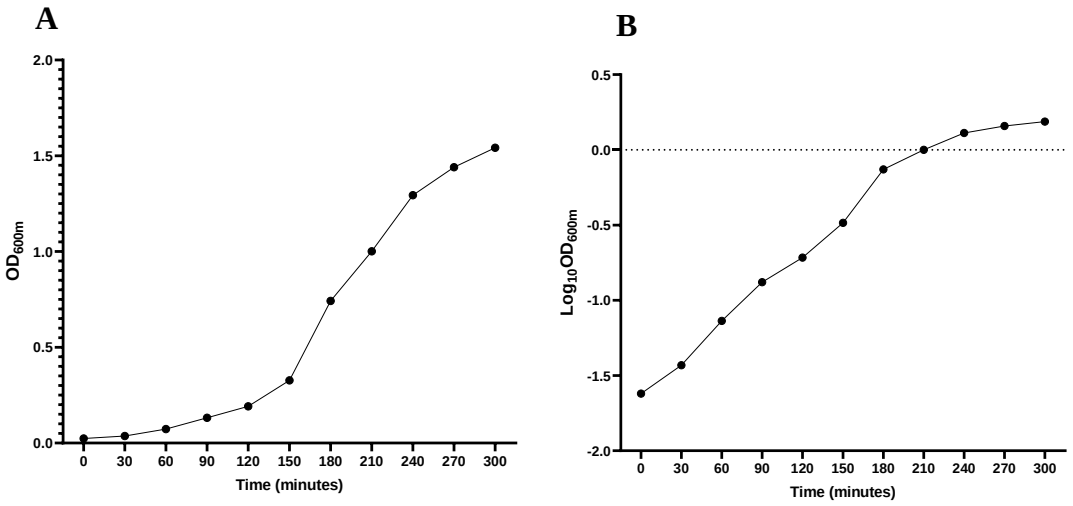
4.3 – Optimising a novel leaky periplasm *E. coli* JE5505 expression system within soft agar overlay assays for screening AMPs

4.3.1 Microbiological cultivation and phagemid transformation of E. coli JE5505

As depicted in **Figure 4.1**, a standard growth curve was produced for *E. coli* JE5505 **(Method 2.2.4),** to precisely pinpoint growth kinetics at optical density 600nm (OD_{600nm}). This supported empirically enumerating *E. coli* JE5505 for use in the overlay assay manipulation (Guralp *et al.*, 2013). On average *E.coli* JE5505 overnight dilutions starting from OD_{600nm} of ~0.025, entered the exponential phase of growth +150mins (2hrs 30mins) (**Figure 4.1**). *E. coli* JE5505 from this point displayed relatively normal growth and multiplication kinetics, similar to that reported in literature (Hirota et al., 1977; Lakshmikanth et al., 2017).

Figure 4.1 Growth curve of wildtype *E. coli* JE5505 (*lpo*-) mutant grown in 2xYT broth, 200rpm at 37ºC.

2xYT media was utilised for the cultivation of three individual *E. coli* JE5505 colonies, and this graph presents the raw optical density at 600nm (OD_{600nm}) values **(A)** as well as the average $Log_{10}OD_{600nm}$ **(B)** observed across 300 mins (5hrs). The average colony forming units at +150mins was; $OD_{600nm:}$ ~0.238 = ~1.92 x 10^7 cfu/mL. ~0.238 OD_{600nm} is a key optical density hallmark of the initiation of exponential growth for *E.coli* JE5505 cultures. This brief period exponential growth is maintained for 30 minutes where *E. coli* JE5505 reaches OD_{600nm} = 0.749 whereupon growth begins to plateau ~ +240 mins.



4.3.2 ***Characterisation of Pln-423wt and mutant variants using a Luria-Bertani (LB) and Tryptone -Soya Broth soft agar E. coli JE5505 overlay assay method***

Inverse PCR primers were designed as summarised in **Method 2.6** for the insertion of Pln-423 wt, and two mutants; Pln-H28R/H34E and Pln-S27D/K36N, into our template *pSD3_RLL_-BspQI_+SpeI_opal stop*. Inversely orientated primers would anneal to a *pSD3* regions which would exclude the *gIII* gene from amplicons. The mis-annealing region of the forward and reverse primers would contain the desired Pln-423 peptide sequence. Pln-423 and related mutants were supplanted into the peptide region generating vectors; *pSD3_Pln423_-BspQI_-gIII* vectors encoding; Pln-423 wt, Pln-H28R/H34E and Pln-S27D/K36N. The agarose gel electrophoresis conducted to confirm correct PCR amplification of Pln-423 wt and respective mutants into the *pSD3* vector is summarised in **Figure 4.2.** Whereas **Table 4.1** shows the handpicked clones sanger sequenced from the transformation of *E. coli* JE5505 with ligated PCR products from **Figure 4.2**.

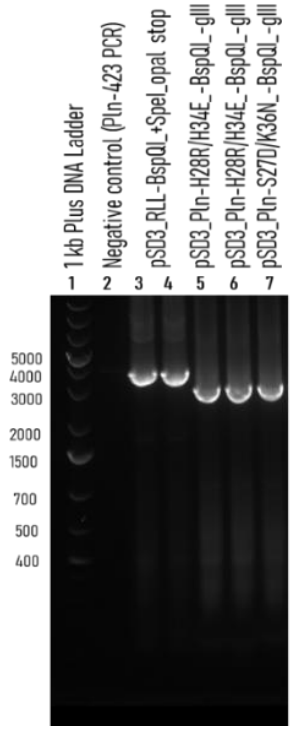


Figure 4.2. Verifying PCR modifications used to generate pSD3_Pln423_-BspQI_-gIII and respective improved mutants by agarose gel electrophoresis visualisation

Lane 3/4: *pSD3_RLL-BspQI_+SpeI_opal stop* library template, containing gIII. PCR amplicons were generated from a separate PCR outlined in Method 2.6. **Lane 5-7:** PCR amplicons of *pSD3_PlnX_-BspQI_-gIII* wherein X equates to Pln-423, Pln-H28R/H34E and Pln-S27D/K36N. (1kb Plus DNA ladder, Thermo Fisher). Degenerate primers were designed in manner where primer annealing omitted the inclusion of *gIII* in amplified vectors. **Lane 2:** Negative control = DNA template, no reverse primer (for Pln-423/variant inverse PCR).

Table 4.1 Sanger sequencing of handpicked *E. coli* JE5505 carrying pSD3_Pln423 and improved mutants

Handpicked *E. coli* JE5505 carrying pSD3_Pln423 and improved mutants			
Exemplar randomly picked clone	**Encodes**	**read length (bp)**	**Sanger sequence read translated**
1	Pln-423 wt	264	XXXPQFYFKETVIMKYLLPTAAAGLLLLAAQPAMA**KYYGNGVTCGKHSCSVNWGQAFSCSVSHLANFGHGKC**-SCQFFWLANNYARG
1	Pln- H28R/H34E	263	XXXPQFYFKETVIMKYLLPTAAAGLLLLAAQPAMA**KYYGNGVTCGKHSCSVNWGQAFSCSVSRLANFGEGKC**-SCQFFWLANNYARG
1	Pln- S27D/K36N	266	XXXAAILFXETVIMKYLLPTAAAGLLLLAAQPAMA**KYYGNGVTCGKHSCSVNWGQAFSCSVDHLANFGHGNC**-SCQFFWLANNYARG
Five randomly selected clones were sanger sequenced for each *E. coli* JE5505 carrying *pSD3* vectors_Pln423; H28R/H34E; S27D/K36N. No PCR templates were identified. Single letter amino acid code utilised, 'X' refers to an unquantified amino acid codon. '-' refers to an opal stop (TAA) Amino acid sequences; *PelB* leader (22Aa); MKYLLPTAAAGLLLLAAQPAMA Aa position 1 10 20 30 37 Pln-423 wt; KYYGNGVTCGKHSCSVNWGQAFSCSVSHLANFGHGKC Pln-H28R/H34E; KYYGNGVTCGKHSCSVNWGQAFSCSVSRLANFGEGKC Pln-S27D/K36N; KYYGNGVTCGKHSCSVNWGQAFSCSVDHLANFGHGNC *Note:* Where Aa = Pln-423 wt occurring Aa and 'Aa' = amino acid substitution mutation			

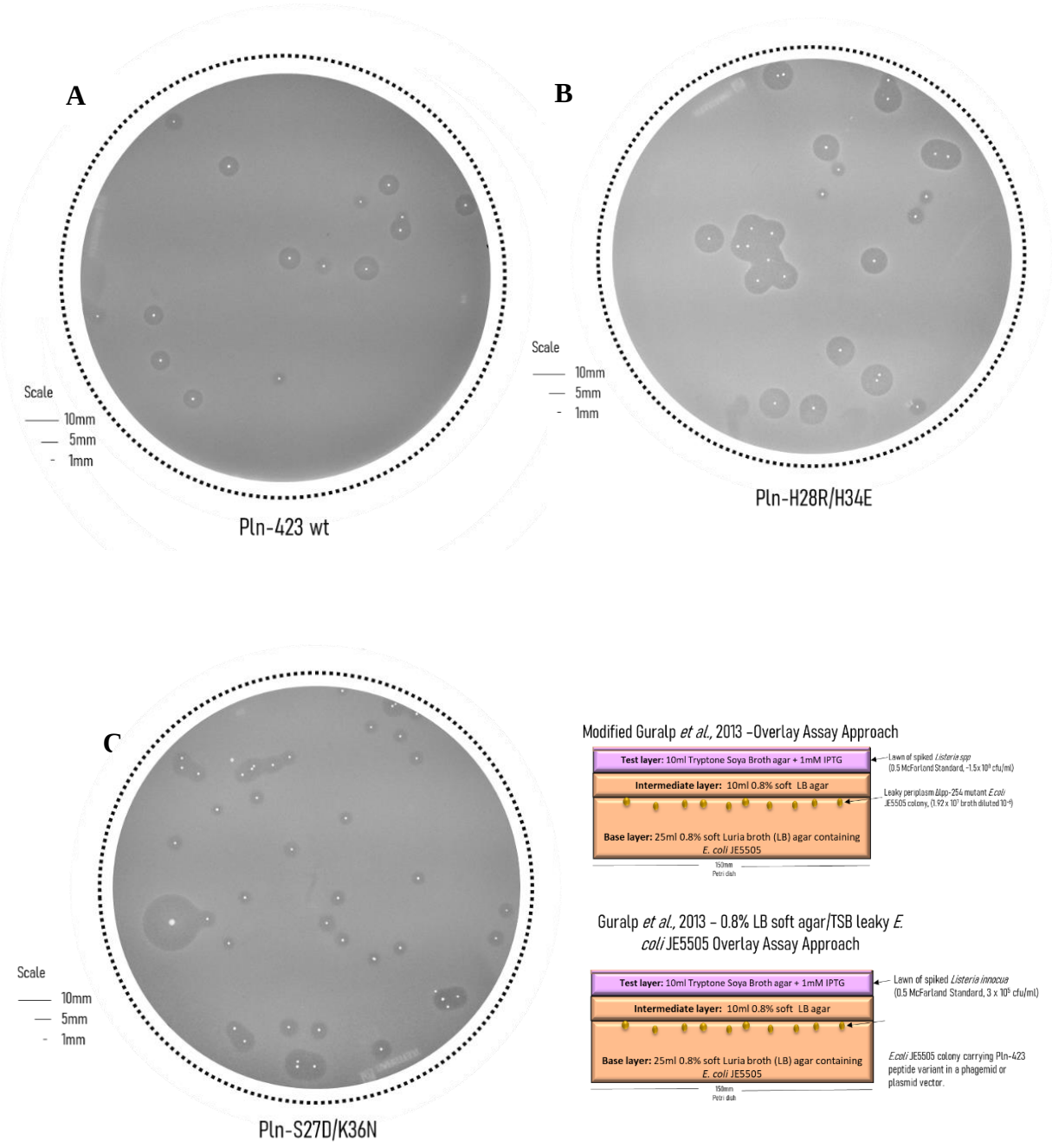
Pln-423 wt peptide sequence was synthesised in crude and tested against *Listeria innocua* in a microdilution MIC as described in **Method 2.5.2.** The MIC value obtained for Pln-423 wt was 0.097µM against *L. innocua* which corresponds with the sub-micromolar range reported in literature (Guralp *et al.,* 2013; Van Reenen et al., 2003). Following this, Guralp *et al.,* (2013) overlay assay method was replicated with only a few exceptions in terms of overnight broth culturing and optical density focal points empirically determined from *E. coli* JE5505 growth curve analysis (**Method 2.5.10).** Both agar and broth media composition were upheld; 0.8% soft Luria-Bertani (LB) agar for *E. coli* JE5505 layers and 0.8% Tryptone -Soya Broth (TSB) agar for the *Listeria* test species lawn layer; either *L. innocua* or *L. monocytogenes.*

E. coli JE5505 overnights were typically diluted to OD_{600nm} ~0.025 in fresh pre-warmed media and gently agitated (200-220rpm) to mid-log phase (OD_{600nm}: 0.6-

0.65). Mid-log phase *E. coli* JE5505 cultures were then diluted to OD_{600nm} ~0.238 (1.92 x 10^7 cfu/mL). 1:10 dilutions series were titrated to enumerate the number of colonies produced when specific volumes of OD_{600nm} ~0.238 *E. coli* JE5505 culture was spread plated and dried on 25mL of LB agar in large (150mm) overlay bioassay plates (**Figure 4.3**). Typically, <40 colonies were produced per overlay assay using this approach within a 3-layered 0.8% LB/TSB. The visualisation of zones of clearing surrounding *E. coli* JE5505 colonies carrying respective *pSD3_Pln423-BspQI,-gIII* or improved Pln-423 mutant vectors against *L. innocua i*n the 0.8% LB/TSB overlay are represented in **Figure 4.3.** To verify activity as derived from the expression of the *pSD3* vector, a wildtype *E. coli* JE5505 mutant with no vector was handled identically to test samples against the two aforementioned *Listeria* species and no zones of clearing were observed. This is a recommended step to ensure *E. coli* JE5505 wt does not encode any antimicrobial substances against screened test species.

Figure 4.3 *E. coli* JE5505 colonies carrying pSD3_Pln423-BspQI,-gIII vectors tested against *L. innucoa* strain in a modified Gurlap *et al.*, (2013) LB and TSB soft agar overlay assay.

0.8% LB/TSB soft agar overlays (150mm petri dishes) were conducted to assess the antimicrobial activity of vector *(pSD3_PlnX-BspQI,-gIII*) encoded expression of Pln-423wt/mutants; **(A)** Pln-423 wt**, (B)** Pln-H28R/H34E and **(C)** Pln-S27D/K36N from *E. coli* JE5505 colonies. As shown in **(D)** the test species in all three overlays was *L. innocua* spiked at 0.5McFarland (1.5×10^8) and 1mM IPTG spiked into this layer. Zones were visualised for A, B, C with all colonies producing zones of inhibition.



Optimising a modification of Guralp *et al.,* (2013), bar a few minor exceptions in broth culturing parameters and plated colony forming units validated the capacity of the *E. coli* JE5505 expression system for the production of recombinant peptides with antimicrobial bioactivity **(Figure 4.3).** The following average zone of inhibition diameters were observed by randomly selecting five *E. coli* JE5505 colonies per assay from triplicate plates; ~5.6mm (Pln-423 wt), ~7.2mm (Pln-H28R/H34E), ~2.9mm (Pln-S27D/K36N) (**Figure 4.3**). Under the results obtained via the modified 0.8% TSB/LB overlay **(Figure 4.3)**, Pln-S27D/K36N would likely not be identified as an improved mutant. Initially, this seems to contrast Guralp *et al.,* (2013), who demonstrated improved mutants generate larger zones of inhibition than the *pFLAG*-encoded Pln-423 wt.

Figure 4.3. demonstrates zones of inhibition generated by colonies of identical vector heritage (*i.e. pSD3_ Pln-S27D/K36N -BspQI,-gII*) can vary. Literature reported results by Guralp *et al.,* (2013) likely represent images from colony clones with the largest zones. As described previously, *E. coli JE5505* colony diameter typically ranged between 1 to 2mm, and often the larger colonies or the close proximity of two or more colonies together produced larger zones (**Figure 4.3**). The generation of different colony sizes might indicate some phenotypic heterogeneity of cell fitness in the original *E. coli* JE550 strain utilised. However, more likely larger colonies originate from cells of the similar fitness, but greater colony mass accumulation occurs due to positional advantages within the agar matrices or early formation on agar medium. Broth cultured *E. coli* JE5505 cells might exhibit varying growth phases at the point of and following agar adaptation as well as be prone to heat sensitivity from the ~45^{o}C soft agar decanting temperature.

4.3.3 ***Optimisation of a leaky E. coli JE5505 pSD3 expression system in a Mueller-Hinton soft agar overlay assay.***

4.3.3.1 *Mueller-Hinton soft agar overlay assay method for AMP scanning studies or the assessment of AMP compatibility with the E. coli JE5505 expression system.*

Figure 4.3, exemplifies the use of 0.8% LB and TSB soft agars to support the growth of *E. coli* JE5505 and *Listeria* spp respectively in overlay assays. LB broth and agar are widely utilised in research settings for *E. coli* recombinant production, whereas TSB media is the prototypical medium used to support the maximum biomass generation of *Listeria* species (Hany *et al.*, 1993; Sang-Hyun Park et al., 2014; Rosano and Ceccarelli, 2014). Nonetheless, in spite of the nutrient richness of LB media, *E. coli* cell growth arrests at relatively low densities due to scarce availability of divalent positive ions, media acidification and/or utilisable carbon sources (*i.e.* carbohydrates) (Rosano and Ceccarelli, 2014). Consequently, richer media than LB/TSB were explored to further optimise an overlay assay method to improve *E. coli* JE5505 cultivation and broaden the scope of well-supported test species applicable for testing.

Richer media than LB can be exploited for improved recombinant peptide production and cell yields, as larger quantities of yeast extract or peptones/tryptone can lead to higher *E. coli* cell densities (Rosano and Ceccarelli, 2014). In comparison to LB agar, 2xYT is particularly rich containing twice the concentration of yeast extracts and can maintain high-density cells cultivations for long periods(Kram and Finkel, 2015). Additionally, 2xYT is the source media for *M13* phage propagation (Kram and Finkel, 2015). Mueller-Hinton is a starch-rich media containing dehydrated beef infusions and hydrolysed casein and this media cultivates of a broad spectrum of bacteria (fastidious and non-fastidious) and fungal species (Nizeta et al., 2017). The significant majority of gold-standard antimicrobial testing methods for

both broth and agar use Mueller-Hinton. For disk-diffusion and antibiotic testing, Mueller-Hinton can often be supplemented with Ca^{2+} and magnesium (Mg^{2+}) cations (Nizeta et al., 2017). Both 2xYT and Mueller-Hinton media are non-selective similar to LB/TSB agar compositions.

The overlay assay culturing, enumerations and plating optimisation previously described were then replicated for both 2xYT and MHB media to verify whether agar composition impacted E. *coli* JE5505 colony formation, zone of inhibition diameters and/or and ease of visualisation in overlays **(Table 4.2).** This was specifically tested with *pSD3_Pln423-BspQI,-gIII* vector carrying *E. coli* JE5505 against *L. innocua* **(Table 4.2).** 1.0% 2xYT or Mueller-Hinton agar as well as 0.7% soft agar equivalents for both were tested. The overlay method was then completed with top layer containing no spiked test species for the ease of colony observation. **Table 4.2.** demonstrates the growth of *E. coli* JE5505 colonies from respective broth and agar plating parameters, when completing a 10^{-6} dilution of OD_{600nm} = 0.238 (~1.92 x 10^7 cfu/mL) broths.

Table 4.2 2xYT and Mueller-Hinton broth cultivation of *E. coli* JE5505 carrying a *pSD3 vector*

2xYT and Mueller-Hinton Broth culturing and plating *E. coli* JE5505 pSD3_Pln423-BspQI,-gIII				
		Number of colonies per across triplicate overlay assays		
Overlay agar media	Plated overlay agar antibiotic supplement	*E. coli* JE5505 broth ~0.238 (OD_{600nm}) dilution plated	Neat broth culture (no supplementation)	Broth culture supplemented with $CARB^{100}$ and 1% glucose
0.7% 2xYT	$Carb_{100}$	10^{-6}	0	0
2xYT	$Carb_{100}$	10^{-6}	41	31
0.7% MHA	$Carb_{100}$	10^{-6}	72	35
MHA	$Carb_{100}$	10^{-6}	62	37
0.7% 2xYT	No antibiotic supplement	10^{-6}	0	0
2xYT agar	No antibiotic supplement	10^{-6}	50	35
0.7% MHA	No antibiotic supplement	10^{-6}	85	31
MHA	No antibiotic supplement	10^{-6}	78	39
E. coli JE5505 were cultivated in broth conditions at 37°C, 200rpm until mid-log phase and then diluted to OD_{600nm} ~0.238 (1.92 x 10^7 cfu/mL). 350 µl of each dilution was plated onto large 150mm petri dishes, air dried and then overlaid. MHA = Mueller Hinton Agar, $Carb_{100}$ defines the working concentration of carbenicillin within media, specifically 100µg/mL				

In comparison to LB and 2xYT, *E. coli* JE5505 demonstrated the best growth when cultured with and plated onto Mueller-Hinton media. **Figure 4.4** depicts the schematic for the final optimised Mueller-Hinton soft agar leaky *E. coli* JE5505 overlay assay. The selection of Mueller-Hinton was additionally supported by the relevancy of this media in downstream AMP candidate characterisation.

The design of the overlay assay agar layers was the next focus, and the triplet layer format was altered for instance, overlay assays designed with the intermediate agar layer omitted. Varying soft Mueller-Hinton agar percentages were tested to verify whether decreasing the relative concentration of agar would impact the diffusion of recombinant peptides and thus zone diameters. Pln-423 peptides seemed inclined to more readily diffuse when the top layer was composed of 0.35% Mueller-Hinton agar. However, the difference compared to 0.7% soft Mueller-Hinton agar was negligible and observational differences were likely a result of impaired *Listeria* lawn formation rather than significant increases zone diameter or improved visualisation. Application of the modified Mueller-Hinton soft agar overlay method (**Figure 4.4**) with *E. coli* JE5505 carrying *pSD3_Pln423-BspQI,-gIII* vectors demonstrated the antimicrobial activity of Pln-423 wt and mutants; Pln-H28R/H34E and Pln-S27D/K36N against *L. innocua* within another overlay assay system **(Figure 4.5).**

Figure 4.4 Optimised Mueller-Hinton soft agar leaky *E. coli* JE5505 overlay assay method for the screening of known or uncharacterised AMPs.

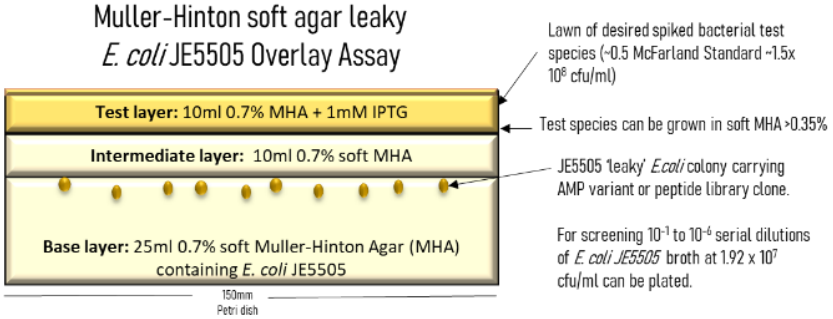
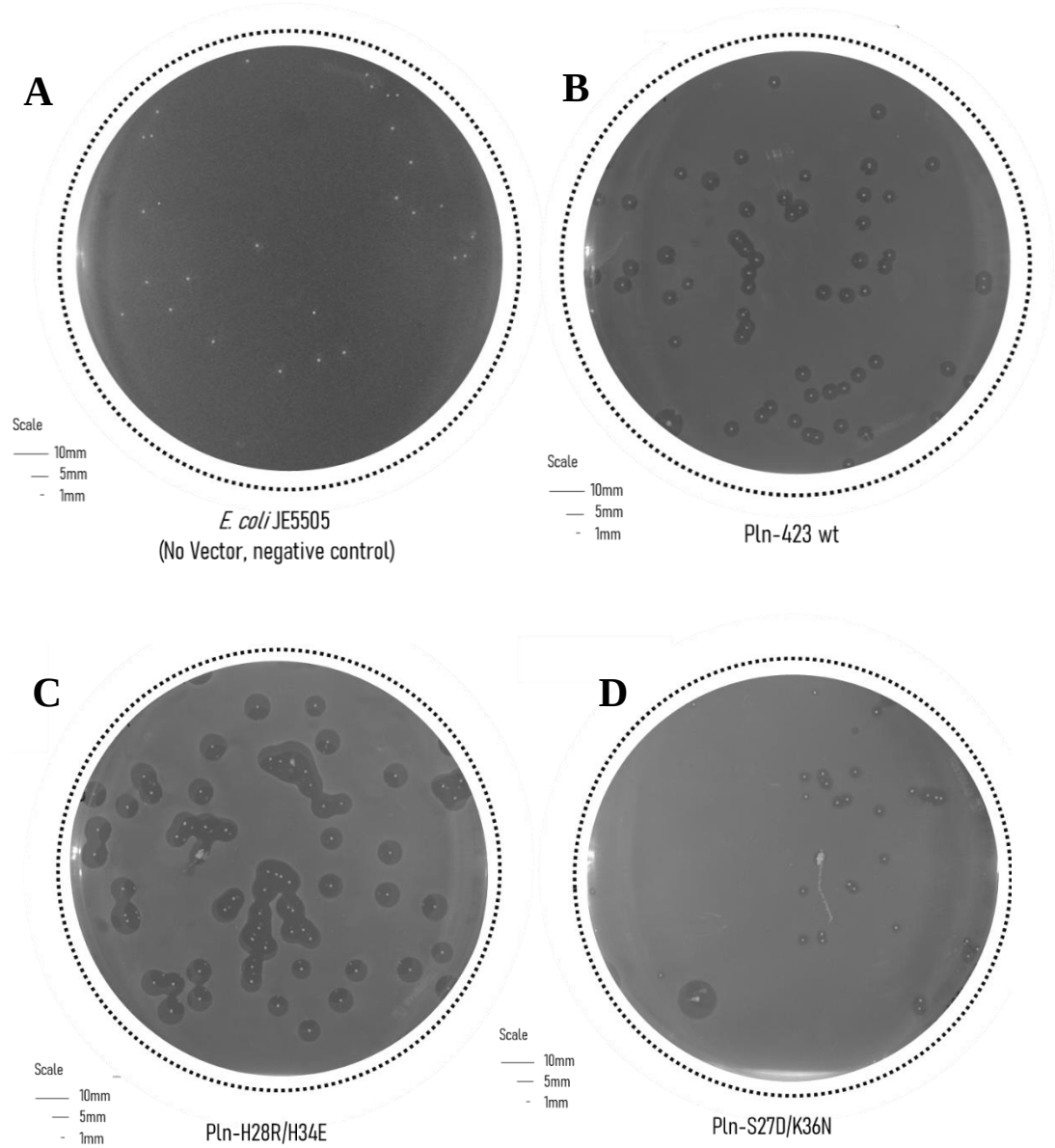


Figure 4.5. *E. coli* JE5505 (*pSD3_3-BspQI,-gIII*) encoding Pln-423 and improved mutants against *L. innocua* in a 0.7% Mueller-Hinton soft agar overlay assay screen.

Triplicate 0.7% Mueller-Hinton soft agar overlays (150mm petri dishes) screened against *L. innocua* using for *E. coli* JE5505 *wt* (No vector, negative control) which produces no visible zones of inhibition (**A**). *E. coli* JE5505 colonies carrying expression vector *pSD3_PlnX-BspQI,-gIII;* where X = Pln-423 wt **(B)**; Pln-H28R/H34E**(C)**; Pln-S27D/K36N**(D)** are shown. The average diameter of zones were equivalent to the modified 0.8%LB/TSB overlay. Neat Mueller-Hinton broth culturing at 37^{o}C, 200rpm for both *E. coli* JE5505 and *L. innocua* test species. 10^{-6} dilution of a *E. coli* JE5055 mid-log broth (~1.92 x 10^{7} cfu/mL) whereas *L. innocua* was spiked at 0.5McFarland (~1.5 x 10^{8} cfu/mL) into 0.7% soft agar containing 1mM IPTG.



4.3.4 ***Growth kinetics of E. coli JE5505 carrying pSD3 vectors cultured with glucose or IPTG.***

Once the overlay assay method was established several literature and repository database AMPs were targeted for recombinant expression in the E. coli JE5505 system. Firstly, the antimicrobial activity of these peptides was characterised in minimum inhibitory concentration (MIC) assays, as shown in **Table 4.3.** During the process of utilising Mueller-Hinton broth to grow both *E. coli* JE5505 *wt* (negative control, no vector) and *pSD3* transformed variants an observable lag in the growth kinetics (>15 mins) of *E. coli* JE5505 broth cultivations carrying AMP *pSD3* vectors (*i.e.* Pln-423/mutants) was identified. Growth curves of JE5505 wt (No pSD3 vector), *E. coli* Je5505 carrying *pSD3_Pln423-BspQI,-gIII* and *pSD3_RLL-BspQI,-gIII* which would express Pln-423 and Peptide RLL (Sainath Rao, Mohan and Atreya, 2013) respectively were conducted. For the aforementioned *E. coli* JE5505, Mueller-Hinton broth culturing supplementation included the presence of 1% glucose or 1mM IPTG. Optical density (OD_{600nm}) changes were subsequently tracked for <7hrs (**Figure 4.6**). 1mM IPTG/Mueller-Hinton broths resulted in similar growth pattern to neat Mueller-Hinton cultivation for all three *E. coli* JE5505 tested in **Figure 4.6**. This possibly indicates IPTG activated expression of recombinant peptides places minimal pressures on the growth of *E. coli* JE5505 and/or even in the absence of IPTG induction, *E. coli* JE5055 exhibits basal expression leakage despite systems of positive control via catabolite repression (Rosano and Ceccarelli, 2014).

Table 4.3 The characterisation of antimicrobial activity of natural or phage-display derived AMPs; Peptide RLL, Pln-423, Bacteriocin E50-52, Cathelicidin-BF and Plectasin defensin.

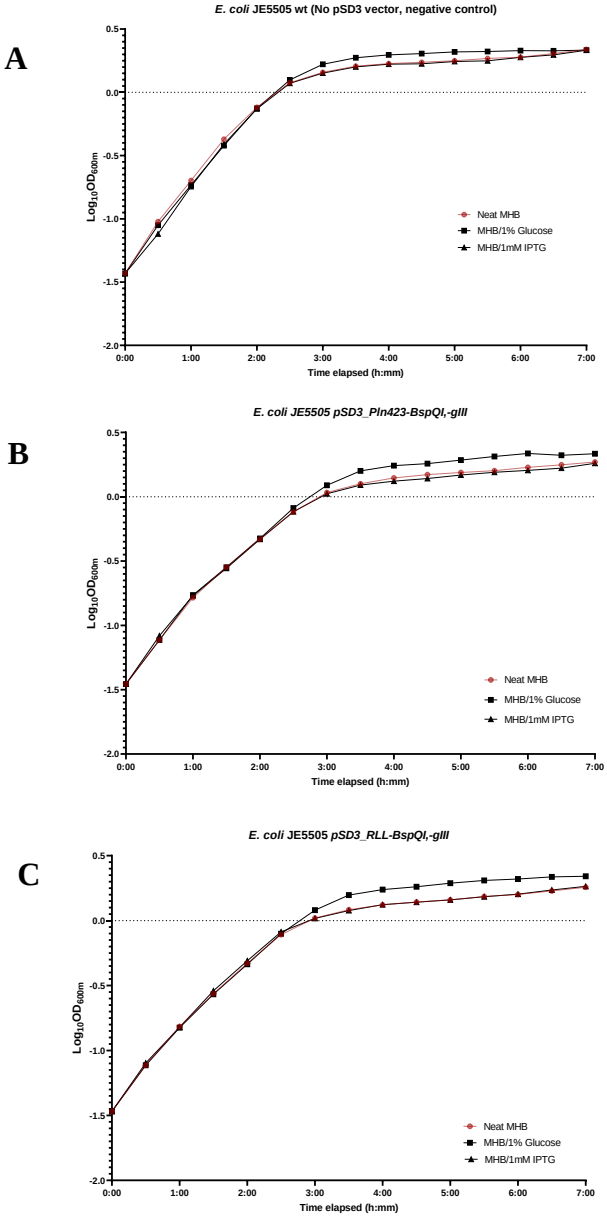
Comparison of the antimicrobial activity of natural and phage displayed AMPs					
	Literature/AMP database annotations				
AMP/(Spectrum of activity)/Reference	**Literature referenced MIC ranges (µM)**[1]	**Test species identified as susceptible to antimicrobial activity**[1]		**MIC test species**	**Observed broth microdilution MIC**
Peptide RLL **(BS)**	4.80 - 154.20	*E. coli*		*E. coli JE5505* *E. coli ETEC* *S. Typhimurium* *L. innocua* *S. suis*	12.5 µM 6.25 µM 25 µM >200 µM 100 µM
Pln-423 wt **(GP)**	0.075	*Listeria spp* *Bacillus spp* *Lactobacilli spp*		*E. coli JE5505* *E. coli ETEC* *S. Typhimurium* *L. innocua* *S. suis*	100 µM ND ND 0.097 µM 1.56 µM
Bacteriocin E50-52 **(BS)**	0.020	*Campylobacter jejuni* *L. innocua* *L. monocytogenes* *S. Typhimurium* *E. coli*		*E. coli JE5505* *E. coli ETEC* *S. Typhimurium* *L. innocua* *S. suis*	 >200 µM >200 µM 1.56 µM >200 µM >200 µM
Cathelicidin-BF **(GN)**	0.275 – 2.190	*S. Typhimurium*		*E. coli JE5505* *E. coli ETEC* *S. Typhimurium* *L. innocua* *S. suis*	0.09 µM ND 1.56 µM ND ND
Plectasin defensin (**GP)**	0.010	*S. suis*		*E. coli JE5505* *E. coli ETEC* *S. Typhimurium* *L. innocua* *S. suis*	>200 µM ND >200 µM >200 µM >20 0µM

Key: MIC refers to Minimum inhibition concentration observed via the CLSI broth microdilution method.

[1] = Refers to literature or database derived MIC values. Note only susceptible test species which are relevant to this study are described. Spectrum of activity was orientated towards; Gram-Negative bacteria (GN), Gram-Positive bacteria (GP) and Broad spectrum (BS)

Figure 4.6 Growth curves of *E. coli* JE5505 wt and pSD3 carrying variants in Mueller-Hinton broth supplemented with 1% Glucose or 1mM IPTG.

E. coli JE5505 carrying *pSD3* encoding Pln-423 wt **(B),** Peptide RLL **(C)** and *E. coli* JE5505 wt with no *pSD3* vector (**A**) were subjected to 7hrs tracking, from a starting OD_{600nm} = ~0.025. Three broth culturing were neat MHB (red line), and MHB supplementation with 1% Glucose or 1mM IPTG. (**A**) *E. coli* JE5505 wt demonstrates improved growth with 1% glucose supplementation, and this supersedes neat MHB and 1mM IPTG broth cultures. Notably, growth kinetic pattern is replicated for *E. coli* JE5505 carrying *pSD3* **(B/C)** however, both suffer a slower rate of growth in all conditions compared to the negative control.



As demonstrated in **Table 4.3**, Peptide RLL exhibited a 12.5µM MIC against *E. coli* JE5505, whereas ~100µM MIC was observed for Pln-423. Despite this contrasting antimicrobial activity against *E. coli* JE5505, the carriage of *pSD3* vectors encoding Pln-423 or Peptide RLL both resulted in a minor lag in the growth kinetics compared to the *E. coli* wt (No vector) (**Figure 4.6**). However, Peptide RLL does exhibit the greatest disparity when comparing neat and 1mM Mueller-Hinton broths to 1% glucose supplementation. Nonetheless, a sharp increase is observed for 1% glucose Mueller-Hinton broth cultivation across all three specimens (**Figure 4.6**). Glucose is the most preferred and firstly consumed carbon source which leads to reduced *lac* operon induction (Bren *et al.*, 2016). Disruption of this feedback loop therefore restricts recombinant peptide expression and energy expenditure from protein production. Therefore, impaired growth of *E. coli* JE5505 *pSD3* carrying variants might simply be caused by the exhaustion of sugars which in turn initiates metabolically demanding recombinant peptide production. The excess availability of 1% glucose in the Mueller-Hinton broth therefore supports improved growth kinetics, similar to *E. coli* JE5505 wt **(Figure 4.6)**. Consequently, the modified method herein, utilises 1% glucose Mueller-Hinton agar and broth media for *E. coli* JE5505 cultivation in all steps proceeding plating onto overlay assay.

4.3.4.1 Mueller-Hinton soft agar overlay assay method for high-throughput screening of peptide libraries.

The empirical enumeration of *E. coli* JE5505 to directly dictate *n* colonies per assay is critically important. Serially diluting (1/10) the *E. coli* JE5505 mid-log broth specifically diluted to OD_{600nm} 0.238 can be optimised to provide *n* colonies per 150mm overlay assay petri dish, depending on volume plated onto the agar and the purpose of screening. Plating 350 µl of the 10^{-6} (1:1000000) dilution of the OD_{600nm}

0.238 *E. coli* JE5505 broth provides approximately <80 colonies per 150mm petri dish, and this is perfectly suitable for analysing antimicrobial activity of a single *E. coli* JE5505 encoded peptide or for scanning studies. To transform the methodology described herein into a high-throughput screening strategy simply centres on; plating 350 µl of the 10^{-3} (1:1000) dilution of the OD_{600nm} 0.238 *E. coli* JE5505 broth (**Method 2.5.10**). This provides generates ~1.4 x 10^4 colonies per plate. Wherein only ten replicate samples would be necessary to achieve a screen coverage of a ~10^5 individual *E. coli* JE5505 colonies. To exemplify this method for library screening, an *E. coli* JE5505 *pSD3_Pln423 and _RLL* co-culture mix was analysed in the overlay.

Specifically, 10^{-3} (1:1000) dilution of OD_{600nm} 0.238 broths for *E. coli* JE5505 carrying *pSD3_RLL-BspQI,-gIII* and *pSD3_Pln-423-BspQI,-gIII* or *pSD3_ Pln-S27D/K36N-BspQI,-gIII* were mixed and co-plated. This co-mixture was tested against *L. innocua* in the 0.7% Mueller-Hinton soft agar overlay assay. *E. coli* JE5505 carrying *pSD3_Pln-423-BspQI,-gIII* or *Pln-S27D/K36N* were spiked (1:100) into the 10mL volume of *E. coli* JE5505 carrying *pSD3_RLL-BspQI,-gIII* (OD_{600nm} 0.238, 10^{-3}). Peptide RLL did not exhibit antimicrobial activity against *L. innocua* in any characterisation assays including the overlay. Therefore, the visualisation of zones from the plated "*E. coli* JE5505 *pSD3_Pln-423orS27D/K36N and _RLL* co-culture mix" was assumed to be due to the production of recombinant Pln-423 or S27D/K36N. The S27D/K36N was specifically of interest as the zones produced by this variant were on average smaller than the wt.

However, to exemplify how this approach could be utilised for library screening and variant isolation this assumption needed a succinct strategy for clone recovery. Guralp *et al.,* (2013) suggested high-throughput adoptions of their strategy should use robotised colony picking and colony PCR to identify peptides encoded by zone

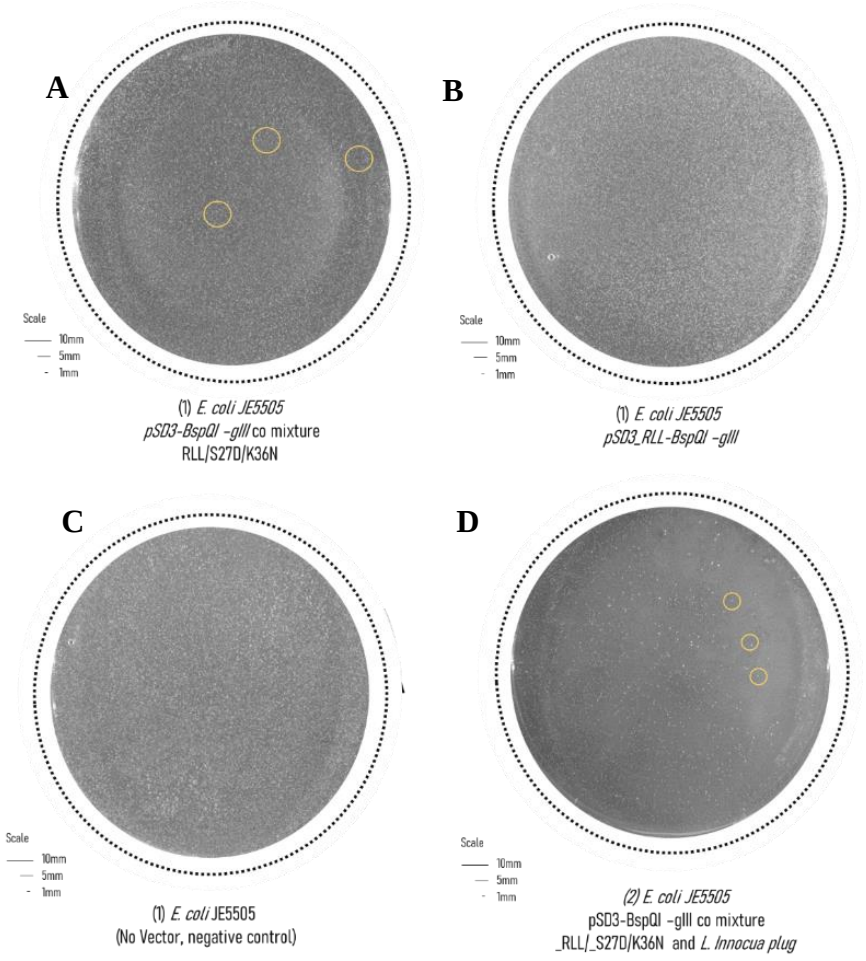
producing *E. coli* JE5505 colonies. This suggestion is supported herein as > 10^4 colonies per assay results in smaller *E. coli* JE5505 colonies and this was observed both in *E. coli* JE5505 wt and p*SD3* transformants.

Small colonies hinder accurately "picking" colonies for a colony-PCR isolation approach. Consequently, the preferred approach required the identification of a region of agar containing a zone of inhibition whereupon a ~0.75cm diameter plug was created and recovered in 1% glucose/MHB overnight. At this stage an overnight glycerol stock was generated as a backup. The plug containing overnight broth is then utilised in a secondary overlay assay screen whereby ≤80 colonies per overlay desired with ≥6 replicate plates; *i.e.* plating ~2.1mL/10mL OD_{600nm}0.238 overnight dilution 10^{-3}. Importantly, this secondary repeat provides the opportunity to verify zone of inhibition formation and/or antimicrobial activity against the same test species. Well-spaced and relatively small numbers of *E. coli* JE5505 colonies facilitate greater ease of plug excision for identification.

Colony PCR cloning is redundant here, as a plug of a singular zone-forming colony can simply be grown overnight. 50% of the overnight can be utilised for glycerol stocks and the reminder miniprepped for the extraction of vector DNA wherein utilisation of *pSD3* sanger sequencing primers support identification. Re-verifying antimicrobial activity with a secondary screen and sanger sequencing plug containing overnights identified Pln-423 ad S27D/K36N as the antimicrobial active variants from the co-mixture. The overlay assays and sanger sequencing which achieved this for S27D/K36N from the "*E. coli* JE5505 *pSD3*_ S27D/K36N and _*RLL* co-culture mix" experiment is shown in **Figure 4.7.**

Figure 4.7. Modified Mueller-Hinton leaky *E. coli* JE5505 overlay assay library screening approach and recovery of zone producing JE5505 colonies for antimicrobial peptide variant identification.

(A/B/C) are overlay assays produced where >10^4 *E. coli* JE5505 colonies were present per triplicate overlay assay condition tested against *L. innoca*. **(A)** Co-mixture of *pSD3_-BspQI_-gIII* carrying *E. coli* JE5505 encoding either Pln-S27D/K36N or Peptide RLL (1:100 ratio). **(B) and (C)** control plates represent plating only *E. coli* JE5505 carrying *pSD3_RLL_-BspQI_-gIII* or *E. coli* JE5505 wt (no *pSD3* vector) respectively and no zones of inhibition were observed. **(D)** Subsequent recovery plate and secondary antimicrobial verification assay generated from the overnight plug broth plated 1.92 x 10^7 cfu/mL diluted 10^{-3}. The plugs were excised from 3 observed zones of inhibition (yellow circles) in assay **(A).** Isolated zone producing *E. coli* JE5505 colonies (yellow circles) were then recovered as plugs**(D)**, then grown overnight in MHB 37ºC, 200rpm. Sanger sequencing of the miniprepped plug broths successfully identified peptide variant Pln-S27D/K36N **(E).**



E

Sanger sequencing to recovery

(i)

NNNNNNNNNNNNNNNNNGNNGCGGCGCAATTCTATTTNNGGAGACAGTCATAATGAAATACTTACTGCCGACTG

CAGCTGCNNNNNNACTGTTACTGGCTGCCCAGCCGGCGATGGCCAAATATTATGGCAACGGCGTTACCTGCGGCA

AACATAGCTGCAGCGTTAACTGGGGCCAGGCGTTTAGCTGCAGCGTTAGCCGTCTGGCGAACTTTGGCGAAGGCA

AATGCTAATCATGCCAGTTCTTTTGGCTAGCTAATAATTATGCTAGAGGTGGC

(ii)

5'3' Frame 3

XXXXXXXGAILFXETVIMKYLLPTAAXXXLLLAAQPAMAKYYGNGVTCGKHSCSVNWGQAFSCSVSRLANFGEGKC-SCQFFWLANNYARGG

4.3.5 ***Evaluating the use of tetrazolium red in broth and agar AMP characterisation assays***

When overlays are highly populated with *E. coli* JE5055 the colonies observed are naturally smaller (typically <1mm). Smaller colonies tended to produce smaller zones which can hamper detection of an AMP-expressing *E. coli* JE5505 library variant. Consequently, the adaption of techniques from colorimetric agar-based assays were explored to improve the visualisation of zones. McLaughlin and Balaa, (2005) utilised 2,3,5-triphenyltetrazolium chloride (tetrazolium red) to enhance the visualisation of phage plaques formed against *Salmonella enterica subsp. enterica* in 0.75% soft trypticase soy agar assays. The use of tetrazolium red for notion of improved zone visualisation in the overlay assay was explored. Several concentrations (w/v) of tetrazolium red were tested against *E. coli JE5505, S. Typhimurium* and *L. innocua* in MHB broth to determine toxicity and plates were read at OD_{450nm} to excite formazan red (**Figure 4.8**). Two different starting cfu/mL were utilised *i)* 0.5 McFarland standard (~1.5 x 10^8 cfu/mL) and *ii)* 5 x 10^5 cfu/mL the recommended starting confluency for brth microdilution MIC assays. Enumerations were calculated for each spiked concentration and compared against negative controls with no tetrazolium red. 0.0019% tetrazolium red (w/v) was the lowest concentration which produced observable red precipitate **(Figure 4.8 and Figure 4.9)**. <0.0039% (w/v) tetrazolium red concentrations were sufficiently non-toxic, exerting negligible effects on bacterial growth and produced colony forming units similar to the negative control **(Figure 4.8 and Figure 4.9)**.

Figure 4.8 *S. Typhimurium* broth cultures of varying colony forming units in a Mueller-Hinton broth tetrazolium red colorimetric assay.

Mid-log *S. Typhimurium* was diluted to approximately 0.5 McFarland standard (1.5×10^8 cfu/mL) and 5×10^5 cfu/mL. One of the duplicates assays was instantly subjected to Tetrazolium red (0.0004% to 1.0% w/v). In the remaining duplicate assay, Polymyxin B was spiked into all non-control wells at 200µM (> well in-excess of MIC) then the assay was incubated static 37ºC <18hrs. Tetrazolium red was spiked into wells (except negative control wells) and incubated at RT for <30mins. Formazan red production was clearly observable down to 0.0019% w/v when polymyxin B was not spiked into wells. Colony forming units remained within low standard deviation under ≤ 0.15625% Tetrazolium red.

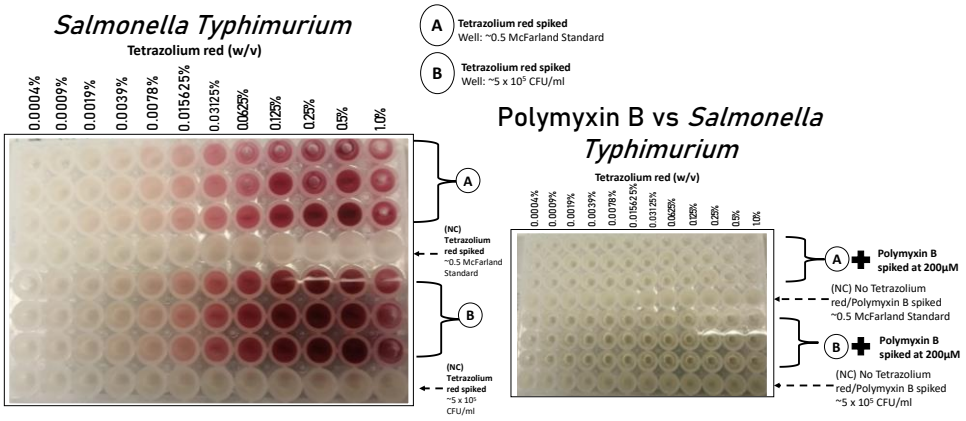
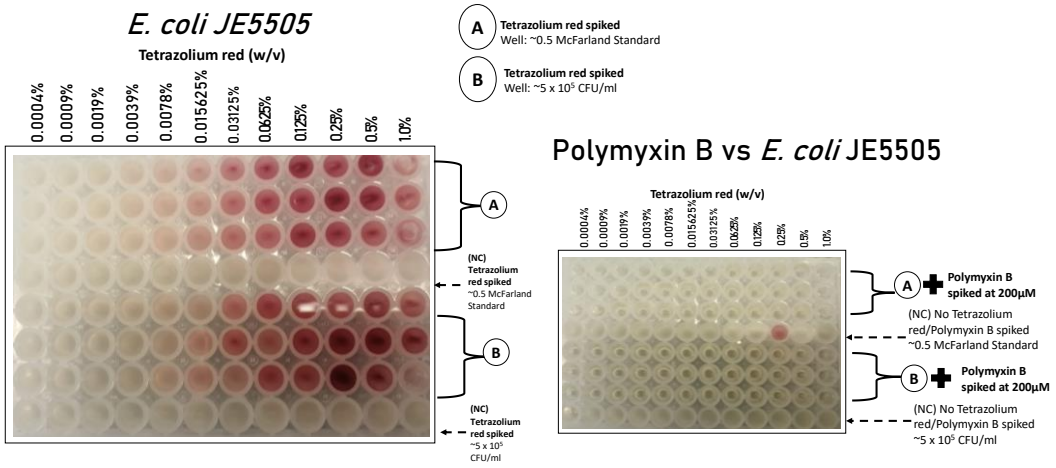


Figure 4.9 *E. coli* JE5505 broth cultures of varying colony forming units in a Mueller-Hinton/tetrazolium red colorimetric assay.

The method used to generate **(Figure 4.12)** was repeated for *E. coli* JE5505. Formazan red production was also clearly observable down to 0.0019% w/v tetrazolium red when 200µM polymyxin B was not spiked into plate wells for both ~ 0.5McFarland broth (1.5×10^8 cfu/mL) and the $5x\ 10^5$ cfu/mL. Colony forming units remained within low standard deviation from the 0.5 McFarland and 5×10^5 cfu/mL controls with concentrations of ≤ 0.15625% Tetrazolium red.



For utilisation in agar-based assays herein higher than >0.0039% concentrations were necessary simply to compensate for higher bacterial content and the impracticality of flooding large volumes of low concentrations of tetrazolium on each assays. In *E. coli* JE5505 *pSD3_Pln-423-BspQI,-gIII* vs *L. innocua* overlay assays (modified 0.7% soft MHA method), tetrazolium red rapidly diffuses through soft agar and, within ~30 minutes *E. coli* JE5055 colonies turn slightly pink in colour. No observable improvements were seen for zone visualisation even when using 100 µl of 1.0% tetrazolium red. The test species lawn is evenly distributed and embedded into the agar matrix, as such formazan red colorimetric shift is less apparent compared to the dense *E. coli* JE5505 colonies. For this purpose, tetrazolium red addition to soft agar overlays was omitted however, 0.0019-0.0039% (w/v) tetrazolium red are putatively beneficial for observational MIC determination in broth assays.

4.3.6 Overlay assay limitation; E. coli expression vs peptide diffusivity;

As mentioned previously, Peptide RLL and several other literature and/or AMP repository database derived AMPs were inserted into the *pSD3_-BspQI,-gIII* vector for expression in *E. coli* JE5505. A selection of these peptides were synthesised crude with no modifications (*i.e.* amidation). Microdilution MIC values were determined for bacterial pig pathogens such as; *S. Typhiumurium*, *S. suis* P1/7 and Enterotoxin producing *E. coli* (ETEC) of porcine origin. The literature peptides with characterised antimicrobial activity under <100µM include broad spectrum acting Peptide RLL (Sainath Rao et al., 2013); 6.25 to 25µM for *E. coli JE5505, ETEC* (porcine host specificity*), S.Typhimurium* and *S. suis P1/7*. Peptide Cathelicidin-BF (YiFan Liu, et al., 2011;Zhang *et al.*, 2015); 0.09-1.56µM (*E. coli JE5505, S.Typhimurium*). Pln-423 wt exerted low micromolar to sub-micromolar MICs 0.097-1.56µM (*L. innocua* and *S. suis*).

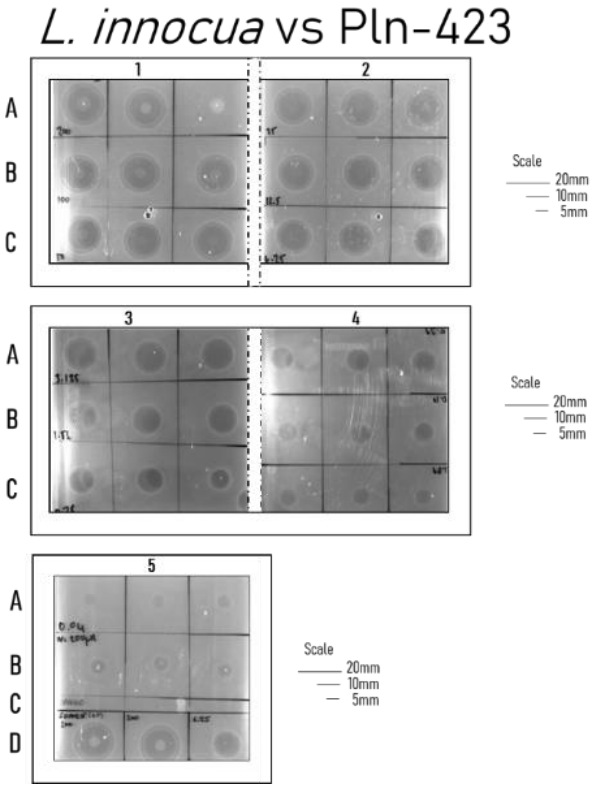
S. suis specific Plectasin was weakly active (>200µM) in broth, and this also extended to their effect on *E. coli* JE5505. Alternatively, peptides such as Cathelicidin-BF and Peptide RLL demonstrated relatively potent activity against *S. typhimurium* and towards *E. coli* strains including; porcine-host specificity ETEC and *E. coli* JE5505. Mueller-Hinton soft agar leaky *E. coli* JE5505 overlays were attempted for the several literature/AMP repository peptides against *S. typhimurium,* and *S. suis* and *E. coli* . No zones of inhibition were observed, and assays were comparable to negative controls irrespective of the *E. coli* JE5505 encoded peptide to test species assay format. Focus was then extended towards investigating the inability to visualise zones of inhibitions for non-Pln423 AMPs in the leaky *E. coli* JE5505 overlay assay. This was investigated in two-prongs, *i)* firstly demonstrate various peptides were in fact able to exert antimicrobial effects on test species in agar. *ii)* Secondly, it was important to evaluate the expression system efficiency outside of the primary visualisation of zones of inhibition from Pln-423 or respective mutant carrying *E. coli* JE5505 against *Listeria* species.

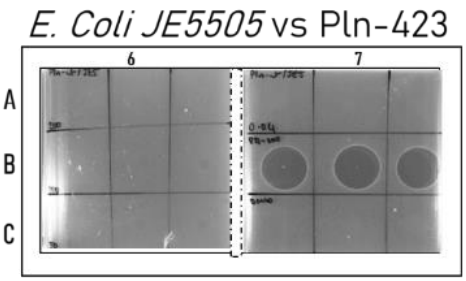
Two agar-based antimicrobial testing methodologies were employed to assess diffusivity in agar; plug diffusion peptide serial dilution agar diffusion assays **(Method 2.5.6).** These agar assays sought to illuminate whether concentrations of crude synthesised peptide equivalents (<200µM) diffuse and produce zones of inhibition when spiked or spotted onto agar containing a lawn of a test species. The results for the peptide serial dilution agar diffusion assay for crude peptides; Pln-423, Cathelicidin-BF, Plectasin, and Peptide RLL are depicted in **Figure 4.10, 4.11 and Table 4.4.** The peptide serial dilution agar diffusion assay once again demonstrated Pln-423 potent activity against *L. innocua* **(Figure 4.10),** and spotting crude Pln-423 at >0.097µM produced zone diameters similar to that observed in the *E. coli* JE5505

overlays. Pln-423 sub-micromolar antimicrobial activity both in broth and agar uniquely distinguishes Pln-423 from the majority of literature/AMP repository peptides otherwise tested herein **(Table 4.3).** Crude Peptide RLL demonstrated the ability to diffuse and inhibit the growth of *S. Typhimurium* in both selective and non-selective plug diffusion assays with brilliant green agar and Mueller-Hinton. Agar-based characterisation outside of the overlay assay *E.coli* JE5505 expression system demonstrated AMPs such as Cathelicidin-BF and Peptide RLL exert antimicrobial effects against bacteria in agar nonetheless, this was at concentrations ≤ 6.25µM

Figure 4.10 Spotting peptide serial dilution agar diffusion assay for Cathelicidin-BF and Peptide RLL tested against *L. innocua* and *E. coli* JE5505.

0.5 McFarland (~1.5 x 10^8 cfu/mL) of *L. innocua* (Plate 1-5) or *E. coli* JE5505 (**Plate 6-7);** was spiked into Mueller-Hinton agar (45ºC) for the production of confluent lawns. 5 µl triplicate spot volumes of peptide concentrations (<200µM). Plates were incubated static at 37ºC for <18hrs. Lowest inhibitory concentration in agar was determined; Pln-423 wt demonstrated activity down to ~0.04µM where faint zones (<5mm) are produced (**Plate 5**). Pln-423 produced extremely faint zones against *E. coli* JE5505 down to 100µM **(Plate 7).** Nisin and Polymyxin B (Plate 5/7) positive controls produced zones whereas neat DMSO (peptide solvent control) produced no visible inhibitory effects.





Peptide concentration for triplicate spots

	1	2	3	4	5	6	7
A	200µM	25µM	3.125µM	0.39µM	0.04µM	200µM	0.04µM
B	100µM	12.5µM	1.56µM	0.19µM	Nil 200µM	100µM	Polymyxin 200µM
C	50µM	6.25µM	0.78µM	0.09µM	DMSO	50µM	DMSO
D	-	-	-	-	200, 200, 6.25µM		

Figure 4.11 Spotting peptide serial dilution agar diffusion assay for Cathelicidin-BF and Peptide RLL tested against *E. coli* JE50505

(**Plate 1-10);** Agar assays were prepared as stated in **Figure 4.14.** Cathelicidin-BF demonstrated activity down to ~0.09µM where faint zones (<5mm) **(Plate 4).** Whereas Peptide RLL demonstrated activity down to 12.5µM **(Plate 7).** This was determined as the lowest inhibitory concentration in agar**.** Polymyxin B positive control produced zones (~15-17mm) whereas neat DMSO (peptide solvent control) produced no visible inhibitory effects.

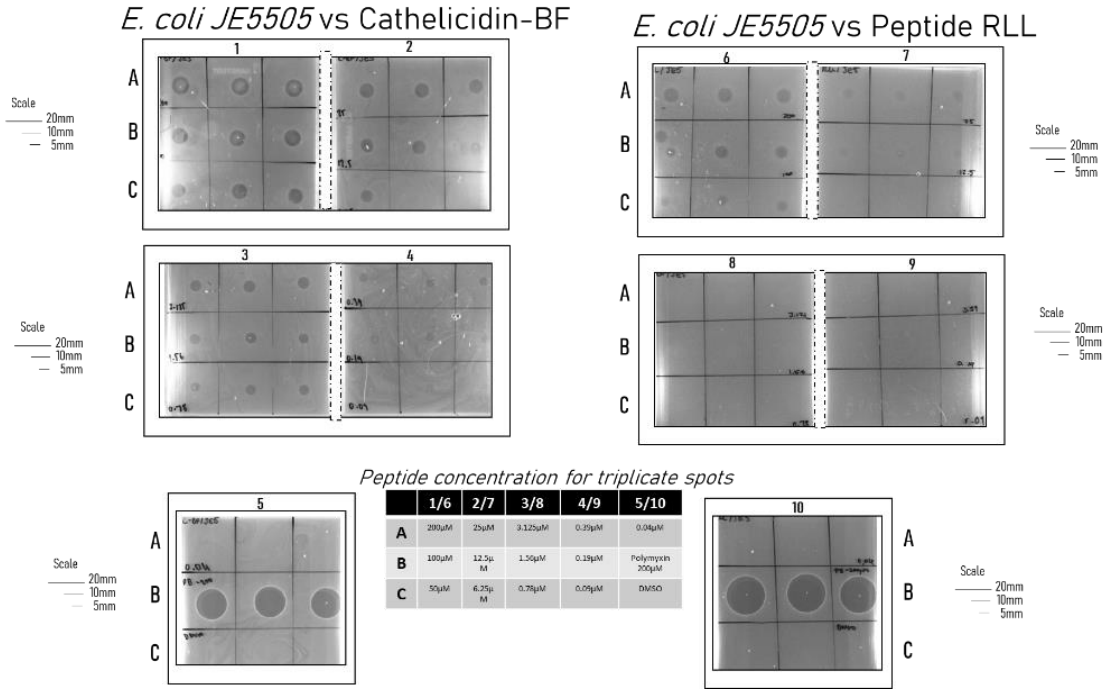


Table 4.4 Summary of the spotting peptide serial dilution agar diffusion assays of several "AMPs" against *L. innocua, E. coli JE5505, S. Typhimurium* and *S. suis.*

Antimicrobial activity of peptides in both broth and agar-based characterisation assays			
AMP (Spectrum activity)	**Test species**	**Observed broth microdilution MIC**	**Lowest inhibitory zone forming concentration**
Peptide RLL **(BS)**	*E. coli JE5505* *E. coli ETEC* *S. Typhimurium* *L. innocua* *S. suis*	12.5µM 6.25µM 25µM >200µM 100µM	0.097µM ND 6.25µM ND No activity
Pln-423 wt **(GP)**	*E. coli JE5505* *E. coli ETEC* *S. Typhimurium* *L. innocua* *S. suis*	100µM ND ND 0.097µM 1.56µM	100µM ND ND 0.039µM No activity
Cathelicidin-BF, Peptide CATHELICIDIN- **(GN)**	*E. coli JE5505* *E. coli ETEC* *S. Typhimurium* *L. innocua* *S. suis*	0.097µM ND 1.56µM ND ND	0.097µM ND 0.039µM ND ND
Plectasin defensin (**GP)**	*E. coli JE5505* *E. coli ETEC* *S. Typhimurium* *L. innocua* *S. suis*	>200µM ND >200µM >200µM >200µM	No activity ND ND ND 0.78µM
Key: Peptides were characterised via broth microdilution (**Method 2.5.2**) and agar diffusion (**Method 2.5.6**). The MIC values determined herein are from triplicate samples (3 well replicate or 3 spots replicates). **ND** = Not determined, "**No activity**" refers to the absence of any zones of inhibition. Spectrum of activity was orientated towards; Gram- Negative bacteria (**GN**), Gram-Positive bacteria (**GP**), Broad spectrum **(BS)**			

Modified co-streaking and cross-streaking methodologies were attempted (**Method 2.6**) on a thick layer of agar containing 1mM IPTG. This was conducted with *E. coli* JE5505 *pSD3_-BspQI,-gIII* encoding peptides; Peptide KPQ, Peptide KFV, Peptide WRE, Peptide SLN and Peptide AFV specifically against *B. subtilis* and *S. Typhimurium*. Co-streaking and cross-streaking aims to create overlapping streaks of the antimicrobial producer bacteria and test species. Inhibitory activity is quantified via striated streaks with gaps where the test species was unable to grow. The antimicrobial activity of *E. coli* JE5505 *pSD3_-BspQI,-gIII* encoding peptides could not be effectively evaluated by these methods as human error when interpretating the start and endpoints of streaks can affect the accuracy of detecting "gaps" in streaking.

Nonetheless, a simplified supernatant spotting experiment was devised **(Method 2.5.5).** Wherein *E. coli* JE5505 carrying *pSD3_-BspQI,-gIII* vectors encoding peptides were grown to mid-log and then IPTG (0.5mM and 1mM) induced and incubated +3hrs to overnight at 37ºC or 30ºC. The overnight IPTG inductions were filtered via 0.025µm pore size membrane filters to produce cell-free supernatant. 5 µl, 10, 50 µl of supernatant triplicates were spotted onto agar containing the test species spiked to produce a confluent lawn as seen in the overlay assay. No zones of inhibition were observed from this spotting strategy.

Utilising the membrane filtered IPTG induction supernatant in a 50:50 mix microdilution MIC with various test species additionally did not provide any evidence of antimicrobial activity. Broth culturing with IPTG did not leach sufficient recombinant peptides into broths and this deterred the potential for broth-based *E. coli* JE5505 screening strategies. Embedment into agar matrices likely improves the capacity of *E. coli* JE5505 to maintain more stable growth and metabolic control as

the colony mass accumulates recombinant peptides which diffuse locally rather than within solution as seen in broth.

4.4 Screening degenerately randomised 16-mer peptide libraries in soft agar overlay assay approach for the identification of library AMP candidates

4.4.1 ***Screening VNN_{15}+$(TTT)_1$ and NNK randomised 16-mer peptide libraries in 0.7% Mueller-Hinton soft agar overlay assay***

A pSD3 phagemid beneficially ensures libraries are compatible with phage display screening, wherein peptide pIII fusions are expressly generated due to *gIII* inclusion downstream of the randomised peptide region. Nevertheless, this study sought recombinant peptides not expressed as *pIII* fusions in order to ensure limited distribution to structural conformations or solubility. Non-pIII fusions were desired as encoded peptides in *E. coli* JE5505 expression systems tended to be <100Da. Additionally, removing the large pIII protein reduces the potential of undue metabolic stresses from the production of this large heterologous protein in *E. coli* JE5505. A simple method for *SpeI* digestion of *pSD3_-BspQI_+SpeI_opal stop* vectors was shown to remove *gIII* . Titrations of *E. coli* JE5505 sub-libraries estimated library sizes were 1 x 10^7 for VNN_{15}+$(TTT)_1$ and 5 x 10^6 for NNK. *E. coli* JE5505.

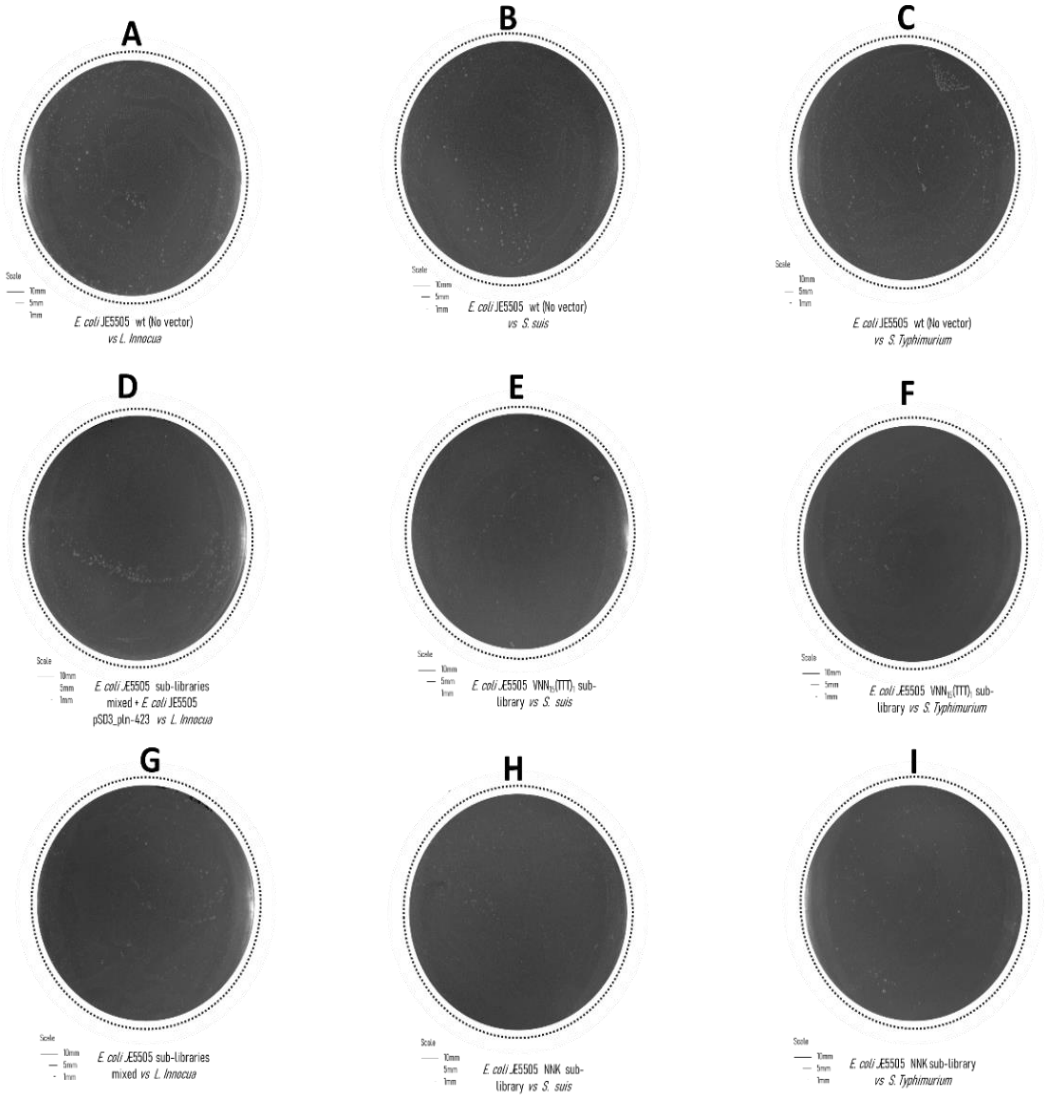
The optimised Mueller-Hinton soft agar leaky *E. coli* JE5505 overlay assay was able to demonstrate the antimicrobial activity of Pln-423 wt and Pln-H28R/H34E and Pln-S27D/K36N mutants against *Listeria* species. Additionally, the isolation of zone producing variant was exemplified from a screening format wherein >10^4 *E. coli* JE5505 *pSD3* carrying colonies were present per overlay assay **(Figure 4.7).** Although, limitations were identified in demonstrating the antimicrobial activity of non-Pln423 peptides within the *E. coli* JE5505 overlay assays. Nonetheless, in alternative overlay method (Tominaga and Hatakeyama, 2006) exemplified the

antimicrobial activity of natural AMP pediocin PA-1 using the *E. coli* JE5505 expression system. Both Pln-423 and PA-1 possess highly potent sub-micromolar MICs against mainly Gram-positive foodborne pathogens such as *Listeria* species. Zone of inhibition formation in the leaky *E. coli* JE5505 overlay assay might therefore favour the detection of sub-micromolar AMPs with a greater propensity for Gram-positive bacteria binding and/or killing effect. This seeming preference of the *E. coli* JE5505 expression system for Gram positive bacteria AMPs, with sub-micromolar broth microdilution MICs, is putatively the main advantage of this overlay method.

Consequently, the next focus aimed to utilise the modified Mueller-Hinton soft agar leaky *E. coli* JE5505 overlay to screen the $VNN_{15}(TTT)_1$ and NNK degenerately randomised peptide libraries**.** *E. coli* JE5505 library glycerol stocks were recovered in MHB/1% glucose broth for use in the aforementioned Mueller-Hinton leaky *E. coli* overlay assay (**Method 2.5.10**). Ten plates containing $>10^4$ colonies were screened per library, and three sets of triplicate controls; wherein *E. coli* JE5505 wt (No pSD3 vector) was the negative control. 1mL of the OD_{600nm} 0.238, 10^{-3} broth dilution for *E. coli* JE5505 library sets was combined with 1mL of a similarly prepared and diluted broth of *E. coli* JE5505 carrying *pSD3 Pln-423_-BspQI,-gIII.* This co-mixture was then plated onto 6 individual overlay assay plates, three of which were overlaid with *L. innocua* and the remaining three with the test species of choice (*S. Typhimurium* or *S. suis*) **(Figure 4.12).** The overlay assay screening of the *E. coli* JE5505 $VNN_{15}(TTT)_1$ and NNK peptide libraries did not lead to the identification any visible zones of inhibition or the identification of AMP candidate library variants.

Figure 4.12. 0.7% soft agar Mueller-Hinton *E. coli* JE5505 overlay assay screening of 16mer peptide libraries degenerately randomised via $VNN_{15}(TTT)_1$ or NNK scheme.

~10^5 screening coverage of *E. coli* JE5505 $VNN_{15}(TTT)_1$ and NNK sub libraries was completed against test species *S. Typhumirum* and *S. suis*, exemplar plates are represented in overlay assays (**E, F, H and I)**. No zones of inhibition were identified across assays **E/F/H and I**, these plates were similar to respective negative control plates (**B-C**) where *E. coli* JE5505 wt (no *pSD3* vector) is screened against the two test species. To ensure antimicrobial activity is demonstrable within a specific library screening overlay assay run**, (A, D and G)** represent additional control overlay assays. *E. coli* JE5505 carrying *pSD3_Pln-423_-BspQI_-gIII* was spiked into a 50:50 mixture of *E. coli* JE5505 $VNN_{15}(TTT)_1$ and NNK sub libraries and tested against *L. innocua* (D), zones were observed for this control assay and absent in overlay **(G)** which is the 50:50 mixture of *E. coli* JE5505 $VNN_{15}(TTT)_1$ and NNK sub libraries absent of Pln-423 encoding *E. coli* JE5505.



4.5 Discussion

4.5.1 Key considerations made for the optimisation of the Mueller-Hinton soft agar E. coli JE5505 overlay assay; media characteristics and assay format

E. coli expression systems strongly confer several advantages including; ease of *E. coli* genetic manipulation underpinned by highly comprehensible genetics, rapid growth rate within affordable fermentation media and crucially the high yield of recombinant proteins (Rosano and Ceccarelli, 2014). Various methods exist to capture recombinant proteins from robust *E. coli* cell factories. The application of osmotic shock, high-pressure, sonication, milling, grinding or enzymatic action (lysozymes *etc*) are such examples (Simpson, 2010). Alternatively, exploitation of unique *E. coli* membrane mutants offers the ability to bypass the necessity of aforementioned recombinant product capture approaches (Mergulhaõ et al., 2005). Recombinant proteins exported to periplasmic spaces are often retained in soluble periplasm fractions or leached extracellularly into media (Mergulhaõ et al., 2005). Kanamori *et al.,* (1998) designed a patented approach for manipulation of the *E. coli* JE5505 *lpp* deletion mutant, with a vector system containing the *trp* promoter and alkaline phosphatase signal peptide fusions for periplasmic secretion. JE5505 (F− *1po his proA argE thi gal lac xyl mtl tsx*) is a *Δlpp-254* mutant which lacks free and bound murein lipoprotein. Kanamori *et al.,* (1998) demonstrated ~50µg/mL of low-molecular weight recombinant human pancreatic secretory trypsin inhibitors (~56mers) were produced by *E. coli* JE5505 in culture supernatant.

Approaches adapted from Miller *et al.,* (1993) such as the enclosed soft Mueller-Hinton agar overlay and other previously reported iterations (Tominaga and Hatakeyama, 2006; Guralp et al., 2013), have demonstrated the *E. coli* JE5505 can produce recombinant peptides with antimicrobial activity in optimised agar-based

assays. Adapted Miller *et al.,* (1993) approaches primarily differ on agar composition/percentage; 0.7-2.5% MHA, LB, TSB and MRS and assay incubation strategy. In concern for the latter, Tominaga and Hatakeyama, (2006) incubated each stage of the overlay assay preparation at 30ºC ≤21hrs instead of 37ºC ≤24hrs as preferred with the present method. Nonetheless, all three modified overlay assay approaches utilise a 3-layer soft agar overlay format, with a thin uppermost layer spiked with 1mM IPTG to activate *lac* operon expression of encoded peptides in vectors. The test species inoculum concentrations devised herein are higher than previously reported for an *E. coli* JE5505 overlay assay (Tominaga and Hatakeyama, 2006; Guralp et al., 2013). However, the utilisation of 10^4 or 10^5 cfu/mL were shown herein and previously reported to generate less dense and uniform lawns (Hossain et al., 2022). Spiking test species at 0.5 McFarland (10^8 cfu/mL) is utilised for standardised antimicrobial characterisation assays (Hudzicki, 2009). Additionally, 0.5 McFarland staring innocula of the test species creates homogeneous confluent lawns for all test species utilised herein and improved zone visualisation.

All bacteriological media compositions under achieve in terms of replicating *in viv*o physiological ionic, nutrient and host immune conditions which downstream can neglect or overestimate antimicrobial activity (Nizet, 2017). Nevertheless, the selection of media for an overlays requires careful consideration of host producer and test species preferences for growth as well as considering antimicrobial diffusivity. As the key aim of this study was to identify novel antimicrobial peptides, the selection of Mueller-Hinton was especially apt for this purpose. More specifically, Mueller-Hinton media is recommended by European Committee on Antimicrobial Susceptibility Testing (EUCAST) and Clinical and Laboratory Standards Institute (CLSI) for antimicrobial testing especially agar disk diffusion (EUCAST, 2022). Therefore,

demonstrating the antimicrobial activity of a novel antimicrobial within Mueller-Hinton media compositions, at the point of discovery, would facilitate efficiencies in downstream characterisation (Ahman et al., 2020; EUCAST, 2022).

4.5.2 Reproducibility of zones of inhibition in leaky E. coli JE5505 overlay assays; semi-quantitative or purely qualitative?

Agar thickness effects zones of inhibition reproducibility. As increased agar depth and weight subsequently leads to higher critical concentrations of antimicrobials which in turn causes observable reductions in zone of inhibition diameters (Flanagan and Steck, 2017). The soft Mueller-Hinton agar overlay utilises ~45mL Mueller-Hinton soft agar per large bioassay dish (150mm), this likely advantageously facilitates antimicrobial diffusion by decreasing the depth and thickness of agar in the assays. The application of soft agar percentages to the Mueller-Hinton *E. coli* JE5505 overlay was intended to improve antimicrobial diffusivity whilst supporting consistent colony formation (**Figure 4.5**). >0.35% soft Mueller-Hinton agar percentages were tested herein and diffusion of Pln-423 wt and/mutant seemingly improved at lower agar percentages. However, this was largely negated by issues of zone formation as *Listeria spp* lawn growth were slightly impaired at lower soft afar percentages. Possibly indicating the importance of anchorage-dependent growth for this test species.

Low concentrations of agar, typically ≤50% reduction from the manufacturer's recommendation ~15-17.5g/L, can improve antimicrobial diffusion in the semi-solid medium, and this remains the principal rule employed to soft agar assay methodologies (Hossain et al., 2022). Nisin a central AMP positive control for Gram-positives herein, diffusion improves as agar percentages reduce from 1.5% to 0.75% when complemented with buffering phosphate salts (Lalpuria et al., 2012). Agar is a

complex of two large polymer structures; principally the neutrally charged agarose and anionic agaropectin which contains negatively charged sugar-subunits due to sulfate groups (Bhattacharjee, 2015). As such cationic antimicrobials and dyes are expected to interact with agaropectin (Bhattacharjee, 2015). However, agarose carries the gelling capacity of agar, thus the addition of agarose to Mueller-Hinton has been explored for reducing agaropectin drug absorption and improving antimicrobial characterisation in agar (Bhattacharjee, 2015; Nyerges et al., 2020). Agarose was optimised with the *E. coli* JE5505 overlay assay however this did not improve peptide diffusivity or zones (data not shown).

Reproducibility of zone formation and inhibition diameters were consistent across replicates of screening *E. coli* JE5505 carrying *pSD3_PeptideX_-BspQI_-gIII* vectors in the Mueller-Hinton 0.7% soft agar overlay, wherein the recombinant peptide (X) was Pln-423 wt, Pln-H28R/H34E or Pln-S27D/K36N (**Figure 4.5**). *E. coli* JE5505 expressing Pln-H28R/H34E produced the largest inhibitory zones, and typically the Pln-S27D/K36N improved mutant would exhibit equivalent or smaller zones of inhibition. Nonetheless, intervariance of zones diameters between *E. coli* JE5505 colony clones carrying the same *pSD3* peptide variant was observed. This observation highlights a slightly indistinct correlation between colony sizes and respective zone diameters in the overlay assay. Demonstrating larger zones cannot be assumed to be the product of merely a more potent *E. coli* JE5505 encoded recombinant peptide.

Another reasoning behind the variance in zones of inhibition diameters might be due to positional depth differences in *E. coli* JE5505 colonies within the assay agar matrix. Tominaga and Hatakeyama, (2006) improved zone reproducibility by 4-16% in secondary activity verification assays by in principle inoculating via a stab culture

approach thus, embedding the position of colony forming cells in the overlay agar assay. Tominaga and Hatakeyama, (2006) hypothesised limitations in zone of inhibition reproducibility across identical *E. coli* JE5505 colony clones arose from the proximity of colonies to the upper test species agar layer. Principally the margin edge of zones of inhibitions indicate agar regions where *E. coli* JE5505 expressed recombinant peptide variant perseveres at insufficient concentrations to inhibit test species growth. Therefore, *E. coli* JE5505 colonies in closer proximity to respective test species layer produce an ostensible but misleading improvement to the rate of diffusion simply by the reduction in distance required to contact the test species.

During the Mueller-Hinton soft agar overlay method, large bioassay plates (150mm) containing ~25mL 0.7% Mueller-Hinton soft agar were spread plated with desired broth dilution of *E. coli* JE5505. Assays were dried at room temperature for ~ <1hr and completion time varied depending on the use of laminar flow hood or benchtop bunsen. Although, broth cultures sufficiently dried the overlaying of the intermediate layer could putatively relocate *E. coli* JE5505 colony forming cells both horizontally across the bottom agar layer and vertically. This does support the hypothesis of Tominaga and Hatakeyama, (2006) which explored *E. coli* JE5505 colony depth differences. Nonetheless, in-keeping with this observation of zone reproducibility it was recognised herein larger zones of inhibition were often coupled with above average sized *E. coli* JE5505 colonies (**Figure 4.5**).

When *E. coli* JE5505 wt and *pSD3* carrying variants were maintained in a 2-layer overlay assay format, the colonies produced were larger than respective overlaid assay samples with an intermediate layer. If it is assumed *E. coli* JE5505 becomes relatively unconstrained in soft agar, especially when molten, and is free to translocate. Larger colonies might arise from a select number of viable cells which rather than being

embedded under or within the intermediate layer, they are advantageously more aerobically exposed and unconstrained by the agar matrix due to above agar colony emergence. During the initial incubation at 37^{o}C for 24hrs, viable colony forming cells positional proximity to the upper regions of the overlaid intermediate agar layer could facilitate the accumulation of larger colony masses over time and thus lead to larger zones of inhibition.

4.5.3 Characterisation of antimicrobial activity of several literature and repository database AMPs in agar-based approaches

Microbiological broth micro- or macro dilutions, time-kill tests and agar-based approaches, such as agar radial diffusion, cups, well and paper disk diffusion techniques, remain as standard methods for assaying antimicrobials (Bonev et al., 2008; Balouiri et al., 2016). The optimised Mueller-Hinton soft agar *E. coli* JE5505 overlay assay approach was able to demonstrate the antimicrobial activity of Pln-423 wt, Pln-H28R/H34E and Pln-S27D/K36N against *Listeria* species. Several literature and repository database AMPs with characterised activity against key bacterial pathogens; *S. Typhimurium*, *S. suis* and *E. coli* were collected. This included AMPs identified via phage display approaches (*e.g.* Peptide RLL) and various natural AMPs isolated from a variety of source organisms (*e.g.* peptide Cathelicidin-BF and Plectasin defensin). To verify the diffusivity of aforementioned literature peptides, a spotting peptide serial dilution agar diffusion assay was optimised (**Method 2.5.6**). This method specifically aimed to replicate the guidelines of the widely utilised Hudzicki, (2009). Kirby–Bauer disk diffusion, however employed spotting 5 µl volumes of crude peptide dilutions. More specifically, the following was strictly adhered to for improved reproducibility; ~4 mm thick Mueller–Hinton agar, <9 spots per assay and spiking test species at 0.5 McFarland (~1.5 x 10^{8} cfu/mL) (Hudzicki, 2009).

Positive controls used in this assay were AMP models; Nisin and Polymyxin B (at 200µM). Peptide serial dilution agar diffusion assays were able to identify antimicrobial activity for Pln-423, Peptide RLL, Cathelicidin-BF and Plectasin defensin in 1.0% Mueller-Hinton agar. The lowest observed concentration which inhibited bacterial growth were as follows; Pln-423 (*E. coli* JE5505, 100µM; *L. innocua* 0.039µM), Peptide RLL (*E. coli* JE5505, 0.097µM; *S. Typhimurium* 6.25µM), Cathelicidin-BF (*E. coli* JE5505, 0.097µM; *S. Typhimurium* 0.039µM) and Plectasin defensin (*S. suis* 0.78 µM). Despite the observation of these sub-micromolar inhibitory concentrations in agar, Peptide RLL, Cathelicidin-BF and Plectasin defensin were not able to demonstrate activity in the *E. coli* JE5505 pSD3 expression system and overlay assay. Pln-423, Peptide RLL and Cathelicidin-BF were able to demonstrate activity in both broth and agar media. Contrastingly, Plectasin broth MIC was observed to be >200µM against *S. suis* however, this shifted to sub micromolar concentrations in agar.

Peptide RLL and Cathelicidin-BF exerting seemingly more potent activity in the agar diffusion assay might be due to the principle of pre-diffusion. Spots of crude AMPs were allowed to absorb and diffuse into the agar medium before assay incubation. Therefore, resulting in higher concentrations of AMPs encountering the initial bacteria seed inoculum. As crude peptide spots were transferred <10mins of agar solidification, spiked test species might be relatively unrecovered from heat exposure in 45^{o}C molten agar (Lalpuria et al., 2012). Subsequently, leading to a lag phase before test species division and growth. In the agar diffusion assay, crude spotted AMPs likely exerted antimicrobial activity and reduced bacterial cell density earlier. Additionally, decreased bacterial cell density has been reported to increase the

diffusivity of antimicrobials in solid medium containing immobilised cells (Lalpuria et al., 2012). If the assumption carries that recombinant peptide production from *E. coli* JE50505 exists but remains relatively negligible in the absence of IPTG induction and sugar source exhaustion. The overlaying of the test species at 0.5 McFarland ~1.5 x 10^8 spiked with 1mM IPTG is the process which initiates the colony mass production of the recombinant AMPs. Unlike the agar diffusion assay, the overlays require *E. coli* expression system to establish and produce encoded recombinant AMPs which must diffuse extracellularly towards the overlaid test species. Recombinant AMPs likely encounter higher starting inoculum cell densities as seen in the agar diffusion assay.

Antimicrobials demonstrate significant deviation in terms of agar diffusibility, and notably effective diffusion becomes increasingly difficult the more hydrophobic and amphipathic properties said antimicrobial possesses (Bonev et al., 2008). Dissipative diffusion of antimicrobials can arise from negative interactions with the diffusion medium. Additionally, zone of inhibition formation can be affected by the loss of substrate via antimicrobial degradation (protease action), protectant biofilms and/or bacterial resistance mechanisms (Bonev et al., 2008). The proteinaceous nature of AMPs leave them especially vulnerable to proteolytic enzymes such as proteases, which often target the three basic residues, for instance serine proteases cleave C-terminal Lys and Arg residues of AMPs (Bonev et al., 2008; Lalpuria et al., 2012).

Mode of action can affect the observance of antimicrobial activity. For instance; antimicrobials such as tetracycline, nisin and subtilin either target internalisation via active transport mechanisms or disrupt outer membrane leaflet components (Kapoor et al., 2017). Whereas antimicrobials such as kanamycin, gentamicin among others

rely upon traversing intact bacteria membranes to access target sites (Bonev et al., 2008; Lalpuria et al., 2012; Kapoor et al., 2017). Varying mode of actions generate contrasting challenges and windows to interact with different resistance mechanisms *e.g.* efflux pumps, membrane associated and/or intracellular antimicrobial modifying enzymes (Bonev et al., 2008; Lalpuria et al., 2012). Additionally, certain AMPs can possess multifaceted concentration-dependent modes of action. For instance, at high concentrations nisin possesses the ability to alternatively adopt surfactant antimicrobial action rather than simply pore forming activity (Lalpuria et al., 2012).

4.5.4 Class II bacteriocins; case study for their quintessential applicability with E. coli JE5505 expression systems

Gram-positive sourced AMPs from bacterial species such as *Lactobacillus* are typically denoted as bacteriocins. Prominent examples herein include the evaluation of class IIa bacteriocin Pln-423 (Lactobacillus plantarum 423) and the use of nisin (*Lactococcus lactis*) as a positive control for peptide characterisation assays (Tingting Zhang et al., 2022; Todorov et al., 2009). Well-studied AMPs such as bacteriocins and enterocins have exemplified antimicrobial ability in a variety of agar-based assays, (Bonev et al., 2008; Tingting Zhang et al., 2022) and diffusion assays have been designed for their specific novel identification (Choyam et al., 2015). The well-studied Pln-423 and mutant variants were able to readily diffuse through three different media compositions; Luria Broth (LB), Tryptic Soy Broth (TSB) and Mueller-Hinton soft agar herein. Pln-423 is reported to be stable across a pH range of 1 – 10, and relatively resistant to heat treatment at 80°C, with only 50% reduction in activity following incubation at 100°C (1hr) (Van Reenen et al., 2003). However, being a bacteriocin AMP, Pln-423 is sensitive to proteolytic enzymes; pepsin, papain, α-chymotrypsin, trypsin and Proteinase K (Van Reenen et al., 2003; Soltani et al., 2021).

Todorov and Dicks (2006) evaluated Pln-423 absorption, wherein test strain cell suspensions were incubated with Pln-423 (37^{o}C, 1hr), and this was repeated with media of varying media compositions, pH and temperature. *Enterococcus spp.*, *Lactobacilli spp.*, *L. lactis, L. innocua 13568* and *Streptoccocus caprinus* ATCC 700066 were evaluated test strains. Membrane filtered cell-free supernatant obtained from each suspension were used in a secondary broth dilution assays, and the relative activity observed was utilised to quantify unbound-Pln 423 from suspensions. Pln-423 was absorbed by sensitive and resistant Gram-positive bacteria. However, Pln-423 susceptible bacteria tended to absorb ≥50% of spiked Pln-423 whereas <20% absorption was common for non-susceptible stains (Todorov and Dicks 2006). The *L. innocua* 13568 cell suspension absorbed at ~67%, and persisted with the joint highest absorption rate across the Gram-positive test species screened by (Todorov and Dicks 2006). For Pln-423 susceptible test species, high absorption percentage values for Pln-423 were maintained across various pH (2-10) and temperatures (4-60^{o}C) with biological temperatures ranges and molten soft agar temperatures (~37^{o}C and 45^{o}C respectively) exerting limiting effects on Pln-423 absorption (Todorov and Dicks 2006).

Similarly, as a class IIa bacteriocin pediocin PA-1 was additionally analysed using the *E. coli* JE5505 overlay assay (Tominaga and Hatakeyama, 2006). Pediocin PA-1 exhibits similar resistance to temperature (100% activity after treatment 100^{o}C 1hr) and susceptibility to certain proteases (Daboura et al., 2009). Although, pediocin PA-1 is more alkaline sensitive than Pln-423 nonetheless PA-1 remains active across broad pH (2-12) (Daboura et al., 2009). Rescuing bioactive recombinant pediocin PA-1 has been exemplified in various *E. coli* expression systems, wherein a significant

portion of the peptide can be extracted from soluble fractions and shares close resemblance to the native peptide (Beaulieu et al., 2007).

Peptide RLL and Cathelicidin-BF demonstrated antimicrobial activity against *E. coli* JE5505 in both broth microdilution and agar diffusion assays. Contrary to this Pln-423 was active at 100µM in both broth and agar characterisation assays (**Figure 4.9 and Figure 4.16**). The sub-micromolar critical agar concentrations of Peptide RLL and Cathelicidin-BF, support the hypothesis that recombinant overexpression within this leaky *E. coli* host could lead to toxicity and/or the formation of inclusion bodies. In *E. coli* heterologous recombinant proteins can be expressed as insoluble aggregated intermediaries classified as inclusion bodies (Rosano and Ceccarelli, 2014). Protein aggregation typically commences when gene expression exceeds physiological levels, and the exposure of hydrophobic surfaces of desired recombinant protein can lead to folding intermediates (Mergulhão et al. 2005).

Various statistical models exist to classify the propensity of inclusion body formation from amino acid composition. Harrison *et al.*, (2000) developed such solubility model, which can be accessed via EMBOSS Pepstats **(Method 2.8).** Under Harrison *et al.*, (2000) model the following solubility percentiles were identified; Pln-423 (+94%), PA-1 (+89.7%), RLL (-731.5%), Cathelicidin-BF (-23.7%) and Plectasin (+95.7%). Exemplifying Gram-positive bacteriocins Pln-423 and PA-1 were highly soluble alongside Plectasin a fungi-sourced AMP. Peptide sequences demonstrating prediction values $\geq$ +90% are typically identified as highly soluble and are likely to be obtained from soluble fractions without the need for a solubility improving fusion tag (Harrison et al., 2000). Phage-display derived Peptide RLL was significantly more insoluble and prone to inclusion body formation under this model compared to the natural AMP Cathelicidin-BF and Plectasin.

Fundamentally, Pln-423 assimilates many of the quintessential absorption, solubility and sub-micromolar antimicrobial characteristics necessary to be an ideal candidate for overlay assay analysis against *Listeria innocua* strains (Guralp *et al.*, 2013). Pln-423 readily diffuses in agar, exhibits tolerance to extreme temperatures or pH and vitally remains relatively highly absorbed by target Gram-positive bacteria withstanding conditions.

4.5.5 The pitfalls and prizes of recombinant production in E. coli JE5505 expression systems

In spite of the *lpp* deletion, *E. coli* JE5505 exhibits relatively unimpaired growth kinetics **(Figure 4.1).** This strain is still sensitive and requires preparation as such in both broth and agar media. Especially, to support recombinant protein production. The Mueller-Hinton soft agar overlay approach utilised 1% glucose media supplementation, where appropriate, to reduce preferred sugar source depletion and inadvertent activation of recombinant peptide production. *E. coli* JE5505 can leach periplasmic enzymes into liquid culture medium (Kanamori et al., 1998). Banka et al., (2014) demonstrated *E. coli* JE5505 transformed with *pFLAG-CTS* encoding *B. subtilis* M015 hemicellulases, xylanase and β-xylosidase (<60kDa), were able to secrete encoded enzymes into lysogeny broth supernatant (~OD_{600nm} broths, 1mM IPTG induced then grow for <16hrs at 37^{o}C).

Orr et al., (2012) evaluated the bioprocessing capacity of *E. coli* JE5505 specifically for the recombinant production of penicillin G acylase under the control a recombinant plasmid pTrcKn (*pac2902, trc, promoter, AmpR, lacI*). 0.1mM IPTG LB media (NaCl 19g/L, Yeast extract 5 g/L and Tryptone g/L) culturing at 30^{o}C, 250rpm for 16hrs was employed. Cell growth was 0.73g for *E. coli* JE5505 and *E. coli* HB101 was quantified at 2.2g under the same culturing conditions (Orr et al., 2012). *E.coli*

JE5505 secretion efficiency (40%) was almost double that of *E. coli* HB101 in 0.1mM IPTD LB media. Recombinant penicillin G acylase activity from *E. coli* JE5505 extracellular fractions were significantly above that produced by the control strain.

Orr et al., (2012) denoted *E. coli* JE5505 carrying the *pTrcKnPAC2902* vector exhibited severely impaired growth in media with comparatively low osmolarity thus MOPS was supplemented into media to rescue growth retardation with improved buffering. Impaired growth of *E. coli* JE5505 in low-ionic strength media has been reported and is common for other *lpp* deletion mutants where the absence of divalent cations such as Ca2+ and Mg2+ can retard growth (Sonntag et al., 1978). The *lpp* deletion in *E. coli* JE5505 results in this strains hypersensitivity EDTA, cationic dyes and detergents (Hirota, 1977). Sensitivity towards cationic dyes in particular might extend negative consequences for the production of recombinant AMPs with high charges, such as Peptide RLL (+6) and Cathelicidin-BF (+7), which demonstrate activity against *E. coli* JE5505 in solution. Optimisation of the Mueller-Hinton media supplementation (buffering and cations) as well as optimising lower IPTG inductions might work towards facilitating achieving improved *E. coli* JE50505 broth cultivation and recombinant secretion efficiency (Choi and Lee, 2004; Mergulha˜o et al. 2005; Orr et al,, 2012).

Orr et al., (2012) was able to demonstrate ~90% recombinant PAC was produced by *E. coli* JE5505 cell factories was synthesised within ≤8hrs of IPTG induction. However, the growth of *E. coli* JE5505 *pSD3_Pln423_-BspQI_-gIII* at 30ºC or 37ºC for <18hrs in MHB supplemented with 0.5mM or 1mM IPTG did not seem to produce any detectable antimicrobial product when cell-free supernatants were spotted onto *L.innocua* lawn plates. This was additionally the case for testing *E. coli* JE5505 *pSD3_RLL_-BspQI_-gIII* IPTG supernatant against *S. Typhimurium.* Adapting the

approach of Todorov and Dicks (2006), cell-free IPTG inductions were utilised in a broth dilution assay alongside the required test strain. However, no inhibitory activity was identified similar to spotting of 5 µl, 10 µl and 50 µl of cell-free IPTG inductions against respective test species. Therefore, the potential of exploiting the *E. coli* JE5505 expression system in a screening approach in broth conditions was deemed to be likely fruitless. Nonetheless for the purpose of high-throughput screening herein, *E. coli* JE5505 recombinant expression and extracellular leakage of encoded peptides in an agar assay was most paramount. This study and previous reports have demonstrated semi-quantitative characterisation of class II bacteriocin AMPs (Pln-423 and PA-1) against several Gram-positive overlays. As aforementioned, Pln-423 and PA-1 are quintessentially apt for overlay assay screening, and partially this reflects both peptides sub-micromolar activity specifically towards Gram-positives and little to no inhibition of *E. coli* JE5505.

Considering this fact, screening $VNN_{15}(TTT)_1$ and NNK peptide randomised 16mer peptide libraries via the overlay assay was executed. Large and diverse peptide libraries provide near innumerable possibilities of combinatorial sequences, and this is further reflected in the variance of peptide properties both as; antimicrobial activity spectrum, solubility, agar diffusivity and compatibility with *E. coli* expression. The lacking identification of antimicrobial variants via the overlay assay screening of $VNN_1(TTT)_1$ and NNK peptide libraries likely reflects the complexity of the *E. coli* JE5505 expression within agar rather than the absence of antimicrobial candidates in each library. However, to further evidence this next generation phage display technologies were used as a secondary screening strategy for both $VNN_{15}(TTT)_1$ and NNK peptide libraries.

Chapter 5: Discovering and characterising novel AMPs identified from the Next-Generation Sequencing Phage Display screening of $VNN_{15}(TTT)_1$ and NNK 16 mer phage peptide libraries.

5.1 Introduction

Peptide phage display technologies are underpinned by five fundamental factors. Firstly (*i*) peptide libraries with substantial numbers of variants and superior diversity facilitate the increased likelihood of successfully isolating ligands with the desired phenotype (Kong et al., 2020). Secondly, (*ii)* the genotype of diverse peptide libraries (DNA or RNA) are reflected in the numerous genetically encoded library candidates, which are subsequently screened by phenotype for selective target binding affinity (López-Pérez et al., 2015). Thirdly, (*iii*) "phenotype information cycle"; the continuity of phenotype selection using iterative panning cycles manifests *in vitro* via phage replicative systems. In essence phage are utilised to facilitate biologically amplifying specific encoded library nucleic acids which are selectively enriched. Several successive iterations of enrichment can identify library peptide candidates with high selectivity for targets (*iv*). Nonetheless, utilising a singular round of panning combined with high-throughput sorting techniques (*i.e.* FACs) can sufficiently identify peptide ligands of interest (Zhang et al., 2011; Rebollo et al., 2014). Lastly, (*v*) phage display platforms are increasingly intertwined with Next Generation Sequencing (NGS) technologies. Exploiting NGS provides unparalleled elucidation of $>10^6$ sequences present in panning round outputs for binder identification (Grada and Weinbrecht, 2013; Shave et al., 2018).

Next generation phage display technologies have been integral to antibody discovery nonetheless, small peptide ligand isolation against small molecules, receptors and whole-cell epitopes targets have been routinely exemplified. In the extant of literature, combinatorial phage peptide libraries for novel AMP identification are $>10^9$ in size and have utilised both display on pIII or pVIII phage proteins (Bishop-hurley et al., 2005; Sainath Rao, Mohan and Atreya, 2013;Flachbartova et al., 2016).

Novel phage display derived AMPs are typically identified from screening repurposed commercially available combinatorial small peptide phage libraries. For example, Sainath Rao, Mohan and Atreya, (2013) screened the combinatorial Ph.D.-12mer phage peptide library (New England BioLabs) and identified RLLFRKIRRLKR (EC5, Peptide RLL), an arginine and lysine-rich AMP with activity against *E. coli* among other bacterial strains.

Interestingly, Pini *et al.,* (2005) constructed an NNK degenerately randomised 10-mer pIII phage peptide library (3+3) strategy; N": A/T/G/C and "K": G/T. Pini *et al.,* (2005) identified peptide QEKIRVRLSA, a novel AMP with broad-spectrum activity (≤107μM). QEKIRVRLSA preferentially exhibited higher potency against multidrug resistant Gram-negative bacteria; *P. aeruginosa* and *Enterobacteriaceae* family members in comparison to Gram-positive test species. Sainath Rao, Mohan and Atreya, (2013) and Pini *et al.,* (2005) both utilised whole-cell phage display methods for the isolation of novel antibacterial AMPs. Whole-cell phage display methods typically entail suspending phage library progeny (~10^{10} to 10^{12} plaque forming units, PFU) with ~10^{6-8} CFU/mL of target bacteria in phosphate buffering saline (PBS). The present chapter aims to identify novel AMPs against key bacterial pathogens of pigs by exploiting whole-cell phage display. Specifically, early hit candidates will be derived from 16 mer peptide libraries randomised with the conventionally exploited NNK approach or the semi-rationale "AMP biased" approach VNN_{15}+(TTT), "V" = A/G/C. The identification of novel antibacterial AMPs herein resides within the broader pursuit of the next generation antimicrobials to aid overcoming forecasted challenges of treatment resistant bacterial pathogens of both pigs.

5.2 Aims:

i. Apply a whole-cell phage display method for the isolation of library peptide ligands enriched against Gram-Negative and Gram-Positive bacteria model species *S. Typhimurium 4/74* and *S. suis P1/7* respectively.

ii. Implement a biopanning strategy and *in silico* next generation sequencing analysis pipeline suite to analyse species-specific and broad-spectrum enrichment of panning output peptides identified against bacteria targets (*i*).

iii. Characterise the antimicrobial susceptibility of key bacterial pathogens with porcine host specificity; *S. Typhimurium 4/74*, *S. suis P1/7* and *E. coli P433* against enriched library binders to verify antimicrobial phenotypes.

iv. Implement *i-iii* for peptide libraries ($>10^9$ size) generated under the NNK and $VNN_{15}+(TTT)_1$ randomisation schemes to assess whether a semi-rational approach to bias amino acid distribution towards that identified in natural AMPs can improve the antimicrobial nature or number of novel AMPs identified.

5.3 ISOLATION OF NOVEL PHAGE-DISPLAY DERIVED ANTIMICROBIAL PEPTIDES AGAINST BACTERIAL PATHOGENS OF PIGS

5.3.1 Whole-cell phage display screening of degenerately randomised phage peptide libraries

Chapter 1 quality control via Next Generation Sequencing (NGS) analysis revealed libraries of high diversity and quality; *E. coli* TG1 libraries constructed herein have an estimated diversity of 1.83 x 10^9 and 2.12 x 10^9 for VNN15+$(TTT)_1$ and NNK libraries respectively (Rebollo et al., 2014). A 3+3 helper bacteriophage display system (Ex-phage, pIII vector system) was exploited. To begin, phage propagations of each library were generated, and the initial phage input titrations were 5 x 10^{12} PFU/mL and 7x10^{11} PFU/mL for the NNK and $VNN_{15}(TTT)_1$ peptide phage libraries respectively. The three-round panning strategy outlined in **Method 2.11.1** was utilised. 10 replicates were conducted for the round 3 panning sample sets, resulting in 80 individual panning output samples. Refraining from pooling, each individual round 3 panning sample was recovered using *E. coli* TG1. Overnights were utilised to firstly generate glycerol stocks and to extract encoded library phagemid vectors via miniprepping. Qubit quantified phagemid extractions were subjected to NGS preparation; two successive rounds of PCR amplification were conducted with 80 different barcoding primers which additionally introduce necessary Ion torrent NGS sequencing compatibility regions (**Method 2.12**). For the ease of explanation, the nomenclature of NGS and panning strategy implemented will be signified as follows; LIB_XX2_XX3 wherein LIB refers to $VNN_{15(}TTT)_1$ or NNK, XX2 and XX3 are the species panned against in round 2 and 3 respectively, "TY" equates to *S. Typhimurium* 4/74 whilst "SS" refers to *S. suis* P1/7.

5.3.2 E. coli JE5505 overlay assay screening of phage display panning outputs

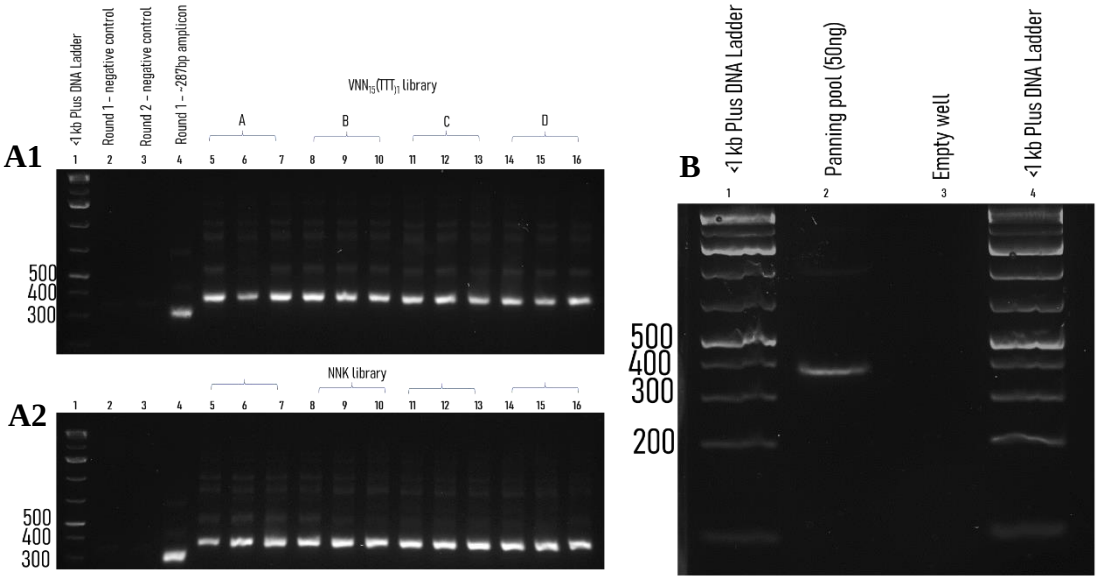
The iterative cycling of panning enriches certain library candidates in phage display peptide libraries however, by the nature of the biological system, multiple phenotypes can be selected for whilst screening. TopN and Z score analysis identified enriched library peptides which exhibited an *S. suis* P1/7 > *S. Typhimurium* 4/74 bias in frequency ratio and reproducibility across panning replicates **(see 5.3.4 onwards).** The lucrative prospect of possible Gram-positive biased AMPs discovery therefore, reinstated the utility of applying the *E. coli* JE5505 overlay. *E. coli* TG1 recovery for VNN_TY2_TY3, VNN_SS2_SS3 and NNK library panning equivalents, were miniprepped and *SpeI* digested then transformed into *E. coli* JE5505. Round three output peptide phage populations were screened in the 0.7% Mueller-Hinton leaky *E. coli* JE5505 overlay assay. However, no zones of inhibition were identified when screening panning *E. coli* JE5505 sub libraries against *S. Typhimurium* 4/74, *S. suis* P1/7 and *E. coli* (ETEC) P433 (data not shown).

5.3.3 Next generation sequencing preparation for panning output samples

NGS PCR Round 1 amplification for both VNN_{15}+$(TTT)_1$ and NNK of extracted DNA from panning samples produce a ~287bp desired PCR amplicon which is later extended by ~60bp in Round 2 PCR amplification. **Figure 5.1** depicts the agarose gel electrophoresis for PCR amplicons generated during NGS sample preparation. Quantified equalised 50ng pool of the 80 barcoded panning samples was re-run on agarose to verify the presence of the desired single desired band (**Figure 5.1).** 200ng of the equalised pool was sent for Next Generation Sequencing via the Ion Torrent platform.

Figure 5.1 Next Generation Sequencing (Ion Torrent) round 3 panning sample preparation

Representative agarose gel electrophoresis for the NGS preparation for panning output samples panning VNN_{15}+$(TTT)_1$ and NNK; *gIII (*present) peptide phage libraries, **A1** and **A2** respectively. A control round 1 sample fromVNN_{15}+$(TTT)_1$ screened against *S. suis*in all rounds was utilised for comparisom (**A1**/**A2)**. **A1/2** A = LIB_SS2_SS3, B = LIB_SS2_TY3, C = LIB_TY2_TY3, D = LIB_TY2_SS3. The first phase generates a ~287bp desired PCR amplicon, and the second PCR amplification increases desired PCR amplicons by ~60bp. Negative controls and <1kb DNA Ladder (NEB) were utilised. Negative controls contained no DNA template to ensure no unspecific amplification occurred. **(B)** 50ng pool of NGS PCR amplified and barcoded 80 panning sample sets was analysed to re-verify the final sample sent for sequencing contained the desired amplicon.



5.3.4 Phage display next generation sequencing analysis pipeline 1; Frequency analysis TopN

Figure 5.2 depicts the Top50 enriched binders identified with replicate reproducibility in the "positive dataset" $p_{dataset}$ = ≥ 40%, and this is shown for LIB_SS2_SS3 vs LIB_SS2_TY2 $p/n_{dataset}$ coupling for both $VNN_{15}(TTT)_1$ and NNK. To corroborate Z score derived "species specific" enriched binders. The Top*N* manual reverse analysis additionally involved analysing the negative datasets wherein the original input phage library was panned against different test species in all three rounds *e.g.*, comparing NKK_SS2_SS3 to NKK_TY2_TY3. In this context the peptides selected adhered to the cut off rule; replicate reproducibility; $p_{dataset}$ = ≥ 40% and

$n_{dataset}$ = ≤ 20%; i.e. Top*N* peptide observed $p_{dataset}$ = ≥ 4 / 10 in the positive replicates and $n_{dataset}$ = ≤ 2 / 10 in the negative replicates.

Figure 5.2A Selection of Top50 enriched binders originating from screening $VNN_{15}(TTT)_1$ phage library vs *S. suis* round 2 output phage against both test species; *S. suis* P1/7 and *S. Typhimurium* 4.74 in the third panning round.

The identification of ligands across panning replicates (10) is demonstrated for the positive panning dataset (VNN_SS2_SS3) wherein, $p_{dataset}$ = ≥ 40% replicate reproducibility and negative panning dataset (VNN_SS2_TY3) $n_{dataset}$ = was ≥ 0% were the cut offs applied. This comparison allows the identification of enrichment discrepancies when the final panning round test species differs. Peptide SYHRLLTYFKCSFSYW (red dashed circle) remains enriched in all 10 third round of panning against *S. suis P1/7*. Whereas the ligand is lost when the same round 2 output phage was panned against *S. Typhimurium 4/74*. Contrastingly, ligands such as ANVLSFSQRHLGFRCI (blue dashed circle) are Top50 peptide highly reproduced against both test species; $p/n_{dataset}$ reproducibility: 100%/100%.

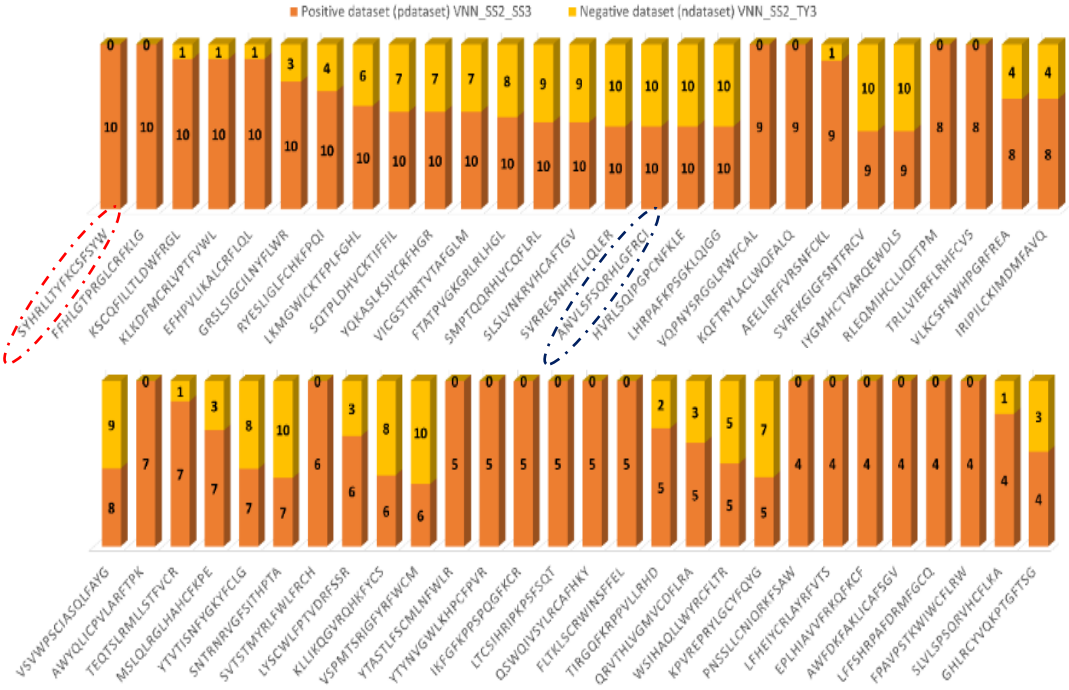


Figure 5.2B Selection of Top50 enriched binders originating from screening NNK phage library vs *S. suis* round 2 output phage against both test species; *S. suis* P1/7 and *S. Typhimurium* 4.74 in the third panning round.

The identification of ligands across panning replicates (10) is demonstrated for the positive panning dataset (NNK_SS2_SS3) wherein, $p_{dataset}$ = ≥ 40% replicate reproducibility and negative panning dataset (NNK_SS2_TY3) $n_{dataset}$ = was ≥ 0% were cut offs applied. In comparison to the $VNN_{15}(TTT)_1$ equivalent **(Figure 5.2A),** the NNK derived Top50 peptides tended to be highly reproduced across replicates for both test species (e.g., more peptides are identified which are enriched against both test species). Whereas $VNN_{15}(TTT)_1$ exhibited numerous peptides which were bias in their enrichment reproducibility for *S. suis* P1/7 > *S. Typhimurium* 4/74.

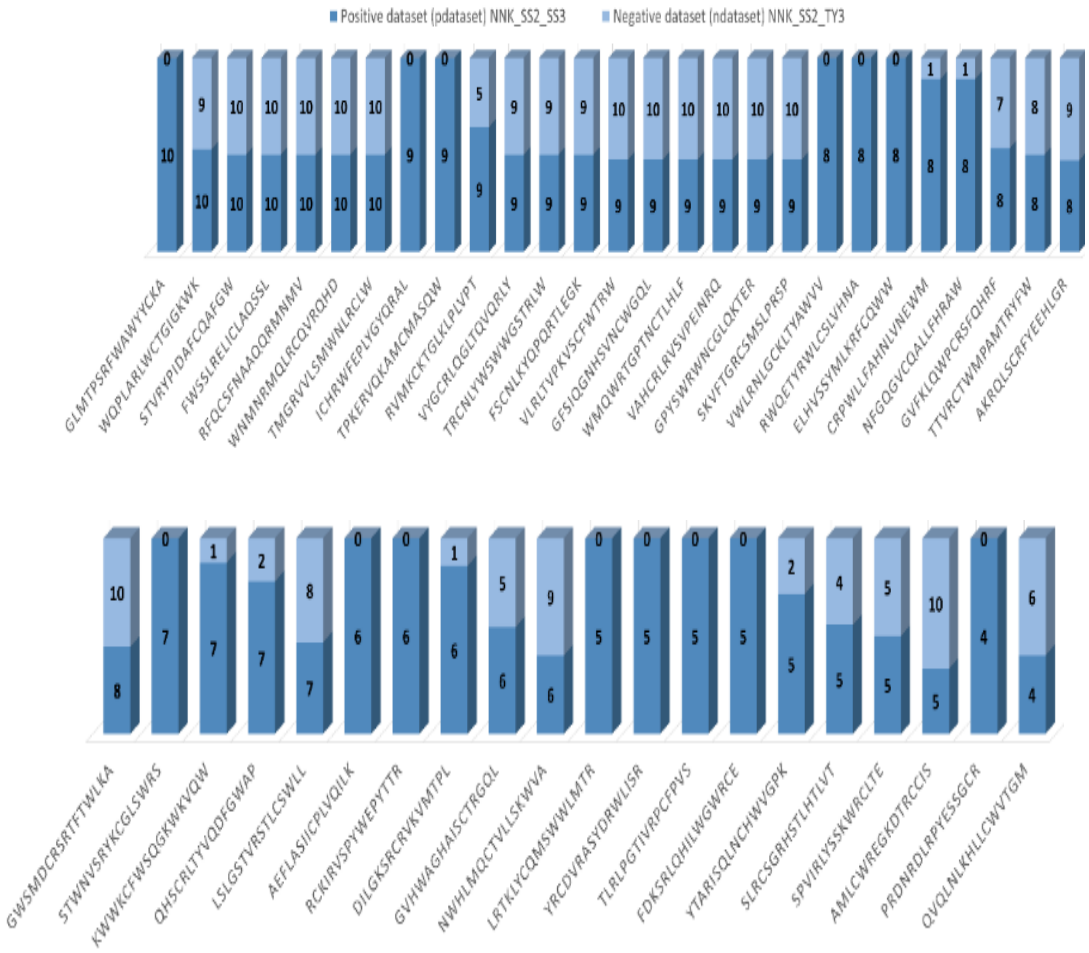


Table 5.1 to **Table 5.4** summarise the 42 phage display derived peptides identified as relatively frequent and enriched binders from both VNN_{15}+$(TTT)_1$ and NNK libraries via the Top50 pipeline. Noteworthy to state the majority of these peptides were identified via the $p/n_{datasets}$ = ≥70% cut off, and exhibit enrichment across both test species (**Figure 5.2**). For both libraries, panning against *S. suis* P1/ in round two, followed by the two-counter species screening in round three identified more binders

with enrichment reproducibility across both test species (**Table 5.1** and **Figure 5.7).** However, the application of this trend was less apparent when *S. Typhimurium* was the round two test species (**Figure 5.2, Table 5.1** and **Figure 5.7).**

Table 5.1. Top50 enriched binders: ≥40% replicate reproducibility identified when panning VNN_{15}+$(TTT)_1$ and NNK respectively against *S. Typhimurium* 4/74 and *S. suis*P1/7

Degenerate library screening parameter [a]:	*1: NNK_S. suis*		*2: NNK_ S. Typhimurium*		*3: VNN_S. suis*		*4: VNN_ S. Typhimurium*	
Test species[b]:	Screen X	Screen Y	Screen X	Screen Y	Screen X	Screen Y	Screen X	Screen Y
FWSSLRELICLAQSSL	E	F	H					
GGRVRCEIIMRPFGSV							D	C
KFSSRVRLSDREFYCT							D	C
GFSIQGNHSVNCWGQL	E	F	H					
ANVLSFSQRHLGFRCI					A	B		
FLLKVQKSGNMTATKE							D	C
FSCNLKYQPQRTLEGK	E	F						
GPYSWRWNCGLQKTER	E	F						
HVRLSQIPGPCNFKLE					A	B		
IYGMHCTVARQEWDLS					A	B		
LHRPAFKPSGKLQIGG					A	B		
LSKYTDTRCFARFRYW							D	C
RFQCSFNAAQQRMNMV	E	F	H					
RTIQVHSFHCPIFDIV							D	C
SKVFTGRCSMSLPRSP	E	F	H					
SLSLVNKRVHCAFTGV					A	B		
SMPTQQRHLYCQFLRL					A	B		
SPLKHMQHSFSCFNQR							D	C
SSVHIGLLSQHSFSCR							D	C
STVRYPIDAFCQAFGW	E	F	H					
SVRFKGIGFSNTFRCV					A	B		
SVRRESNHKFLLQLER					A	B		
TMGRVVLSMWNLRCLW	E	F	H					
TPKERVQKAMCMASQW	E*		H*					
TRCNLYWSWWGSTRLW	E	F						
TVCNMKIAPHQSFRVL							D	C
VAHCRLRVSVPEINRQ	E	F						

Key:

[a] Degenerate library scheme; $VNN_{15}(TTT)_1$, "VNN" or NNK; Screen X = S.Typhimiurm and Y = S. suis
- 2 panning rounds *S. suis* p1/7; NNK **(1),** $VNN_{15}(TTT)_1$ **(3)**
- 2 panning rounds *S. Typhimurium* 4/74; NNK **(2),** $VNN_{15}(TTT)_1$ **(4)**

[b] Round 3 panning test species, Screen **X** (*S. suis* p1/7), Screen **Y** (*S. Typhimurium* 4/74)

[c] ≥40% replicate reproducibility in the Top50 peptides (10 replicates per round 3 screening strategy). Any peptide identified across three independent datasets (grey fill)

Set i) VNN_SS2_SS3 **(A);**VNN_SS2_TY3 **(B);** VNN_TY2_TY3 **(C);**VNN_TY2_SS3**(D)**

Set ii) NNK_SS2_SS3 **(E)**, NNK_SS2_TY3 **(F)**, NNK_TY2_TY3 **(G**, *absent***)**, NNK_TY2_SS3 **(H)**

(X)* control species specific peptide from TopN

Table 5.2 Summary of the physiochemical and charged residue composition of Top50 enriched VNN_{15}+$(TTT)_1$ and NNK library binders:

Peptide sequence (16 mers)	Lib [a]	Species specific? (N/Y)	Charge	Hydrophobicity (%) [b]	*pI* [c]	Charged residues [d]
FWSSLRELICLAQSSL	NNK1	N	0.00	56.25	5.84	1
GGRVRCEIIMRPFGSV	VNN1	N	2.00	68.75	10.28	3
KFSSRVRLSDREFYCT	VNN2	N	2.00	37.50	9.27	4
GFSIQGNHSVNCWGQL	NNK2	N	0.50	56.25	6.97	1
ANVLSFSQRHLGFRCI	VNN3	N	2.50	56.25	10.37	3
FLLKVQKSGNMTATKE	VNN4	N	2.00	56.25	10.50	3
FSCNLKYQPQRTLEGK	NNK3	N	2.00	43.75	9.23	3
GPYSWRWNCGLQKTER	NNK4	N	2.00	50.00	9.27	3
HVRLSQIPGPCNFKLE	VNN5	N	1.50	56.25	8.03	3
IYGMHCTVARQEWDLS	VNN6	N	-0.50	56.25	5.43	2
LHRPAFKPSGKLQIGG	VNN7	N	3.50	62.50	11.82	4
LSKYTDTRCFARFRYW	VNN8	N	3.00	50.00	10.00	4
RFQCSFNAAQQRMNMV	NNK5	N	2.00	50.00	10.37	2
RTIQVHSFHCPIFDIV	VNN9	N	1.00	56.25	7.16	3
SKVFTGRCSMSLPRSP	NNK6	N	3.00	50.00	11.48	3
SLSLVNKRVHCAFTGV	VNN10	N	2.50	56.25	9.67	3
SMPTQQRHLYCQFLRL	VNN11	N	2.50	50.00	9.33	3
SPLKHMQHSFSCFNQR	VNN12	N	3.00	37.50	9.67	4
SSVHIGLLSQHSFSCR	VNN13	N	2.00	43.75	8.06	3
STVRYPIDAFCQAFGW	NNK7	N	0.00	68.75	5.77	1
SVRFKGIGFSNTFRCV	VNN14	N	3.00	56.25	11.48	3
SVRRESNHKFLLQLER	VNN15	N	2.50	31.25	11.38	5
TMGRVVLSMWNLRCLW	NNK8	N	2.00	68.75	10.37	2
TPKERVQKAMCMASQW	NNK9	Y	2.00	50.00	9.51	3
TRCNLYWSWWGSTRLW	NNK10	N	2.00	56.25	9.32	2
TVCNMKIAPHQSFRVL	VNN16	N	2.50	56.25	9.67	3
VAHCRLRVSVPEINRQ	NNK11	N	2.50	50.00	10.28	4

Key;

[a] Degenerate library scheme used for randomising library 16mer peptide regions ; $VNN_{15}(TTT)_1$, "VNN" or NNK; Top50 enriched binders: ≥40% replicate reproducibility (10 replicates in total used).

[b] Percentage of hydrophobic residues; "AMP Amino acid grouping"; G, A, V, L, I, P, F, M, W, C, and Y.

[c] *pI* = Isoelectric point

[d] = AMP Amino acid grouping cationic (basic) charged residues; K, R and H only.

Table 5.3. Top50 enriched binders: ≥40% replicate reproducibility identified when panning VNN15+(TTT)1 and NNK respectively against *S. Typhimurium* 4/74 and *S. suis*P1/7 continued

Degenerate library screening parameter :	1: NNK_S. suis		2: NNK_ S. Typhimurium		3: VNN_S. suis		4: VNN_ S. Typhimurium	
Test species:	Screen X	Screen Y	Screen X	Screen Y	Screen X	Screen Y	Screen X	Screen Y
VLRLTVPKVSCFWTRW	E	F	H					
VQKAVVYRCNLRFSDL							D	C
VYGCRLQGLTQVQRLY	E	F						
WMQWRTGPTNCTLHLF	E	F						
WNMNRMQLRCQVRQHD	E	F						
WQPLARLWCTGIGKWK	E	F	H					
WVRCNLQVHGLSFLWR							D	C
GVFKLQWPCRSFQHRF			H	G				
LSLGSTVRSTLCSWLL			H	G				
RNEMRSFKCFLGSVRE			H					
SPISMPWWIWCTVLFK			H					
STFPVSARWMHMLRCK			H					
TTVRCTWMPAMTRYFW			H					
AIKMRGPLHRCNFAVP							D	C
AMLCWREGKDTRCCIS		F	H					

Key: Any peptide was identified across three independent datasets (grey fill). Screen X = S.Typhimiurm and Y = S. suis
Set i) VNN_SS2_SS3 **(A);**VNN_SS2_TY3 **(B);** VNN_TY2_TY3 **(C);**VNN_TY2_SS3**(D)**
Set ii) NNK_SS2_SS3 **(E)**, NNK_SS2_TY3 **(F)**, NNK_TY2_TY3 **(G)** NNK_TY2_SS3 **(H)**

Table 5.4. Summary of the physiochemical and charged residue composition of Top50 enriched VNN_{15}+$(TTT)_1$ and NNK library binders continued

Peptide sequence (16 mers)	Lib [a]	Species specific? (N/Y)	Charge	Hydrophobicity (%) [b]	*pI* [c]	Charged residues [d]
VLRLTVPKVSCFWTRW	NNK12	N	3	62.5	11.48	3
VQKAVVYRCNLRFSDL	VNN17	N	2	56.25	9.27	3
VYGCRLQGLTQVQRLY	NNK13	N	2	62.5	9.15	2
WMQWRTGPTNCTLHLF	NNK14	N	1.5	56.25	8.03	2
WNMNRMQLRCQVRQHD	NNK15	N	2.5	37.5	10.28	4
WQPLARLWCTGIGKWK	NNK16	N	3	68.75	10.79	3
WVRCNLQVHGLSFLWR	VNN18	N	2.5	62.5	10.37	3
GVFKLQWPCRSFQHRF	NNK17	N	3.5	56.25	11.48	4
LSLGSTVRSTLCSWLL	NNK18	N	1	56.25	8.00	1
RNEMRSFKCFLGSVRE	NNK19	N	2	43.75	9.66	4
SPISMPWWIWCTVLFK	NNK20	N	1	75	8.00	1
STFPVSARWMHMLRCK	NNK21	N	3.5	56.25	11.48	4
TTVRCTWMPAMTRYFW	NNK22	N	2	62.5	9.32	2
AIKMRGPLHRCNFAVP	VNN19	N	3.5	68.75	11.48	4
AMLCWREGKDTRCCIS	NNK23	N	1	56.25	7.68	3

Key; [a] Degenerate library scheme used for randomising library 16mer peptide regions ; $VNN_{15}(TTT)_1$, "VNN" or NNK; Top50 enriched binders: ≥40% replicate reproducibility (10 replicates in total used).
[b] Percentage of hydrophobic residues; "AMP Amino acid grouping"; G, A, V, L, I, P, F, M, W, C, and Y. [c]*pI* = Isoelectric point, [d] = AMP Amino acid grouping cationic (basic) charged residues; K, R and H only.

55% (23/42) and 45% (19/42) of the Top50 shortlisted peptides were obtained from screening the NNK and VNN_{15}+$(TTT)_1$ libraries respectively against *S. Typhimurium* 4/74 and *S. suis* P/17. The slight increase in Top50 peptides selected from the NNK library likely represents the slightly improved output of binders observed against *S. Typhimurium* 4/74 in comparison to VNN_{15}+$(TTT)_1$ library panning. This is exemplified by the occurrence of NNK peptide binders which occur across triplet replicates in **Table 5.1** and **Figure 5.2.** Specifically, this includes eight peptides FWSSLRELICLAQSSL; "Pep_NNK/1", GFSIQGNHSVNCWGQL; "Pep_NNK/2", RFQCSFNAAQQRMNMV;"Pep_NNK/5",SKVFTGRCSMSLPRSP; "Pep_NNK/6" ,STVRYPIDAFCQAFGW;"Pep_NNK/7", TMGRVVLSMWNLRCLW; "Pep_NNK/8", VLRLTVPKVSCFWTRW;"Pep_NNK/12" and WQPLARLWCTGIGKWK; "Pep_NNK/16"

Only 34% of the Top50 NNK-derived peptides possessed this triplet enrichment identification. Pep_NNK/1, 2, 5, 6, 7, 8, 12 and 16 are seemingly enriched in samples originating from *S. suis* round two output phage (NNK_SS2), both after being panned against *S. suis* (3^{rd} iterative panning cycle). Intriguingly a "bounce back" to Top50 ranking is observed for the aforementioned peptides when NNK_TY2 output phage was panned against *S. suis* (NNK_TY2_SS3). However, these peptides vanish from Top50 enrichment classification when NNK_TY2 was panned against *S. Typhimurium* 4/74*;* NNK_TY2_TY3 (3^{rd} iterative panning cycle). The loss and gain of Top50 ranking indicates, peptides were maintained in the pool <2 panning rounds whereupon biased binding enrichment for *S. suis* P1/7 and a lesser extent *S. Typhimurium* 4/74 constrains representation accordingly in round 3.

VNN_{15}+$(TTT)_1$ Top50 peptides included were all identified as duplicate couplings from the same round 2 output phage (**Table 5.1 and Figure 5.2**). The

absence of the "bounce back" might therefore indicate either *i*) two iterative rounds of panning of $VNN_{15}+(TTT)_1$ library against *S. suis* was sufficient for species-specific binder enrichment or alternatively as mentioned previously $VNN_{15}+(TTT)_1$ library contains less library candidates with *S. Typhimurium* 4/74 binding affinity. A combination of the two might be at play, especially as compared to *S. Typhimurium* panning; $VNN_{15}+(TTT)_1$ library was able to identify several peptides with enrichment reproducibility against *S. suis* when the cut-off for Top50 analysis was $p_{dataset} = \geq 40\%$ and $n_{dataset} = \leq 20\%$.

5.3.5 Phage display next generation sequencing analysis pipeline 2; Z scores

Figure 5.3 demonstrates the number of Z score ≥2 cut-off peptides identified across the eight panning datasets which were observed in ≥40% of $p_{datasets}$ sample set replicates. The $VNN_{15}+(TTT)_1$ panning subsets tend to produce more Z score ≥2 cut-off peptides binders with reproducibility $p_{datasets} = \geq 90\%$ of replicates and this was primarily in Gram-positive *S. suis* P1/7 centred panning samples (**Figure 5.3)**. In most instances when *S. Typhimurium* 4/74 was either used for round 1/2 panning or the counter test species in round 3 this reduced the observed number of peptides with Z score: ≥cut-off in comparison to *S. suis* (**Figure 5.3**). For instance, VNN_SS2_SS3 vs VNN_SS2_TY3 exemplifies this decline. Whereas NNK_TY2_TY3 and NNK_TY2_SS3 the decline is markedly reversed, with panning against *S. suis* P1/7 round 3 test species improving the number of peptides ≥ Z score cut-off with high reproducibility (**Figure 5.3)**. VNN_SS2_SS3 and VNN_TY2_SS3 provide ~70% of the Z score: ≥cut-off peptides found in 90-100% of positive sample replicates, and the dominance of peptide representation across replicates is demonstrated in (**Figure 5.3**). Specifically, when panning was completed for three rounds against *S. suis* P1/7 the VNN_SS2_SS3 dataset identified ~8.25 times : ≥cut-off peptides enriched across all

10 replicates than the NNK library equivalent; NNK_SS2_SS3. Reverse analysis validated peptides identified from LIB_XX2_XX3 vs negative pool outputs with Z score: ≥cut-off were indeed not present in respective designated negative pool control. Numerous peptides in **Table 5.5 and Figure 5.3** however, indicated species specific binder enrichment in both the positive dataset and comparison control with the same test species panned in round 3.

Figure 5.3 VNN_{15}+$(TTT)_1$ and NNK derived library peptides enriched over three rounds of panning and possess Z scores ≥ 2 and ≥40% positive dataset sample reproducibility.

The number of library peptides identified in various LIB_XX2_XX3 panning samples are represented when this dataset was the "positive dataset"; $VNN_{15}(TTT)_1$ library (**A**) and NNK **(B)**. All peptides have Z scores ≥ 2 ±SD and were anlaysed for their presence in 10 respective positive dataset samples (**Figure 5.9**). Peptides identified in 4/10 panning samples are depicted which is referred herein to as; ≥40% positive dataset sample reproducibility. The number of peptides increases as reproducibility falls, and only 13.4% of peptides were identified with >90% reproducibility.

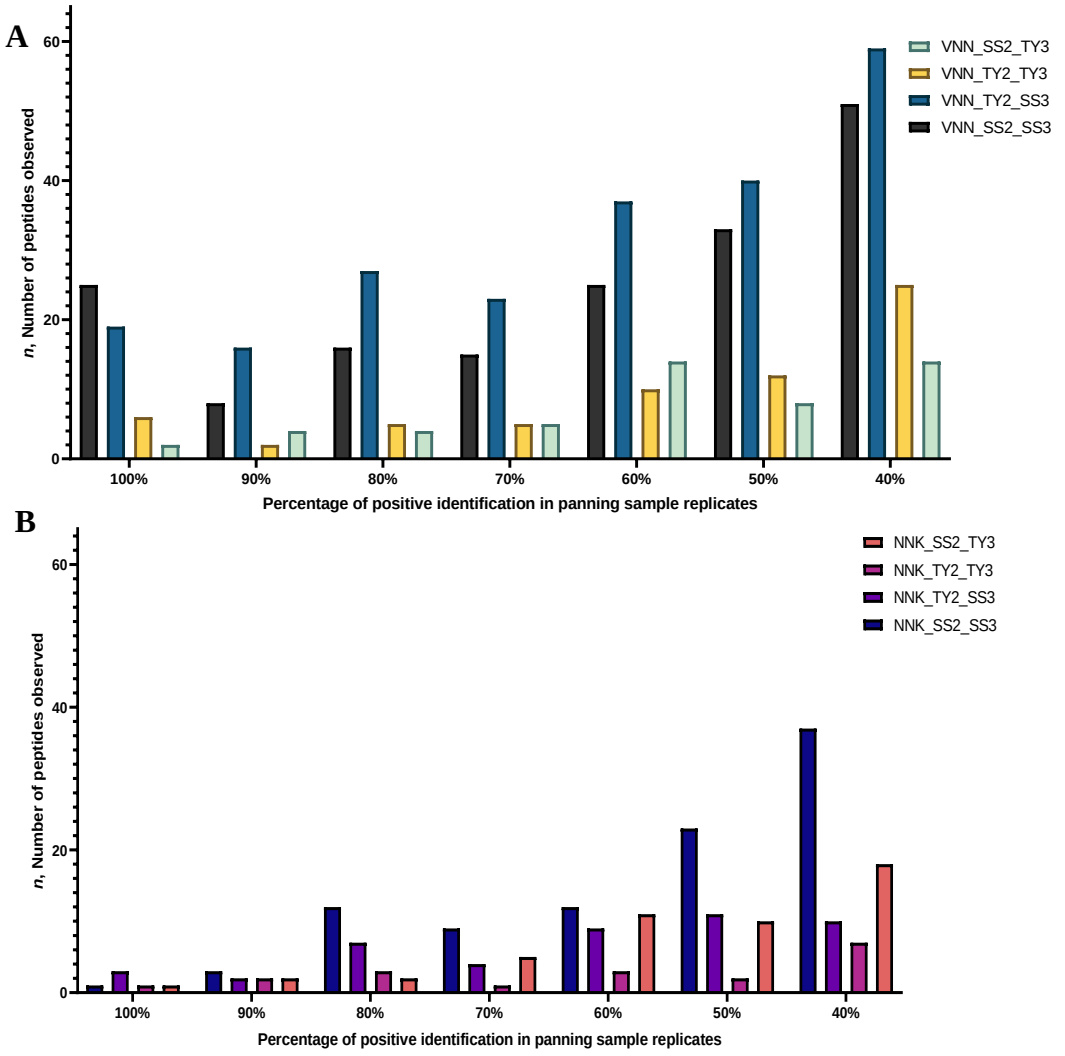


Table 5.5 Quantitative and qualitative presentation of ranked peptides with Z scores ≥2 identified in $VNN_{15}(TTT)_1$ and NNK panning sample sets against *S. suis*P1/7 and *S. Typhimurium* 4/74

Peptides with Z scores ≥2 identified when panning $VNN_{15}(TTT)_1$ and NNK phage libraries *S. suis*P1/7 and *S. Typhimurium* 4/74					
		Dataset identification			
Z scores ≥ 2- 2.5 enriched binders: ≥90% replicate reproducibility [c]:	KSCQFILLTLDWFRGL	A (Z: 4.50, SD± 1.16)			
	KLKDFMCRLVPTFVWL	A (Z: 4.43 SD± 0.76)	D (Z: 3.83, SD± 0.64)		
	SYHRLLTYFKCSFSYW	A (Z: 4.62, SD± 1.06)	D (Z: 3.23, SD± 0.57)		
	AWYQLICPVLARFTPK	A (Z: 5.18, SD± 0.80)	D (Z: 5.11, SD± 1.07)		
	AWFDKFAKLICAFSGV	A (Z: 4.23, SD± 0.70)			
	KQFTRYLACLWQFALQ	A (Z: 4.20, SD± 0.35)			
	YQKASLKSIYCRFHGR	A (Z: 5.61, SD± 1.27)			
	TRLLVIERFLRHFCVS	A (Z: 4.37, SD± 1.03)			
	AWSEYLCWVLRYFPSQ	A (Z: 3.63, SD± 0.46)	D (Z: 1.60, SD± 0.67)		
	APFTALCRLFEIFRSA	A (Z: 3.97, SD± 0.79)			
	VQPNYSRGGLRWFCAL	A (Z: 3.73, SD± 0.69)			
	GRSLSIGCILNYFLWR	A (Z: 3.84, SD± 0.79)			
	FLTKLSCRWINSFFEL	A (Z: 3.77, SD± 0.48)			
	KSFLSRWLCGIMFIGV	A (Z: 3.45, SD± 0.62)			
	FLARFTTPLRCLFINV	A (Z: 3.87, SD± 0.53)	D (Z: 2.28, SD± 0.86)		
	FLDNPFVSLIRCFIAR	A (Z: 3.50, SD± 0.44)			
	TWPQSINRWFCIFRGI	A (Z: 3.29, SD± 0.62)			
	GSVLSVMCRAWAFIAA	A (Z: 3.69, SD± 0.43)			
	SVTSTMYRLFWLFRCH	A (Z: 4.03, SD± 1.09)			
	KRNNVALTIVCRFLKW	A (Z: 4.02, SD± 0.60)			
	NAESRRPQWLHWFPLC	A (Z: 3.54, SD± 0.65)	D (Z: 1.47, SD± 1.11)		
	FFHLGTPRGLCRFKLG	A (Z: 4.62, SD± 1.43)			
	EFHPVLIKALCRFLQL	A (Z: 4.10, SD± 0.86)			
	SFSESISGWIRKFLCS	A (Z: 3.40, SD± 0.78)			
	ISSFWQSSACRTFGAL	A (Z: 3.83, SD± 0.72)			
	YTYNVGWLKHPCFPVR	A (Z: 3.66, SD± 1.08)			
	DQRWFSSYLCKAFTSR	A (Z: 3.00, SD± 0.66)			
	KLLGKECVVRFPFPNR	A (Z: 3.61, SD± 0.97)			
	FPAVPSTKWIWCFLRW	A (Z: 3.58, SD± 1.13)			
	TKPSRGFAREQIEKIL	A (Z: 3.84, SD± 0.68)			
	SIVGMITRLYCNFVSG	A (Z: 2.93, SD± 0.56)	D (Z: 1.95, SD± 0.53)		
	AEELIRFFVRSNFCKL	A (Z: 3.72, SD± 1.24)			

Key:

[c] ≥90% replicate reproducibility in the Z score peptides (10 replicates per round 3 screening strategy). The Z score for the specific dataset is stated as (Z: X.XX, SD±X.XX), wherein "Z" refers to Z score, and "SD" is the standard deviation across replicates

Set i) VNN_SS2_SS3 **(A)**, VNN_SS2_TY3 **(B)**, VNN_TY2_TY3 **(C)**, VNN_TY2_SS3 **(D)**

Set ii) NNK_SS2_SS3 **(E)**, NNK_SS2_TY3 **(F)**, NNK_TY2_TY3 **(G**, *absent***)**, NNK_TY2_SS3 **(H)**

Table 5.6. Physiochemical properties and relevant amino acid compositional characteristics of Z score identified peptide binders.

Characteristics of Z-score identified panning peptide ligands.					
Peptide sequence (16 mers)	**Lib** [a]	**Charge**	**Hydrophobicity (%)** [b]	***pI*** [c]	**Charged residues** [d]
KSCQFILLTLDWFRGL	VNN20	1.0	62.50	8.00	2
KLKDFMCRLVPTFVWL	VNN21	2.0	68.75	9.51	3
SYHRLLTYFKCSFSYW	VNN22	2.5	56.25	9.04	3
AWYQLICPVLARFTPK	VNN23	2.0	75.00	9.28	2
AWFDKFAKLICAFSGV	VNN24	1.0	75.00	8.00	2
KQFTRYLACLWQFALQ	VNN25	2.0	62.50	9.28	2
YQKASLKSIYCRFHGR	VNN26	4.5	50.00	10.45	5
TRLLVIERFLRHFCVS	VNN27	2.5	56.25	10.28	4
AWSEYLCWVLRYFPSQ	VNN28	0.0	68.75	5.84	1
APFTALCRLFEIFRSA	VNN29	1.0	68.75	8.00	2
VQPNYSRGGLRWFCAL	VNN30	2.0	68.75	9.32	2
GRSLSIGCILNYFLWR	VNN31	2.0	68.75	9.32	2
FLTKLSCRWINSFFEL	VNN32	1.0	56.25	8.00	2
KSFLSRWLCGIMFIGV	VNN33	2.0	75.00	9.67	2
FLARFTTPLRCLFINV	VNN34	2.0	68.75	10.37	2
FLDNPFVSLIRCFIAR	VNN35	1.0	68.75	8.00	2
TWPQSINRWFCIFRGI	VNN36	2.0	62.50	10.37	2
GSVLSVMCRAWAFIAA	VNN37	1.0	81.25	8.00	1
SVTSTMYRLFWLFRCH	VNN38	2.5	56.25	9.33	3
KRNNVALTIVCRFLKW	VNN39	4.0	56.25	11.65	4
NAESRRPQWLHWFPLC	VNN40	1.5	56.25	8.03	3
FFHLGTPRGLCRFKLG	VNN41	3.5	68.75	11.48	4
EFHPVLIKALCRFLQL	VNN42	1.5	68.75	8.03	3
SFSESISGWIRKFLCS	VNN43	1.0	50.00	8.00	2
ISSFWQSSACRTFGAL	VNN44	1.0	56.25	8.00	1
YTYNVGWLKHPCFPVR	VNN45	2.5	68.75	9.13	3
DQRWFSSYLCKAFTSR	VNN46	2.0	43.75	9.27	3
KLLGKECVVRFPFPNR	VNN47	3.0	62.50	10.77	4
FPAVPSTKWIWCFLRW	VNN48	2.0	75.00	9.67	2
TKPSRGFAREQIEKIL	VNN49	2.0	43.75	10.77	4
SIVGMITRLYCNFVSG	VNN50	2.0	68.75	7.99	1
AEELIRFFVRSNFCKL	VNN51	1.0	56.25	8.00	3

Key; [a] Degenerate library scheme used for randomising library 16mer peptide regions ; $VNN_{15}(TTT)_1$, "VNN" or NNK; Top50 enriched binders: ≥40% replicate reproducibility (10 replicates in total used).
[b] Percentage of hydrophobic residues; "AMP Amino acid grouping"; G, A, V, L, I, P, F, M, W, C, and Y.
[c] *pI* = Isoelectric point
[d] = AMP Amino acid grouping cationic (basic) charged residues; K, R and H only.

Table 5.7. Quantitative and qualitative presentation of ranked peptides with Z scores ≥2 identified in $VNN_{15}(TTT)_1$ and NNK panning sample sets against *S. suis*P1/7 and *S. Typhimurium* 4/74 continued

Peptides with Z scores ≥2 identified when panning VNN15(TTT)1 and NNK phage libraries *S. suis*P1/7 and *S. Typhimurium* 4/74 continued					
		Dataset identification			
	ICHRWFEPLYGYQRAL			E(Z: 5.20, SD± 1.73)	
	GLMTPSRFWAWYYCKA			E (Z: 3.36, SD± 0.92)	H(Z: 2.19, SD± 0.97)
	QQRSLRVIFVCQFLTG		C(Z: 4.78, SD± 0.83)		
	NVHLRCSIAQQTFPKL		C(Z: 3.25, SD± 0.68)		
	CCRHILHDRRLLFTIC		C(Z: 3.29, SD± 0.63)		
	SFPRGQIGIICSFSRS		C(Z: 3.21, SD± 0.45)		
	SCIDVQIQATPSFSLR		C(Z: 2.87, SD± 0.62)		
	GQQDLMCTLALPFDLR		C(Z: 2.48, SD± 0.77)		
	TTALTVQCTVNSMGLS		G(Z: 3.22, SD± 0.84)		
	MLLPAELLKHGKQMVC		G(Z: 1.96, SD± 1.10)		

Key:
[c] ≥90% replicate reproducibility in the Z score peptides (10 replicates per round 3 screening strategy). The Z score for the specific dataset is stated as (Z: X.XX, SD±X.XX), wherein "Z" refers to Z score, and "SD" is the standard deviation across replicates
Set i) VNN_SS2_SS3 **(A)**, VNN_SS2_TY3 **(B)**, VNN_TY2_TY3 **(C)**, VNN_TY2_SS3 **(D)**
Set ii) NNK_SS2_SS3 **(E)**, NNK_SS2_TY3 **(F)**, NNK_TY2_TY3 **(G**, *absent***)**, NNK_TY2_SS3 **(H)**

Table 5.8 .Physiochemical properties and relevant amino acid compositional characteristics of Z score identified peptide binders continued

Characteristics of Z-score identified panning peptide ligands. continued						
Peptide sequence (16 mers)	**Lib [a]**	**Species specific? (N/Y)**	**Charge**	**Hydrophobicity (%) [b]**	***pI* [c]**	**Charged residues [d]**
ICHRWFEPLYGYQRAL	NNK24	Y	1.5	68.75	8.02	3
GLMTPSRFWAWYYCKA	NNK25	Y	2.0	75.00	9.13	2
QQRSLRVIFVCQFLTG	VNN52	Y	2.0	56.25	10.37	2
NVHLRCSIAQQTFPKL	VNN53	Y	2.5	50.00	9.67	3
CCRHILHDRRLLFTIC	VNN54	Y	3.0	56.25	8.30	5
SFPRGQIGIICSFSRS	VNN55	Y	2.0	56.25	10.37	2
SCIDVQIQATPSFSLR	VNN56	Y	0.0	50.00	5.77	1
GQQDLMCTLALPFDLR	VNN57	Y	-1.0	62.50	4.11	1
TTALTVQCTVNSMGLS	NNK26	Y	0.0	50.00	5.53	0
MLLPAELLKHGKQMVC	NNK27	Y	1.5	68.75	8.03	3

Key; [a] Degenerate library scheme used for randomising library 16mer peptide regions ; $VNN_{15}(TTT)_1$, "VNN" or NNK; Top50 enriched binders: ≥40% replicate reproducibility (10 replicates in total used).
[b] Percentage of hydrophobic residues; "AMP Amino acid grouping"; G, A, V, L, I, P, F, M, W, C, and Y.
[c] *pI* = Isoelectric point
[d] = AMP Amino acid grouping cationic (basic) charged residues; K, R and H only.

As aforementioned the Z score pipeline focused on identifying likely species-specific affinity binders, and for both test species the vast majority of peptides akin to this were obtained from the $VNN_{15}(TTT)_1$ library. Manual analysis of Z-score outut across **Table 5.5** and **Table 5.7**, identified nine peptides exhibiting the previous bounce-back ability. Among the nine was six $VNN_{15}(TTT)_1$ library peptides including KLKDFMCRLVPTFVWL "Pep_VNN/21" and AWYQLICPVLARFTPK; "Pep_VNN/23" which were seen in $p/n_{datasets}$ = 100% of VNN_SS2_SS3 and VNN_TY2_SS3 replicates (**Table 5.5).** The remaining three peptides exhibiting this apparent binder signal bounce back with *S. suis* were NNK-library derived including Peptide NNK/25 (GLMTPSRFWAWYYCKA) Contrastingly, Z score analysis of the sample sets against S. *Typhimurium* 4/74, identified only one peptide candidate which fit this parameter specifically; Pep_NNK/2 (GFSIQGNHSVNCWGQL) (**Figure 5.3**).

Pep_NNK/1 (FWSSLRELICLAQSSL),was observed in 100% of NNK_SS2_SS3 and NNK_TY2_SS3 replicates, and Peptide NNK/1 observed an average Z score of 9.21 and 9.54 respectively. Manual reverse analysis using both Z scores and Top50 outputs identified Pep_NNK/1 and four other peptides as highly frequent and enriched library binder candidates against both test species herein. Namely, Pep_NNK/2, Pep_NNK/16, Pep_VNN/1 and Pep_VNN/2; and both VNN peptides were identified for *S. Typhimurium* 4/74 species specific enrichment with average Z scores of ~5.9 and ~4.3 respectively. Whereas, Pep_NNK/2 and Pep_NNK/16 possess contrasting bias for test species with average Z scores of ~3.8 (NNK_TY2_TY2) and ~2.9 – 4.2 (NNK_XX2_SS3) respectively.

5.3.6 Characterisation for shortlisted enriched binders against key bacteria pathogens of pigs

5.3.6.1 Physiochemical properties of shortlisted phage display derived peptides.

Completing the *in silico* NGS analysis pipeline sufficiently identified frequency ratio differences in peptide enrichment across panning sample sets. Z score: ≥cut-off peptides identified under the species-specific analysis pipeline held high interest. VNN_{15}+$(TTT)_1$ and NNK library peptides were therefore peptides with Z score ≥ 2 cut-off, $p_{dataset}$ = ≥ 90% and $n(pool)_{dataset}$ = ≤20% were shortlisted (**Table 5.5 to Table 5.8**). Whereas broad-spectrum enriched peptides were mainly identified via analysis using Top50 (**Table 5.1 to Table 5.4)**and Top100 (data not shown) with the cut-off set to $p_{dataset}$ = ≥ 90% and $n_{dataset}$ = ≥50% replicate reproducibility.

The eighty-five peptides shortlisted for crude synthesis present a broad range of physiochemical proprieties, and many align towards the generalised characteristics of AMPs (data not shown). On average peptides possessed an overall charge of ~ +2.00, isoelectric point, *pI* ~9.1, and were significantly composed of hydrophobic amino acids (~59% of Aa's on average) and cationic charged residues; Lys, Arg and His (~ 17% of Aa's on average). Charge is influenced by not only the presence of cationic residues but the absence of negatively charged counterparts. For instance, Z score shortlisted GQQDLMCTLALPFDLR, "Pep_VNN/57" possessed an overall -1 charge due to the richness of negatively charged aspartic acid (D), whereas YQKASLKSIYCRFHGR, "Pep_VNN/26", has a +4.5 charge and is abundantly rich in all three cationic Aa's (~31%). All peptides selected possessed ≥ 37.5% hydrophobicity. Excluding Phe, hydrophobic amino acids such as Leu, Val, Cys, Gly, and Thr were highly represented across both VNN and NNK library peptides. Additionally, peptides tended to align to the previous observations in the

ADP3_AMP_PigPathogenDataset, wherein AMPs 11-20mers in length typically contain 2-5 hydrophobic stretches.

5.3.6.2 *Screening for the antibacterial phenotype of* $VNN_{15}+(TTT)_1$ *and NNK phage display derived binders.*

The crude synthesis of the eighty-five Top50 and Z score shortlisted peptides entailed one modification, namely C-terminal amidation. Different termini modifications can be used to improve peptides and their adaption to physiological or experimental conditions. Amidated C-terminals are used for synthesised AMPs to maintain high-net positive charges across the peptide and reduce susceptibility to degradative action *e.g.* carboxypeptidase (Hansen et al., 2020). As shown in **Figure 5.4, 5.5, 5.6 and 5.7** shortlisted peptides were initially screened in broth microdilution MICs, in triplicate wells at 200µM against *S. Typhimurium* 4/74, *S. suis* P1/7 and secondary Gram-negative pig pathogen test species of interest *E. coli* P433 (porcine ETEC).

Figure 5.4. Broth microdilution 200µM screening of shortlisted phage display derived peptides against key Gram-negative bacterial pathogens of pigs; *S. Typhimurium* 4/74.

MIC assay Average OD_{600nm} absorbance change of triplicate wells were calculated **(See Method 2.5.3).** Shortlisted peptides key box; **red** $VNN_{15}(TTT)_1$, **blue** NNK. 96-well plates (polypropylene); 100 µl/per well of *S. Typhimurium 4/74* at 5 x 10^5 CFU/mL were spiked with crude library peptides at 200µM (incubated static <18hrs, 37ºC and aerobic conditions). The average absorbance of negative controls "**NCs**" is depicted by the red dashed line. Pep NNK/17 at 200µM "labelled G" reduces the OD_{600nm} absorbance change of *S. Typhimurium 4/74* in a manner equivalent to positive controls wells; Peptide RLL (**PC1**) and Polymyxin B (**PC2**) at 200µM. Interestingly, spiking Pep NNK/20 "labelled K" lead to a higher average change in OD_{600nm} absorbance than the negative control wells.

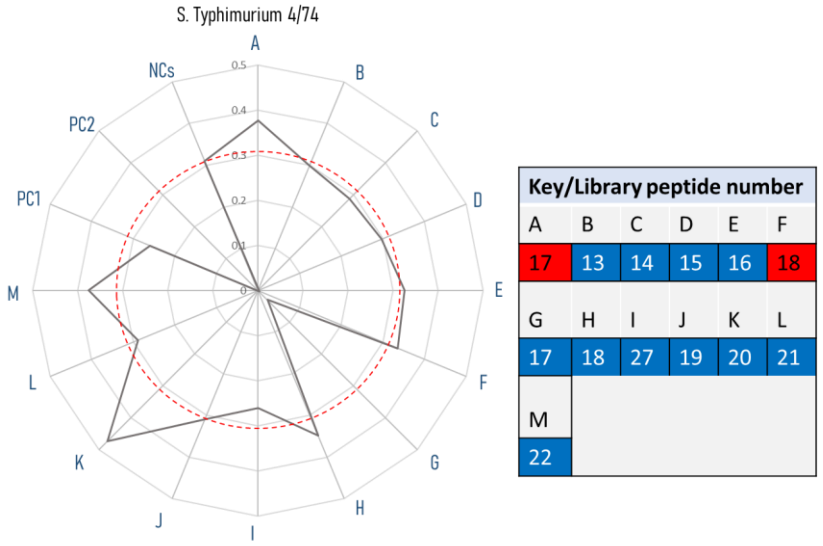


Figure 5.5 Broth microdilution 200µM screening of shortlisted phage display derived peptides against key Gram-negative bacterial pathogens of pigs; *E.coli* P433 (porcine ETEC)

MIC assay Average OD_{600nm} absorbance change of triplicate wells were calculated **(See Method 2.5.3).** Shortlisted peptides key box; **red** $VNN_{15}(TTT)_1$, **blue** NNK. 96-well plates (polypropylene); 100 µl/per well of *E. coli* P433 at 5 x 10^5 CFU/mL were spiked with crude library peptides at 200µM (incubated static <18hrs, 37ºC and aerobic conditions). The average absorbance of negative controls "**NCs**" is depicted by the red dashed line. Peptides either increase the average absorbance change of *E. coli* (Pep_NNK/19, Pep_NNK/20) or reduce the absorbance change to <0.1 O_{D600nm} (Pep_NNK/17 and Pep_NNK/21). Positive controls wells; Peptide RLL (**PC1**) and Polymyxin B (**PC2**).

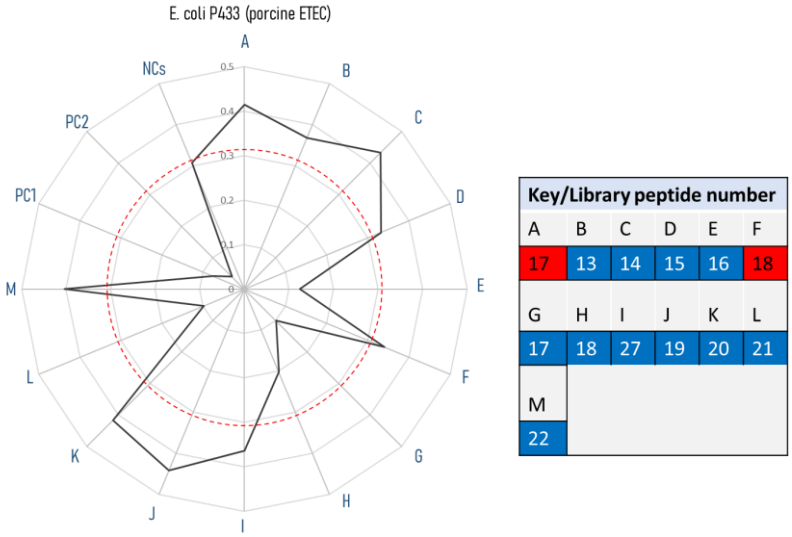
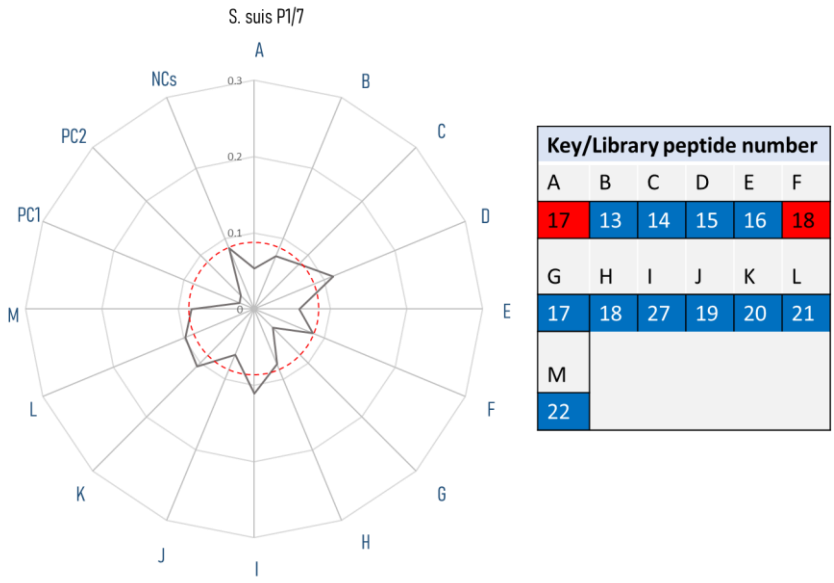


Figure 5.6 Broth microdilution 200µM screening of shortlisted phage display derived peptides against key Gram-positive bacterial pathogens of pigs; *S. suis*P1/7 Screen A

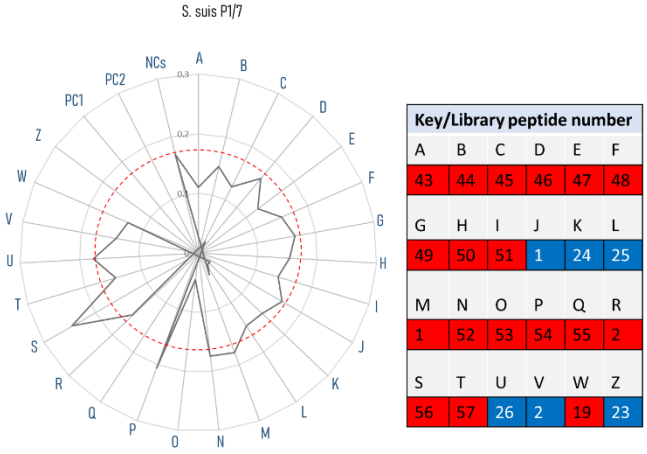
MIC assay Average OD_{600nm} absorbance change of triplicate wells were calculated **(See Method 2.5.3).** Shortlisted peptides key box; **red** $VNN_{15}(TTT)_1$, **blue** NNK. 96-well plates (polypropylene); 100 µl/per well of *S. suis P1/7* at 5 x 10^5 CFU/mL were spiked with crude library peptides at 200µM (incubated static <18hrs, 37ºC and aerobic conditions). The average absorbance of negative controls "**NCs**" is depicted by the red dashed line, and *S. suis* slower overnight growth compared to Gram-negative test species is exemplified. Four peptides reduce absorbance change; *i.e.* Pep_NNK/17, Pep_NNK/16 however, this is not comparable to positive controls wells; Nisin (**PC1**) and Kanamycin (**PC2**).



The broth microdilution screening of shortlisted panning peptides at 200µM provided an early indication of peptides exhibiting the desired antibacterial phenotype. For instance, the Top50 derived GVFKLQWPCRSFQHRF, "Pep_NNK/17" was identified from panning replicates against both test species and demonstrated antimicrobial activity at 200µM across all three test species; *S. Typhimurium* 4/74 (**Figure 5.4**), *E. coli* P433 (**Figure 5.5**) and *S. suis* (**Figure 5.6**). *E. coli* P433 was additionally susceptible to; STFPVSARWMHMLRCK "Pep_NNK/21" and WQPLARLWCTGIGKWK "Pep_NNK/16" (+3 charge, *pI* 11.48 and 10.79 respectively). Interestingly, both Pep_NNK/16 and NNK/21 share a C-termini cationic-hydrophobic moiety "MLRCK" and "…GIGKWK" respectively. A feature

which was characteristic of many of the ADP3_AMP_PigDataset natural AMPs. Pep_NNK/21 shares the same hydrophobicity as Pep_NNK/17 (56.25%), whereas Pep_NNK/16 was one of the most hydrophobic shortlisted peptides; ~68.75%, with hydrophobic amino acids generating a core consisting of a 7-Aa hydrophobic stretch. Four peptides in total exemplified activity against *S. suis* at 200µM, and this included two $VNN_{15}(TTT)_1$ peptides NVHLRCSIAQQTFPKL, "Pep_VNN/53" and SFPRGQIGIICSFSRS, "Pep_VNN/55" as well as AMLCWREGKDTRCCIS "Pep_NNK/23" (**Figure 5.7**). All three peptides were comparable to the average physiochemical properties of the shortlisted peptides. Peptide_NNK/23 was one of several Top50 bounce back -broad spec peptides identified in panning datasets; NNK_TY2_SS3 and NNK_SS2_TY2. This suggests peptides with weak antibacterial and broad-spectrum target affinity, can degrade to low frequency ratios or loss due lacking the sufficient binding affinity to be amplified under continued iterations of panning against the same target.

Figure 5.7 Broth microdilution 200µM screening of shortlisted phage display derived peptides against key Gram-positive bacterial pathogens of pigs; *S. suis* P1/7 Screen B.
Observed optical density radar chart of absorbance change observed (-minus assay seed bacteria contributing absorbance). 96-well plates (polypropylene); 100 µl/per well of *S. suis* P1/7 at 5 x 105 CFU/mL. Top50/Z score finalised peptides were spiked into triplicate wells at 200µM (incubated static <18hrs, 37oC and aerobic conditions). Shortlisted peptides key box; **red** $VNN_{15}(TTT)_1$, **blue** NNK. The average absorbance of negative controls "NCs" is depicted by the red dashed line, interestingly the majority of peptides were able to reduce absorbance change below that seen in the NCs. However, VNN/55 and VNN/53 were able to demonstrate inhibitory effects similar to positive controls wells; Nisin (**PC1**) and Kanamycin (**PC2**).



Key/Library peptide number					
A	B	C	D	E	F
43	44	45	46	47	48
G	H	I	J	K	L
49	50	51	1	24	25
M	N	O	P	Q	R
1	52	53	54	55	2
S	T	U	V	W	Z
56	57	26	2	19	23

Defining the minimum inhibitory concentrations of peptides identified with demonstrable activity at 200µM against any of the three was then prioritised. **Table 5.9** summarises the MICs observed against *S. Typhimurium* 4/74, *S. suis* P1/7, *E. coli* P433 strains. When tested in crude during the initial 200µM screen, several shortlisted panning peptides were able to register comparative test species OD_{600nm} absorbance diminishment to positive controls; Peptide RLL, Polymyxin B, Nisin or Kanamycin (**Figure 5.4 to Figure 5.7**). However, the characterised AMP positive controls utilised demonstrated their superiority in potency as the vast majority of shortlisted panning peptides exhibited MIC values at ≥ 200µM under further examination (**Table 5.9**).

Table 5.9. Minimum inhibitory concentrations for VNN15+(TTT)1 and NNK library phage display derived peptides against *S.Typhimurium* 4/74, *S. suis* P1/7 and *E. coli* P433 (ETEC)

Minimum inhibitory concentrations and species-specific activity				
Test bacteria species	**Peptides with activity ≤200µM (Repeat 1)**		**Peptides with activity ≤200µM (Repeat 2)**	
	Peptide	**Observed MIC (µM)**	**Peptide**	**Observed MIC (µM)**
***E. coli* P433 (Porcine, enterotoxin producing strain, "PETEC")**	Pep_VNN/19 (SS) Pep_NNK/16[BS] Pep_NNK/17 (BS) Pep_NNK/21 (SS)	200µM 200µM 50-100µM 50µM	Pep_VNN/19 (SS) Pep_NNK/16 [BS] Pep_NNK/17 (BS) Pep_NNK/21 (SS) Pep_NNK/18 [BS]	200µM 200µM 50-100µM 50µM ~200µM
***S. Typhimurium* (Rob strain?)**	Pep_NNK/17 (BS)	200µM	Pep_NNK/17 (BS)	200µM
***S. suis*P1/7**	Pep_VNN/53 Pep_VNN/55 [BS] Pep_NNK/23 [BS] Pep_VNN/43 [BS] Pep_VNN/47 [BS] Pep_NNK/17 (BS) Pep_VNN/35 (SS)	>200µM 200µM 100µM 100µM 200µM 200µM 200µM	Pep_VNN/53 Pep_VNN/55 [BS] Pep_NNK/23 [BS] Pep_VNN/43 [BS] Pep_VNN/47 [BS] Pep_NNK/17 (BS) Pep_VNN/35 (SS)	>200µM 100µM 100µM 100µM 200µM 200µM ~200µM

(BS) Broad spectrum peptide identified across Gram-positive and negative bacteria in initial and follow-up screens, [BS] specifically denotes identification of broad-spectrum activity at x2 or x4 MIC (data not shown). **(SS)** refers to peptides verified as "species-specific" is within the constraints of the three test species tested, *i.e.* peptide only exhibited activity against one test species.

~XµM or > XµM approximation of MIC via comparison with OD_{600nm} of negative control wells (data not shown).X-XµM (*i.e.* Peptide 79) reflects 2/3 clear wells observed for both 50 and 100µM, true MIC likely between these values.

200µM, equates to ~320µg/mL for 16mer peptide, when each amino acid is considered to be ~100Da.

Table 5.10 provides a summary of the properties of the peptides of most interest. Pep_NNK/17; GVFKLQWPCRSFQHRF (+3.5 charge, hydrophobicity; 56.25%, 4 charged residues, pI: 11.48) was one of the most cationic shortlisted peptides. Pep_NNK/17; was derived from Top50 analysis of the NNK_TY2_TY3/SS3 duplicate round 3 panning sampl sets. Pep_NNK/17 apparent broad-spectrum activity from the initial screen (**Figure 5.4 to Figure 5.7**) was further validated with MIC

values ranging between ~50-200µM MIC against the two Gram-negative (*E. coli* P433 and *S. Typhimurium* 4/74) and Gram-positive (*S. suis* P1/7) test species herein (**Table 5.9**). Several other shortlisted panning peptides were identified with the MIC value range of Pep_NNK/17 (**Table 5.9)**. Quantifying the MIC values were vital for potency determination however, determined MICs were additionally useful for demonstrating the species-specific activity of peptides. Peptides in **Table 5.9** were screened against all three test species at *x*2 and *x*4 MIC values to determine the concentration limits of "species specificity" as utilised herein to mean activity against *S. Typhimurium* 4/74, *S. suis* P1/7 and *E. coli* P433(data not shown).

Table 5.10 The physiochemical, hydrophobic and structural estimates of NNK and $VNN_{15}(TTT)_1$ phage display derived peptides.

The properties of phage display derived peptides					
Phage display derived peptide	**Physiochemical properties**	***μH* moment**	**Structure content**	**Solubility index**	**Observed MICs**
Pep_VNN/43	+1, 50.00%, pI ~8	0.65	α-helical/ coil	97.3%	100 µM (*S. suis*)
Pep_NNK/23	+2, 56.25%, pI ~10	0.43	coil /β-sheet	67.7%	
Pep_VNN/55	+2, 56.25%, pI ~10	0.34	β-sheet / coil	91.4%	100-200 µM (*S. suis*)
Pep_NNK/17	+3.5, 56.25%, pI ~11	0.45	coil / β-sheet	97.9%	50-200 µM (*E. coli, S. suis, S. Typhimurium*)
Pep_NNK/21	+1, 56.25%, pI ~8	0.44	coil	97.9%	50 µM (*E. coli*)

Amino acid sequences (16 mers)	Pep_VNN/43 SFSESISGWIRKFLCS *Identified:* Z score; 3.40 (VNN_SS2_SS3) Pep_NNK/17 GVFKLQWPCRSFQHRF *Identified:* Top50; (NNK_TY2_ TY3 and SS3) Pep_NNK/23 AMLCWREGKDTRCCIS *Identified:* Top50; (NNK_SS2_ TY3 and NNK_TY2_SS3)	Pep_VNN/55* SFPRGQIGIICSFSRS *Identified:* Z score; 3.21 (VNN_TY2_TY3) Pep_NNK/21 STFPVSARWMHMLR CK *Identified:* Top50; (NNK_TY2_ SS3)
Structure predictions; SFSESISGWIRKFLCS CCCCHHHHHHHHHHCC (Confidence level ~60%)		**Structure predictions;** SFPRGQIGIICSFSRS CCCCCEEEEEEEECCC (Confidence level ~89%)
Structure predictions; GVFKLQWPCRSFQHRF CCEECCCCCCCCCCCC (Confidence level ~75%)		**Structure predictions;** STFPVSARWMHMLRCK CCCCCCHHHHEEECCC (Confidence level ~70%)
Structure predictions; AMLCWREGKDTRCCIS CCEEECCCCCEEEEEC (Confidence level ~75%)		C = Coil content H = Helix content E= Beta sheet content

Key:
- Physiochemical properties; + or - (charge), (hydrophobicity %), *(pI,* isoelectric point)
- *μH* moment at 100° angle – peptides R package
- Structure prediction: Conducted on Proteus structure prediction server 2.0 http://www.proteus2.ca/proteus2/
- Solubility index; ≥ +90% are typically identified as highly soluble and are likely to be obtained from soluble fractions (Harrison et al., 2000).
- Observed MICs: conventional CLSI broth microdilution determined MIC
- LIB_XX2_XX3, LIB = VNN15(TTT)1 or NN, XX2 and XX3 are round 2 and 3 panning test species respectively.

Three peptides were identified as "species – specific" against *E. coli* P433 or *S. suis* P1/7 under the aforementioned defined test species context. This included two *E. coli* P433 specific AMPs from both the NNK and $VNN_{15}(TTT)_1$ library; Pep_VNN/19; AIKMRGPLHRCNFAVP (MIC:200µM) and Pep_NNK/21 (MIC:50µM) ; STFPVSARWMHMLRCK. Pep_NNK/21 was the most potent of the two and interestingly, was a randomly picked clone identified when sanger sequencing the NNK *E. coli* JE5505 library used for overlay screening. Both peptides Pep_VNN/19 and Pep_NNK/21; were identified from analysing Top50 LIB_SS2_SS3 and LIB_SS2_TY3 panning samples. Contrastingly, Z score derived Pep_VNN/35 (Z: ~3.50 SD ±0.44); FLDNPFVSLIRCFIAR (MIC 200µM) was identified only in VNN_SS2_SS3 panning sample set, and characterisation identified specific antimicrobial activity against *S. suis* P1/7 (**Table 5.9**). This bias for demonstrating activity was matched by the majority of $VNN_{15}(TTT)_1$ peptides reviewed during this study (**Table 5.9**). Verifying the activity of *x*2 and *x*4 MIC concentrations of crude peptides demonstrated the concentration dependent broad-spectrum activity of three peptide. Specifically, this included; Top50 derived Pep_NNK/23, and two Z-score derived peptides Pep_VNN/43 SFSESISGWIRKFLC (Z score ~3.40) and Pep_VNN/55 SFPRGQIGIICSFSRS (Z score ~3.21), which were obtained from VNN_SS2_SS3 and VNN_TY2_TY2 panning outputs respectively. Pep_VNN/43 (x2 MIC) and Pep_NNK/23 (x4 MIC) demonstrated activity against *E. coli* P433 at above MIC concentrations, whereas Pep_VNN/55 was shown to be active against *S. typhimurium* 4/74 (x4 MIC). In conclusion the present study demonstrated the ability to identify AMPs with specific-specific biases and broad-spectrum antibacterial activity.

5.4 Discussion

Natural AMPs are ≤ 60 mers, highly hydrophobic and often amphiphilic in nature, and likely carry an overall cationic charge (+2 to +9) due to excess representation of the triplet basic (positively-charged side chains) polar amino acids: Lys, Arg, and His. Antibacterial AMPs interact and disrupt anionic bacterial membranes, and as such have gained attention for potential clinical applications. In spite of multi-disciplinary progresses in antimicrobial drug discovery, the unfettered emergence of antimicrobial resistance necessitates further successes in novel AMP development for this class of peptide to indeed supplant conventional antibiotics.

Developments in site-directed degenerate mutagenesis and next generation sequencing (NGS) phage display technologies both possess the ability to enhance the research and discovery of novel AMPs for clinical exploitation. Nevertheless, the identification of novel antibacterial AMPs via degenerately randomised peptide libraries and pIII phage display systems still requires further investigation. Alongside both innovations in degenerate schemes utilised and phage display panning implementation. The present study aimed to design an AMP discovery pipeline which exploited a novel $VNN_{15}(TTT)_1$ degenerate scheme in the hopes of generating peptide libraries which provide more library candidates with antimicrobial phenotypes than presently highly exploited NNK scheme.

Five of the most promising library panning peptides in crude form possessed MICs ranges between; 50 to 100 µM against bacterial pathogens with sufficient porcine host specificity to elicit enteric, respiratory or systematic diseases in pigs; *S. suis* P1/7, *S. Typhiumurium* 4/74 and *E. coli* P433. Nonetheless, noteworthy to mention characterised $VNN_{15}(TTT)_1$ and NNK phage display derived peptides herein, observed ≥40-fold higher MICs values than characterised literature AMPs. This fold

difference was both in the context of; antibacterial phage display derived peptides from similar NNK randomised phage display platforms (Pini *et al.,* 2012) and primary natural AMP antibiotics (Nisin, Polymyxin B and Gramicidin S) which demonstrate activity against numerous species within the same Gram-spectrum, some of which have typified antibiotic resistance profiles (Hancock and Chapple, 1999).

Our work essentially demonstrates peptides with MICs ≥100µM can be effectively enriched against *S. suis* P1/7 via whole-cell phage display, and when synthesised with C-termini amidation peptides exhibit phenotypic antibacterial inhibition. Interestingly, *S. Typhimurium* 4/74 showed little susceptibility to panning peptides. As the most potent phage display derived peptide identified against *Typhimurium* 4/74 pathogen possessed an MIC ~200µM (Pep_NNK/17). Susceptibility differences are underpinned by the compositional frequency of membrane targets and antimicrobial resistance mechanisms which exist across bacteria species and strains. For instance, efflux pumps, membrane component altering enzymes and biofilms are within the defensive arsenal of E. *coli, S. Typhimurium* and *S. suis* test species.

However, this illustrates iterative cycles of whole bacteria cell panning against phage display libraries can inadvertently enrich peptides for bacteria test species not tested. Even highly unrelated bacteria can share similar membrane components, and this was likely exploited during iterative panning rounds to give rise to *E. coli* P433 AMPs identified. Target specificity of AMPs can be more rationally exploited via approaches centred on panning against integral characterised membrane components. For instance, Yuxia Xionga ZhibangYanga *et al.,* (2019) conducted three rounds of panning against recombinant ArsRS gastric acid receptor and screening ELISA

verified panning binders thus increased the depth of data generated prior to screening against ArsRS-rich *H. pylori*

Herein an initial subtractive panning round against beneficial microbiota lactic acid bacteria, such as *L. acidiophillus* ATCC 43561, was omitted. Specifically, this was in favour of conducting *in silico* subtraction within the NGS analysis strategy and in essence replicating a subtractive screen in the final panning round. Essentially this uncomplicated the panning process. However, more importantly the panning structure opted for was to ensure any unquantified pool of putative Gram-positive specific peptide ligands in libraries were not inadvertently diminished prior to panning against *S. suis* P1/7. Consequently, we generated two groupings of final round panning datasets*; a)* samples where input $VNN_{15}(TTT)_1$ and NNK phage libraries were panned against each test species for 3 iterative enrichment cycles, *b)* panning samples where round 2 output phage-sub libraries enriched against the two test species were bilaterally cross-panned.

This facilitated conscientiously analysing the enrichment and/or diminishment discrepancy of Top50 and Z score > cut-off peptide ligands across *S. Typhimurium* 4/74 and *S. suis* P1/7. Manual analysis of this disparity led to the shortlisting of Pep_NNK/23 which intriguingly diminished past the >cut-off frequency representation in panning samples originating from three iterative panning rounds against each test species . Whereas Pep_NNK/23 emerged prevalent in panning output samples wherein *S. Typhimurium* 4/74 and *S. suis* P1/7 were the test species panned against for ≤2 rounds (NNK_SS2_ TY3 and NNK_TY2_SS3). TopN and Z score analysis pipelines both corroborated the “bounce back” of enriched peptides into the realms of high frequency classification. Peptides exhibiting this pattern of representation in panning output, possibly target bacterial membrane components

which exist somewhat broad spectrum. However, equally they exert weaker than necessary binding affinity to be sustained in the phage phenotype-information cycle when subjected to further iterations of enrichment. Our results therefore suggests a window of *n* number of panning rounds, where n = ≤ 3 might be necessary to observe the loss AMPs with ~100µM activity *in vitro*.

AMP *in silico* prediction would be the obvious solution to narrowing down antibacterial candidates in panning outputs. However, these tools lack validation due to limited comprehension of predicting AMP activity (Tucker et al., 2018). For instance, when assessing seven AMP prediction tool and models on CAMPr3 and DBAASP, all five peptides with MICs $\leq$ 200 µM herein registered AMP classification in $\leq$ 1/7 tools. The diversification of *in-silico* frequency analysis, via both TopN and Z scores, supported the identification and verification of enriched panning peptide ligands. Arguably, the fundamental question is not which approach efficiently identifies binders but does the output of analysing phage display peptides in this manner improve AMP identification. Crude synthesised shortlisted peptides demonstrate the qualitative and quantitative data derived from *in-silico* frequency analysis of binder enrichment weakly correlates to the identification of the desired antibacterial phenotype. Although, the expense of producing AMPs for in vitro assays limits the expansive characterisation that can be employed nevertheless techniques such as SPOT-AMP synthesis might facilitate cheaper and more widespread synthesis of candidates of interest identified from peptide libraries (Hilpert and Hancock, 2007).

The inability to characterise antimicrobial activity at the point of screening is a bottleneck for phage display technologies (Tucker et al., 2018). Peptide pIII fusion can restrict recombinantly encoded peptides from demonstrating antibacterial action, and this can arise for several reasons. By the nature of the pIII phage fusion system,

native AMP structural conformations are likely affected, recombinant AMP-pIII fusion can sustain low virion display percentages below accumulation thresholds necessary to kill bacteria via modes of action such as the detergent carpet model. However, integral to this assumption is the compatibility of the AMP recombinant pIII with the phage replicative systems. Library encoded peptides which disrupt phage infection or assembly are unfavoured due to reduced phage fitness. Potentially, library peptides with more potent antimicrobial activity than the panning ligands characterised herein, were lost from the library and phage pool. Consequently, improving AMP identification from phage display systems might require the NGS analysis of starting library material; ligated vector DNA. Sequencing from this ligated library DNA could then possibly be compared to bacteria transformations (*E. coli* TG1) or phage propagation (M13), and this might provide insight for understanding various biological censorships exhausted by the phage display approach.

Chapter 6: Final Discussion

The principal goal of this research study was to develop a novel AMP discovery pipeline by exploiting peptide library technologies, semi-rational degenerate codon scheme mutagenesis and high-throughput screening approaches. 187 natural AMPs with activity against key bacterial pathogens of pigs were collected from one of the largest and well-recognised AMP repository databases; ADP3 database "APD3_ AMP_PigPathogenDataset" (Mishra and Wang 2012). Natural AMPs are typically ≤ 60 mers, highly hydrophobic and often amphiphilic in nature, with an overall cationic charges (+2 to +9). The natural AMPs present in APD3_ AMP_PigPathogenDataset presented physicochemical and amino acid biases which confirmed aforementioned generalisations of this class of peptides. APD3_ AMP_PigPathogenDataset contained peptides which were highly cationic and amphipathic, prominent Aa's across the AMP amino acid groupings were Gly/Leu/Ala (29.9%), Lys (12.8%) and Ser (4.4%), and these Aa biases tend to be present across natural AMP databases (Mishra and Wang 2012; Pirtskhalava et al., 2021). $VNN_{15}(TTT)_1$ scheme was theorised to bring the presentation of all three AMP amino acids groupings towards that observed in the AMP_PigPathogenDataset.

Following the identification of the $VNN_{15}(TTT)_1$ scheme, this study focused on designing a library construction strategy. A phagemid vector *pSD3* was selected and subsequent modifications entailing restriction enzyme site additions (*SpeI*), facilitated vector system suitability for screening technologies reliant on *E. coli* strains; including importantly *E. coli* TG1 in phage display approaches or leaky periplasm *E. coli* mutants in agar-based assays. NNK and $VNN_{15}(TTT)_1$ degenerately randomised 16mer peptide libraries were constructed from relatively simple modifications to the

approaches outlined by Kong *et al.,* (2020) and Tsoumpeli *et al.,* (2022). The whole-construct inverse PCR approach designed herein was able to generate large and diverse 16mer peptide phage libraries; specifically, 1.83 x 10^9 and 2.21 x 10^9 for $VNN_{15}(TTT)_1$ and NNK schemes respectively. A promising result as uncomplicating the process of library construction is likely to lead to greater applications of these schemes in broader AMP research. Additionally, *SpeI* digestion facilitated the removal of the phage *gIII* with relative ease. *SpeI* digested libraries were ligated and subsequently transformed into *E. coli* JE5505 electrocompetent cells generating sub-libraries for screening in the overlay assay. The *E. coli* JE5505 sub-libraries generated possessed diversities of 9.21 x 10^6 and 4.6 x 10^6 which is the typical library size exploited for AMP scanning studies (Tominaga and Hatakeyama, 2006; Guralp et al., 2013).

NGS quality control demonstrated VNN_{15}+(TTT)1 naïve peptide libraries (*E. coli* TG1, *E. coli* JE5505, M13 phage), observed frequency of amino acids and stop codons (TAG, TGA, TAA) deviated by ~ ± 3% from the theoretical expected frequency. NNK naïve peptide library equivalents observed amino acids frequency which deviated by ~ ± 2% from theoretical expected. This range of deviation existed within the bounds of that reported for previous degenerate peptide library strategies (Tsoumpeli et al., 2022). The two libraries possessed differing patterns of over and under-represented amino acid or stops. VNN_{15}+$(TTT)_1$ *E. coli* TG1 library overrepresented Aa's Val, Gly, Gln but underrepresented Pro, Arg, His, Thr. The residual inclusion of cysteine and <0.6% stops likely signify reintroductions from primers where the VNN scheme was inappropriately applied. The amino acids overrepresented in the NNK *E. coli* TG1 library were; Lys, Trp, Met and amber stops whereas hydrophobic or cationic Aa's Ala, Pro, His, and Thr were underrepresented.

Interestingly, the deviations of NNK from expected likely provided addition benefits of representing amino acids such as Lys and Trp which with further examination in this study would prove prominent in library peptides displaying antimicrobial activity.

Large peptide libraries can be generated by numerous means however, utility is derived from high-throughput screening of library candidates to identify ligands for defined targets. Agar-based assays which exploit *E. coli* mutants with leaky periplasm have been utilised for AMP research, and importantly improved library mutants were identified (Tominaga and Hatakeyama, 2006; Guralp et al., 2013). This alluring prospect led to development of *E. coli* JE5505 0.7% Mueller-Hinton soft agar overlay assay for screening antimicrobial activity of pSD3 encoded periplasm tagged recombinant proteins. The assay beneficially utilises the most applicable antimicrobial susceptibility testing media (Mueller-Hinton) and supports testing against Gram-positive and Gram-negative bacteria including pig pathogens; *S. suis* P1/7, *S. Typhimurium* 4/74 and *E. coli* P433.

This study extends the current knowledge-base by exemplifying the *pSD3_PelB* phagemid vector system (*-gIII*) is compatible with the *E. coli* JE5505 expression system. Under 1mM IPTG induction, *E. coli* JE5505 colonies were able to leach recombinant Pln-423 and Guralp *et al.,* (2013) improved mutants into the agar creating zones of inhibition against *Listeria* species. Nonetheless, further investigation revealed characterised phage display AMPs literature and natural AMPs from repository database peptides such as; Peptide RLL (EC5), CATHELICIDIN-Bf were unable to generate zones of inhibition in the *E. coli* JE5505 expression system despite their relatively potent antimicrobial *in vitro* properties in broth assay and spotting crude peptide diffusion agar assays.

Agar diffusion methods rely on an implausible assumption of antimicrobial free diffusion. This creates the major bottleneck of agar-based techniques, as antimicrobial activity and/or potency is directly influenced by the diffusibility of active substances (Choyam et al., 2015). The leaky *E. coli* JE5505 overlay assay is therefore restrained by the complexities of balancing both biological and media diffusion biases. *E. coli* JE5505 expression of peptides with activity against the strain, likely exerts undue antimicrobial effects, or contributes to high metabolic loads, increasing toxicity, inclusion body formation and/or stunting growth kinetics. Whereas Class-II bacteriocins utilised within overlays generally were predicted to present in soluble fractions of bacteria more readily and have weak activity against *E. coli* JE5505.

The investigation of several other AMPs in the *E. coli* JE5505 expression system demonstrated sub-micromolar antimicrobial activity of AMPs such as Pln-423 are likely additionally necessary to produce observable zone of inhibitions. Fundamentally, the *E. coli* JE5505 expression overlay system is weakly compatible with screening libraries of degenerately randomised with uncharacterised peptides properties. Rather, the applicability of overlays remains within the narrow spectrum of identifying improved mutants from the site-directed mutagenesis of known and characterised bacteriocins or enterocins.

Nevertheless, screening via reliable and highly exploited phage display technologies was maintained in library design. NNK and $VNN_{15}(TTT)_1$ were screened via a whole-bacteria cell phage display approach, with a three-round panning strategy against representative Gram-negative and Gram-positive model bacteria with porcine host specificity; *S. Typhimurium* 4/74 and *S. suis* P1/7 respectively. *In silico* NGS analysis focused on analysing the enrichment and/or diminishment discrepancy of peptide ligands identified as highly frequent in designated positive panning sample

datasets. Panning samples were subjected to two-tiered TopN and Z scores frequency bioinformatic analysis pipelines. Manual analysis of the frequency analysis output from the two *in silico* pipelines; TopN and Z score identified ~235 library peptide candidates as enriched binders with 90% and 100% reproducibility across samples.

85/235 peptides were shortlisted for antimicrobial susceptibility testing against bacteria pathogens with porcine host specificity; *S. Typhimurium* 4/74 *S. suis* P1/7 and *E. coli* P433. Antimicrobial susceptibility testing of the shortlisted phage -display derived peptides demonstrated the phenotype-information cycle of phage selects for multiple phenotypes including antimicrobial activity. For instance, NNK_20 increased growth of all three aforementioned test species in the MIC) assays. Whereas <6% of peptides shortlisted registered the antimicrobial phenotype, including; three NNK-dervived peptides: Pep_NNK/17, Pep_NNK/23 and Pep_NNK/21 in addition to two peptides Pep_VNN/43 and Pep_VNN/55. Generally, these peptides were cationic (< +3) and highly hydrophobic (50 - 56.25%) with numerous predicted secondary structures including α-helices, β-sheets and coils. Nonetheless, both $VNN_{15}(TTT)_1$ and NNK phage display derived peptides were unable to demonstrate activity lower than 50μM. This more broadly reflects identifying novel AMPs which supersede the activity of conventional antibiotics including clinical AMPs such as polymyxin still remains a significant challenge. Nevertheless, Pep_NNK/17, Pep_NNK/23, Pep_NNK/21, Pep_VNN/43 and Pep_VNN/55 provide the necessary proof-of-concept to conclude the discovery pipeline formulated herein can identify early AMP hit candidates for developing novel therapies against bacterial pathogens of pigs.

Chapter 7: Conclusion and future perspectives

The NNK degenerate codon scheme approach aims to limit redundancy while encoding all twenty natural amino acids. Contrastingly, the $VNN_{15}(TTT)_1$ scheme sought to prioritise both limiting redundancy and restricting representation of amino acids towards the *in silico* derived distribution of natural AMPs. Diverse libraries offer innumerable possibilities of combinatorial peptide sequences nonetheless, rationally designed libraries are presently limited in their present applicability to AMPs. As designing AMPs solely through *in silico* means remains a challenging task given the limited comprehension of the relationship between amino acid distribution and peptide *in vitro* characterised antimicrobial activity (Tucker et al., 2018). Consequently, high-throughput screening approaches are crucially required for identifying library peptide variants with desired phenotypes. For the *E. coli* JE5505 overlay assay, the divergent agar diffusibility and solubility properties of a randomised and uncharacterised peptide library, limited the suitability of this approach for identifying novel AMPs from $VNN_{15}(TTT)_1$ and NNK peptide libraries designed herein. (Bonev et al., 2008; Choyam et al., 2015; Balouiri et al., 2016). Phage display has exemplified the identification of several short (<20 mers) and novel AMPs, with MICs which tend to be ≥1 µM (Pini et al., 2005; Bishop-hurley et al., 2005; Sainath Rao, Mohan and Atreya, 2013;Flachbartova et al., 2016).

Whole-cell next generation phage display herein against *S. Typhimurium* and *S. suis* successfully identified AMPs with MIC values ≤200µM from both $>10^9$ diverse $VNN_{15}(TTT)_1$ and NNK degenerately randomised 16 mer peptide libraries. The conventional NNK degenerate library demonstrated the ability to produce more potent peptides with the desired antibacterial phenotype against bacteria pathogens of pigs;

S. Typhimurium 4/74, *S. suis* P1/7 and *E. coli* P433. Pep_NNK/21 was the most potent AMP identified, exhibiting an MIC of 50µM against *E. coli* P433 (ETEC, isolated from porcine host) nevertheless Pep_NNK/21 is less potent than previously identified *E. coli* targeting phage display AMPs (Pini et al., 2012; Sainath Rao, Mohan and Atreya, 2013). $VNN_{15}(TTT)_1$ scheme derived phage display peptides were able to *in vitro* inhibit the growth of *S. suis P1/7* at 100µM, extending the present knowledge of phage display derived AMP candidates for *S. suis*. Additionally, the $VNN_{15}(TTT)_1$ scheme produced $\leq x$ 9 more enriched panning peptides replicated with 90% or 100% sample reproducibility with Z scores ≥2. Intriguingly indicating $VNN_{15}(TTT)_1$ derived peptides possessed a higher propensity to be enriched against the whole-cells of bacterial pathogens. Fundamentally, this study further consolidates phage display screening of degenerately randomised libraries can identify peptide ligands including antimicrobials against bacterial pathogens of pigs with porcine host specificity. A minuscule fraction of library variants which are enriched binders against panned bacteria species exhibit potent desired antimicrobial phenotype. Nevertheless, hit-to-lead development can be exploited downstream to further characterise initial hits and improve potency, selectivity and pharmacological properties for downstream novel AMP therapies.

Future work:

- Principial downstream development would involve *in vitro* screening $VNN_{15}(TTT)_1$ and NNK phage display derived peptides against larger panels of *S. Typhimurium, S. suis, E. coli* strains among other bacterial pathogens of pigs, including strains with antimicrobial resistance profiles.
- The enrichment of library peptide variants relies upon their ability to bind selectively and with high-affinity to whole-cell bacterial targets. *in silico* NGS

analysis indicated higher enrichment of $VNN_{15}(TTT)_1$ ligands. Consequently, downstream work will endeavour to elucidate the comparative bacterial target binding affinity of $VNN_{15}(TTT)_1$ and NNK library peptides. Phage displayed peptides can be exploited in ELISA formats as previously described by Pini et al., (2005) if stringent bacterial fixation methods are optimised to maintain bacterial integrity, preserve antigenicity and assay reproducibility.

- Kim et al., 2020 ; Peng Tan et al., 2020 exemplified relatively weakly active AMPs with strong species-specific binding affinity can be tethered to potent partner AMPs to generate novel dimerised constructs with superior activity than singular constituents. Naturally, many of the peptides within this study could be explored for their species-specific binding candidacy in novel STAMP constructs.
- Tanaka, Kokuryu and Matsunaga, (2008) engineered phage display derived peptides using rational mutagenesis, resulting in the identification of mutants with greater antimicrobial activity. Phage display derived peptides identified with activity herein, could be candidates for second round site-directed mutagenesis and scanning studies. Specifically, strategies to identify improved mutants could involve either targeting saturating termini regions with hydrophobic and cationic amino acids or alternatively, VNN or NNK degenerate randomisation of specific regions across candidates.
- Lastly, varying degenerate codon schemes or approaches can be devised and adapted for use with the reproducible vector design, library construction and screening strategy outlined herein. The present study prioritised analysing the $VNN_{15}(TTT)_1$ approach for randomising a 16 mer peptide region. However, minor amendments such as partnering VNN randomisation with an encoded

amino acid "$VNN_{15}(XXX)_1$" not merely to reintroduce a codon scheme restricted amino acid such as phenylalanine "TTT" but to strategically oversaturate other key amino acids such as lysine i.e. $VNN_{15}(AAA)_1$, or the other cationic amino acids and/or more highly represented hydrophobic amino acids found in AMPs (Glycine, Leucine, Alanine)

- Other display platforms such as SLAY tethering of AMPs to membrane components can directly screen antimicrobial phenotypes of peptides (Tucker et al., 2018). The study herein designed a strategy which was optimised for phage display and overlay assays, future work will likely consider replacing overlay approaches with SLAY tethering screening.

Moving forward peptide phage display libraries randomised utilising trinucleotide "TRIM" technology such as; M13 phage TriCo 16 mer and 20 mer™ Phage Display Peptide Libraries (Creative Biolabs) might be useful commercial libraries to repurpose for screening against bacterial pathogens. Nonetheless, the needle in the haystack analogy is perfectly apt for novel antimicrobial peptide discovery. Perhaps, at this present state of multi-disciplinary progress, the size of the haystack (*i.e.* large peptide libraries) is not the obfuscating factor. Indeed, the bottleneck for novel AMP identification are the strategies to identify the needle in the haystack, as the AMP research field deserpartely requires innovative high-throughput screening methods which directly identify novel candidates based on their antimicrobial phenotype.

Chapter 8: References

A. De Lucia, F. Ostanello; Large Animal Review 2020; 26: 133-140. (n.d.).

Abee T. Delves-Broughton J. Bacteriocins—nisin. In: Russell NJ, editor; Gould GW, editor. Food Preservatives. New York City: Springer; 2003. pp. 146–169.

Abu Hatab, A., Cavinato, M. E., & Lagerkvist, C. J. (2019). Urbanization, livestock systems and food security in developing countries: A systematic review of the literature. Food Security, 11(2), 279-299. doi:10.1007/s12571-019-00906-1

Acevedo-Rocha, C. G., Reetz, M. T., & Nov, Y. (2015). Economical analysis of saturation mutagenesis experiments. Scientific Reports, 5(1). doi:10.1038/srep10654.

Åhman, J., Matuschek, E., & Kahlmeter, G. (2020). EUCAST evaluation of 21 brands of Mueller–Hinton dehydrated media for disc diffusion testing. Clinical Microbiology and Infection, 26(10). doi:10.1016/j.cmi.2020.01.018

Ackermann, H. (2003). Bacteriophage observations and evolution. Research in Microbiology, 154(4), 245-251. doi:10.1016/s0923-2508(03)00067-6

Amaral, L., Martins, A., Spengler, G., & Molnar, J. (2014). Efflux pumps of gram-negative bacteria: What they do, how they do it, with what and how to deal with them. Frontiers in Pharmacology, 4. doi:10.3389/fphar.2013.00168

Andrews, J. M. (2001). Determination of minimum inhibitory concentrations. Journal of Antimicrobial Chemotherapy, 48(suppl_1), 5–16. doi:10.1093/jac/48.suppl_1.5

Animal and Plant Health Agency: The Pig Disease Surveillance Dashboard, 2022). (n.d.). Retrieved October 31, 2022, from http://apha.defra.gov.uk/vet-gateway/surveillance/scanning/disease-dashboards.htm

Ao, T. T., Feasey, N. A., Gordon, M. A., Keddy, K. H., Angulo, F. J., & Crump, J. A. (2015). Global burden of invasive nontyphoidal salmonella disease, 20101. Emerging Infectious Diseases, 21(6), 941-949. doi:10.3201/eid2106.140999

Apley, M. D., Bush, E. J., Morrison, R. B., Singer, R. S., & Snelson, H. (2012). Use estimates of in-feed antimicrobials in swine production in the United States. Foodborne Pathogens and Disease, 9(3), 272-279. doi:10.1089/fpd.2011.0983

Arenz, S., & Wilson, D. N. (2016). Bacterial protein synthesis as a target for antibiotic inhibition. Cold Spring Harbor Perspectives in Medicine, 6(9). doi:10.1101/cshperspect.a025361

Arunachalam, T. S., Wichert, C., Appel, B., & Müller, S. (2012). Mixed oligonucleotides for random mutagenesis: Best way of making them. Organic & Biomolecular Chemistry, 10(24), 4641. doi:10.1039/c2ob25328c

Ashraf, M., Frigotto, L., Smith, M., Patel, S., Hughes, M., Poole, A., . . . Hine, A. (2013). ProxiMAX randomization: A new technology for non-degenerate saturation mutagenesis of contiguous codons. Biochemical Society Transactions, 41(5), 1189-1194. doi:10.1042/bst20130123

Bahar, A., & Ren, D. (2013). Antimicrobial peptides. Pharmaceuticals, 6(12), 1543-1575. doi:10.3390/ph6121543

Bahar, A., & Ren, D. (2013). Antimicrobial peptides. Pharmaceuticals, 6(12), 1543-1575. doi:10.3390/ph6121543

Bahnson, P., Fedorka-Cray, P., Ladely, S., & Mateus-Pinilla, N. (2006). Herd-level risk factors for salmonella enterica subsp. enterica in U.S. market pigs. Preventive Veterinary Medicine, 76(3-4), 249-262. doi:10.1016/j.prevetmed.2006.05.009

Bakshi, C. S., Singh, V. P., Wood, M. W., Jones, P. W., Wallis, T. S., & Galyov, E. E. (2000). Identification of Sope2, a salmonella secreted protein which is highly homologous to sope and involved in bacterial invasion of epithelial cells. Journal of Bacteriology, 182(8), 2341-2344. doi:10.1128/jb.182.8.2341-2344.2000

Balouiri, M., Sadiki, M., & Ibnsouda, S. K. (2016). Methods for in vitro evaluating antimicrobial activity: A Review. Journal of Pharmaceutical Analysis, 6(2), 71-79. doi:10.1016/j.jpha.2015.11.005

Barrett, C. B. (2010). Measuring food insecurity. Science, 327(5967), 825-828. doi:10.1126/science.1182768

Bäumler, A. J., Tsolis, R. M., Ficht, T. A., & Adams, L. G. (1998). Evolution of host adaptation in salmonella enterica. Infection and Immunity, 66(10), 4579–4587. doi:10.1128/iai.66.10.4579-4587.1998

Baums, C. G., & Valentin-Weigand, P. (2009). Surface-associated and secreted factors ofstreptococcus suisin epidemiology, pathogenesis and Vaccine development. Animal Health Research Reviews, 10(1), 65-83. doi:10.1017/s146625230999003x

Bellido-Carreras, N., Argüello, H., Zaldívar-López, S., Jiménez-Marín, Á, Martins, R. P., Arce, C., . . . Garrido, J. J. (2019). Salmonella typhimuriuminfection along the porcine gastrointestinal tract and associated lymphoid tissues. Veterinary Pathology, 56(5), 681-690. doi:10.1177/0300985819843682

Bergšpica, I., Kaprou, G., Alexa, E. A., Prieto, M., & Alvarez-Ordóñez, A. (2020). Extended spectrum β-lactamase (ESBL) producing escherichia coli in pigs and pork meat in the European Union. Antibiotics, 9(10), 678. doi:10.3390/antibiotics9100678

Bergšpica, I., Kaprou, G., Alexa, E. A., Prieto, M., & Alvarez-Ordóñez, A. (2020). Extended spectrum β-lactamase (ESBL) producing escherichia coli in pigs and

pork meat in the European Union. Antibiotics, 9(10), 678. doi:10.3390/antibiotics9100678

Bertani, B., & Ruiz, N. (2018). Function and biogenesis of lipopolysaccharides. EcoSal Plus, 8(1). doi:10.1128/ecosalplus.esp-0001-2018

Betic, N., Klun, I., Djordjevic, V., Brankovic Lazic, I., Baltic, T., Vasilev, D., & Karabasil, N. (2021). Toxoplasma gondii in pork and pigs in Serbia – a real food safety hazard. IOP Conference Series: Earth and Environmental Science, 854(1), 012008. doi:10.1088/1755-1315/854/1/012008

Bishop-Hurley, S. L., Rea, P. J., & McSweeney, C. S. (2010). Phage-displayed peptides selected for binding to campylobacter jejuni are antimicrobial. Protein Engineering Design and Selection, 23(10), 751-757. doi:10.1093/protein/gzq050

Biswaro, L. S., Da Costa Sousa, M. G., Rezende, T. M., Dias, S. C., & Franco, O. L. (2018). Antimicrobial peptides and nanotechnology, recent advances and challenges. Frontiers in Microbiology, 9. doi:10.3389/fmicb.2018.00855

Blair, J. M., Richmond, G. E., & Piddock, L. J. (2014). Multidrug efflux pumps in gram-negative bacteria and their role in antibiotic resistance. Future Microbiology, 9(10), 1165-1177. doi:10.2217/fmb.14.66

Blyth, G. A., Connors, L., Fodor, C., & Cobo, E. R. (2020). The network of Colonic host defense peptides as an innate immune defense against enteropathogenic bacteria. Frontiers in Immunology, 11. doi:10.3389/fimmu.2020.00965

Boyen, F., Pasmans, F., Van Immerseel, F., Donné, E., Morgan, E., Ducatelle, R., & Haesebrouck, F. (2009). Porcine in vitro and in vivo models to assess the virulence of salmonella enterica serovar typhimurium for pigs. Laboratory Animals, 43(1), 46-52. doi:10.1258/la.2007.007084

Bozovičar, K., & Bratkovič, T. (2019). Evolving a peptide: Library platforms and diversification strategies. International Journal of Molecular Sciences, 21(1), 215. doi:10.3390/ijms21010215

Brenner, F. W., Villar, R. G., Angulo, F. J., Tauxe, R., & Swaminathan, B. (2000). Salmonella nomenclature. Journal of Clinical Microbiology, 38(7), 2465-2467. doi:10.1128/jcm.38.7.2465-2467.2000

Campbell, E. A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A., & Darst, S. A. (2001). Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. Cell, 104(6), 901-912. doi:10.1016/s0092-8674(01)00286-0

Campos, J., Mourão, J., Peixe, L., & Antunes, P. (2019). Non-typhoidal salmonella in the pig production chain: A comprehensive analysis of its impact on human health. Pathogens, 8(1), 19. doi:10.3390/pathogens8010019

Canibe, N., Steien, S. H., Overland, M., & Jensen, B. B. (2001). Effect of K-diformate in starter diets on acidity, microbiota, and the amount of organic acids in the

digestive tract of piglets, and on gastric alterations. Journal of Animal Science, 79(8), 2123. doi:10.2527/2001.7982123x

Cao, M., & Helmann, J. D. (2004). The bacillus subtilis extracytoplasmic-function σ factor regulates modification of the cell envelope and resistance to cationic antimicrobial peptides. Journal of Bacteriology, 186(4), 1136-1146. doi:10.1128/jb.186.4.1136-1146.2004

Casas, V., Rodríguez-Asiain, A., Pinto-Llorente, R., Vadillo, S., Carrascal, M., & Abian, J. (2017). Brachyspira hyodysenteriae and B. Pilosicoli proteins recognized by Sera of Challenged Pigs. Frontiers in Microbiology, 8. doi:10.3389/fmicb.2017.00723

Casey, P. G., Gardiner, G. E., Casey, G., Bradshaw, B., Lawlor, P. G., Lynch, P. B., . . . Hill, C. (2007). A five-strain probiotic combination reduces pathogen shedding and alleviates disease signs in pigs challenged with salmonella enterica serovar typhimurium. Applied and Environmental Microbiology, 73(6), 1858-1863. doi:10.1128/aem.01840-06

Castel, G., Chtéoui, M., Heyd, B., & Tordo, N. (2011). Phage display of combinatorial peptide libraries: Application to antiviral research. Molecules, 16(5), 3499-3518. doi:10.3390/molecules16053499

Carvajal, F. (1947). Screening tests for antibiotics. Mycologia, 39(1), 128. doi:10.2307/3755295

Cevallos-Almeida, M., Martin, L., Houdayer, C., Rose, V., Guionnet, J., Paboeuf, F., . . . Kerouanton, A. (2019). Experimental infection of pigs by Salmonella Derby, S. typhimurium and monophasic variant of S. Typhimurium: Comparison of colonization and Serology. Veterinary Microbiology, 231, 147-153. doi:10.1016/j.vetmic.2019.03.003

Chantziaras, I., Boyen, F., Callens, B., & Dewulf, J. (2013). Correlation between veterinary antimicrobial use and antimicrobial resistance in food-producing animals: A report on seven countries. Journal of Antimicrobial Chemotherapy, 69(3), 827-834. doi:10.1093/jac/dk*T4*43

Che, L., Hu, Q., Wang, R., Zhang, D., Liu, C., Zhang, Y., . . . Gao, F. (2019). Inter-correlated gut microbiota and scfas changes upon antibiotics exposure links with rapid body-mass gain in weaned piglet model. The Journal of Nutritional Biochemistry, 74, 108246. doi:10.1016/j.jnutbio.2019.108246

Chen, M. M., Snow, C. D., Vizcarra, C. L., Mayo, S. L., & Arnold, F. H. (2012). Comparison of random mutagenesis and semi-rational designed libraries for improved cytochrome P450 BM3-catalyzed hydroxylation of small alkanes. Protein Engineering Design and Selection, 25(4), 171-178. doi:10.1093/protein/gzs004

Chen, C.-Y., Nace, G. W., & Irwin, P. L. (2003). A 6×6 drop plate method for simultaneous colony counting and MPN enumeration of campylobacter jejuni, listeria

monocytogenes, and escherichia coli. Journal of Microbiological Methods, 55(2), 475–479. doi:10.1016/s0167-7012(03)00194-5

Chopin, M., Rouault, A., Ehrlich, S. D., & Gautier, M. (2002). Filamentous phage active on the gram-positive bacterium propionibacterium freudenreichii. Journal of Bacteriology, 184(7), 2030-2033. doi:10.1128/jb.184.7.2030-2033.2002

Citron, D. M., Tyrrell, K. L., & Leoncio, E. S. (2017). In vitro activity of Pexiganan and 10 comparator antimicrobials against 234 isolates, including 93 Pasteurella species and 50 anaerobic bacterial isolates recovered from animal bite wounds. Antimicrobial Agents and Chemotherapy, 61(6). doi:10.1128/aac.00246-17

Clark, C. G., Landgraff, C., Robertson, J., Pollari, F., Parker, S., Nadon, C., . . . Nash, J. (2020). Distribution of heavy metal resistance elements in Canadian salmonella 4,[5],12:I:- populations and association with the monophasic genotypes and phenotype. PLOS ONE, 15(7). doi:10.1371/journal.pone.0236436

CLSI (2019). Performance Standards for Antimicrobial Susceptibility Testing. M100, ED29. Wayne, PA: Clinical and Laboratory Standards Institute.

Costa, S., Ntokou, E., Martins, A., Viveiros, M., Pournaras, S., Couto, I., & Amaral, L. (2010). Identification of the plasmid-encoded QACA efflux pump gene in meticillin-resistant Staphylococcus aureus (MRSA) strain HPV107, a representative of the MRSA Iberian clone. International Journal of Antimicrobial Agents, 36(6), 557-561. doi:10.1016/j.ijantimicag.2010.08.006

Courvalin, P. (2006). Vancomycin resistance in gram-positive cocci. Clinical Infectious Diseases, 42(Supplement_1). doi:10.1086/491711

Crossley, B. M., Bai, J., Glaser, A., Maes, R., Porter, E., Killian, M. L., … Toohey-Kurth, K. (2020). Guidelines for Sanger sequencing and Molecular Assay Monitoring. Journal of Veterinary Diagnostic Investigation, 32(6), 767–775. doi:10.1177/1040638720905833

Currin, A., Swainston, N., Day, P. J., & Kell, D. B. (2015). Synthetic Biology for the directed evolution of protein biocatalysts: Navigating sequence space intelligently. Chemical Society Reviews, 44(5), 1172-1239. doi:10.1039/c4cs00351a

Côté, S., Letellier, A., Lessard, L., & Quessy, S. (2004). Distribution of Salmonella in tissues following natural and experimental infection in pigs. Distribution of Salmonella in Tissues following Natural and Experimental Infection in Pigs, (68), 4th ser., 241–248.

Da Costa, J. P., Cova, M., Ferreira, R., & Vitorino, R. (2015). Antimicrobial peptides: An alternative for innovative medicines? Applied Microbiology and Biotechnology, 99(5), 2023-2040. doi:10.1007/s00253-015-6375-x

Deane, A., Murphy, D., Leonard, F. C., Byrne, W., Clegg, T., Madigan, G., . . . Prendergast, D. M. (2022). Prevalence of salmonella spp. in slaughter pigs and

carcasses in Irish abattoirs and their antimicrobial resistance. Irish Veterinary Journal, 75(1). doi:10.1186/s13620-022-00211-y

Desiree, K., Mosimann, S., & Ebner, P. (2021). Efficacy of phage therapy in Pigs: Systematic Review and meta-analysis. Journal of Animal Science, 99(7). doi:10.1093/jas/skab157

Destoumieux-Garzón, D., Mavingui, P., Boetsch, G., Boissier, J., Darriet, F., Duboz, P., . . . Voituron, Y. (2018). The One Health Concept: 10 years old and A long road ahead. Frontiers in Veterinary Science, 5. doi:10.3389/fvets.2018.00014

D'Incau, M., Salogni, C., Giovannini, S., Ruggeri, J., Scali, F., Tonni, M., . . . Alborali, G. L. (2021). Analysis of the comparative occurrence of salmonella typhimurium and its monophasic variant (4,[5],12:I:-) in healthy and clinically ill pigs in northern Italy. doi:10.21203/rs.3.rs-127918/v1

Domenyuk, V., Loskutov, A., Johnston, S. A., & Diehnelt, C. W. (2013). A technology for developing synbodies with antibacterial activity. PLoS ONE, 8(1). doi:10.1371/journal.pone.0054162

Domingues, M. M., Inácio, R. G., Raimundo, J. M., Martins, M., Castanho, M. A., & Santos, N. C. (2012). Biophysical characterization of polymyxin B interaction with LPS aggregates and membrane model systems. Biopolymers, 98(4), 338-344. doi:10.1002/bip.22095

Drlica, K., & Zhao, X. (1997). DNA gyrase, Topoisomerase IV, and the 4-quinolones. Microbiology and Molecular Biology Reviews, 61(3), 377-392. doi:10.1128/mmbr.61.3.377-392.1997

Dutkiewicz, J., Sroka, J., Zając, V., Wasiński, B., Cisak, E., Sawczyn, A., . . . Wójcik-Fatla, A. (2017). Streptococcus suis: A re-emerging pathogen associated with occupational exposure to pigs or pork products. part I – epidemiology. Annals of Agricultural and Environmental Medicine, 24(4), 683-695. doi:10.26444/aaem/79813

E. Greber, K., & Dawgul, M. (2016). Antimicrobial peptides under clinical trials. Current Topics in Medicinal Chemistry, 17(5), 620-628. doi:10.2174/1568026616666160713143331

Ebrahimizadeh, W., & Rajabibazl, M. (2014). Bacteriophage vehicles for phage display: Biology, mechanism, and application. Current Microbiology, 69(2), 109-120. doi:10.1007/s00284-014-0557-0

Epand, R. M., & Epand, R. F. (2009). Lipid domains in bacterial membranes and the action of antimicrobial agents. Biochimica Et Biophysica Acta (BBA) - Biomembranes, 1788(1), 289-294. doi:10.1016/j.bbamem.2008.08.023

Evangelopoulou, G., Kritas, S., Christodoulopoulos, G., & Burriel, A. R. (2015). The commercial impact of pig salmonella spp. infections in border-free markets during

an economic recession. Veterinary World, 8(3), 257-272. doi:10.14202/vetworld.2015.257-272

Fanzo, J., Davis, C., McLaren, R., & Choufani, J. (2018). The effect of climate change across food systems: Implications for Nutrition Outcomes. Global Food Security, 18, 12-19. doi:10.1016/j.gfs.2018.06.001

Farzan, A., & Friendship, R. M. (2009). Vaccination to control salmonella shedding and improve growth in Pigs. International Conference on the Epidemiology and Control of Biological, Chemical and Physical Hazards in Pigs and Pork. doi:10.31274/safepork-180809-801

Ferreira Amaral, M., Frigotto, L., & Hine, A. (2017). Beyond the natural proteome. Methods in Enzymology, 111-133. doi:10.1016/bs.mie.2016.10.005

Fittipaldi, N., Segura, M., Grenier, D., & Gottschalk, M. (2012). Virulence factors involved in the pathogenesis of the infection caused by the swine pathogen and Zoonotic Agent streptococcus suis. Future Microbiology, 7(2), 259-279. doi:10.2217/fmb.11.149

Frigotto, L., Smith, M., Brankin, C., Sedani, A., Cooper, S., Kanwar, N., . . . Hine, A. (2015). Codon-precise, synthetic, antibody fragment libraries built using automated hexamer codon additions and validated through Next Generation Sequencing. Antibodies, 4(2), 88-102. doi:10.3390/antib4020088

Fukunaga, K., & Taki, M. (2012). Practical tips for construction of custom peptide libraries and affinity selection by using commercially available Phage Display Cloning Systems. Journal of Nucleic Acids, 2012, 1-9. doi:10.1155/2012/295719

Garcia, S. N., Osburn, B. I., & Jay-Russell, M. T. (2020). One Health for Food Safety, food security, and Sustainable Food Production. Frontiers in Sustainable Food Systems, 4. doi:10.3389/fsufs.2020.00001

García, P., Malorny, B., Rodicio, M. R., Stephan, R., Hächler, H., Guerra, B., & Lucarelli, C. (2016). Horizontal acquisition of a multidrug-resistance module (R-type assut) is responsible for the monophasic phenotype in a widespread clone of salmonella serovar 4,[5],12:I:-. Frontiers in Microbiology, 7. doi:10.3389/fmicb.2016.00680

García-Gil, A., Galán-Enríquez, C. S., Pérez-López, A., Nava, P., Alpuche-Aranda, C., & Ortiz-Navarrete, V. (2018). SOPB activates the AKT-yap pathway to promote salmonella survival within B cells. Virulence, 9(1), 1390-1402. doi:10.1080/21505594.2018.1509664

Gavin, C., Simons, R., Berriman, A., Moorhouse, D., Snary, E., Smith, R., & Hill, A. (2018). A cost-benefit assessment of salmonella-control strategies in pigs reared in the United Kingdom. Preventive Veterinary Medicine, 160, 54-62. doi:10.1016/j.prevetmed.2018.09.022

Gaytán, P., Contreras-Zambrano, C., Ortiz-Alvarado, M., Morales-Pablos, A., & Yáñez, J. (2009). Trimerdimer: An oligonucleotide-based saturation mutagenesis approach that removes redundant and stop codons. Nucleic Acids Research, 37(18). doi:10.1093/nar/gkp602

Glukhov, E., Stark, M., Burrows, L. L., & Deber, C. M. (2005). Basis for selectivity of cationic antimicrobial peptides for bacterial versus mammalian membranes. Journal of Biological Chemistry, 280(40), 33960-33967. doi:10.1074/jbc.m507042200

Goyette-Desjardins, G., Auger, J., Xu, J., Segura, M., & Gottschalk, M. (2014). Streptococcus suis, an important pig pathogen and emerging zoonotic agent—an update on the worldwide distribution based on serotyping and sequence typing. Emerging Microbes & Infections, 3(1), 1-20. doi:10.1038/emi.2014.45

Grada, A., & Weinbrecht, K. (2013). Next-generation sequencing: Methodology and application. Journal of Investigative Dermatology, 133(8), 1–4. doi:10.1038/jid.2013.248

GROVES, M. L., PETERSON, R. F., & KIDDY, C. A. (1965). Polymorphism in the red protein isolated from milk of individual cows. Nature, 207(5000), 1007-1008. doi:10.1038/2071007a0

Græsbøll, K., Larsen, I., Clasen, J., Birkegård, A. C., Nielsen, J. P., Christiansen, L. E., . . . Folkesson, A. (2019). Effect of tetracycline treatment regimens on antibiotic resistance gene selection over time in Nursery Pigs. BMC Microbiology, 19(1). doi:10.1186/s12866-019-1619-z

Guralp, S. A., Murgha, Y. E., Rouillard, J., & Gulari, E. (2013). From design to screening: A new antimicrobial peptide discovery pipeline. PLoS ONE, 8(3). doi:10.1371/journal.pone.0059305

Guyomard, H., Darcy-Vrillon, B., Esnouf, C., Marin, M., Russel, M., & Guillou, M. (2012). Eating patterns and food systems: Critical knowledge requirements for policy design and implementation. Agriculture & Food Security, 1(1). doi:10.1186/2048-7010-1-13

Hammerum, A. M., Larsen, J., Andersen, V. D., Lester, C. H., Skovgaard Skytte, T. S., Hansen, F., . . . Agersø, Y. (2014). Characterization of extended-spectrum β-lactamase (esbl)-producing escherichia coli obtained from Danish pigs, pig farmers and their families from farms with high or no consumption of third- or fourth-generation cephalosporins. Journal of Antimicrobial Chemotherapy, 69(10), 2650-2657. doi:10.1093/jac/dku180

Hansen, J. L., Ippolito, J. A., Ban, N., Nissen, P., Moore, P. B., & Steitz, T. A. (2002). The structures of four macrolide antibiotics bound to the large ribosomal subunit. Molecular Cell, 10(1), 117-128. doi:10.1016/s1097-2765(02)00570-1

Harris, F., & Pierpoint, L. (2011). Photodynamic therapy based on 5-aminolevulinic acid and its use as an antimicrobial agent. Medicinal Research Reviews, 32(6), 1292-1327. doi:10.1002/med.20251

Helke, K. L., Nelson, K. N., Sargeant, A. M., Jacob, B., McKeag, S., Haruna, J., . . . Hollinger, C. (2015). Background pathological changes in minipigs. Toxicologic Pathology, 44(3), 325-337. doi:10.1177/0192623315611762

Hirota, Y., Suzuki, H., Nishimura, Y., & Yasuda, S. (1977). On the process of cellular division in escherichia coli: A mutant of E. coli lacking a murein-lipoprotein. Proceedings of the National Academy of Sciences, 74(4), 1417–1420. doi:10.1073/pnas.74.4.1417

Hoffman, D. (2019). The double burden of malnutrition: Research agenda for reversing global trends. Annals of Nutrition and Metabolism, 75(2), 149-152. doi:10.1159/000503676

Huan, Y., Kong, Q., Mou, H., & Yi, H. (2020). Antimicrobial peptides: Classification, design, application and research progress in multiple fields. Frontiers in Microbiology, 11. doi:10.3389/fmicb.2020.582779

Hughes, J., Nghia, H., Taylor, W., Schultsz, C., & Wertheim, H. (2009). Streptococcus suis:an emerging human pathogen. Clinical Infectious Diseases, 48(5), 617-625. doi:10.1086/596763

Hughes, M. D., Nagel, D. A., Santos, A. F., Sutherland, A. J., & Hine, A. V. (2003). Removing the redundancy from randomised gene libraries. Journal of Molecular Biology, 331(5), 973-979. doi:10.1016/s0022-2836(03)00833-7

Hudzicki J. Kirby-Bauer disk diffusion susceptibility test protocol. American Society for Microbiology; Washington, DC: [Online.] http://www.microbelibrary.org/component/resource/laboratory-test/3189-kirby-bauer-disk-diffusion-susceptibility-test-protocol.

Højberg, O., Canibe, N., Poulsen, H. D., Hedemann, M. S., & Jensen, B. B. (2005). Influence of dietary zinc oxide and copper sulfate on the gastrointestinal ecosystem in newly weaned piglets. Applied and Environmental Microbiology, 71(5), 2267-2277. doi:10.1128/aem.71.5.2267-2277.2005

Innis, M. A., Gelfand, D. H., Sninsky, J. J., & White, T. J. (1990). PCR protocols a guide to methods and applications. Burlington: Elsevier Science.

Jenssen, H., Hamill, P., & Hancock, R. E. (2006). Peptide antimicrobial agents. Clinical Microbiology Reviews, 19(3), 491-511. doi:10.1128/cmr.00056-05

Jiang, Z., Anwar, T. M., Peng, X., Biswas, S., Elbediwi, M., Li, Y., . . . Yue, M. (2021). Prevalence and antimicrobial resistance of salmonella recovered from pig-borne food products in Henan, China. Food Control, 121, 107535. doi:10.1016/j.foodcont.2020.107535

Kapoor, G., Saigal, S., & Elongavan, A. (2017). Action and resistance mechanisms of antibiotics: A guide for clinicians. Journal of Anaesthesiology Clinical Pharmacology, 33(3), 300. doi:10.4103/joacp.joacp_349_15

Kaspar, A. A., & Reichert, J. M. (2013). Future Directions for Peptide Therapeutics Development. Drug Discovery Today, 18(17-18), 807-817. doi:10.1016/j.drudis.2013.05.011

Kasper, C. A., Sorg, I., Schmutz, C., Tschon, T., Wischnewski, H., Kim, M. L., & Arrieumerlou, C. (2010). Cell-cell propagation of NF-KB transcription factor and MAP kinase activation amplifies innate immunity against bacterial infection. Immunity, 33(5), 804-816. doi:10.1016/j.immuni.2010.10.015

Katsuda, K., Kohmoto, M., Kawashima, K., & Tsunemitsu, H. (2006). Frequency of enteropathogen detection in suckling and weaned pigs with diarrhea in Japan. Journal of Veterinary Diagnostic Investigation, 18(4), 350-354. doi:10.1177/104063870601800405

Khondker, A., & Rheinstädter, M. C. (2020). How do bacterial membranes resist polymyxin antibiotics? Communications Biology, 3(1). doi:10.1038/s42003-020-0803-x

Kille, S., Acevedo-Rocha, C. G., Parra, L. P., Zhang, Z., Opperman, D. J., Reetz, M. T., & Acevedo, J. P. (2012). Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. ACS Synthetic Biology, 2(2), 83-92. doi:10.1021/sb300037w

Kim, S. I., Kim, S., Kim, E., Hwang, S. Y., & Yoon, H. (2018). Secretion of salmonella pathogenicity island 1-encoded type III secretion system effectors by outer membrane vesicles in salmonella enterica serovar typhimurium. Frontiers in Microbiology, 9. doi:10.3389/fmicb.2018.02810

King, S. J., Leigh, J. A., Heath, P. J., Luque, I., Tarradas, C., Dowson, C. G., & Whatmore, A. M. (2002). Development of a multilocus sequence typing scheme for the pig pathogen streptococcus suis : Identification of virulent clones and potential capsular serotype exchange. Journal of Clinical Microbiology, 40(10), 3671-3680. doi:10.1128/jcm.40.10.3671-3680.2002

Kiss, G., & Michl, H. (1962). Uber Das giftsekret Der Gelbbauchunke, Bombina variegata L. Toxicon, 1(1), 33-34. doi:10.1016/0041-0101(62)90006-5

Kontermann, R., & Dübel, S. (2010). Antibody engineering. Berlin: Springer.

Kosikowska, P., & Lesner, A. (2016). Antimicrobial peptides (AMPS) as drug candidates: A Patent Review (2003–2015). Expert Opinion on Therapeutic Patents, 26(6), 689-702. doi:10.1080/13543776.2016.1176149

Kralik, Z., Kralik, G., Grčević, M., Suchý, P., & Straková, E. (2012). Effects of increased content of organic selenium in feed on the selenium content and fatty acid

profile in broiler breast muscle. Acta Veterinaria Brno, 81(1), 31–35. doi:10.2754/avb201281010031

Kretz, C. A., Tomberg, K., Van Esbroeck, A., Yee, A., & Ginsburg, D. (2018). High throughput protease profiling comprehensively defines active site specificity for thrombin and ADAMTS13. Scientific Reports, 8(1). doi:10.1038/s41598-018-21021-9

Krzyściak, W., Pluskwa, K. K., Jurczak, A., & Kościelniak, D. (2013). The pathogenicity of the streptococcus genus. European Journal of Clinical Microbiology & Infectious Diseases, 32(11), 1361-1376. doi:10.1007/s10096-013-1914-9

Lane, M. D., & Seelig, B. (2014). Advances in the directed evolution of proteins. Current Opinion in Chemical Biology, 22, 129-136. doi:10.1016/j.cbpa.2014.09.013

Lapointe, G., Skepper, C. K., Holder, L. M., Armstrong, D., Bellamacina, C., Blais, J., . . . Rivkin, A. (2021). Discovery and optimization of DNA gyrase and topoisomerase IV inhibitors with potent activity against fluoroquinolone-resistant gram-positive bacteria. Journal of Medicinal Chemistry, 64(9), 6329-6357. doi:10.1021/acs.jmedchem.1c00375

Larsson, O. (2002). Quantitative codon optimisation of DNA libraries encoding sub-random peptides: Design and characterisation of a novel library encoding transmembrane domain peptides. Nucleic Acids Research, 30(23). doi:10.1093/nar/gnf133

Lazzaro, B. P., Zasloff, M., & Rolff, J. (2020). Antimicrobial peptides: Application informed by evolution. Science, 368(6490). doi:10.1126/science.aau5480

Lekagul, A., Tangcharoensathien, V., & Yeung, S. (2019). Patterns of antibiotic use in global pig production: A systematic review. Veterinary and Animal Science, 7, 100058. doi:10.1016/j.vas.2019.100058

Lekagul, A., Tangcharoensathien, V., Mills, A., Rushton, J., & Yeung, S. (2020). How antibiotics are used in pig farming: A mixed-methods study of pig farmers, Feed Mills and veterinarians in Thailand. BMJ Global Health, 5(2). doi:10.1136/bmjgh-2019-001918

Li, B., Yin, F., Zhao, X., Guo, Y., Wang, W., Wang, P., . . . Wang, X. (2020). Colistin resistance gene MCR-1 mediates cell permeability and resistance to hydrophobic antibiotics. Frontiers in Microbiology, 10. doi:10.3389/fmicb.2019.03015

Lin, L., Chi, J., Yan, Y., Luo, R., Feng, X., Zheng, Y., . . . Pan, X. (2021). Membrane-disruptive peptides/peptidomimetics-based therapeutics: Promising systems to combat bacteria and cancer in the drug-resistant era. Acta Pharmaceutica Sinica B, 11(9), 2609-2644. doi:10.1016/j.apsb.2021.07.014

Liu, Y., Han, F., Xie, Y., & Wang, Y. (2011). Comparative antimicrobial activity and mechanism of action of bovine lactoferricin-derived synthetic peptides. BioMetals, 24(6), 1069–1078. doi:10.1007/s10534-011-9465-y

Lunha, K., Chumpol, W., Samngamnim, S., Jiemsup, S., Assavacheep, P., & Yongkiettrakul, S. (2022). Antimicrobial susceptibility of streptococcus suis isolated from diseased pigs in Thailand, 2018–2020. Antibiotics, 11(3), 410. doi:10.3390/antibiotics11030410

Macheboeuf, P., Contreras-Martel, C., Job, V., Dideberg, O., & Dessen, A. (2006). Penicillin binding proteins: Key players in bacterial cell cycle and drug resistance processes. FEMS Microbiology Reviews, 30(5), 673-691. doi:10.1111/j.1574-6976.2006.00024.x

Mai-Prochnow, A., Hui, J. G., Kjelleberg, S., Rakonjac, J., McDougald, D., & Rice, S. A. (2015). 'big things in small packages: The genetics of filamentous phage and effects on fitness of their host'. FEMS Microbiology Reviews, 39(4), 465-487. doi:10.1093/femsre/fuu007

Majowicz, S., Musto, J., Scallan, E., Angulo, F., Kirk, M., O'Brien, S., . . . Hoekstra, R. (2010). The global burden of nontyphoidal salmonella gastroenteritis. Clinical Infectious Diseases, 50(6), 882-889. doi:10.1086/650733

Makarova, O., Johnston, P., Rodriguez-Rojas, A., El Shazely, B., Morales, J. M., & Rolff, J. (2018). Genomics of experimental adaptation of Staphylococcus aureus to a natural combination of insect antimicrobial peptides. Scientific Reports, 8(1). doi:10.1038/s41598-018-33593-7

Mangoni, M. L., & Casciaro, B. (2020). Development of antimicrobial peptides from amphibians. Antibiotics, 9(11), 772. doi:10.3390/antibiotics9110772

Matiašovic, J., Nedbalcová, K., Žižlavský, M., Fleischer, P., Pokludová, L., Kellnerová, D., . . . Šlosárková, S. (2021). Streptococcus suis isolates—serotypes and susceptibility to antimicrobials in terms of their use on selected repopulated Czech pig farms. Pathogens, 10(10), 1314. doi:10.3390/pathogens10101314

McLaughlin, M. R., & Balaa, M. F. (2006). Enhanced contrast of bacteriophage plaques in salmonella with ferric ammonium citrate and sodium thiosulfate (FACST) and Tetrazolium Red (TZR). Journal of Microbiological Methods, 65(2), 318–323. doi:10.1016/j.mimet.2005.08.008

Mercer, D. K., Torres, M. D., Duay, S. S., Lovie, E., Simpson, L., Von Köckritz-Blickwede, M., . . . Angeles-Boza, A. M. (2020). Antimicrobial susceptibility testing of antimicrobial peptides to better predict efficacy. Frontiers in Cellular and Infection Microbiology, 10. doi:10.3389/fcimb.2020.00326

Mesonero-Escuredo, S., Strutzberg-Minder, K., Casanovas, C., & Segalés, J. (2018). Viral and bacterial investigations on the aetiology of recurrent pig neonatal

diarrhoea cases in Spain. Porcine Health Management, 4(1). doi:10.1186/s40813-018-0083-8

Miles, A. A., Misra, S. S., & Irwin, J. O. (1938). The estimation of the bactericidal power of the blood. Epidemiology and Infection, 38(6), 732–749. doi:10.1017/s002217240001158x

Miller, K. W., Schamber, R., Osmanagaoglu, O., & Ray, B. (1998). Isolation and characterization of Pediocin Ach chimeric protein mutants with altered bactericidal activity. Applied and Environmental Microbiology, 64(6), 1997–2005. doi:10.1128/aem.64.6.1997-2005.1998

Mittal, D., Grakh, K., Prakash, A., Moudgil, P., Devi, B., & Jadhav, V. (2018). Isolation, identification and characterization of enteric bacteria from post weaning diarrheic pigs and their resistance to multiple antibiotics. International Journal of Current Microbiology and Applied Sciences, 7(12), 2377-2384. doi:10.20546/ijcmas.2018.712.269

Molchanova, N., Wang, H., Hansen, P. R., Høiby, N., Nielsen, H. M., & Franzyk, H. (2019). Antimicrobial activity of α-peptide/β-peptoid lysine-based peptidomimetics against colistin-resistant pseudomonas aeruginosa isolated from cystic fibrosis patients. Frontiers in Microbiology, 10. doi:10.3389/fmicb.2019.00275

Monger, X. C., Gilbert, A., Saucier, L., & Vincent, A. T. (2021). Antibiotic resistance: From pig to meat. Antibiotics, 10(10), 1209. doi:10.3390/antibiotics10101209

Moura, E. A., Silva, D. G., Turco, C. H., Sanches, T. V., Storino, G. Y., Almeida, H. M., . . . Oliveira, L. G. (2021). Salmonella bacterin vaccination decreases shedding and colonization of salmonella typhimurium in Pigs. Microorganisms, 9(6), 1163. doi:10.3390/microorganisms9061163

Naqid, I. A., Owen, J. P., Maddison, B. C., Gardner, D. S., Foster, N., Tchórzewska, M. A., . . . Gough, K. C. (2015). Prebiotic and probiotic agents enhance antibody-based immune responses to salmonella typhimurium infection in Pigs. Animal Feed Science and Technology, 201, 57-65. doi:10.1016/j.anifeedsci.2014.12.005

Nhung, N., Cuong, N., Thwaites, G., & Carrique-Mas, J. (2016). Antimicrobial usage and antimicrobial resistance in animal production in Southeast Asia: A Review. Antibiotics, 5(4), 37. doi:10.3390/antibiotics5040037

Niemeyer-van der Kolk, T., Wall, H., Hogendoorn, G. K., Rijneveld, R., Luijten, S., Alewijk, D. C., . . . Doorn, M. B. (2020). Pharmacodynamic effects of topical Omiganan in patients with mild to moderate atopic dermatitis in a randomized, placebo-controlled, phase II trial. Clinical and Translational Science. doi:10.1111/cts.12792

Niemi, J., Bennett, R., Clark, B., Frewer, L., Jones, P., RimmLer, T., & Tranter, R. (2020). A value chain analysis of interventions to control production diseases in the

intensive pig production sector. PLOS ONE, 15(4). doi:10.1371/journal.pone.0231338

Nilsson, N., Malmborg, A., & Borrebaeck, C. A. (2000). The phage infection process: A functional role for the distal linker region of bacteriophage protein 3. Journal of Virology, 74(9), 4229-4235. doi:10.1128/jvi.74.9.4229-4235.2000

O. (n.d.). Pigmeat production. Retrieved July 15, 2022, from https://ourworldindata.org/grapher/pigmeat-production-tonnes

Oh, M., Joo, H., Hur, B., Jeong, Y., & Cha, S. (2007). Enhancing phage display of antibody fragments using GIII-amber suppression. Gene, 386(1–2), 81–89. doi:10.1016/j.gene.2006.08.009

Olvera, A., Segalés, J., & Aragón, V. (2007). Update on the diagnosis of haemophilus parasuis infection in pigs and novel genotyping methods. The Veterinary Journal, 174(3), 522-529. doi:10.1016/j.tvjl.2006.10.017

Opriessnig, T., Giménez-Lirola, L. G., & Halbur, P. G. (2011). Polymicrobial respiratory disease in Pigs. Animal Health Research Reviews, 12(2), 133-148. doi:10.1017/s1466252311000120

Otvos, L., & Wade, J. D. (2014). Current challenges in peptide-based drug discovery. Frontiers in Chemistry, 2. doi:10.3389/fchem.2014.00062

Papich, M. (2016). Important contact information for veterinary drugs. Saunders Handbook of Veterinary Drugs. doi:10.1016/b978-0-323-24485-5.00056-5

Perez-Escamilla, R., Bermudez, O., Buccini, G. S., Kumanyika, S., Lutter, C. K., Monsivais, P., & Victora, C. (2018). Nutrition disparities and the global burden of malnutrition. BMJ. doi:10.1136/bmj.k2252

Petrocchi-Rilo, M., Martínez-Martínez, S., Aguarón-Turrientes, Á, Roca-Martínez, E., García-Iglesias, M., Pérez-Fernández, E., . . . Gutiérrez-Martín, C. (2021). Anatomical site, typing, virulence gene profiling, antimicrobial susceptibility and resistance genes of streptococcus suis isolates recovered from pigs in Spain. Antibiotics, 10(6), 707. doi:10.3390/antibiotics10060707

Petrovska, L., Mather, A. E., AbuOun, M., Branchu, P., Harris, S. R., Connor, T., . . . Kingsley, R. A. (2016). Microevolution of monophasicsalmonellatyphimurium during epidemic, United Kingdom, 2005–2010. Emerging Infectious Diseases, 22(4), 617-624. doi:10.3201/eid2204.150531

Piddock, L. J. (2006). Multidrug-resistance efflux pumps ? not just for resistance. Nature Reviews Microbiology, 4(8), 629-636. doi:10.1038/nrmicro1464

Pluske, J. R., Pethick, D. W., Hopwood, D. E., & Hampson, D. J. (2002). Nutritional influences on some major enteric bacterial diseases of pig. Nutrition Research Reviews, 15(2), 333-371. doi:10.1079/nrr200242

Pokharel, S., Shrestha, P., & Adhikari, B. (2020). Antimicrobial use in food animals and human health: Time to implement ‘one health’ approach. Antimicrobial Resistance & Infection Control, 9(1). doi:10.1186/s13756-020-00847-x

Prosekov, A. Y., & Ivanova, S. A. (2018). Food security: The challenge of the present. Geoforum, 91, 73-77. doi:10.1016/j.geoforum.2018.02.030

Pushpanathan, M., Gunasekaran, P., & Rajendhran, J. (2013). Antimicrobial peptides: Versatile biological properties. International Journal of Peptides, 2013, 1-15. doi:10.1155/2013/675391

Qi, H., Lu, H., Qiu, H., Petrenko, V., & Liu, A. (2012). Phagemid vectors for phage display: Properties, characteristics and construction. Journal of Molecular Biology, 417(3), 129-143. doi:10.1016/j.jmb.2012.01.038

Quax, T., Claassens, N., Söll, D., & Van der Oost, J. (2015). Codon bias as a means to fine-tune gene expression. Molecular Cell, 59(2), 149-161. doi:10.1016/j.molcel.2015.05.035

Rakonjac, & Bennett,. (2009). Filamentous bacteriophage: Biology, phage display and nanotechnology applications. Current Issues in Molecular Biology. doi:10.21775/cimb.013.051

Rami, A., Behdani, M., Yardehnavi, N., Habibi-Anbouhi, M., & Kazemi-Lomedasht, F. (2017). An overview on application of Phage Display Technique in immunological studies. Asian Pacific Journal of Tropical Biomedicine, 7(7), 599-602. doi:10.1016/j.apjtb.2017.06.001

Reetz, M. T., Kahakeaw, D., & Lohmer, R. (2008). Addressing the numbers problem in directed evolution. ChemBioChem, 9(11), 1797-1804. doi:10.1002/cbic.200800298

Reetz, M. T., Prasad, S., Carballeira, J. D., Gumulya, Y., & Bocola, M. (2010). Iterative saturation mutagenesis accelerates laboratory evolution of enzyme stereoselectivity: Rigorous comparison with traditional methods. Journal of the American Chemical Society, 132(26), 9144-9152. doi:10.1021/ja1030479

Rumio, C., Besusso, D., Palazzo, M., Selleri, S., Sfondrini, L., Dubini, F., . . . Balsari, A. (2004). Degranulation of Paneth cells via toll-like receptor 9. The American Journal of Pathology, 165(2), 373-381. doi:10.1016/s0002-9440(10)63304-4

Saar-Dover, R., Bitler, A., Nezer, R., Shmuel-Galia, L., Firon, A., Shimoni, E., . . . Shai, Y. (2012). D-alanylation of lipoteichoic acids confers resistance to cationic peptides in group B streptococcus by increasing the cell wall density. PLoS Pathogens, 8(9). doi:10.1371/journal.ppat.1002891

Sabat, A. J., Tinelli, M., Grundmann, H., Akkerboom, V., Monaco, M., Del Grosso, M., . . . Friedrich, A. W. (2018). Daptomycin resistant Staphylococcus aureus clinical strain with novel non-synonymous mutations in the MPRF and VRAS genes: A new

insight into daptomycin resistance. Frontiers in Microbiology, 9. doi:10.3389/fmicb.2018.02705

Sachdeva, K., & Sundaramurthy, V. (2020). The interplay of host lysosomes and intracellular pathogens. Frontiers in Cellular and Infection Microbiology, 10. doi:10.3389/fcimb.2020.595502

Sainath Rao, S., Mohan, K. V., & Atreya, C. D. (2013). A peptide derived from phage display library exhibits antibacterial activity against E. coli and pseudomonas aeruginosa. PLoS ONE, 8(2). doi:10.1371/journal.pone.0056081

Sarrazin, S., Joosten, P., Van Gompel, L., Luiken, R. E., Mevius, D. J., Wagenaar, J. A., . . . Stärk, K. (2018). Quantitative and qualitative analysis of antimicrobial usage patterns in 180 selected farrow-to-finish pig farms from nine European countries based on single batch and purchase data. Journal of Antimicrobial Chemotherapy, 74(3), 807-816. doi:10.1093/jac/dky503

Schierack, P., Walk, N., Reiter, K., Weyrauch, K. D., & Wieler, L. H. (2007). Composition of intestinal Enterobacteriaceae populations of healthy domestic pigs. Microbiology, 153(11), 3830-3837. doi:10.1099/mic.0.2007/010173-0

Schmitz, U., Versmold, A., Kaufmann, P., & Frank, H. (2000). Phage display: A molecular tool for the generation of antibodies— a review. Placenta, 21. doi:10.1053/plac.1999.0511

Scott, H. M., Acuff, G., Bergeron, G., Bourassa, M. W., Gill, J., Graham, D. W., . . . Wittum, T. E. (2019). Critically important antibiotics: Criteria and approaches for measuring and reducing their use in food animal agriculture. Annals of the New York Academy of Sciences, 1441(1), 8-16. doi:10.1111/nyas.14058

Segura, M., Calzas, C., Grenier, D., & Gottschalk, M. (2016). Initial steps of the pathogenesis of the infection caused bystreptococcus suis: Fighting against nonspecific defenses. FEBS Letters, 590(21), 3772-3799. doi:10.1002/1873-3468.12364

Shai, Tossi, A., & Zelezetsky, I. (2006). Alpha-helical antimicrobial peptides—using a sequence template to guide structure–activity relationship studies. Biochimica Et Biophysica Acta (BBA) - Biomembranes, 1758(9), 1436-1449. doi:10.1016/j.bbamem.2006.03.021

Shippy, D. C., Bearson, B. L., Holman, D. B., Brunelle, B. W., Allen, H. K., & Bearson, S. M. (2018). Porcine response to a multidrug-resistant salmonella enterica serovar I 4,[5],12:I:- outbreak isolate. Foodborne Pathogens and Disease, 15(5), 253-261. doi:10.1089/fpd.2017.2378

Sieber, T., Hare, E., Hofmann, H., & Trepel, M. (2015). Biomathematical description of synthetic peptide libraries. PLOS ONE, 10(6). doi:10.1371/journal.pone.0129200

Sierra, J. M., Fusté, E., Rabanal, F., Vinuesa, T., & Viñas, M. (2017). An overview of antimicrobial peptides and the latest advances in their development. Expert Opinion on Biological Therapy, 17(6), 663-676. doi:10.1080/14712598.2017.1315402

Silhavy, T. J., Kahne, D., & Walker, S. (2010). The bacterial cell envelope. Cold Spring Harbor Perspectives in Biology, 2(5). doi:10.1101/cshperspect.a000414

Sjölund, M., Postma, M., Collineau, L., Lösken, S., Backhans, A., Belloc, C., . . . Dewulf, J. (2016). Quantitative and qualitative antimicrobial usage patterns in farrow-to-finish pig herds in Belgium, France, Germany and Sweden. Preventive Veterinary Medicine, 130, 41-50. doi:10.1016/j.prevetmed.2016.06.003

Sohlenkamp, C., & Geiger, O. (2015). Bacterial membrane lipids: Diversity in structures and pathways. FEMS Microbiology Reviews, 40(1), 133-159. doi:10.1093/femsre/fuv008

Spencer, P. S., & Barral, J. M. (2012). Genetic code redundancy and its influence on the encoded polypeptides. Computational and Structural Biotechnology Journal, 1(1). doi:10.5936/csbj.201204006

Stellwagen, S. D., Sarkes, D. A., Adams, B. L., Hunt, M. A., Renberg, R. L., Hurley, M. M., & Stratis-Cullum, D. N. (2019). The next generation of biopanning: Next Gen sequencing improves analysis of bacterial display libraries. BMC Biotechnology, 19(1). doi:10.1186/s12896-019-0577-8

Stratton, R. J., King, C. L., Stroud, M. A., Jackson, A. A., & Elia, M. (2006). ‘malnutrition universal screening tool’ predicts mortality and length of hospital stay in acutely ill elderly. British Journal of Nutrition, 95(2), 325-330. doi:10.1079/bjn20051622

Suchsland, R., Appel, B., & Müller, S. (2018). Preparation of trinucleotide phosphoramidites as synthons for the synthesis of gene libraries. Beilstein Journal of Organic Chemistry, 14, 397-406. doi:10.3762/bjoc.14.28

Sun, L., Yang, S., Deng, Q., Dong, K., Li, Y., Wu, S., & Huang, R. (2020). Salmonella effector SpvB disrupts intestinal epithelial barrier integrity for bacterial translocation. Frontiers in Cellular and Infection Microbiology, 10. doi:10.3389/fcimb.2020.606541

Söderström, L., Rosenblad, A., Thors Adolfsson, E., & Bergkvist, L. (2017). Malnutrition is associated with increased mortality in older adults regardless of the cause of death. British Journal of Nutrition, 117(4), 532-540. doi:10.1017/s0007114517000435

Tassinari, E., Bawn, M., Thilliez, G., Charity, O., Acton, L., Kirkwood, M., . . . Kingsley, R. A. (2020). Whole-genome epidemiology links phage-mediated acquisition of a virulence gene to the clonal expansion of a pandemic salmonella

enterica serovar typhimurium clone. Microbial Genomics, 6(11). doi:10.1099/mgen.0.000456

Thanki, A. M., Brown, N., Millard, A. D., & Clokie, M. R. (2019). Genomic characterization of jumbo salmonella phages that effectively target United Kingdom pig-associated salmonella serotypes. Frontiers in Microbiology, 10. doi:10.3389/fmicb.2019.01491

Tikhonova, E., Yamada, Y., & Zgurskaya, H. (2011). Sequential mechanism of assembly of multidrug efflux pump AcrAB-tolc. Chemistry & Biology, 18(4), 454-463. doi:10.1016/j.chembiol.2011.02.011

Tikunova, N. V., & Morozova, V. V. (2009). Phage display on the base of filamentous bacteriophages: Application for recombinant antibodies selection. Acta Naturae, 1(3), 20-28. doi:10.32607/20758251-2009-1-3-20-28

Tominaga, T., & Hatakeyama, Y. (2006). Determination of essential and variable residues in pediocin PA-1 by NNK scanning. Applied and Environmental Microbiology, 72(2), 1141-1147. doi:10.1128/aem.72.2.1141-1147.2006

Tonikian, R., Zhang, Y., Boone, C., & Sidhu, S. S. (2007). Identifying specificity profiles for peptide recognition modules from phage-displayed peptide libraries. Nature Protocols, 2(6), 1368-1386. doi:10.1038/nprot.2007.151

Torres, M. D., Sothiselvam, S., Lu, T. K., & De la Fuente-Nunez, C. (2019). Peptide design principles for antimicrobial applications. Journal of Molecular Biology, 431(18), 3547-3567. doi:10.1016/j.jmb.2018.12.015

Tran, T. H., Everaert, N., & Bindelle, J. (2016). Review on the effects of potential prebiotics on controlling intestinal enteropathogenssalmonellaandescherichia coliin pig production. Journal of Animal Physiology and Animal Nutrition, 102(1), 17-32. doi:10.1111/jpn.12666

Trimble, M. J., MLynárčik, P., Kolář, M., & Hancock, R. E. (2016). Polymyxin: Alternative Mechanisms of action and resistance. Cold Spring Harbor Perspectives in Medicine, 6(10). doi:10.1101/cshperspect.a025288

Tsekouras, N., Athanasakopoulou, Z., Diezel, C., Kostoulas, P., Braun, S. D., Sofia, M., . . . Papatsiros, V. G. (2022). Cross-sectional survey of antibiotic resistance in extended spectrum β-lactamase-producing Enterobacteriaceae isolated from pigs in Greece. Animals, 12(12), 1560. doi:10.3390/ani12121560

Tsoumpeli, M. T., Gray, A., Parsons, A. L., Spiliotopoulos, A., Owen, J. P., Bishop, K., . . . Gough, K. C. (2022). A simple whole-plasmid PCR method to construct high-diversity synthetic phage display libraries. Molecular Biotechnology, 64(7), 791-803. doi:10.1007/s12033-021-00442-4

Tucker, A. T., Leonard, S. P., DuBois, C. D., Knauf, G. A., Cunningham, A. L., Wilke, C. O., … Davies, B. W. (2018). Discovery of next-generation antimicrobials through

bacterial self-screening of surface-displayed peptide libraries. Cell, 172(3). doi:10.1016/j.cell.2017.12.009

Turner, J., Cho, Y., Dinh, N.-N., Waring, A. J., & Lehrer, R. I. (1998). Activities of LL-37, a cathelin-associated antimicrobial peptide of human neutrophils. Antimicrobial Agents and Chemotherapy, 42(9), 2206–2214. doi:10.1128/aac.42.9.2206

Van Boeckel, T. P., Brower, C., Gilbert, M., Grenfell, B. T., Levin, S. A., Robinson, T. P., . . . Laxminarayan, R. (2015). Global trends in antimicrobial use in Food Animals. Proceedings of the National Academy of Sciences, 112(18), 5649-5654. doi:10.1073/pnas.1503141112

VanderWaal, K., & Deen, J. (2018). Global trends in infectious diseases of swine. Proceedings of the National Academy of Sciences, 115(45), 11495-11500. doi:10.1073/pnas.1806068115

Virnekas, B., Ge, L., Plu€kthun, A., Schneider, K., Wellnhofer, G., & Moroney, S. E. (1994). Trinucleotide phosphoramidites: Ideal reagents for the synthesis of mixed oligonucleotides for random mutagenesis. Nucleic Acids Research, 22(25), 5600-5607. doi:10.1093/nar/22.25.5600

Vishnepolsky, B., Gabrielian, A., Rosenthal, A., Hurt, D. E., Tartakovsky, M., Managadze, G., . . . Pirtskhalava, M. (2018). Predictive model of linear antimicrobial peptides active against gram-negative bacteria. Journal of Chemical Information and Modeling, 58(5), 1141-1151. doi:10.1021/acs.jcim.8b00118

Vötsch, D., Willenborg, M., Weldearegay, Y. B., & Valentin-Weigand, P. (2018). Streptococcus suis – the “two faces” of a pathobiont in the porcine respiratory tract. Frontiers in Microbiology, 9. doi:10.3389/fmicb.2018.00480

W. (n.d.). Climate change and nutrition: A case for acting now (pp. 1-2) (S. Bashir, Ed.). World Food Programme (WFP).

Webber, M. A. (2002). The importance of efflux pumps in bacterial antibiotic resistance. Journal of Antimicrobial Chemotherapy, 51(1), 9-11. doi:10.1093/jac/dkg050

Wiegand, I., Hilpert, K., & Hancock, R. E. (2008). Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. Nature Protocols, 3(2), 163–175. doi:10.1038/nprot.2007.521

Weidenmaier, C., & Peschel, A. (2008). Teichoic acids and related cell-wall glycopolymers in gram-positive physiology and host interactions. Nature Reviews Microbiology, 6(4), 276-287. doi:10.1038/nrmicro1861

Wilson, D. R., & Finlay, B. B. (1998). Phage display: Applications, innovations, and issues in phage and host biology. Canadian Journal of Microbiology, 44(4), 313-329. doi:10.1139/w98-015

Wittum, T., & Dewey, C. (1996). Partial budget analysis of sow Escherichia coli vaccination. Partial Budget Analysis of Sow Escherichia Coli Vaccination, 4(1), 9-13.

Woolhouse, M., Ward, M., Van Bunnik, B., & Farrar, J. (2015). Antimicrobial resistance in humans, livestock and the wider environment. Philosophical Transactions of the Royal Society B: Biological Sciences, 370(1670), 20140083. doi:10.1098/rstb.2014.0083

Xiong, Y., Yang, Z., Zhang, J., Li, J., Chen, P., & Xiang, Y. (2019). Panning using a phage-displayed random peptide library to identify peptides that antagonize the Helicobacter pylori arss acid-sensing domain. Microbial Pathogenesis, 135, 103614. doi:10.1016/j.micpath.2019.103614

Yagodkin, A., Azhayev, A., Roivainen, J., Antopolsky, M., Kayushin, A., Korosteleva, M., . . . Mackie, H. (2007). Improved synthesis of trinucleotide phosphoramidites and generation of randomized oligonucleotide libraries. Nucleosides, Nucleotides and Nucleic Acids, 26(5), 473-497. doi:10.1080/15257770701426260

Yang, M., & Zhang, C. (2022). Important role of acute care surgery during pandemic time. World Journal of Gastrointestinal Surgery, 14(6), 626-628. doi:10.4240/wjgs.v14.i6.626

Yang, S., Deng, Q., Sun, L., Zhu, Y., Dong, K., Wu, S., . . . Li, Y. (2021). Salmonella effector SpvB inhibits NF-KB activity via KEAP1-mediated downregulation of ikkβ. Frontiers in Cellular and Infection Microbiology, 11. doi:10.3389/fcimb.2021.641412

Yu, Z., Zhu, Y., Qin, W., Yin, J., & Qiu, J. (2017). Oxidative stress induced by polymyxin E is involved in rapid killing of paenibacillus polymyxa. BioMed Research International, 2017, 1-12. doi:10.1155/2017/5437139

Yuan, H., Perry, C. N., Huang, C., Iwai-Kanai, E., Carreira, R. S., Glembotski, C. C., & Gottlieb, R. A. (2009). LPS-induced autophagy is mediated by oxidative signaling in cardiomyocytes and is associated with cytoprotection. American Journal of Physiology-Heart and Circulatory Physiology, 296(2). doi:10.1152/ajpheart.01051.2008

Yumoto, H., Hirota, K., Hirao, K., Ninomiya, M., Murakami, K., Fujii, H., & Miyake, Y. (2019). The pathogenic factors from oral streptococci for systemic diseases. International Journal of Molecular Sciences, 20(18), 4571. doi:10.3390/ijms20184571

Zavascki, A. P., Goldani, L. Z., Li, J., & Nation, R. L. (2007). Polymyxin B for the treatment of multidrug-resistant pathogens: A critical review. Journal of Antimicrobial Chemotherapy, 60(6), 1206–1215. doi:10.1093/jac/dkm357

Zeineldin, M., Aldridge, B., & Lowe, J. (2019). Antimicrobial effects on swine gastrointestinal microbiota and their accompanying antibiotic resistome. Frontiers in Microbiology, 10. doi:10.3389/fmicb.2019.01035

Zeya, H. I., & Spitznagel, J. K. (1963). Antibacterial and enzymic basic proteins from leukocyte lysosomes: Separation and identification. Science, 142(3595), 1085-1087. doi:10.1126/science.142.3595.1085

Zhang, C., & Yang, M. (2022). Antimicrobial peptides: From design to clinical application. Antibiotics, 11(3), 349. doi:10.3390/antibiotics11030349

Zhang, H., Torkamani, A., Jones, T. M., Ruiz, D. I., Pons, J., & Lerner, R. A. (2011). Phenotype-information-phenotype cycle for deconvolution of combinatorial antibody libraries selected against Complex Systems. Proceedings of the National Academy of Sciences, 108(33), 13456–13461. doi:10.1073/pnas.1111218108

Zimmerman, J. J. (2017). Diseases of swine / edited by Jeffrey J. Zimmerman... et al.. (Vol. 10). Chichester: Wiley-Blackwell. doi:ISBN: 9780813822679

Chapter 9: Appendix

9.1 ADP3 database "PigDataset"

Designated peptide No.	APD3 Reference	Peptide Sequence	Length (mers)	Overall net charge	Activity against
1	AP00002	YVPLPNVPQPGRRPFPTFPGQGPFNPKIKWPQGY	34	4	*E. coli*
2	AP00006	GNNRPVYIPQPRPPHPRI	18	4	*E. coli, S. Typhimurium*
3	AP00120	GLLDTIKGVAKTVAASMLDKLKCKISGC	28	3	*E. coli*
4	AP00205	ITSISLCTPGCKTGALMGCNMKTATCHCSIHVSK	34	3	*C. jejuni, E. coli, S. Typhimurium, Clostridium spp*
5	AP00341	FIGLLISAGKAIHDLIRRRH	20	4	*E.coli*
6	AP00454	FLPAIVGAAGKFLPKIFCAISKKC	24	4	*E. coli*
7	AP00525	ILGPVLSMVGSALGGLIKKI	20	3	*E. coli*
8	AP00832	ILGPVISTIGGVLGGLLKNL	20	2	*E. coli*
9	AP01016	GFFGKMKEYFKKFGASFKRRFANLKKRL	28	10	*E. coli*
10	AP01246	FLPKTLRKFFCRIRGGRCAVLNCLGKEEQIGRCSNSGRKCCRKKK	45	12	*E.coli, S. Typhimurium*
11	AP01634	INWKKIFEKVKNLV	14	4	*E. coli*
12	AP01781	GFMDTAKNVAKNVAATLLDKLKCKITGGC	29	3	*E. coli*
13	AP02981	FKFGSFIKRMWRSKLAKKLRAKGKELLRDYANRVLSPEEEAAAPAPYPA	49	7	*E. coli, Salmonella spp*
14	AP02615	GIFSKFAGKGLKNLFMKGAKTIGKEVGMDVLRTGIDIAGCKIKGEC	46	5	*E.coli*
15	AP01591	KCWNLRGSCREKCIKNEKLYIFCTSGKLCCLKPKFQPNMLQR	42	8	*E. coli*
16	AP01248	INMKASAAVAKKLL	14	3	*E. coli*
17	AP01778	DSMGAVKLAKLLIDKMKCEVTKAC	24	2	*E. coli*
18	AP01903	GFGSFLGKALKAALKIGANVLGGAPEQ	27	2	*E. coli*
19	AP01977	FLFSLIPSAISGLISAFK	18	2	*E.coli*
20	AP02048	DHYLCVKNEGICLYSSCPSYTKIEGTCYGGKAKCCK	36	2	*E.coli*
21	AP02218	IFKAIWSGIKSLF	13	2	*E. coli*
22	AP02456	KKCKFFCKVKKKIKSIGFQIPIVSIPFK	28	9	*E. coli*
23	AP02700	KKVWFIFHVCPKLKQRIL	18	5	*E. coli*
24	AP02806	VVYALKRNGRTLYGF	15	3	*E. coli*
25	AP02909	RRRRRFRRVIRRIRLPKYLTINTE	24	10	*E. coli*
26	AP00281	GLLRKGGEKIGEKLKKIGQKIKNFQKLVPQPEQ	34	6	*E. coli, S. Typhimurium*
27	AP00600	GLLRASSVWGRKYYVDLAGCAKA	23	3	*E. coli*
28	AP00964	GLWSKIKEAAKAAGKAALNAVTGLVNQGDQPS	32	2	*E. coli*
29	AP02107	FIGPVLKMATSILPTAICKGFKKC	24	4	*E. coli*
30	AP02268	GLLDSVKEGLKKVAGQLLDTLKCKISGCTPA	31	2	*E. coli*
31	AP02368	FLPIIGKLLSGLL	13	1	*E. coli*
32	AP02555	IWLTALKFLGKNLGKHLAKQQLAKL	25	6	*E. coli*
33	AP02633	KGFKKIEKLGRHIRDGVVKAGPAVGVVGQASSIMG	35	5	*E. coli, Salmonella spp*
34	AP00969	GLWSKIKEAAKTAGLMAMGFVNDMV	25	1	*E. coli*
35	AP01434	FFGSVLKLIPKIL	13	3	*E. coli*
36	AP01925	GLLDAIKDTAQNLFANVLDKIKCKFTKC	28	2	*E. coli*
37	AP02127	IWSAIWSGIKGLL	13	2	*E. coli*
38	AP02266	GFWDSVKEGLKNAAVTILNKIKCKISECPPA	31	2	*E. coli*
39	AP02354	KINNPVSCLRKGGRCWNRCIGNTRQIGSCGVPFLKCCKRK	40	10	*E. coli*
40	AP02508	HHHHRFGKIGHELHKGVKKVEKVTHDVNKVTSGVKKVASSIEKAKNV	47	7	*E. coli*
41	AP00121	GLMDTVKNVAKNLAGHMLDKLKCKITGC	28	3	*E. coli*
42	AP00163	ALWKDILKNVGKAAGKAVLNTVTDMVNQ	28	3	*E. coli*
43	AP00570	SIITMTKEAKLPQLWKQIACRLYNTC	26	3	*E. coli*
44	AP00781	FLGALIKGAIHGGRFIHGMIQNHH	24	3	*E. coli, S. Typhimurium*
45	AP02325	GIPCGESCVFIPCTVTALLGCSCKDKVCYKN	31	1	*E. coli*
46	AP02445	QSHLSLCRYCCNCCRNKGCGYCCKF	25	4	*E. coli*
47	AP02565	FLPLLAGLAANFLPKIFCKITRK	23	4	*E. coli*
48	AP00430	ILGKIWEGIKSLF	13	2	*S. Typhimurium, E. coli*
49	AP00493	NLVSGLIEARKYLEQLHRKLKNCKV	25	4	*E. coli*
50	AP02321	TNYGNGVGVPDAIMAGIIKLIFIFNIRQGYNFGKKAT	37	3	*C. jejuni, S. Typhimurium, E. coli, C. perfringens. Salmonella spp*
51	AP02929	LPRRNRWSKIWKKVVTVFS	19	7	*C. jejuni, S. Typhimurium, E. coli*
52	AP02930	HLRRINKLLTRIGLYRHAFG	20	6	*C. jejuni, S. Typhimurium, E. coli*
53	AP02931	NRFTARFRRTPWRLCLQFRQ	20	7	*C. jejuni, S. Typhimurium, E. coli*
54	AP01003	FKSWSFCTPGCAKTGSFNSYCC	22	2	*C. jejuni, S. Typhimurium, Clostridium spp*
55	AP02249	FISQIISTARI	11	1	*C. difficle, E. coli*
56	AP02956	AGCICSGSVAVANSHNAGPAYCVGYCGNNGAVTRNANANLARTA	44	2	*C. perfringens*
57	AP00516	IWLTALKFLGKHAAKHLAKQQLSKL	25	6	*E. coli*
58	AP00517	KIKWFKTMKSIAKFIAKEQMKKHLGGE	27	6	*E. coli*
59	AP00895	KRFKKFFKKLKNSVKKRAKKFFKKPRVIGVSIPF	34	15	*E.coli*
60	AP01457	GLKDIFKAGLGSLVKGIAAHVAN	23	3	*E. coli, Salmonella spp*
61	AP00691	GFFKKAWRKVKHAGRRVLDTAKGVGRHYVNNWLNRYR	37	10	*E.coli, C. perfringens*
62	AP02638	GLPWILLRWLFFRG	14	2	*S. Typhimurium, E. coli*
63	AP00007	GNNRPVYIPQPRPPHPRL	18	4	*S. Typhimurium, E. coli*
64	AP00009	RFRPPIRRPPIRPPFYPPFRPPIRPPIFPPIRPPFRPPLGPFP	43	9	*S. Typhimurium, E. coli*
65	AP00344	GNNRPIYIPQPRPPHPRL	18	3	*S. Typhimurium, E. coli*
66	AP00362	VDKPDYRPRPRPPNM	15	2	*S. Typhimurium, E. coli*
67	AP02237	FFRLLFHGVHHGGGYLNAA	19	2	*S. Typhimurium, E. coli*
68	AP02238	GWKKWFTKGERLSQRHFA	18	4	*S. Typhimurium, E. coli*
69	AP02239	GFLGILFHGVHHGRKKALHMNSERRS	26	4	*S. Typhimurium, E. coli*
70	AP00041	QGVRNHVTCRIYGGFCVPIRCPGRTRQIGTCFGRPVKCCRRW	42	10	*E.coli, Salmonella spp*
71	AP00368	GLFRRLRDSIRRGQQKILEKARRIGERIKDIFRG	34	8	*E.coli, S. Typhimurium, Salmonella spp*

			Length (mers)	Overall net charge	Activity against
72	AP00678	RLKELITTGGQKIGEKIRRIGQRIKDFFKNLQPREEKS	38	6	*E. coli, S. Typhimurium*
73	AP01325	QPFIPRPIDTCRLRNGICFPGICRRPYYWIGTCNNGIGSCCARGWRS	47	6	*E. coli*
74	AP02193	YSKSLPLSVLNP	12	1	*E.coli*
75	AP02677	FWQKMSFA	8	1	*E.coli, S. Typhimurium*
76	AP00001	GLWSKIKEVGKEAAKAAAKAAGKAALGAVSEAV	33	4	*E. coli*
77	AP00037	VRNHVTCRINRGFCVPIRCPGRTRQIGTCFGPRIKCCRSW	40	10	*E. coli*
78	AP00038	QGVRNHVTCRINRGFCVPIRCPGRTRQIGTCFGPRIKCCRSW	42	10	*E. coli*
79	AP00054	GIGGALLSAGKSALKGLAKGLAEHFAN	27	3	*E. coli*
80	AP00055	IIGPVLGMVGSALGGLLKKI	21	2	*E. coli*
81	AP00056	LIGPVLGLVGSALGGLLKKI	21	3	*E. coli*
82	AP00058	GIGTKILGGVKTALKGALKELASTYAN	27	4	*E. coli*
83	AP00059	GIGTKILGGVKTALKGALKELASTYVN	27	4	*E. coli*
84	AP00073	FLPLLAGLAANFLPKIFCKITRKC	24	4	*E. coli*
85	AP02685	GLPICGETCFKTKCYTKGCSCSYPVCKRN	29	4	*E. coli*
86	AP02362	NLKAIAALAKKLL	13	3	*E. coli*
87	AP02549	GVWDWLKKTAKNVWNSDIVKQLKGKAINAAKNYVAEKIGATPS	43	6	*E. coli*
88	AP01706	GFMDTAKNVAKNVAVTLLDKLKCKITGGC	29	3	*E. coli*
89	AP01458	GLKEIFKAGLGSLVKGIAAHVAN	23	3	*E. coli*
90	AP00580	GLFGKILGVGKKVLCGLSGMC	21	3	*E. coli*
91	AP00473	FFHHIFRGIVHVGKTIHRLVTG	22	3	*E. coli, S. Typhimurium*
92	AP00172	GKPRPYSPRPTSHPRPIRV	19	5	*E. coli*
93	AP00210	GMASKAGAIAGKIAKVALKAL	21	5	*E. coli*
94	AP00466	FLPAIAGVAAKFLPKIFCAISKKC	24	4	*E. coli*
95	AP01496	IPPFIKKVLTTVF	13	3	*E. coli*
96	AP00894	GLLDFVTGVGKDIFAQLIKQI	21	1	*E. coli*
97	AP01440	FFPIIAGMAAKVICAITKKC	20	3	*E. coli*
98	AP01640	INWLKLGKMVIDAL	14	1	*E. coli*
99	AP01695	GFGSLLGKALRLGANVL	17	3	*E. coli*
100	AP02013	FIGKLISAASGLLSHL	16	2	*E. coli*
101	AP00357	FFPIGVFCKIFKTC	14	2	*E. coli*
102	AP00431	TWLKKRRWKKAKPP	14	7	*E. coli*
103	AP00455	FFPIVAGVAGQVLKKIYCTISKKC	24	4	*E. coli*
104	AP01455	FFPLALLCKVFKKC	14	3	*E. coli*
105	AP00663	GFSSIFRGVAKFASKGLGKDLARLGVNLVACKISKQC	37	6	*E. coli*
106	AP00830	GLLLDTLKGAAKDIAGIALEKLKCKITGCKP	31	3	*E. coli*
107	AP00964	GLWSKIKEAAKAAGKAALNAVTGLVNQGDQPS	32	2	*E. coli*
108	AP01400	RPKHPIKHQGLPQEVLNENLLRF	23	2	*E. coli*
109	AP01457	GLKDIFKAGLGSLVKGIAAHVAN	23	3	*E. coli*
110	AP01476	ACDTATCVTHRLAGLLSRSGGVVKNNFVPTNVGSKAF	37	4	*E. coli*
111	AP01753	GIWSSIKNLASKAWNSDIGQSLRNKAAGAINKFVADKIGVTPSQAAS	47	4	*E. coli*
112	AP01925	GLLDAIKDTAQNLFANVLDKIKCKFTKC	28	2	*E. coli*
113	AP01927	GLFNVFKKVGKNVLKNVAGSLMDNLKCKVSGEC	33	4	*E. coli*
114	AP02001	GMATKAGTALGKVAKAVIGAAL	22	4	*E. coli*
115	AP02003	GFWTTAAEGLKKFAKAGLASILNPK	25	4	*E. coli*
116	AP02125	FWGKLWEGVKNAI	13	2	*E. coli*
117	AP02224	GLLLDTVKGAAKNVAGILLNKLKCKMTGDC	30	3	*E. coli*
118	AP02328	GSVIKCGESCLLGKCYTPGCTCSRPICKKD	30	3	*E. coli*
119	AP02449	GVSKILHSAGKFGKAFLGEIMKS	23	3	*E. coli*
120	AP02483	GIGKFLHSAGKFGKAFLGEVMKS	23	3	*E. coli*
121	AP02490	GFWSSALEGLKKFAKGGLEALTNPK	25	2	*E. coli*
122	AP02491	GLASTIGSLLGKFAKGGAQAFLQPK	25	3	*E. coli*
123	AP02521	MTPLWRVMGNKPFGAYCQDHVECSTGICKGGHCITSQPIKS	41	2	*E. coli*
124	AP02599	GVFDIIKGAGKQLIAHAMGKIAEKVGLNKDGN	32	2	*E. coli*
125	AP02697	GTFIKQQRKQKQQRHHTSGTRKRMAK	26	9	*E. coli*
126	AP02698	VGRKHSILNCIPYLKKKKIMRL	22	7	*E. coli*
127	AP02717	AVNIPFKVHLRCKAAFC	17	3	*E. coli*
128	AP02765	NGNLLGGLLRPVLGVVKGLTGGLGKK	26	4	*E. coli*
129	AP02813	ILELAGNAARDNKKTRIIPRHLQL	24	3	*E. coli*
130	AP02888	VHISHREARGPSFRICVGFLGPRWARGCSTGN	32	4	*E. coli*
131	AP02889	GDVPPGIRNTICRMQQGICRLFFCHSGTGQQHRQRCG	37	4	*E. coli*
132	AP02962	RRIRFRPPYLPRPGRRPRFPPPFPIPRIPRIP	32	10	*E. coli*
133	AP02976	NLVSALIEGRKYLKNVLKKLNRLKEKNKAKNSKENN	36	8	*E. coli*
134	AP00134	SWLSKTAKKLENSAKKRISEGIAIAIQGGPR	31	5	*S. Typhimurium, E. coli*
135	AP00150	ILPWKWPWWPWRR	13	4	*E. coli, S. Typhimurium*
136	AP00166	GWGSFFKKAAHVGKHVGKAALTHYL	25	4	*E. coli, S. Typhimurium*
137	AP00176	ACYCRIPACIAGERRYGTCIYQGRLWAFCC	30	3	*S. Typhimurium, E. coli*
138	AP00200	LKLKSIVSWAKKVL	14	5	*E. coli, S. Typhimurium, Salmonella spp*
139	AP00276	VFQFLGKIIHHVGNFVHGFSHVF	23	5	*S. Typhimurium, E. coli*
140	AP02262	FLGGLLASLLGKI	13	1	*E. coli*
141	AP00307	AGRGKQGGKVRAKAKTRSSRAGLQFPVGRVHRLLRKGNY	39	12	*S. Typhimurium, E.coli*
142	AP00366	GRFKRFRKKFKKLFKKLSPVIPLLHLG	27	10	*S. Typhimurium, E.coli*
143	AP00367	GGLRSLGRKILRAWKKYGPIIVPIIRIG	27	7	*S. Typhimurium, E.coli*
144	AP00369	RIIDLLWRVRRPQKPKFVTVWVR	23	6	*S. Typhimurium, E.coli*
145	AP00396	RRRPRPPYLPRPRPPPFFPPRLPPRIPPGFPPRFPPRFP	39	11	*E. coli, S. Typhimurium*
146	AP00507	GLLSKVLGVGKKVLCGVSGLC	21	3	*S. typhimurium*
147	AP00508	GLLDSIKGMAISAGKGALQNLLKVASCKLDKTC	33	3	*S. Typhimurium*
148	AP00771	GIGKFLHSAGKFGKAFVGEIMKS	23	3	*S. typhimurium, E.coli*
149	AP00777	GKGRWLERIGKAGGIIIGGALDHL	24	3	*S. Typhimurium*
150	AP00778	WLRRIGKGVKIIGGAALDHL	20	4	*S. typhimurium, E.coli*

Appendix Table 1 continued: The antimicrobial peptides identified from ADP3 with activity against several bacteria pathogens of pigs "ADP3_ AMPDatabase PigPathogens".

Designated peptide No.	APD3 Reference	Peptide Sequence	AMP characteristic and activity.		
			Length (mers)	Overall net charge	Activity against
151	AP00779	GRRKRKWLRRIGKGVKIIGGAALDHL	26	9	*S. typhimurium, E.coli*
152	AP00782	GWGSIFKHGRHAAKHIGHAAVNHYL	25	4	*E. coli*
153	AP00783	RWGKWFKKATHVGKHVGKAALTAYL	25	7	*E. coli, S. Typhimurium*
154	AP00784	FFRLLFHGVHHVGKIKPRA	19	5	*S. Typhimurium, E. coli*
155	AP00785	GWKSVFRKAKKVGKTVGGLALDHYL	25	6	*S. Typhimurium, E. coli*
156	AP00788	AGWGSIFKHIFKAGKFIHGAIQAHND	26	3	*E. coli, S. Typhimurium*
157	AP00789	GFWGKLFKLGLHGIGLLHLHL	21	3	*S. Typhimurium, E.coli*
158	AP00790	GWKKWLRKGAKHLGQAAIK	19	7	*E. coli, S. Typhimurium*
159	AP00791	GWKKWLRKGAKHLGQAAIKGLAS	23	7	*E. coli, S. Typhimurium*
160	AP00792	FLGLLFHGVHHVGKWIHGLIHGHH	24	2	*E. coli, S. Typhimurium*
161	AP01545	FFGHLFKLATKIIPSLFQ	18	2	*E. coli, S. Typhimurium*
162	AP01578	GIHDILKYGKPS	12	2	*S. Typhimurium, E.coli*
163	AP01714	KVFLGLK	7	2	*S. Typhimurium*
164	AP02049	FFHHIFRGIVHVGKTIHKLVTGT	23	3	*S. typhimurium, Salmonella spp, E.coli*
165	AP02178	LRVRRTLQCSCRRVCRNTCSCIRLSRSTYAS	31	8	*S. typhimurium, E.coli*
166	AP02228	RKGWFKAMKSIAKFIAKEKLKEHL	24	7	*S. Typhimurium*
167	AP02621	GKLTKDKLKRGAKKALNVASKVAPIVAAGASIAR	34	9	*S. Typhimurium, E.coli*
168	AP02629	KRFWQLVPLAIKIYRAWKRR	20	7	*S. Typhimurium, E.coli*
169	AP02642	AIPWIWIWWLLRKG	14	2	*S. Typhimurium, E.coli*
170	AP02640	RIRFPWPWRWPWWRRVRG	18	6	*S. Typhimurium, E.coli*
171	AP02643	AIPWSIWWRLLFKG	14	2	*S. Typhimurium*
172	AP02767	APKGVQGPNG	10	1	*E. coli, S. Typhimurium*
173	AP02926	ASVVNKLTGGVAGLLK	16	2	*E. coli, S. Typhimurium*
174	AP02854	GYFPGRPPFPRPFPRPPSRPFPRPPFPGPFPRPYPWR	37	7	*S. Typhimurium, E.coli*
175	AP02890	INLKAIAALAKKLF	14	3	*E. coli, S. Typhimurium, Salmonella spp*
176	AP02501	GWLKKIGKKIERVGQHTRDASIQAIGIAQQAANVAATARG	40	5	*E. coli*
177	AP00074	FLPVLAGIAAKVVPALFCKITKKC	24	4	*E. coli*
178	AP00075	GLLDSLKGFAATAGKGVLQSLLSTASCKLAKTC	33	3	*E. coli*
179	AP00080	GIFSKLGRKKIKNLLISGLKNVGKEVGMDVVRTGIDIAGCKIKGEC	46	6	*E.coli*
180	AP00083	GILSLVKGVAKLAGKGLAKEGGKFGLELIACKIAKQC	37	5	*E. coli*
181	AP00085	SLFSLIKAGAKFLGKNLLKQGACYAACKASKQC	33	6	*E. coli*
182	AP00146	GIGAVLKVLTTGLPALISWIKRKRQQ	26	6	*E. coli*
183	AP00470	FLPIIASVAAKVFSKIFCAISKKC	24	4	*E.coli*
184	AP00498	GLVRKGGEKFGEKLRKIGQKIKEFFQKLALEIEQ	34	4	*E.coli*
185	AP00500	GLGSVLGKALKIGANLL	17	3	*E.coli*
186	AP00504	LAHQKPFIRKSYKCLHKRCR	20	7	*E. coli*
187	AP00521	ILGTILGLLKGL	12	2	*E.coli*

* "Activity" is defined from MIC/MBC values from antimicrobial assays often derived from literature article which is the source of the AMP referenced in ADP3.

9.2 Next Generation Sequencing Barcode Primers

Provided below is the complete list of NGS barcode primers used for sequencing library and panning samples.

Table 2.13 extended
PCR Round 2 "Barcoding & Adapter" NGS primers for pSD3 _16mer library

Primer	Oligonucleotide primer sequence [1] 5' → 3'	(bp)	Features
P1prim-linker2	**CCTCTCTATGGGCAGTCGGTGATCTAG AACATTTCACTTAC**	46	T_m = 68 °C GC%= 46%
Akey-BC1-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTAAGGTAACCGTAATCCTTGTGGTATCG		
Akey-BC2-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTAAGGAGAACCGTAATCCTTGTGGTATCG		
Akey-BC3-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGAAGAGGATTCCGTAATCCTTGTGGTATCG		
Akey-BC4-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTACCAAGATCCGTAATCCTTGTGGTATCG		
Akey-BC5-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCAGAAGGAACCGTAATCCTTGTGGTATCG		
Akey-BC6-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGCAAGTTCCGTAATCCTTGTGGTATCG		

Akey-BC7-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCGTGATTCCGTAATCCTTGTGGTATCG
Akey-BC8-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCCGATAACCGTAATCCTTGTGGTATCG
Akey-BC9-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGAGCGGAACCGTAATCCTTGTGGTATCG
Akey-BC10-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGACCGAACCGTAATCCTTGTGGTATCG
Akey-BC11-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCTCGAATCCGTAATCCTTGTGGTATCG
Akey-BC12-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTAGGTGGTTCCGTAATCCTTGTGGTATCG
Akey-BC13-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTAACGGACCGTAATCCTTGTGGTATCG
Akey-BC14-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTGGAGTGTCCGTAATCCTTGTGGTATCG
Akey-BC15-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTAGAGGTCCGTAATCCTTGTGGTATCG
Akey-BC16-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTGGATGACCGTAATCCTTGTGGTATCG
Akey-BC17-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTATTCGTCCGTAATCCTTGTGGTATCG
Akey-BC18-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGGCAATTGCCGTAATCCTTGTGGTATCG
Akey-BC19-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTAGTCGGACCGTAATCCTTGTGGTATCG
Akey-BC20-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCAGATCCATCCGTAATCCTTGTGGTATCG
Akey-BC21-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCGCAATTACCGTAATCCTTGTGGTATCG
Akey-BC22-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCGAGACGCCGTAATCCTTGTGGTATCG
Akey-BC23-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGCCACGAACCGTAATCCTTGTGGTATCG
Akey-BC24-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGAACCTCATTCCGTAATCCTTGTGGTATCG
Akey-BC25-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCTGAGATACCGTAATCCTTGTGGTATCG
Akey-BC26-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTACAACCTCCGTAATCCTTGTGGTATCG
Akey-BC27-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGAACCATCCGCCGTAATCCTTGTGGTATCG
Akey-BC28-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGATCCGGAATCCGTAATCCTTGTGGTATCG
Akey-BC29-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCGACCACTCCGTAATCCTTGTGGTATCG
Akey-BC30-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGAGGTTATCCGTAATCCTTGTGGTATCG

[1]All primers were manufactured by Sigma-Aldrich, the barcoding primers were at least 58bp in length. "XXX" relates to the A-key sequence, whereas "XXX" is the unique barcode and "XXX" region of the linker from the reverse primer which the barcode primer anneals to in the Round 1 PCR products.

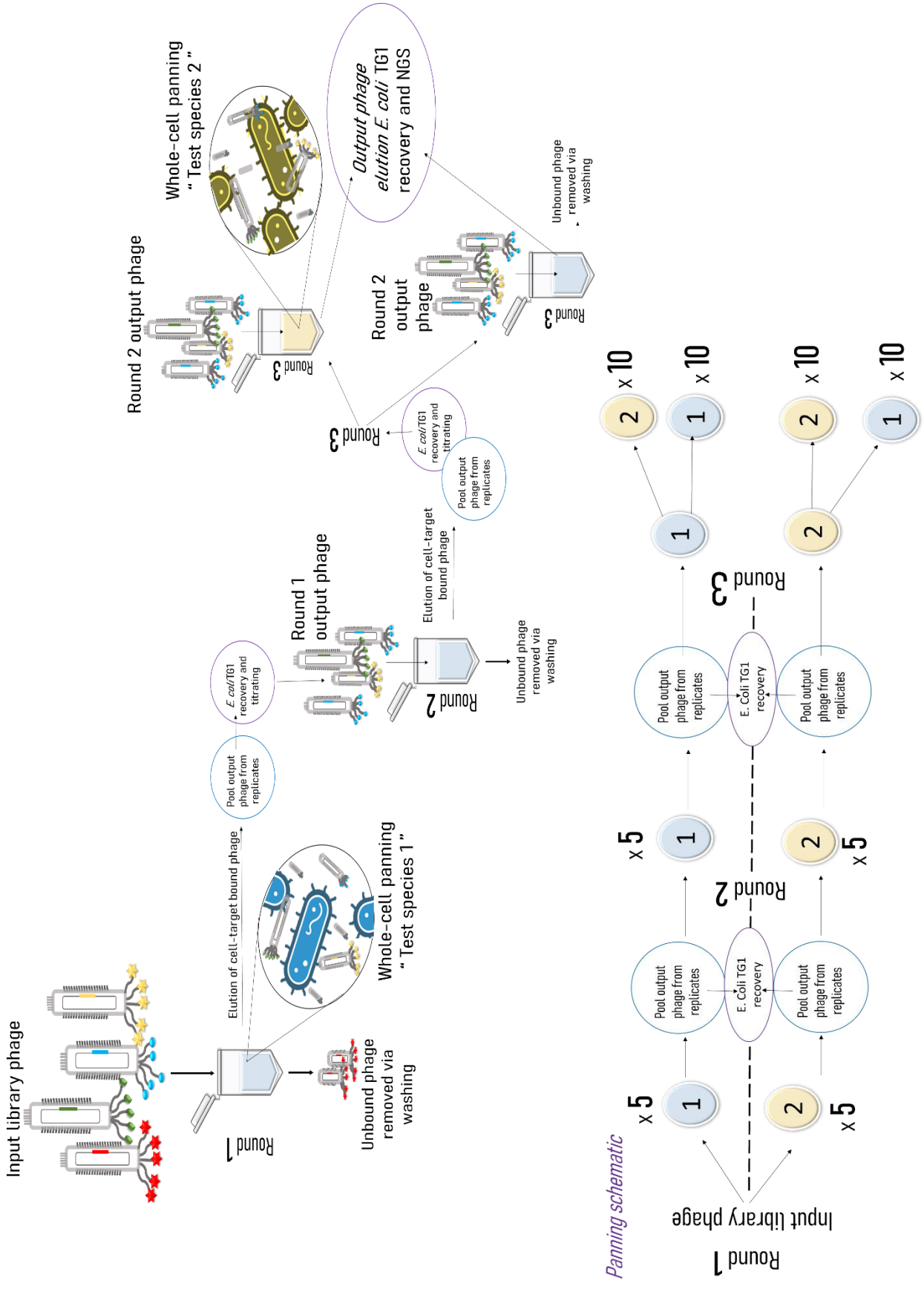
Table 2.13 PCR Round 2 "Barcoding & Adapter" NGS primers for pSD3 _16mer library

Primer	Oligonucleotide primer sequence [1] 5' → 3'	(bp)	Features
Akey-BC31-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCAAGCTGCCGTAATCCTTGTGGTATCG		
Akey-BC32-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTTACACACCGTAATCCTTGTGGTATCG		
Akey-BC33-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCTCATTGAACCGTAATCCTTGTGGTATCG		
Akey-BC34-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCGCATCGTTCCGTAATCCTTGTGGTATCG		
Akey-BC35-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTAAGCCATTGTCCGTAATCCTTGTGGTATCG		

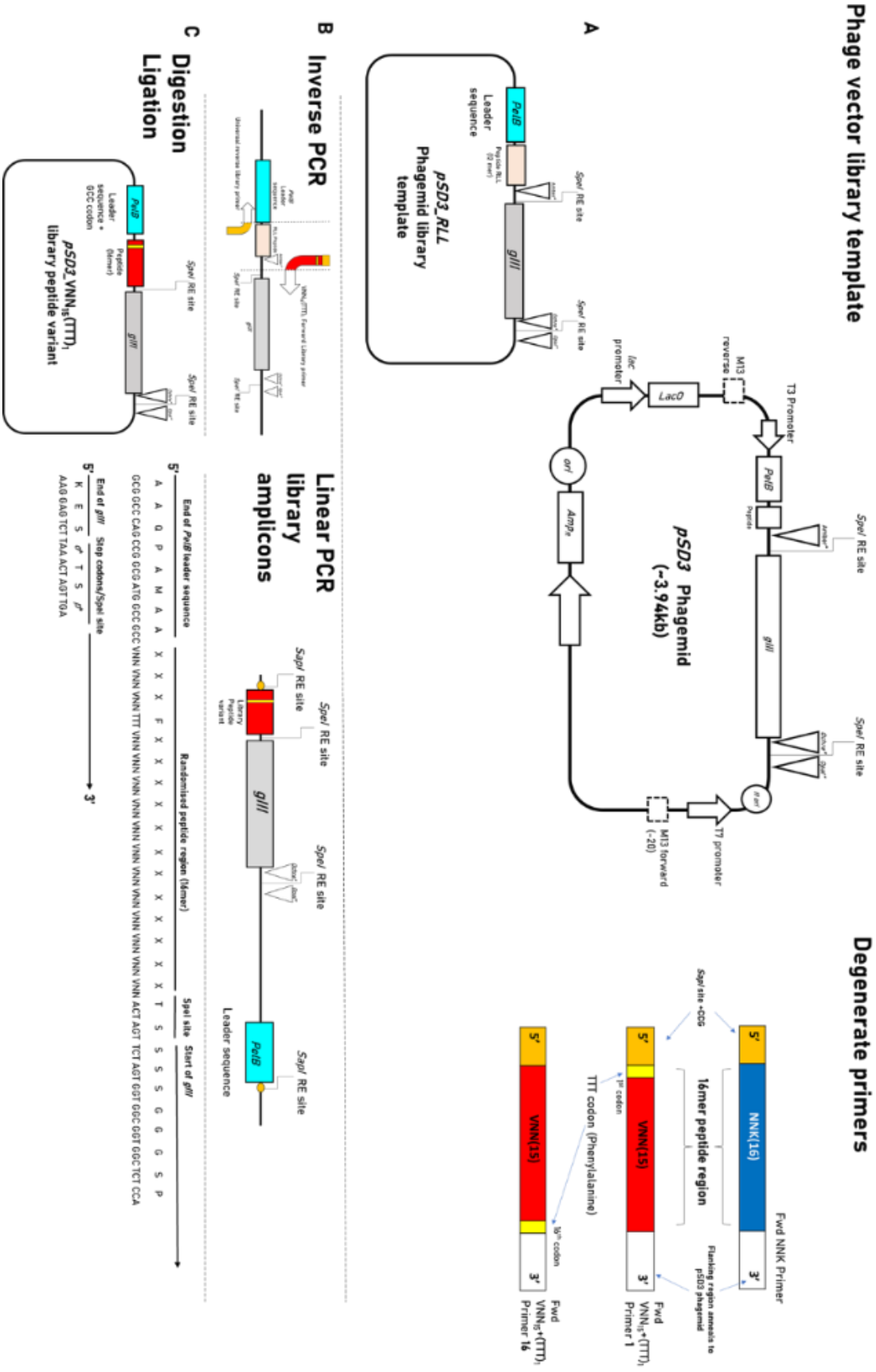
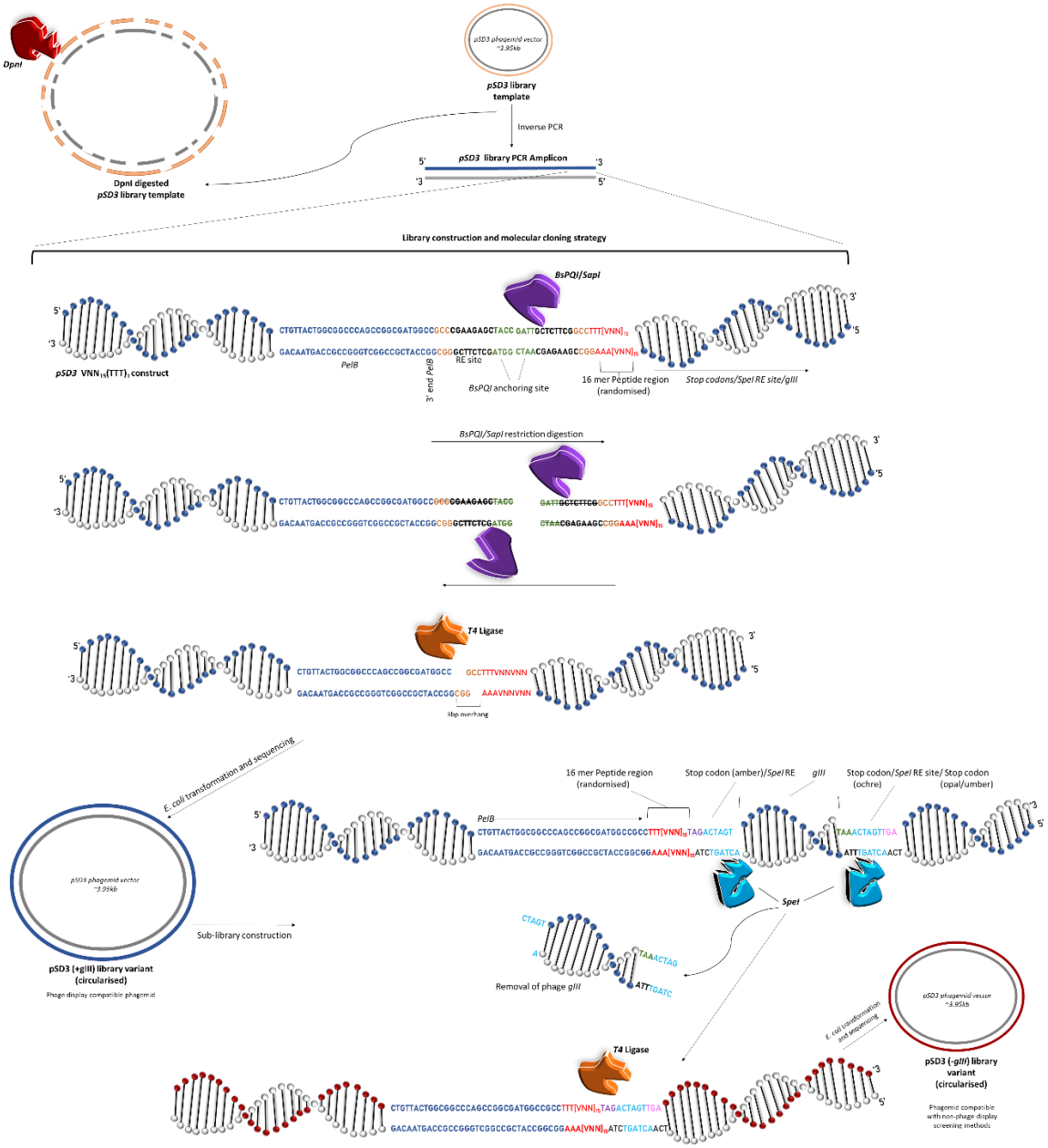
Akey-BC36-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGAAGGAATCGTCCGTAATCCTTGTGGTATCG
Akey-BC37-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTTGAGAATGTCCGTAATCCTTGTGGTATCG
Akey-BC38-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGGAGGACGGACCGTAATCCTTGTGGTATCG
Akey-BC39-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTAACAATCGGCCGTAATCCTTGTGGTATCG
Akey-BC40-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGACATAATCCGTAATCCTTGTGGTATCG
Akey-BC41-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCCACTTCGCCGTAATCCTTGTGGTATCG
Akey-BC42-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGCACGAATCCGTAATCCTTGTGGTATCG
Akey-BC43-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTTGACACCGCCGTAATCCTTGTGGTATCG
Akey-BC44-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTGGAGGCCAGCCGTAATCCTTGTGGTATCG
Akey-BC45-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGGAGCTTCCTCCGTAATCCTTGTGGTATCG
Akey-BC46-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCAGTCCGAACCGTAATCCTTGTGGTATCG
Akey-BC47-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTAAGGCAACCACCGTAATCCTTGTGGTATCG
Akey-BC48-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCTAAGAGACCGTAATCCTTGTGGTATCG
Akey-BC49-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCTAACATAACCGTAATCCTTGTGGTATCG
Akey-BC50-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGGACAATGGCCGTAATCCTTGTGGTATCG
Akey-BC51-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTGAGCCTATTCCGTAATCCTTGTGGTATCG
Akey-BC52-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCGCATGGAACCGTAATCCTTGTGGTATCG
Akey-BC53-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGGCAATCCTCCGTAATCCTTGTGGTATCG
Akey-BC54-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCGGAGAATCGCCGTAATCCTTGTGGTATCG
Akey-BC55-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCACCTCCTCCGTAATCCTTGTGGTATCG
Akey-BC56-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCAGCATTAATTCCGTAATCCTTGTGGTATCG
Akey-BC57-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTGGCAACGGCCGTAATCCTTGTGGTATCG
Akey-BC58-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCTAGAACACCGTAATCCTTGTGGTATCG
Akey-BC59-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCTTGATGTTCCGTAATCCTTGTGGTATCG
Akey-BC60-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTAGCTCTTCCGTAATCCTTGTGGTATCG
Akey-BC61-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCACTCGGATCCGTAATCCTTGTGGTATCG

Table 2.13 PCR Round 2 “Barcoding & Adapter” NGS primers for pSD3 _16mer library			
Primer	**Oligonucleotide primer sequence**	**(bp)**	**Features**
Akey-BC62-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCCTGCTTCACCGTAATCCTTGTGGTATCG		
Akey-BC62-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCTTAGAGTTCCGTAATCCTTGTGGTATCG		
Akey-BC64-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGAGTTCCGACCGTAATCCTTGTGGTATCG		
Akey-BC65-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCTGGCACATCCGTAATCCTTGTGGTATCG		
Akey-BC66-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCGCAATCATCCGTAATCCTTGTGGTATCG		
Akey-BC67-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCCTACCAGTCCGTAATCCTTGTGGTATCG		
Akey-BC68-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCAAGAAGTTCCGTAATCCTTGTGGTATCG		
Akey-BC69-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCAATTGGCCGTAATCCTTGTGGTATCG		
Akey-BC70-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCTACTGGTCCGTAATCCTTGTGGTATCG		
Akey-BC71-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGAGGCTCCGACCGTAATCCTTGTGGTATCG		
Akey-BC72-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGAAGGCCACACCGTAATCCTTGTGGTATCG		
Akey-BC73-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTGCCTGTCCGTAATCCTTGTGGTATCG		
Akey-BC74-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGATCGGTTCCGTAATCCTTGTGGTATCG		
Akey-BC75-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCAGGAATACCGTAATCCTTGTGGTATCG		
Akey-BC76-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGGAAGAACCTCCGTAATCCTTGTGGTATCG		
Akey-BC77-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGAAGCGATTCCGTAATCCTTGTGGTATCG		
Akey-BC78-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCAGCCAATTCTCCGTAATCCTTGTGGTATCG		
Akey-BC79-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCTGGTTGTCCGTAATCCTTGTGGTATCG		
Akey-BC80-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCGAAGGCAGGCCGTAATCCTTGTGGTATCG		
Akey-BC81-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCTGCCATTCGCCGTAATCCTTGTGGTATCG		
Akey-BC82-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTGGCATCTCCGTAATCCTTGTGGTATCG		
Akey-BC83-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTAGGACATTCCGTAATCCTTGTGGTATCG		
Akey-BC84-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTTCCATAACCGTAATCCTTGTGGTATCG		
Akey-BC85-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCAGCCTCAACCGTAATCCTTGTGGTATCG		
Akey-BC86-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTTGGTTATTCCGTAATCCTTGTGGTATCG		
Akey-BC87-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTGGCTGGACCGTAATCCTTGTGGTATCG		
Akey-BC88-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCGAACACTTCCGTAATCCTTGTGGTATCG		
Akey-BC89-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCTGAATCTCCGTAATCCTTGTGGTATCG		
Akey-BC90-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTAACCACGGCCGTAATCCTTGTGGTATCG		
Akey-BC91-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGGAAGGATGCCGTAATCCTTGTGGTATCG		
Akey-BC92-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTAGGAACCGCCGTAATCCTTGTGGTATCG		
Akey-BC93-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTTGTCCAATCCGTAATCCTTGTGGTATCG		
Akey-BC94-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCGACAAGCCGTAATCCTTGTGGTATCG		
Akey-BC95-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGGACAGATCCGTAATCCTTGTGGTATCG		
Akey-BC96-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTAAGCGGTCCGTAATCCTTGTGGTATCG		

9.3 Panning strategy, enlarged schematic diagram



9.4 Whole-construct PCR cloning method, enlarged schematic



9.5 Molecular cloning strategy, restriction enzyme digestion; enlarged image