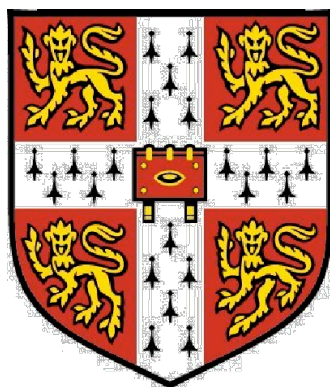


A Biosynthetic Approach to the Discovery of Novel Bioactive Peptides



University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Oliver Evan Wright



Emmanuel College

2011

DECLARATION

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text.

I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

This dissertation contains approximately 62,000 words and 105 figures, and thus does not exceed the prescribed limits of 65,000 words and 150 figures laid down by the Department of Engineering.

Oliver Evan Wright
December 2011

SUMMARY

Peptides represent a source of novel therapeutics for recalcitrant human diseases, but screening for bioactivity from natural or synthetic sources can be uneconomic. In contrast, *in vivo* expression of peptides from DNA libraries in a heterologous host such as *Escherichia coli* may combine production with screening. This dissertation aimed to use such an approach to discover novel bioactive peptides in a high-throughput and cost-effective manner, with a focus on antimicrobials and antiaggregants as proof-of-principle.

Antimicrobial peptides (AMPs) are innate defence effectors that may combat antibiotic-resistant pathogens. An inducible, autocleaving fusion tag was utilised to produce the model murine cathelicidin K2C18, along with a number of variants, which exhibited varying degrees of antimicrobial activity against a panel of microbes. Importantly, K2C18 also exhibited a bacteriostatic effect *in vivo* when secreted to the periplasm. This allowed for the implementation of an *in vivo* whole-cell screen for novel AMPs, using genomic DNA libraries as an input. One putative hit, the peptide S-H4, showed similar *in vivo* behaviour to K2C18 and was active when added exogenously to microbial cultures.

A second *in vivo* screen was constructed to search for inhibitors of A β 42 aggregation, a process implicated in Alzheimer's disease. The aggregation state of A β 42 was coupled to the fluorescence of a chromophore fusion partner, and used to screen co-expressed peptides from a random DNA library for putative antiaggregants. Additionally, the system incorporated an internal fluorescent reference to allow ratiometric comparison between samples. Several hits were identified and further validated using flow cytometry, with work ongoing to assess their activity *in vitro*.

Proof-of-principle of these two screens was achieved, indicating that such *in vivo* approaches to bioactive peptide discovery could lead to the development of new and useful therapeutics.

PUBLICATIONS

The following publications resulted from the work contained in this dissertation:

- Wright, O., Yoshimi, T. and Tunnacliffe, A. (2012). Recombinant production of cathelicidin-derived antimicrobial peptides in *Escherichia coli* using an inducible autocleaving enzyme tag. *New Biotechnol* **29** (3): 352-8, doi:10.1016/j.nbt.2011.11.001
- Wright, O., Yoshimi, T. and Tunnacliffe, A. (2011, February 7). *Recombinant production of cathelicidin-derived AMPs*. Poster presented at Proteins, peptides and peptidomimetics: applications in drug discovery and drug development, Uppsala, Sweden: <http://www.farmfak.uu.se/conference/>
- Wright, O. and Tunnacliffe, A. (2011, June 15). *High-throughput bacterial screen for inhibitors of A β 42 aggregation*. Poster presented at SB5.0: the fifth international meeting on synthetic biology, Palo Alto, USA: <http://sb5.biobricks.org/>

ACKNOWLEDGEMENTS

For all my whānau

A big thank-you to my supervisor, (now Prof) Alan Tunnacliffe, for allowing me to work in his laboratory. I've learnt an immense amount over the past four years, and would like to thank him for having such an open door when I had enquiries (banal or otherwise). Also, many thanks for having such an eye for detail with regards to editing: discerning a non-italicised number in an italicised reference is quite a feat! As an aside, I estimate I racked up ~£18,000 in consumables over the years, not to mention the cost of a sojourn in China. Expensive stuff this science, but hopefully worthwhile with regards to my overall education.

I'd also like to thank the Tunnacliffe lab members for putting up with me, and more importantly for helping me out over the years: Chiara, Matt B, Sohini, Rashmi, Sooraj, Tom, Purbani, Lukasz, Matt W (thanks for being an editor) and Neil. Other guest appearances include Guobao, Hanrui, Juan Jesús and Esther. Finally, special mention must go to Tatsuya for working so closely with me at various points – a lot of sweat and toil!

The technical and admin team at the Institute of Biotechnology also deserve a mention; I hope I wasn't too much of a pain. John Lester (DNA sequencing), Peter Sharratt (amino acid analysis), Len Packman (mass spectrometry), Nigel Miller and Nick Holmes (flow cytometry) have also been immensely helpful during my work – thanks!

I owe a large gratitude to the funding bodies that have made my time in Cambridge possible. The generosity of the Cambridge Commonwealth Trust, an Overseas Research Studentship, the C. T. Taylor Studentship, the Searle Scholarship and Emmanuel College has been much appreciated. The Cambridge Philosophical Society and European Research Council also kindly contributed towards the end of my work.

Finally, a huge amount of love to Mum & Dad and of course to my long-suffering Steph. I couldn't have made it without your support over the years. Thanks also to Jim for the editing eye. A shout out too for the Kiwis in London, who added another dimension to my time here. Chur.

Thanks for the good times Cambridge – all the people I've met, experiences had, opportunities taken – a real privilege.

ABBREVIATIONS

| | | | |
|-----------------|--|----------|--|
| Aav | <i>Aphelenchus avenae</i> | mCherry | Monomeric cherry fluorescent protein |
| aa | Amino acids | MFI | Median fluorescence intensity |
| AAA | Amino acid analysis | MIC | Minimum inhibitory concentration |
| A β | Amyloid beta | mRNA | Messenger RNA |
| AMP | Antimicrobial peptide | MW | Molecular mass |
| ara | L-arabinose | MWCO | Molecular weight cut-off |
| β -ME | Beta-mercaptoethanol | NB | Nutrient broth |
| BLAST | Basic local alignment search tool | NCBI | National Center for Biotechnology Information |
| bp | Base pairs | Ni-NTA | Nickel nitrilotriacetic acid |
| cfu | Colony forming units | NP-40 | Nonidet P-40; octylphenoxy polyethoxy ethanol |
| CPD | Cysteine protease domain | OD | Optical density |
| CPEC | Circular polymerase extension cloning | ORF | Open reading frame |
| CRAMP | Cathelin-related antimicrobial peptide | PAGE | Polyacrylamide gel electrophoresis |
| Da | Daltons | PBS | Phosphate buffered saline |
| DNA | Deoxyribonucleic acid | PCR | Polymerase chain reaction |
| DMSO | Dimethyl sulfoxide | PEG | Polyethylene glycol |
| DTT | Dithiothreitol | Pep2 | Peptide 2 |
| EGFP | Enhanced green fluorescent protein | PMBN | Polymyxin B nonapeptide |
| FACS | Fluorescence-activated cell sorting | RBS | Ribosome binding site |
| Fmoc | 9-fluorenylmethyloxycarbonyl | rcf | Relative centrifugal force |
| FPLC | Fast protein liquid chromatography | RNA | Ribonucleic acid |
| gDNA | Genomic DNA | RoI | Region of interest |
| GFP | Green fluorescent protein | SDS | Sodium dodecyl sulphate |
| HPLC | High-performance liquid chromatography | SPPS | Solid-phase peptide synthesis |
| IP ₆ | Inositol hexakisphosphate | TAE | Tris-acetate-EDTA |
| IPTG | Isopropyl β -D-1-thiogalactopyranoside | TBS-T | Tris buffered saline with Tween |
| kbp | Kilobase pairs | TFA | Trifluoroacetic acid |
| kDa | Kilodaltons | ThT | Thioflavin T |
| LB | Lysogeny broth | Tween 20 | Polyoxyethylene (20) sorbitan monolaurate |
| LEA | Late embryogenesis abundant | X-gal | 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside |
| LPS | Lipopolysaccharide | | |
| MBC | Minimum bactericidal concentration | | |

CONTENTS

| | |
|--|----|
| Preface | i |
| Abbreviations | iv |
| Contents | v |
| Chapter 1 – General introduction | |
| 1.1 Bioactive peptides | |
| 1.1.1 Brief history | 1 |
| 1.1.2 Properties of bioactive peptides | |
| 1.1.2.1 Specificity | 3 |
| 1.1.2.2 Stability | 4 |
| 1.1.2.3 Bioavailability | 6 |
| 1.1.2.4 Toxicity | 7 |
| 1.1.3 Potential uses and market | |
| 1.1.3.1 Potential uses | 8 |
| 1.1.3.2 Market | 11 |
| 1.1.4 Sources | |
| 1.1.4.1 Natural products | 15 |
| 1.1.4.2 Chemical synthesis | 15 |
| 1.1.4.3 Biosynthetic approaches | 17 |
| 1.2 Screening for bioactivity | |
| 1.2.1 Brief history | 19 |
| 1.2.1.1 High-throughput screening | 20 |
| 1.2.1.2 <i>In silico</i> screening | 21 |
| 1.2.1.3 Relevance to bioactive peptides | 21 |
| 1.2.2 Bioprospecting | 22 |
| 1.2.2.1 Screening for bioactive peptides | 23 |
| 1.2.2.2 Advantages of bioprospecting | 24 |
| 1.2.2.3 Use of DNA as an encoding source for peptides | 26 |
| 1.2.3 Biosynthetic approaches | 28 |
| 1.2.3.1 Recombinant peptide screens | 28 |
| 1.2.3.2 Advantages and disadvantages of recombinant peptide screens | 30 |
| 1.2.3.3 Screen considerations | 31 |
| 1.3 Proof of principle | |
| 1.3.1 Antimicrobial peptides | 32 |
| 1.3.2 Antiaggregation peptides | 33 |
| 1.3.3 Summary and aims | 33 |
| Chapter 2 – Materials and methods | |
| 2.1 Materials | |
| 2.1.1 Chemicals | 35 |
| 2.1.2 Oligonucleotides | 35 |
| 2.1.3 Bacterial expression strains and plasmids | 35 |
| 2.1.4 Chemically-synthesised peptides and cell-permeabilizing reagent | 36 |
| 2.2 General cloning methods | |
| 2.2.1 Common enzymatic manipulations of DNA | 36 |
| 2.2.2 Purification of DNA | 37 |

| | | |
|------------|---|----|
| 2.2.3 | Determination of DNA concentration | 37 |
| 2.2.4 | DNA sequencing | 37 |
| 2.2.5 | Amplification of DNA using polymerase chain reaction (PCR) | |
| 2.2.5.1 | General PCR | 38 |
| 2.2.5.2 | Colony PCR of plasmid inserts | 38 |
| 2.2.6 | Circular polymerase extension cloning (CPEC) | 39 |
| 2.2.7 | Agarose gel electrophoresis | 40 |
| 2.2.8 | Transformation of <i>E. coli</i> | 40 |
| 2.2.9 | Growth of <i>E. coli</i> | 40 |
| 2.3 | General protein methods | |
| 2.3.1 | Induction of protein expression in <i>E. coli</i> | 41 |
| 2.3.2 | Extraction of protein from <i>E. coli</i> | 41 |
| 2.3.3 | Polyacrylamide gel electrophoresis (SDS-PAGE) | |
| 2.3.3.1 | Glycine-based SDS-PAGE | 41 |
| 2.3.3.2 | Tricine-based SDS-PAGE | 42 |
| 2.3.3.3 | SDS-PAGE visualisation | 42 |
| 2.3.3.4 | Western blotting | 42 |
| 2.3.4 | Amino acid analysis | 43 |
| 2.3.5 | Mass spectrometry analysis | 43 |
| 2.4 | Recombinant AMP production methods | |
| 2.4.1 | Construction of pET28-AMP-CPD vectors | 43 |
| 2.4.1.1 | Construction of pET28-CPD | 44 |
| 2.4.1.2 | Construction of pET28-K2C18-CPD _{WT} | 45 |
| 2.4.1.3 | Construction of pET28-AMP-CPD variants | 45 |
| 2.4.2 | CPD on-column cleavage | 45 |
| 2.4.3 | Further AMP purification | |
| 2.4.3.1 | C18 hydrophobic resin purification of AMPs | 46 |
| 2.4.3.2 | Molecular weight cut-off (MWCO) purification of AMPs | 46 |
| 2.4.4 | Physical properties of peptides | 47 |
| 2.5 | Antimicrobial activity assays | |
| 2.5.1 | Microbial test strains | 47 |
| 2.5.2 | Agar radial diffusion assay | 48 |
| 2.5.3 | Antimicrobial liquid culture assay | 48 |
| 2.6 | Novel AMP screening methods | |
| 2.6.1 | Construction of AMP screen vectors | 48 |
| 2.6.1.1 | Construction of pBADm | 49 |
| 2.6.1.2 | Insertion of K2C18 and K2C38 into pBAD/gIII-A and pBADm | |
| 2.6.1.2.1 | Cloning of K2C18 | 49 |
| 2.6.1.2.2 | Cloning of K2C38 | 50 |
| 2.6.1.3 | Insertion of <i>hok</i> and <i>hokFS</i> into pBAD/gIII-A and pBADm | 50 |
| 2.6.1.4 | Construction of pAMP/S and pAMP | 50 |
| 2.6.2 | Growth curves | 51 |
| 2.6.3 | Preparation and insertion of genomic DNA (gDNA) into vectors | |
| 2.6.3.1 | Shearing of gDNA | 51 |
| 2.6.3.2 | Insertion of gDNA into pAMP/S and pAMP | 52 |
| 2.6.4 | Replica plate screening of pAMP/S and pAMP libraries | 52 |
| 2.6.5 | Identification of pAMP/S and pAMP library hits | 52 |
| 2.6.6 | Bioinformatic analysis of pAMP/S and pAMP library hits | 53 |
| 2.6.7 | Modification of pAMP/S library hits | 53 |
| 2.6.7.1 | Site-directed mutagenesis to introduce a frame-shift | 54 |
| 2.6.7.2 | Antibiotic resistance cassette exchange | 54 |

| | | |
|--|--|----|
| 2.6.7.3 | Secretion tag removal | 55 |
| 2.6.8 | Insertion and expression of AMP hits in pET28-CPD | |
| 2.6.8.1 | Insertion of selected AMP hits into pET28-CPD | 55 |
| 2.6.8.2 | Expression of selected AMP hits from pET28-CPD | 56 |
| 2.7 | Novel antiaggregant screening methods | |
| 2.7.1 | Construction of pAG2-Aβ42 | 57 |
| 2.7.1.1 | Insertion of mCherry operon | 57 |
| 2.7.1.2 | Construction and insertion of Library operon | 57 |
| 2.7.1.3 | Insertion of A β 42-EGFP into the <i>araBAD</i> site | 58 |
| 2.7.2 | Construction of pAG2-Aβ42 controls | 58 |
| 2.7.2.1 | Construction of pBADm-EGFP | 59 |
| 2.7.2.2 | Insertion of EGFP into Library site of pAG2-A β 42 | 59 |
| 2.7.2.3 | Construction of pAG2-GM6 | 59 |
| 2.7.2.4 | Insertion of Peptide 2 (Pep2) and AavLEA1 into pAG2-A β 42 | 59 |
| 2.7.3 | Insertion of random DNA library into pAG2-Aβ42 | 60 |
| 2.7.4 | Microscopy | 60 |
| 2.7.5 | Flow cytometry | |
| 2.7.5.1 | Flow cytometry analysis of <i>E. coli</i> fluorescence | 61 |
| 2.7.5.2 | Preparation of libraries for fluorescence-activated cell sorting | 61 |
| 2.7.5.3 | Fluorescence-activated cell sorting (FACS) | 62 |
| 2.7.5.4 | Calculations | 62 |
| 2.7.6 | Identification of pAG2-Aβ42 library hits on solid medium | 63 |
| 2.7.7 | Bioinformatic analysis of pAG2-Aβ42 library hits | 63 |
| Chapter 3 – Biosynthetic production of bioactive peptides: antimicrobials | | |
| 3.1 | Introduction | |
| 3.1.1 | Brief history | 64 |
| 3.1.2 | Structure and function | |
| 3.1.2.1 | Structures | 65 |
| 3.1.2.2 | Functions | 67 |
| 3.1.3 | Sources | 71 |
| 3.1.4 | Therapeutic potential | 72 |
| 3.1.5 | Production | |
| 3.1.5.1 | Chemical synthesis | 73 |
| 3.1.5.2 | Recombinant production | 74 |
| 3.1.5.2.1 | K2C18 as a model AMP | 75 |
| 3.1.5.2.2 | Current recombinant production methods | 75 |
| 3.1.5.2.3 | Use of an inducible, autocleaving tag | 76 |
| 3.1.6 | Summary | 76 |
| 3.2 | Results | |
| 3.2.1 | Construction of pET28-CPD | 77 |
| 3.2.2 | Expression and activity of recombinant K2C18-CPD | |
| 3.2.2.1 | Construction of pET28-K2C18-CPD _{WT} | 78 |
| 3.2.2.2 | Expression of pET28-K2C18-CPD _{WT} | 79 |
| 3.2.2.3 | Initial purification and activity of recombinant K2C18 | 79 |
| 3.2.3 | Design, construction and expression of AMP variants | |
| 3.2.3.1 | Design of AMP variants | 81 |
| 3.2.3.1.1 | K2C18 analogues | 82 |
| 3.2.3.1.2 | MCC18 analogues | 84 |
| 3.2.3.2 | Construction and expression of AMP variants | 85 |
| 3.2.4 | Purification of AMP variants | |
| 3.2.4.1 | Polishing of AMPs by C18 resin | 88 |

| | | |
|---|---|------------|
| 3.2.4.2 | Assessment of purified AMP concentration and purity | 89 |
| 3.2.4.3 | Yields of purified AMPs | 90 |
| 3.2.5 | Measurement of antimicrobial activity | 91 |
| 3.3 | Discussion | 93 |
| Chapter 4 – Screening for bioactive peptides: antimicrobials | | |
| 4.1 | Introduction | |
| 4.1.1 | Brief history | 99 |
| 4.1.2 | Recombinant screen approaches | 100 |
| 4.1.2.1 | Whole-cell screening | 101 |
| 4.1.2.2 | <i>Cis</i> or <i>trans</i> screens | 101 |
| 4.1.3 | Use of an AMP-sensitive production host | 102 |
| 4.1.3.1 | Exogenous versus endogenous hits | 102 |
| 4.1.3.2 | Secretion systems | 103 |
| 4.1.3.3 | Replica plating | 104 |
| 4.1.4 | Other screen considerations | 104 |
| 4.1.5 | Summary | 105 |
| 4.2 | Results | |
| 4.2.1 | Analysis of K2C18-CPD activity as a fusion protein | 106 |
| 4.2.1.1 | Purification and activity of crude K2C18-CPD _{WT} | 106 |
| 4.2.2 | Endogenous activity of K2C18 and K2C38 expressed in <i>E. coli</i> | 107 |
| 4.2.2.1 | Construction of expression vectors | 107 |
| 4.2.2.2 | Growth of <i>E. coli</i> expressing K2C18 and K2C38 on solid medium | 108 |
| 4.2.2.3 | Growth curves of <i>E. coli</i> expressing K2-CRAMP constructs | 110 |
| 4.2.2.4 | Western blot of <i>E. coli</i> expressing K2-CRAMP constructs | 111 |
| 4.2.2.5 | Morphology of <i>E. coli</i> expressing K2-CRAMP constructs | 112 |
| 4.2.2.6 | Seed growth curves of <i>E. coli</i> expressing K2-CRAMP constructs | 113 |
| 4.2.2.7 | Comparison to the persister effect | 115 |
| 4.2.3 | Construction of vectors for AMP screen | 117 |
| 4.2.3.1 | Construction of pAMP/S | 117 |
| 4.2.3.2 | Construction of pAMP | 117 |
| 4.2.4 | Screening for novel putative AMPs | |
| 4.2.4.1 | Use of bdelloid rotifer and human genomic DNA libraries | 118 |
| 4.2.4.2 | Replica plating of pAMP/S and pAMP gDNA libraries | 119 |
| 4.2.4.3 | Growth curves of putative AMP library hits | 120 |
| 4.2.4.4 | Bioinformatic analysis of pAMP/S and pAMP gDNA library hits | 123 |
| 4.2.5 | Further exploration of novel putative AMP hits | |
| 4.2.5.1 | Frame-shift of pAMP/S library hits | 130 |
| 4.2.5.2 | Removal of the secretion signal from pAMP/S library hits | 131 |
| 4.2.5.3 | Use of a different antibiotic selection marker | 131 |
| 4.2.6 | Validation of novel putative AMP hits | |
| 4.2.6.1 | Recombinant production of putative AMPs using CPD system | 132 |
| 4.2.6.2 | Activity of chemically-synthesised putative AMPs | 134 |
| 4.2.6.3 | Synergy of hits with cell wall permeabilizing reagents | 135 |
| 4.3 | Discussion | 136 |
| Chapter 5 – Screening for other activities: antiaggregation | | |
| 5.1 | Introduction | |
| 5.1.1 | Brief history | 144 |

| | | |
|---------|---|-----|
| 5.1.2 | Amyloid beta peptide (A β) | |
| 5.1.2.1 | A β 40 and A β 42 | 145 |
| 5.1.2.2 | A β 42 in Alzheimer's disease | 147 |
| 5.1.3 | Screens for A β 42 antiaggregants | 149 |
| 5.1.3.1 | Small molecules and antibodies as antiaggregants | 150 |
| 5.1.3.2 | <i>In vitro</i> screening and animal models | 150 |
| 5.1.3.3 | Peptides as antiaggregants | 151 |
| 5.1.3.4 | Recombinant screen approaches | 152 |
| 5.1.4 | Recombinant screen considerations | |
| 5.1.4.1 | Antiaggregation controls | 155 |
| 5.1.4.2 | Use of flow cytometry | 156 |
| 5.1.4.3 | Random oligonucleotide libraries | 158 |
| 5.1.5 | Summary | 158 |
| 5.2 | Results | |
| 5.2.1 | Construction of vectors for antiaggregant screen | |
| 5.2.1.1 | Construction of pAG2-A β 42 | 159 |
| 5.2.1.2 | Construction of pAG2-A β 42 controls | 160 |
| 5.2.2 | Profiling of pAG2-A β 42 and controls | |
| 5.2.2.1 | Expression of pAG2-A β 42 and controls | 162 |
| 5.2.2.2 | Fluorescence microscopy of pAG2-A β 42 and controls | 164 |
| 5.2.2.3 | Growth rates of pAG2-A β 42 and controls | 166 |
| 5.2.2.4 | Flow cytometry of pAG2-A β 42 and controls | 168 |
| 5.2.3 | Using FACS to screen for novel antiaggregant peptides | |
| 5.2.3.1 | Creation of random DNA library | 173 |
| 5.2.3.2 | Plasmid stability and assessment of library insertion | 174 |
| 5.2.3.3 | FACS of pAG2-A β 42-74 and pAG2-A β 42-74 libraries | 177 |
| 5.2.3.4 | Further analysis of FACS data | 181 |
| 5.2.4 | Using solid medium to screen for novel antiaggregant peptides | 184 |
| 5.2.4.1 | Fluorescence of colonies expressing pAG2-A β 42 and controls | 184 |
| 5.2.4.2 | Screening for novel antiaggregant peptides | 188 |
| 5.2.4.3 | Validation through flow cytometry | 189 |
| 5.2.4.4 | Fluorescence microscopy of library hits | 195 |
| 5.2.4.5 | Bioinformatic analysis of library hits | 196 |
| 5.2.4.6 | Selection of putative antiaggregant peptides for <i>in vitro</i> analysis | 204 |
| 5.3 | Discussion | 205 |
| | Chapter 6 – Final discussion | 213 |
| | References | 218 |
| | Appendices | 233 |

CHAPTER 1 – GENERAL INTRODUCTION

1.1 Bioactive peptides

1.1.1 Brief history

Peptides are ubiquitous in living systems. In their simplest form, these constitute oligomers that are ribosomally synthesised from a pool of 20 common amino acids. The particular sequence order is encoded via a universal genetic code specified by nucleic acid templates. Organisms typically use peptides as signalling molecules to convey information regarding cellular state between one another at a unicellular level, or to coordinate global tissue responses to physiological cues at a multicellular level. Sewald & Jakube (2009) have written an excellent textbook that serves as an introduction to the peptide field.

The development of novel therapeutics is a driver of interest in these common, yet key, compounds. Whether it is the dysregulation of a normal biological process in humans, treatment of a pathogenic or undesired species, or a more cosmetic application, being able to remediate or alter such processes underpins modern medicine. Peptides exhibiting bioactivity (that is, provoking a biological response) are a relatively new class of therapeutics (Sewald & Jakubke, 2009). Rather it is small molecule drugs, typically with a molecular mass of 500 daltons or less, that have traditionally been used as medicines. For example, Hippocrates mentions the use of willow bark preparations as an analgesic in the 5th century BC, and the active ingredient, salicylic acid, has now been in use for two centuries as aspirin (Awtry & Loscalzo, 2000).

Medicine had to wait until the 20th century, however, for peptides to become feasible therapeutics, due to the greater technical capabilities required to synthesise or purify these compounds in comparison to small molecule drugs. Insulin is the most famous example of a bioactive peptide. Used by the body as a global hormone to promote the uptake of blood glucose in liver, muscle and adipose tissue, this 51 amino acid entity (5.8 kDa; two subunits joined by disulphide bonds) is produced by the islets of Langerhans (β -cells) in the pancreas of humans via processing of the proinsulin precursor (Steiner & Oyer, 1967). Lack of insulin production through autoimmune destruction of these cells, or an insufficient

signalling response to insulin itself, are the underlying physiological causes of diabetes mellitus (Steiner et al., 2009).

Diabetes, and the discovery of insulin, pioneered much peptide research last century (Sneider, 2001). Fredrick Banting and Charles Best first isolated extracts of the mature peptide while experimenting on canine pancreases, and successfully used natively purified insulin (sourced from foetal calf pancreas) to treat paediatric patients for the then fatal disease in 1922 (Best, 1962). Porcine insulin was soon found to be a better heterologous match for humans, but due to immunogenic concerns and source limitations, deriving a human version of this life-saving peptide became a goal. Insulin became the first protein to have its amino acid sequence elucidated in 1951 (Sanger, 1959) and was subsequently chemically-synthesised in its human form in the 1960s (Katsoyannis & Tometsko, 1966). Finally, in 1978 human biosynthetic insulin was successfully produced recombinantly in *Escherichia coli* by Genentech Inc. (Genentech, 1978), and its improved variants have been mass-produced economically for medical use from the 1980s onwards (Johnson, 1983).

More modern examples of useful bioactive peptides include Eli Lilly and Company's teriparatide (trade name Forteo, 34 aa, 4.1 kDa), a truncated recombinant version of the human hormone parathyroid which is used to treat some forms of osteoporosis via regulating calcium metabolism and stimulating osteoblast activity (Saag et al., 2007); and Hoffmann-La Roche Ltd.'s enfuvirtide (trade name Fuzeon, 36 aa, 4.5 kDa), a human immunodeficiency virus fusion inhibitor which is employed as a therapy of last resort and acts as an antagonist of the viral transmembrane protein gp41 to inhibit its activation by other key viral entry proteins (Kilby et al., 1998).

Although peptides and proteins are both composed of amino acids, they are differentiated on the basis of molecular size. It is commonly accepted that the peptide label may be applied to proteinaceous compounds with 100 or fewer amino acid residues (Lax, 2010). However, short peptides (e.g. 2 aa) are more akin to small molecule drugs in overall size, while large peptides (e.g. greater than 50 aa) are more similar to small proteins. Both are peptides, and although this appears to

be a semantic issue, it has regulatory implications. Small molecule drugs and proteins are approved for therapeutic use in different manners by safety authorities (Lax, 2010), resulting in a fuzziness with regards to exact peptide size definition.

Because of amino acid diversity, even short peptides allow an enormous number of possible sequence variations. For instance, if just the natural complement of 20 amino acids is considered (non-natural residues are discussed in *Section 1.1.4.2*), a 4-residue peptide has 20^4 possible permutations (i.e. ~160,000 distinct compounds), while a 14 residue peptide has 20^{14} (~ 1.64×10^{18} compounds). In practice these theoretical compounds will not be as diverse as the numbers suggest, as some amino acids exhibit similar properties e.g. the hydrophobicity of leucine, isoleucine and valine. The use of random combinations of residues in a short peptide will typically lead to a linear, unstructured and consequently unconstrained compound. A lot of bioactive peptides, however, possess some form of structure (Watt, 2006). Secondary structures, such as α -helix, β -sheet, disulfide bonds and cyclisation are possible in even short chains of amino acids (Sewald & Jakubke, 2009). Dimerisation and oligomerisation is also possible (Sato et al., 2006a). For example, insulin is a two-chain compound cleaved from a single precursor, possessing three disulphide bonds and a large amount of α -helix (Steiner et al., 2009). The significance of secondary structure is discussed later in *Sections 1.1.2.2* and *1.2.2.2*.

1.1.2 Properties of bioactive peptides

Bioactive peptide research focuses on therapeutic applications. The following sections introduce some of the properties, advantages and challenges of using peptides as therapeutic drugs.

1.1.2.1 Specificity

A number of bioactive peptides are truncations derived from the most active part of a parent protein. For instance, the peptide therapeutics teriparatide (34 aa) and enfuvirtide (36 aa) mentioned earlier are the effector regions of their parent proteins – parathyroid hormone (84 aa) and gp41 (345 aa) respectively (Watt, 2006). Because of their large number of hydrogen bond donor/acceptor moieties, these

peptide truncations offer greater specificity than a small molecule drug, and so consequently exhibit less off-target binding (Vlieghe et al., 2010). An advantage of this higher specificity is that typically only a low amount of peptide (μg to a few mg per day) needs to be administered to a patient for an effective outcome (Lax, 2010), which may reduce toxicity (see *Section 1.1.2.4*).

1.1.2.2 Stability

A significant drawback of using natural peptides is their short half-life *in vivo*. Firstly, their small size means they are susceptible to being removed from the blood stream via filtration in the kidneys (McGregor, 2008). Secondly, their unstructured linear form makes them particularly labile to the activity of peptidases present in the blood stream, liver and kidney, giving them a typical half-life of only minutes in the blood stream (Werle & Bernkop-Schnürch, 2006). Overcoming protease susceptibility is important for parenterally administered peptides, and is especially relevant if a peptide is delivered orally. Even if the acidic pH of the stomach is safely traversed, the lumen of the small intestine poses the greatest challenge – the pancreas secretes peptidases such as chymotrypsin, trypsin, elastase and carboxypeptidases in large quantities for the purpose of complete protein digestion (Vlieghe et al., 2010).

Peptide instability can be both an advantage and disadvantage: if treatment only requires a short systemic residency time, a rapid turnover of the bioactive peptide gives less opportunity for undesirable side-effects to occur (Loffet, 2002). This is mostly applicable to hormonal analogues (e.g. insulin), whose continued systemic presence is undesirable with regards to bodily homeostasis. Despite this, most bioactive peptides would be preferred in a more stable, long-lasting format. The use of polyethylene glycol (PEG) as a conjugate is one way of increasing peptide half-life, as protease function may be inhibited via conferred steric hindrance near the catalytic site (Lien & Lowman, 2003). Furthermore, PEG exhibits high solubility, low immunogenicity and a lack of toxicity, and seems to confer these properties on its conjugates (Werle & Bernkop-Schnürch, 2006). For example, the use of PEGylation increased the half-life of glucagon-like peptide-1 (a gut hormone to promote insulin production) by 16-fold in rat models (Lee et al., 2005a), and 330-fold when conjugated to interferon α -2-b (a hepatitis C anti-viral) (Ramon et al., 2005).

Polymer chains, typically ranging from 5 to 40 kDa in molecular mass, may be attached at a specific peptide residue, or spread over multiple sites (Sato et al., 2006a). However, using multiple PEGylation sites complicates the task of creating a homogenous peptide batch, something which is desirable for therapeutic compounds (McGregor, 2008). Another positive side-effect of bulking the peptide mass up is the reduction of clearance from the bloodstream via the kidneys, as molecules with a mass over 50 kDa generally require alternative routes of breakdown (i.e. by the liver) before being disposed of (Sato et al., 2006a). Adding carbohydrate chains such as glucose or xylose moieties is an alternative to PEGylation (Vlieghe et al., 2010).

A simpler modification to improve protease resistance is to acetylate the N-terminal residue or amidate the C-terminus of a peptide. Such measures have been shown to reduce susceptibility to exopeptidases in particular (Werle & Bernkop-Schnürch, 2006). Another approach involves the swapping of one or more natural L-amino acids with their D-amino enantiomers. For example, Novartis Pharmaceuticals AG's octreotide (trade name Sandostatin, a somatostatin mimetic to inhibit growth hormone release) is an 8-residue truncated version of human somatostatin, with a key tryptophan in the D configuration: half-life in blood circulation was increased from a few minutes to 1.5 hours (Harris, 1994). Because such enantiomer swapping changes the structure of the peptide, it may no longer properly fit a protease active site and thus confers proteolytic resistance (McGregor, 2008). Fully "retroinverso" peptides, where the original peptide is produced in reverse with residues in the D-form, have also been explored (Watt, 2009).

Structurally constraining the peptide is another method of reducing susceptibility to protease action. Introducing cross-linking disulfide bonds through cysteine side-chains, or cyclising the peptide head-to-tail (i.e. chemically-ligating the N- and C-terminals) may prevent it from fitting the cleaving site of a protease (Vlieghe et al., 2010). Providing a more rigid structure in this manner may also help constrain residue side-chains in a more "active" conformation, which may not be thermodynamically favoured in the linear form (Sato et al., 2006a). This potentially results in a higher binding affinity for the desired target.

Lastly, a peptide may be attached to a scaffold that already possesses innate stability. The F_C region of an antibody is one such candidate, which has a serum retention time of several days, thanks in part to its systemic recycling via the neonatal F_C receptor (Roopenian & Akilesh, 2007). Another scaffold that may be harnessed is the abundant serum protein albumin, which has a half-life of 19 days in humans (Lien & Lowman, 2003). Several companies, such as Genentech Inc. and Dyax Corp., are investigating using these well-described proteins as scaffolds to carry bioactive peptides (McGregor, 2008). Steric hindrance is a concern with regards to interfering with peptide activity, especially considering that such a scaffold will be 10 to 30 times greater in size than its cargo (Sato et al., 2006a).

Finally, it should be emphasised that no matter what approach to improving peptide stability is used, any structural change to the active peptide may of course affect function. Such modifications need to be empirically validated.

1.1.2.3 Bioavailability

A major caveat to using peptides as therapeutics is that their targets should be in or exposed to the extracellular milieu – it is much more difficult for a peptide to access intracellular targets, as this involves crossing one or more biological membranes. While cell-penetrating peptides are being developed (Sebbage, 2009) and are discussed as a potential application in *Section 1.1.3.1*, any fusions of this sort are context-specific and need to be evaluated as such. Small molecule drugs, depending on their solubility and lipophilicity, seem to have a greater innate ability to passively diffuse into cells (McGregor, 2008). Endocytosis of peptides by cells still fails to solve the problem, as this merely traps them in another biological compartment (endosomes) (Foerg & Merkle, 2008). Even assuming that the peptide can escape this vesicle and its associated proteases, it may still have to cross additional biological membranes if the target is located inside an organelle e.g. a mitochondrion. However, there are many extracellular receptors that may be targeted, and activating or suppressing their particular signal cascades is one manner of influencing intracellular events. This is reflected in the targets of current peptides on the market (see *Section 1.1.3.2*).

Delivery of a peptide to its site of action, even if it is extracellular, may still pose difficulties. Parenteral injection is the most common form of delivery – the examples of insulin, teriparatide and enfuvirtide presented earlier are all administered in this manner. However, patient discomfort and compliance are issues when it comes to injections. The most desirable route of administration is oral (Vlieghe et al., 2010). Despite the severe stability issues that face a peptide traversing the gut as well as poor penetration of the intestinal mucosa to enter systemic circulation, many researchers have spent the last decade trying to develop an orally-compatible form of insulin (Lien & Lowman, 2003). It is envisaged that solving this problem will allow the oral delivery of other peptides in a generic manner. However, to date nothing has appeared on the market as trials continue. Alternative delivery approaches to the oral route include transdermal patches for passive diffusion through the skin; iontophoresis for active electromotive administration; sonophoresis for ultrasound-mediated administration; and aerosols for inhalation and subsequent access to the blood stream through the pulmonary system (Lien & Lowman, 2003; Pichereau & Allary, 2006). The successful and mainstream employment of one of these methods will be a boon to the development of future bioactive peptides, as delivery remains one of the key obstacles to pharmaceutical development (Lax, 2010). One current solution, if the bioavailability of a peptide is low through either poor stability or delivery, is to increase the dosage (if safe to do so) via more frequent administration – or more preferably, through a sustained delivery system (i.e. slow constant release) (Werle & Bernkop-Schnürch, 2006). However, this leads to an increased cost of treatment (Lien & Lowman, 2003).

1.1.2.4 Toxicity

Bioactive peptides are usually composed of natural amino acids or their enantiomers, and thus seem more natural to the body than some synthetic small molecule drugs (Vlieghe et al., 2010). Metabolic break-down products are amino acids, and will be dealt with by the body just as normal protein turnover is, thus limiting undesired side-reactions. Where side-effects are seen it is usually related to dosage amounts, or, if injected, as a localised effect around the site of administration (Lax, 2010). This lack of systemic toxicity is seen as one of the key strengths of bioactive peptides (Vlieghe et al., 2010).

Despite this, potential immunogenicity and eventual hypersensitivity to the therapeutic remain as concerns (Sato et al., 2006a). However, it is postulated that peptides, because of their short length, do not exhibit major histocompatibility complex class II epitopes (i.e. are not recognised by key lymphocytes such as macrophages or B cells), and thus fail to invoke an adaptive immune response (Vlieghe et al., 2010). PEGylation might also be used to confer the non-immunogenic properties of PEG to a peptide (Werle & Bernkop-Schnürch, 2006). Non-human derived peptide examples exist that have proven their lack of immunogenicity in clinical trials, such as enfuvirtide (39 aa, a HIV fusion inhibitor described earlier) and salmon calcitonin (31 aa, used to treat osteoporosis) (Loffet, 2002). Monitoring of potential immunogenicity, however, should remain standard practice during the development of systemically administered peptides.

1.1.3 Potential uses and market

1.1.3.1 Potential uses

The earliest therapeutic use of bioactive peptides fell in the hormonal category, e.g. insulin to remedy insulin deficiency in diabetes mellitus type 1, or insulin-like growth factor 1 to treat growth deficiencies in children having an imbalance in this hormone (Rosenbloom, 2006). While hormonal mimetics are the most popular class of bioactive peptides, now that stability and bioavailability issues are being addressed there is a large range of other potential applications: allergies, analgesics, asthma, arthritis, baldness, cardiovascular diseases, diabetes, gastrointestinal dysfunction, growth problems, haemostasis, immune disorders, impotence, incontinence, infective diseases, inflammation, obesity, oncology, and osteoporosis (Vlieghe et al., 2010).

Vaccines are another potential use of peptides, but represents conflicting aims. The short length of peptides is thought responsible for their low immunogenicity (discussed previously) (Vlieghe et al., 2010), but with vaccines a peptide of longer length (>69 aa; Watt [2006]) is meant to serve as an epitope to train the acquired immune system. An example is the experimentation with peptides as an anti-cancer vaccine, not only prophylactically, but also to treat pre-existing cancerous cells (Reichert & Wenger, 2008). This may be problematic if the immune system is already tolerant of these rogue cells. Although some clinical trials have been

performed with little report of peptide toxicity, any significant anti-cancer effect has yet to be found (Lien & Lowman, 2003). To overcome a lack of immune response, it may be that coupling the desired epitope to a larger scaffold protein may be a solution (see *Section 1.1.2.2*). However, this approach risks raising antibodies that do not recognise the native cancer epitope when it presents itself independently.

In relation to the above, peptides could be used to modulate the immune system in other ways. For example, LL-37 (an antimicrobial peptide; see *Section 1.3.1*) was found to enhance the adaptive immune response when added as a co-translated adjuvant to a DNA vaccine against tumour cells in mice (An et al., 2005). This is likely related to the ability of the peptide to bind DNA and allow its subsequent recognition and processing by dendritic cells (primarily responsible for antigen presentation to T-cells) (Lande et al., 2007). As LL-37 has also been shown to manipulate dendritic cell activation and differentiation (Davidson et al., 2004), the use of analogues to therapeutically modulate the immune system could be possible.

Cell-penetrating peptides (CPPs), while not strictly the active therapeutic, are a potential solution to poor peptide bioavailability. As mentioned earlier, peptides, as well as potentially therapeutic oligonucleotides, proteins and other large macromolecules, cannot readily cross the plasma membranes of cells. The hydrophobic nature of lipid-based membranes acts as a non-permeable barrier to hydrophilic and charged compounds (i.e. most soluble peptides), and this selectivity is vital for correct cellular function (Stewart et al., 2008). While membranes do contain active transport systems (e.g. membrane-spanning pore proteins) for the uptake and secretion of physiologically important products, these are hard to take advantage of in a manner generic enough to transport compounds of disparate size, charge and structure. It would be very beneficial if such generic trans-membrane transport were possible, as this would open the intracellular environment to therapeutic targeting. Research into CPPs has been ongoing since 1988, when a 14-residue truncation of the Tat protein (a transcription-activating factor coded for by the human immunodeficiency virus type 1) was found to be sufficient to promote its uptake into HeLa cells in a potentially non-specific and non-toxic manner (Green & Loewenstein, 1988). Tat truncations and other CPPs, such as a 16-residue truncation of the Antennapedia transcription factor from *Drosophila melanogaster*

(Joliot et al., 1991), have since been fused or cross-linked to various proteins, peptides and oligonucleotides and trialled for mammalian intracellular delivery (Lindgren et al., 2000). There are some examples to prove their potential: the transport of small interfering RNA to inhibit translation of a luciferase reporter in mammalian cell cultures (Muratovska & Eccles, 2004); the successful transfer of Cre recombinase (38 kDa) to the nucleus to mediate DNA recombination via LoxP sites (Wadia et al., 2004); and the remarkable delivery of β -galactosidase (116 kDa) across the blood-brain barrier in a mouse model (Schwarze et al., 1999). Being able to deliver passenger compounds across the blood-brain barrier (via passive or active membrane uptake) is a key goal in itself (Vlieghe et al., 2010). While their exact mode of entry is still debated, i.e. direct penetration versus endocytosis and subsequent intracellular escape, it is most likely CPP-specific (Stewart et al., 2008). To complicate matters further, both modes could be utilised by a single peptide: LL-37 (mentioned previously) has not only been proposed to enter epithelial cells via specific receptors (Lau et al., 2005), but also to co-transport extracellular plasmid DNA via lipid raft endocytosis (Sandgren et al., 2004). More research is required on CPPs, as structure/function relationships are still unclear (Lindgren et al., 2000; Foerg & Merkle, 2008). It is worth noting, however, that the CPP field is moving away from the concept of a “generic” transporter (Foerg & Merkle, 2008). The development of cell- or organelle-specific penetrating peptides would be of value, especially with regards to partnering with other bioactive peptides that are membrane-impermeable on their own.

One area where bioactive peptides could become important is against the so-called “undruggable” targets, i.e. protein-protein interactions (Watt, 2009). Small molecule drugs have been particularly successful against targets with defined binding pockets, such as cell membrane signal receptors e.g. the G-protein coupled receptor family, or enzyme active sites e.g. angiotensin-converting enzyme inhibitors for treatment of hypertension (Drews, 2000). They have not, however, had much success against targets lacking defined binding pockets, such as transcription factors or other intracellular global regulators (Drews, 1996). As a consequence, only 417 protein products have been therapeutically targeted in humans i.e. ~2% of the potential proteomic total (Drews, 1996). A number of protein-protein interactions occur over large, “flat” surfaces, as opposed to surface involution “hot-spots” which

small molecule drugs favour (Watt, 2009; Moellering et al., 2009). Bioactive peptides, which are typically an order of magnitude larger with respect to a small molecule drug, may be better suited to such binding – an increased number of potential hydrogen bond partners (i.e. increased avidity) can cumulatively overcome the unfavourable thermodynamics of a flat surface, leading to sufficient binding affinity for an effect (Sato, 2006a). Aileron Therapeutics Inc. has developed examples of peptides that disrupt protein-protein interactions. Chemically-synthesised peptides that incorporate two non-natural side-chains are subsequently “stapled” together, locking the peptide into an α -helix secondary structure (Schafmeister, 2000). Such constrained peptides are then used to hit “undruggable” targets: one such example acts as an antagonist to disrupt the assembly of the NOTCH transcription factor complex (implicated to be over-active in T-cell acute lymphoblastic leukaemia) (Moellering et al., 2009), while another acts as an agonist to activate the pro-apoptotic BAX protein in mammalian cells to induce cell death (Gavathiotis et al., 2008).

Another use for bioactive peptides is in the “cosmeceutical” industry – adding biologically active compounds to cosmetics. In 2006 the global anti-ageing cosmeceutical market was estimated to be worth US\$3.5 billion (Pichereau & Allary, 2006). Lipotec SA are developing peptide derivatives of the botulinum neurotoxin to treat wrinkles, and Therapeutic Peptides Inc. are developing peptides that stimulate collagen production (Pichereau & Allary, 2006).

1.1.3.2 Market

The market for bioactive peptides has slowly been growing over the past forty years. Research on hormonal homologues such as insulin in the 1970s and 1980s took 30 years to diversify into areas such as oncology, analgesics, osteopathy, haematology and virology (Reichert et al., 2008). Part of this slow development can be attributed to a dogma of pharmaceutical companies that entities with a molecular mass over 600 Da would not be useful as drugs due to poor bioavailability (Loffet, 2002). Furthermore, peptides do not meet a number of other noted physiochemical criteria for drug-like properties formalised a decade ago by Lipinski and co-workers (i.e. size, lipophilicity, a limited number of hydrogen bond donors/acceptors)

(Lipinski et al., 2001). An increase in peptide therapeutic growth has been driven by both a demand for new treatments for conditions or diseases in which small molecules are underperforming, and by pharmaceutical companies looking for new molecular entities to replenish emptying development pipelines (Vlieghe et al., 2010). The past decade has also seen a number of approaches taken to mitigate the perceived failures of bioactive peptides, such as overcoming bioavailability, stability and delivery issues (reviewed in *Section 1.1.2*). Production costs of chemically-synthesising short peptides (<50 aa) have also fallen as greater demand for contract manufacturing has led to more competitive pricing. For instance, for large-scale production the cost may now be less than US\$1 per gram per amino acid residue (Bray, 2003).

Recently the Peptide Therapeutics Foundation (founded 2008; <http://www.peptidetherapeutics.org/>) has started to track peptide-based drugs from the pre-clinical stage through to approval by the relevant regulatory authority (usually the United States Food and Drug Administration or European Medicines Agency). *Figure 1.1* shows the rapidly increasing number of peptides entering clinical trials each year.

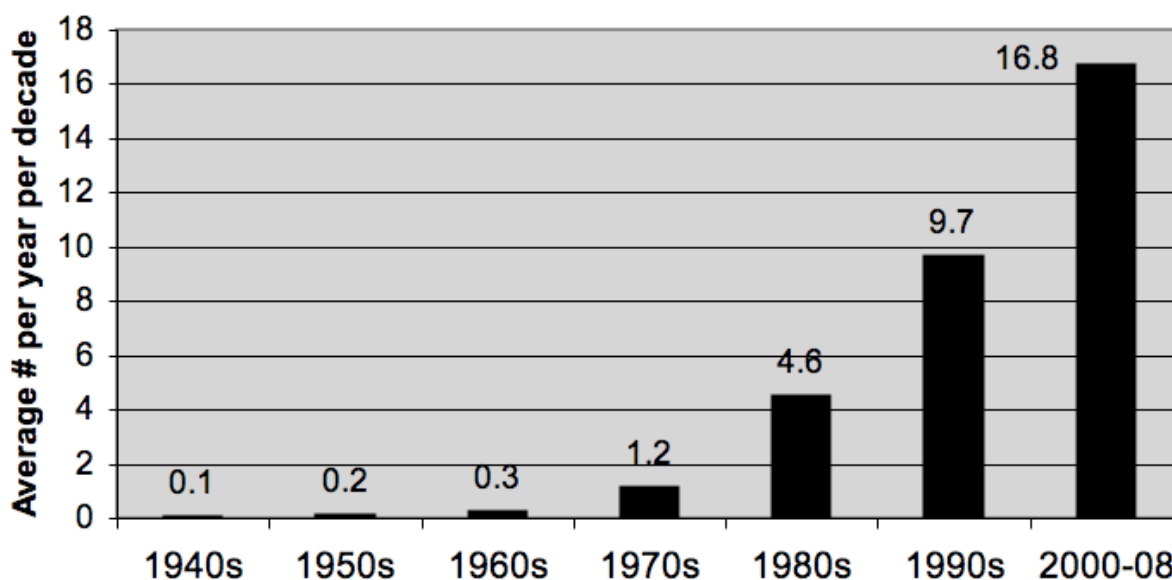


Figure 1.1: Average number of new peptide drug candidates per year, per decade. Numbers entering clinical trials, figure adapted from Reichert et al. (2010).

In 2001 there were approximately 150 peptide therapeutics under known development; in 2004, this increased to 400, with 30 peptides past clinical trials and approved for market (bringing in sales of US\$6 billion) (Pichereau & Allary, 2006). As of 2010, 54 approved peptide drugs were on the market, and accounted for annual sales of approximately US\$13 billion (Lax, 2010; Reichert et al., 2010). While this represents only ~1.5% of the total drug market in that year (~US\$870 billion), previous predictions estimated this would not be achieved until 2013 (Pichereau & Allary, 2006). Growth in peptide drugs is progressing, and on current trends is projected to be at least 7% per annum for several years (Lax, 2010). In 2010 approximately 140 bioactive peptides were in clinical trials (15 were at Phase 3 or under regulatory review), with approximately 400 in the pre-clinical stage (Reichert et al., 2010; Lax, 2010).

The success rate for peptides that enter clinical trials is relatively high, with 23-26% of candidate entries between 1984 and 2000 being approved (Reichert et al., 2010). This is approximately twice the success rate achieved by small molecule drugs (Reichert et al., 2008). However, as higher numbers of peptides have entered clinical trials over the last decade, this success percentage may decrease in the post-2000 cohort. The main failing of bioactive peptides during clinical trials seems to be a lack of efficacy at the Phase 2 stage. Only ~45% of peptide candidates proceed from Phase 2 to Phase 3 (Reichert et al., 2008). This issue of efficacy relates to the twin problems of stability and bioavailability, discussed in *Sections 1.1.2.2* and *1.1.2.3*. As these impediments are being addressed, it is predicted that interest in developing bioactive peptides will continue to grow (Reichert et al., 2010).

Certain target classes are heavily favoured in the small number of entities already on the market (54 as of 2010). As outlined earlier, the vast majority are extracellular (>90%), reflecting the difficulties associated with getting a peptide across membrane(s) to access intracellular targets. For example, just under half of the total approved peptides are agonists for members of the G-protein coupled receptor family (Reichert et al., 2010). Of these, the gonadotropin-releasing hormone receptor (a global regulator for controlling sexual hormone levels, manipulated to treat breast and prostate cancers among other indications) is a favoured target

(Loffet, 2002). Other targets include somatostatin receptors (to inhibit growth hormone release), calcitonin receptors (to lower blood calcium levels), platelet aggregate inhibitors (to prevent blood clotting) and viral proteins (Pichereau & Allary, 2006).

There is potential for some of these peptides to be worth a great deal economically, with a number already generating sales reaching US\$1 billion per year. In 2008, Teva Pharmaceutical Industries Ltd.'s glatiramer acetate (trade name Copaxone, a random polymer of glutamic acid, lysine, alanine and tyrosine used to treat multiple sclerosis by acting as a decoy myelin basic protein to the immune system) had sales of US\$3.18 billion; Novartis Pharmaceuticals AG's octreotide acetate (trade name Sandostatin, a somatostatin mimetic to inhibit growth hormone release) had sales of US\$1.12 billion; Eli Lilly and Company's teriparatide (trade name Forteo, a parathyroid hormone analogue to treat varieties of osteoporosis) had sales of US\$780; and Amylin's exenatide (trade name Byetta, an incretin mimetic to aid insulin production) had sales of US\$751 million (Reichert et al., 2010). This proof that bioactive peptides are capable of earning substantial revenue will help drive further interest in developing new peptides. The peptides listed above are all administered via injection (i.e. uncomfortably), so the pharmaceutical industry seems to accept that there is still a viable market for drugs that cannot be taken orally (Loffet, 2002; Lax, 2010).

In comparison to well-described small molecule and antibody scaffolds, an additional advantage of bioactive peptides is that the intellectual property space is relatively unrestricted. While these alternative two classes rely on tweaking variant moieties as part of defined scaffolds, which are heavily patented, *de novo* peptide sequences are free to explore (Watt, 2008). A newcomer to the market provides an example of this. Bicycle Therapeutics Ltd. screens 17 amino acid motifs (12 residues are randomised) that are structurally constrained through subsequent chemical crosslinking by 3 cysteine residues (Heinis et al., 2009). The aim is to mimic the structural conformation constraints of an antibody's complementarity determining region (responsible for antigen recognition) while avoiding the associated intellectual property.

1.1.4 Sources

Bioactive peptides are obtained from three primary sources: naturally occurring products, chemical synthesis, or recombinant biosynthesis.

1.1.4.1 Natural products

Proteins play an essential role in the plethora of different activities a cell performs on a daily basis. These include cell signalling, enzymatic catabolism and anabolism, cellular replication, and homeostasis and repair. Because different organisms cope with the selection pressures placed on them by varying environments in different ways, it follows that many organisms will possess proteins with functions that are novel and perhaps useful as therapeutics e.g. antimicrobial peptides (see *Section 1.3.1*). While *Section 1.2.2* goes into more detail, it is sufficient to note that most bioactive peptides have historically been sourced natively from plants, animals or insects. In particular, bioactive peptides have been isolated from the active domains of various hormones or other proteins that possess the desired biological activity (McGregor, 2008).

Recent examples include The Medicines Company's bivalirudin (trade name Angiomax, 20 aa, 2.2 kDa), a thrombin inhibitor, which is a variant of the anticoagulant hirudin found in the saliva of medicinal leeches (Warkentin et al., 2008); and Amylin Pharmaceuticals Inc.'s exenatide (trade name Byetta, 39 aa, 4.2 kDa), found in the saliva of the Gila monster (*Heloderma suspectum*), a homologue of human glucagon-like peptide 1 that is used to stimulate insulin production in diabetes mellitus type 2 patients (Leader et al., 2008). However, once these peptides have been isolated from their natural source and their activity profiled, they are typically produced commercially using either recombinant or more commonly chemical synthesis technology due to economies of scale.

1.1.4.2 Chemical synthesis

Solid-phase peptide synthesis (SPPS), pioneered in the 1960s (Merrifield, 1963), is still the gold-standard method for chemical synthesis of bioactive peptides today. Fmoc (9-fluorenylmethyloxycarbonyl) mediated chemistry allows for the sequential building of a peptide from protected amino acid residues starting at the C-terminus (attached to a solid resin support) and working "back" towards the N-terminus. In

brief, synthesis follows a cycle for the addition of each amino acid: coupling, wash (of unreacted compounds), deprotecting (of the amino group so the next residue can couple to the chain), and another wash (Guzmán et al., 2007). Such a process can be highly automated, and several peptides can be co-synthesised over a period of a few days (Bray, 2003).

Peptides with a length of less than 30 amino acids are more economically manufactured in comparison to longer peptides using SPPS (Lax, 2010). This is because the longer the peptide chain desired, the lower the yield achieved: there is a greater chance of product side-reactions or racemisation (i.e. heterogeneity between individual peptides due to some residues adopting the D-enantiomer form) (Bray, 2003). In addition, some peptides are prone to aggregation during synthesis, or may fold to obscure access to the N-terminus, thus making further extension difficult (Bray, 2003). This is especially true for peptides containing a high proportion of hydrophobic residues, and thus the specific properties of the target sequence will determine the efficiency, purity and yield of peptide synthesis. One method for overcoming these difficulties is the use of microwaves. As the polar peptide backbone constantly tries to align itself to the applied electromagnetic field, it works against the aggregative forces imparted by various amino acid side-chains, allowing full-length product synthesis to be achieved (Palasek et al., 2007).

For synthesis of longer peptides (typically >100 aa), two common approaches are taken (Bray, 2003). Convergent synthesis is a mixture of solid-phase and solution-phase synthesis – amenable short sequences are built up using traditional solid-phase synthesis, and these purified fragments are subsequently condensed together in a liquid-phase reaction (Vlieghe et al., 2010). Chemical ligation is a variation on this technique, occurring prior to fragment purification: specific N- or C-terminal residues (usually cysteine) are coupled chemoselectively, followed by rearrangement to form an amide bond prior to purification of the final product (Guzmán et al., 2007).

One advantage of chemical synthesis over recombinant production (discussed below) is the ability to incorporate orthogonal residues into a peptide (McGregor, 2008). Such peptidomimetics may include D-enantiomers, non-natural chemical

side-chains (such as norleucine or homoserine), and other labels such as chromophores (e.g. fluorescein isothiocyanate), chemical ligands for subsequent binding (e.g. biotin labels to be captured by streptavidin), or nucleic acid conjugates (Vlieghe et al., 2010). As discussed in *Section 1.1.2*, these non-natural peptides may exhibit improved stability, bioavailability and perhaps novel function. Orthogonal peptide backbones may also be used to produce protease-resistant products. Peptoids are perhaps the best described example of this, with the side-chains coupled to the nitrogen atom of the peptide backbone instead of the usual α -carbon (Simon et al., 1992).

Another advantage of chemical synthesis over other production routes is that quality assurance and regulatory compliance are simpler to achieve (Lax, 2010). Native or recombinant approaches involve purification from a heterogeneous mix that may include potentially toxic compounds that require removal (i.e. endotoxins in bacterial recombinant systems), whereas chemical synthesis starts clean from a set of known compounds. While removal of endotoxins and other contaminants is feasible, it does add to the production costs associated with native and recombinant peptide purification.

Because of the mostly generic manner in which SPPS can occur, there are now a number of specialised contract manufacturers to which chemical synthesis of a lead bioactive peptide can be outsourced, freeing up researchers to concentrate on optimisation instead (Pichereau & Allary, 2006). The amounts of peptide synthesised can vary from anywhere in the mg range, typical for laboratory work, to ~100 kg/year for drugs that are exceptionally popular (Lax, 2010). The cost of synthesis is also lower in comparison to recombinant production (Vlieghe et al., 2010). In comparison to a 500 Da small molecule drug, however, the production costs of a 5,000 Da peptide will be approximately 10-fold higher (Bray, 2003).

1.1.4.3 Biosynthetic approaches

A biosynthetic approach to peptide production involves harnessing recombinant DNA technology (i.e. the coding sequence for a particular peptide) and inserting this into a production host, such as bacteria, yeast, insect cells, mammalian cells, plants or transgenic animals (Structural Genomics Consortium et al., 2008). Such systems

have been used extensively in the past thirty years to produce industrially or medically useful proteins, including proteases, amylases and lipases for laundry powders, and monoclonal antibodies for therapeutic use. This *in vivo* methodology is also complemented by *in vitro* transcription/translation techniques, which essentially use the required molecular machinery to produce a protein from a recombinant DNA template sans biological host (Katzen et al., 2005). While recombinant production approaches are suitable for proteins (i.e. over 10 kDa), it is important to note that SPPS (Section 1.1.4.2) has considerable advantage over recombinant techniques with regards to short peptides.

Despite this, recombinant techniques are still widely used for some bioactive peptides, such as insulin (51 aa, 5.8 kDa), salmon calcitonin (31 aa, 3.4 kDa, used to treat osteoporosis) and glucagon (29 aa, 3.5 kDa, used to affect a raise in blood glucose levels) (Lax, 2010). Recombinant production is not readily amenable to the incorporation of non-natural amino acids, although some work has been carried out on freeing redundant codons to allow for such a possibility (Young & Schultz, 2010). Still, peptides that require glycosylation, phosphorylation or other post-translational modifications to create their active form are more suited to recombinant production. Given a suitable choice of production host, recombinant approaches can achieve such modifications in a more cost-effective manner than synthetic production (Vlieghe et al., 2010).

While SPPS may be preferred for industrial production of bioactive peptides, recombinant techniques are of importance to this subject in two ways (Raventós et al., 2005). Firstly, recombinant production may be harnessed to produce a bioactive peptide and its variants on a small scale and in a short timeframe using common molecular biology resources. Secondly, and most importantly, the *in vivo* production of a peptide can potentially be coupled with a screen for a desired bioactivity in a high-throughput manner. Such concurrent production and screening may lead to the identification of novel bioactive peptides.

1.2 Screening for bioactivity

1.2.1 Brief history

The field of screening for bioactive peptides, interwoven with the rise of molecular biology over the past 40 years, can trace its roots back to the beginnings of drug discovery research in the 19th century (Sneader, 2005). During the industrial revolution, by-products such as coal-tar (rich in aromatic and aliphatic compounds) were studied as a source of useful products such as dyes. The specificity of some dyes for certain biological tissues was soon noted, with Paul Ehrlich postulating that these compounds bound cell-specific chemoreceptors and that this could form the basis of selective therapy (Bosch & Rosich, 2008). Research proceeded in a disparate manner, however, as university laboratories, pharmacies and industrial dye companies did not individually possess the required knowledge or skill set for drug development. As pharmacology became a distinct field, new ways of finding, characterising and developing drugs led the new industry of drug discovery.

Moving forward, serendipity has played a large part in the discovery of new biological drugs (Kubinyi, 1999), the most famous example being Alexander Fleming's discovery of penicillin after observing inhibition of a *Staphylococcus* culture's growth on agar plates contaminated with a white mould (Fleming, 1929). Subsequent work on this secondary metabolite, and the microbiological and chemical engineering advances required for mass production during World War II, drove the screening of other microbes to see if they too produced antibiotics or other therapeutically valuable compounds (Li & Vederas, 2009). This provided one of the first instances of bioprospecting (*Section 1.2.2*). Work proceeded on a case-by-case basis, with drugs like ivermectin (an antiparasitic to treat nematode worm infection) and cyclosporine (an immunosuppressant drug used in organ transplants) being identified (Drews, 2000). As research into the modes of action of successful drugs increased, biochemistry reinforced the notion that enzymes and cell signal receptors (which Ehrlich had postulated) were good drug targets, as well as elucidating more examples of them for target consideration (Sneader, 2005). Following this, the core scaffolds of drug compounds began to be produced synthetically and were subsequently derivatised by chemists to improve their activity, specificity and bioavailability (Kubinyi, 1999; Schreiber, 2000).

This core base of drug discovery knowledge led to an increased collaboration between chemists and biologists with an emphasis on understanding the relationship between the structure of a drug and its biological function. In turn, this led to the design and implementation of screens to identify new lead compounds of certain function from combinatorial libraries (Schreiber, 2000). These libraries consisted of a large number of different compounds in which the chemical composition or structure was subtly altered from a known natural compound. More random chemical libraries were also trialled to explore a larger chemical space (Bleicher, 2003).

1.2.1.1 High-throughput screening

The advent of high-throughput screening – cell-based or *in vitro* roboticised screens of combinatorial libraries on the microlitre scale – was predicted to revolutionise drug discovery, be it small molecule or peptide based (Drews, 2000). Initial compound “hits” would translate through into “leads” that exhibited the desired effect in more complicated models, such as in animal trials (Bleicher, 2003). Modification of this lead could in turn allow for further exploration of the sequence space and an increase in the number of novel bioactive compounds. High-throughput screening, in combination with combinatorial compound libraries, led to a huge increase in data collection by pharmaceutical firms – from roughly 200,000 discrete data points for a typical large trial in the early 1990s, to 50 million by the end of the decade (Drews, 2000). However, it soon became apparent that this large increase in numbers still needed to be matched by critical scientific analysis (Kubinyi, 2003). Many hits failed to translate into leads. Designs of combinatorial libraries focused on generating as much structural diversity as possible to explore the theoretical chemical space (Bleicher, 2003), but led to the generation of compounds that, while soluble and active in the solvents or dilute form tested during high-throughput screening, were too lipophilic or large to have any efficacy when subjected to a more intensive clinical test (Kubinyi, 2003). While some commentators heralded the lack of successful new drugs as a refutation of high-throughput screening (Bleicher, 2003), it merely shows that inappropriate questions give meaningless answers (Drews, 2000). This is not only with reference to the poor choice of compounds tested, but also to the design of the screen itself. Screens

should be designed to be as close to physiologically relevant as possible, so that potential subtleties of drug action are not missed.

1.2.1.2 *In silico* screening

To rectify the initial problems with high-throughput screening, *in silico* methods have been employed to pre-screen members of a combinatorial library before synthesis (McInnes, 2007). Virtual screening is used to avoid structures that may have undesirable properties (such as solubility issues), and to focus on compounds that are predicted to have tight binding affinities for a particular target (Shoichet, 2004). In this manner a compound library may be targeted, and certain “privileged” structures enriched, i.e. those commonly found experimentally to interact with the target class in question (Bleicher, 2003). Problems with *in silico* screening are still common: an inability to estimate the strength of hydrogen bonds in different solvent conditions; the effect of freezing conformation/degrees of freedom of parts of a compound; and the effects of water molecules at hydrophobic residues can all lead to problems when the compound is tested *in vivo* (Kubinyi, 1999). Another application of *in silico* screening is to try to find potential structures that may fit a target’s binding site (usually modelled from nuclear magnetic resonance spectroscopy or x-ray crystallography data), and thus guide what a combinatorial library should look like when testing experimentally (McInnes, 2007). This approach still exhibits problems, as computer models usually accept compound structures as imported by the user. Thus the effect of compound and binding site flexibility may be dismissed, hydrogen bond donors and acceptors may be incorrectly assigned, and so poor *in silico* docking of a compound to the defined three-dimensional binding pocket may erroneously result (or *vice versa*) (Kubinyi, 2003). *In silico* screening of compounds and their potential interactions should therefore be used with a knowledge of its limitations. A “black-box” approach where compound structures are fed in, and a program predicts their potential as a drug, should only inform decisions along with other *a priori* knowledge about the potential target.

1.2.1.3 Relevance to bioactive peptides

The lessons learnt from early approaches to combinatorial chemical libraries, high-throughput screening, and *in silico* modelling hold true for bioactive peptides. Nevertheless, screening for novel bioactive peptides started slowly. With the advent

of recombinant protein production in hosts such as *E. coli* in the late 1970s, molecular biology was initially only seen as a way to produce therapeutically desired proteins as opposed to purifying them from a native source (e.g. insulin) (Drews, 2000). But the power of coupling genotype with phenotype – i.e. DNA with its encoded protein product – soon led to recombinant protein libraries being screened for activity in their own right. This coupling of the nucleic acid input with a peptide output is an important strength. It enables high-throughput screening of libraries where, if a hit is found, the nucleotides coding for it are readily purified, and the active peptide composition hence deduced. The most well known example of this is phage display (Smith, 1985), where large peptide libraries ($>10^9$ entities) are screened for binding affinity to a target molecule (reviewed later in *Section 1.2.3.1*). Identification of new peptides, as well as optimisation of current peptides, is possible using the same technique. While interest in bioactive peptides from pharmaceutical companies waned in the 1990s as issues with delivery, bioavailability and stability became apparent (see *Section 1.1.2*), peptides as therapeutics are now making a comeback (see *Section 1.1.3.2*). As with combinatorial chemistry, random or rationally designed peptide libraries can be employed, but the technique of bioprospecting may also be used.

1.2.2 Bioprospecting

The term “bioprospecting” refers to the search for industrially or therapeutically useful compounds from natural sources in a systematic manner (Macilwain, 1998). At its simplest level, this may involve investigating and isolating the active component(s) of a traditional medicine (usually a plant extract). For example, the isolation of salicin (metabolised to salicylic acid [aspirin] when ingested) from meadowsweet or willow bark, and that of artemisinin (an anti-malarial agent) from the sweet wormwood plant, was achieved in this manner (Awtry & Loscalzo, 2000; Ro et al., 2006). Another approach is to bioprospect with no *a priori* information about potential compound function – this is akin to grinding up an organism, separating its constituent compounds as much as possible, and individually (or via compound pools) testing for the desired activity (Macilwain, 1998). One such example of this approach is the discovery of halichondrin B, a compound isolated from the sea sponge *Halichondria okadai* that showed potent anti-cancer activity when screened against murine tumour cells *in vitro* and also *in vivo* (Hirata &

Uemura, 1986). Unfortunately the yield of halichondrin B from this sponge is low at 21 mg/tonne of animal. The yield of the same compound from another sponge species (*Lissodendoryx*) has been reported at 300 mg/metric ton of animal, but given that tens of grams of the compound are needed for clinical trials, and that the entire natural population of this sponge is estimated at 280 tons, progress has been hampered (Vogel, 2008).

Because of such environmental and economic concerns, the biological source of an active compound and its actual production are usually separated where possible. In the above examples, salicylic acid (a relatively simple phenolic acid) was chemically-synthesised by Bayer AG in 1897 (Awtry & Loscalzo, 2000). Artemisinic acid (a precursor to artemisinin) has been produced biosynthetically in yeast by introducing two sweet wormwood (*Artemisia annua*) enzymes (amorphaadiene synthase and a cytochrome P450 monooxygenase) into the precursor mevalonate biosynthesis pathway (Ro et al., 2006). Halichondrin B (a large polyether macrolide, 1.1 kDa) was difficult to synthesise chemically, but eventually led to the commercialisation of the analogue eribulin (trade name Halaven; Eisai Inc.) (Towle et al., 2001). Artificial synthesis (either chemically or biosynthetically) is therefore usually desirable as the native host may not produce sufficient amounts of the natural compound, or even if it does, it may not be feasible to cultivate the organism commercially (Li & Vederas, 2009).

1.2.2.1 Screening for bioactive peptides

Bioprospecting for novel bioactive peptides can proceed as described in *Section 1.2.1* and *1.2.2*, instead focusing on proteins and their peptide truncations. Furthermore, given that there is now a large knowledgebase concerning peptide synthesis and optimisation (see *Section 1.1.4*), final production problems are less of an issue than with complex small molecules (such as artemisinin and halichondrin B outlined in *Section 1.2.2*).

The natural world, derived by millions of years of evolutionary selection pressure, provides the biggest rationally designed protein library known to man. This is the key advantage that bioprospecting for novel bioactive peptides offers. As noted earlier, bioactive peptides may be suitable candidates to address non-classical

targets such as protein-protein interactions, where small molecules are largely ineffective due to the thermodynamic barriers associated with binding “flat” surfaces (Watt, 2006). Successful examples of such peptides include interfacial peptides that inhibit dimerisation of human immunodeficiency virus 1 integrase (Zhao et al., 2003) and others that dissociate the same virus’s protease (Park & Raines, 2000). However, rather than screening peptides derived from other known interacting domains (as per the first example), the second example involved screening random peptides of 9 residues in length *in vivo* attached to a thioredoxin domain, giving a hit rate of 1 per million (Park & Raines, 2000).

It is worth noting that naturally occurring peptides, rather than truly random ones, may have a higher hit rate due to residue combinations that are more physiologically relevant, i.e. selected for during evolution. This is referred to as the “stacked” deck of nature as opposed to the “straight” deck of random libraries (Watt, 2006). For example, Phylogica Ltd. specialises in bioprospecting from such naturally-occurring peptides, and claims a 10-fold increased hit rate cf. Park & Raines (2000) above, e.g. in blocking the MyD88 adaptor-like protein which is involved in Toll-like receptor 2 signalling mediated inflammation (Watt, 2009).

1.2.2.2 Advantages of bioprospecting

Assuming that screens are suitably designed to have a negligible false-positive rate, what rationalises a 10-fold increase of natural versus random peptide hits? Random peptides may not be able to adopt stable secondary structures, which are frequently needed for a suitable interaction, if not just biostability. In contrast, peptides derived from existing proteins may form defined motifs, or “folds” (Watt, 2006). In this context it is also worth noting that divergent primary sequences can still lead to a similar secondary structure, e.g. amphipathicity, which may be more important for biological recognition of particular sizes, shapes and charges. This reduces the true structural diversity of random peptides with regards to pure numerical diversity. Following this, it has been predicted that the number of naturally occurring protein structural motifs (i.e. independent of primary residue sequence) is between 1,000 and 10,000 (Wolf, 2000; Watt, 2006). Such motifs typically span a length of 15 to 30 residues (Riechmann & Winter, 2000), thus fitting into the peptide category.

Furthermore, the overall structural diversity exhibited by peptides in isolation is greater than that from scaffolded entities (which are used in their entirety for therapeutic treatment), where only a small portion of the total structure is different between library members (*Figure 2.1*). Examples of commonly used scaffolds include antibodies, designed ankyrin repeat proteins and anticalins (Binz et al., 2005).

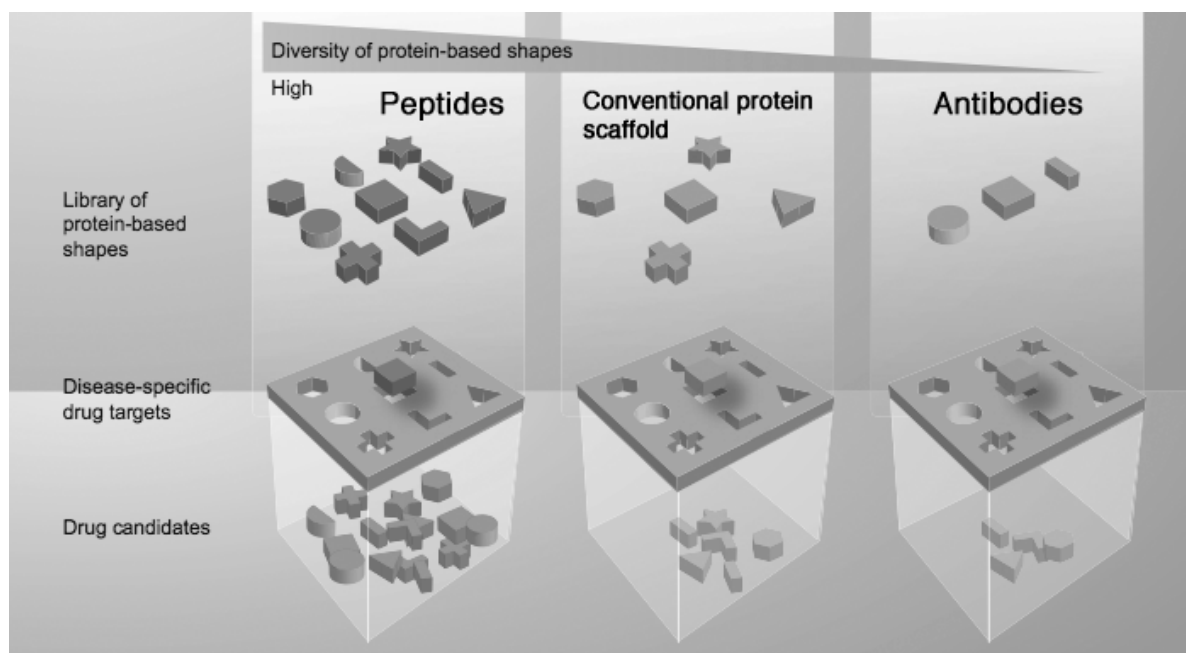


Figure 2.1: Structural diversity increases from antibodies (right) through to peptides (left). Antibodies exhibit diversity at their complementary determining regions, while smaller scaffolds such as designed ankyrin repeat proteins or anticalins are less sterically hindered by their constant regions. Peptides, which may form their own motifs, exhibit the highest diversity, and thus represent a good source of potential bioactive compounds. Figure adapted from Watt (2009).

Peptides derived from bioprospecting, when used as therapeutics, are typically utilised out of context (Watt, 2006). Obviously, depending on the source organism, the majority of naturally occurring peptides are segregated in different biological compartments or habitats from the therapeutic target organism. For instance, it is hard to think of an occasion when peptides produced by an organism dwelling in a deep-sea vent would naturally come into contact with a human cancer. In addition, some peptide motifs may perhaps not be fully accessible for interaction in their native protein form (e.g. hidden in protein folds). Because of this segregation, novel

peptide-protein interactions may be found without regard for the original niche or function of a peptide (Watt, 2006).

1.2.2.3 Use of DNA as an encoding source for peptides

As discussed at the beginning of *Section 1.2.2*, the inability to harvest large amounts of protein/peptide from an organism during bioprospecting should not become an overall limiting factor, due to existing synthetic techniques of production (see *Section 1.1.4*). Problems with bioprospecting remain, however. Not all proteins will be expressed at all times in an organism, and given that a number of therapeutically valuable products may only be induced under particular stimuli (e.g. stress, the presence of pathogens), the harvested proteome of any given organism may be incomplete. Furthermore, while macroscopic organisms may be relatively easy to obtain in bulk, some microorganisms may not be sampled in significant quantities, and therefore not enough active compound will be present to be detected in a screen. There are probably millions of unidentified bacterial, fungal and protozoan species, but most of these are under-represented in environmental samples due to a small number of species dominating the biosphere (Sogin et al., 2006). If the proteomes of species could be accessed without regard to their environmental abundance, an enormous universe of protein and peptide variation would become available. A method of providing such access involves the isolation of genomic DNA.

Genomic DNA libraries, representing the nucleic acid blueprint of an organism, are a rich source of input for a recombinant bioactive peptide screen. Such DNA is easily purified from environmental samples, whether it is a jar of water containing microbes, a tissue sample or a single animal. Furthermore, genomic DNA will encode peptide motifs that have been sculpted by evolution, and are consequently biologically relevant (see *Section 1.2.2.2*). However, there is still an element of randomness in screening a genomic library. The source DNA is commonly sheared into fragments that are of peptide-encoding size, e.g. 45 to 300 bp to encode peptides of 15 to 100 amino acids, and inserted into a recombinant expression system. Usually the orientation of this insertion is uncontrolled, resulting in six possible reading frames, i.e. three forward and three reverse. In-frame stop codons may also occur. Even if a piece of DNA does code for a bioactive peptide, it is not

guaranteed that it will be expressed in its correct reading frame in the screen. However, when Watt (2009) screened a genomic library for peptides that bound the MyD88 adaptor-like protein (see *Section 1.2.2.1*), it was found that while only approximately 14% of the library was inserted in its natural reading frame (when compared to its annotated genomic sequence), ~41% of peptide hits were from this subsection. While this shows that naturally-encoded peptides from DNA libraries may be enriched in function, it also indicates that random, non-naturally occurring peptides from the same DNA template may also exhibit activity. Being unable to control DNA library insertion orientation or reading frame therefore combines the advantages of screening “evolved”, naturally occurring sequences with a more random selection of sequences, i.e. those from “unnatural” reading frames.

The use of complementary DNA libraries may further enrich for peptide fragments that are produced by the organism *in vivo*. This DNA, which is derived from an organism’s messenger RNA (mRNA) transcripts, represents actively-transcribed parts of the genome (i.e. its transcriptome). Furthermore, cis (or even trans) splicing of such transcripts is possible in eukaryotes, revealing open reading frames that are not discernible from genomic DNA alone (Di Segni et al., 2008). While a large amount of this DNA will go on to be translated into protein *in vivo*, the same issues regarding the lack of control of orientation and reading frame when inserted into a recombinant expression screen still apply. In addition, not every mRNA transcript may be present at the time of library generation, as many proteins are not constitutively expressed but rather generated in response to environmental signals.

In addition to using genomic or complementary DNA as an input source for recombinant peptide production, random synthetic DNA may also be used. Random single-stranded oligonucleotides may be obtained commercially, with defined 5’ and 3’ terminal sequences so that they may be amplified into double-stranded DNA via polymerase chain reaction (PCR) techniques. A defined length of oligonucleotide allows for potential peptide length to be controlled. Discounting internal stop codons, this is typically up to 25 amino acids, depending on the maximum oligonucleotide length capable of being reliably synthesised. Furthermore, this approach is inexpensive, costing only tens of pence per base from a commercial synthesis company (see *Section 2.1.2*).

1.2.3 Biosynthetic approaches

Once a nucleic acid source is identified, the question becomes how to express it as a peptide for screening. As outlined earlier in *Section 1.2.1.3*, a key requirement in using a DNA source to power a recombinant screen for bioactive peptides is that genotype and phenotype are linked. Each member of the peptide library must be discretely coupled to the DNA that coded for it, so that if a hit in the screen is observed, the active amino acid residue sequence can be inferred. In addition, the ease of DNA manipulation makes it straightforward to create mutants of any hit peptide for further structure/function studies.

1.2.3.1 Recombinant peptide screens

Phage display is one of the better-known techniques for screening, enriching and identifying bioactive peptides (Smith, 1985). The DNA library is inserted into a phage vector in-frame with the coding region for a capsid protein (typically pIII if the filamentous fd or M13 bacteriophage is used), so that when expressed in *E. coli* the corresponding peptide is fused to the capsid protein. When new phage are produced, they are assembled with the library-encoded peptide displayed on the outer tip of the virus. The DNA library plasmid also contains the necessary elements to ensure that it too is packaged in the new phage during virion assembly, thus ensuring that genotype and phenotype are physically linked – one phage, one peptide displayed, one genetic sequence coding for it. This library of phage is then washed over an immobilised target molecule. Those displaying a peptide that interacts are retained, and can then be recycled to infect fresh bacteria to multiply and enrich further for rare members with a high binding affinity to the target in question. Phage display libraries numbering 10^{11} different peptides have reportedly been produced (McGregor, 2008). This technique was used to discover bioactive peptides such as Amgen Inc.'s romiplostim (trade name Nplate, 41 aa scaffolded to a human IgG F_C region), an agonist of the thrombopoietin receptor to promote blood platelet production, and Affymax Inc.'s peginesatide (trade name Hematide, 14 aa in a PEGylated and cyclic configuration, under clinical development [Wrighton et al., 1996]), a mimetic of erythropoietin to treat anaemia.

Another popular recombinant screen is the two-hybrid approach. Initially developed in the yeast *Saccharomyces cerevisiae*, two functionally separate domains of a

reporter protein are taken (e.g. the DNA binding domain and transcription activation region of the GAL4 transcription factor) and fused to a “bait” and “prey” protein respectively (Fields & Song, 1989). If the bait and prey interact, the biological circuit is complete and an output occurs. In the GAL4 system, the transcription/translation of a reporter gene such as β -galactosidase is used, which converts exogenous X-gal from a colourless compound into a blue product. Blue yeast colonies therefore indicate a successful interaction. A similar system utilising part of a native RNA polymerase has been used in *E. coli* (Joung et al., 2000). For bioactive peptide screening, the prey protein gene may be replaced by a DNA library to find potential peptides that interact with a defined bait protein; or an existing protein-protein (i.e. bait-prey) interaction may be targeted for disruption by an additionally expressed peptide library. Peptides found to bind the MyD88 adaptor-like protein (see *Section 1.2.2.3*) were identified using such a two-hybrid approach (Watt, 2009). However, false positive rates are reportedly high, presumably due to non-specific bait-prey interactions (Brückner et al., 2009).

Other recombinant screening techniques include mRNA display (Roberts & Szostak, 1997), ribosome display (Zahnd et al., 2007) and CIS display (Odegrip et al., 2004). All of these techniques involve *in vitro* transcription and translation as opposed to phage display. In mRNA display, a DNA library is used to generate an mRNA library (no stop codons included), which subsequently has a puromycin moiety (a mimic of tyrosyl-tRNA) ligated to its 3' end. *In vitro* translation is then conducted, and once the ribosome reaches the incorporated puromycin, translation is disrupted (puromycin serves as an antibiotic in this manner). Crucially, the puromycin is ligated by the ribosome to the C-terminus of the peptide, thus physically coupling genotype with phenotype. Ribosome display proceeds in a similar manner, except that a spacer sequence is used instead of puromycin at the 3' end of the DNA library, and a lack of stop codon in the resulting transcript causes the ribosome to stall on this spacer. However, the peptide has enough “chaser” sequence to ensure it is properly displayed rather than being masked inside the ribosome exit tunnel. Finally, CIS display (CIS being derived from cis-acting) couples the peptide directly to its library DNA template via fusing the coding sequence upstream to that of the bacterial plasmid protein RepA, which when translated recognises and binds a particular sequence also included on the DNA

template. If transcription and translation are coupled *in vitro*, the nascent RepA protein is able to perform this function while the DNA template, RNA polymerase, mRNA and ribosome are in a complex; thus each library peptide is directly linked to its DNA template. An alternative library DNA/corresponding peptide association can also be achieved through *in vitro* compartmentalisation techniques such as water/oil droplet emulsions that encapsulate single library DNA members (Miller et al., 2006). Example uses of these *in vitro* approaches include screening for single-chain antibodies with improved target binding affinities (Hanes & Plückthun, 1997; Fukuda et al., 2006), as well as finding peptide ligands that bind enzymes such as lysozyme (Odegrip et al., 2004).

Peptides produced by all of these methods may then be screened for specific binding or function, and their genotype recovered for further analysis or subsequent enrichment (i.e. screening again). It is possible to use such methods to isolate peptides with very low dissociation constants (<10 nM) to a particular target (Sato et al., 2006a).

1.2.3.2 Advantages and disadvantages of recombinant peptide screens

There are general problems to consider related to transcribing and translating peptide libraries: sequence length, promoter efficiency, codon bias, product solubility and activity in differing physiological conditions are all contributing factors (Ingham & Moore, 2007). The choice of expression system can help with this, as well as the use of larger protein scaffolds or tags in order to aid stability, solubility or downstream purification of a peptide. In addition, such a single-peptide approach necessarily limits the complexity of what may be identified. Any peptide requiring co-factors for activity (unless serendipitously provided by the expression system) or post-translation modification will be missed, as well as those that may show synergistic action with other peptides. Along with the limitations of uncontrolled insertion orientation and reading frame, this leads to a narrowed window of discovery when using DNA libraries to screen for bioactive peptides. Despite this, such an approach is still capable of finding useful peptides, as previous examples cited in *Section 1.2.3.1* have illustrated. Screens will inherently ignore peptides that are poorly expressed or insoluble, which serves to remove entities unsuitable for further development. Any peptide that is identified as a hit should therefore be

physiochemically suitable for future production. It is also worth noting that discovery of even a single peptide with limited function may be an important first step in elucidating little-known or new pathways to do with the activity being assayed for.

1.2.3.3 Screen considerations

Because of a peptide's limited bioavailability and stability (at least before optimisation; see *Section 1.1.2*), it may be more worthwhile pursuing peptide agonists than antagonists. An agonist of a particular receptor is typically only required at low concentration for activation, and a short half-life is not problematic due to the ensuing signal cascade that such an activation typically evokes (Lien & Lowman, 2003). In contrast, an antagonist is competing with the native ligand for its binding site, and thus requires higher concentrations and binding site residence time for efficacy. In general, an antagonist must occupy over 50% of the receptor population to have the desired effect, in comparison to 5-20% occupation for an agonist (Vlieghe et al., 2010). Despite this knowledge, many screens are targeted towards identifying antagonists rather than agonists because screening for inhibitors of natural ligand binding is more straightforward than the functional screens that agonists require (Lien & Lowman, 2003). Common antagonist targets are specific receptors or enzymes, which have ligand binding constants measured *in vitro* by utilising techniques such as surface-plasmon resonance or isothermal titration calorimetry (Rich & Myszka, 2000; Pierce et al., 1999). In contrast, finding peptides that provoke a desired response through other modes of action, like disrupting protein-protein interactions or allosteric activation/inhibition, require more complicated, whole-cell functional screens. This dissertation is concerned with whole-cell approaches.

Whole-cell phenotypic screening is a more holistic way of looking for potential bioactive peptides. Much like classical forward genetic screens, such an approach looks for the desired phenotype/output and then studies the agent responsible (Watt, 2009). For recombinantly-produced bioactive peptides, where the input genotype is linked to an output phenotype in a discrete manner, identification of the casual agent is straightforward. If the interaction partner(s) of an active peptide is unknown, modern proteomic techniques such as chemical cross-linking, pull-down assays and mass spectrometry can be used to aid identification (Puig et al., 2001).

A screen relying on positive selection, i.e. an output is given or its level is increased, is more desirable than negative selection, i.e. a decrease in an output from an observed baseline, as this allows for rare peptides to be more easily identified from large pools. For instance, if the desired activity in an *in vivo* screen is coupled to an output of antibiotic resistance, organisms that can grow when challenged by this antibiotic represent hits – those that are negative fail to appear. Coupling the desired peptide activity to a selectable phenotype in this manner gives a screen enormous resolution power (cf. phage display in *Section 1.2.3.1*). In contrast, locating and isolating clones that fail to grow is more difficult. Such a positive selection approach may be required if the hit rate is low (i.e. less than one per million) (Watt, 2009).

1.3 Proof of principle

Potential screens for bioactive peptides are only limited by the ingenuity of the researcher and the resources available. In this dissertation, whole-cell *in vivo* assays are studied and employed to find novel bioactive peptides. In particular, peptides displaying antimicrobial or antiaggregation activity were sought, principally as a proof of principle to demonstrate the feasibility of the strategy adopted.

1.3.1 Antimicrobial peptides

The use of peptides as antimicrobials is a special focus of this research. Found in most species of life, antimicrobial peptides serve as part of the first line of defence against pathogens (Zasloff, 2002). Rather than acting as an agonist/antagonist against a specific target, they can exert their effect directly via lysis of microbial membranes, although also through other immunomodulatory functions (Bowdish et al., 2005a). Their broad-spectrum activity against non-proteinaceous targets suggests a low probability of target pathogens developing resistance (Peschel & Sahl, 2006). This makes them attractive candidates to treat “superbugs” such as methicillin-resistant *Staphylococcus aureus* and vancomycin-resistant *Enterococcus faecalis*. The properties of antimicrobial peptides and their applicability to a biosynthetic screen will be discussed in detail in *Chapter 3* and *Chapter 4*.

1.3.2 Antiaggregation peptides

Another potential application of bioactive peptides may be as therapeutic antiaggregants. Inappropriate aggregation and subsequent amyloid formation (i.e. insoluble protein fibres) are characteristic of a number of human neuropathies (Treusch et al., 2009). In Alzheimer's disease, the amyloid β peptide, which is normally 40 residues in length, undergoes alternative cleavage from its amyloid precursor protein leading to two additional residues being retained at its C-terminus (Walsh & Selkoe, 2007). This mutant isoform is especially prone to aggregating via the formation of anti-parallel β -sheets (Hilbich et al., 1991). In Huntington's disease, triplet repeat mutations in exon I of the huntingtin gene leads to an extended polyglutamine tract being coded for with 42-66 residues instead of the typical 11-35 (Huntington's Disease Collaborative Research Group, 1993). This extended "sticky" repeat leads to fragments of the large huntingtin protein (348 kDa) forming aggregates with itself or other proteins (Cattaneo et al., 2005). For both amyloid β and huntingtin, other peptides may be able to prevent their aggregation and potentially alleviate the disease state. For example, both the 15-residue Peptide 2 (Baine et al., 2009) and 11-residue polyglutamine binding protein 1 (Nagai et al., 2000) have been shown to inhibit aggregation of their target proteins *in vitro*. A screen to identify peptides able to interfere with amyloid β aggregation will be discussed in detail in *Chapter 5*.

1.3.3 Summary and aims

Peptides and proteins are exquisitely tuned by evolution to perform a huge range of biological functions. Peptides, which consist of 100 amino acid residues or fewer, are often employed *in vivo* as cell signalling molecules and hence are implicated in a number of disease settings. After falling out of favour due to issues with delivery, stability and bioavailability, peptides are now being re-examined with increased interest by biotechnology and pharmaceutical companies. Some economically valuable peptide drugs now exist. Furthermore, it is clear that the potential uses for peptides can expand past traditional cell signal membrane receptors to more diverse bioactivities, such as antimicrobials and antiaggregants. The extraordinary diversity of organisms in the natural world represents a vast pool of different proteins, all of which have been optimised via evolutionary pressures to exhibit

particular activities. Tapping this diversity by using the DNA that encodes it to produce potentially bioactive peptides for screening may lead to the identification of bioactive peptides with novel therapeutic functions.

The overall aim of this project is to explore the utility of *in vivo* systems to produce, screen and identify existing and novel bioactive peptides, focusing on antimicrobial and antiaggregation activities. For the first of these, the ability of *E. coli* to produce antimicrobial peptides will be assessed, with the aim of generating a recombinant system that can produce antimicrobials for further characterisation. The knowledge gained will then be used to design an *in vivo* screen of DNA libraries for novel peptides exhibiting antimicrobial function. For the second activity, another *in vivo* screen will be designed and tested in order to find peptides that possess antiaggregation activity against the amyloid β peptide implicated in Alzheimer's disease.

CHAPTER 2 – MATERIALS AND METHODS

2.1 Materials

2.1.1 Chemicals

Chemicals were purchased from Bio-Rad Laboratories Ltd. (Hemel Hempstead, UK), Melford Laboratories Ltd. (Ipswich, UK), MP Biomedicals Ltd. (Illkirch, France), Sigma-Aldrich Company Ltd. (Gillingham, UK) or Fisher Scientific UK Ltd. (Loughborough, UK). All other sources are as indicated.

2.1.2 Oligonucleotides

Oligonucleotides (i.e. primers) were obtained from Eurofins MWG Operon GmbH (Ebersberg, Germany) at the standard, salt-free purity. Long oligonucleotides (see *Section 2.7.3*) were obtained from Sigma-Aldrich Company Ltd. at a high purity (PAGE-extracted). When used for constructing genes, all oligonucleotide sequences were codon-optimised for use in *Escherichia coli* as per Nakamura et al. (2000).

2.1.3 Bacterial expression strains and plasmids

E. coli TOP10 chemically-competent cells were obtained from Invitrogen, a division of Life Technologies Ltd. (Paisley, UK), while DH10B chemically-competent cells were obtained from New England Biolabs (UK) Ltd. (Hitchin, UK). TOP10 and DH10B share the same genotype, being *araBAD^C*, i.e. unable to metabolise the pBAD induction agent L-arabinose. *E. coli* Novagen BL21(DE3) Star and BL21(DE3)pLysS Rosetta cells were obtained from Merck KGaA (Darmstadt, Germany). *Vibrio cholerae* El Tor N16961 was kindly provided by Dr Gillian Fraser (Department of Pathology, University of Cambridge). Plasmid pET28a(+) was obtained from Merck KGaA; plasmids pBAD/gIII-A, pBAD/gIII-Calmodulin and pUC19 were Invitrogen-brand and obtained from Life Technologies Ltd.; plasmid pCC1FOS from Epicentre Biotechnologies Inc. (Madison, USA); plasmid pKG1010 (Pecota et al., 2003) was kindly provided by Prof Kenn Gerdes (Institute for Cell and Molecular Biosciences, Newcastle University); plasmid pA β 42-EGFP (Baine et al., 2009) was kindly provided by Prof David Moffet (Department of Chemistry and Biochemistry, Loyola Marymount University, USA); plasmid pSAlect-GM6 (Fisher et al., 2006) was kindly provided by Prof Matthew DeLisa (School of Chemical and

Biomolecular Engineering, Cornell University); plasmids pSB3K3-mCherryGFP1plus3, pET28-A7-EGFP and pET15-AavLEA1 were previously made in the Tunnacliffe laboratory. Cultures of bacterial strains, and of those containing the various plasmids, were stored at -80°C as glycerol stocks (15%) (Sambrook & Russell, 2001).

2.1.4 Chemically-synthesised peptides and cell-permeabilizing reagent

Table 2.1 lists the AMP-related peptides purchased during this work. Peptides came lyophilised in 1 mg aliquots. K2C18 and S-H4 were resuspended in sterile water; S-H1, S-H5, H4 and R13 in neat DMSO to aid solubility. Polymyxin B nonapeptide was purchased from Sigma-Aldrich Company Ltd. (catalogue #P2076).

| Name | Peptide (# aa) | MW | Hydrophobic | Purity |
|-------|--------------------------------------|--------|-------------|--------|
| K2C18 | GKKLKKIGQKIKNFFQKL (18) | 2146.6 | 33% | 88.9% |
| S-H1 | TMSRNAVESGRESNGEADGMQW (22) | 2412.5 | 18% | 75.6% |
| S-H4 | KRYLGTRCLSLMWAKGDLRQIKMGKRGDKKE (31) | 3667.4 | 29% | 91.8% |
| S-H5 | TMSFSKFLFLLHFIHLIFIH (20) | 2492.0 | 65% | 86.7% |
| H4 | MTMSTWPILAAALGFRTVFIFFGEEVM (26) | 2995.6 | 54% | 86.4% |
| R13 | MTMSILQILYRVIFILNVIVVHIHYV (26) | 3128.9 | 73% | 79.8% |

Table 2.1: Chemically-synthesised peptides purchased for this work. K2C18 purchased from Proimmune Ltd. (Oxford, UK); rest from GenScript USA Inc. (Piscataway, USA). Hydrophobic, % of residues (F, I, L, N, V, W & Y); Purity, as measured by HPLC analysis (data not shown).

2.2 General cloning methods

Sambrook & Russell (2001) comprehensively outline the following molecular biology techniques.

2.2.1 Common enzymatic manipulations of DNA

To site-specifically digest DNA, restriction enzymes from New England Biolabs (UK) Ltd. or Fermentas GmbH (St. Leon-Rot, Germany) were used. To dephosphorylate the 5' ends of linear DNA in order to prevent self-ligation during ligase treatment, calf intestinal alkaline phosphatase or Antarctic phosphatase from New England Biolabs (UK) Ltd. was used, or alternatively FastAP thermosensitive alkaline

phosphatase from Fermentas GmbH. To phosphorylate the 5' ends of linear DNA to allow them to ligate during ligase treatment, T4 polynucleotide kinase from New England Biolabs (UK) Ltd. or Fermentas GmbH was used. To ligate ends of linear DNA together, T4 DNA ligase or a Quick Ligation kit from New England Biolabs (UK) Ltd. was used, or alternatively T4 DNA ligase from Fermentas GmbH. All enzymes were used according to the manufacturer's instructions, with any modification indicated.

2.2.2 Purification of DNA

For plasmid purification, a QIAprep Spin Miniprep kit was used. For PCR product purification, a QIAquick PCR Purification kit was used, or alternatively a QIAquick Nucleotide Removal kit as indicated. For agarose gel extraction of DNA, a QIAquick Gel Extraction kit was used, with agarose gels lacking ethidium bromide being immersed for 10 min in a solution of 1 µg/mL ethidium bromide before visualisation at 365 nm (see *Section 2.2.7*). All kits were used according to the manufacturer's instructions (QIAGEN Ltd., Crawley, UK).

2.2.3 Determination of DNA concentration

DNA concentration and purity was determined using a ND-1000 spectrophotometer as per the manufacturer's instructions (NanoDrop Technologies Inc., Wilmington, USA).

2.2.4 DNA sequencing

To verify that a constructed plasmid contained the desired sequence, or to determine the sequence of an unknown insert, 1 µg of plasmid was analysed using appropriate sequencing primers (10 µM; indicated) on a 3730xl DNA Analyser (Applied Biosystems Inc., Foster City, USA) by the DNA Sequencing Facility, Department of Biochemistry, University of Cambridge. Sequence data was analysed using 4Peaks (v1.7.2) and EnzymeX (v3.1) software (Mekentosj BV, Amsterdam, Netherlands; accessed at <http://www.mekentosj.com/>).

2.2.5 Amplification of DNA using polymerase chain reaction (PCR)

2.2.5.1 General PCR

Advantage 2 polymerase mix (Takara Bio Inc., Otsu, Japan), Phusion High-Fidelity DNA polymerase (Finnzymes Oy, Espoo, Finland) or *Taq* DNA polymerase (Roche Diagnostics Ltd., Burgess Hill, UK) was used for PCR reactions (typically 20 μ L) as indicated. PCR reaction compositions and cycling conditions were performed as per polymerase manufacturer's instructions (typically 500 nM primer concentration), with primer-specific annealing temperatures indicated. In addition, touchdown PCR (Don et al., 1991) was sometimes performed over the first 10 cycles as indicated, with the remaining cycles (typically 20) using a constant annealing temperature. A PCR Express (Hybaid Ltd., Ashford, UK) or GS482 (G-Storm Ltd., Somerton, UK) thermocycler was used.

2.2.5.2 Colony PCR of plasmid inserts

To confirm correct plasmid construction, a bacterial colony swab was inoculated directly into 10 μ L of 1x ReddyMix PCR master mix (Fisher Scientific UK Ltd.) containing the appropriate primers (see *Table 2.2*). Primers T7-Fwd and T7-Rev were used to PCR amplify (annealing at 51°C) insert sequences from the multi-cloning site region of pET28-based plasmids. Primers pBADmFWD and pBADgIII-reverse were used to PCR amplify (annealing at 57°C) insert sequences from the *araBAD* multi-cloning site region of pBAD-based plasmids. Primers pAGLib-FWD and pAGLib-REV were used to PCR amplify (annealing at 54°C) insert sequences from the Library cloning site region of pAG2-A β 42-based plasmids.

| Primer name | Sequence (5' to 3') |
|------------------|-----------------------|
| T7-Fwd | TAATACGACTCACTATAGGG |
| T7-Rev | TAGTTATTGCTCAGCGGTGG |
| pBADmFWD | TTAGCGGATCCTACCTGACGC |
| pBADgIII-reverse | TATCAGACCGCTTCTGCGTTC |
| pAGLib-FWD | AGCGGATTTGAACGTTGCG |
| pAGLib-REV | GATGCCTGGCAGTTTATGG |

Table 2.2: Primers used for colony PCR.

2.2.6 Circular polymerase extension cloning (CPEC)

A modified form of CPEC (Quan & Tian, 2009), named CPEC Viaduct, was developed to insert and/or delete DNA (ranging from single bp to kbp) from a circular plasmid, i.e. without requiring restriction enzyme treatment. This approach is analogous to site-directed mutagenesis (Sambrook & Russell, 2001). *Figure 2.1* outlines the approach.

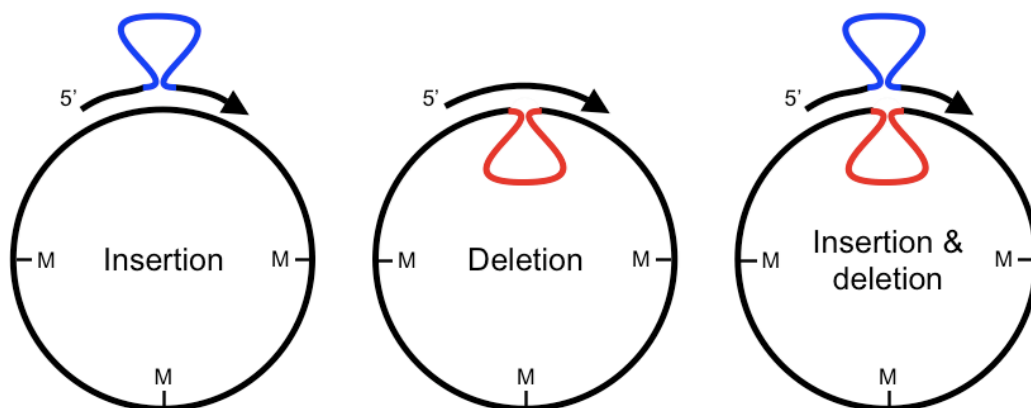


Figure 2.1: CPEC Viaduct may be used to insert and/or delete a sequence (1 bp to kbp) from a plasmid in a site-specific manner. Only one strand of the dsDNA insert/deletion product is shown for clarity. The inserted region is shown in blue, deleted in red. M, methyl groups on template plasmid that allow for selective digestion with DpnI, leading to only *de novo* plasmid being viable.

In brief, the dsDNA sequence that was desired between two regions of a plasmid was constructed (typically by PCR amplification or annealing of complimentary oligonucleotides) so that its 5' and 3' ends contained at least 20 bp of homologous overlap with the upstream and downstream plasmid regions respectively. 100 ng of plasmid was mixed with dsDNA at a vector:dsDNA molar ratio of 1:3 in a 10 μ L Phusion polymerase reaction (see *Section 2.2.5.1*). Typical cycling conditions were as follows: 98°C for 30 s; 3x cycles 98°C for 30 s, 55°C for 60 s, 72°C for 30 s/kbp final plasmid size; hold at 10°C. The reaction was then treated with the methylation-specific *DpnI* (2 U) to digest the template plasmid (*de novo* plasmid produced by CPEC Viaduct contains no GAmTC methyl motifs) at 37°C for 15 min, and an aliquot subsequently transformed into *E. coli*.

2.2.7 Agarose gel electrophoresis

To confirm approximate linear DNA size, electrophoresis of DNA was carried out using agarose gels (0.8 to 1.5%, containing 0.5 µg/mL ethidium bromide) in Tris-acetate-EDTA buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA, pH 8.0). DNA samples were mixed with 5x GelPilot DNA loading dye (QIAGEN Ltd.) to 1x final concentration (typical final sample volume of 10 µL) and electrophoresed at between 60 to 110 V in a Mini-Sub Cell GT (Bio-Rad Laboratories Ltd.). DNA was visualised at 312 nm using a Doc-008.XD transilluminator system (UVItec Ltd., Cambridge, UK).

2.2.8 Transformation of *E. coli*

Aliquots of purchased chemically-competent TOP10 cells and DH10B cells were transformed as per the manufacturer's instructions. For BL21(DE3) strains, chemically-competent cells were prepared using the calcium chloride method (Sambrook & Russell, 2001) and stored at -80°C. Aliquots were thawed on ice and ligation mixture added to a maximum of 1/10th the cell aliquot volume (typically 0.5 to 2 µL). Cells were incubated on ice for 30 min, heat-shocked at 42°C for 30 s, returned to ice for 3 min, then 250 µL of lysogeny broth (LB) medium (171 mM NaCl; Sambrook & Russell [2001]) added and out-growth performed at 37°C, ~225 rpm for 1 h prior to plating aliquots on solid medium.

2.2.9 Growth of *E. coli*

To obtain colonies on solid medium, bacteria were typically grown on LB agar (1.5%) plates at 37°C for ~16 h. For liquid culture, bacteria were typically grown in LB medium at 37°C, ~225 rpm. Carbenicillin (100 µg/mL), kanamycin (50 µg/mL), chloramphenicol (34 µg/mL) or streptomycin (50 µg/mL) was added to the medium as appropriate for particular plasmid selection to give the indicated final concentrations. pBAD-, pKG-, pET15- and pSB3K3-based plasmids encoded carbenicillin resistance; pET28-based plasmids encoded kanamycin resistance; pCDF-based plasmids, i.e. pAβ42-EGFP, encoded streptomycin resistance; and pSAlect-based plasmids encoded chloramphenicol resistance. Chloramphenicol was also required to maintain the pLysS plasmid in the BL21(DE3)pLysS Rosetta cell line.

2.3 General protein methods

2.3.1 Induction of protein expression in *E. coli*

Starter cultures (grown overnight) were used to inoculate fresh medium at a dilution of 1:100. Cultures were grown to mid-log growth phase ($OD_{600} = 0.4$ to 0.6 ; using a path length of 10 mm in a DU 800 spectrophotometer [Beckmann Coulter Inc., Brea, USA]) and induced as appropriate (1 mM IPTG for pET plasmids; 0.05% L-arabinose for pBAD plasmids) at a defined temperature for a set time (indicated).

2.3.2 Extraction of protein from *E. coli*

All steps were carried out at 4°C or on ice where possible. Induced cells were collected by centrifugation at $\sim 13,700$ rcf for 10 min, and the cell pellet resuspended in binding buffer (33 mM NaH_2PO_4 , pH 8.0, 425 mM NaCl, 25 mM imidazole) to $1/17^{\text{th}}$ the original culture volume and subsequently frozen at -80°C . Thawed resuspensions were further lysed by sonication (amplitude 10, 10x cycles of 15 s on, 30 s off, using a Soniprep 150 [SANYO Electric Ltd., Moriguchi, Japan]). Whole cell samples were taken at this point for SDS-PAGE analysis. Insoluble debris was removed by centrifugation at $\sim 20,000$ rcf for 30 min, and the supernatant retained. Clarified lysate samples were taken at this point for SDS-PAGE analysis.

For lysis of cells in pAG2-A β 42 experiments, cell pellets were resuspended in $1/15^{\text{th}}$ culture volume of Genlantis SoluLyse (in 50 mM NaHPO_4 , pH7.4 [Gene Therapy Systems Inc., San Diego, USA]) and nutated at 23°C for 10 min in lieu of sonication.

2.3.3 Polyacrylamide gel electrophoresis (SDS-PAGE)

2.3.3.1 Glycine-based SDS-PAGE

Glycine-based SDS-PAGE (Laemmli, 1970), used to estimate protein size, was carried out in a Mini-PROTEAN 3 Cell (Bio-Rad Laboratories Ltd.) as per the manufacturer's instructions. A 5% stacking gel was used, with resolving gel acrylamide concentration as indicated (typically 12% or 15%). Protein samples were mixed with 5x sample dye (60 mM Tris pH 6.8, 25% glycerol, 10% SDS, 0.1% bromophenol blue, 5% β -mercaptoethanol [freshly added]) to 1x final concentration, boiled in a water bath for 2 min, and allowed to cool prior to gel loading. PageRuler Plus Prestained Protein Ladder, Spectra Multicolor Low Range Protein Ladder

(Fermentas GmbH), Precision Plus Protein Dual Color (Bio-Rad Laboratories Ltd.) or MagicMark XP Western Protein Standard (Life Technologies Ltd.) was used as a molecular mass standard. Gels were typically electrophoresed at 60 V constant during the stacking phase, then 120 V constant during the resolving phase until the sample dye front had reached the bottom of the gel.

2.3.3.2 Tricine-based SDS-PAGE

Tricine-based SDS-PAGE (Schägger, 2006), used to estimate peptide size, was carried out in an Invitrogen XCell SureLock Mini-Cell (Life Technologies Ltd.) as per the manufacturer's instructions. Invitrogen Novex tricine pre-cast gels were obtained from Life Technologies Ltd., using a 4% stacking gel with resolving gel acrylamide concentration as indicated (typically 10-20% gradient). Protein samples were mixed with Invitrogen 2x Novex tricine sample buffer (Life Technologies Ltd.; with 100 mM dithiothreitol [freshly added]) to 1x final concentration, then boiled and loaded as per *Section 2.3.3.1*. Gels were typically electrophoresed at 125 V constant until the sample dye front had reached the bottom.

2.3.3.3 SDS-PAGE visualisation

After electrophoresis, gels were fixed/stained in warm Coomassie stain (25% propan-2-ol, 10% acetic acid, 0.05% Coomassie brilliant blue R-250) for 30 min followed by destaining in destain solution (40% methanol, 10% acetic acid). Gels were subsequently scanned (Epson Perfection 3490 Photo flatbed scanner, Seiko Epson Corp., Nagano, Japan) or photographed (UVIdoc HD2/20LM transilluminator, UVItec Ltd.).

2.3.3.4 Western blotting

After SDS-PAGE (*Section 2.3.3.1*), the gel was washed in water and then equilibrated in transfer buffer (25 mM Tris pH 8.3, 192 mM glycine, 20% methanol) for 10 min. Amersham Hybond-P polyvinylidene fluoride membrane (GE Healthcare UK Ltd., Little Chalfont, UK) was equilibrated in methanol for 1 min, water for 5 min and then transfer buffer for 10 min. Proteins were transferred to the membrane using a Trans-Blot SD Semi-Dry Electrophoretic Transfer Cell (Bio-Rad Laboratories Ltd.) as per the manufacturer's instructions (typically run at 15 V for 30 min). After transfer, the membrane was blocked in 5% non-fat milk powder TBS-T

(Tris 50 mM pH 7.5, NaCl 150 mM, 0.1% Tween 20) for 1 h on a rocker, rinsed briefly with 1% non-fat milk powder TBS-T, and then 3.3 μ L of Novagen Anti-His-Tag monoclonal mouse IgG antibody (Merck KGaA) was added to 5 mL fresh 1% non-fat milk powder TBS-T (1:1,500 dilution). After at least 1 h incubation with the primary antibody, the membrane was washed 5x with 1% non-fat milk powder TBS-T over 30 min. 1.3 μ L of ECL Anti-mouse IgG horseradish-peroxidase-linked sheep antibody (GE Healthcare UK Ltd.) was then added to 5 mL fresh 1% non-fat milk powder TBS-T (1:4,000 dilution) and the membrane incubated for 1 h, followed by 5x washes as above. Signal was visualised using ECL Plus Western Blotting Detection Reagents and Hyperfilm ECL (GE Healthcare UK Ltd.) as per the manufacturer's instructions. Films were developed using a X-OMAT 1000 Processor (Eastman Kodak Co., Rochester, USA).

2.3.4 Amino acid analysis

To verify peptide concentration, as well as qualitatively measure purity and composition, ion-exchange-ninhydrin analysis was carried out on 0.1 to 1 nmoles of sample hydrolysate on a Biochrom 30 amino acid analyser (Biochrom Ltd., Cambridge, UK) by the Protein & Nucleic Acid Chemistry Facility, Department of Biochemistry, University of Cambridge.

2.3.5 Mass spectrometry analysis

To determine the mass of a peptide for comparison to its expected mass, matrix-assisted laser desorption/ionisation of 2 to 5 pmoles of sample was analysed by time-of-flight mass spectrometry using a Maldi Micro MX and Q-ToF micro (Waters UK Ltd., Elstree, UK) by the Protein & Nucleic Acid Chemistry Facility, Department of Biochemistry, University of Cambridge.

2.4 Recombinant AMP production methods

2.4.1 Construction of pET28-AMP-CPD vectors

The oligonucleotides listed in *Table 2.3* were used during vector construction as outlined in this section.

| Primer name | Sequence (5' to 3') | Notes |
|-------------|--|---|
| CPD-Via1 | GGACAGCAAATGGGTGCGGATCCGGAAAAATACTCCATAATCAAATG | <i>Bam</i> HI site underlined |
| CPD-Via2 | GGTGGTGGTGGTGGTGCCTCGAGACCTTGCGCGTCCCAGCTTAG | <i>Xho</i> I site underlined |
| K2C18 | GGTAAGAAACTGAAAAAGATCGGTACAGAAATCAAGAACTTCTCCAGAACTG | Sequence of K2C18 |
| CPD-Via3 | <u>AATCAAGA</u> ACTTCTCCAGAACTGGCGGATGGAAAAATACTCCATAATC | 3' of K2C18 underlined |
| K2Build-FWD | ACAGAAGTGACGACCATGGGT | Common FWD primer, <i>Nco</i> I underlined |
| K2Build-REV | ACTGTCAGAGAGGATCCAG | Common REV primer, <i>Bam</i> HI underlined |
| K2front#1 | <u>ACAGAAGTGACGACCATGGGT</u> AAGAACTGAAAAAGATCCTGCA | FWD primer site underlined |
| K2front#2 | <u>ACAGAAGTGACGACCATGGGT</u> AAGAACTGAAAAAGATCGGTCA | " |
| K2front#3 | <u>ACAGAAGTGACGACCATGGGT</u> AAGAACTGAAAAAGATCGGTAA | " |
| K2front#4 | <u>ACAGAAGTGACGACCATGGGT</u> CGTCGTGCGTCGTATCGGTCA | " |
| K2front#5 | <u>ACAGAAGTGACGACCATGGGT</u> AAGAACTGAAAAACATCGGTAA | " |
| K2front#6 | <u>ACAGAAGTGACGACCATGGGT</u> AAGAACTGAAAAACATCCTGAA | " |
| K2front#7 | <u>ACAGAAGTGACGACCATGGGT</u> GAAAACTGAAAAAGATCGGTCA | " |
| K2rear#1 | ATCAAGAACTTCTCCAGAACTGGGATCCTCTCTGACAGT | REV primer site underlined |
| K2rear#2 | ATCAAGAACAAATCCAGCGTCTGGGATCCTCTCTGACAGT | " |
| K2rear#3 | ATCCGTAACCTTCTCCAGCGTCTGGGATCCTCTCTGACAGT | " |
| K2rear#4 | ATCAAAAAATTCTCCAGAACTGGGATCCTCTCTGACAGT | " |
| K2rear#5 | ATCAAGGATTTCTGAAAAACTGGGATCCTCTCTGACAGT | " |
| K2mid#1aR | GTTTCTGGAAGAAGTTCTTGATTTTCTGACCGATCTTTTCAGTTTCTT | Overlap "Front" and "Rear" |
| K2mid#1R | GTTTCTGGAAGAAGTTCTTGATTTTCTGCAGGATCTTTTCAGTTTCTT | " |
| K2mid#2R | GTTTCTGGAATTTGTTCTTGATTTTCTGACCGATCTTTTCAGTTTCTT | " |
| K2mid#3R | GTTTCTGGAAGAAGTTCTTGATTTTCTTACCGATCTTTTCAGTTTCTT | " |
| K2mid#4R | GACGCTGGAAGAAGTTACGGATACGCTGACCGATACGACGACAGAC | " |
| K2mid#5R | GTTTCTGGAAGAATTTTGGATCTGTTTACCGATGTTTTCAGTTTCTT | " |
| K2mid#6R | GTTTCTGGAAGAATTTTGGATCTGTTTACCGATGTTTTCAGTTTCTT | " |
| K2mid#7R | GTTTTTTCAGGAAATCCTTGATACGCTGACCGATCTTTTCAGTTTCTT | " |
| K2mid#8R | GTTTTTTCAGGAAATCCTTGATACGCTGACCGATCTTTTCAGTTTCTT | " |
| K2mid#9R | GTTTTTTCAGGAAATCCTTGATACGTTTACCGATCTTTTCAGTTTCTT | " |

Table 2.3: Primers used to construct the various pET28-AMP-CPD vectors.

2.4.1.1 Construction of pET28-CPD

Primers CPD-Via1 and CPD-Via2 were used to PCR amplify (Phusion polymerase, annealing at 72°C) the cysteine protease domain (CPD) from the *rtxA* gene (NCBI reference sequence NB_231094.1, amino acids 3440-3650) of *V. cholerae* El Tor N16961. The 667 bp product was inserted into pET28a by CPEC Viaduct (see Section 2.2.6) so that the CPD coding region was between the *Bam*HI and *Xho*I sites of the plasmid, giving pET28-CPD. A C-terminal hexa-histidine tag was encoded in-frame by the pET28a vector downstream of the *Xho*I site.

2.4.1.2 Construction of pET28-K2C18-CPD_{WT}

The 54 bp sequence of K2C18 (see *Table 2.3*) had been previously inserted into pET28a by CPEC Viaduct (see *Section 2.2.6*) so that it lay between the *NcoI* and *NheI* sites of the plasmid (pET28-K2C18). The CPD sequence was PCR amplified (as per *Section 2.4.1.1*) using primers CPD-Via2 and CPD-Via3. The 674 bp product was inserted into pET28-K2C18 by CPEC Viaduct so that the fused K2C18-CPD coding region was between the *NcoI* and *XhoI* sites of the plasmid.

2.4.1.3 Construction of pET28-AMP-CPD variants

The 54 bp sequences encoding the K2C18 and MCC18 variants were stitched together from oligonucleotides (*Figure 2.2*) using polymerase cycling assembly (Phusion polymerase, annealing at 59°C) as per Stemmer et al. (1995), then subsequently inserted into pET28-CPD via the *NcoI/BamHI* sites. The assembly key was as follows: K2C18, Front#2, Mid#1a, Rear#1; K2C18(G8L), Front#1, Mid#1, Rear#1; K2C18(F14K), Front#2, Mid#2, Rear#2; K2C18(Q9K), Front#3, Mid#3, Rear#1; RC18, Front#4, Mid#4, Rear#3; K2C18-2, Front#5, Mid#5, Rear#4; K2C18-2(G8L), Front#6, Mid#6, Rear#4; MCC18, Front#7, Mid#7, Rear#5; MCC18(E2K), Front#2, Mid#8, Rear#5; MCC18(E2K,Q9K), Front#3, Mid#9, Rear#5.

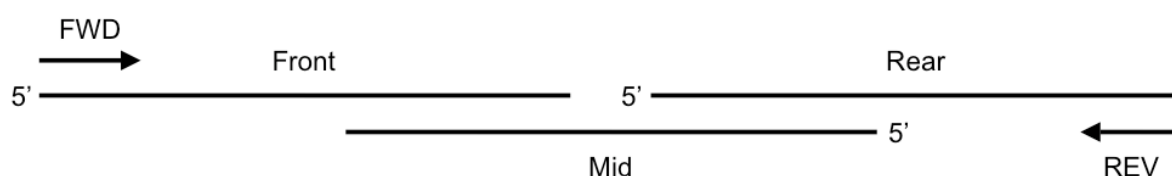


Figure 2.2: Schematic of the polymerase cycling assembly approach used to construct the K2C18 and MCC18 analogue sequences. Oligonucleotide sequences are outlined in *Table 2.3*.

2.4.2 CPD on-column cleavage

All steps were carried out at 4°C or on ice where possible. Ni-NTA agarose beads (50% slurry, ~25 mg binding capacity/mL; QIAGEN Ltd.) were added to cell clarified lysate (see *Section 2.3.2*) at 1 mL per L original culture volume, and CPD-hexahistidine tagged protein allowed to bind for at least 1 h while agitating on a roller. Ni-NTA beads were collected by centrifugation at ~100 rcf for 2 min, placed in a disposable column, and washed with at least 15 bead volumes of wash buffer

(20 mM NaH₂PO₄, pH 8.0, 500 mM NaCl, 10 mM imidazole). On-column cleavage of CPD was induced via the addition of 1 bead volume wash buffer containing 1 mM IP₆ (catalogue #593648, Sigma-Aldrich Company Ltd.), sealing the column, and incubating for at least 1 h while agitating on a rotating wheel. Free peptide was collected by washing the column with an additional 5 bead volumes of wash buffer in 1 bead volume fractions. The cleaved CPD-His tag was eluted by washing the column with 1 bead volume of wash buffer containing 500 mM imidazole.

2.4.3 Further AMP purification

2.4.3.1 C18 hydrophobic resin purification of AMPs

TFA was added to pooled free AMP fractions (*Section 2.4.2*, typically ~4.5 to ~5.5 mL) to a final concentration of 0.1% before application onto two Sep-Pak Vac C18 cartridges (50 mg resin, ~2 mg binding capacity [Waters UK Ltd.], pre-equilibrated with 0.8 mL 65% acetonitrile/0.07% TFA followed by 1.6 mL 2% acetonitrile/0.07% TFA). A syringe was used to pull samples through. AMPs bound to the C18 resin were washed with 1.6 mL 2% acetonitrile/0.07% TFA, then eluted with 0.4 mL 15%/30%/45%/65% acetonitrile/0.07% TFA in a step-wise manner. Fractions were dried using a VacuFuge 5301 (Eppendorf AG, Hamburg, Germany) at ~30°C and resuspended in sterile water, then stored at -80°C. This procedure was performed in collaboration with Dr Tatsuya Yoshimi.

2.4.3.2 Molecular weight cut-off (MWCO) purification of AMPs

Pooled free AMP fractions (*Section 2.4.2*) were applied onto a Vivaspin 2 10 kDa MWCO polyethersulfone filtration spin column (Sartorius Stedim Biotech S.A., Aubagne, France) and filtered as per the manufacturer's instructions. The low-molecular mass flow-through was collected and dialysed using Spectra/Por 7 2 kDa MWCO regenerated cellulose membrane (Spectrum Laboratories Inc., Rancho Dominguez, USA) against 0.1 M ammonium acetate at 4°C for ~16 h while stirring. Dialysate was lyophilised using a FTS Dura-Stop MP freeze dryer (SP Industries Inc., Warminster, USA) as per the manufacturer's instructions, resuspended in sterile water, and stored at -80°C.

2.4.4 Physical properties of peptides

The charge (Q) of a peptide was determined by counting the number of positively-charged amino acids (arginine, lysine and histidine) and subtracting the number of negatively-charged amino acids (aspartic acid and glutamic acid). The positively-charged N-terminal and negatively-charged C-terminal negated each other. Cationic face angle (Φ) of an α -helix was determined using a helical wheel plot, with 20° between each position assuming an ideal α -helix conformation. Mean hydrophobicity (H) was derived from Eisenberg's consensus hydrophobicity scale (1984). The relative hydrophobic moment (μ_{Rel}) of an α -helix represented the vector sum of the hydrophobicities of all residues (assuming an ideal α -helix) expressed as a % of μ_{max} (0.80 – determined by an 18 residue peptide consisting of 9 arginines and 9 isoleucines in an ideal α -helix), and was calculated as per Eisenberg (1984). H and μ_{Rel} were calculated using HydroMCalc (Antimicrobial Peptides Laboratory, University of Trieste; accessed at <http://www.bbcm.univ.trieste.it/~tossi/HydroCalc/HydroMCalc.html>). Molecular masses and other properties of proteins and peptides were predicted using ProtParam (Swiss Institute of Bioinformatics; Gasteiger et al. [2003]; accessed at <http://web.expasy.org/protparam/>).

2.5 Antimicrobial activity assays

2.5.1 Microbial test strains

Escherichia coli (DH10B), *Pseudomonas putida* (American Type Culture Collection #47054), *Bacillus subtilis* (American Type Culture Collection #6051 derivative [#NZ8900, NIZO Food Research B.V., Ede, Netherlands]) and *Candida albicans* (laboratory strain) were used in antimicrobial activity assays. Starter cultures were grown overnight in LB medium at ~225 rpm, 37°C for *E. coli* and *P. putida*; 30°C for *B. subtilis* and *C. albicans*. The medium for *C. albicans* was supplemented with 0.05% glucose. These cultures were used to inoculate fresh medium and grown as above for several hours until logarithmic growth was achieved (OD₆₀₀ between 0.4 to 0.6). An OD₆₀₀ equivalent to 1 was assumed to equate to approximately 5x10⁸ cfu/mL for *E. coli*, *P. putida* and *B. subtilis* (Domínguez et al., 2001; Fraile et al., 2001; Garbisu et al., 1998), and was approximated experimentally by plating culture dilutions. For *C. albicans*, an OD₆₀₀ of 1 was assumed to equate to approximately 3x10⁷ cfu/mL (Jackson et al., 2007).

2.5.2 Agar radial diffusion assay

5 mL of cooled molten LB top agar (0.7%) was inoculated with 15 μ L overnight culture of *E. coli* TOP10 and poured over an LB agar plate. Wells (~2.5 mm diameter) were punched into agar and 10 μ L of AMP fraction added. Plates were incubated for ~16 h upright at 37°C. Zones of growth inhibition around a well indicated antimicrobial activity.

2.5.3 Antimicrobial liquid culture assay

Logarithmic stage cultures of test microbes (see *Section 2.5.1*) were diluted in fresh medium to achieve approximate cell concentrations of 5×10^4 cfu/mL (*E. coli*, *P. putida* [37°C] and *B. subtilis* [30°C]) or 6×10^3 cfu/mL (*C. albicans* [30°C]). Unless otherwise indicated, nutrient broth (NB; 0.5% peptone, 0.3% meat extract) was used for *E. coli* and *C. albicans*, and LB for *P. putida* and *B. subtilis*. In a 96-well plate (Nunc brand, Nunclon Δ surface; Fisher Scientific UK Ltd.), 90 μ L of microbe was added to a duplicate dilution series of each AMP in sterile water. The plate was sealed with Parafilm (Pechiney Plastic Packaging Inc., Chicago, USA) and incubated (stationary) at the appropriate temperature for each microbe (see above), and OD₆₀₀ measured periodically using an EnVision 2104 plate reader spectrophotometer (PerkinElmer Inc., Waltham, USA). Minimum inhibitory concentration (MIC) was determined as the concentration of peptide that completely inhibited microbial growth after 18 h (bacteria) or 23.5 h (yeast).

2.6 Novel AMP screening methods

2.6.1 Construction of AMP screen vectors

The oligonucleotides listed in *Table 2.4* were used during vector construction as outlined in this section.

| Primer name | Sequence (5' to 3') | Notes |
|--------------------------|--|---|
| pBADmFWD | TTAGCGGATCCTACCTGACGC | <i>Bam</i> HI site underlined |
| pBADmREV | GAGCTCCATGGTCATGGTTAATTCCTCC | <i>Nco</i> I site underlined |
| <i>Nco</i> I-CRAMP-FWD | GCA CC ATGGGTAAGAACTGAAAAAG | <i>Nco</i> I site underlined |
| <i>Bst</i> BI-CRAMP-REV2 | AGC TT CGAACGTACAGTTTCTGGAAGAAG | <i>Bst</i> BI site underlined |
| <i>Bst</i> BI-CRAMP-REV1 | AGC TT CGAACTACAGTTTCTGGAAGAAG | <i>Bst</i> BI site underlined, stop codon in bold |
| K2CF-A | TAGCACCATGGCGATTAGTCGTCTGGCGGGTCTGCTGCGTAAAG | <i>Nco</i> I site and overlap underlined |
| K2CF-B | ACCAATTTTTTTCAGTTTTTTTACCAATTTTTTCACCACCTTTACGCAGCAGACCCG | Overlaps underlined |
| K2CF-C | GTAAAAAACTGAAAAAAATTGGTCAGAAAATTA AAAACTTCTTCCAGAACTGGTA | Overlaps underlined |
| K2CF-D(Tag) | GAAAGCTTTCGAATTTCCGGCTGCGGTACCAGTTTCTGGAAGAAGT | <i>Bst</i> BI site and overlap underlined |
| K2CF-D(*) | GAAAGCTTTCGAAT TC ATTCCGGCTGCGGTACCAGTTTCTGGAAGAAGT | <i>Bst</i> BI site and overlap underlined, stop codon in bold |
| Hok-FWD1 | ATTCCATGGGTAAACTACCACGAAGTTCCCTTG | <i>Nco</i> I site underlined |
| Hok-REV | ATTTTCGAACTACTTACCGGATTCGTAAGCC | <i>Bst</i> BI site underlined, stop codon in bold |
| Hok-FWD2 | ATTCCATGGTAAACTACCACGAAGTTCCCTTG | <i>Nco</i> I site underlined |
| LacZ-FWD | GCGCAACGCAATTAATGTGAGTTAGC | |
| LacZ-REV | CAGGGTTATTGTCTCATGAGCGG | |
| pBAD-Afelstop-FWD | CATGAGCGCTAAGTAAGTAAC | <i>Afe</i> I site underlined, stop codons in bold |
| pBAD-Afelstop-REV | CATGGTTACTTACTTAGCGCT | " |

Table 2.4: Primers used to construct the various pAMP vectors.

2.6.1.1 Construction of pBADm

Primers pBADmFWD and pBADmREV were used to PCR amplify (Advantage 2 polymerase, annealing at 59°C) a 100 bp product from pBAD/gIII-A spanning from the unique *Bam*HI site to the start codon of the pIII secretion signal. This was subsequently digested with *Bam*HI/*Nco*I and ligated into dephosphorylated pBAD/gIII-A via the same sites, resulting in the removal of the pIII-coding sequence from the plasmid.

2.6.1.2 Insertion of K2C18 and K2C38 into pBAD/gIII-A and pBADm

2.6.1.2.1 Cloning of K2C18

Primers *Nco*I-CRAMP-FWD and *Bst*BI-CRAMP-REV2 were used to PCR amplify (Advantage 2 polymerase, annealing at 54°C) the K2C18 sequence from the 54 bp oligonucleotide K2C18 (see Table 2.3). In addition, primer *Bst*BI-CRAMP-REV1 was used in conjunction with primer *Nco*I-CRAMP-FWD to include a stop codon immediately after the C-terminal leucine of K2C18, thus preventing read-through to

the C-terminal hexa-histidine tag in the pBAD vectors. These PCR products were subsequently digested with *NcoI/BstBI* and ligated into dephosphorylated pBAD/gIII-A via the same sites, resulting in pBAD/gIII-K2C18-Tag and pBAD/gIII-K2C18* respectively.

2.6.1.2.2 *Cloning of K2C38*

The 114 bp sequence of K2C38 was stitched together using polymerase cycling assembly (*Taq* polymerase, annealing at 56°C) as per Stemmer et al. (1995) from the overlapping oligonucleotides K2CF-A, K2CF-B, K2CF-C and K2CF-D(Tag). In addition, oligonucleotide K2CF-D(*) was used instead of K2CF-D(Tag) to include a stop codon immediately after the C-terminal glutamic acid of K2C38, thus preventing read-through to the C-terminal hexa-histidine tag in the pBAD vectors. These DNA products were subsequently digested with *NcoI/BstBI* and ligated into dephosphorylated pBAD/gIII-A and pBADm via the same sites, resulting in pBAD/gIII-K2C38-Tag, pBAD/gIII-K2C38*, pBADm-K2C38-Tag and pBADm-K2C38* respectively. See *Section 4.2.2.1* for an overview of these plasmids.

2.6.1.3 Insertion of *hok* and *hokFS* into pBAD/gIII-A and pBADm

Primers Hok-FWD1 and Hok-REV were used to PCR amplify (Phusion polymerase, annealing at 60°C) the *hok* sequence from pKG1010. In addition, primer Hok-FWD2 was used in conjunction with primer Hok-REV to produce a frame-shift (FS) mutation immediately after the start codon. These ~160 bp PCR products were subsequently digested with *NcoI/BstBI* and ligated into dephosphorylated pBAD/gIII-A and pBADm via the same sites, resulting in pBAD/gIII-Hok, pBAD/gIII-HokFS, pBADm-Hok and pBADm-HokFS respectively.

2.6.1.4 Construction of pAMP/S and pAMP

Primers LacZ-FWD and LacZ-REV were used to PCR amplify (Advantage 2 polymerase, annealing at 58°C) the *lacZ* operon from pUC19. The 5' ends of this 757 bp product were phosphorylated and blunt-end ligated into *BsaAI*-digested/dephosphorylated pBAD/gIII-A and pBADm, resulting in pBAD/gIII-LacZ and pBADm-LacZ respectively. For the new *araBAD* cloning site, complimentary oligonucleotides pBAD-*Afelstop*-FWD and pBAD-*Afelstop*-REV were designed to

have *Nco*I-compatible 5' overhangs, and annealed by heating 100 μ L of a 25 μ M solution to 72°C for 10 min and subsequently cooling to 53°C over 10 min. This product was phosphorylated and ligated into *Nco*I-digested/dephosphorylated pBAD/gIII-LacZ and pBADm-LacZ, resulting in pAMP/S and pAMP respectively. See *Section 4.2.3* for a schematic of these plasmids.

2.6.2 Growth curves

Growth curves were carried out in sterile 24-well plates (Nunc brand, non-treated surface; Fisher Scientific UK Ltd.). 2 mL of fresh LB (containing appropriate antibiotic) was inoculated with a 1:50 dilution of overnight starter culture, which was then split into duplicate 1 mL aliquots in separate wells. Plates were sealed using a sterile polydimethylsiloxane mat and binder clips, and incubated flat at 37°C, ~225 rpm for 1.5 h before L-arabinose (0.05%) was added to one of each duplicate well. OD₆₀₀ was measured periodically using an EnVision 2104 plate reader spectrophotometer (PerkinElmer Inc., Waltham, USA).

2.6.3 Preparation and insertion of genomic DNA (gDNA) into vectors

2.6.3.1 Shearing of gDNA

Genomic DNA extracts from human (*Homo sapiens*) and bdelloid rotifer (*Adineta ricciae*) sources were gifts from Prof Alan Tunnacliffe and Dr Chiara Boschetti respectively. 2 to 4 μ g of gDNA was added to 500 μ L nebulisation buffer (Tris-EDTA pH 8.0, 40% glycerol) and placed in an Invitrogen disposable nebuliser as per the manufacturer's instructions (Life Technologies Ltd.). Nebulisation proceeded on ice in a fume hood using nitrogen gas at 3.45 bar for 2 min. The nebuliser was briefly centrifuged at 1000 rcf, and the sheared gDNA solution removed and concentrated via ethanol precipitation (Sambrook & Russell, 2001). To remove potential 5' or 3' overhangs from the sheared gDNA that may have resulted from nebulisation, and to ensure phosphorylation of 5' ends prior to ligation, an End-It DNA End-Repair kit was used as per the manufacturer's instructions (Epicentre Biotechnologies Inc.). Sheared and end-repaired gDNA was subsequently purified using a QIAquick Nucleotide Removal kit (see *Section 2.2.2*).

2.6.3.2 Insertion of gDNA into pAMP/S and pAMP

Fragmented gDNA was ligated into 50 ng of *AfeI*-digested/dephosphorylated pAMP/S and pAMP at a vector:insert molar ratio of 1:5 using concentrated T4 DNA ligase (1000 U, New England Biolabs (UK) Ltd.) at 16°C for ~16 h (10 µL total volume). Ligase was inactivated prior to transformation via storage of ligation mixture at -20°C. For each desired agar plate, 1 µL of ligation mixture was transformed into ~25 µL of chemically-competent *E. coli* TOP10.

2.6.4 Replica plate screening of pAMP/S and pAMP libraries

Replica plating followed a modified version of that outlined by Sambrook & Russell (2001). In brief, transformations were plated on sterile 85 mm mixed cellulose ester membranes (0.45 µm pore size, white; Advantec MFS Inc. [Dublin, USA]) on LB agar (including 50 µg/mL X-gal) and incubated at 30°C until pinhead-sized colonies were visible (typically 16 h). This master membrane was laid colony side up on Whatman 3MM Chr blotting paper (GE Healthcare UK Ltd.) on a thick glass plate, covered with a fresh membrane (pre-wetted on a fresh agar plate), followed by additional blotting paper and another glass plate. Pressure was evenly applied to ensure uniform colony transfer, the apparatus disassembled and the membranes asymmetrically keyed using a needle dipped in indelible ink before being separated by tweezers. This first replica filter was placed on LB agar containing 0.05% arabinose (+ara membrane). A second replica membrane was made and placed on LB agar lacking arabinose (-ara membrane). The master membrane was returned to its agar plate, and all plates were further incubated at 37°C until large colonies appeared (typically 8 h).

2.6.5 Identification of pAMP/S and pAMP library hits

Replica membranes for each plate (master, +ara and -ara) were scanned (Epson Perfection 3490 Photo flatbed scanner) and the images imported into Photoshop v7.0.1 (Adobe Systems Inc., San Jose, USA) where they were converted to black & white via the “threshold” function: colonies became black, the background white. Each membrane was then coloured using the “screen” function: the master membrane blue, +ara black, and -ara red. The master and +ara images were overlaid on the -ara image as separate layers, aligned to the asymmetric key, and growth inhibition of colonies on the +ara membrane identified using the “darken”

function. Where no or little growth inhibition had occurred, the black colony on the +ara membrane obscured its red replica on the -ara membrane. Where growth was inhibited, the red -ara colony was visible and easily detected by eye. Any identified hits were then compared to the master membrane, and validated manually by inspecting the colonies themselves.

2.6.6 Bioinformatic analysis of pAMP/S and pAMP library hits

Insert gDNA sequences from selected pAMP/S and pAMP hits were compared to the National Centre for Biotechnology Information's (NCBI, Bethesda, USA) nr database (all GenBank+EMBL+DDBJ+PDB sequences [but no EST, STS, GSS, environmental samples or phase 0, 1 or 2 HTGS sequences]) using NCBI's translated query versus translated database basic local alignment search tool (tBLASTx v2.2.25+; Altschul et al. [1997]; accessed at <http://blast.ncbi.nlm.nih.gov/>). Default search settings were used, with low complexity regions left unfiltered. To search for putative antisense RNA partners to endogenous *E. coli* transcripts, the reverse complement of an insert sequence was compared against the *E. coli* DH10B genome (NCBI taxid:316385; Durfee et al. [2008]) using NCBI's nucleotide query versus nucleotide database (BLASTn v2.2.25+; Altschul et al. [1997]). Default search settings were used, except the "expect" threshold was raised from 10 to 100, and low complexity regions left unfiltered. RNA secondary structure prediction was performed using RNAfold v2.0.0 (Hofacker [2003]; default options used, accessed at <http://rna.tbi.univie.ac.at/>). Peptide secondary structure prediction was performed using Multivariate Linear Regression Combination (SOPMA-GOR4-SIMPA) (Guermeur et al. [1999]; accessed at <http://npsa-pbil.ibcp.fr/>).

2.6.7 Modification of pAMP/S library hits

The oligonucleotides listed in *Table 2.5* were used during vector modification as outlined in this section.

| Primer name | Sequence (5' to 3') | Notes |
|-------------|---|---------------------------------|
| pAMP/S-SDM1 | GCTAACCAA <u>ACCGG</u> TAAACCCCGC | <i>AgeI</i> site underlined |
| pAMP/S-SDM6 | CGAACAGCAGTTTTTTT <u>QC</u> ATGGTTAATTCC | Frameshift insertion underlined |
| pAMP/S-SDM5 | GGAATTAACCATGGAAAAAACTGCTGGTTCCG | " |
| pAMP/S-SDM2 | ACCAGCTGCAGAT <u>CTCG</u> AGCTCC | <i>XhoI</i> site underlined |
| CamR-FWD2 | GATA <u>TCATGA</u> ATAAATACCTGTGACGGAAG | <i>BspHI</i> site underlined |
| CamR-REV | AGAT <u>TCATGA</u> GCTAACTCACATTAATTGCGTTGC | " |

Table 2.5: Primers used to modify selected pAMP/S hit vectors.

2.6.7.1 Site-directed mutagenesis to introduce a frame-shift

Site-directed mutagenesis was carried out as per overlap extension PCR (Sambrook & Russell, 2001). For a frame-shift downstream of the start codon (ATG to ATGG), primers pAMP/S-SDM1 and pAMP/S-SDM6 were used to PCR amplify (Phusion polymerase, annealing at 56°C) the region from the frame-shift upstream to the *AgeI* site of pAMP/S; primers pAMP/S-SDM5 and pAMP/S-SDM2 were used to PCR amplify the region from the frame-shift downstream to the *XhoI* site. 5 ng of the upstream and downstream product was then mixed and PCR amplified (as previously) using the flanking primers pAMP/S-SDM1 and pAMP/S-SDM2. The full-length product was purified, digested with *AgeI/XhoI*, and ligated into pAMP/S that had been digested in the same manner and dephosphorylated.

2.6.7.2 Antibiotic resistance cassette exchange

The primers CamR-FWD2 and CamR-REV were used to PCR amplify (Phusion polymerase, annealing at 58°C) the chloramphenicol acetyl transferase operon (*chl^R*) from pCC1FOS. The product was purified, digested with *BspHI*, and ligated into pBADm (see *Section 2.6.1.1*) that had been digested in the same manner and dephosphorylated; this replaced the β -lactamase operon (*amp^R*), giving the plasmid pBADmC. Flanking primers pAMP/S-SDM1 and pAMP/S-SDM2 were then used to PCR amplify (Phusion polymerase, annealing at 60°C) selected pAMP/S hits; the products were purified, digested with *AgeI/XhoI*, and ligated into pBADmC that had been digested in the same manner and dephosphorylated.

2.6.7.3 Secretion tag removal

CPEC Viaduct (see *Section 2.2.6*) was used to remove the 60 bp between the start codon of the pAMP/S *araBAD* site and the start of the insert sequence, i.e. the *gIII* secretion signal sequence and the downstream 9 bp encoding Thr-Met-Ser. *Table 2.6* outlines the primers used for each plasmid; reverse complements of each were also utilised, but are omitted for clarity.

| Plasmid name | Sequence (5' to 3') |
|--------------|---|
| pAMP/S-H2 | GGGCTAACAGGAGGAATTAACCATG <u>TTACTTGTCTTATTTCAAATAAAT</u> |
| pAMP/S-H3 | GGGCTAACAGGAGGAATTAACCATG <u>GTTTTAAATGAATTCAAAAGTCAG</u> |
| pAMP/S-H4 | GGGCTAACAGGAGGAATTAACCATG <u>AAAAGGTACCTTGGTACCCGATGC</u> |
| pAMP/S-H5 | GGGCTAACAGGAGGAATTAACCATG <u>TTCTCTAAATTTCTGTTCTTGCTTC</u> |
| pAMP/S-H8 | GGGCTAACAGGAGGAATTAACCATG <u>GTGATTGTACTAATTTACATTTCTA</u> |
| pAMP/S-R14 | GGGCTAACAGGAGGAATTAACCATG <u>AAATTAATATGCAAATAATCATTG</u> |
| pAMP/S-R15 | GGGCTAACAGGAGGAATTAACCATG <u>CGGAACTCATTTTCAATGCTTCTAC</u> |
| pAMP/S-R17 | GGGCTAACAGGAGGAATTAACCATG <u>GAAGATTATTTTCGACGAACGTTG</u> |

Table 2.6: Primers used to remove *gIII* sequence from several pAMP/S hits via CPEC Viaduct. Start codon in bold, hit insert sequence underlined. Reverse complements of each primer also used, not shown.

2.6.8 Insertion and expression of AMP hits in pET28-CPD

2.6.8.1 Insertion of selected AMP hits into pET28-CPD

The primers outlined in *Table 2.7* were used to PCR amplify (Phusion polymerase, annealing at 62°C) the following pAMP/S hit sequences (primer combination in parentheses): pAMP/S-H2 (CPD-Via5/6), pAMP/S-H3 (CPD-Via7/8), pAMP/S-H4 (CPD-Via9/10), pAMP/S-R17 (CPD-Via11/12), pAMP/S-R22 (CPD-Via22/23), pAMP/S-R25 (CPD-Via24/25), pAMP/S-R32 (CPD-Via32/33), pAMP/S-R33 (CPD-Via34/35) and pAMP/S-R39 (CPD-Via42/43). CPEC Viaduct (see *Section 2.2.6*; annealing at 51°C) was then used to insert these products (all encoding C-terminal leucine codons) into pET28-CPD between the *NcoI*-provided start codon and the *BamHI* site. The resulting plasmids were named pET28-S-xxx-CPD accordingly.

| Primer name | Sequence (5' to 3') |
|-------------|--|
| CPD-Via5 | AAC TTAAGAAGGAGATATACC ATG <u>TACTTGTCTTATTTCAAATAAAT</u> |
| CPD-Via6 | TATGGAGTATTTTTCCATCCGCC CAG <u>ACGAGTTCTATTCCAAAACCTATT</u> |
| CPD-Via7 | AAC TTAAGAAGGAGATATACC ATG <u>GTTTTAAATGAATTCAAAGTCAGT</u> |
| CPD-Via8 | TATGGAGTATTTTTCCATCCGCC CAG <u>CTTACTTAGCGGAATACTAGGCATA</u> |
| CPD-Via9 | AAC TTAAGAAGGAGATATACC ATG <u>AAAAGGTACCTTGGTACCCGATG</u> |
| CPD-Via10 | TATGGAGTATTTTTCCATCCGCC CAG <u>CTCCTTCTTGTCCTCTTTTT</u> |
| CPD-Via11 | AAC TTAAGAAGGAGATATACC ATG <u>GGAAGATTATTTTCGACGAACGTTGA</u> |
| CPD-Via12 | TATGGAGTATTTTTCCATCCGCC CAG <u>TTTCTCCACGGTATCACCATTGATC</u> |
| CPD-Via22 | AAC TTAAGAAGGAGATATACC ATG <u>TACAGCAAAACAATCACTGTAAAGA</u> |
| CPD-Via23 | GGAGTATTTTTCCATCCGCC CAG <u>CTGGTAAGTCAATCCGATGATTATG</u> |
| CPD-Via24 | AAC TTAAGAAGGAGATATACC ATG <u>TTATTATATTCATTGAGACGTCGTT</u> |
| CPD-Via25 | GGAGTATTTTTCCATCCGCC CAG <u>TTCCAAAAGATCCACGAGGTTTG</u> |
| CPD-Via32 | AAC TTAAGAAGGAGATATACC ATG <u>CTCTTTTCCCTTCTCTCTCTC</u> |
| CPD-Via33 | GGAGTATTTTTCCATCCGCC CAG <u>TTTCTTGTTTCGTTTCGTAAAAGAAAAAT</u> |
| CPD-Via34 | AAC TTAAGAAGGAGATATACC ATG <u>TGTTGGATTTTTTGCATTCCCTT</u> |
| CPD-Via35 | GGAGTATTTTTCCATCCGCC CAG <u>CTTAGCACATTTTCGCTGATCG</u> |
| CPD-Via42 | AAC TTAAGAAGGAGATATACC ATG <u>GACCGTGTTACTGCGTTTTGAG</u> |
| CPD-Via43 | GGAGTATTTTTCCATCCGCC CAG <u>GTTTGAATAACTTGGCGATCGAAT</u> |

Table 2.7: Primers used to transfer several pAMP/S hit sequences to pET28-CPD via CPEC Viaduct. Start codon (ATG) or critical CPD cleavage Leu codon (CAG, reverse complement) in bold, hit insert sequence underlined. Used in combination as outlined above.

2.6.8.2 Expression of selected AMP hits from pET28-CPD

Expression occurred as per Section 2.3.1, with pET28-S-H2-CPD, pET28-S-H3-CPD, pET28-S-H4-CPD, pET28-S-R17-CPD transformed into *E. coli* BL21(DE3) Star and 500 mL induced at 30°C for 5 h; pET28-S-R22-CPD, pET28-S-R25-CPD, pET28-S-R32-CPD, pET28-S-R33-CPD and pET28-S-R39-CPD were transformed into BL21(DE3)pLysS Rosetta cells and 250 mL induced at 16°C for 18 h.

2.7 Novel antiaggregant screening methods

2.7.1 Construction of pAG2-A β 42

The oligonucleotides listed in *Table 2.8* were used during vector construction as outlined in this section; BioBrick part sequences (BBa; Canton et al. [2008]) are available at the Registry of Standard Biological Parts (accessed at <http://partsregistry.org/>).

| Primer name | Sequence (5' to 3') | Notes |
|-------------|---|--------------------------------------|
| pAG-Via1 | GCCAGTATACACTCCGCTATCGCTTTACAGCTAGCTCAGTCCTAG | mCherry, 5' of promoter underlined |
| pAG-Via2 | GGTGTCGGGGCGCAGCCATGACCTATAAACGCAGAAAGGCCAC | mCherry, 3' of terminator underlined |
| pAG-Via3 | GGAGCGGATTTGAACGTTGCGAAGCITGACAGCTAGCTCAGTCCTAGGTA | Library, 5' of promoter underlined |
| pAG-Via4 | ICTCCTCTTTGCGGAGTAGCTAGCATTATACCTAGGACTGAGCTAGCTGT | Library, overlaps underlined |
| pAG-Via19 | TAGCTACTCCGCAAAGAGGAGAAAAGTAGTATGAGCGCTGACTAGGTGAG | Library, overlaps underlined |
| pAG-Via6 | ATGAGCGCTGACTAGGTGAGACGTCAGGCGACGTGCCCTGGCGGCAGTAGC | Library, overlaps underlined |
| pAG-Via7 | GATGCCTGGCAGTTTATGGCGGAGGAGAGCGTTCACCGACA | Library, 3' of terminator underlined |
| pAG-Via8 | GGGCTAACAGGAGGAATTAACCATGGATGCGGAATTTCCGCCATG | A β 42-EGFP, 5' underlined |
| pAG-Via9 | CAGCCAAGCTGGAGACCGTTTACTTTGTACAGCTCGTCCATG | A β 42-EGFP, 3' underlined |

Table 2.8: Primers used to construct pAG2-A β 42.

2.7.1.1 Insertion of mCherry operon

Primers pAG-Via1 and pAG-Via2 were used to PCR amplify (Phusion polymerase, touchdown 61°C to 52°C, then annealing at 65°C) the mCherry operon (BBa_J06702, partnered with the medium-strength promoter BBa_J23101) from pSB3K3-mCherryGFP1plus3. This 958 bp product was inserted into pBADm near its *Bsa*I site by CPEC Viaduct (see *Section 2.2.6*) so that the mCherry coding region lay between the *amp^R* operon and the origin of replication, resulting in pAG-A.

2.7.1.2 Construction and insertion of Library operon

The Library site consisted of the strong promoter BBa_J23119, ribosome binding site BBa_B0034, a start codon straddled by *Spel* and *AfeI* sites, stop codons for all three reading frames, and the *rrnB* T1/T2 terminator region (for more detail see *Appendix 8*). Primers pAG-Via3 and pAG-Via4, which overlapped, were extended by PCR amplification (Phusion polymerase, touchdown 62°C to 53°C, then annealing at 72°C), giving a 78 bp product (named A). Primers pAG-Via6 and pAG-Via7 were used to PCR amplify (as above) the *rrnB* T1/T2 terminator region

from pBADmC (see *Section 2.6.7.2*), giving a 249 bp product (named B). Product B was re-amplified (as above) using Primers pAG-Via19 and pAG-Via 7, giving a 279 bp product (named C). Products A and C, which overlapped, were stitched together by PCR amplification (as above) using primers pAG-Via3 and pAG-Via7, giving the full-length Library site. This 335 bp product was inserted into pAG-A by CPEC Viaduct (see *Section 2.2.6*; annealing at 62.5°C) between the end of the *araBAD* site's transcriptional terminator and the start of the *amp^R* operon, resulting in pAG2-B.

2.7.1.3 Insertion of A β 42-EGFP into the *araBAD* site

Primers pAG-Via8 and pAG-Via9 were used to PCR amplify (Phusion polymerase, touchdown 62°C to 53°C, then annealing at 72°C) A β 42-EGFP from pA β 42-EGFP. This 929 bp product was inserted into pAG2-B by CPEC Viaduct (see *Section 2.2.6*; annealing at 60°C) between the *araBAD* site's start codon and the *PmeI* site, resulting in pAG2-A β 42. The linker between the A β 42 and EGFP domains was GSAGSAAGSGESHMV (15 aa; as per Baine et al., [2009]). See *Section 5.2.1.1* for a schematic of pAG2-A β 42; the full annotated sequence is shown in *Appendix 7*.

2.7.2 Construction of pAG2-A β 42 controls

The oligonucleotides listed in *Table 2.9* were used during vector construction as outlined in this section.

| Primer name | Sequence (5' to 3') | Notes |
|-------------|---|--|
| EGFP-Via1 | CCATGGAGCTCATGGTGAGCAAGGGCCGAGGAGCTGTTACACCGGGG | EGFP, 5' underlined |
| EGFP-Via2 | GTTTTTGTCTAGATACTTGTACAGCTCGTCCATGCCGAGAGTGATCCC | EGFP, 3' underlined |
| pAG-Via21 | GCTACTCCGCAAAGAGGAGAAAACTAGTATGGTGAGCAAGGGCCGAGGAGC | EGFP, 5' underlined, <i>SpeI</i> in bold |
| Cherry-REV2 | TCATTCTAGATACTTGTACAGCTCGTCC | EGFP, 3' underlined, <i>XbaI</i> in bold |
| pAG-Via8 | GGGCTAACAGGAGGAATTAACCATGGATGCCGAATTCGCCATG | GM6, 5' underlined |
| pAG-Via10 | CGCAATCACCCACGCCG | GM6, 3' underlined |
| p2F | CATGCAGAAACTGGATGTGGTAGCGGAAGATGCTGGCTCTAACAAATAAG | Pep2, overlap underlined |
| p2R | GATCCTTATTTGTTAGAGCCAGCATCTTCCGCTACCACATCCAGTTTCTG | Pep2, overlap underlined |
| pAG-Via23 | CTCCGCAAAGAGGAGAAAACTAGTATGCAGAAACTGGATGTGGTAGCG | Pep2, 5' underlined |
| pAG-Via16 | CGTCGCCTGACGTCTCACCTAGTCATTTGTTAGAGCCAGCATCTTCCG | Pep2, 3' underlined |
| pAG-Via25 | CTCCGCAAAGAGGAGAAAACTAGTATGTCCCTCTCAGCAGAACCAGAAC | AavLEA1, 5' underlined |
| pAG-Via18 | CGTCGCCTGACGTCTCACCTAGTCAGTCGCGGCCCTTGACCTT | AavLEA1, 3' underlined |

Table 2.9: Primers used to construct pAG2-A β 42 controls.

2.7.2.1 Construction of pBADm-EGFP

As a single (green) positive control, EGFP was inserted into the *araBAD* site of pBADm (see *Section 2.6.1.1*). Primers EGFP-Via1 and EGFP-Via2 were used to PCR amplify (Phusion polymerase, touchdown 65°C to 56°C, then annealing at 72°C) EGFP from pET28-A7-EGFP. This 744 bp product was inserted into pBADm by CPEC Viaduct (see *Section 2.2.6*; annealing at 65°C) between the *NcoI* and *XbaI* sites of the *araBAD* region, resulting in pBADm-EGFP. The sequence ran on into the pBADm C-terminal hexa-histidine tag (17 aa extra).

2.7.2.2 Insertion of EGFP into Library site of pAG2-A β 42

To test the activity of the Library site, EGFP was inserted as a reporter. Primers pAG-Via21 and Cherry-REV2 were used to PCR amplify (Phusion polymerase, touchdown 60°C to 51°C, then annealing at 58°C) EGFP from pET28-A7-EGFP (the same sequence as Baine et al. [2009]), incorporating a 5' *SpeI* and 3' *XbaI* site. This 757 bp product was subsequently digested with *SpeI/XbaI* (which share compatible sticky ends) and ligated into dephosphorylated pAG2-B via the *SpeI* Library site, resulting in pAG2-B/EGFP.

2.7.2.3 Construction of pAG2-GM6

pAG2-A β 42 was modified to incorporate the non-aggregating A β 42 mutant GM6 (Wurth et al., 2002) as a positive control. Primers pAG-Via8 and pAG-Via10 were used to PCR amplify (Phusion polymerase, touchdown 62°C to 53°C, then annealing at 72°C) GM6 from pSALect-GM6. This 151 bp product was inserted into pAG2-A β 42 by CPEC Viaduct (see *Section 2.2.6*; annealing at 51°C) to replace the A β 42 domain, resulting in pAG2-GM6.

2.7.2.4 Insertion of Peptide 2 (Pep2) and AavLEA1 into pAG2-A β 42

The coding sequences for Pep2 (Baine et al., 2009) and AavLEA1 (Browne et al., 2002) were inserted into the Library site of pAG2-A β 42 to use as antiaggregant controls. To create the Pep2 coding sequence, complimentary oligonucleotides p2F and p2R were annealed by heating 20 μ L of a 25 μ M solution to 95°C for 5 min and subsequently cooling from 80°C to 60°C at 0.5°C/min. This 50 bp product was PCR amplified (Phusion polymerase, touchdown 62°C to 53°C, then annealing at 72°C)

using primers pAG-Via23 and pAG-Via16. The resulting 92 bp product was inserted into pAG2-A β 42 by CPEC Viaduct (see *Section 2.2.6*; annealing at 51°C) between the Library site's start codon and first stop codon, resulting in pAG2-A β 42-Pep2.

Primers pAG-Via25 and pAG-Via18 were used to PCR amplify (as above) AavLEA1 from pET15-AavLEA1. This 476 bp product was inserted into pAG2-A β 42 as per Pep2 above, resulting in pAG2-A β 42-AavLEA1.

2.7.3 Insertion of random DNA library into pAG2-A β 42

Oligonucleotides (see *Table 2.10*) coding for a maximum of either 12 or 24 random amino acids (13 or 25 including the initiating methionine) were PCR amplified (from 10 pmol of template; Phusion polymerase, annealing at 57°C) using primers pAGIns-FWD and pAGIns-REV. Only 15 amplification cycles were used in order to minimise any loss of product diversity. These 74 and 110 bp products were subsequently digested with *SpeI/AfeI* and ligated into 200 ng of dephosphorylated, gel extracted pAG2-A β 42 via the same sites at a vector:insert molar ratio of 1:5 using T4 DNA ligase (Fermentas GmbH) at 23°C for 1 hour (40 μ L total volume). Ligase was inactivated prior to transformation via storage of ligation mixture at -20°C. The resulting libraries were named pAG2-A β 42-74 (13 aa peptide) and pAG2-A β 42-110 (25 aa peptide). For each desired agar plate, 1.25 μ L of ligation mixture was transformed into ~12.5 μ L of chemically-competent *E. coli* DH10B. Each transformation was plated on LB agar containing 0.05% arabinose.

| Primer name | Sequence (5' to 3') | Notes |
|-------------|--|---|
| pAG-Ins74 | GCACGCTGCG <u>ACTAGTATG</u> (NNN) ₁₂ TGAAGCGC TIGATCCGATGC | 13 aa peptide; <i>SpeI/AfeI</i> underlined, start/stop codons in bold |
| pAG-Ins110 | GCACGCTGCG <u>ACTAGTATG</u> (NNN) ₂₄ TGAAGCGC TIGATCCGATGC | 25 aa peptide; <i>SpeI/AfeI</i> underlined, start/stop codons in bold |
| pAG-InsFWD | GCACGCTGCGACTAGTATG | 5' of pAG-Ins oligonucleotides |
| pAG-InsREV | GCATCGGATCAGCGCTTCA | 3' of pAG-Ins oligonucleotides |

Table 2.10: Oligonucleotides used to create random DNA libraries. N, any base pair.

2.7.4 Microscopy

Bright field and fluorescence microscopy were undertaken on an Optiphot-2 microscope with an EFD-3 epifluorescence system (Nikon Corp., Tokyo, Japan).

Images were captured using a DS-2Mv camera system (Nikon Corp.). EGFP had peak absorption and emission wavelengths of 488 nm and 509 nm, while mCherry was 587 nm and 610 nm respectively (Takara Bio Inc.). For EGFP, a B-2A filter block was used (excitation 470 +/- 20 nm, emission >520 nm); for mCherry, G-2A (excitation 535 +/- 25 nm, emission >590 nm). Bleed-through of mCherry fluorescence was observed when using the B-2A filter block for long exposure times; such excitation catches the extreme left-hand shoulder of mCherry's absorption spectrum, and its full emission spectrum is able to transit the longpass emission filter. No bleed-through of EGFP fluorescence was observed using the G-2A filter block.

2.7.5 Flow cytometry

2.7.5.1 Flow cytometry analysis of *E. coli* fluorescence

Cultures (3 mL) containing the various induced (see *Section 2.3.1*) constructs were diluted to approximately 2×10^7 cfu/mL in phosphate buffered saline (PBS; 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄, 2.7 mM KCl, 137 mM NaCl, pH 7.4) and analysed using a DXP 8 FACScan II flow cytometer (Cytex Development Inc., Fremont, USA) as per the manufacturer's instructions. Any cultures exhibiting cell clumping were thoroughly vortexed immediately prior to analysis to aid dispersion. Cells were initially gated on a dot plot by log Forward Scatter (FS) versus log Side Scatter (SS) to select for a homogenous cell population (G1), and G1 subsequently gated by log FS area versus log FS width to select for single cells only, i.e. doublet discrimination (G2). Cells selected by G1 and G2 were individually interrogated for EGFP or mCherry fluorescence using the appropriate single-positive control strains, and a compensation matrix applied to correct for any spectral overlap observed between the two (minimal). For EGFP, excitation at 488 nm was used, with emission monitored at 530 +/- 15 nm; for mCherry, excitation at 561 nm was used, with emission monitored at 615 +/- 12.5 nm. A total of 100,000 events were recorded per sample. Data was analysed using FlowJo v7.6.4 (Tree Star Inc., Ashland, USA).

2.7.5.2 Preparation of libraries for fluorescence-activated cell sorting

Approximately 25,000 colonies (30°C incubation for ~16 h after transformation [as per *Section 2.2.8*]) from each of the pAG2-Aβ42-74 and pAG2-Aβ42-110 libraries were resuspended and pooled in 1 mL of LB medium, and the cell density

subsequently normalised by dilution to $OD_{600} = 2.5$ (*Section 2.3.1*). The same was performed for approximately 25,000 colonies containing pAG2-A β 42, which was then split into three aliquots and spiked separately with 3 colonies of pAG2-A β 42-Pep2, pAG2-A β 42-AavLEA1 or pAG2-GM6. These five resuspensions were used as inoculums for induction of expression as per *Section 2.3.1*.

2.7.5.3 Fluorescence-activated cell sorting (FACS)

Cultures (3 mL) to be analysed were pelleted and resuspended in PBS prior to being diluted further as per *Section 2.7.5.1*, and sorted on a DakoCytomation MoFlo MLS cell sorter (Beckmann Coulter Inc., Brea, USA) as per the manufacturer's instructions under the supervision of Mr Nigel Miller (Flow Cytometry Facility, Department of Pathology, University of Cambridge). Cell population gating was as per *Section 2.7.5.1*, with spectral compensation performed manually by adjusting photomultiplier tube (PMT) gain. For EGFP, excitation at 488 nm was used, with emission monitored at 530 +/- 15 nm; for mCherry, excitation at 568 nm was used, with emission monitored at 613 +/- 10 nm. At least 20,000 total events were recorded per sample. Data was analysed using Summit v4.3.02 (Beckmann Coulter Inc., Brea, USA). For cell sorting, approximately 1×10^4 positive events (indicated) from 1×10^6 cells were selected and pooled into 3 mL fresh LB medium. For further enrichment, these pooled cultures were directly grown and induced as per *Section 2.3.1* before undergoing additional rounds of FACS (indicated). After the final round, cells were plated out on LB agar.

2.7.5.4 Calculations

To give a measure of the spread of data obtained during flow cytometry, the coefficient of variation (CV) was used (Huber, 2005). CV represents a normalised standard deviation, i.e. the standard deviation divided by the mean, and is commonly converted to a percentage. A more preferred statistic is the robust CV (rCV), which is not as skewed as CV by outlying data, and was calculated using *Equation 2.1* using the flow cytometer manufacturer's software (*Section 2.7.5.1*).

$$\%rCV = 100 * \frac{1}{2} (Intensity [at 84.13 percentile] - Intensity [at 15.87 percentile]) / Median$$

Equation 2.1: %rCV, robust coefficient of variation (as a percentage).

However, the rCV calculation was not available for the MoFlo MLS cell sorter (Section 2.7.5.3), so CV was used instead.

2.7.6 Identification of pAG2-A β 42 library hits on solid medium

EGFP fluorescence from induced colonies was photographed (3 s exposure) using an ImageQuant LAS 4000 system (GE Healthcare UK Ltd.). Excitation was at ~460 nm (blue epifluorescence), emission at 510 nm (510DF10 filter). Images were imported into ImageJ v1.41o (National Institutes of Health, Bethesda, USA), and the “threshold” function applied to highlight the colonies with the brightest fluorescence. Any colony hits identified were manually verified by microscopy for both EGFP and mCherry fluorescence.

2.7.7 Bioinformatic analysis of pAG2-A β 42 library hits

Peptide sequences were interrogated for pairwise alignment to the A β 42 sequence using the European Molecular Biology Open Software Suite (EMBOSS) Matcher v2.0u4 (Rice et al. [2000]; default options used, with 3 alternative matches shown; accessed at http://www.ebi.ac.uk/Tools/psa/emboss_matcher/), or for multiple sequence alignment to each other using ClustalW2 v2.1 (Larkin et al. [2007]; default options used; accessed at <http://www.ebi.ac.uk/Tools/msa/clustalw2/>). Open Reading Frame Finder (NCBI; bacterial genetic code used; accessed at <http://www.ncbi.nlm.nih.gov/projects/gorf/>) was used to search for alternate start codons in insert sequences with short open reading frames originating from the Library site start codon. To predict the percentage of sequences from a random DNA insert library that should incorporate an internal stop codon, Equation 2.2 was utilised (Walker et al., 2001).

$$\% \text{ peptides with an internal stop codon} = 1 - \left(\frac{61 \text{ encoding codons}}{64 \text{ total codons}} \right)^n$$

Equation 2.2: n, number of amino acids in peptide.

CHAPTER 3 – BIOSYNTHETIC PRODUCTION OF BIOACTIVE PEPTIDES: ANTIMICROBIALS

3.1 Introduction

3.1.1 Brief history

Antimicrobial peptides (AMPs) are evolutionarily ancient compounds harnessed by a wide variety of organisms to repel microbial infection. Insects, fish, plants, amphibians and mammals all use these gene-encoded antimicrobials as a first line of defence against pathogens (Zasloff, 2002). Bacteria and fungi may also produce AMPs to inhibit competing species (Rossi et al., 2007). While their presence is of no surprise in organisms that lack an adaptive immune system, AMPs are also commonly found in higher organisms (including humans) that possess more intricate defences (i.e. an adaptive immune system). Ever since extracts from human neutrophils in the 1950s revealed low molecular mass proteinaceous components with broad-spectrum antimicrobial activities (Skarnes & Watson, 1957), the AMP field has blossomed – to date, more than 1,700 AMPs and their derivatives have been described (Wang et al., 2009).

The increasing prevalence of antibiotic resistance is a major cause of concern for healthcare systems worldwide. Methicillin-resistant *Staphylococcus aureus*, vancomycin-resistant *Enterococcus faecalis*, carbapenem-resistant *Escherichia coli*, multidrug-resistant *Pseudomonas aeruginosa* – all these examples of “superbugs” are resistant to “last resort” antibiotics, and can lead to increased incidences of mortality amongst the immuno-compromised (Mulvey & Simor, 2009; Fairlamb & Cole, 2011). The majority of new antimicrobials to combat this have been generated by derivatising existing drug scaffolds (Monaghan & Barrett, 2006). Unfortunately, further resistance may rapidly develop due to modifications of existing microbial resistance mechanisms, i.e. those that were effective against the parent scaffold (Fischbach & Walsh, 2009). Entirely new classes of antibiotics are required, but despite this pressing need only the oxazolidinones have come onto the market since 1962 (Rossi et al., 2007). Antimicrobial peptides exhibit a number of characteristics that may make them suitable as a new antibiotic class.

3.1.2 Structure and function

3.1.2.1 Structures

In general, mature AMP sequences consist of between 10 to 60 residues, in which cationic (i.e. lysine and arginine) and hydrophobic (e.g. leucine and isoleucine) amino acids are especially prominent (Zasloff, 2002; Brogden, 2005). While this leads to most AMPs having a positive charge of +2 or greater (Wang et al., 2009), the primary sequences of AMPs are varied, and do not contain conserved motifs, even between closely related species (Peschel & Sahl, 2006). As organisms produce more than one AMP species in certain tissues (Zasloff, 2002), it is thought that such variation is tolerated due to functional redundancy, thus leading to a greater evolutionary flexibility in the classical “arms race” with pathogens (Hancock & Sahl, 2006). What seems to matter more than primary AMP sequence is overall secondary structure. Three main structural classes are apparent: α -helical, with an amphipathic nature (i.e. hydrophobic and cationic side chains segregated to distinct faces of the helix); β -sheet, with a similar amphipathic nature but with multiple disulphide bonds acting as stabilisers; and an extended conformation, lacking clear secondary structure in which certain residues are favoured (e.g. tryptophan [39%] in indolicidin, proline [49%] and arginine [26%] in PR-39) (Gallo & Huttner, 1998; Brogden, 2005). *Figure 3.1* gives examples of peptides from each of these classes. In addition, cyclic peptides in a head-to-tail circular configuration, such as the θ -defensin RTD-1, have also been described (Tang et al., 1999).

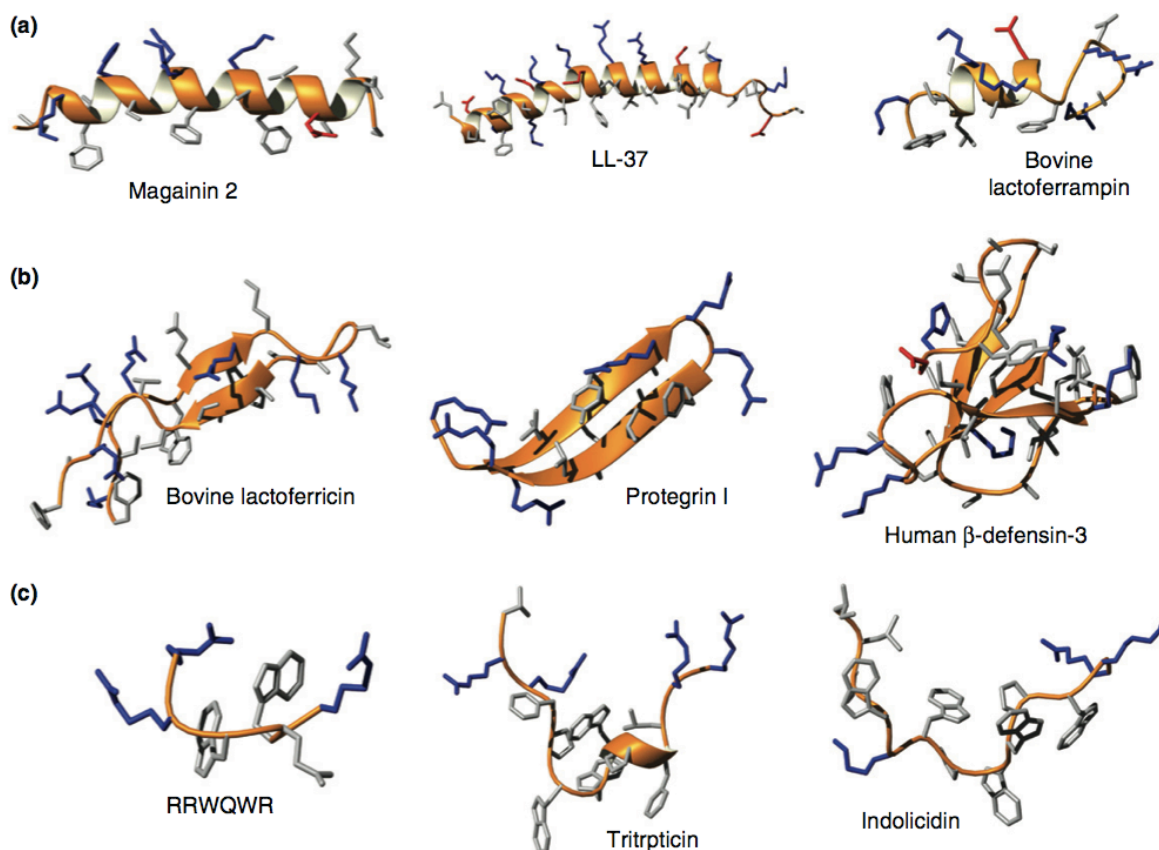


Figure 3.1: Examples of various AMP structures. Common structural classes include **A**, α -helical; **B**, β -sheet; and **C**, extended. Cationic residues are shown in blue, while anionic are shown in red. Selected peptide details (amino acids [aa], origin, Protein Data Base entry [PDB ID]) are as follows: magainin 2 (23 aa, African clawed frog, PDB ID 2MAG), LL-37 (37 aa, human, PDB ID 2K6O), lactoferricin (25 aa, bovine, PDB ID 1LFC), protegrin 1 (18 aa, porcine, PDB ID 1PG1), β -defensin-3 (45 aa, human, PDB ID 1KJ5), tritrpticin (13 aa, synthetic, PDB ID 1D6X), and indolicidin (13 aa, bovine, PDB ID 1G89). Figure modified from Nguyen et al. (2011), PDB accessible at <http://www.pdb.org> (Berman et al., 2000).

What unites many AMP molecules is their amphiphilicity when bound to lipid membranes. Positively charged and hydrophobic residues cluster in distinct regions of an AMP, be it on opposite sides of an α -helix or faces of a β -sheet (Toke, 2005). Of particular relevance to this work is the structure of cathelicidins, a class of mammalian AMPs (12 to 39 aa) so named because of the conserved pre-pro regions that they share prior to the cleavage event that gives rise to the active peptide (Gennaro & Zanetti, 2000; Ramanathan et al., 2002). Expressed in a variety of epithelial and lymphatic tissues (Dorschner et al., 2001; Nizet et al., 2001), a subgroup of these peptides are linear and unstructured in solution, but circular dichroism and nuclear magnetic resonance spectroscopy studies (Yu et al., 2002;

Park et al., 2003) have revealed that they form an α -helix when associated with model phospholipid membranes (e.g. the human LL-37, *Figure 3.1*) (Toke, 2005).

3.1.2.2 Functions

Traditional antibiotics often have well-defined targets: for example, the β -lactams bind bacterial cell wall transpeptidases, while macrolides bind bacterial 50S ribosome subunits to inhibit protein synthesis (Devasahayam et al., 2010). AMPs, on the other hand, predominantly act on less specific targets, such as biological membranes (van't Hof et al., 2001). Such a lytic mode of action has been elucidated by several *in vitro* studies (Saiman et al., 2001; Park et al., 2001; Fantner et al., 2010), including examining the release of the contents of mimetic membranes and microbes themselves, as well as electron microscopy (Toke, 2005). The cytoplasmic membranes from both Gram-negative and Gram-positive bacteria contain anionic head groups in their phospholipid make up, resulting in an overall negative charge (Zaslhoff, 2002). In addition, Gram-positives possess a number of anionic teichoic and lipoteichoic acid moieties in their cell wall, while Gram-negatives exhibit anionic lipopolysaccharides (LPS) on their outer membrane (Hale & Hancock, 2007; van't Hof et al., 2001). Both these features contribute to the electrostatic attraction of positively-charged AMPs, with LPS being especially important with regards to Gram-negatives. These negative moieties provide a passage to the inner cell membrane via a “self-mediated” uptake mechanism, in which the competitive displacement of LPS-associated divalent cations (e.g. Ca^{2+} , Mg^{2+}) allows outer membrane permeabilisation and subsequent AMP translocation to occur (Sawyer et al., 1988).

Once associated with a bacterial membrane, unstructured AMPs typically undergo a conformational change such as the coil-helix transition exhibited by LL-37 mentioned previously (Gennaro & Zanetti, 2000). For α -helical cathelicidins, this conformation is energetically favourable with regards to membrane insertion due to a reduced energy cost imparted by intramolecular NH-CO hydrogen bonding along the peptide backbone (Dathe & Wieprecht, 1999). The realignment of residues into an amphipathic configuration allows the hydrophobic face to insert into the membrane between the phospholipid head groups, while shielding the polar peptide

backbone from the lipid membrane interior (Oren & Shai, 1998; Brogden, 2005). Cationic AMP insertion events are further facilitated by the internally negative membrane potential of a bacterium (Toke, 2005). When a critical threshold peptide concentration is reached (which may be near complete saturation of the membrane [Melo et al., 2009]), the integrity of the bilayer is compromised. Several mechanisms for this have been proposed, none of which are mutually exclusive (see *Figure 3.2*) (Hale & Hancock, 2007). Bacterial death results within minutes due to the loss of membrane polarisation and overall integrity (Brogden, 2005). *In silico* molecular dynamics simulations have been performed to try to further understand such mechanisms of action, and may give additional insight as modelling times progress from nanosecond bursts to microseconds (Bond & Khalid, 2010).

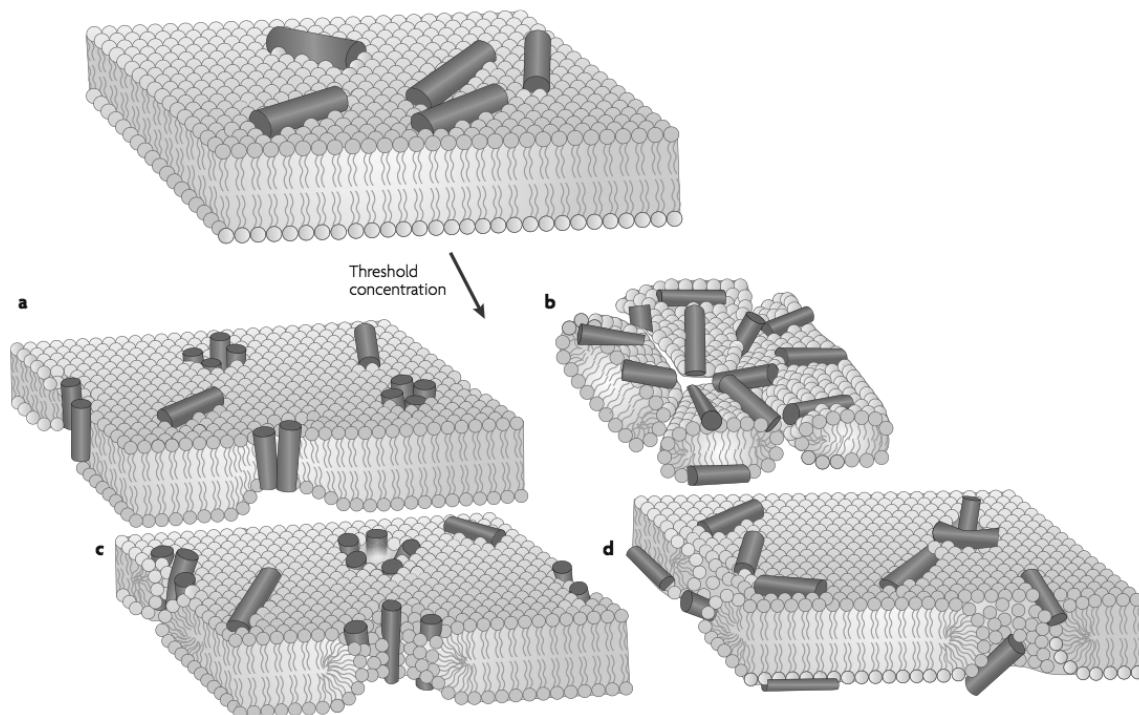


Figure 3.2: Cartoon of proposed mechanisms of AMP-mediated membrane destabilisation. Once a critical AMP threshold concentration has been reached, membrane lysis can occur via **A**, barrel-stave formation by multiple AMP molecules, where multiple hydrophilic faces associate to form a transmembrane pore; **B**, a “carpet” mechanism akin to the effect of a detergent; **C**, toroidal pore formation, where AMPs insert between phospholipid head groups and promote membrane curvature much like the inside of a doughnut; or **D**, disordered toroidal pore formation (less structured with respect to a conventional toroidal pore). Figure modified from Melo et al. (2009).

In contrast to bacterial membranes, host membranes such as those of mammalian cells are more neutral in charge, containing zwitterionic outer leaflets that attract

and bind cationic AMPs poorly (Oren & Shai, 1998). In addition, their membranes are typically less fluid than bacterial membranes due to the “stiffening” properties conferred by cholesterol, further reducing the susceptibility to AMP-mediated lysis (Toke, 2005). It must be emphasised, however, that mammalian cells are not impervious to lytic activity – AMPs frequently exhibit undesirable haemolytic properties. For example, melittin, a 26 aa α -helical peptide from the honeybee, exhibits equal bactericidal and haemolytic activity at its minimum inhibitory concentration (MIC) due to possession of a hydrophobic N-terminal domain coupled with a cationic C-terminal domain (van't Hof et al., 2001). Mutational studies with different AMPs give conflicting evidence as to which property is most important (Shin et al., 2000; Travis et al., 2000; Yang et al., 2003; Hilpert et al., 2006; Chen et al., 2007; Jiang et al., 2008), but adjusting the charge, the proportion of hydrophobic/cationic residues, and the residue distribution over a peptide can abrogate haemolytic activity (Dathe & Wieprecht, 1999). Overall, haemolysis seems to be context dependent for each peptide, although high absolute hydrophobicity leads to greater haemolytic activity due to an increased affinity for zwitterionic membranes (van't Hof et al., 2001; Frecer et al., 2004; Toke, 2005; Matsuzaki, 2009). It appears that there is a subtle balance between amino acid composition, charge, lipophilicity, amphipathicity and structure.

AMPs have also been shown to have alternative (or concurrent) targets to the cell membrane (Brogden, 2005). Clues to this came from the observation that, when used at their MIC, some AMPs were able to kill a target without leading to obvious membrane disruption. For example, analogues of pleurocidin, a 25 aa α -helical peptide from the winter flounder, only lead to membrane depolarisation at concentrations ten-fold higher than its MIC (Patrzykat et al., 2002). It makes sense that intracellular-acting AMPs still exhibit lytic activity at higher concentrations, as they are required to cross the cell membrane to access potential targets. Because of their cationic properties, such targets are often anionic in nature (Hale & Hancock, 2007). Examples of AMPs with intracellular targets include PR-39 (39 aa, porcine origin, rich in proline and arginine, extended structure), which has been shown to inhibit protein synthesis in *E. coli* (Boman et al., 1993); indolicidin (mentioned previously), which caused filamentation of *E. coli* cells by altering

cytoplasmic membrane septum formation (Subbalakshmi & Sitaram, 1998); and buforin II (21 aa, toad origin, α -helix structure), which is able to accumulate in the *E. coli* cytoplasm and bind DNA (Park et al., 1998).

Another alternative function of some AMPs is an ability to modulate a host's immune response (e.g. LL-37; see *Section 1.1.3.1*) (Hancock & Sahl, 2006). As mentioned above, cationic AMPs bind LPS moieties during their self-mediated uptake in Gram-negative bacteria. Bacterial lysis results in liberation of these endotoxins, which provoke a strong systemic immune response in humans that may lead to septic shock (Zanetti, 2005). Among other AMPs, LL-37 has been shown to bind free LPS, and thus limit its ability to stimulate the production of pro-inflammatory cytokines such as tumour necrosis factor- α (Bowdish et al., 2005b). Furthermore, the presence of serum seems to abrogate the membrane lytic activity of LL-37 (Gennaro & Zanetti, 2000; van't Hof et al., 2001). It has therefore been hypothesised that, although LL-37 shows good antimicrobial activity *in vitro*, its primary function *in vivo* may be as an immunomodulator (Bowdish et al., 2005b). In addition to endotoxin binding, LL-37 and other cationic AMPs (such as defensins) may also act as chemokines to attract macrophages; translocate into lymphocytes (presumably via their innate membrane penetration properties) to induce expression of anti-inflammatory genes; and promote wound healing in general (Bowdish et al., 2005a). Because of the above, AMPs that possess an immunomodulatory function in addition to antimicrobial activity have been re-christened "host defence peptides" in order to reflect this increased functional workload (Bowdish et al., 2005a).

There are concerns that the use of AMPs in a clinical setting may lead to the eventual rise of resistance, which may have the undesirable consequence of also promoting resistance to endogenously produced AMPs (Perron et al., 2006). However, because the primary mode of action of many AMPs is the general disruption of the cell membrane, it has been posited that resistance to such a broad mechanism of action is unlikely to arise *in vivo*, i.e. a fundamental membrane redesign would be required (Zasloff, 2002). Some specific resistance mechanisms do exist, however, such as the use of D-alanine amino groups by *Staphylococcus aureus* to reduce the negative charge of surface anionic teichoic acid molecules

(Peschel et al., 1999), and the use of outer membrane proteins as potential target decoys (e.g. adhesin A in the Gram-negative bacterium *Yersinia enterocolitica* [Visser et al., 1996]). The Gram-negative bacterium *Burkholderia cepacia*, an opportunistic pathogen involved in cystic fibrosis, is also extremely resistant to AMPs due to lipid A modifications (Mahenthiralingam et al., 2005). Other resistance strategies include protease production (particularly effective against linear AMPs), the use of capsule polysaccharides to prevent AMP access to the membrane, and increasing the rigidity of the membrane to make it more difficult for AMP insertion (Peschel & Sahl, 2006). Despite this, AMPs have remained effective antimicrobials over millions of years in many species, probably due to a multitude of targets and the production of several different peptides in a single organism (Zasloff, 2002). Consistent with this, it takes many passages of a microbe at sub-lethal AMP concentrations to induce resistance (Gennaro & Zanetti, 2000; Perron et al., 2006): cross-resistance to other AMPs is limited (Samuelsen et al., 2005), and animal models lacking certain endogenous AMPs have shown little increase in susceptibility to infection (Hancock & Sahl, 2006).

In summary, the large number of different activities that AMPs can possess has led to the idea that AMPs may be “dirty drugs”, i.e. hit multiple targets with varying affinities, and such a multifaceted approach could lead to new microbial treatments (Peschel & Sahl, 2006).

3.1.3 Sources

A wide variety of organisms including plants, animals and insects (see *Section 3.1.1*) have been found to produce AMPs, and thus bioprospecting is well suited to their discovery (see *Section 1.2.2*). In addition, fragments of larger proteins have been shown to possess antimicrobial activity, e.g. lactoferricin (*Figure 3.1*), a naturally occurring pepsin digestion product from the 80 kDa lactoferrin, which is itself implicated in the immune response regarding the sequestration of iron (Gifford et al., 2005). Native purification, due to limitations such as source availability or amenability to cultivation (discussed in *Section 1.2.2*) is not economically viable (Hancock, 1997). Further to this, while a large number of mature AMPs are actively secreted, for example magainin 2 on the skin of the frog *Xenopus laevis* (Zasloff, 1987), expression is usually only upregulated upon tissue damage or challenge by

a pathogen (Zanetti, 2005). Chemical synthesis or recombinant production is required instead (see *Section 3.1.5* below). There is, however, one well-known exception – nisin, a lantibiotic (34 aa in length, post-translationally modified with lanthionine groups [thioester-linked alanine side chains]), is naturally produced in large quantities by *Lactococcus lactis*, and has been harnessed as a food preservative against Gram-positive bacteria for many years (Cotter et al., 2005).

3.1.4 Therapeutic potential

AMPs represent a potential new route to dealing with recalcitrant pathogens that current antibiotics are ineffective against. Synthetic α -helical AMPs have been shown to be active against “superbugs” *in vitro* (Tiozzo et al., 1998), and even protozoan parasites such as *Plasmodium falciparum* (the causative agent of malaria) and *Leishmania* species (involved in leishmaniasis) are being examined as targets (Bell, 2011; Luque-Ortega et al., 2008). However, despite initial promise in the 1990s (Hancock, 1997), the use of systemically-administered AMPs in the clinic has failed to materialise to date. For example, two non-ribosomally synthesised bacterial AMPs, gramicidin S (10 aa, produced by *Bacillus brevis*, cyclic structure incorporating D-phenylalanine and ornithine) and the polymyxins (B and E, produced by *Bacillus polymyxa*, cyclic cationic head with a lipid tail) have potent antimicrobial activity, but they both exhibit systemic toxicity, thus constraining their regular use to topical applications in wound creams, eye and eardrops (Marr et al., 2006). Although most AMPs can be modified to reduce haemolytic activity (see *Section 3.1.2* above), no such peptide has currently been approved for systemic administration, perhaps due to potential additional activities such as over-stimulation of the immune system (e.g. induction of apoptosis or mast-cell degranulation) (Bowdish et al., 2005a; Hancock & Sahl, 2006). Despite these issues, topical use is still a worthy goal – gramicidin S and polymyxins are used for treatment of *Pseudomonas aeruginosa* in cystic fibrosis patients, and against *Acinetobacter baumannii* in other nosocomial incidences.

Some AMPs have nearly made it into the clinic (Marr et al., 2006). Pexiganin, a magainin 2 derivative originally developed by Magainin Pharmaceuticals Inc., was denied Phase 3 clinical approval for the topical treatment of diabetic foot ulcers as it could not better the efficacy of current treatments (Lipsky et al., 2008). Omiganan, a

bovine bactenecin derivative (12 aa in length, linear as opposed to the native cyclic form) originally developed by Migenix Inc, also failed to pass Phase 3 clinical trials when used topically for the prevention of local catheter site infection due to trial inconsistencies. Other candidate AMPs are still in the pipeline, but their future is by no means certain (Marr et al., 2006). More work is required to accurately assess *in vivo* function, stability and side effects (Nguyen et al., 2011).

Other alternative uses for AMPs are going forward. As mentioned above, many AMPs, such as LL-37, indolicidin and Bac2A (12 aa, a linear form of bovine bactenecin) possess immunomodulatory functions both *in vitro* and *in vivo* (Bowdish et al., 2005c). Future trials may focus on exploiting such properties, rather than innate antimicrobial activity *per se* (Yeung et al., 2011). One example of such an approach is the synthetic peptide IDR-1 (aka IMX00C1), which while devoid of direct antimicrobial activity *in vitro* is still able to invoke a successful immune response to *Staphylococcus aureus* and *Salmonella typhimurium* through monocytes and macrophages in mouse models (Scott et al., 2007).

Finally, the use of AMPs to coat medical implants or other surfaces is also being explored. Derivatives of indolicidin and Bac2A successfully retained activity when tethered by their C-terminus to a cellulose support (Hilpert et al., 2009). While an analogue of Bac2A failed clinical trials when used to treat catheter-related infections in a gel format (Omiganan, mentioned above), further work by the Hancock laboratory is focusing on attaching similar peptides directly to the catheter itself – initial results seem promising (R. Hancock, unpublished data). Another approach has even looked at using AMPs in an antimicrobial paint (Fulmer & Wynne, 2011), with the aim of employing this on hospital walls or children's toys.

3.1.5 Production

3.1.5.1 Chemical synthesis

AMPs are well suited to solid-phase peptide synthesis (reviewed in *Section 1.1.4.2*). However, cyclic peptides, or those containing multiple disulphide bonds (such as defensins) will require more effort to ensure correct product formation. The cost of synthesis is not inconsequential, and numerous studies have indicated that a cheaper manufacturing alternative is desired (van't Hof et al., 2001; Marr et al.,

2006; Hancock & Sahl, 2006; Rossi et al., 2007; Li et al., 2010). This is especially relevant to the initial research phase, as “good manufacturing practice” compliance of chemical synthesis for final production is not so relevant when the primary goal is to analyse a number of AMP variants cheaply and rapidly. Advances have allowed microarrays of short (6 to 18 aa) peptides to be produced on a cellulose solid support in μmol amounts, with subsequent release into soluble form being possible (Hilpert et al., 2007). This scaled-down approach has been used to synthesise and test single residue mutants of Bac2A (12 aa in length) (Hilpert et al., 2005). However, such an approach is technically demanding, and traditional solid-phase peptide synthesis is required to confirm any findings (Hilpert et al., 2007).

3.1.5.2 Recombinant production

In comparison with chemical synthesis, the input costs of microbe and feed are relatively small for recombinant production of a peptide, AMP or otherwise (Lax, 2010). As long as only natural amino acids are required (i.e. the 20 common L-forms), it is relatively straightforward to build the short coding sequence necessary from synthetic oligonucleotides. This can then be inserted into an expression vector of a common heterologous host, such as the prokaryotes *E. coli* and *L. lactis* or the eukaryotes *Saccharomyces cerevisiae* and *Pichia pastoris* (Li et al., 2010). Secretion of a small number of AMPs into the culture medium has been shown in limited amounts, as has the ability for disulphide bonds to be correctly formed in the oxidising environment of the *E. coli* periplasm (i.e. for the production of defensins) (Raventós et al., 2005). Levels of expression and final production yields, however, depend on the AMP in question. It is also important to note that, because of an AMP's intrinsic antimicrobial activity, it may be toxic to a microbial production host. Therefore AMPs are typically fused to larger carrier proteins in order to prevent or alleviate toxicity during expression, presumably through steric hindrance (Ingham & Moore, 2007). Additionally, the presence of a stable fusion partner could protect labile AMPs from native protease attack (Taguchi et al., 1994; Walker et al., 2001). Fusion partners are often also used as a target for affinity chromatography, thus also affording simple purification of the fusion protein. Common fusion partners include thioredoxin (12 kDa), glutathione S-transferase (26 kDa), and chitin-binding domain (6 kDa) (Baneyx, 1999). Linkers between the AMP and the fusion partner typically incorporate a chemical or enzymatic cleavage

site, which can be utilized to separate the two during subsequent downstream processing *in vitro*. Furthermore, given the small size of most AMPs and the difficulty in detecting low-level expression via SDS-PAGE, the fusion partner allows for expression levels to be more readily assessed. While recombinant production of AMPs may have drawbacks, such as a lack of host toleration of certain inserted nucleic acid sequences and final protein product solubility (Ingham & Moore, 2007), it represents a straightforward and rapid method for the production of crude peptide for initial analysis.

3.1.5.2.1 *K2C18 as a model AMP*

The murine-sourced CRAMP (cathelin-related AMP, expressed in neutrophils and bone marrow) is a typical cathelicidin, being unstructured in solution but forming an amphipathic α -helix when associated with membranes (Gallo et al., 1997). This 38 aa peptide has been shown to have broad-spectrum activity against Gram-negative (*E. coli*, *S. typhimurium*, *P. aeruginosa*) and Gram-positive (*Bacillus subtilis*, *Streptococcus pyogenes*, *S. aureus*) bacteria, as well as some yeast (*Candida albicans* and *Aspergillus fumigatus*), and displays low haemolytic activity (Shin et al., 2000). Furthermore, a series of 18-residue truncations of this AMP also exhibited similar properties to the full-length peptide, with circular dichroism and nuclear magnetic resonance studies confirming their ability to form amphipathic helices upon interaction with membranes (Park et al., 2003). One of these truncations, K2C18 (including a glutamic acid to lysine mutation at position 2), was chosen as a model linear AMP to use as a positive control in this work (Shin et al., 2000).

3.1.5.2.2 *Current recombinant production methods*

A number of approaches exist to produce linear AMPs recombinantly using *E. coli* in conjunction with fusion tags. Recent examples include fusion of the N-terminus of an AMP to a glutathione S-transferase domain (Moon et al., 2006); a split-intein/chitin binding domain (Hong et al., 2010); a hexa-histidine-tagged thioredoxin domain (Krahulec et al., 2010); or the use of a hexa-histidine-tagged small ubiquitin-like modifier domain (Bommarius et al., 2010). While these have been shown to be effective methods, overnight cleavage steps are required using expensive enzymes (like enterokinase) or other reagents to liberate the AMP from its fusion partner.

Furthermore, some of these protocols require additional chromatography steps to remove the cleaving enzyme or fusion partner prior to any further polishing of the desired AMP.

3.1.5.2.3 Use of an inducible, autocleaving tag

RtxA (~500 kDa), a virulence factor produced by the pathogen *Vibrio cholerae*, contains a 23 kDa cysteine protease domain (CPD) that is responsible for the autoprocessing of this large toxin into its active subunits via a cysteine/histidine catalytic dyad (Lin et al., 1999; Satchell, 2007). This 209 residue protein domain has been shown to cleave itself from an upstream (i.e. N-terminal) protein region when induced allosterically by the sugar inositol hexakisphosphate (IP₆) (Prochazkova & Satchell, 2008). While IP₆ is commonly found in eukaryotic cells as a storage unit of phosphate (Zhou & Erdman, 1995), it is lacking in prokaryotes such as *E. coli*. Bogoyo and coworkers exploited these properties by using a hexahistidine (His) tagged version of CPD as a fusion partner for the recombinant production of a number of proteins (ranging from 14 to 35 kDa) in *E. coli* (Shen et al., 2009a). After expression of the full-length fusion product, cleavage of the CPD-His tag could successfully be performed *in vitro* via the addition of IP₆. Such an inducible, autocleaving enzyme tag could be useful for linear AMP production. CPD is an order of magnitude larger than most AMPs and thus should afford some protection from toxicity towards the *E. coli* host. In addition, downstream purification of an AMP could be streamlined in comparison with current recombinant production techniques.

3.1.6 Summary

There is an increased need for new classes of antibiotics to combat the era of antibiotic resistance that clinicians find themselves in. The broad-spectrum, multi-target activities of AMPs indicate a promising potential for their use as therapeutics, as does their proven worth in Nature over millions of years. While the first cohort of AMPs to enter clinical trials did not perform well, this setback should not be discouraging. Rather, it illustrates that additional fundamental research into their structure and function is required, as are larger numbers of candidates. In addition,

more straightforward recombinant production techniques are required in order to facilitate this in a cost-effective manner.

To expand upon current knowledge, the use of *E. coli* to produce a model AMP, K2C18, was investigated. A CPD-His tag was used as an inducible, autocleaving fusion partner, and a number of K2C18 variants were produced. Furthermore, novel mammalian cathelicidin consensus sequences were produced, along with selected derivatives. This recombinant production system was explored, and the AMP products were tested against a number of bacterial strains to demonstrate efficacy, to ascertain their minimum inhibitory concentration, and to try to elucidate structure and function relationships.

3.2 Results

3.2.1 Construction of pET28-CPD

The CPD sequence was cloned into the IPTG-inducible pET28a plasmid between its *Bam*HI and *Xho*I sites so that the C-terminus was in-frame with the hexahistidine tag provided by the vector (see Section 2.4.1.1). Figure 3.3 outlines the properties of the plasmid, as well as the purification protocol designed for this work.

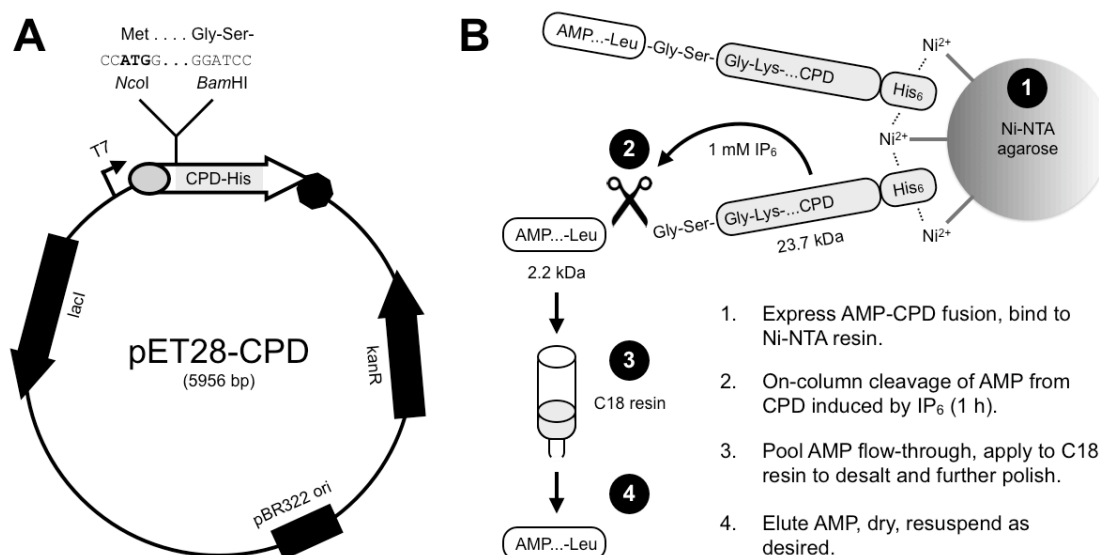


Figure 3.3: The pET28-CPD expression vector and AMP purification protocol. A: The desired AMP coding sequence may be inserted upstream of the CPD domain between the *Nco*I and *Bam*HI sites. The start codon is indicated in bold. The codon immediately preceding the *Bam*HI site must code for a leucine to allow CPD cleavage to occur. **B:** Schematic of the AMP-CPD purification protocol. *lac*I, *lac* operon repressor; T7, RNA polymerase promoter; *kanR*, kanamycin resistance; *ori*, origin of replication.

pET28-CPD is similar in design to the pET2b-CPD_{BamHI} vector created by Shen and co-workers (Shen et al., 2009a), except that the start codon is contained within an *NcoI* site. The wild-type cleavage recognition sequence of Leu[^]Ala-Asp was changed to Leu[^]Gly-Ser in order to incorporate a *BamHI* site to facilitate cloning of DNA inserts ([^] denotes the scissile bond). As only the leucine residue is essential for cleavage (Shen et al., 2009b), it was not expected that such a change would affect CPD activity. The use of an *NcoI* site as the initiation codon requires that the second codon begins with a G, limiting the encoded amino acid to either a valine, alanine, aspartic acid, glutamic acid or glycine. However, the above constraints pose no problem for K2C18, as its N- and C-terminal residues are glycine and leucine respectively.

3.2.2 Expression and activity of recombinant K2C18-CPD

3.2.2.1 Construction of pET28-K2C18-CPD_{W/T}

The coding sequence for K2C18 was cloned into pET28-CPD via the *NcoI* and *BamHI* sites (see Section 2.4.1.3), as shown in Figure 3.3. In parallel to this, a version of pET28-K2C18-CPD was also constructed to incorporate the wild-type (W/T) CPD cleavage recognition sequence of Leu[^]Ala-Asp (named pET28-K2C18-CPD_{W/T}; see Section 2.4.1.2). This version of the plasmid was used to initially explore the production of active K2C18. The N-terminal regions of these three plasmids are compared in Table 3.1.

| Plasmid name | N-terminal region |
|--------------------------------|---|
| pET28-CPD | MGSSHHHHHHSSGLVPRGSHMASMTGGQQMGR <u>GSGKIL</u> ... |
| pET28-K2C18-CPD | MGKKLKKIGQKIKNFFQKL <u>GSGKIL</u> ... |
| pET28-K2C18-CPD _{W/T} | MGKKLKKIGQKIKNFFQKL <u>ADGKIL</u> ... |

Table 3.1: N-terminal regions encoded by pET28-CPD vectors. Amino acids downstream of CPD cleavage site in bold; residues modified by incorporation of *BamHI* site are underlined.

3.2.2.2 Expression of pET28-K2C18-CPD_{WT}

The pET28-CPD and pET28-K2C18-CPD_{WT} vectors were transformed into *E. coli* BL21(DE3) Star cells. Expression was induced by the addition of IPTG (see Section 2.3.1), and protein products analysed using SDS-PAGE after 5 h (Figure 3.4).

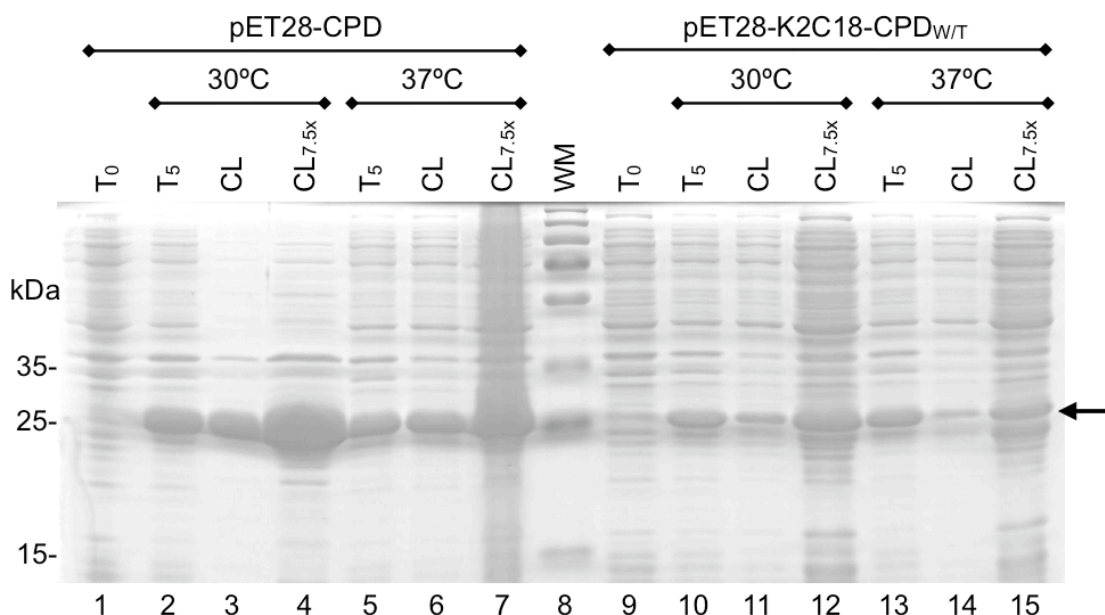


Figure 3.4: Expression from pET28-CPD and pET28-K2C18-CPD_{WT}. SDS-PAGE (glycine 15%) of induction trial in *E. coli*, CPD product expected to run at 26 kDa (arrowed). Equal volumes of cell culture were loaded (normalised to the same density via OD₆₀₀). The clarified lysate (CL) fractions represent soluble protein. T₀, uninduced (whole-cell); T₅, 5 h induction (whole-cell); CL_{7.5x}, 7.5x loaded; WM, molecular mass marker.

K2C18-CPD_{WT} was found to express at both 30°C and 37°C, but not at the same level as CPD alone. In addition, the K2C18-CPD_{WT} fusion protein appears to be less soluble than CPD alone (compare lanes 13 & 14 with 5 & 6). An expression temperature of 30°C was thus chosen for K2C18-CPD_{WT} in order to ensure maximal accumulation of soluble product.

3.2.2.3 Initial purification and activity of recombinant K2C18

The soluble fraction from induced pET28-K2C18-CPD_{WT} *E. coli* cells was applied to Ni-NTA agarose in a batch manner, and K2C18 eluted by autocatalytic CPD cleavage induced by the addition of IP₆ as described in Section 2.4.2. Fractions from the purification process were analysed by SDS-PAGE (Figure 3.5).

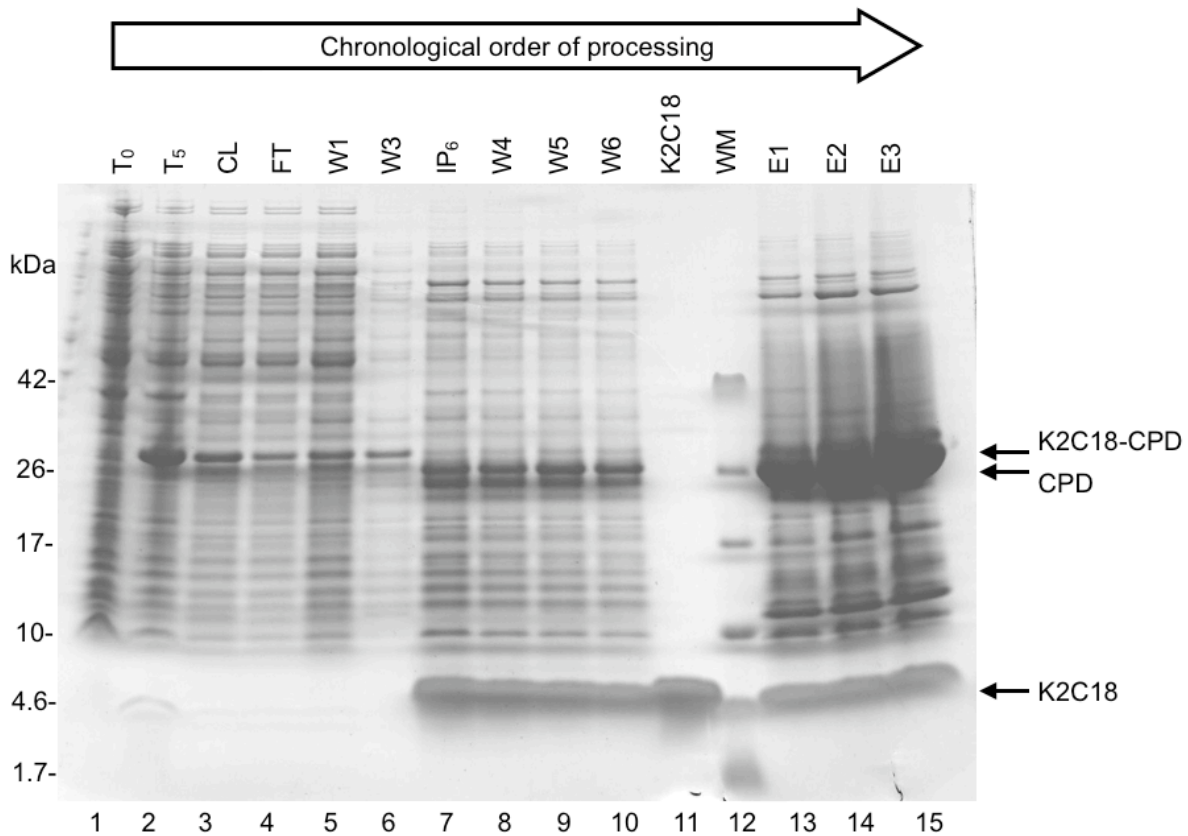


Figure 3.5: Fractions from on-column cleavage of K2C18-CPD_{WT}. SDS-PAGE (tricine 10-20%) of cleavage products shows that IP₆ induces CPD cleavage to give a product band similar in size to chemically-synthesised K2C18. T₀, uninduced *E. coli*; T₅, 5 hours induction at 30°C; CL, clarified lysate; FT, flow-through; W, wash, six in total with W1 to W3 prior to cleavage, W4 to W6 conducted post-cleavage; IP₆, cleavage induced; K2C18, 2.5 µg synthetic peptide; WM, molecular mass marker; E, 250 mM imidazole elution.

K2C18 was successfully cleaved from the CPD tag. Although both chemically-synthesised (lane 11) and recombinantly-produced K2C18 (e.g. lane 7) have a molecular mass of ~2.2 kDa, assuming that *E. coli* removes the initiating methionine after translation (Hirel et al., 1989), they were seen to run near the 4.6 kDa marker. This apparent anomaly is most likely due to the highly charged nature of K2C18 (+7), resulting in an altered migration towards the anode due to ineffective masking of charge by SDS moieties. Mass spectrometry analysis of purified recombinant K2C18 (see *Appendix 1*) confirmed the molecular mass to be ~2.2 kDa, as well as the removal of the initiating methionine.

The cleaved K2C18 does not elute cleanly, being spread over all elution fractions (lane 7 onwards). It appears that this peptide “sticks” to the column after cleavage, requiring several washes to elute. Several higher molecular mass contaminants,

including CPD, also wash off the column after cleavage is induced. This indicates that additional purification is required for further studies; however, the peptide was pure enough for an initial analysis of antimicrobial activity with the appropriate controls.

The fractions shown in *Figure 3.5* were tested for antimicrobial activity against *E. coli* cells by agar radial diffusion assay (see *Section 2.5.2*) and liquid culture assay (see *Section 2.5.3*). *Figure 3.6* shows that antimicrobial activity is observed in the cleaved fractions, i.e. those containing free K2C18.

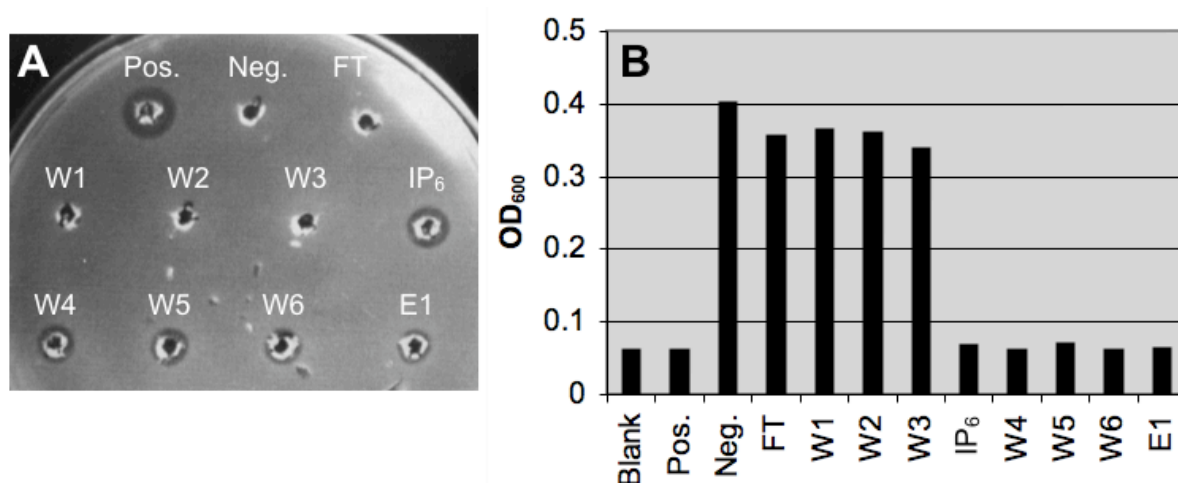


Figure 3.6: Antimicrobial activity of fractions from initial K2C18 purification. 10 μ L of each fraction was added to: **A**, wells punched into *E. coli* TOP10 soft LB agar overlay; and **B**, 90 μ L of $\sim 5 \times 10^4$ cfu/mL *E. coli* TOP10 in LB medium. Photo/OD₆₀₀ reading taken after 18 h incubation at 37°C. Pos., wash buffer including 100 μ M IP₆ and 225 μ M synthetic K2C18; Neg., wash buffer including 100 μ M IP₆; Blank, LB medium only.

Fraction E1 contains 250 mM imidazole, which is toxic to *E. coli* even when diluted to 25 mM in the liquid culture assay (Simonetti et al., 2001). The imidazole in the wash buffer was of sufficiently low concentration to be tolerated, however (see Neg. fraction; 2 mM final concentration when diluted in the assay).

3.2.3 Design, construction and expression of AMP variants

3.2.3.1 Design of AMP variants

The experiments described in *Section 3.2.2* showed that the CPD system can be used to produce K2C18 in an active form. Therefore a number of constructs containing variant AMPs were designed to further explore the potential of this

approach for producing recombinant AMPs for investigation of structure/function relationships.

3.2.3.1.1 *K2C18 analogues*

Using a helical wheel plot (Schiffer & Edmundson, 1967), several variants were designed to explore the relationship between the amphipathic nature of K2C18 and its antimicrobial activity (see *Figure 3.7*). Mutants were created that extended the hydrophobic face of the α -helix through a glycine to leucine mutation at residue 8, i.e. K2C18(G8L), while extension of the cationic face performed by a phenylalanine to lysine mutation at residue 14, i.e. K2C18(F14K). An increase of the positive charge on the cationic face was achieved by substituting glutamine with lysine at residue 9, i.e. K2C18(Q9K), and the effect of utilising arginine instead of lysine as the cationic residue was explored, i.e. RC18. To ascertain the effect of charge dispersal around the cationic face (as opposed to clumping lysine residues together), K2C18-2 was modelled on CRAMP-18-2, a peptide from a previous study (Kim & Cha, 2010). In addition, a mutant of this with an expanded hydrophobic face was designed, i.e. K2C18-2(G8L). As a negative control, a random mutant of K2C18 that did not exhibit antimicrobial activity was chosen, i.e. A6c (isolated in the Tunnacliffe laboratory by Dr Tatsuya Yoshimi, unpublished data).

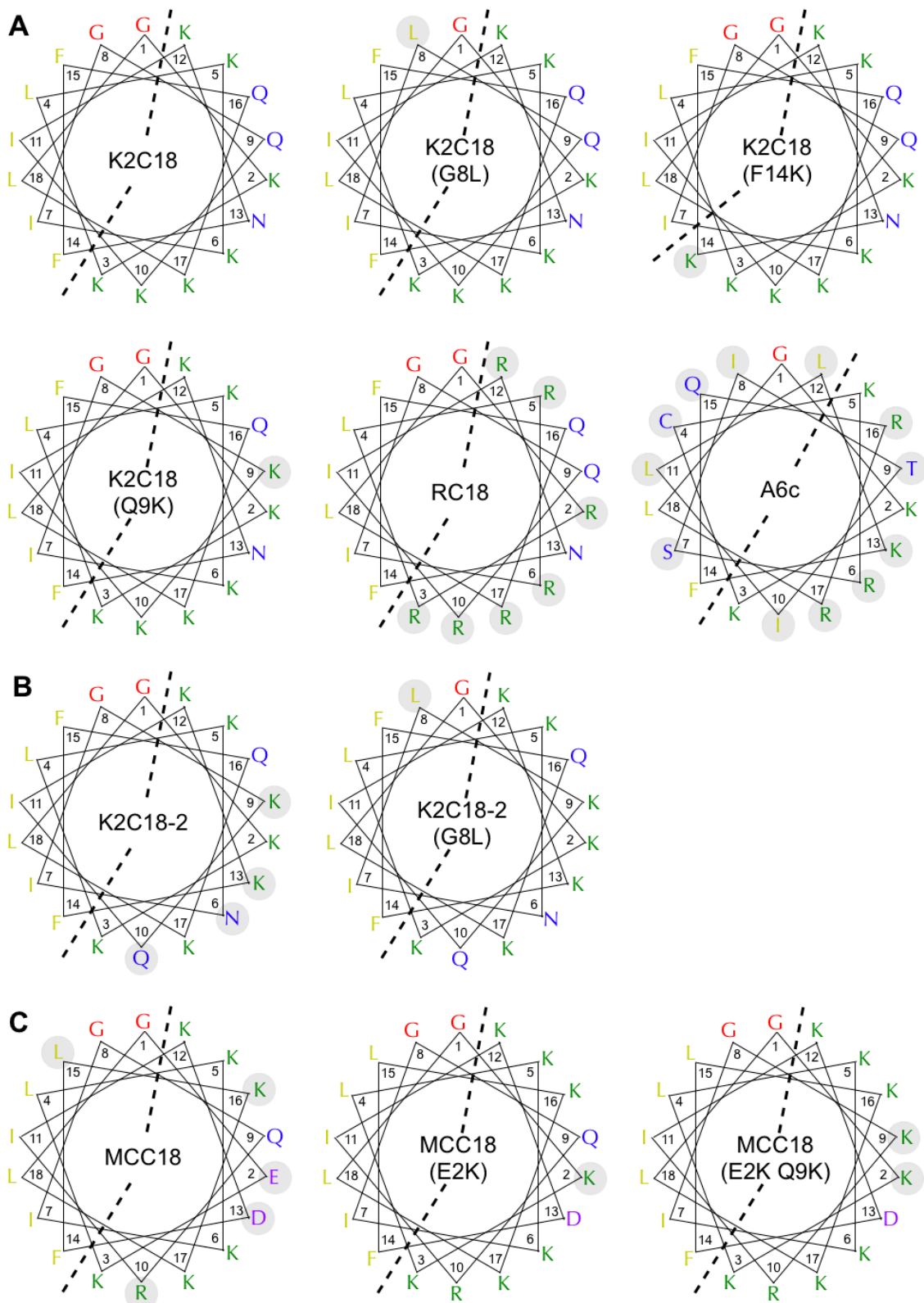


Figure 3.7: Helical wheel projections of designed AMPs. Amino acids mutated from parent AMP are circled. **A**, K2C18 and derivatives. A6c serves as a negative control with an interrupted hydrophobic face. **B**, K2C18-2 and a derivative with dispersed cationic face. **C**, Mammalian consensus MCC18 and derivatives. The boundary between the cationic and hydrophobic faces is indicated by a dashed line, amino acids are colour coded: cationic (green), hydrophobic (yellow), anionic (purple) and uncharged hydrophilic (blue). Differences between variants and the type sequence are shown by a grey background.

3.2.3.1.2 MCC18 analogues

As a further novelty, a consensus of mammalian CRAMP-18 sequences (MCC18) was generated from an alignment of 14 AMP sequence regions that are similar to K2C18 (Table 3.2 and Figure 3.8). Two mutants of MCC18 were created (Figure 3.7), in which the cationic face was altered sequentially in order to increase its charge, with a glutamic acid to lysine mutation at residue 2 (MCC18(E2K)), and a glutamine to lysine mutation at residue 9 (MCC18(E2K Q9K)).

| ADP # | Species | AMP name | Full sequence | Consensus region |
|---------|---------------|-------------|--|------------------------------|
| AP00281 | Mouse | mCRAMP | GLLRKGGKEKI GEK LKKIGQKIKNFFQKLVLPQPEQ | GEK LKKIGQKIKNFFQKL |
| AP00310 | Human | LL-37 | LLGDFFRKSKSEKI GKEFKRIVQR IKDFLRNLVPRTES | GKEFKRIVQR IKDFLRNL |
| AP00366 | Cow | BMAP-27 | GRFKRFRKKFKKLFK KLSPVIPLHLHG | GRFKRFRKK FKKLFK KL |
| AP00371 | Pig | PMAP-37 | GLLSRLRDFLSDR GRR LGEKIERIGQKIKDLSEFFQS | GRR LGEKIERIGQKIKDL |
| AP00418 | Rabbit | CAP18 | GLRKRLRKFRNKI KEK LKKIGQKIQGFVPKLAPRTDY | KEK LKKIGQKIQGFVPKL |
| AP00498 | Rat | rCRAMP | GLVRKGGKEF GEK LRKIGQKIKI EFFQKL ALEIEQ | GEK LRKIGQKIKI EFFQKL |
| AP00676 | Rhesus monkey | RL-37 | RLGNFFRKVKEKI GG LK VGQKIK DFLGNLVPRTAS | GG LKVGQKIK DFLGNL |
| AP00677 | Guinea pig | CAP11 | GLRKKFRKTRKRIQ KLGRKIKGTGRKVKWAWREYGGQIPYPCRI | R KKFRKTRKRIQ KLGRKI |
| AP00678 | Dog | Caninacidin | RLKELITGGQKIG EKIR RIGQR IKDFFKNL QPREKS | GEK IRRIGQR IKDFFKNL |
| AP00680 | Sheep | SMAP-34 | GLFGRLRDSLQR GGQKILE KAERIWCKIKDIFR | G QKILEKAERIWCKIKDI |
| AP00733 | Orangutan | ppyLL-37 | LLGDFFRKAREKI GEEFKRIVQR IKDFLRNLVPRTES | G EEFKRIVQR IKDFLRNL |
| AP00734 | Gibbon | hmdSL-37 | SLGNFFRKARKKI GEEFKRIVQR IKDFLQHLIPRTEA | G EEFKRIVQR IKDFLQHL |
| AP00735 | Leaf monkey | pobRL-37 | RLGNFFRKAKKKI GRGL KKIGQKIKDFLGNLVPRTES | GR GLKKIGQKIK DFLGNL |
| AP00736 | Marmoset | cjaRL-37 | RLGDILQKAREKI EGGL KKLVQKIKDFFGKFPARTES | EG GLKLVQKIK DFFGKF |

Table 3.2: Consensus sequences of mammalian α -helical cathelicidins. Fourteen mammalian cathelicidin sequences with α -helical propensity were collated from the Antimicrobial Peptide Database (ADP, Wang et al. [2009], accessible at <http://aps.unmc.edu/AP/main.php>). An 18-residue consensus region most similar to the K2C18 sequence was aligned (indicated in bold).

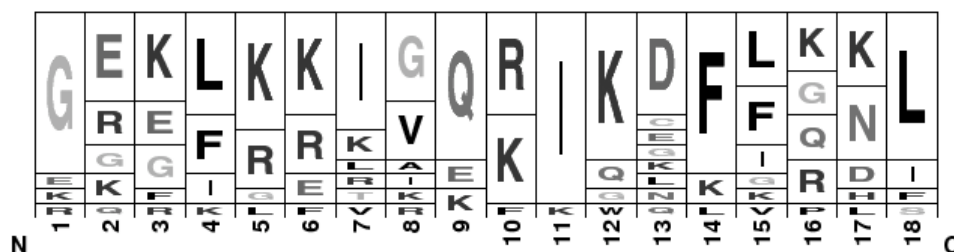


Figure 3.8: Overall consensus sequence of mammalian α -helical cathelicidins. The 14 mammalian cathelicidin consensus regions identified in Table 3.2 were used to generate an 18-residue consensus sequence (MCC18) using WebLogo (Crooks et al. [2004], accessible at <http://weblogo.berkeley.edu/>). The height of the letters indicates their relative frequency. Residue position is indicated.

3.2.3.2 Construction and expression of AMP variants

The coding sequences for the K2C18 and MCC18 variants (Figure 3.7) were “stitched” together from oligonucleotides and cloned into pET28-CPD via the *Nco*I and *Bam*HI sites (see Section 2.4.1.3), as shown in Figure 3.3. Expression in *E. coli* BL21(DE3) Star was induced by the addition of IPTG (see Section 2.3.1), and protein products analysed using SDS-PAGE. Expression levels of some AMP-CPD fusions was found to be poor at 30°C (cf. Section 3.2.2.2). Induction temperature was reduced to 16°C, induction time extended to ~16 h, and *E. coli* BL21 Rosetta(DE3)pLysS cells were trialled in order to alleviate potential translational problems. Rosetta cells express six rare tRNAs to help overcome codon bias (Novagen User Protocol TB009). There was a marked increase in the expression of K2C18-CPD when induced under these conditions (see Figure 3.9).

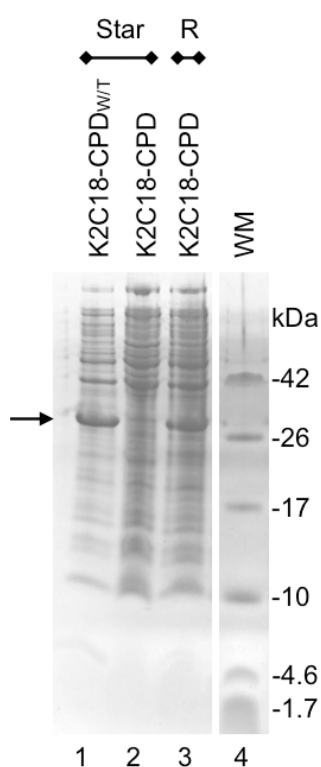


Figure 3.9: pET28-K2C18-CPD_{WT} versus pET28-K2C18-CPD expression. SDS-PAGE (tricine 16%) of *E. coli* whole-cell extract, induced at 16°C for ~18 h at ~225 rpm. Equal volumes of cell culture were loaded (normalised to the same density via OD₆₀₀). K2C18-CPD expression is not seen in BL21(DE3) Star cells (Star), but is in BL21 Rosetta(DE3) pLysS cells (R). K2C18-CPD band size arrowed. WM, molecular mass marker.

While K2C18-CPD_{WT} expresses in BL21(DE3) Star cells (lane 1), K2C18-CPD does not (lane 2). It is unclear why there is such a disparity between the expression levels of K2C18-CPD in pET28-K2C18-CPD_{WT} versus pET28-K2C18-CPD under the conditions used. The only difference between the two constructs is the presence of a *Bam*HI site (GGATCC) immediately downstream of the cleavage-critical leucine in pET28-K2C18-CPD (the wild-type sequence is GCGGAT). This 6 bp change alters the encoded CPD cleavage site from Leu^{Ala}-Asp to Leu^{Gly}-Ser (see *Table 3.1*). While the glycine (GGA) and serine (TCC) codons conferred by the incorporation of the *Bam*HI site are not favoured for translation by *E. coli* (Nakamura et al., 2000), pET28-CPD alone, which also translates the *Bam*HI site, is readily expressed (see *Figure 3.4*). The use of *E. coli* BL21 Rosetta(DE3)pLysS cells at 16°C for ~18 h seems to alleviate this *Bam*HI-related expression problem (lane 3), so these conditions were employed for subsequent experiments.

K2C18 and MCC18 analogues were expressed in 1.5 L culture volumes to increase total protein yields, and purified as per *Section 2.4.2*. *Figure 3.10* shows SDS-PAGE analysis of AMP-CPD fractions after on-column cleavage.

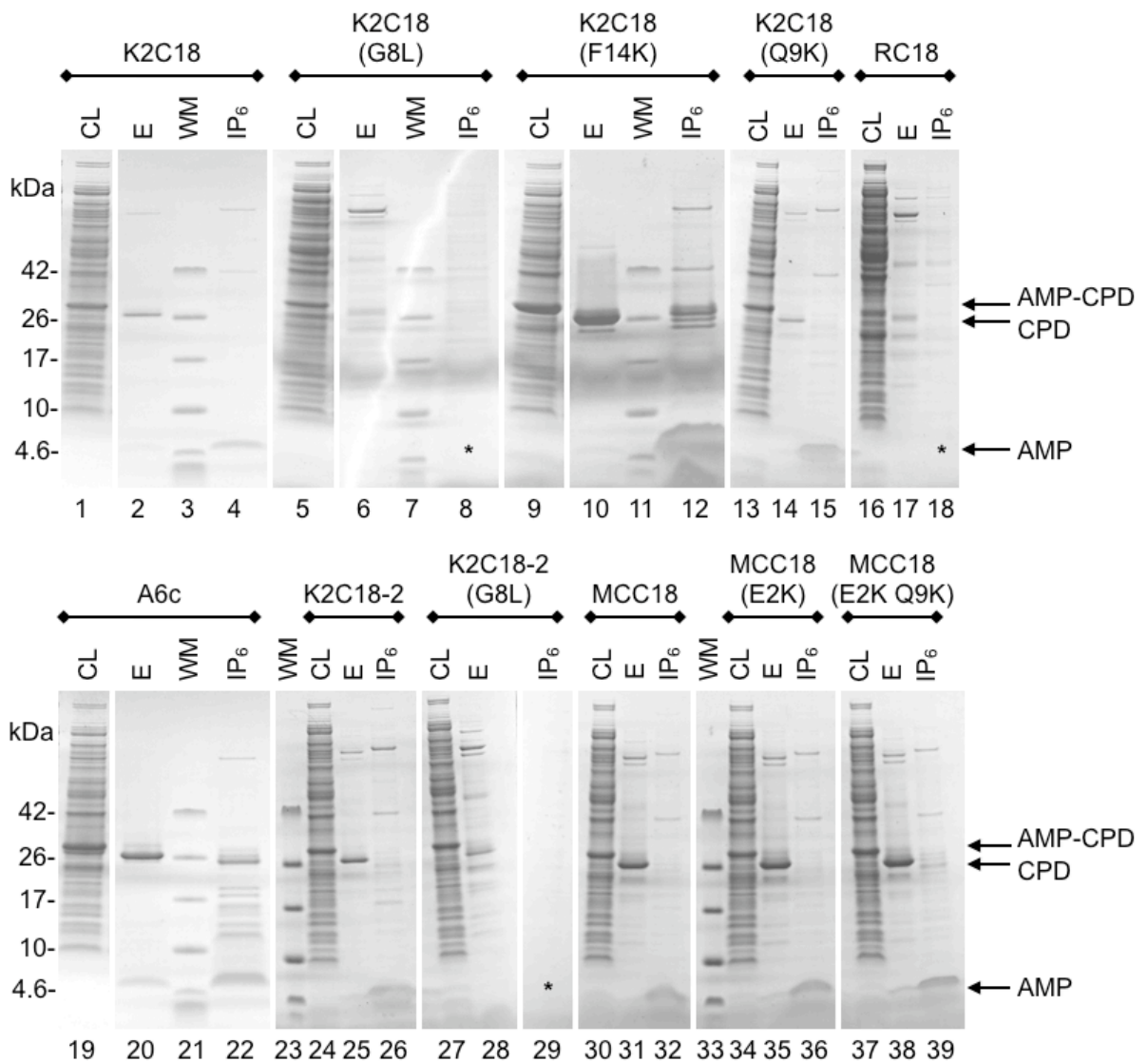


Figure 3.10: Expression and on-column cleavage of K2C18 and MCC18 analogues. SDS-PAGE (tricine 10-20%) of equivalent protein loads of purification fractions from the 11 K2C18 and MCC18 analogues. *, peptides that are poorly expressed; CL, clarified lysate; E, cleaved CPD eluate (250 mM imidazole); WM, molecular mass marker; IP₆, pooled cleaved AMP.

Of the 11 K2C18 and MCC18 analogues that were expressed, 8 gave discrete AMP bands after cleavage was induced. K2C18(F14K) was expressed particularly well (lane 12). However, K2C18(G8L) and K2C18-2(G8L), while seen as a CPD fusion in the clarified lysate fractions (lanes 5 & 27), did not exhibit free AMP bands (lanes 8 and 29). Faint cleaved CPD bands are seen in these samples (lanes 6 & 28), but it may be that the addition of the hydrophobic leucine at position 8 of both peptides leads to aggregation and a poor yield in general. Any free peptide may have been retained on the column during the post-cleavage wash steps, as all AMPs studied appeared to wash off the Ni-NTA column slowly over many column volumes (cf.

Figure 3.5). A6c, while seen in the cleaved AMP fraction (lane 22), aggregated into an unusable form when further purified. Finally, RC18-CPD expression was seen to be poor (lane 16), and no free RC18 post-cleavage was observed. Reasons for this are unclear. Due to these difficulties, the study of K2C18(G8L), K2C18-2(G8L), RC18 and A6c was discontinued.

3.2.4 Purification of AMP variants

3.2.4.1 Polishing of AMPs by C18 resin

The seven cleaved and soluble K2C18 and MCC18 analogues (see Figure 3.10) were further purified using hydrophobic C18 resin as per Section 2.4.3.1. This served to desalt the peptides prior to vacuum drying, as well as removing larger protein contaminants that, due to a greater number of hydrophobic residues, remained bound to the C18 resin during peptide elution over an acetonitrile step gradient. Figure 3.11 shows this process for a typical purification, as well as samples of all the recombinant AMPs that were purified.

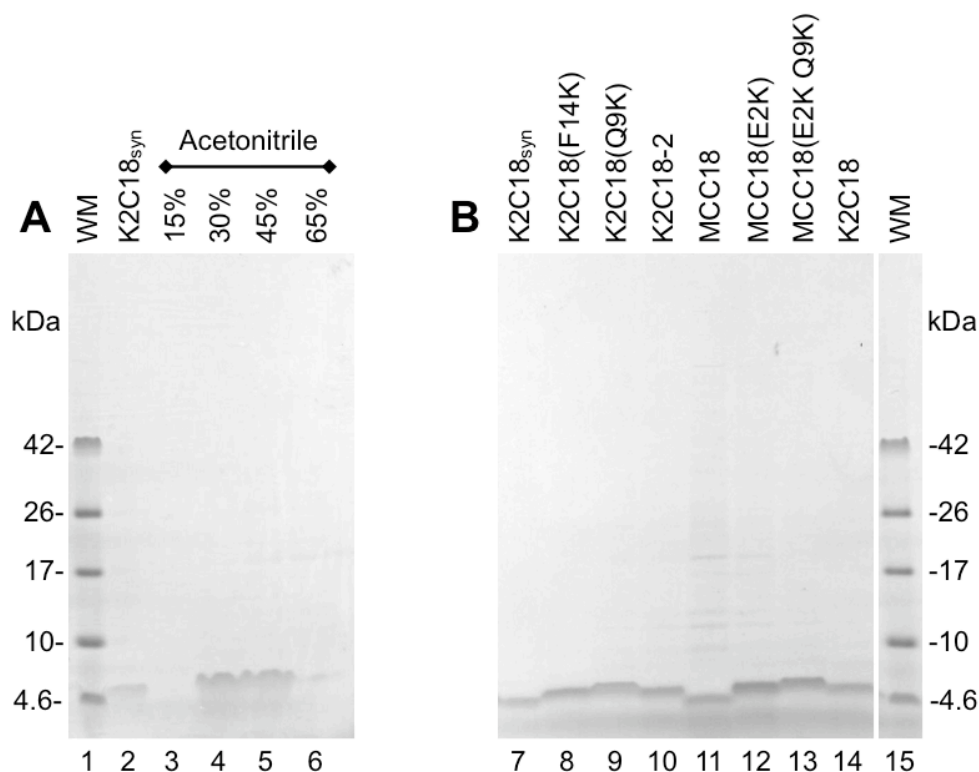


Figure 3.11: C18 polishing of K2C18 and MCC18 analogues. SDS-PAGE (tricine 10-20%) of **A**, acetonitrile elutions of K2C18 from C18 resin, from which pure fractions were pooled (30% and 45% in this example); and **B**, purified K2C18 and MCC18 analogues after drying and resuspension in water. K2C18_{syn}, 0.5 µg chemically-synthesised K2C18.

All peptides appear to be relatively pure as observed by SDS-PAGE with Coomassie staining. MCC18 appears to have a low level of contamination, visible as faint protein bands running at higher molecular masses (lane 11). MCC18(E2K) may also be slightly contaminated (lane 12). Amino acid analysis confirms the presence of contaminating protein through the observed amino acid ratio diverging from that expected (see *Section 3.2.4.2* below, and *Appendix 2*). In both cases, however, the AMP band is by far the most dominant, and the level of purity should be sufficient for activity studies. The AMP bands do not run at the same molecular mass; those that are more highly charged, such as MCC18(E2K Q9K) with a charge of +8 (lane 13), run higher than those with a lower net charge i.e. MCC18 with a charge of +5 (lane 11). This further confirms the anomalous SDS-PAGE running mass of these peptides as observed previously in *Section 3.2.2.3*.

3.2.4.2 Assessment of purified AMP concentration and purity

Samples of the K2C18 and MCC18 analogues were submitted for amino acid analysis (as per *Section 2.3.4*) in order to determine their concentration and to assess their purity. As an example, a comparison of recombinant and chemically-synthesised K2C18 is shown in *Figure 3.12*. The remainder of the amino acid analysis results are shown in *Appendix 2*.

| A AAA Results K2C18 | | | | B AAA Results K2C18 | | | |
|---|----------------|-----------------|------------------------------------|---|----------------|-----------------|------------------------------------|
| Sample: <i>Integer fit of measured mole ratios to expected values</i> | | | | Sample: <i>Integer fit of measured mole ratios to expected values</i> | | | |
| Amino acid | Expected value | Observed value | Closeness of fit to expected value | Amino acid | Expected value | Observed value | Closeness of fit to expected value |
| Cys | 0 | 0.00 | - | Cys | 0 | 0.00 | - |
| Asp | 1 | 1.05 | better than 5% | Asp | 1 | 1.06 | within 5-10% |
| Thr | 0 | 0.00 | - | Thr | 0 | 0.00 | - |
| Ser | 0 | 0.00 | - | Ser | 0 | 0.00 | - |
| Glu | 2 | 2.01 | better than 5% | Glu | 2 | 2.05 | better than 5% |
| Gly | 2 | 2.00 | better than 5% | Gly | 2 | 2.13 | within 5-10% |
| Ala | 0 | 0.00 | - | Ala | 0 | 0.00 | - |
| Val | 0 | 0.00 | - | Val | 0 | 0.00 | - |
| Met | 0 | 0.00 | - | Met | 0 | 0.00 | - |
| Ile | 2 | 1.93 | better than 5% | Ile | 2 | 1.90 | within 5-10% |
| Leu | 2 | 1.95 | better than 5% | Leu | 2 | 1.93 | better than 5% |
| Norleu std | 0 | 0.00 | - | Norleu std | 0 | 0.00 | - |
| Tyr | 0 | 0.00 | - | Tyr | 0 | 0.00 | - |
| Phe | 2 | 1.98 | better than 5% | Phe | 2 | 1.88 | within 5-10% |
| His | 0 | 0.00 | - | His | 0 | 0.00 | - |
| Lys | 7 | 7.08 | better than 5% | Lys | 7 | 7.05 | better than 5% |
| Arg | 0 | 0.00 | - | Arg | 0 | 0.00 | - |
| Pro | 0 | 0.00 | - | Pro | 0 | 0.00 | - |
| Trp | 0 | 0.00 | (not determined) | Trp | 0 | 0.00 | (not determined) |
| Total | 18 | residues | | Total | 18 | residues | |
| Total sample | 0.421 | nmoles | | Total sample | 0.225 | nmoles | |
| | 0.91 | ug | | | 0.47 | ug | |
| Concentration | 21.05 | nmoles/ml | | Concentration | 11.26 | nmoles/ml | |
| | 45.25 | ug/ml | | | 23.65 | ug/ml | |

Figure 3.12: Amino acid analysis of recombinant and chemically-synthesised K2C18. Ion exchange ninhydrin analysis of hydrolysates of **A**, recombinant K2C18; and **B**, chemically-synthesised K2C18 peptide. The closeness of fit between the expected and observed amino acid mole ratios is used as a qualitative indication of purity; unexpected amino acids, i.e. those not in the peptide, were excluded from the analysis. Analyses were carried out in duplicate, single runs shown.

Recombinant K2C18 is of an equivalent, if not greater purity than the chemically-synthesised version (commercially obtained, 88.88% purity according to manufacturer's HPLC analysis). If the majority of observed values are within 10% of the expected values, it indicates a peptide is relatively pure. The seven purified K2C18 and MCC18 analogues met this criteria, including the slightly contaminated MCC18 and MCC18(E2K) from *Figure 3.11*.

3.2.4.3 Yields of purified AMPs

The yield of each AMP from the CPD purification protocol was calculated from the amino acid analysis results, and is summarised in *Table 3.3*.

| Name | Sequence (18 aa) | MW | μM | μL | μg | Yield/L |
|----------------|---|--------|---------------|---------------|---------------|--------------------|
| K2C18 | GKKLKKIGQKIKNFFQKL | 2146.6 | 1426 | 68 | 208 | 139 μg |
| K2C18(F14K) | GKKLKKIGQKIKN K FQKL | 2127.6 | 9024* | 80 | 1536 | 1024 μg |
| K2C18(Q9K) | GKKLKKIG K KIKNFFQKL | 2146.7 | 955 | 78 | 160 | 107 μg |
| K2C18-2 | GKKL K NIG K QIK K FFQKL | 2146.6 | 1061 | 78 | 178 | 118 μg |
| MCC18 | G E KLKKIG Q R I K D FL K KL | 2142.6 | 2429 | 48 | 250 | 167 μg |
| MCC18(E2K) | G K KLKKIG Q R I K D FL K KL | 2141.7 | 1107 | 78 | 185 | 123 μg |
| MCC18(E2K Q9K) | G K KLKKIG K R I K D FL K KL | 2141.7 | 849 | 78 | 142 | 95 μg |

Table 3.3: Purification yields of K2C18 and MCC18 analogues. Mutated residues indicated in bold. MW, molecular mass in Daltons; μM , peptide concentration as determined by amino acid analysis; μL , resuspended peptide volume; μg , total yield from 1.5 L *E. coli* culture; *, K2C18(F14K) was not polished using C18 resin, rather filtered through 10 kDa MWCO membrane and dialysed against 0.1 M ammonium acetate before lyophilisation and resuspension (Section 2.4.3.2).

Yields of peptide per litre of induced *E. coli* culture ranged from ~100 μg for MCC18(E2K Q9K) to ~1 mg for K2C18(F14K), but more typically fell into the 100 to 150 $\mu\text{g/L}$ range. As an example, the expression of K2C18-CPD in *E. coli* BL21 Rosetta(DE3)pLysS cells was estimated at 20 mg/L (see Figure 3.9). Given that K2C18 comprises ~8% of the predicted mass of the fusion protein, a maximal yield of approximately 1.7 mg/L of AMP was estimated. The actual yield of 139 $\mu\text{g/L}$ of K2C18 represents approximately 8% of this. Other peptides, such as K2C18(F14K), were expressed at higher levels (see Figure 3.10), and hence lead to a greater final yield. Although these final yields were not high, enough AMP was purified to assay for antimicrobial activity.

3.2.5 Measurement of antimicrobial activity

Laboratory strains of the Gram-negative bacteria *E. coli* and *P. putida*, the Gram-positive bacterium *B. subtilis*, and the yeast *C. albicans* (see Section 2.5.1; in exponential-phase growth) were challenged with the K2C18 and MCC18 analogues at final concentrations ranging from 16 μM to 0.3 μM (~34 $\mu\text{g/mL}$ to ~0.7 $\mu\text{g/mL}$). Minimum inhibitory concentration (MIC) was determined as the concentration of AMP that prevented microbial growth after an overnight incubation (Table 3.4). MIC is distinct from minimum bactericidal concentration (MBC), which requires

confirmation of cell death (i.e. not just the cessation of growth). The physical characteristics of each AMP i.e. charge (Q), cationic face angle of the α -helix (Φ), mean hydrophobicity (H) and relative hydrophobic moment (μ_{Rel}) were also determined for each peptide as per *Section 2.4.4*. These are common descriptors used to compare α -helical AMPs (Dathe & Wieprecht, 1999).

| Name | Q | Φ | H | μ_{Rel} | MIC <i>E. coli</i> | MIC <i>P. putida</i> | MIC <i>B. subtilis</i> | MIC <i>C. albicans</i> |
|-------------------|----|--------|-------|--------------------|--|-------------------------|--|--|
| MCC18(E2K Q9K) | +8 | 200° | -0.40 | 0.62 | 1.0 μM | <0.3 μM | 2.0 μM | 4.0 μM |
| K2C18(Q9K) | +8 | 200° | -0.33 | 0.59 | 0.8 μM | 0.8 μM | 5.0 μM | 4.0 μM |
| MCC18(E2K) | +7 | 200° | -0.38 | 0.59 | 1.0 μM | 1.5 μM | 5.0 μM | 8.0 μM |
| K2C18-2 | +7 | 200° | -0.31 | 0.58 | 1.0 μM^2 | 2.0 μM | 8.0 μM | 4.0 μM |
| K2C18 (synthetic) | +7 | 200° | -0.31 | 0.56 | 0.6 μM (3.1 μM) ¹ | 1.0 μM | 6.0 μM (3.1 μM) ¹ | 4.0 μM (6.3 μM) ¹ |
| K2C18 (recomb.) | +7 | 200° | -0.31 | 0.56 | 1.0 μM | 2.0 μM | 16.0 μM | 16.0 μM |
| MCC18 | +5 | 200° | -0.35 | 0.56 | 2.0 μM | >16 μM | >16 μM | 16.0 μM |
| K2C18(F14K) | +8 | 220° | -0.40 | 0.53 | 8.0 μM | 1.5 μM | >16 μM | >16 μM |

Table 3.4: Structural characteristics and antimicrobial activities of K2C18 and MCC18 analogues, ranked by relative hydrophobic moment. Minimum inhibitory concentration (MIC) was determined as the concentration of peptide that completely inhibited microbial growth after 18 h (bacteria) or 23.5 h (*C. albicans*). Values represent duplicate average. MICs reported in the literature are shown in italics: 1, Shin et al. (2000); 2, Kim & Cha (2010), no MIC data reported. Structural parameters: Q, charge; Φ , angle subtended on the helix by the cationic face; H, mean hydrophobicity; μ_{Rel} , relative hydrophobic moment.

The K2C18 and MCC18 analogues, expressed and purified using the CPD system, all displayed varying degrees of antimicrobial activity. While recombinant and synthetic K2C18 had a very similar effective concentration against *E. coli* and *P. putida*, a ~2.5 fold difference in MIC was observed for *B. subtilis* and *C. albicans*. The reason for this discrepancy is unclear; racemisation or oxidation of residues in one of the peptides may be an explanation, but the most susceptible amino acids (e.g. methionine, cysteine, histidine, tryptophan and aspartic acid) are not found in K2C18 (Palasek et al., 2007; Sewald & Jakubke, 2009). Of the K2C18 variants, K2C18(Q9K) was the most effective overall against the microbes tested. MCC18 was similar in potency to K2C18 against *E. coli*, but was ineffective at the concentrations tested against *P. putida*. Interestingly, the MCC18(E2K) and

MCC18(E2K Q9K) variants were more effective against *P. putida*, *B. subtilis* and *C. albicans* than their parent peptide MCC18.

Although the active AMP variants all exhibit similarly low MIC values (i.e. same order of magnitude), some structure/function relationships are indicated. Peptides with a larger relative hydrophobic moment, i.e. a greater amphipathic propensity, show more potent activity. MCC18(E2K Q9K) exhibited the greatest antimicrobial activity in comparison to the other analogues tested, while K2C18(F14K) showed the lowest. A higher overall charge also correlated with a lower observed MIC.

3.3 Discussion

While solid-phase peptide synthesis remains the standard for commercial-level synthesis of bioactive peptides (Lax, 2010), there is still a place for recombinant production in a research and development setting. For AMPs, small-scale and rapid creation of a peptide and a variety of analogues for structure/function analysis is desirable (Hancock & Sahl, 2006). The use of the inducible, auto-cleaving CPD tag, first described by Shen et al. (2009a), is shown in this work to adequately suit this purpose. K2C18, a truncation of the extensively studied murine α -helical cathelicidin CRAMP (Gallo et al., 1997; Shin et al., 2000; Yu et al., 2002), was successfully produced and shown to exhibit antimicrobial activity. Several variants, including a novel consensus sequence of CRAMP-like mammalian AMPs, were also purified, with mutations made to the primary amino acid sequences in an attempt to ascertain basic information on structure/function relationships.

The CPD purification approach is cost-effective, as the fusion tag is also the cleaving enzyme that releases free AMP. This negates the need to purchase a separate protease, and the cleavage-inducing compound IP₆ is inexpensive (£9 per gram, Sigma-Aldrich Company Ltd.). In addition, because a 1:1 ratio of AMP to CPD is inherent, complete cleavage can occur rapidly within 1 h or less. This gives the CPD system an advantage over other recent recombinant AMP production techniques. The glutathione S-transferase tag approach utilised by Moon et al. (2006) requires buffer exchange prior to Factor Xa cleavage treatment over two days; the split-intein/chitin binding domain of Hong et al. (2010) requires overnight

incubation in β -mercaptoethanol to induce cleavage; the cleavage of the thioredoxin tag used by Krahulec et al. (2010) requires a day of incubation at 37°C with an expensive enterokinase; and the small ubiquitin-like modifier domain harnessed by Bommarius et al. (2010) exhibited issues with the specificity and dissociation of the cleaving enzyme sumoase.

Expression of the AMP-CPD fusions in *E. coli* did not appear to induce toxicity in the host cells. Although a slight depression in growth rate was generally observed in comparison to cells expressing pET28-CPD alone (data not shown), this varied for each AMP expressed, and showed no correlation with the MIC values that were finally determined. This lack of AMP-fusion toxicity during intracellular expression is in line with that seen in previous recombinant production methods (Moon et al., 2006; Bommarius et al., 2010).

The use of the modified CPD cleavage site, mutated by the introduction of a *Bam*HI site from Leu[^]Ala-Asp to Leu[^]Gly-Ser, did not inhibit CPD cleavage. This confirms the work of Shen et al. (2009a). However, the expression level of identical AMP fusions was influenced by this change. The poor expression from pET28-K2C18-CPD (*Bam*HI cleavage site) in *E. coli* BL21(DE3) Star cells in comparison to pET28-K2C18-CPD_{WT} (wild-type cleavage site) is not readily explainable, as pET28-CPD, which possesses the modified cleavage site in the correct reading frame, expresses well. However, using the wild-type cleavage sequence in the production of additional peptides did not guarantee significant expression (see *Section 4.2.6.1*). Use of *E. coli* BL21(DE3)pLysS Rosetta cells, along with a reduction of induction temperature and extended incubation time, alleviated the expression problem somewhat. It appears that, while the CPD domain is soluble and readily expressed, N-terminal fusions alter this in a context-dependent manner. Indeed, some AMP-CPD fusions were not amenable to purification at all. These included RC18, which did not express; and K2C18(G8L) and K2C18-2(G8L), which expressed but did not produce free AMP. A6c, while expressed and readily cleaved by CPD, aggregated into an unusable form.

The yields of purified K2C18 and MCC18 variants ranged from 0.1 to 1 mg/L. This represents 5 to 50% of the total estimated production level of 1.7 mg/L, i.e.

~20 mg/L AMP-CPD fusion is produced with the AMP portion forming ~8% of this. Initial purification protocols involved filtering the cleaved AMP through a 10 kDa MWCO membrane to remove larger protein contaminants, followed by dialysis to desalt and subsequent lyophilisation. However, allowing the peptides to adsorb to two sets of membranes reduced yields to an undesirable level. Apart from K2C18(F14K), which was strongly expressed, the yields of other peptides using this approach were less than 100 µg/L (data not shown). The use of hydrophobic C18 resin, commonly used in HPLC for separating peptides, overcame this problem by allowing for desalting and contaminant removal in one step. Production yields might further be improved upon by harnessing FPLC and HPLC for the Ni-NTA and C18 purification steps respectively. However, the aim of this work was to produce a system that produces linear AMPs such as K2C18 in a cheap, straightforward and timely manner. The use of Ni-NTA resin and C18 disposable columns is in line with this approach.

While the AMP yields (0.1 to 1 mg/L) seem low, they are commensurate with the yields reported in previous studies that utilised *E. coli*. Moon et al. (2006) reported a yield of 0.3 mg/L for LL-37 (37 aa) expressed as a fusion to the C-terminus of glutathione S-transferase; Hong et al. (2010) did not report their yield of dermcidin (48 aa, an anionic AMP) from a fusion to the C-terminal of a split-intein/chitin binding domain, but did state that they purified a total of 10 mg of AMP. Krahulec et al. (2010) used a fermentor (as opposed to shake flask incubation used in this work and the other previously mentioned studies) to achieve 40 mg/L for LL-37 expressed as a fusion to the C-terminal of thioredoxin. Bommaris et al. (2010) also used a fermentor to produce IDR-1 (13 aa, an artificial host defense peptide) from a C-terminal fusion to a small ubiquitin-like modifier domain at a yield of 45 mg/L, and E6 (12 aa, a bacterenecin analogue) in the same manner at a yield of 8 mg/L. However, the use of a fermentor is not amenable to the straightforward production of multiple AMPs in parallel.

Amino acid analysis was carried out on the purified AMPs to determine the net peptide content (i.e. concentration) of the resuspended peptide solutions. This is important as a dried peptide still contains residual salt, so calculating peptide concentration by resuspended dry-weight over-estimates the actual peptide

concentration (Bommarius et al., 2010). While HPLC analysis was not conducted on the recombinantly purified peptides, SDS-PAGE and amino acid analysis qualitatively showed that the AMPs were relatively pure. Although MCC18 exhibited some contamination, it was deemed sufficiently pure for use in the antimicrobial activity assay.

The MICs of K2C18 in this work are comparable to those determined by Shin et al. (2000). Discrepancies may be explained by different experimental protocols. In this work, 90 μL of 5×10^4 cfu/mL bacteria ($\sim 4,500$ cells) or 6×10^3 cfu/mL fungi (~ 540 cells) were used as starting cultures for the antimicrobial assay; Shin and co-workers used 100 μL of 2×10^6 cfu/mL bacteria ($\sim 20,000$ cells) and 2×10^4 cfu/mL fungi ($\sim 2,000$ cells). Standardisation of antimicrobial assays is lacking in the antimicrobial field, with each lab adopting their own approach (Wiegand et al., 2008). However, as long as assays are carried out in parallel under the same conditions, comparisons can be made between different AMPs under examination.

The antimicrobial activity of the various purified K2C18 and MCC18 analogues confirms and extends previous studies of AMP structure/function relationships. Shin et al. (2000) showed K2C18 to have a potent MIC in the low μM range against a number of Gram-negative and Gram-positive bacteria, as well as against the fungus *C. albicans*. Extending the cationic face angle (K2C18(F14K)) appears to reduce overall effectiveness; in magainin variants, similar mutations lead to an abrogation of antimicrobial activity, alongside a reduction in overall hydrophobicity (Dathe & Wieprecht, 1999). It has been posited that this is due to peptides with wide cationic faces exhibiting a greater stability for associating with negative phospholipid heads, reducing membrane insertion and disruption; complete insertion becomes disfavoured due to electrostatic repulsion between adjacent peptide cationic faces (Toke, 2005). Furthermore, Park et al. (2003) showed that the removal of both phenylalanines (positions 14 and 15) from the K2C18 analogue CRAMP-18 by alanine substitution led to a decrease in antimicrobial activity against *E. coli*, indicating a key role for these hydrophobic moieties (Park et al., 2003). This broadly agrees with the results seen in this work, although K2C18(F14K) still retains a similar activity to K2C18 against *P. putida*.

The analogue K2C18-2 was designed to have the same amino acid composition as K2C18 but with the lysine residues interspersed (rather than clustered) around the cationic face. This was predicted to enhance the interaction of the cationic face of the α -helix with bacterial anionic phospholipids, and hence further promote membrane insertion (Kim & Cha, 2010). While Kim and Cha (2010) found that CRAMP-18-2 (similar to K2C18-2) exhibited a more potent lytic activity against *E. coli* than CRAMP-18 (similar to K2C18), the results presented here did not show any such improvement in MIC. This may be due to the different nature of the protocols, with the liquid culture assay used in this work proceeding overnight rather than for 30 min at a fixed concentration as per Kim & Cha (2010).

MCC18, a consensus sequence derived from 14 mammalian cathelicidin K2C18-like sequences (the 18-residue consensus region from LL-37 correlated well with a previous investigation into its minimum active region [Li et al., 2006]), was examined for antimicrobial activity along with two variants. MCC18 possesses five amino acid changes in comparison to K2C18, of which three are non-homologous: glutamic acid at position 2 (replacing a lysine) and aspartic acid at position 13 reduce the overall charge of the peptide, while lysine at position 16 replaces the uncharged glutamine. MCC18 shows antimicrobial activity at a MIC similar to K2C18 against *E. coli*, but is ineffective against *P. putida*, *B. subtilis* and *C. albicans* over the concentration range tested. While the Gram-negative bacterium *P. putida* has similar membrane compositions as *E. coli* (Pinkart & White, 1997; Ruiz et al., 2006), the ability for *P. putida* to modify its membrane fatty acid and protein content in response to membrane-active substances may be responsible for the lack of effect seen by MCC18 at the concentrations tested (Heipieper & de Bont, 1994). However, when the overall charge of MCC18 (+5) is increased, i.e. to +7 in MCC18(E2K) and +8 in MCC18(E2K Q9K), activity is seen against *P. putida*. This is in line with previous studies regarding the importance of cationic charge (Dathe & Wieprecht, 1999; Jiang et al., 2008; Matsuzaki, 2009). However, the charge of these AMPs is not the only parameter to change between the MCC18 analogues. Overall hydrophobicity and relative hydrophobic moment are also altered concurrently, as it is very difficult to vary one parameter in isolation from others (Dathe & Wieprecht, 1999). This increase in relative hydrophobic moment, due to additional lysine residues on the cationic face, has been shown to enhance the

activity of an AMP against Gram-positive bacteria, as well as other microbes in general (Matsuzaki, 2009). Antimicrobial activity against all microbes thus increases for the MCC18 analogues, as their relative hydrophobic moment increases from 0.56 to 0.62.

The presence of the anionic aspartic acid at position 13 in the MCC18 consensus peptide is interesting, as this residue is situated in the middle of the AMP's cationic face (see *Figure 3.7*). As mentioned previously, a dispersed cationic charge is also favourable to activity (Kim & Cha, 2010). Perhaps an anionic residue at this position is favourable as it simultaneously acts to reduce the overall positive charge while giving an interrupted and hence dispersed cationic face.

While haemolytic activity of the K2C18 and MCC18 variants was not assessed in this study, some general comment may be made. Many AMPs that exhibit a high positive charge (+7 to +9; such as those studied here) do not tend to lyse red blood cells (Dathe & Wieprecht, 1999; Jiang et al., 2008). Confirmation of this, as well as observation of other effects on mammalian cells, would need to be conducted. Encouragingly, the K2C18 and MCC18 variants exhibited activity against the various microbes in high-salt medium (i.e. in LB containing 171 mM NaCl; data not shown), indicating that they may not be antagonised by the high ion concentrations found in human blood (Bowdish et al., 2005a). However, divalent ions such as magnesium and calcium, which may be more potent antagonists of AMP activity (Hancock & Sahl, 2006), were not tested, nor was a more complex medium mimicking physiological conditions (such as serum, which contains a number of peptidases [McGregor, 2008]) used.

In conclusion, the utility of the inducible, autocleaving CPD-hexa-histidine tag as a fusion partner for recombinant production and purification of linear AMPs in *E. coli* has been shown. These AMPs were purified at small scale in an inexpensive and timely manner, and displayed varying antimicrobial activity when tested *in vitro*. It is hoped that this production technique will be useful for the future production of antimicrobials for structure/function studies. The next step is to try to utilise a similar recombinant system to screen for new AMPs.

CHAPTER 4 – SCREENING FOR BIOACTIVE PEPTIDES:

ANTIMICROBIALS

4.1 Introduction

4.1.1 Brief history

The identification of new antimicrobial peptides (AMPs; reviewed in *Section 3.1*) could prove a boon for the treatment of human pathogens that are recalcitrant to existing antibiotics (Rossi et al., 2007). However, identifying and investigating AMPs directly from environmental samples is both economically and environmentally unfavourable (see *Section 1.2.2*). Alternative screening strategies for novel AMPs are therefore desired.

One screening approach is to use *in silico* techniques to identify likely candidates (Fjell et al., 2007). Although AMPs exhibit a large amount of structural diversity, a number contain largely conserved N-terminal pre-pro-regions that require enzymatic cleavage to release the active peptide (Hancock & Sahl, 2006). These conserved regions allow for new AMP sequences to be elucidated by reverse transcriptase PCR from mRNA transcripts, or by bioinformatic database mining as new transcriptome or genome sequences become available (Gennaro & Zanetti, 2000; Fjell et al., 2007). However, such identification of AMPs occurs piecemeal, and misses peptides that do not share similar pre-pro-regions.

Other *in silico* approaches to identifying putative AMPs revolve around quantitative structure-activity relationships, in which descriptor values (such as charge, amphipathicity, hydrophobicity) are given to known peptides that have had their minimum inhibitory concentration validated experimentally (Frederic et al., 2004; Raventós et al., 2005). These data, along with more complicated descriptors (i.e. on the atomic scale for each specific amino acid position), can be fed into computer simulations in order to predict or design AMP sequences *de novo* (Loose et al., 2006; Cherkasov et al., 2009; Fjell et al., 2009). After virtual libraries have been screened in such a manner, the best predictions can be synthesised and experimentally validated. While such *in silico* approaches show great promise (Blondelle & Lohner, 2010), they currently appear limited to the analysis of short

AMP lengths (10 or fewer residues), and their use falls outside the work presented here.

Chemically-synthesised peptides may be harnessed as an input into a high-throughput screen for new AMPs (Blondelle & Lohner, 2010). The associated costs per peptide, however, make screening an unconstrained library uneconomic (see *Section 1.4.1*). Recent use has instead focused on optimising known “hit” peptides. For example, Rathinakumar et al. (2009) screened ~16,000 rationally designed peptides (5 to 9 aa in length) where just four residues were allowed to differ. To reduce cost, Hilpert & co-workers (2007) miniaturised peptide synthesis levels to the μmol level via microarrays, and were able to screen such small amounts for antimicrobial activity via the use of a *Pseudomonas aeruginosa* test-strain expressing a luciferase cassette (Hilpert et al., 2006). In this sensitive setup, if bacterial energy levels are disturbed (i.e. through AMP toxicity), luminescent output falls commensurately. However, as outlined in *Section 3.1.5.1*, this peptide synthesis approach is technically demanding.

In contrast to the natural-product, *in silico* and chemical synthesis screening approaches outlined above, recombinant screening for putative AMPs has the potential to be economically viable, technically straightforward and cost-effective respectively (Raventós et al., 2005).

4.1.2 Recombinant screen approaches

Most approaches to recombinant screening for AMPs have been concerned with optimising known peptides, rather than searching for *de novo* sequences (Blondelle & Lohner, 2010). Phage display (reviewed in *Section 1.2.3.1*) has been employed to select for peptide inhibitors of peptidoglycan synthesis enzymes in the microbial cell wall (El Zoeiby et al., 2003), and for peptide binding to bacterial mimetic liposomes (Tanaka et al., 2008). However, overall antimicrobial activity of isolated peptides when tested against microbial cultures was poor. A more holistic, whole-cell approach is required.

4.1.2.1 Whole-cell screening

Whole-cell screening, where compounds are screened for their ability to provoke a specific phenotype in a cell culture without *a priori* knowledge of a particular target, is reviewed in *Section 1.2.3.3*. Because AMPs can act on generalised targets, such as bacterial membranes (van't Hof et al., 2001), it seems logical to employ this technique as a screening method. The simplest approach is to add purified AMP aliquots to a microbial culture over a concentration gradient, i.e. as per *Section 3.2.5*. Although such a negative selection screen is less desirable than a positive selection approach with regards to resolving power (see *Section 1.2.3.3*), it is the obvious route for novel AMP identification. But while robotisation can aid with the throughput of such a screen (Raventós et al., 2005), prior peptide production is still required.

An alternative is to combine *in vivo* both the screening and peptide production functions. DNA libraries, utilised as a diverse input for subsequent transcription and translation into peptide (see *Section 1.2.2.3* and *Section 1.2.3*), are contained within discrete host cells. Hence, the phenotypic effect that an encoded peptide may have is linked with its genotype, which is recoverable through isolation of the DNA. With regards to an *in vivo* whole-cell screen for AMPs, a novel hit will lead to a loss of growth in the host cell when induced in liquid or solid media. This approach is known as *cis*-screening (see *Section 4.1.2.2*). Past examples of this approach include the use of a vital staining methodology on cells expressing putative AMPs, whereby trypan blue only stains colonies with compromised cell membranes (Loit et al., 2008; Cheng et al., 2009). However, this method appears non-specific (giving a high false-positive rate of 67% to 75%), and does not differentiate between exogenous and endogenous hits (see *Section 4.1.3.1* below). In addition, Novozymes A/S (Basævrð, Denmark) has compared the *in vivo* activity of mutants of a SMAP29 truncation (18 aa, α -helical cathelicidin) in order to optimize its potency (Steinstraesser et al., 2002; Raventós et al., 2005).

4.1.2.2 *Cis* or *trans* screens

In an *in vivo* whole-cell screen for novel AMPs, both *cis* and *trans* approaches are possible (Raventós et al., 2005). The *cis*-based screening approach, in which the producing organism is also the target of the AMP it produces, is outlined in *Section*

4.1.2.1. In a *trans*-based screen, the producing cell is largely immune to the AMP expressed, and an additional AMP-sensitive organism is used to evaluate activity.

The optimisation of plectasin (40 aa, a defensin from the fungus *Pseudoplectania nigrella*) serves as a typical example of a *trans*-based screen (Mygind et al., 2005). Peptide variants were constructed and secreted from colonies of the yeast *Saccharomyces cerevisiae*, and activity assessed via radial diffusion assay using an agar overlay of the test organism *Staphylococcus carnosus* (Raventós et al., 2005). Additional examples of *cis*-based screens using *E. coli* as the host include investigation of the critical residues of apidaecin 1b (Taguchi et al., 1994) and thantoin (Taguchi et al., 2000), as well as those listed above in Section 4.1.2.1. The work described in this chapter is based around utilising *E. coli* as an AMP-sensitive production host in a *cis*-based screen.

4.1.3 Use of an AMP-sensitive production host

A *cis*-acting screen for novel AMPs requires that the producing organism is AMP-sensitive. Previous work has shown that, when patch-plated on inductive medium, clones encoding AMPs, such as PR-39 and bactenecin 5 (43 aa, bovine source), with known intracellular targets fail to grow (Raventós et al., 2005).

4.1.3.1 Exogenous versus endogenous hits

AMPs with a known membrane-specific activity, such as LL-37 and indolicidin, only exhibited a reduction in colony growth when targeted to the periplasm (Raventós et al., 2005). This indicates that membrane localisation is necessary for their activity. Furthermore, it shows promise for an *in vivo* whole-cell screen. If constructed to ensure secretion to the periplasm, identification of novel AMPs that possess exogenous antimicrobial activity should be possible, i.e. activity is retained when purified and added exogenously to a microbe culture.

This is an important detail, as the origin of an AMP produced *in vivo* is the cytoplasm, i.e. endogenous to the target cell. Thus, any AMP hit identified using an *in vivo* whole-cell screen may result from an interaction with a cytoplasmic target. An AMP that is added exogenously is more relevant, as it reflects the clinical reality for any putative drug. It is more than likely that identified intracellular AMP hits

would not possess the required cell-penetrating activity to access the cytoplasm if added exogenously, as the microbial cell membrane is a potent barrier (Stewart et al., 2008). While such hits may be modified to allow cell-penetration (Eriksson et al., 2002), it makes greater sense to focus on isolating exogenous AMP hits in the first instance.

4.1.3.2 Secretion systems

Targeting of known membrane-active AMPs to the periplasm of *E. coli* during recombinant production has been shown to lead to toxicity (discussed in *Section 4.1.3.1*). Although *E. coli* possesses several different secretion systems, most are protein-specific rather than generic in function, and true secretion through both the inner and outer membranes is a complex process (Saier, 2006). Furthermore, modification of these systems for heterologous protein secretion gives unpredictable results, with any success requiring a heuristic approach (Ni & Chen, 2009). Nevertheless, the most commonly harnessed secretion system in *E. coli* is the Sec pathway (Choi & Lee, 2004). In this system, a short (10 to 20 aa) N-terminal tag is recognised by the SecA ATPase and threaded through the inner membrane via the SecYEG complex, where the secretion tag is subsequently cleaved by signal peptidase I (Natale et al., 2008). While there are a number of Sec secretion signals available (Choi & Lee, 2004), a truncation of the gIII gene, which encodes the 18 aa secretion signal of the minor capsid protein pIII from filamentous phage fd (Rapoza & Webster, 1993), was chosen for this work. This commonly used N-terminal tag is sufficient for peptide targeting to the Sec-translocase complex, but may also be aided by the chaperone SecB, which binds and keeps nascent peptides in an unfolded state prior to periplasmic translocation (Natale et al., 2008).

Extracellular release once a pIII-tagged peptide reaches the periplasm is, however, not the norm. While a few recombinant proteins have been observed to accumulate in the extracellular milieu, this may be because of natural cell death and lysis, and seems dependent on culture conditions (Ni & Chen, 2009). Strategies to promote release exist, including: osmotic shock; chemical treatment, e.g. Triton X-100 detergent; co-expression of a lysis promoting peptide, e.g. bacteriocin release protein; or the use of L-form cells, which lack an outer membrane (Gumpert &

Hoischen, 1998; Choi & Lee, 2004). However, such approaches result in a general loss of cell viability, and consequently would lead to an AMP effect being indistinguishable from the control.

4.1.3.3 Replica plating

Although the patch-plating approach discussed at the beginning of *Section 4.1.3* may be suitable for manual analysis of a small number of AMPs expressed by *E. coli in vivo*, a different method is required for high-throughput screening of a DNA library. Replica plating, commonly used for profiling microbial colony mutation libraries on different selective growth media (Lederberg & Lederberg, 1952), may be harnessed as a negative screen. This simple and robust technique may be used to transfer a library of putative AMP-encoding colonies to inductive medium; comparison of subsequent outgrowth colonies can isolate clones that fail to grow due to the toxic effect of a produced putative AMP.

4.1.4 Other screen considerations

As mentioned in *Section 4.1.2.1*, DNA is the input utilised for a recombinant *in vivo* AMP screen. For proof-of-principle in this work, genomic DNA (gDNA) extracts from human (*Homo sapiens*) and bdelloid rotifer (*Adineta ricciae*) were used. The haploid human genome contains approximately 3 billion bp, with 1.2% of this coding for approximately 20,000 to 25,000 discrete proteins (International Human Genome Sequencing Consortium, 2004). The *A. ricciae* genome is in the process of being annotated, but it has a predicted size of 120 to 135 million bp (C. Boschetti, personal communication). Both sources of gDNA should serve as good sources of sequence diversity, especially that of *A. ricciae*, as bdelloid rotifer genomes have been shown to contain large numbers of horizontally transferred genes (Gladyshev et al., 2008).

Promoter selection and potential codon bias in the peptide expression vector should also be considered (Ingham & Moore, 2007). For the former, the tightly-regulated, arabinose-inducible *araBAD* promoter was chosen, as it is suitable for the cloning and expression of toxic genes such as AMPs (Guzman et al., 1995). For the latter, the codon bias (Nakamura et al., 2000) of the human and rotifer gDNA inserts may affect their translation, and hence final peptide yield, in *E. coli*. However, as the

screen will presumably “skip” such reticent sequences, codon bias should not be an issue for any putative AMP hits identified.

Lastly, fusion partners to an AMP are commonly utilised during recombinant production, both in order to alleviate toxicity to their host and to aid in downstream purification (reviewed in *Section 3.1.5.2*). Results from *Chapter 3* indicated that *E. coli* tolerates endogenous expression of an AMP-CPD fusion, and that active AMP can be successfully released upon addition of IP₆. This property of inducible CPD cleavage could be utilised in an *in vivo* whole-cell screen. If truly inactive, an AMP-CPD fusion could be tagged for secretion in either a *cis* or *trans*-based screen, and once a critical threshold of AMP-CPD was reached, cleavage could be induced and active AMP released. Exogenous antimicrobial activity could then be observed, either against the host or by using the culture medium to challenge other microbes. To ascertain whether this would be possible, the first step is to establish if the positive control K2C18-CPD_{WT} lacks antimicrobial activity in its fusion state.

4.1.5 Summary

Natural-product and chemically-synthesised peptide library screens for new AMPs have issues with environmental sustainability, technical expertise required and cost-effectiveness. Although only focusing on peptide-based entities, rather than the wide range of secondary metabolites that many organisms are capable of producing, screens for AMPs are inherently more amenable to high-throughput methodologies. While chemically-synthesised libraries have shown promise in recent years, recombinant whole-cell approaches can potentially couple AMP synthesis with assaying for activity in the same step and at a lower cost. Harnessing a *cis*-based screen, in which the microbial host is sensitive to the putative AMP produced, could lead to the rapid identification of novel AMPs.

To demonstrate proof-of-principle, the effect of production in *E. coli* of the model AMP K2C18 was investigated for amenability with regards to a replica plate negative screen. Furthermore, genomic DNA libraries of human and rotifer origin were employed in a screen for novel AMPs with exogenous or endogenous activity. Identified hits were further investigated through bioinformatic and mutational means, and several hits produced recombinantly and synthetically. Their efficacy when

added exogenously back to their production host was also investigated in order to confirm the ability of the screen to identify novel AMPs that have bioactive potential.

4.2 Results

4.2.1 Analysis of K2C18-CPD activity as a fusion protein

To determine if an AMP-CPD fusion was inactive when added exogenously to an *E. coli* culture, i.e. before cleavage and release of the active AMP, K2C18-CPD_{WT} (see Section 2.4.1.2) was utilised as a model control.

4.2.1.1 Purification and activity of crude K2C18-CPD_{WT}

Expression and purification of pET28-K2C18-CPD_{WT} in *E. coli* was carried out as described earlier in Section 3.2.2.2, except CPD cleavage was not induced; rather, 3 bead volumes of wash buffer containing 500 mM imidazole were used for elution of the fusion protein. Imidazole, which is toxic to *E. coli* (Simonetti et al., 2001), was subsequently removed by dialysis (see Section 2.4.3.2) against wash buffer without imidazole. Dilutions of this K2C18-CPD_{WT} dialysate, with or without 500 μ M IP₆ pre-treatment for 1 h to induce the release of free K2C18, were tested for antimicrobial activity against *E. coli* DHB10 by liquid culture assay (see Section 2.5.3).

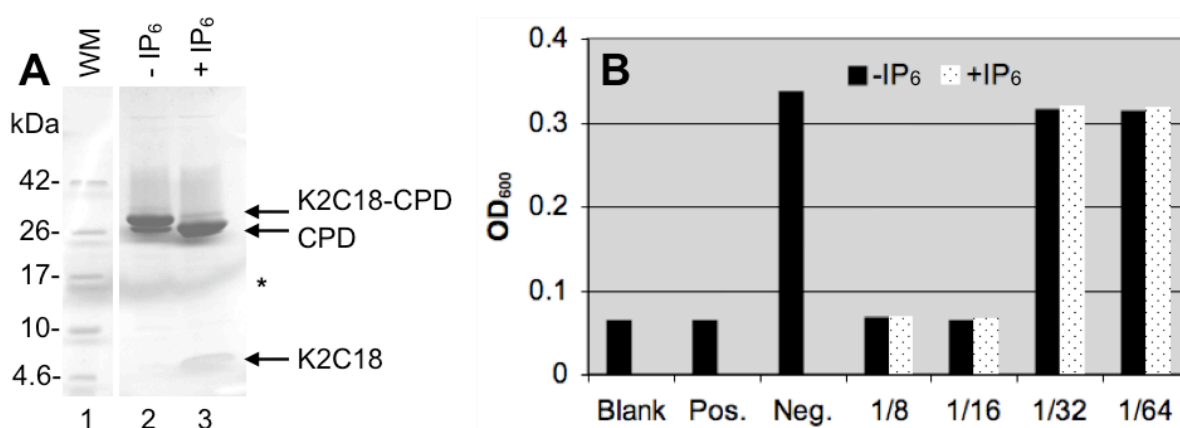


Figure 4.1: K2C18-CPD fusion exhibits antimicrobial activity in liquid culture assay. **A**, SDS-PAGE (tricine 10-20%) of dialysed K2C18-CPD_{WT} Ni-NTA eluate; and **B**, 10 μ L dialysate serially diluted (with/without 500 μ M IP₆ pre-treatment for 1 h) and added to 90 μ L of 5×10^4 cfu/mL *E. coli* DH10B. OD₆₀₀ reading taken after 24 h incubation at 37°C, representative experiment shown from duplicate. WM, molecular mass marker; Blank, LB medium only; Pos., elution buffer (250 mM imidazole); Neg., wash buffer including 1 mM IP₆; *, gel artefact. Representative experiment shown (n = 2).

Both cleaved and uncleaved K2C18-CPD_{WT} (Figure 4.1 A) showed an equivalent antimicrobial activity over the dilutions tested (Figure 4.1 B). There is an estimated 500 ng/μL liberated K2C18 in the IP₆ treated gel sample (lane 3, with reference to the molecular mass marker), which equates to an approximate final concentration of 1.4 μM when diluted 160-fold (16-fold initially, then another 10-fold when added to the antimicrobial liquid culture assay). This is in line with the minimum inhibitory concentration (MIC) observed for purified K2C18 against *E. coli* (see Section 3.2.5).

Even if the untreated K2C18-CPD_{WT} contained a small amount of liberated K2C18 from unstimulated background CPD cleavage (not observed in lane 2), there would be at least a two-fold observable difference between the inhibitory concentration of the plus and minus IP₆ samples if the free peptide was solely responsible for the antimicrobial effect. This was not observed. Overall, the use of “inactive” membrane-targeted AMP-CPD fusions appears unsuitable for *E. coli*, as the fusion may retain antimicrobial activity.

4.2.2 Endogenous activity of K2C18 and K2C38 expressed in *E. coli*

As the strategy of harnessing an “inactive” AMP-CPD fusion proved unfeasible, the use of an AMP’s innate toxicity (without a fusion partner) towards its production host was attempted as a screening approach instead (Raventós et al., 2005). K2C18, as well as K2C38, the full-length CRAMP parent from which K2C18 was derived (Shin et al., 2000), were used as model AMPs in order to validate this method and aid in the construction of a screen for novel AMPs.

4.2.2.1 Construction of expression vectors

The K2C18 and K2C38 sequences were cloned into the arabinose-inducible pBAD/gIII-A plasmid between the vector-encoded N-terminal pIII secretion tag and C-terminal hexa-histidine tag (see Section 2.6.1.2). pBAD/gIII-A was also modified to remove the pIII-encoding secretion tag, resulting in pBADm (see Section 2.6.1.1). K2C38 was cloned into pBADm in a similar manner as per pBAD/gIII-A. Versions of K2C18 and K2C38 were also constructed to incorporate a stop codon, so that the C-terminal hexa-histidine tag was not translated (see Section 2.6.1.2). Table 4.1 summarises the plasmids constructed, which were transformed into *E. coli* TOP10 for expression studies.

| Plasmid name | Encoded product | MW (kDa) |
|---------------------|--|----------|
| pBAD/gIII-A | <u>MKKLLFAIPLVVPFYSHSTMELEICSWYHMGIRSFLEQKLISEEDL</u> <u>NSAVDHHHHHH</u> | 6.8 |
| pBADm | MTMELEICSWYHMGIRSFLEQKLISEEDLNSAVDHHHHHH | 4.8 |
| pBAD/gIII-K2C18-Tag | <u>MKKLLFAIPLVVPFYSHSTMGKKLKKIGQKIKNFFQKLYVRSFLEO</u> <u>KLISEEDLNSAVDHHHHHH</u> | 7.7 |
| pBAD/gIII-K2C18* | <u>MKKLLFAIPLVVPFYSHSTMGKKLKKIGQKIKNFFQKL</u> | 4.5 |
| pBAD/gIII-K2C38-Tag | <u>MKKLLFAIPLVVPFYSHSTMAISRLAGLLRKGGEKIGKKLKKIG</u> <u>QKIKNFFQKLVQPEIRSFLEQKLISEEDLNSAVDHHHHHH</u> | 9.8 |
| pBAD/gIII-K2C38* | <u>MKKLLFAIPLVVPFYSHSTMAISRLAGLLRKGGEKIGKKLKKIG</u> <u>QKIKNFFQKLVQPE</u> | 6.7 |
| pBADm-K2C38-Tag | MTMA ISRLAGLLRKGGEKIGKKLKKIGQKIKNFFQKLVQPE IR <u>SFLEQKLISEEDLNSAVDHHHHHH</u> | 7.8 |
| pBADm-K2C38* | MTMA ISRLAGLLRKGGEKIGKKLKKIGQKIKNFFQKLVQPE | 4.7 |

Table 4.1: pBAD plasmid constructs encoding K2C18 and K2C38. pBAD/gIII vectors encode the N-terminal pIII secretion signal (underlined), while pBADm vectors do not. Expression is induced by L-arabinose. K2C18/K2C38 indicated in bold, C-terminal hexa-histidine tag region (Tag) also underlined. Asterisk, denotes the lack of hexa-histidine tag region; MW, molecular mass.

4.2.2.2 Growth of *E. coli* expressing K2C18 and K2C38 on solid medium

As replica plating was to be utilised as a screening methodology, the size of *E. coli* colonies expressing K2C18 and K2C38 on solid medium was assessed initially. Ideally, bacteria expressing toxic AMPs would fail to grow, although only partial growth inhibition has been reported previously (Raventós et al., 2005). *E. coli* TOP10 containing the plasmids outlined in Table 4.1 were streaked out on LB agar containing the inducing agent arabinose in concentrations ranging from 0.005% to 2%, and the resulting colony morphology observed (see Figure 4.2).

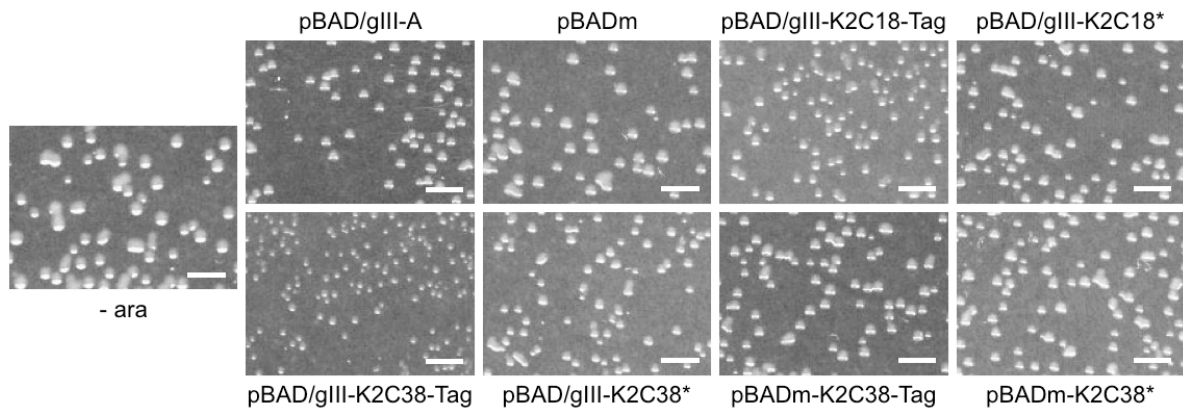


Figure 4.2: Colonies of induced *E. coli* containing pBAD/gIII-K2C18-Tag and pBAD/gIII-K2C38-Tag are smaller in size than the controls. Photos taken after ~16 h at 37°C on LB medium, 0.05% arabinose. All uninduced cultures exhibited the same morphology, representative photo shown (- ara). Scale bar, 2 mm. Representative photos shown (n = 2).

No AMP constructs failed to grow on arabinose-containing medium; rather, most of the induced K2C18 and K2C38 constructs exhibited a similar colony size to that of the pBAD/gIII-A and pBADm controls. Furthermore, all induced colonies were slightly smaller than their uninduced controls, indicating a general effect on growth by arabinose. However, the pIII/hexa-histidine tagged versions of K2C18 and K2C38 gave a discernible further reduction in colony size, with pBAD/gIII-K2C38-Tag being the most marked. All ranges of arabinose concentrations tested (0.005% to 2%) resulted in similar colony sizes to those shown in *Figure 4.2*. As 0.05% was seen to ensure maximal expression in the literature (Guzman et al., 1995), this concentration was utilised for subsequent experiments.

Because the difference in size between the induced pBAD/gIII-A control and pBAD/gIII-K2C18-Tag colonies was not dramatic, colony PCR (see *Section 2.2.5.2*; using K2C18-specific primers) was used to confirm whether or not the smaller colonies did contain the K2C18 plasmid. A mixed culture of pBAD/gIII-A (~66%) and pBAD/gIII-K2C18-Tag (~33%) was plated on LB agar containing 0.05% arabinose, and the resultant small and large colonies analysed (see *Figure 4.3*).

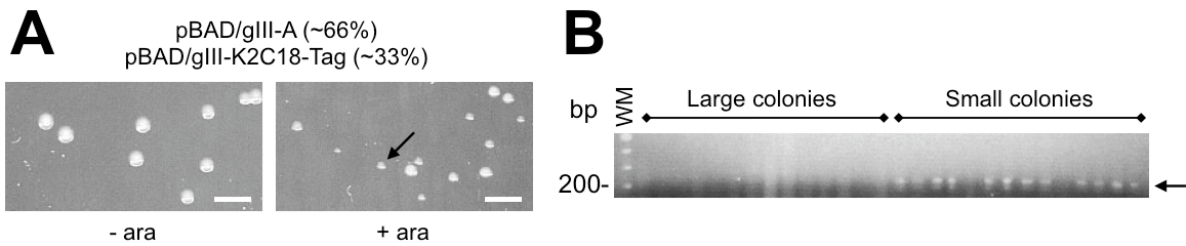


Figure 4.3: Small colonies from induced mix of pBAD/gIII-A and pBAD/gIII-K2C18-Tag *E. coli* are found to contain the pBAD/gIII-K2C18-Tag plasmid. **A:** Photos taken after ~16 h at 37°C on LB medium, example small colony arrowed. Scale bar, 2 mm. **B:** Colony PCR shows a K2C18-specific product (191 bp) in no large colonies but in 12 out of 14 small colonies assayed (arrowed). Ara, arabinose (0.05%); WM, DNA ladder.

Figure 4.3 indicates that the size difference observed between the colonies of induced control and K2C18-containing *E. coli* can be used to identify colonies that contain the AMP-encoding plasmid. This shows that colony size may be used as a selection criterion in a screen for endogenously produced AMPs.

4.2.2.3 Growth curves of *E. coli* expressing K2-CRAMP constructs

The effect of induction of K2C18 and K2C38 expression in liquid LB was also explored. Cultures of *E. coli* TOP10 containing the various plasmids outlined in Table 4.1 were induced during early exponential-phase growth as per Section 2.3.1 and OD₆₀₀ monitored over time.

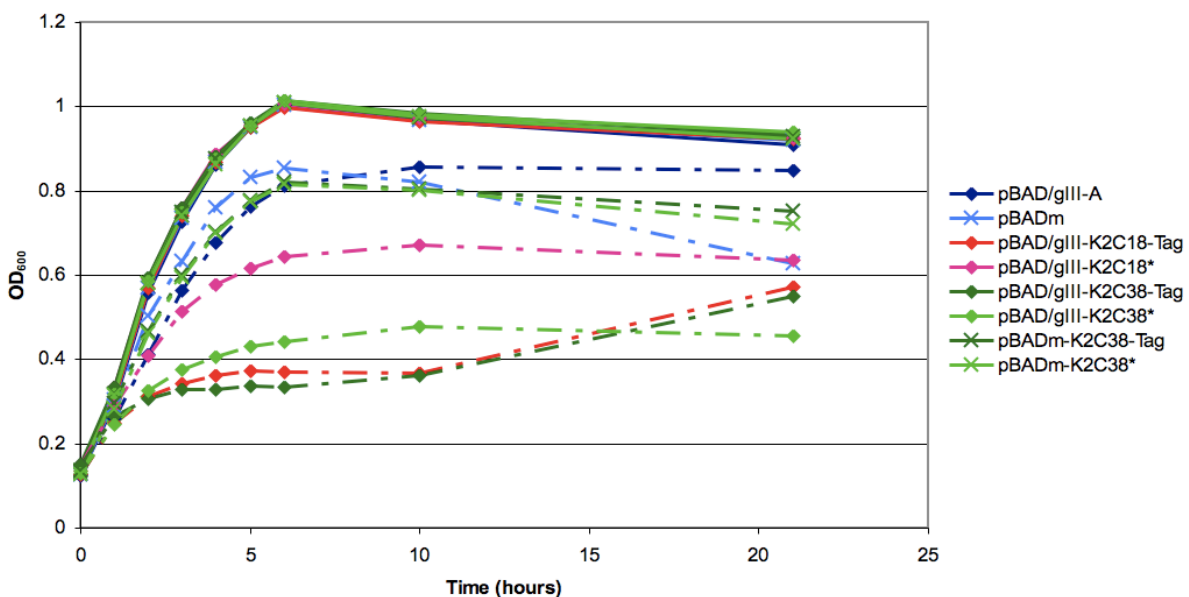


Figure 4.4: K2C18 and K2C38, when targeted to the periplasm, inhibit *E. coli* growth. Arabinose (0.05%) added to induce cultures in LB medium at 0 h (dashed lines), incubated at 37°C at ~225 rpm. Representative experiment shown (n = 3).

All uninduced strains grow uniformly (*Figure 4.4*), while the pBAD/gIII-A and pBADm controls (possessing or lacking the pIII secretion tag, respectively) show slightly reduced growth. This general induction effect on growth was seen to be typical for non-toxic products, such as pIII-tagged calmodulin (see *Appendix 3*). Both pBAD/gIII-K2C18-Tag and pBAD/gIII-K2C38-Tag exhibit a similar level of inhibition of growth. This planes out over the first 3 h of induction, but growth resumes after 10 h. Removal of the C-terminal hexa-histidine tag region, however, reduces the effect of K2C18 (and that of K2C38 slightly). This was also seen on solid medium (*Figure 4.2*). It may be that simply increasing the length of the encoded peptide confers stability and thus increases half-life, allowing more time for a deleterious effect to occur (Maurizi, 1992). Strikingly, removal of the N-terminal pIII secretion signal from K2C38 leads to complete removal of growth inhibition, indicating that this AMP needs to be targeted to the periplasm for its antimicrobial activity to be successfully exerted.

4.2.2.4 Western blot of *E. coli* expressing K2-CRAMP constructs

The expression of the pIII-K2C18-Tag, pIII-K2C38-Tag and K2C38-Tag peptides in liquid culture was analysed by Western blotting using an anti-hexa-histidine antibody (see *Section 2.3.3.4*).

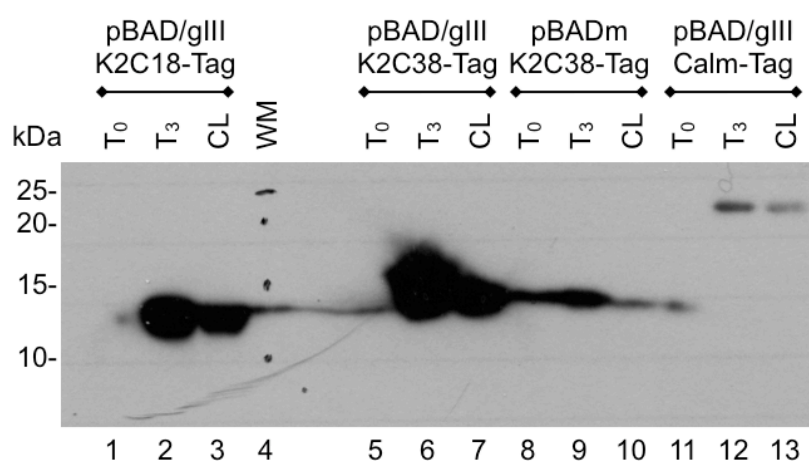


Figure 4.5: Western blotting confirms expression of K2C18 and K2C38 in *E. coli*. SDS-PAGE (glycine 15%) of hexa-histidine tagged constructs from Table 4.1. pBAD/gIII-Calm-Tag encodes calmodulin (22 kDa), used as a control. T₀, uninduced; T₃, 3 h induction (0.05% arabinose); CL, clarified lysate; WM, molecular mass marker.

While *Figure 4.5* confirms the production of K2C18 and K2C38 (lanes 2, 6 & 9), the corresponding bands were not visible on a Coomassie-stained polyacrylamide gel (data not shown). This indicates that overall expression levels are low, i.e. in the order of ng/mL culture. This is unsurprising, as such short peptides may be especially susceptible to proteases (Maurizi, 1992; Taguchi et al., 1994). Given that the MIC of recombinant K2C18 was observed to be 1 μ M (\sim 2.2 μ g/mL, see *Section 3.2.5*), it is likely that the AMP is being produced at sub-MIC levels. Lastly, although K2C18 and K2C38 run higher than their expected molecular masses (see *Table 4.1*; explanation in *Section 3.2.2.3*), pIII-K2C38-Tag and K2C38-Tag exhibit bands running at the same molecular mass (compare lanes 7 & 9). It appears therefore that the N-terminal pIII secretion signal (2.1 kDa) is efficiently cleaved from pIII-K2C38-Tag.

4.2.2.5 Morphology of *E. coli* expressing K2-CRAMP constructs

Following from *Section 4.2.2.3*, *Figure 4.6* shows photos taken of *E. coli* cells containing the various plasmids (*Table 4.1*) after 4 h induction. Constructs that exhibited negligible growth inhibition in *Figure 4.4* exhibit a short, rod-like morphology, which was considered normal (Justice et al., 2008). In contrast, cells that targeted K2C18 and K2C38 to the periplasm, i.e. with the pIII secretion tag, are elongated. This increase in length is likely due to cell division being halted, an effect that has been observed during treatment of *E. coli* with AMPs such as PR-39, indolicidin and microcin 25 (Broden, 2005).

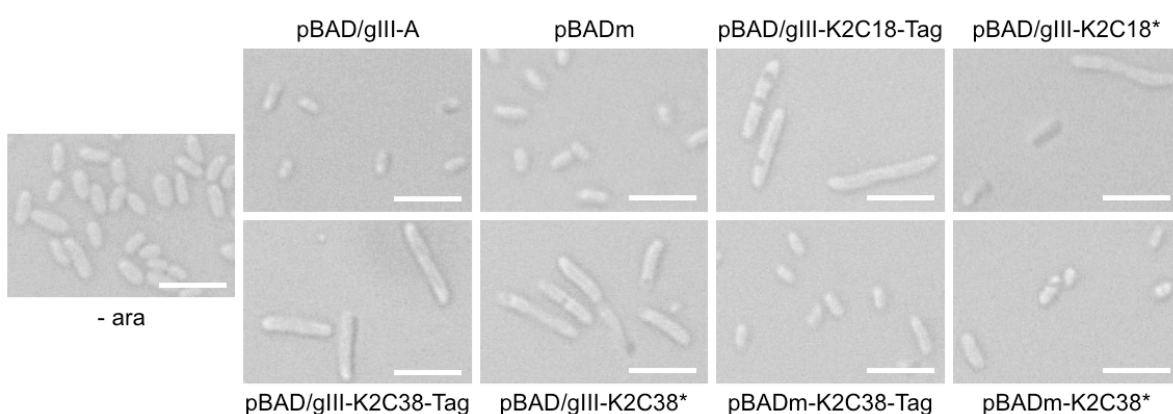


Figure 4.6: *E. coli* expressing growth-inhibiting K2C18 or K2C38 constructs have an elongated morphology. Photos taken after 4 h arabinose induction (0.05%) in LB medium at 37°C at \sim 225 rpm. All uninduced cultures exhibited the same morphology, representative photo shown (- ara). Scale bar, 5 μ m.

Cells from *Figure 4.6* were inspected again after 21 h induction. While longer cells were still visible in the pBAD/gIII-K2C18 and pBAD-K2C38 cultures, these were now accompanied by cells with a more normal, short rod morphology. Due to the observed recovery of growth, K2C18 and K2C38 appear to exert a bacteriostatic rather than bactericidal effect when produced endogenously. Further indication of this was obtained by plating aliquots of cultures that had been induced for 4 h (cf. *Figure 4.4*) on arabinose-free LB agar containing carbenicillin to select for plasmid retention; preliminary data indicated that colony forming units (cfu) per OD₆₀₀ unit were within the same order of magnitude for all constructs (data not shown).

4.2.2.6 Seed growth curves of *E. coli* expressing K2-CRAMP constructs

The apparent recovery of cultures expressing K2C18 and K2C38 was further explored by using samples induced for 21 h (see *Figure 4.4*) to seed fresh growth curves (*Figure 4.7 A*). Furthermore, uninduced samples were then taken from *Figure 4.7 A* after 21 h growth and used to seed a third growth curve (*Figure 4.7 B*), in the hope that uninduced cultures in *Figure 4.7 A* may have had time to “reset” in the absence of arabinose, and thus become responsive to induction once more.

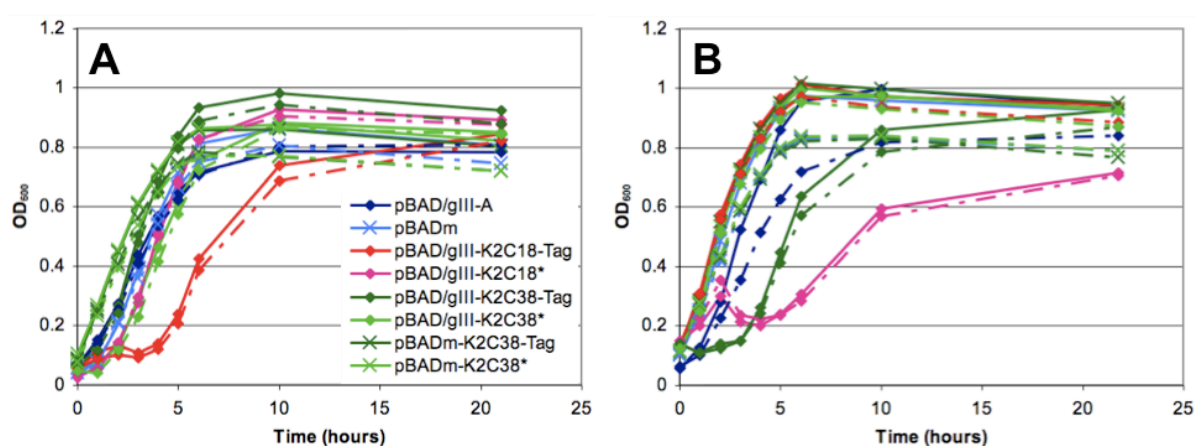


Figure 4.7: *E. coli* from pre-induced gIII-K2C18/K2C38 cultures does not respond to further arabinose induction (A), and are not reset to arabinose sensitivity after recovery in arabinose-free LB medium (B). Arabinose (0.05%) added to induce cultures at 0 h (dashed lines), incubated at 37°C at ~225 rpm. Representative experiment shown (n = 2).

The pBAD/gIII-A, pBADm, pBADm-K2C38-Tag and pBADm-K2C38* control constructs, which showed no inhibition of *E. coli* growth upon induction (*Figure 4.4*), gave similar growth curve profiles in both *Figure 4.7 A* and *B*. However, the

inhibiting constructs pBAD/gIII-K2C18 and pBAD/gIII-K2C38 no longer display retarded growth, but rather track their uninduced controls in both seeded growth curves. Arabinose-induced growth inhibition is therefore not reset in these constructs after recovery in arabinose-free medium.

Some constructs (both induced and uninduced) exhibited initial growth lags in the seed growth curves. For example, pBAD/gIII-K2C18-Tag cells in *Figure 4.7 A*, when analysed microscopically, exhibited a number of filamented cells amongst normal short rods after 4 h induction (*Figure 4.8*). pBAD/gIII-K2C18* cells (*Figure 4.7 B*) had a similar mixed phenotype. In both cases, uninduced and induced samples were morphologically similar.

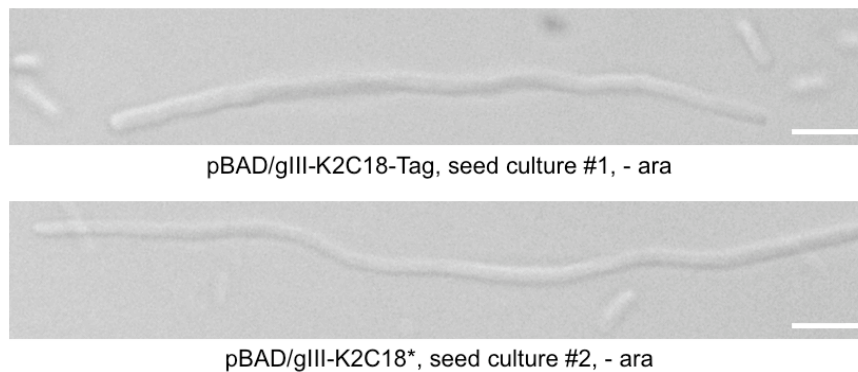


Figure 4.8: *E. coli* from pre-induced cultures that exhibit growth curve lags contain filamented cells. Photos taken after 4 h post-inoculation in LB medium at 37°C at ~225 rpm. Uninduced (- ara) cultures shown, induced cultures exhibited the same morphology. Scale bar, 5 μ m.

Lastly, the addition of fresh carbenicillin (to select for plasmid retention) and/or arabinose during the early induction period, i.e. prior to 10 h, did nothing to restore growth-inhibition of bacteria containing the pBAD/gIII-K2C18-Tag construct (data not shown). However, if induced cultures were plated on arabinose-free LB agar after 4 h induction, and resultant colonies used as an inoculum for a new growth curve, the growth-inhibitory phenotype was restored (*Figure 4.9*).

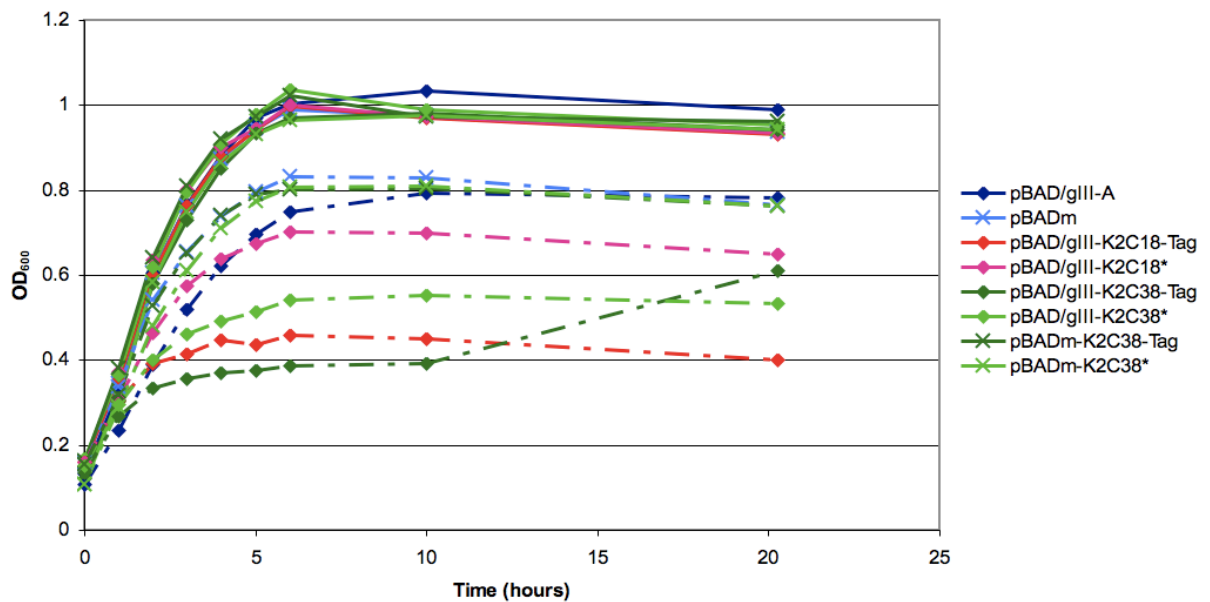


Figure 4.9: *E. coli* from induced *gIII-K2C18/K2C38* cultures are reset to arabinose sensitivity after recovery on arabinose-free LB agar. Cultures were induced for 4 h (cf. Figure 4.4), recovered on LB agar, and single colonies used as starter culture inoculums for this figure. Arabinose (0.05%) added to induce cultures at 0 h (dashed lines), incubated at 37°C at ~225 rpm.

The growth curves in Figure 4.9 are similar to those initially observed in Figure 4.4 (although the pBAD/gIII-A induced control lags slightly). This ability for the bacteria to “reset” on solid medium indicates that plasmid or genomic mutations are not responsible for the recovery of growth observed past ~10 h. It appears that *E. coli* is capable of adapting to the production and presence of toxic K2C18 and K2C38. This may be related to the persister effect, in which a small, stochastically generated sub-population of microbes in a genetically homogenous culture becomes tolerant to antimicrobial action (Lewis, 2007).

4.2.2.7 Comparison to the persister effect

The persister effect has been previously reported in *E. coli* harbouring a plasmid encoding the hok (host killing) peptide (Bej et al., 1988). Hok is a 52 amino acid peptide that kills the expressing host through inner cell membrane depolarisation, and forms part of the well-described hok/sok toxin-antitoxin pair that is responsible for retention of the R1 virulence plasmid in *E. coli* (Pecota et al., 2003). To determine if this effect is similar to that observed for K2C18 and K2C38 in Section 4.2.2.6, the hok gene was cloned into pBAD/gIII and pBADm (see Section 2.6.1.3; a stop codon was included to prevent C-terminal hexa-histidine tag translation).

Furthermore, frame-shift mutations were generated after the start codon of the *hok* gene to serve as non-toxic controls. Growth curves in *E. coli* TOP10 (Figure 4.10 A), as well as seed growth curves (Figure 4.10 B and C) were performed as for the K2C18 and K2C38 constructs previously.

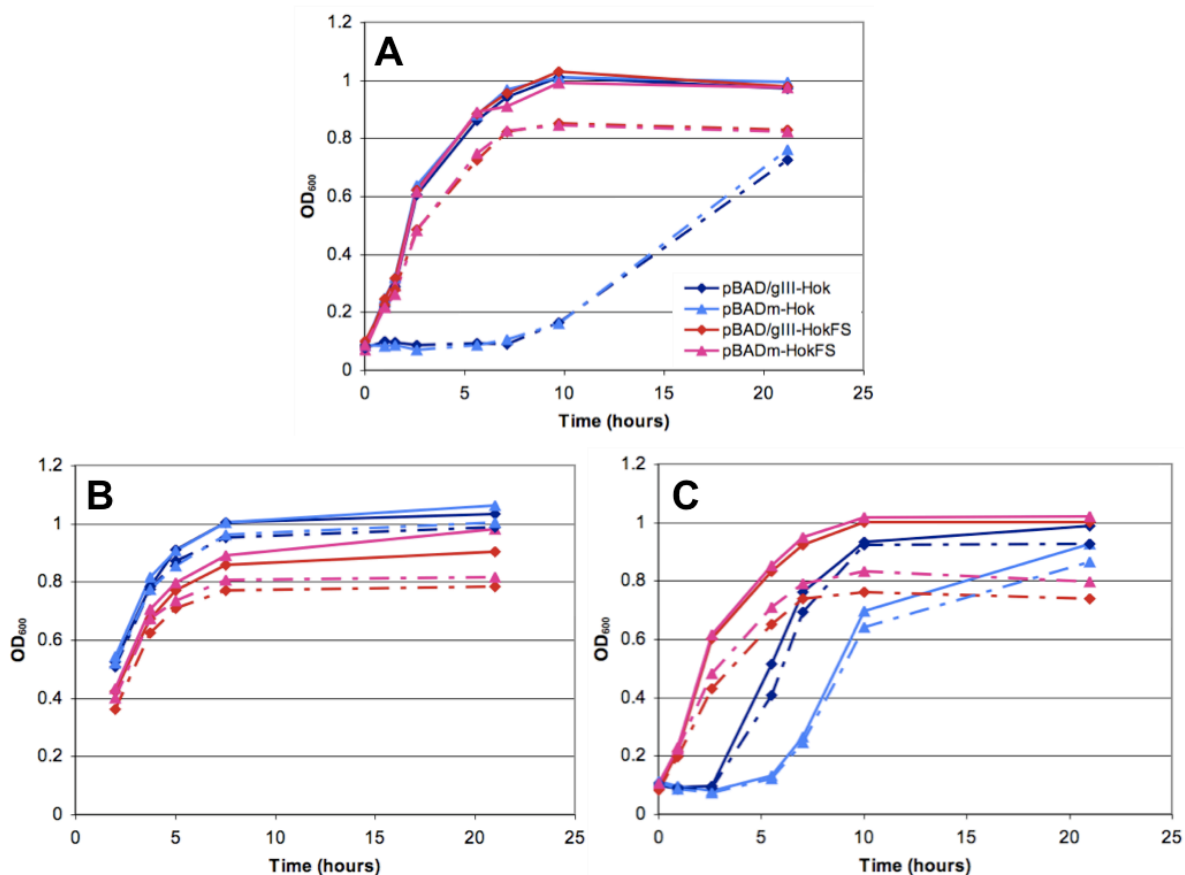


Figure 4.10: The persister effect also occurs during the induction of *hok* in *E. coli*. Arabinose (0.05%) added to induce cultures in LB medium at 0 h (dashed lines), incubated at 37°C at ~225 rpm. **A:** Induction of *hok* suppresses *E. coli* growth initially before recovery results. **B:** *E. coli* from pre-induced *hok* cultures does not respond to further arabinose induction. **C:** *E. coli* containing *hok* plasmids are not reset to arabinose sensitivity after recovery in arabinose-free medium. *HokFS*, frame-shift negative control. Representative experiments shown ($n = 2$).

The secreted and non-secreted forms of *hok* (blue/light blue lines in Figure 4.10) exhibited the same phenomenon as seen with K2C18 and K2C38 in Figure 4.7. A growth-inhibitory effect was seen in the initial growth curve (Figure 4.10 A), which is somewhat stronger and more immediate than that seen with K2C18 and K2C38 in Figure 4.4. However, using these arabinose-treated cultures to seed a fresh growth curve leads to a loss of response to fresh induction (Figure 4.10 B). When these uninduced cultures were used to seed a third growth curve, no resetting of a

response to arabinose was observed (*Figure 4.10 C*). Rather, both uninduced and induced hok versions grew at a similar rate, with an initial lag period seen before outgrowth. As with K2C18 and K2C38 in *Figure 4.7*, reasons for this are unclear. The frame-shift mutants (red/pink lines), on the other hand, exhibited the same non-inhibitory profile across all growth curves.

4.2.3 Construction of vectors for AMP screen

The endogenous expression of the AMPs K2C18 and K2C38 was characterised in *Section 4.2.2*. On solid medium, induced colonies were noticeably smaller in size than their controls when the AMP was targeted to the periplasm. In liquid medium, although recovery of growth during induction was observed after several hours, inhibition was seen during the early growth phase. These properties may be exploited to screen DNA libraries for novel AMPs if an appropriate set of expression vectors is created.

4.2.3.1 Construction of pAMP/S

The pBAD/gIII-A plasmid was modified (see *Section 2.6.1.4*) in order to make it suitable for the insertion of random lengths of DNA, as well as amenable to replica plating (*Figure 4.11*). For the first property, a new araBAD cloning site was constructed to allow for blunt-end insertion of DNA fragments via an *AfeI* site. Downstream stop codons in all three reading frames were also incorporated to ensure translation would be halted no matter what reading frame the DNA insert presented itself in. To aid replica plating, the *lacZ* operon was incorporated. Constitutively expressed, the encoded β -galactosidase can cleave exogenously supplied X-gal to give an insoluble blue product (Davies & Jacob, 1968). This gives the normally beige *E. coli* colonies an excellent contrast when plated on white membranes on X-gal medium, making them suitable for scanning and post-image processing.

4.2.3.2 Construction of pAMP

To screen for AMPs that only exhibit endogenous antimicrobial activity, i.e. act on intracellular targets, pAMP/S was modified (see *Section 2.6.1.4*) to remove the region coding for the N-terminal pIII secretion signal. *Figure 4.11* shows schematics of the pAMP/S and pAMP vectors.

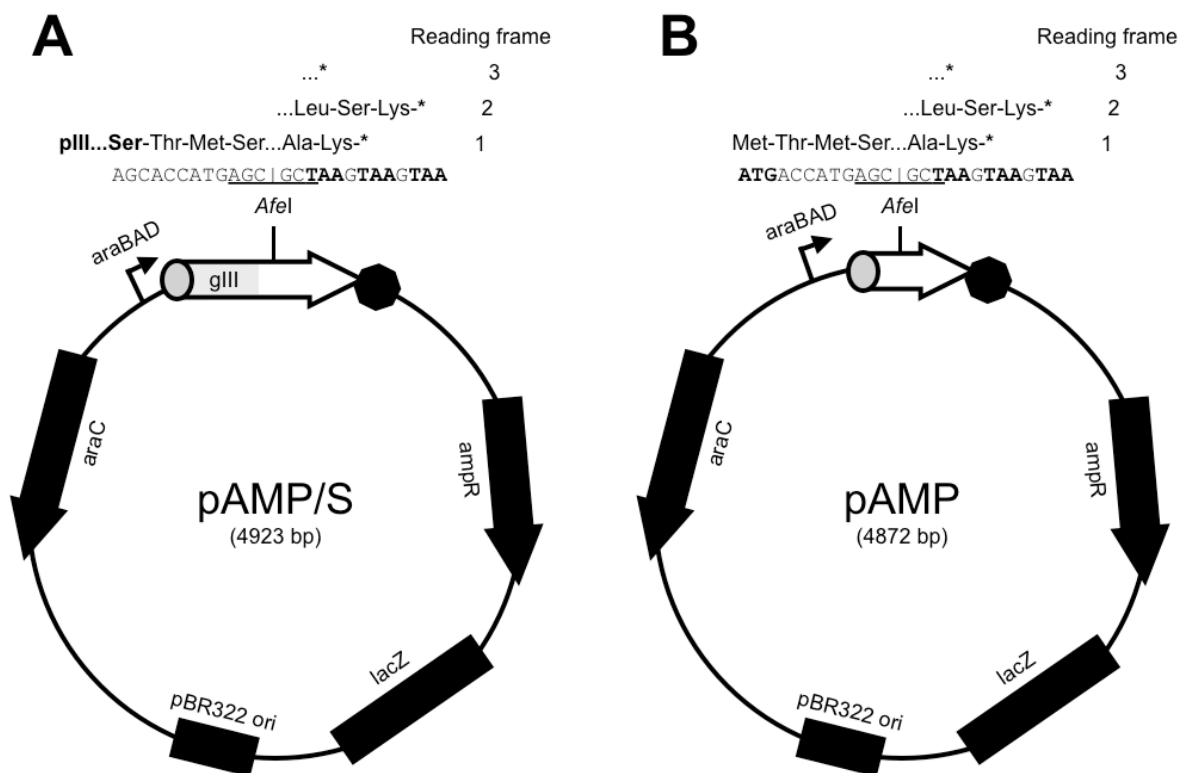


Figure 4.11: The pAMP/S and pAMP vectors used to screen for putative AMPs. Insert DNA may be blunt-end ligated into the *AfeI* site (underlined). **A:** pAMP/S, encodes the N-terminal *pIII* for targeting a peptide to the periplasm. The C-terminal serine residue of this 18 aa secretion signal is indicated in bold (cleaved after this residue post-secretion). **B:** pAMP, lacks the *pIII* secretion signal, start codon indicated in bold. Stop codons are included for all three reading frames, indicated in bold (possible C-terminal residues also shown).

4.2.4 Screening for novel putative AMPs

4.2.4.1 Use of bdelloid rotifer and human genomic DNA libraries

Genomic DNA (gDNA) from the bdelloid rotifer *A. ricciae*, as well as from a human source, was utilised as input DNA for ligation into pAMP/S and pAMP. Fragmentation of this gDNA to a suitable size range for encoding peptides was required: nebulisation (see Section 2.6.3) was employed due to its ability to induce extremely random DNA breakage without requiring specialist equipment (Quail, 2005). Figure 4.12 A shows a representative agarose gel of the achieved fragment size range (100 to 900 bp, clustered around 350 bp), which is in line with the smallest achieved in the literature (Margulies et al., 2005). Ligation of these fragments into pAMP gave inserts with a typical size range of approximately 60 to 400 bp, as revealed by colony PCR (Figure 4.12 B). If no internal stop codons fall in-frame, peptides between approximately 20 and 130 amino acids in length would be coded for.

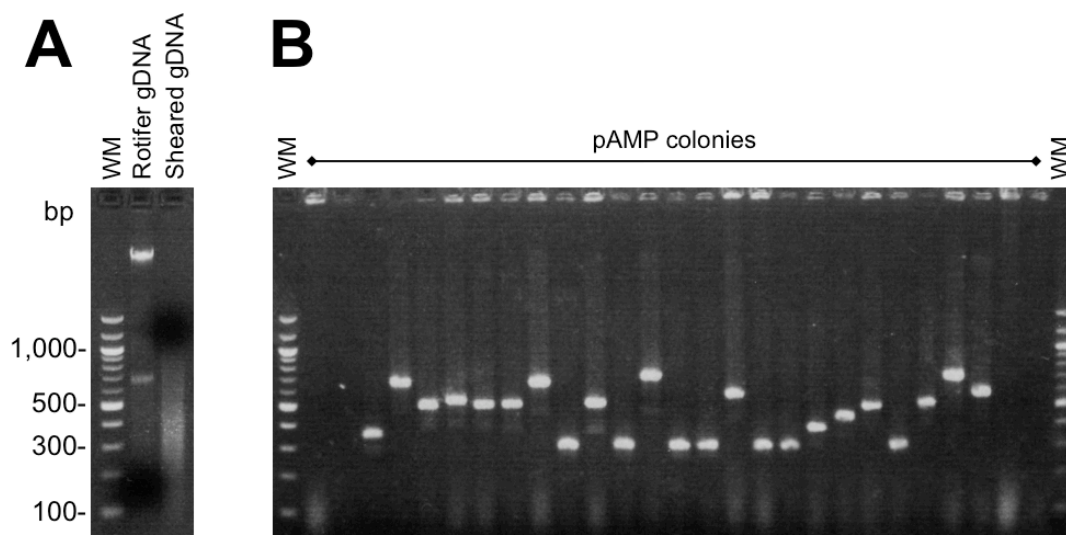


Figure 4.12: Nebulisation of gDNA gives fragments centred around 350 bp in size, leading to inserts ranging from approximately 60 to 400 bp. **A:** Representative agarose gel of sheared *A. ricciae* gDNA. **B:** Colony PCR (see Section 2.2.5.2) shows the presence of inserts in 16 out of 23 colonies assayed. Empty plasmid results in a 300 bp band. WM, DNA ladder.

4.2.4.2 Replica plating of pAMP/S and pAMP gDNA libraries

Replica plating was carried out on white membranes using pAMP/S and pAMP gDNA libraries that had been transformed into *E. coli* TOP10 (see Section 2.6.4). To identify putative novel AMP hits, colony sizes were compared between replica membranes on plus/minus arabinose LB media (see Section 2.6.5). Those that exhibited a reduced size when induced were selected for further analysis. The protocol was designed to minimise false-positives derived from poor replica plating. For a colony to be considered a hit, its corresponding uninduced replicas on both the master and minus arabinose membranes were required to grow. A lack of proper colony transfer between the master and plus arabinose membrane could be ruled out if a colony grew on the minus arabinose membrane, which was the last to be replicated and hence the most likely to reveal experimental error. Figure 4.13 shows a representative of replica membrane processing and the identification of hits.

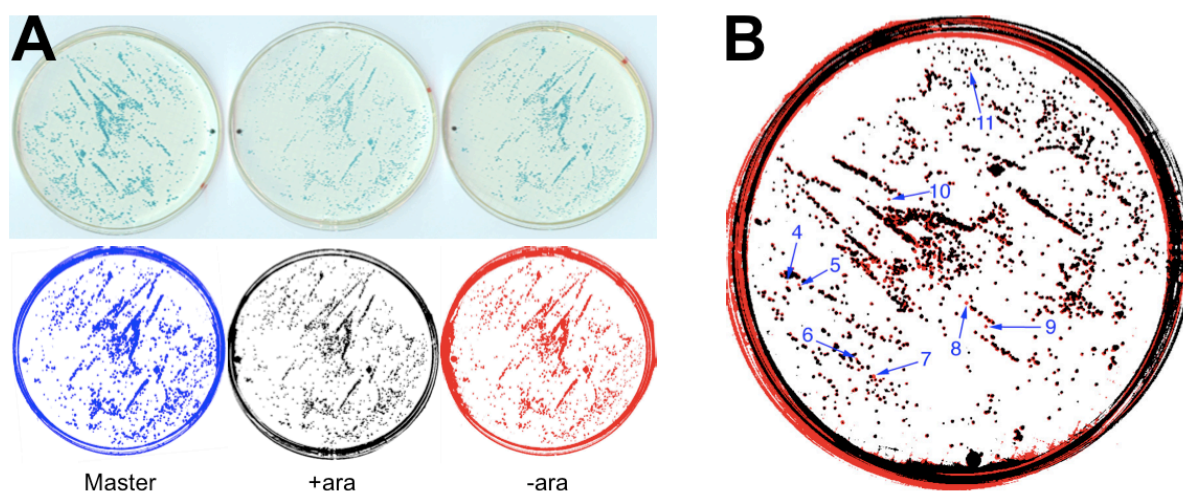


Figure 4.13: Image processing of replica membranes to reveal colonies that may express a novel AMP. Representative experiment ($n = 15$) showing a pAMP/S human gDNA library. **A:** Scans of master, +ara and -ara replica membranes and subsequent image manipulation. **B:** Image overlay of +ara on -ara replica membrane. Potential hits arrowed. Ara, arabinose.

Colonies that were not dispersed adequately on a replica membrane, i.e. those touching neighbours, were ignored. In total, approximately 3,500 discrete pAMP/S gDNA (~800 human, ~2,700 rotifer) and 1,400 discrete pAMP gDNA (~700 human, ~700 rotifer) library colonies were analysed. A number of hits were observed. For the pAMP/S libraries, 11 human and 30 rotifer hits were selected for further analysis; for the pAMP libraries, 11 human and 8 rotifer hits were chosen.

4.2.4.3 Growth curves of putative AMP library hits

To verify that the 60 hit putative AMPs selected in *Section 4.2.4.2* did inhibit *E. coli* growth like the model AMPs K2C18 and K2C38 in *Section 4.2.2*, and to further remove false-positives, the effect of induction in liquid LB was investigated. Cultures of *E. coli* TOP10 pAMP/S and pAMP hits were induced as per *Section 2.3.1* and OD_{600} monitored over time. *Figure 4.14* shows a representative set of growth curves obtained.

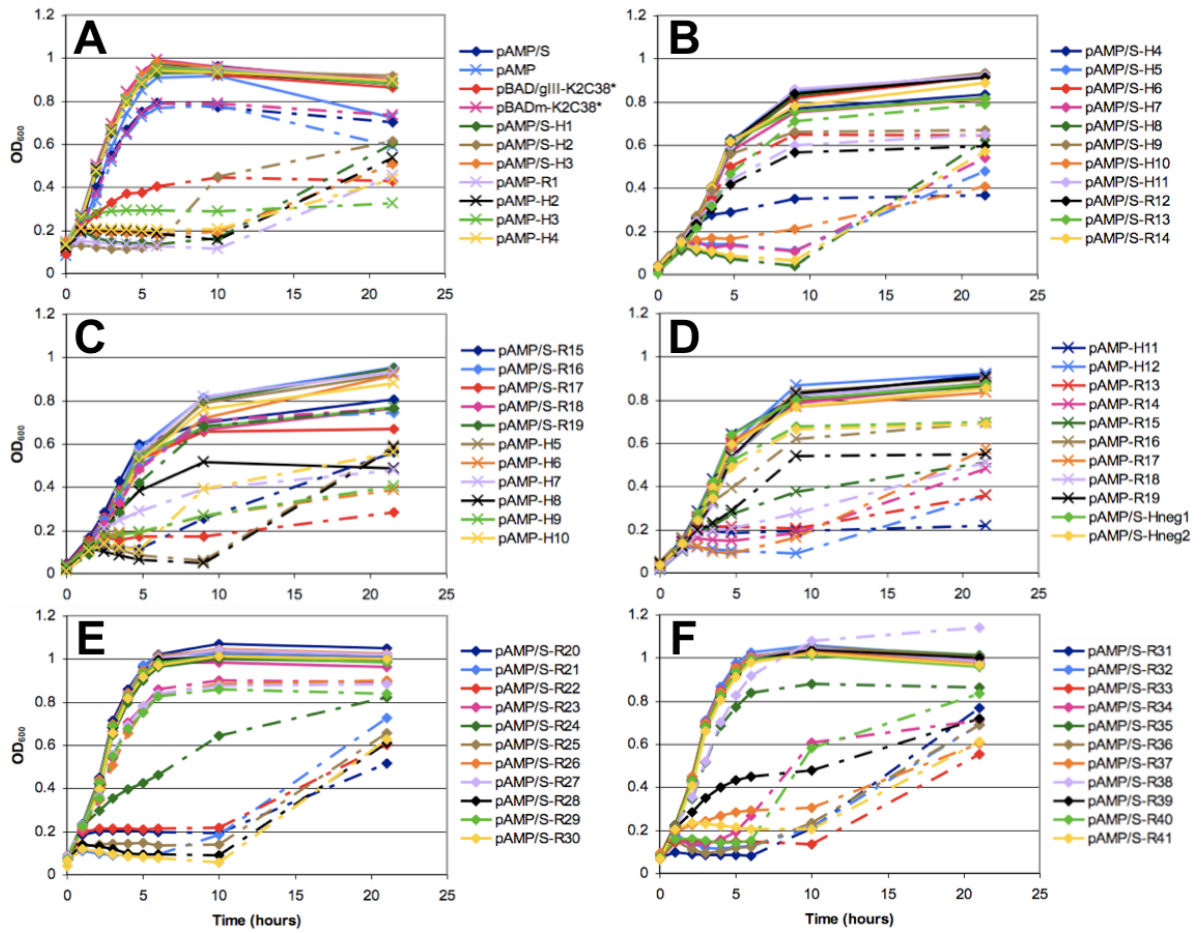


Figure 4.14: Growth curves of pAMP/S and pAMP gDNA library replica plate hits verify that ~77% possess inhibitory activity. Arabinose (0.05% final) added to induce cultures in LB medium at 0 h (dashed lines), incubated at 37°C at ~225 rpm. H, human gDNA library hit; R, rotifer gDNA library hit.

The empty pAMP/S and pAMP vector controls showed no toxicity when induced (see Figure 4.14 A), and the controls pBAD/gIII-K2C38* (targeted to the periplasm) and pBADm-K2C38* (retained in the cytoplasm) behaved as seen previously (cf. Figure 4.4). In addition, two colonies that exhibited no growth inhibition during a pAMP/S human gDNA library screen did not show any effect on growth in liquid culture (pAMP/S-Hneg1 & 2, Figure 4.14 D). However, 14 out of the 60 replica plate hits tested (~23%) also show no inhibition when induced, and thus appear to be false-positives. Although high, this false-positive rate is better than other *in vivo* AMP studies, where rates between 67% and 75% were observed (Loit et al., 2008; Cheng et al., 2009). Many of the putative AMP hits in this screen, be they targeted to the periplasm or not, showed a more immediate impact on growth than pBAD/gIII-K2C38* (Figure 4.14 A) in a manner more similar to Hok (cf. Figure 4.10 A). Restoration of growth after approximately 10 h post-induction, also as seen with

Hok, was commonly observed. pAMP-H3 and pAMP-H11, however, did not recover. Furthermore, isolation of the plasmids from the hits in *Figure 4.14 A* and subsequent retransformation gave the same growth curve profile (see *Appendix 3*), confirming that the inhibitory agent is plasmid-encoded.

Cells from *Figure 4.14 A* showed a variety of morphologies when examined after 4 h induction (see *Figure 4.15*). While pAMP/S-H1 exhibited some long, filamented cells (cf. pBAD/gIII-K2C38*), most were only twice the normal rod length. The induced hit cultures contained cells that appeared slightly swollen, and double the length of the control strains; pAMP-H2 appeared to have visible inclusion bodies at each cell pole.

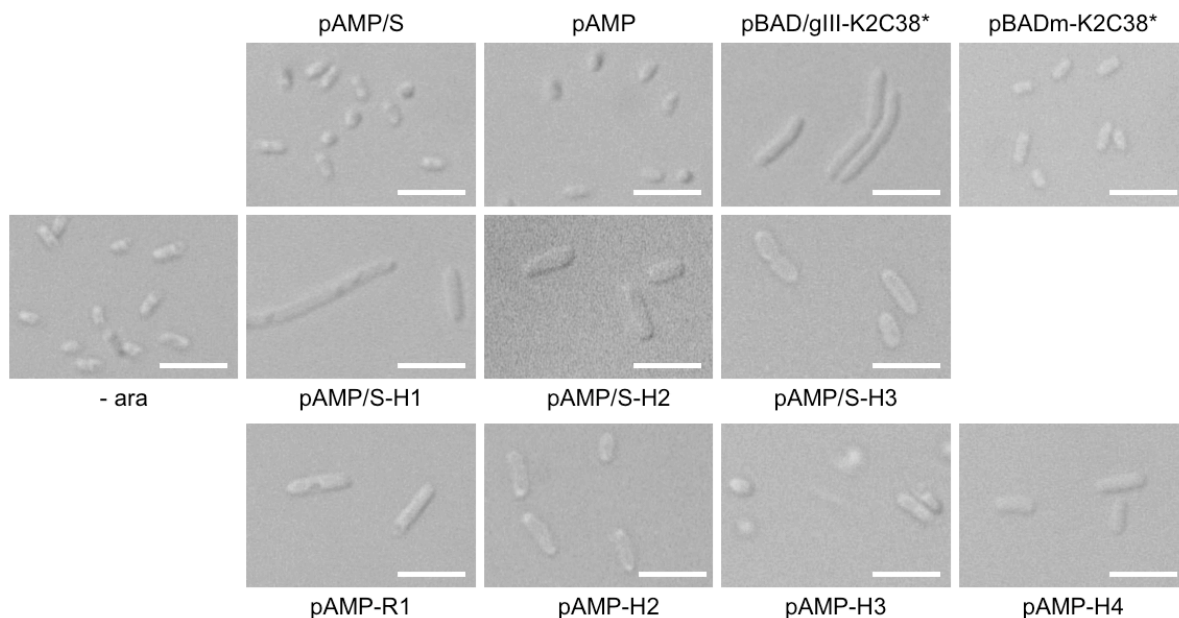


Figure 4.15: E. coli expressing putative AMP hits exhibit varied morphology. Photos taken after 4 h arabinose induction (0.05%) in LB medium at 37°C at ~225 rpm. All uninduced cultures exhibited the same morphology, representative photo shown (- ara). Scale bar, 5 μ m.

Furthermore, some of the pAMP/S and pAMP library hits showed indications of being initially bactericidal. Preliminary data indicated that although all cultures recovered after approximately 10 h induction, plating aliquots of cultures that had only been induced for 4 h on arabinose-free LB agar resulted in an approximate 100-fold reduction in cfu for pAMP/S-1, pAMP/S-2, pAMP-H3 and pAMP-H4 in

comparison to the pAMP/S and pAMP controls (data not shown). Further investigation is required to confirm this effect.

4.2.4.4 Bioinformatic analysis of pAMP/S and pAMP gDNA library hits

A number of putative AMP hits from the replica plate screening (*Section 4.2.4.2*) that had been further validated by liquid culture growth curves (*Section 4.2.4.3*) had their gDNA insert isolated and sequenced. In total, 26 pAMP/S hits (8 from the human gDNA libraries, 18 rotifer) and 14 pAMP hits (9 human, 5 rotifer), representing an initial hit rate of approximately 0.8% of all library colonies examined (~5,000), were analysed as per *Section 2.6.6*. *Table 4.2* summarises the findings; full DNA insert sequences are shown in *Appendix 4*.

| Name | bp | Peptide sequence (# aa) | Q | H | α | A | Accession # | Name | E-value | Reading frame |
|------------------|------|---|-----|-------|----------|-----|-------------|--|----------|---------------------|
| pBAD/gIII-K2C18* | 114 | <u>TMGKKLKKIGQKIKNFFQKL</u> (20) | +7 | -0.27 | + | +++ | AF035680.1 | CRAMP gene (<i>Mus musculus</i> [mouse]) | 2.60E-02 | In-frame (aa 3-20) |
| pBAD/gIII-K2C38* | 54 | <u>TMAISRLAGLLRKGGEKIGKKLKKIGQKIKNFFQKLVQPPE</u> (41) | +9 | -0.22 | + | +++ | AF035680.1 | CRAMP gene (<i>M. musculus</i>) | 6.00E-16 | In-frame (aa 4-41) |
| pAMP/S-H1 | 389 | <u>TMSRNAVESGRESNGEADGMQW</u> (22) | -2 | -0.30 | | | EF445003.1 | NOL1/NOP2/Sun (NSUN6) gene (<i>Homo sapiens</i>) | 1.00E-97 | Intronic |
| pAMP/S-H2 | 130 | <u>TMSLLVLFQNKYDLKIEAIVQIEFWNRRL</u> (30) | +1 | -0.09 | | | NG_007966.1 | Cytochrome P450 (CYP4X1) gene (<i>H. sapiens</i>) | 4.00E-21 | Intronic |
| pAMP/S-H3 | 218 | <u>TMSVLNEFKSQCHLVFHAFIHQTLPECLPWAICYTEQGRNGHSQLPHSRISSSSGGVRKGNKHKMFSGILRMPISPLSK</u> (79) | +12 | -0.13 | + | + | AL591603.4 | Chromosome 6 (<i>H. sapiens</i>) | 2.00E-44 | Non-coding |
| pAMP/S-H4 | 136 | <u>TMSKRYLGTRCLSLMWAKGDLRQIKMGKRGDKKE</u> (34) | +7 | -0.36 | + | + | AL158165.15 | Chromosome 10 (<i>H. sapiens</i>) | 4.00E-23 | Non-coding |
| pAMP/S-H5 | 179 | <u>TMSFSKFLFLLHFIHLIFIH</u> (20) | +4 | 0.23 | + | | AL35462.23 | Chromosome 9 (<i>H. sapiens</i>) | 3.00E-32 | Non-coding |
| pAMP/S-H7 | 105 | <u>TMSERSRHERSYIVEFHLYAMFRLSQSTEIESR</u> (33) | +2 | -0.30 | + | | AC245260.1 | Chromosome 8 (<i>H. sapiens</i>) | 3.00E-15 | Non-coding |
| pAMP/S-H8 | 384 | <u>TMSVIVLIYISTKKVQGFIFLHILASICVARHSDISHLNWGDASHCSFDLHFSIGDIEHLFICLFAIFMSSFEEKRFRSFAHFFNEVIRFFFLQS</u> (96) | +7 | 0.04 | + | + | AC207610.3 | Chromosome 12 (<i>H. sapiens</i>) | 3.00E-84 | Non-coding |
| pAMP/S-H10 | 1186 | <u>TMSLWSDSMADVILILLNLLRLALWQSMWLILEYGPCTDEKDAYSMVEGYSVDVY</u> (55) | -6 | 0.02 | + | | AC067849.6 | Chromosome 8 (<i>H. sapiens</i>) | 0.00E+00 | Non-coding |
| pAMP/S-R14 | 171 | <u>TMSKLNMQIIIGLVLIDIALFKSLEICKNICCKSNSLRKALLDICLFLAAMICISAT</u> (58) | +3 | 0.05 | + | | | | | |
| pAMP/S-R15 | 175 | <u>TMSRNSFSMLLLAINLIFLIVGLISIIINNGKMENIFLQVNPKVILFTWYKQARRIYELVVIC</u> (62) | +4 | 0.04 | + | + | | | | |
| pAMP/S-R17 | 235 | <u>TMSEDYFRRTLNIQVDIRLIDENEIEAASIPNLEEILPGKPLIHFRHEPMISIRLINRQPYSGHFEWTIPVINGDTVEKLG</u> (82) | -3 | -0.14 | | | BC006060.1 | Leucyl-tRNA synthetase gene (<i>M. musculus</i>) | 3.00E-06 | In-frame (aa 6-81) |
| pAMP/S-R20 | 164 | <u>TMSLHQITVMCRKLLLNLFQGFIFGCISEFYQRR</u> (35) | +4 | -0.02 | + | | AK371870.1 | Hypothetical protein gene (<i>Hordeum vulgare</i> [barley]) | 5.00E-15 | Wrong reading frame |
| pAMP/S-R21 | 367 | <u>TMSFEVI</u> (7) | -1 | 0.15 | | | | | | |
| pAMP/S-R22 | 116 | <u>TMSYSKTITVKTNIIIGLTYQ</u> (21) | +2 | -0.02 | | | | | | |
| pAMP/S-R25 | 175 | <u>TMSLLYSLRRRCWPFVICQRAMLLLTSAVGVFRIIIRLFFDLFSQTSWIFLE</u> (53) | +4 | 0.02 | + | | | | | |

| Name | bp | Peptide sequence (# aa) | Q | H | α | A | Accession # | Name | E-value | Reading frame |
|------------|-----|--|-----|-------|----------|---|--------------------|--|----------|---------------------|
| pAMP/S-R28 | 347 | <u>TMSEWKCTYSEYLAVERSVFETQMNTRKVLFPFYTRGIVDQQLSLFLSLSLFLS</u> ERASFK (62) | 0 | -0.15 | + | | | | | |
| pAMP/S-R30 | 145 | <u>TMSLATMLQIGAMLIFFTGHIIDTIGRRRSIHLITALLLITSLITQACLQFG</u> (52) | +4 | 0.09 | + | | | | | |
| pAMP/S-R31 | 373 | <u>TMSIEIDLVLIIHIHTHTRVSLGIEIDLVLIIHIHTHTPVS LGIEIDLVLIIHIHT</u> HTPVSLGIEIDLVLIIHTYPHTPVSLGIEIDLVLIIHTYPHTPVSLGIEIDLVLII HTYPHTPVSLGIEIDC (128) | +2 | 0.12 | | | AC139335.3 | Chromosome 10 (<i>M. musculus</i>). Similar due to repetitiveness. | 1.00E-17 | Non-coding |
| pAMP/S-R32 | 238 | <u>TMSLFSLLSLSLSLFFFYERNKK</u> (24) | +2 | -0.01 | + | | | | | |
| pAMP/S-R33 | 108 | <u>TMSCWIFLHSLLYSIQVSFLRRFLFLIPVPDKSDQRKCAK</u> (41) | +5 | -0.07 | + | | | | | |
| pAMP/S-R34 | 168 | <u>TMSYFELNIRPCRSASARRRRSIAAFSFRSMGRSSLYISVSGYKNERMLCILKTF</u> SFKAK (61) | +11 | -0.23 | + | | NM_00103259 0.1 | Protofilament ribbon protein (pf-rib) gene (<i>Ciona intestinalis</i> [sea squirt]) | 9.00E-10 | Wrong reading frame |
| pAMP/S-R36 | 199 | <u>TMSIATFCMLSLIFVTFHKPLVRNIWGTIKSLLNFIFYMRKKK</u> (43) | +8 | -0.01 | + | + | | | | |
| pAMP/S-R37 | 132 | <u>TMSSHIIHFYSMELFHLCSCLFLDVGQYRIFVKSVTQHDNRARHIPLAK</u> (49) | +7 | -0.08 | + | | | | | |
| pAMP/S-R39 | 123 | <u>TMSDRVTAFEFHNVYVLSNFRYRLDVQEQYIRSPSYSN</u> (38) | +1 | -0.19 | | | | | | |
| pAMP/S-R40 | 222 | <u>TMSELISKLFILYRSSELLISLSITSNSASLACLVEIEPFVPSKLSNLVLFES</u> CSLI (60) | -1 | 0.03 | + | | | | | |
| pAMP/S-R41 | 161 | <u>TMSNLLFQLCLNKKFNIVFSNSNVNMNLVLPVVVDDKVVSLLPACK</u> (48) | +3 | -0.01 | + | | | | | |
| pAMP-R1 | 337 | <u>TMSFYVSFLLSFFALVTEQNIYQCSFVCLSPLLRFLFIPFVDI</u> (42) | -1 | 0.14 | + | | | | | |
| pAMP-H2 | 383 | <u>TMSLFLFFFFF</u> (11) | 0 | 0.41 | | | AC090383.4 | Chromosome 18 (<i>H. sapiens</i>) | 3.00E-86 | Non-coding |
| pAMP-H3 | 334 | <u>TMSSLAEHYIQLKENVVSWNTDFGDDAQPRNRI</u> (33) | -1 | -0.22 | + | + | AC026887.9 | Chromosome 10 (<i>H. sapiens</i>) | 2.00E-70 | Non-coding |
| pAMP-H4 | 268 | <u>TMSTWPILAALGFRTVFIFFGEEVM</u> (25) | -1 | 0.15 | + | | AC020593.6 | Chromosome 4 (<i>H. sapiens</i>) | 6.00E-56 | Non-coding |
| pAMP-H5 | 273 | <u>TMSNKNGCYQKPVPLLAHVHCYLITMMLSAGYGLPLTTAEAWTLFAQMIFSSSF</u> CFNYVHEGSGWRRPGRGRSGYSTEHNVYFFSILLTSIKAK (96) | +8 | -0.04 | + | | AC016877.11 | Chromosome 8 (<i>H. sapiens</i>) | 5.00E-58 | Non-coding |
| pAMP-H6 | 362 | <u>TMSGGQGRGR</u> (10) | +2 | -0.37 | | | GU324919.1 | Myosin heavy chain 6 (MYH6) gene (<i>H. sapiens</i>) | 3.00E-80 | Intronic |

The insert length of the gDNA in each hit varied (range of 105 to 1,186 bp), as did the length of peptide encoded (range of 3 to 141 aa) (see *Figure 4.16*). The bimodal nature of *Figure 4.16 A* can be attributed to different gDNA nebulisation preparations, i.e. different lengths of insert. As a number of gDNA inserts contained in-frame stop codons, the median encoded peptide length of 48 aa was less than that predicted by the median gDNA insertion length of ~220 bp (approximately 73 aa). Approximately 75% of the encoded products fell between the 10 to 62 aa peptide range, giving a suitable peptide size spread.

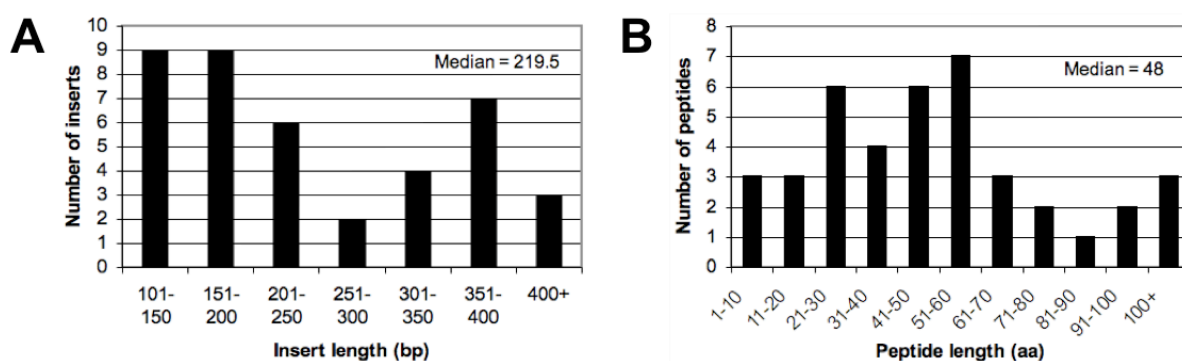


Figure 4.16: Range of gDNA library insertion lengths and peptide lengths coded for (see Table 4.2). A: Insert DNA length. B: Peptide sequence length.

Peptides in *Table 4.2* with a mean hydrophobicity of -0.1 or less were considered to be truly hydrophilic, with 33% of the peptides meeting this criterion. These hydrophilic peptides were expected to be more amenable to downstream purification, i.e. soluble, as opposed to other very hydrophobic peptides (e.g. pAMP-H2, which is mostly phenylalanine and leucine). No significant bias for hydrophobicity was observed between the screening vectors or gDNA insert sources used. The charges of the peptides ranged from the anionic (-6 for pAMP/S-H10) to the cationic (+12 for pAMP/S-H3); for peptides <50 aa, the extremes were -2 for pAMP/S-H1 and +7 for pAMP/S-H4. Approximately 73% of the peptides analysed were cationic in charge.

Of the 40 putative AMP sequences in *Table 4.2*, 30 are predicted to encode a peptide that contains an α -helical region. Peptides >25 aa were also predicted to contain unstructured regions, i.e. random coil and extended strand. However, only 7 peptides were considered to possess a potentially amphipathic α -helical region

when projected on a helical wheel; most of these weakly so (see *Figure 4.17*). This indicates that the growth inhibitory mechanism of the putative AMPs may not be the same as that of the model amphipathic α -helical K2C18 control.

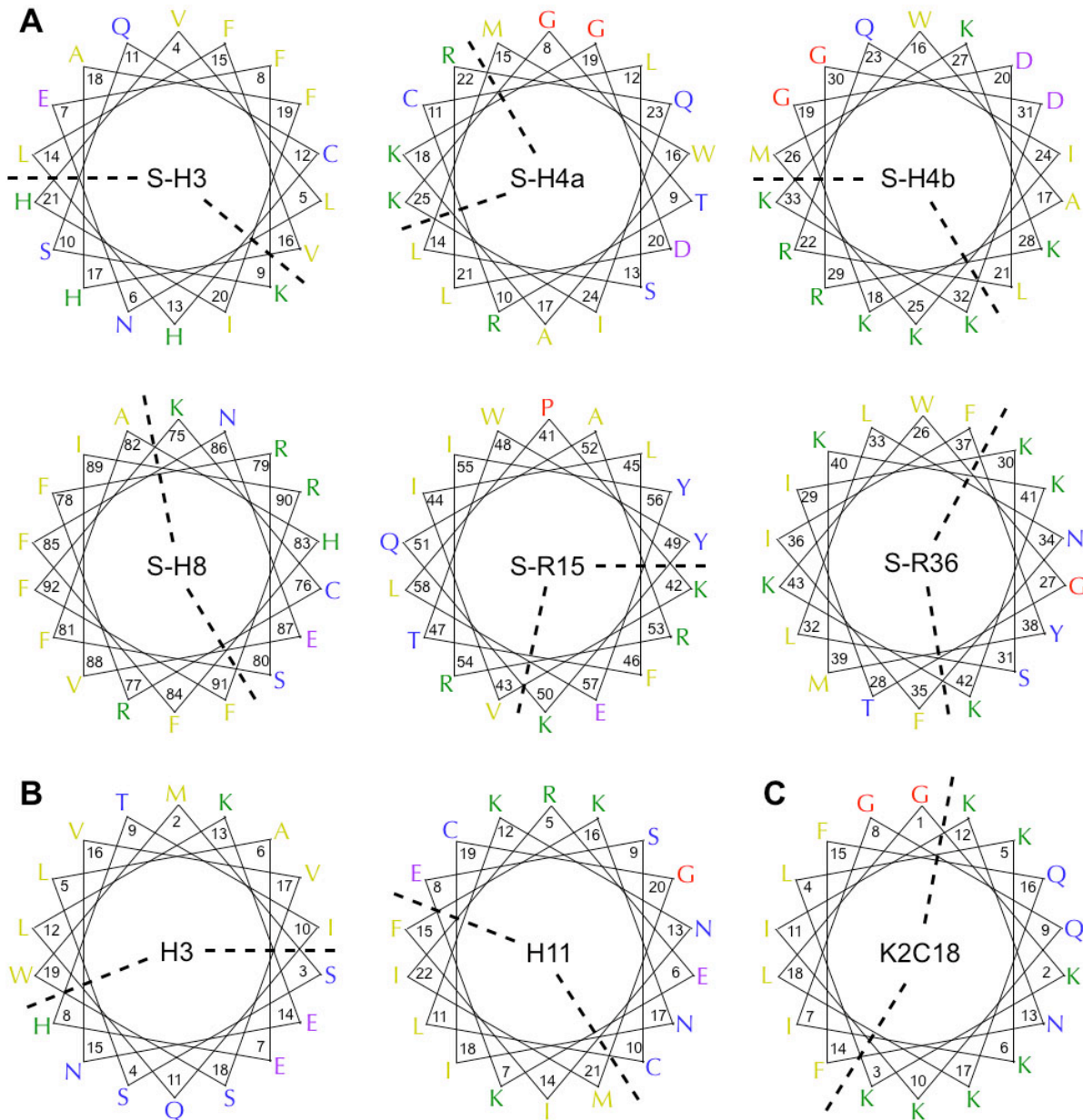


Figure 4.17: Helical wheel projections of selected putative amphipathic AMP hits. Peptide sequences taken from Table 4.2, residue positions indicated. **A:** pAMP/S hits, i.e. targeted to the periplasm, including two regions of S-H4 (34 aa). **B:** pAMP hits. **C:** K2C18 as an ideal amphipathic AMP reference. Hydrophilic face boundary indicated with dashed line, amino acids are colour coded: cationic (green), hydrophobic (yellow), anionic (purple) and uncharged hydrophilic (blue).

BLAST analysis (tBLASTx; see Section 2.6.6) of the putative AMP encoding sequences in Table 4.2 was able to find genomic matches for all human-derived

DNA inserts (17 from 17), as opposed to the rotifer-derived set (7 from 23 gave moderately significant matches). As the genome of *A. ricciae* is not currently annotated (C. Boschetti, personal communication), this is unsurprising. Of those 24 peptide sequences successfully matched, few were naturally in-frame with a known coding region; 19 were from non-coding or intronic regions instead. Three rotifer-derived sequences were in the wrong reading frame, while two were in-frame (pAMP/S-R17 matching a mouse transfer RNA synthetase; pAMP-R17 a nematode hypothetical protein).

Four peptides in *Table 4.2* were predicted to encode very short peptides, which may be false-positives (<12 aa; pAMP/S-R21, pAMP-H2, pAMP-H6 and pAMP-R18). For example, pAMP-R18 possesses a stop codon as its first codon, with the next potential start codon (including the alternatives GTG and TTG [Blattner et al., 1997]) being too far from the ribosome binding site (31 bp) (Ringquist et al., 1992; Chen et al., 1994). pAMP-R18 thus only encodes the tripeptide Thr-Met-Ser; in contrast, the empty pAMP vector encodes Thr-Met-Ser-Ala-Lys, which has no effect on *E. coli* growth (see *Figure 4.14 A*). It was thought that perhaps the gDNA inserts of the above hits (available in *Appendix 4*) exerted their inhibitory effect as RNA transcripts instead, either by complementing and preventing the translation of an endogenous *E. coli* transcript (Gottesman, 2004; Morita et al., 2006) or by folding into an active secondary structure (Isaacs et al., 2006). While no significant transcript matches between the hit sequences and the *E. coli* DH10B genome were found, all transcript sequences were predicted to form a significant amount of secondary structure, i.e. intrasequence base pairing (*Table 4.3*).

| Name | aa | bp | BLASTn | RNA 2° |
|------------|----|-----|--------|--------|
| pAMP/S-R21 | 7 | 367 | - | + |
| pAMP-H2 | 11 | 383 | - | + |
| pAMP-H6 | 10 | 362 | - | + |
| pAMP-R18 | 3 | 129 | - | + |

Table 4.3: Analysis of pAMP/S and pAMP hits that may be RNA-based effectors. aa, putative peptide length; bp, length of gDNA insert (see Appendix 4 for sequence); BLASTn; significant reverse complement nucleotide query versus *E. coli* DH10B genome (see Section 2.6.6); RNA 2°, prediction of a RNA secondary structure by RNAfold (Hofacker, 2003).

4.2.5 Further exploration of novel putative AMP hits

4.2.5.1 Frame-shift of pAMP/S library hits

Analysis of the putative AMP hits in *Section 4.2.4.4* revealed that some of the selected pAMP/S and pAMP plasmids only encoded very short peptides, indicating that their effect may instead be transcript related. To investigate whether or not this was also true for constructs that encoded longer peptides, several pAMP/S library hits were mutated to introduce a frame-shift mutation immediately downstream of the start codon (ATG to ATGG; see *Section 2.6.7.1*). It was reasoned that, if a frame-shifted construct maintained a negative effect on the growth rate of *E. coli* when induced, the wild-type encoded peptide would not be the cause. Cultures of *E. coli* TOP10 with frame-shifted pAMP/S hits were induced as in *Section 4.2.4.3* and OD₆₀₀ monitored over time (*Figure 4.18*).

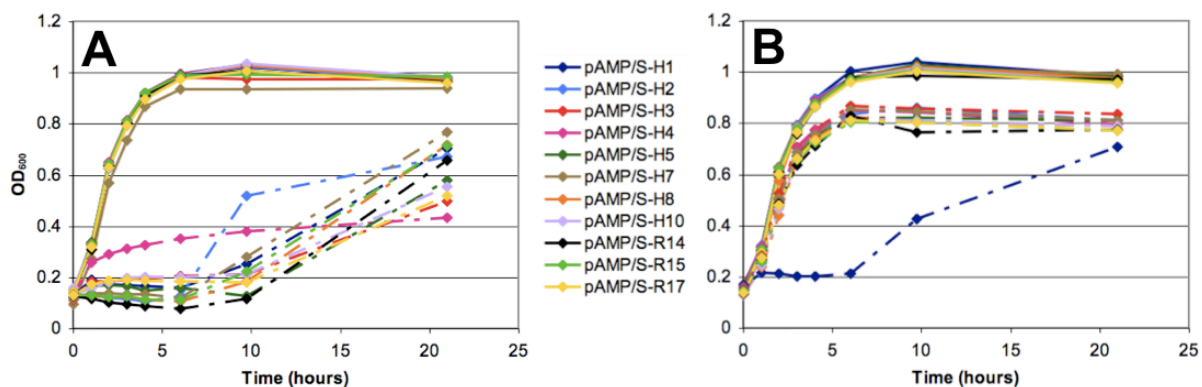


Figure 4.18: Frame-shift mutation of several pAMP/S hits reveals that pAMP/S-H1 retains inhibition of *E. coli* growth. Arabinose (0.05%) added to induce cultures in LB medium at 0 h (dashed lines), incubated at 37°C at ~225 rpm. **A:** Wild-type growth curves. **B:** Frame-shift mutant growth curves. Representative experiments shown ($n = 2$).

All of the frame-shifted pAMP/S hits analysed lost their growth inhibitory effect (other than the slight reduction in growth typically seen during arabinose induction) except for pAMP/S-H1, which retained activity. While this construct should encode a 22 aa peptide (see *Table 4.2*), it appears to have an RNA-based effect instead. Point-mutations to remove the start codon (ATG to CTG) from the wild-type constructs listed in *Figure 4.18* were also made, and resulted in similar growth curves to those observed above for the frame-shift mutations (data not shown). Analysis of the pAMP/S-H1 sequence as a possible RNA effector was carried out as per *Section 4.2.4.4* above, but no potential *E. coli* antisense sequence or secondary structure was found.

4.2.5.2 Removal of the secretion signal from pAMP/S library hits

The N-terminal pIII secretion signal used in pAMP/S allows for complete synthesis of the downstream peptide prior to secretion to the periplasm (Natale et al., 2008). This window of cytoplasmic retention time allows for the possibility that pAMP/S peptide hits have a cytoplasmic, rather than periplasmic, target. Previously it was established that K2C38 required the pIII secretion signal to exert an effect on *E. coli* growth; when this was removed, no inhibition was observed (see Figure 4.4). Removal of the pIII-encoding sequence (i.e. gIII) from the pAMP/S hits could thus confirm whether or not its target was periplasmic. Mutagenesis was conducted as per Section 2.6.7.3, and the resulting growth curves (as per Section 4.2.4.3) shown in Figure 4.19.

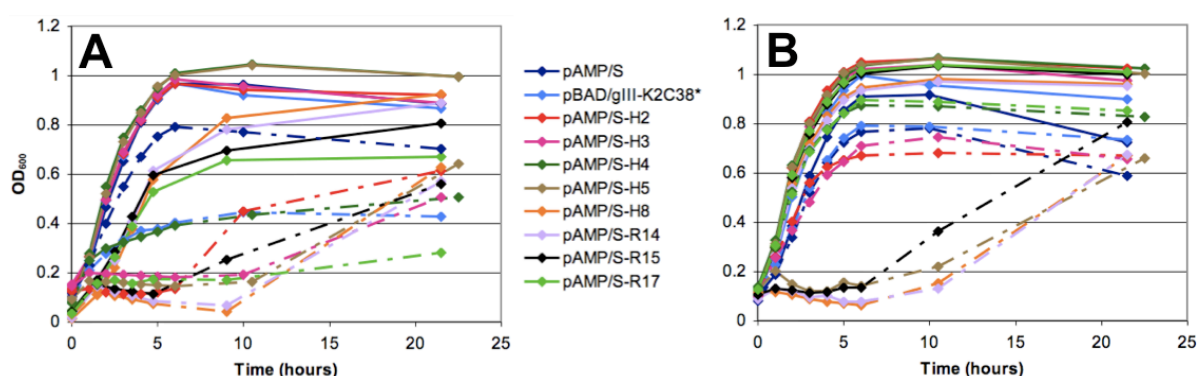


Figure 4.19: Removal of the pIII secretion signal from several pAMP/S hits reveals some that retain inhibition of *E. coli* growth. Arabinose (0.05%) added to induce cultures in LB medium at 0 h (dashed lines), incubated at 37°C at ~225 rpm. **A:** Wild-type growth curves. **B:** Mutants lacking the pIII secretion signal.

Removal of the pIII secretion signal from pAMP/S-H2, pAMP/S-H3, pAMP/S-H4 and pAMP/S-R17 led to the loss of growth inhibition (see Figure 4.18 B), indicating that access to the periplasm is required for their activity. These may be truly exogenous hits. In contrast, pAMP/S-H5, pAMP/S-H8, pAMP/S-R14 and pAMP/S-R15 retained their original level of inhibitive activity without the pIII secretion signal. This indicates that the pAMP/S screen does not solely select for putative AMPs that act in the periplasm; intracellular activity is also isolated.

4.2.5.3 Use of a different antibiotic selection marker

The identified AMP hits, rather than having a truly antimicrobial function against their *E. coli* host, could instead interfere with the production or function of the

plasmid-encoded β -lactamase antibiotic resistance marker. This could lead to false positives in the AMP screening system by restoring *E. coli* sensitivity to carbenicillin. To briefly explore this, the β -lactamase operon (*ampR*) was replaced with the chloramphenicol acetyl transferase operon (*chlR*) in pAMP/S-H1 and pAMP/S-H5 as per Section 2.6.7.2. Growth curves were then performed as in Section 4.2.4.3; no differences were observed with regards to the *ampR* controls (Figure 4.20).

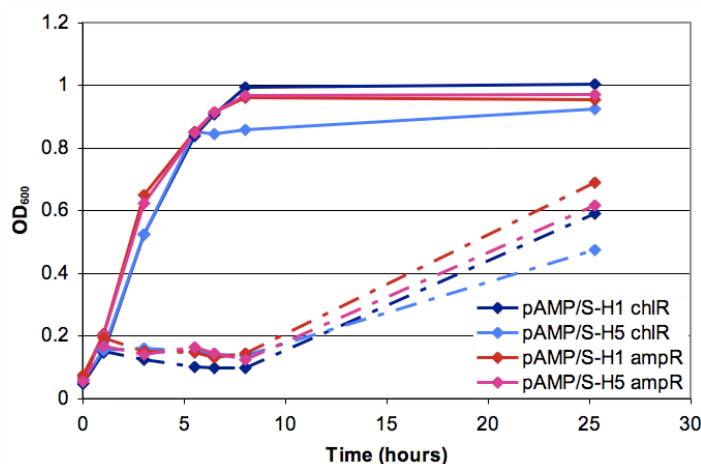


Figure 4.20: pAMP/S-H1 & pAMP/S-H5 still inhibit *E. coli* growth when a different antibiotic resistance marker is used. Arabinose (0.05%) added to induce cultures in LB medium at 0 h (dashed lines), incubated at 37°C at ~225 rpm. *chlR*, chloramphenicol resistant; *ampR*, carbenicillin resistant. Representative experiment shown ($n = 2$).

4.2.6 Validation of novel putative AMP hits

4.2.6.1 Recombinant production of putative AMPs using CPD system

The CPD system outlined in Chapter 3 was utilised in order to produce several of the pAMP/S hits identified in Section 4.2.4. Although their status as *bona fide* peptide hits that act periplasmically was not confirmed (see Section 4.2.5), it was reasoned that if these putative AMPs could be purified, and retained activity when added exogenously to *E. coli*, their target would most likely be membrane-relevant. To explore this, the coding sequences for nine pAMP/S hits were transferred to the pET28-CPD expression vector (see Section 2.6.8.1). The wild-type CPD cleavage sequence of Leu-Ala-Asp was retained in order to try to ensure suitable levels of recombinant protein production (see discussion in Section 3.3). After expression and purification (see Section 2.6.8.2 and 2.4.2 respectively), initial levels of antimicrobial activity for each putative AMP against *E. coli* TOP10 were assessed

using the antimicrobial liquid culture assay (see Section 2.5.3). Table 4.4 outlines the results.

| Plasmid name | Peptide (# aa) | H | Hyphb. | Expression | Purification | Activity |
|-----------------|---|-------|--------|------------|--------------|----------|
| pET28-S-H2-CPD | LLVLFQNKYDLKIEAIVQIEFWNRTRL (27) | -0.10 | 52% | + | - | |
| pET28-S-H3-CPD | VLNEFKSQCHLQVHAFIHQTLPECLPWAIYCT EQGRNGHSQQLPHSRISSSSGGVRKGNKHKMKS GILRMPSIPLSKL (77) | -0.07 | 31% | + | + | - |
| pET28-S-H4-CPD | KRYLGTRCLSLMWAKGDLRQIKMGKRGDKKEL (32) | -0.36 | 31% | + | + | - |
| pET28-S-R17-CPD | EDYFRRTLNIQVDIRLIDENEIEAASIPNLE EILPGKPLIHFRHEPMISIRLINRQPYSGHFE WTIPVINGDTVEKL (78) | -0.18 | 37% | + | + | - |
| pET28-S-R22-CPD | YSKTITVKTNIIGLTYQL (19) | 0.01 | 47% | + | - | |
| pET28-S-R25-CPD | LLYSLRRRCWPFVICFQRAMLLLSAVGVFRI IIRLFFDLFSQTSWIFLEL (51) | 0.01 | 57% | - | | |
| pET28-S-R32-CPD | LFSLLSLSLLEFFFYERNKKL (22) | 0.01 | 59% | - | | |
| pET28-S-R33-CPD | CWIFLHSLLYSIQVSPFLRRFLLFLIPVPDKSD QRKCAKL (39) | -0.05 | 49% | - | | |
| pET28-S-R39-CPD | DRVTAFEFHNVYVLSNFRYRLDVQEYIRSPS YSNL (36) | -0.19 | 44% | + | - | |

Table 4.4: Several pAMP/S hits were transferred to the CPD purification system, but no exogenous activity was observed. Common N-terminal residues Thr-Met-Ser removed, non-native C-terminal leucine (critical for CPD cleavage) indicated in bold. H, mean hydrophobicity (see Section 2.4.4); Hyphb., % of hydrophobic residues (F, I, L, N, V, W & Y); Expression, whole cell production of AMP-CPD fusion assessed by SDS-PAGE (data not shown); Purification, free AMP band detected by SDS-PAGE after CPD cleavage (data not shown); Activity, determined by any purification fraction completely inhibiting *E. coli* DH10B growth in LB medium after 18 h at 37°C.

Of the nine pAMP/S hits transferred to the pET28-CPD expression vector, only six were seen to successfully express the AMP-CPD fusion by SDS-PAGE analysis (data not shown). Of these, only three gave appreciable amounts of free AMP after on-column CPD cleavage, i.e. detected by SDS-PAGE (data not shown). Table 4.4 indicates the mean hydrophobicity and percentage of hydrophobic residues of each peptide. It may be that overly hydrophobic peptides, with H values greater than -0.1 (e.g. S-R25 and S-R32), are unsuitable for recombinant production.

None of the putative AMPs observed in significant amounts post-cleavage (S-H3, S-H4 or S-R17) showed antimicrobial activity when added exogenously to *E. coli*. In contrast, the model AMP K2C18 completely inhibited growth (see Section 3.2.2.3). These putative AMPs may be inactive, or simply not concentrated enough to exert

an antimicrobial effect; concentration estimates in the initial cleavage fractions ranged from ~10 μM (S-R17) to ~60 μM (S-H4). In the antimicrobial liquid assay, this would equate to a working concentration range of ~1 to ~6 μM . To investigate further, chemically-synthesised versions of selected putative AMPs were purchased.

4.2.6.2 Activity of chemically-synthesised putative AMPs

Several AMP hits from *Section 4.2.4* were chemically synthesised (see *Section 2.1.4*) and used to challenge *E. coli* DH10B in an antimicrobial liquid culture assay (see *Section 2.5.3*). Due to synthesis costs, short peptides (<35 aa) were focused on. S-H1, which was discovered to exert its effect as a transcript rather than a peptide (see *Figure 4.17*), was used as a negative control. K2C18 and S-H4 were resuspended in sterile water, while S-H1, S-H5, H4 and R13 were resuspended in neat DMSO to aid solubility. DMSO alone was tolerated to 2.5% in the antimicrobial liquid culture assay, and therefore final dilutions below this concentration were used. *Table 4.5* outlines the peptides, their properties, and any antimicrobial activity observed over a concentration range of ~50 μM to ~0.1 μM .

| Name | Peptide (# aa) | Q | H | Hyphb. | MIC <i>E. coli</i> |
|-------|---|----|-------|--------|-----------------------|
| K2C18 | GKKLKKIGQKIKNFFQKL (18) | +7 | -0.31 | 33% | 0.6 μM |
| S-H1 | <u>TMSR</u> NAVESGRESNGEADGMQW (22) | -2 | -0.30 | 18% | >50 μM |
| S-H4 | KRYLGTRCLSLMWAKGDLRQIKMGKRGDKKE (31) | +7 | -0.39 | 29% | 7.8 μM |
| S-H5 | <u>TMSF</u> SKFLFLHLHFIFIH (20) | +4 | 0.23 | 65% | >50 μM |
| H4 | <u>MTM</u> STWPILAALGFRTVFIFFGEEVM (26) | -1 | 0.15 | 54% | >50 μM |
| R13 | <u>MTMS</u> ILQILYRVIFILNVIVVHIHYV (26) | +3 | 0.21 | 73% | >50 μM |

Table 4.5: S-H4 exerts an antimicrobial effect against E. coli. Minimum inhibitory concentration (MIC) was determined as the concentration of peptide that completely inhibited *E. coli* DH10B growth in NB medium after 18 h at 37°C. Experiment repeated at least in duplicate, average shown. Underlined residues encoded by pAMP/S or pAMP vector. Q, charge; H, mean hydrophobicity (see *Section 2.4.4*); Hyphb., % of hydrophobic residues (F, I, L, N, V, W & Y).

The K2C18 positive control exhibited a potent MIC against *E. coli* as seen previously (*Section 3.2.5*); S-H1, acting as a negative control, showed no effect.

S-H4 (for amino acid analysis, see *Appendix 5*), previously established to require targeting to the periplasm for antimicrobial activity when expressed endogenously (*Section 4.2.5.1*), was found to inhibit growth. Agar plating of *E. coli* aliquots treated with S-H4 (31 μ M) after 21 h gave no viable colonies in comparison to the negative control (S-H1), indicating cell death (data not shown; further confirmation is required to determine the minimal bactericidal concentration). H4 and R13, sourced from the intracellular pAMP library, showed no activity. This was not surprising, as the cell membrane was expected to block access to their putative intracellular targets. S-H5, which retained growth inhibitory properties in spite of the removal of its secretion signal (*Section 4.2.5.2*), also lacked antimicrobial activity for the same reason when added exogenously. Additionally, S-H4 was also observed to exert an antimicrobial effect on the Gram-negative bacterium *Pseudomonas putida* (MIC of 11.6 μ M) and the Gram-positive bacterium *Bacillus subtilis* (MIC of 31 μ M), but not against the yeast *Candida albicans*.

4.2.6.3 Synergy of hits with cell wall permeabilizing reagents

The putative AMPs S-H5, H4 and R13 were hypothesised to require access to the cytoplasm for any antimicrobial effect to be seen. In order to facilitate this, the use of reagents that permeabilize the cell wall of *E. coli* was explored. The non-ionic detergent Nonidet P-40 (NP-40) is commonly used in the purification of membrane complexes (Fricke et al., 1999). Polymyxin B nonapeptide (PMBN), a truncation of the cationic polymyxin B antibiotic (produced by *Bacillus polymyxa*), has been used to selectively increase the permeability of the outer membrane of Gram-negative bacteria to other compounds while retaining cell viability (Vaara, 1992; Li et al., 1998). Lastly, the model AMP K2C18, which is known to disrupt the outer membrane during self-mediated uptake (Sawyer et al., 1988), was selected. The three cell-permeabilizing reagents had their individual MICs against *E. coli* DH10B determined using the antimicrobial liquid culture assay (see *Section 2.5.3*): NP-40 was toxic at 0.25%, PMBN at 4 μ g/mL, and chemically-synthesised K2C18 at 0.6 μ M. These compounds were therefore used at the sub-lethal concentrations of 0.01% (NP-40), 1 μ g/mL (PMBN) and 0.23 μ M (K2C18). Antimicrobial liquid culture assays were performed (see *Section 2.5.3*) in combination with the AMPs listed in *Table 4.6*.

| Name | Peptide (# aa) | Neg. | NP-40 (0.01%) | PMBN (1 µg/mL) | K2C18 (0.23 µM) |
|---------|---|------|------------------|-------------------|--------------------|
| Control | GKKLKKIGQKIKNFFQKL (18) | + | + | + | + |
| S-H1 | <u>TMSR</u> NAVESGRESNGEADGMQW (22) | + | + | + | - |
| S-H4 | KRYLGTRCLSLMWAKGDLRQIKMGKRGDKKE (31) | - | - | - | - |
| S-H5 | <u>TMSF</u> SKFLFLLHFIHLIFIH (20) | + | + | +/- | + |
| H4 | <u>MTMST</u> WPILAALGFRTVFIFFGEEVM (26) | + | + | + | + |
| R13 | <u>MTMSILQ</u> ILYRVIFILNVIVVHIHYV (26) | + | + | + | + |

Table 4.6: Effect of putative AMPs on growth of *E. coli* when used synergistically with cell-permeabilizing reagents. Putative AMP peptides used at ~50 µM; growth of *E. coli* DH10B (+ or -) was determined after incubation in NB medium for 18 h at 37°C. Experiment repeated at least in duplicate, identical results achieved. Control, no peptide; Neg., no cell-permeabilizer.

S-H4, as seen previously in Table 4.5, prevented *E. coli* growth under all conditions tested. While H4 and R13 showed no toxicity when used in combination with the cell-permeabilizers, S-H5 showed some potential when used in conjunction with PMBN. However, after an additional incubation past 18 h, turbidity was observed. This was corroborated by agar plating of S-H5 culture aliquots after 21 h; colonies resulted (data not shown). Curiously, the negative control S-H1 consistently showed activity when used with sub-lethal levels of K2C18. However, after an extended lag phase, growth was observed after an additional incubation as per S-H5 above. S-H4, on the other hand, retained its lethal effect under the same conditions.

4.3 Discussion

Earlier work revealed the potential of the inducible, autocleaving CPD domain as a fusion tag for the production of AMPs (see Chapter 3). However, it was found that the uncleaved K2C18-CPD fusion still possessed antimicrobial activity when added exogenously to *E. coli*. Although activity of K2C18-CPD and K2C18 at various concentrations was not compared, it is surprising that the fusion protein exhibits antimicrobial properties at all. Perhaps the entire fusion is able to transit the outer membrane via K2C18-promoted self-mediated uptake (see Section 3.1.2.2) in order to reach its inner membrane site of action (Gallo et al., 1997; Park et al., 2003). Or it may be that the CPD tag sterically hinders this translocation, and rather an

electrostatic interaction of the K2C18 domain with the outer membrane is sufficient to lead to cell death (Sawyer et al., 1988). Although the AMP domain forms only ~8% of the fusion protein, the result in the work presented here is not without precedent. Tethering of magainin and bacteriocin derivatives to solid phase supports did not ameliorate their antimicrobial activity (Haynie et al., 1995; Hilpert et al., 2009). Furthermore, the fusion of a cellulose-binding domain (11.7 kDa) to the N-terminus of enterocin A (47 aa, a bacteriocin) did not significantly reduce antimicrobial activity (Klocke et al., 2005). More research is required to elucidate how an AMP may still function when part of a larger complex (Hilpert et al., 2009), especially with regard to the proposed mechanisms outlined in *Figure 3.2*.

It thus appears that the lack of toxicity seen when expressing the K2C18-CPD fusion endogenously (*Chapter 3*) is primarily because of a lack of access to the periplasm (discussed further below), rather than any protective effect endowed by the C-terminal CPD domain. Confirmation of this was observed when the pIII secretion signal was inserted upstream of K2C18-CPD; when induced *in vivo*, a similar level of growth inhibition to pBAD/gIII-K2C38* (*Figure 4.4*) was observed (see *Appendix 6*). The fact that an AMP-CPD fusion may still exhibit activity is not without some value, especially when directed to the periplasm. A gDNA-CPD fusion library could still be very useful for screening, with any observed AMP hits being amenable to scaled-up recombinant production (minus secretion signal), as their viable expression and activity as a CPD fusion would presumably demonstrate. Additionally, the presence of a stable fusion partner could protect the AMP from host protease attack (Taguchi et al., 1994; Walker et al., 2001). However, with a randomly-sized DNA insert, there is no way to control for the correct read-through to the C-terminal CPD tag; only one in three inserts would allow for CPD to be translated in its correct reading frame. The presence of in-frame stop codons would also be unavoidable. For these reasons, further development appears limited to the optimisation of previously identified AMPs.

Using the sensitive production host strategy and replica plating methodology, 40 putative AMP hits were identified. Of the 24 that can be matched to sequences in the NCBI database, only 2 are found to be putatively in-frame with a natural protein (1 of which was hypothetical [pAMP-R17]). Although the sample number is small,

this indicates that the screen does not enrich for naturally-occurring reading frames with regard to AMP function. In contrast, Watt (2009) found that 15 out of 41 hits in a yeast two-hybrid screen for binders to the Mal intracellular adaptor protein were in their natural reading frame; however, the starting genomic source was 80% protein-coding. In contrast, the human gDNA library used in this work is only approximately 1.2% protein-coding sequence (International Human Genome Sequencing Consortium, 2004); the rotifer gDNA library protein-coding percentage is unknown, but expected to be higher (C. Boschetti, personal communication).

The putative AMP hits pAMP/S-R21, pAMP-H2, pAMP-H6 and pAMP-R18 code for very short peptides (<12 aa). These may be false-positives, in that the encoded peptide is not responsible for the observed inhibition of growth (especially for the 3 aa peptide encoded by pAMP-R18). It was hypothesised that the encoded mRNA transcript, which includes the entirety of the insert sequence (100s of bp), may instead be the effector entity. In addition, mutation studies revealed that the 22 aa encoded peptide of pAMP/S-H1 was not responsible for growth inhibition *in vivo*. No putative endogenous *E. coli* transcript binding partners were found for any of these five sequences, indicating against an antisense RNA effect (Gottesman, 2004; Morita et al., 2006). However, secondary structure for all but the pAMP/S-H1 RNA transcript was predicted. It may be that a putative secondary structure interferes with a critical cell function, or acts as a riboregulator using different parts of its primary sequence (Isaacs et al., 2006). In either case, the primary sequence of the transcript may not point towards an obvious binding partner or function, and further investigation is needed to determine how such transcripts affect *E. coli*.

It is not entirely clear why the expression *in vivo* of an AMP does not lead to a bacteriostatic effect past approximately 10 h of induction. The persister phenomenon is one possible explanation (Lewis, 2007). The formation of these dormant cells, which possess a global decrease in metabolism and lack of cell division, can be encouraged in a homogenous culture through the overproduction of proteins that are toxic to a cell and inhibit growth, such as an AMP. Supporting this, the persister effect has been observed by others when *E. coli* is induced to express the toxic hok peptide (K. Gerdes, unpublished observation). It may be that intracellular expression of a toxic AMP causes an oscillation of protein synthesis

activity. As AMP production reaches toxic levels, global cellular processes and energy levels are affected and transcriptional/translational activity is reduced. As AMP levels fall within the cell, normal function resumes and the AMP is produced once more. The culture, in effect, becomes bacteriostatic. After ~10 h post-induction, however, it appears that this stasis is broken: either production of the AMP is permanently muted, or the host cell becomes tolerant to its presence. The presence of an AMP during this “recovery” phase may be ascertained via Western blotting; however, this was not performed in the work presented here. Raventós et al. (2005) also observed a similar recovery phenomenon during AMP production in *E. coli*. Such an effect appears to last through seeding an induced culture into fresh medium and re-inducing, but cells appear slow to enter exponential growth after a further passage. The expression plasmid is not altered, as isolation and re-transformation into fresh bacteria gave the same result. In addition, the recovery of colonies on solid medium and subsequent use as a fresh inoculum restored the inhibitory growth curve profile, ruling out genomic adaptation. This epigenetic inheritance of tolerance to endogenous AMP production is similar to that of persister cell formation during treatment with ampicillin (Lewis, 2007). Both are a temporary state that is restored to sensitivity once the antimicrobial threat has been removed (on solid media at least in the work presented here). While a comprehensive bactericidal effect upon AMP induction *in vivo* was not observed, the initial bacteriostatic effect correlated with the exogenous activity of some peptides, i.e. K2C18 and the novel S-H4 (discussed below).

The filamentation of *E. coli* seen when expressing periplasmically-targeted K2C18 and K2C38 may be a typical stress response. This elongation is commonly observed in a variety of bacteria suffering DNA damage, the presence of an antibiotic, or nutritional limitation (Justice et al., 2008). For unknown reasons, this may promote survival of a sub-population of cells. In addition, LL-37 (closely related to K2C38; see *Table 3.1*) has recently been shown to target septum formation in the periplasm of *E. coli* (Sochacki et al., 2011). A lack of cell division may lead to tolerance, much as in the persister phenomenon outlined above.

S-H4, a pAMP/S hit, was found to exhibit antimicrobial activity when added exogenously to *E. coli*. Out of the ~3,500 discrete pAMP/S colonies originally

screened, this represents an overall hit rate of approximately 0.03%. However, S-H4 was the only chemically-synthesised peptide expected to exhibit exogenous activity after frame-shift mutation and secretion signal removal investigations were performed (*Section 4.2.5*). The other hits, consisting of S-H2, S-H3 and S-R17, were not chemically-synthesised, and should be considered for further investigation. Although larger screens are required, proof-of-principle of this *in vivo* screening approach for AMPs has been shown. In previous work, Loit et al. (2008) claimed a hit rate of 0.002% when screening a random oligonucleotide library for AMPs, while Cheng et al. (2009) claimed a hit rate of 0.0004% when screening a wasp transcript library (i.e. complementary DNA). Both of these screens, however, had a large initial false-positive rate (67-75%) in comparison to the work presented here (~23%).

S-H4, when expressed in *E. coli*, had a similar effect on the growth curve of *E. coli* as K2C38 when targeted to the periplasm (*Figure 4.14*). However, when produced recombinantly using the CPD system, S-H4 failed to show activity (at an estimated final concentration of 6 μM) against *E. coli* when tested in LB medium. While the chemically-synthesised version had a slightly higher MIC of 7.8 μM against *E. coli* grown in NB medium, this doubled to 15.5 μM when used in LB. It may be that the high salt concentration (171 mM) in LB mitigates the potency of S-H4. This has been seen for other AMPs, such as the cathelicidin LL-37 (Bowdish et al., 2005a). Thus, the lack of inhibitory effect seen when using the cleaved AMP fractions of recombinant S-H4 was probably due to a lack of sufficient peptide concentration. Interestingly, four of the pAMP/S hits encode peptides with cationic tails, of which S-H4 is one (see *Table 4.7*). This makes them structurally similar to the membrane-lytic honeybee AMP melittin (26 aa).

| Name | Peptide (# aa) |
|----------|---|
| Melittin | GIGAVL K VLTTGLAPALISWI KRKR QQ (26) |
| S-H4 | K RYLGTRCLSLMWAKGDLR Q IKMG K RGD K KE (31) |
| S-R32 | LFSLLSLSLSLLFFFYER NKK (21) |
| S-R33 | CWIFLHSLLYSIQVSFL RR FLLFLIPVPD KSDQRKCAK (38) |
| S-R36 | IATFCMLSLIFVTFH K PLVRNIWGTI K SLLNFIFYMR KKK (40) |

Table 4.7: Peptide sequences of melittin and the novel putative AMP S-H4. Cationic residues in bold.

Rather than possessing a regular distribution of hydrophobic and cationic residues over its length (such as K2C18), melittin has a largely hydrophobic N-terminus with a C-terminus enriched in cationic residues, and yet is highly potent as an AMP (Dathe & Wieprecht, 1999). It is important to recognise that AMPs that are not amphipathic along the entire length of an α -helix may be identified during the screening process, as well as AMPs that may not form a defined secondary structure at all.

While it is easy to understand why an AMP that is toxic when produced endogenously may not be toxic when administered exogenously (i.e. the bacterial cell wall acts as a barrier to its target), it is not so clear with regards to an AMP that targets the inner membrane. A prime example is K2C18, which has had its inner membrane lytic ability confirmed both *in vivo* against *E. coli* and *in vitro* against vesicles that mimic the inner membrane (Gallo et al., 1997; Park et al., 2003). This negatively charged membrane is symmetrical in its leaflet composition (Ruiz et al., 2006), so the origin of cationic AMP attack, either periplasmic or cytoplasmic, should not matter. Unless K2C38 (from which K2C18 is derived) is targeted to the periplasm with the pIII secretion signal, however, *in vivo* toxicity does not arise. The pI of K2C18 is predicted to be ~10.8, indicating that it will still be cationic in both the *E. coli* cytoplasm (pH 7.2 to 7.8) and periplasm (reflecting the LB medium, pH 5 to 6.7) (Wilks & Slonczewski, 2007). It may be that the SecB chaperone, which can associate with nascent protein that is tagged with the pIII secretion signal (Natale et al., 2008), protects the AMP from cytoplasmic protease-mediated degradation (Maurizi, 1992). Furthermore, an increase in peptide length may lend such stability;

this could explain the lesser effect on growth of pIII-K2C18* versus pIII-K2C18-Tag in *Figures 4.2* and *4.4*. Subsequent secretion to the periplasm, which contains lower protease activity (Choi & Lee, 2004), may also give enough time for inhibition of growth to occur. Tentatively supporting this, less K2C38 is observed in a Western blot when the pIII secretion signal is removed (*Figure 4.5*).

A number of putative AMP hits were identified that act cytoplasmically, both from the secreted and non-secreted screens. As mentioned above, such hits are not expected to show activity when added to *E. coli* exogenously, as it is unlikely they possess cell-penetrating activity. Indeed, this was observed for those putative AMPs purified and tested in this work, as well as previously for the toxic hok peptide (Pecota et al., 2003). The simultaneous addition of cell-permeabilizing reagents such as NP-40, PMBN or sub-lethal K2C18 did not lead to a toxic phenotype. In any case, synergy between an AMP and a cell-permeabilizing reagent is unlikely to be useful in practical applications due to unwanted side effects (Good et al., 2000). It would be better to focus on AMPs that exhibit intrinsic (and specific) lytic or cell penetrating activity.

Future work could utilise different secretion signals to favour the isolation of periplasmically-active hits. For instance, a number of integral membrane proteins such as DsbA, TorT and TolB utilise the Sec pathway via the signal recognition particle (SRP) (Thie et al., 2008). SRP recognises and binds the secretion signal as it emerges from the ribosome, then pauses translation until the whole complex docks with the membrane-bound Sec-translocase complex (Natale et al., 2008). Translation resumes with the nascent protein being threaded into the periplasm. With this approach there would be little chance for an AMP to interact with potential cytoplasmic targets. In addition, knockouts of periplasmic proteases, such as DegP (Jones et al., 2002) and OmpT (Stumpe et al., 1998), may further allow accumulation of AMP in the periplasm. However, it is not strictly necessary, as a novel AMP has been successfully identified in the work presented here. Protease knockouts may lead instead to the identification of less potent entities, i.e. those requiring a greater concentration for efficacy.

In addition to the results presented in *Section 4.2*, the use of the yeasts *Saccharomyces cerevisiae* (West et al., 1984) and *Pichia pastoris* (Lee et al., 2005b) to produce and secrete K2C18 in a *trans*-based assay was also explored (data not shown). The aim was to use the well-described *S. cerevisiae* α mating factor secretion signal (Daly & Hearn, 2005) to drive secretion of a putative AMP library from yeast colonies plated on solid medium. A subsequent top agar overlay containing AMP-sensitive bacteria (e.g. *E. coli*) would reveal yeast clones producing novel AMPs, i.e. indicated by zones of growth inhibition. Not only could new AMPs be discovered via this *trans*-screen, but the eukaryotic yeast host (presumably tolerant to the AMP it produces) could then also be used for scaled-up peptide production. However, there are few previous examples of successful AMP secretion from yeast using episomal vectors, i.e. that do not integrate into the genome, as such integration is not amenable to high-throughput screening. Pediocin (12 aa, a bacteriocin) was produced in *S. cerevisiae* (Schoeman et al., 1999); and LL-37 and human α -defensin 5 (32 aa) in *P. pastoris* (Hong et al., 2007; Hsu et al., 2009). Unfortunately secreted K2C18 could not be successfully produced in either yeast: no growth inhibition of *E. coli* during top agar overlay of yeast colonies was seen.

In conclusion, a recombinant production strategy in a sensitive bacterial host was used to screen genomic DNA libraries for peptides with AMP function. A number of hits were identified, and one (S-H4) proved successful when added exogenously to the host *E. coli* strain. The effect of endogenous AMP production on cell growth was assessed, and found to induce a persister-like phenomenon that lead to the recovery of cell viability. Further modifications to the *in vivo* screen for novel AMPs are required to minimise false-positives, i.e. effects mediated by transcripts, rather than peptides. The use of cell-permeabilizing reagents was not found to restore the function of AMPs that had putative intracellular targets. Periplasmically-active AMP hits should be focused on in the future, as these represent the best candidates for clinical development.

CHAPTER 5 – SCREENING FOR OTHER ACTIVITIES:

ANTIAGGREGATION

5.1 Introduction

5.1.1 Brief history

Proteins, the molecular workhorses of biology, predominantly require ordered folding to ensure correct function. While organisms have evolved a number of systems to aid this process, or remove misfolded products, i.e. chaperones and proteases (Plempner & Wolf, 1999; Schlieker et al., 2002), errors still occur and can lead to protein aggregation. This agglomeration, usually due to the exposure of hydrophobic side-chain regions in the hydrophilic cellular environment, may prove toxic and ultimately fatal to the cell or organism concerned (Dobson, 1999). In humans, a number of diseases are associated with protein misfolding and aggregation. For example, in Huntington's disease the expansion of a codon repeat region in exon I of the huntingtin gene leads to an extended "sticky" patch of polyglutamine being coded for, which ultimately drives aggregation inside neuronal cells (Huntington's Disease Collaborative Research Group, 1993). Similarly, an increased polyalanine run length in the transcription factor Hox-D13 and subsequent aggregation leads to aggregation, culminating in the addition or fusion of digits (synpolydactyly syndrome) (Amiel et al., 2004). Other diseases in which aggregation is implicated include Parkinson's disease, where the presynaptic-associated protein α -synuclein can undergo aggregation, initiating Lewy body formation (aggregates) within neurons (Galvin et al., 1999); Creutzfeldt-Jakob disease, which is caused by the naturally produced prion protein being induced to change conformation by the disease prion variant and subsequent aggregation into amyloid plaques (Prusiner, 1998); and cystic fibrosis, in which a deletion of a single phenylalanine residue ($\Delta F508$) from the cystic fibrosis transmembrane conductance regulator protein leads to improper folding, aggregation and a loss-of-function (Thomas et al., 1992).

Another prominent aggregation-related condition is Alzheimer's disease, a debilitating neuronal disorder that causes dementia and ultimately death (Selkoe, 2002). The primary protein implicated in this disease is a peptide known as amyloid

β ($A\beta$), which may aggregate and lead to extracellular amyloid plaque formation in the brain (Walsh & Selkoe, 2007). While $A\beta$ is seen as the predominant effector of Alzheimer's disease, aggregation of the microtubule-associated protein tau is also associated with the condition (Brunden et al., 2009; Ittner & Götz, 2011).

In 2009, 35.6 million cases of Alzheimer's disease were reported worldwide, with this number expected to double by 2050 (Ittner & Götz, 2011). Current therapies, while able to alleviate the effects of the disease for a time, are unable to halt its progression (Gravitz, 2011). Therapies that are able to target the cause of Alzheimer's disease are thus desired.

5.1.2 Amyloid beta peptide ($A\beta$)

5.1.2.1 $A\beta$ 40 and $A\beta$ 42

$A\beta$ is produced in two main variations of length, the predominant form being 40 amino acids long ($A\beta$ 40), the other having an additional 2 residues at the C-terminus ($A\beta$ 42) (Walsh & Selkoe, 2007; DaSilva et al., 2010). Shorter lengths, such as $A\beta$ 39, are also possible (Jarrett et al., 1993). These small variations in length, heavily implicated in the aetiology of the disease (see *Section 5.1.2.2*), are formed by alternate processing of the Alzheimer's precursor protein (APP; 87 kDa), a trans-membrane protein found in neurons and a number of other cell types (Walsh & Selkoe, 2007). Although the precise biological function of APP remains unknown, stepwise cleavage by membrane-bound aspartyl proteases processes the protein into smaller subunits, one of which is $A\beta$ (De Strooper et al., 2010). At the first step of processing, β -secretase cleaves after residue 671 at the cell surface, releasing a substantial proportion of the N-terminus of APP into the extracellular milieu and leaving behind a C-terminal stub (De Strooper et al., 2010). This stub, in turn, is further cleaved at additional sites by the γ -secretase complex within the membrane bilayer, resulting in the release of $A\beta$ species of varying length (Carter et al., 2010). Point mutations around these cleavage regions are associated with familial Alzheimer's, but the vast majority of cases are sporadic (Walsh & Selkoe, 2007). Indeed, production of $A\beta$ is not abnormal in itself, as the peptides are found naturally throughout a human's life at nanomolar concentrations in brain tissue or cerebrospinal fluid (Selkoe, 2002).

Exactly how A β peptides cause Alzheimer's disease is still a matter of contention (see Section 5.1.2.2), but their propensity to self-recognise and aggregate is strongly implicated. Indeed, these peptides derive their name from their ability to generate amyloid fibrils, quaternary structures formed by protofilaments rich in β -sheet stacked in a parallel manner along an axis (DaSilva et al., 2010). These protofilament structures of A β may be further stabilised by intermolecular hydrogen bonding (Lührs et al., 2005). In Alzheimer's disease, mature fibrils aggregate further to form amyloid plaques, the classic histological hallmark of the condition (Treusch et al., 2009). Such insoluble aggregates are essentially indestructible under physiological conditions due to a large number of hydrogen bonds shared between adjacent β -sheet regions (Dobson, 1999).

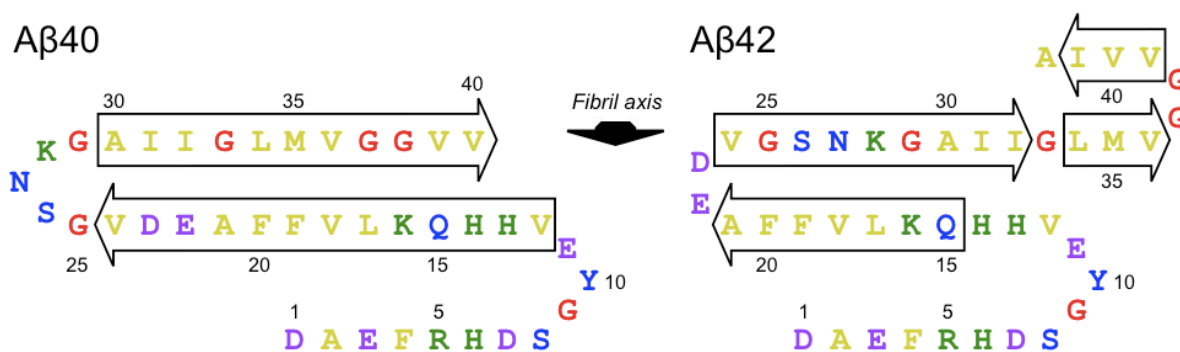


Figure 5.1: A β 40 and A β 42 peptides with proposed secondary structures. Potential β -folds, formed by the central and C-terminal hydrophobic regions, are indicated by open arrows. When stacked, the axis of the protofilament protrudes from the page. The C-terminal dipeptide Ile-Ala, which A β 40 lacks, is thought to adopt a different, or at least highly stabilised, β -fold confirmation. Amino acids are colour coded: cationic (green), hydrophobic (yellow), anionic (purple) and uncharged hydrophilic (blue). Figure adapted from DaSilva et al. (2010).

Because of their propensity to aggregate both *in vivo* and *in vitro*, and consequent insolubility, fine-scale structures of A β have not been possible to obtain (Barrow & Zagorski, 1991). However, a mixture of circular dichroism spectroscopy and solid-state nuclear magnetic resonance spectroscopy, combined with mutagenesis studies, has revealed some details (Hilbich et al., 1991; Lührs et al., 2005; DaSilva et al., 2010). Two main β -fold regions are formed in A β by the hydrophobic residues encompassing positions 17-21 and 30-40 (see Figure 5.1) (DaSilva et al., 2010). Compared to A β 40, A β 42 has a greater propensity to aggregate, at *in vitro* rates of hours versus days respectively (Jarrett et al., 1993). This is due to the additional

hydrophobic dipeptide Ile-Ala at the C-terminus of A β 42, which may lead to a greater stabilisation of the dominant β -sheet region (Urbanc et al., 2004), or perhaps a second β -sheet confirmation being favoured (shown in *Figure 5.1*) (DaSilva et al., 2010). Finally, while the charged N-terminus of A β is not directly implicated in β -sheet formation, mutations to this region can prevent mature fibrils from forming, indicating that electrostatic interactions between the N-terminus and the β -fold regions may help stabilise mature fibril formation (Hilbich et al., 1991; DaSilva et al., 2010).

5.1.2.2 A β 42 in Alzheimer's disease

As mentioned in *Section 5.1.2.1*, aggregation of A β is heavily implicated in Alzheimer's disease. Early observations that the brains of deceased sufferers were rich in amyloid plaque led to the hypothesis that the plaque itself was the causative agent of neuronal disruption (Roth et al., 1966). While the overproduction of A β 40, or increased levels of the relatively low-abundance but aggressive aggregator A β 42, is the root cause of plaques (Hardy & Selkoe, 2002), it is not clear that the plaques themselves are the effectors (Treusch et al., 2009; Ittner & Götz, 2011). Most frequently cited is the weak correlation observed between amyloid plaque levels and the onset of dementia, both in humans and in APP transgenic mouse models (Walsh & Selkoe, 2007). In humans, plaques are almost always present in the brains of healthy octogenarians, and even detectable in ~20% of cognitively normal people at an earlier age (Schnabel, 2011). But in neuronal cell cultures, aggregation of exogenously-added A β into fibrils was associated with toxicity (Pike et al., 1993), and brain extracts from APP transgenic mice (that exhibited amyloid plaques) induced amyloidosis and neuron pathogenesis in naïve mice when injected intracerebrally (Eisele et al., 2010).

One proposed explanation for this is that low-order oligomeric A β species are instead responsible for toxicity, rather than mature amyloid fibrils (Walsh et al., 2002). As these two A β forms are not mutually exclusive, such an explanation is not incompatible with the findings above (Walsh & Selkoe, 2007). Furthermore, the presence of soluble A β oligomers has been shown to be more closely correlated with memory loss (Kayed et al., 2003; Treusch et al., 2009). The oligomerisation of

A β represents a second pathway of aggregation that is distinct from amyloidogenesis (Figure 5.2). It may be that these oligomers display a larger surface area of hydrophobic patches than A β fibrils, bestowing a greater opportunity to interact and interfere with other neuronal components (Treusch et al., 2009). To tie these two aggregation pathways together, the formation of insoluble amyloid may instead serve as a protective mechanism to rid cells of the more reactive A β oligomers (Dobson, 1999; Hardy & Selkoe, 2002). Neuronal toxicity may result, however, if the cellular capacity to sequester A β oligomers into amyloid is exceeded, or the proteasome is overloaded and can no longer degrade soluble A β (Treusch et al., 2009).

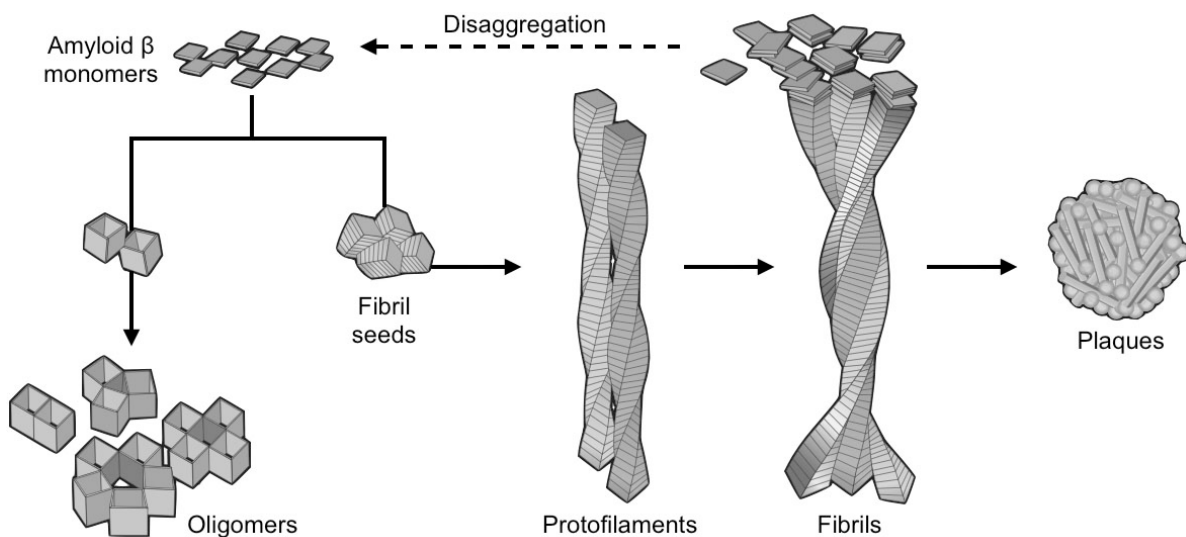


Figure 5.2: Two alternative pathways of amyloid β aggregation, which are not mutually exclusive. One pathway leads from A β monomers to aggregation and amyloidosis; the other to small A β oligomers, which are now thought to be the main toxic species in Alzheimer's disease. Figure adapted from Schnabel (2011).

Early protofilaments of A β behave as intermediates between plaques and oligomers in that they can form both mature fibrils or dissociate back to low molecular weight species of A β (Walsh & Selkoe, 2007). A dynamic equilibrium between these aggregation states exists, with nucleation or "seed" events from monomers being the rate-limiting step (Selkoe, 2002; DaSilva et al., 2010). Furthermore, it has been observed that A β monomers are able to slough off mature fibrils to start this seeding process anew (Jan et al., 2008), and that the less abundant but more aggressively aggregating A β 42 may seed the aggregation of A β 40 (Jarrett et al., 1993). Indeed,

given that the major A β species in the brain is A β 40, a small increase in the level of A β 42 may be enough to set off an aggregation cascade into either oligomers or amyloid (DaSilva et al., 2010). The exact toxic oligomeric form(s), however, is hard to determine given the difficulty of isolating and studying discrete oligomeric states (Walsh et al., 2002; Bitan et al., 2005a). In mouse models, behavioural changes are associated with the appearance of A β nonamers and dodecamers (Lesné et al., 2006). Other lower-order soluble oligomers, such as dimers (Shankar et al., 2008), have also been reported to be effectors (Carter et al., 2010). Given the above, isolating compounds that inhibit the initial stages of A β aggregation seems an attractive strategy, as it covers the gamut of possible aggregation states.

While the form of A β that causes Alzheimer's disease is still debated, so too are the exact mechanisms of neuronal cell death (Carter et al., 2010; Ittner & Götz, 2011). The synapse membrane and associated receptors appear to be a sensitive target (Selkoe, 2002; Ittner & Götz, 2011). A β association with synapse receptors, such as those for the neurotransmitters acetylcholine and glutamate, may lead to a loss of long-term potentiation and eventual neuronal death (Selkoe, 2002). There is even evidence that A β may function as a lytic antimicrobial peptide (Soscia et al., 2010), which could provide an immunoinflammatory explanation for neuronal membrane disruption (Carter et al., 2010).

5.1.3 Screens for A β 42 antiaggregants

Some treatments for the symptoms of Alzheimer's disease already exist, such as the small molecule drugs donepezil (trade name Aricept; Pfizer Inc.) and memantine (trade name Namenda; Forest Laboratories Inc.), which alter levels or responses to the neurotransmitters acetylcholine and glutamate respectively (Carter et al., 2010). Both mask the progression of dementia, but cannot ultimately cure Alzheimer's disease. For this to occur, new treatments are needed that target the underlying causes of Alzheimer's disease, i.e. protein aggregation. Possible approaches include decreasing A β production; stimulating the immune system to clear toxic A β forms from the brain; or inhibiting the aggregation process (all reviewed by Carter et al. [2010]). The work presented here focuses on generating a screen for the latter approach.

5.1.3.1 Small molecules and antibodies as antiaggregants

A number of small molecules have been shown to inhibit aggregation of A β *in vitro*. For example, inositol-based compounds, as well as polyphenols from green tea (epigallocatechin-3-gallate) and turmeric (curcumin), inhibit fibril formation (Sato et al., 2006b; DaSilva et al., 2010). Furthermore, *scyllo*-inositol (a stereoisomer) is currently undergoing Phase II clinical trials after showing promising results in mouse models of Alzheimer's disease (McLaurin et al., 2006). However, such small molecule antiaggregants are not the focus of the work presented here, and are reviewed elsewhere (Carter et al., 2010).

Passive immunotherapy using antibodies raised against A β is also being pursued as a treatment option. This approach has shown promise in APP transgenic mice: synaptic dysfunction was somewhat prevented, even though amyloid plaque levels were not decreased (Dodart et al., 2002). Furthermore, A β toxicity towards neuronal cell cultures was ameliorated (Kayed et al., 2003). As a consequence, A β antibodies are now in clinical trials. Bapineuzumab (Pfizer Inc.), a monoclonal antibody that targets the N-terminus of A β , was seen to lower plaque levels in Phase II trials, however little overall positive effect on cognition was observed (Gravitz, 2011). This may be due to a lack of specificity towards oligomeric A β . To try to counter this, another monoclonal antibody, solanezumab (Eli Lilly and Company), has been raised against soluble A β (i.e. oligomeric) and is targeted against its central region (Gravitz, 2011).

5.1.3.2 *In vitro* screening and animal models

A β 40 and A β 42 may be chemically-synthesised and their aggregation observed *in vitro*, commonly by the monitoring of binding of thioflavin T (ThT), a dye that intercalates with amyloidic cross- β structures in a concentration-dependent manner (LeVine, 1993). For *in vitro* screening, putative antiaggregant compounds may be added to a freshly prepared A β solution and any reduction of ThT binding (and hence fibril formation) observed. However, some antiaggregants, e.g. inositol-based compounds, may allow for alternative A β aggregation into a fibril structure that ThT does not recognise, thus resulting in a "false" positive (Sato et al., 2006b; DaSilva et al., 2010). Techniques less amenable to throughput but more directly indicative of

aggregation state include electron microscopy, atomic force microscopy, and size-exclusion chromatography (DaSilva et al., 2010). A large obstacle, however, when conducting *in vitro* studies of A β (especially A β 42) is irreproducibility: truly monomeric A β is hard to isolate, as pre-existing aggregates are very difficult to remove from working stocks (Bitan & Teplow, 2005; Kim et al., 2006). In addition, chemically-synthesised A β is expensive to manufacture due to its propensity to aggregate and subsequent loss of solubility, currently costing ~£140 per mg (Sigma-Aldrich Company Ltd.). Alternatives are thus desired for high-throughput antiaggregation assays.

Although animal models for the effects of A β aggregation serve an important purpose as pre-clinical validation tools, e.g. the APP transgenic mice mentioned in *Section 5.1.2.2*, these are not envisaged as high-throughput test beds. Animal models with a shorter generation time than mice do exist, but these are still limited (as yet) to profiling small sets of putative antiaggregants. For instance, various A β 42 mutants have been expressed in the brains of the fruit fly *Drosophila melanogaster*, with a positive correlation being observed between their *in vitro* aggregation rate and *in vivo* toxicity (Luheshi et al., 2007). Similar mutational studies have been carried out in the nematode *Caenorhabditis elegans*, with muscle-expressed A β 42 aggregating and eventually causing paralysis (Fay et al., 1998). Harnessing these model organisms in high-throughput screens is a worthwhile goal, but simpler systems such as *E. coli* (see *Section 5.1.3.4*) may have greater immediate utility.

5.1.3.3 Peptides as antiaggregants

As A β self-recognises to aggregate, a number of attempts have been made to rationally design peptides that interfere with this process (Soto & Estrada, 2005). Known as β -breakers, these peptides are often homologous to the β -fold regions of A β (see *Figure 5.1*) and thus competitively interfere with intermolecular interactions between adjacent β -sheets (DaSilva et al., 2010). This, in effect, may inhibit the polymerisation of additional A β peptides at the end of a growing protofilament (Lührs et al., 2005). Alternatively, peptides may disrupt the cross- β -sheet hydrogen

bonding between A β protofilaments that is necessary for mature fibril formation (Sato et al., 2006b).

Because of their homology to A β , however, a number of raw β -breaker sequences themselves aggregate, and thus are useless as therapeutics (DaSilva et al., 2010). To counter this, the addition of charged residues to such recognition elements can act as disrupters of further aggregation. When the pentapeptide KLVFF, homologous to amino acids 16 to 20 of A β , was C-terminally tagged with hexalysine, it slowed the aggregation kinetics of A β 39 *in vitro* and gave protection to neuronal cell line cultures in comparison to KLVFF alone (Pallitto et al., 1999). Furthermore, using a scrambled β -breaker sequence (VLFKF) also worked, indicating that an exact sequence match to A β was not required. An alternative to the use of a charged tag is to incorporate the disrupter element into the β -breaker recognition sequence itself. For example, incorporating a spatially constrained proline residue into the central hydrophobic recognition sequence of LVFFA (mutated to LPFFD) inhibited and dissolved amyloid formation *in vitro*, and prevented neuronal cell culture death on exposure to A β (Soto & Estrada, 2005).

No peptide-based antiaggregation inhibitors, however, are currently in clinical trials due to stability and bioavailability reasons (DaSilva et al., 2010), with particular emphasis on their requirement to negotiate the blood-brain barrier (see *Section 1.1.3.1*). In order to overcome these problems, further work has been carried out to alter the peptide backbone of lead peptides, e.g. *N*-methylation (Gordon et al., 2001; Kokkoni et al., 2006), or to switch amino acids to their D-enantiomers (Findeis et al., 1999). It is hoped that such peptides will eventually prove clinically useful for the treatment of Alzheimer's disease. In the meantime, it may be possible to find novel antiaggregant peptides through an *in vivo* screen if a suitable reporter system for the aggregation state of A β can be harnessed.

5.1.3.4 Recombinant screen approaches

In vitro solutions of chemically-synthesised A β are expensive and may already contain aggregates (see *Section 5.1.3.2*), so alternative approaches are desired. *In vivo* screening in systems such as *E. coli* is attractive: recombinant expression is

cheap and monomeric A β production must initially occur, thus allowing an opportunity for antiaggregants to interact at this key stage. Although using *E. coli* as a host by no means captures the native physiochemical environment of A β , it may prove useful as a first screen to identify putative antiaggregants. These may then be verified by additional *in vitro* and animal model experiments.

A common approach to *in vivo* analysis of protein folding in *E. coli* (and hence solubility) is to couple the aggregation state of the protein of interest to the function of a reporter. Fluorescent proteins have been employed as C-terminal fusions for this purpose: aggregation of upstream proteins may lead to chromophore misfolding and consequent abrogation of fluorescence (Waldo et al., 1999). Supporting this, Waldo et al. (1999) found a strong correlation between the fluorescent output of green fluorescent protein (GFP) and the solubility state of aggregation-prone fusions such as bullfrog H-subunit ferritin or gene V protein mutants.

Hecht and co-workers exploited this fluorescent folding reporter to study the aggregation of A β 42 when expressed in *E. coli* (Figure 5.3). They found a good correlation between the *in vitro* aggregation state of A β 42 and the *in vivo* fluorescent output of GFP, i.e. insolubility and a lack of fluorescence respectively (Wurth et al., 2002). Furthermore, mutations of A β 42 that reduced its propensity to aggregate led to the restoration of chromophore fluorescence, and this also correlated with improved solubility as assessed by SDS-PAGE. Further work used this *in vivo* approach to screen a library of ~1,000 triazine-based small molecule drugs, and led to the identification of two compounds that inhibited A β 42 aggregation *in vitro* via the ThT assay and electron microscopy (Kim et al., 2006).

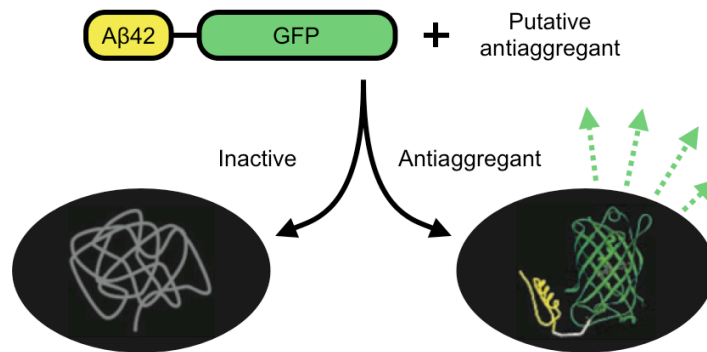


Figure 5.3: Fluorescence-based screen for antiaggregants using an A β 42-GFP fusion. If a compound exhibits no antiaggregant activity, the A β 42-GFP fusion protein misfolds due to A β 42 aggregation, and no fluorescence output is observed. If A β 42 aggregation is inhibited, GFP (indicated in green) is able to fold and consequently fluoresce. Figure adapted from Kim et al. (2006).

Moffet and co-workers used this A β 42-GFP system to screen A β 42-like peptides co-expressed *in vivo* for antiaggregant properties (Baine et al., 2009). As per the β -blocker approach (discussed in Section 5.1.3.3), these peptides were truncated mutants of the central or C-terminal hydrophobic regions of A β 42 (see Figure 5.1) and incorporated aspartic acid mutations to interfere with further cross- β structures. A peptide that could partially restore A β 42-GFP fluorescence was isolated (discussed below in Section 5.1.4.1).

Other similar *in vivo* screens for A β solubility have also been used. Thomas and co-workers fused the α -fragment of β -galactosidase to the C-terminus of A β 42, and found that aggregates were unable to complement the ω -fragment in *E. coli* and hence could not cleave X-gal to give a blue output (Wigley et al., 2001). However, a downside of this approach is that the screen requires a fine degree of optimisation due to the α -fragment remaining accessible for complementation in large aggregates; thus, the range of sensitivity is small (3-fold at best). In another approach, DeLisa and co-workers utilised the twin-arginine translocation system (Tat; a Type II secretion pathway [Brüser, 2007]) as a reporter in *E. coli*. As the Tat system only exports properly folded and soluble proteins, A β 42 (with an N-terminal Tat secretion signal) was fused to the N-terminus of β -lactamase (Bla), which confers resistance to ampicillin when localised to the periplasm. Aggregation of the A β 42-Bla fusion prevented periplasmic export, and thus conferred ampicillin sensitivity (Fisher et al., 2006). Such a system can be used as a positive screen for

antiaggregants: if A β 42-Bla is solubilised, growth in ampicillin-containing media can occur. Again, however, sensitivity was an issue, and consequently a fluorescent substrate of Bla had to be harnessed instead. This approach isolated four hits from the same combinatorial library of ~1,000 triazine scaffold variants used by Hecht and co-workers, one of which was the same (Lee et al., 2009).

The above examples indicate the promise of utilising *in vivo* screens to isolate putative antiaggregants in a timely and cost effective manner. The work presented here aims to utilise such recombinant approaches to identify novel antiaggregant peptides.

5.1.4 Recombinant screen considerations

5.1.4.1 Antiaggregation controls

For a non-aggregating positive control, the A β 42 mutant GM6 was selected. Isolated by Wurth et al. (2002) in their original GFP fusion screen (see *Section 5.1.3.4*), GM6 contains a phenylalanine to serine mutation at position 19 and a leucine to proline mutation at position 34, both of which serve to disrupt the core hydrophobic β -fold regions of A β 42 (see *Figure 5.1*). As a consequence, the GM6-GFP fusion does not aggregate, and fluoresces well *in vivo*.

In their A β 42 antiaggregant peptide screen, Baine et al. (2009) isolated Peptide 2 (Pep2), a 14 aa variant of the central hydrophobic region of A β 42, which was able to double the fluorescence output of A β 42-GFP when co-expressed. With a sequence of GKLDVVAEDAGSNK (mutations compared to A β underlined), this peptide likely complements A β 42 but prevents further aggregation of cross- β structures through the steric hindrance imparted by the aspartic acid residue (see *Section 5.1.2.3*). Not only was Pep2 observed to depress A β 42 aggregation *in vitro* through the ThT assay, it was also seen to disaggregate pre-existing fibrils (Baine et al., 2009). This peptide sequence was therefore picked as a positive peptide control for a *de novo* antiaggregant screen.

As a novelty, a 16 kDa protein from the anhydrobiotic nematode *Aphelenchus avenae*, related to plant group 3 late embryogenesis abundant (LEA) proteins

(Browne et al., 2002), was also selected for study in this chapter. LEA proteins are implicated in the desiccation-tolerance of a large number of plants (specifically their seeds), as well as in some animals, and act to preserve cell component integrity in the “dry” state (Tunnacliffe & Wise, 2007). Functionally, LEA proteins may behave less like traditional chaperones and more as “molecular shields”, forming physical barriers between partially unfolded proteins to prevent their aggregation during desiccation (Goyal et al., 2005a; Chakrabortee et al., 2012). The LEA identified in *A. avenae* (designated AavLEA1; 143 aa) is highly hydrophilic and unstructured in solution (Goyal et al., 2003). Given its ability to prevent whole-cell proteome aggregation *in vitro* (Chakrabortee et al., 2007), its properties might allow it to interfere with A β aggregation if co-expressed in *E. coli*. Lending further evidence towards this, the use of rapeseed group 3 LEA proteins as C-terminal fusion tags of recalcitrant recombinant proteins somewhat alleviates inclusion body formation during expression in *E. coli* (Singh et al., 2009).

5.1.4.2 Use of flow cytometry

Previous approaches to the A β 42-GFP screen in *E. coli* have involved analysing bacterial fluorescence in liquid medium using 96-well plates (Kim et al., 2006), or as colonies on solid medium (Baine et al., 2009). Although antiaggregant hits were identified by both Kim et al. (2006) and Baine et al. (2009), the dynamic range of signal output was low. At best, a two-fold increase in fluorescence output over background A β 42-GFP was achieved. The use of flow cytometry may improve on this. Rather than relying on an average fluorometric reading from an entire cell population, this method assesses the output of cells on an individual basis, allowing for the possible concentration effects of *E. coli* cell size and replication state to be taken into account (Shapiro, 2003). This may give a greater sensitivity to the effect an antiaggregant may have on A β 42-GFP fluorescence output. Importantly, such a single-cell approach is very much compatible with *in vivo* DNA library screening (discussed in *Section 1.2.3*), in which each cell encodes a different member of a putative antiaggregant peptide library. Fluorescence-activated cell sorting (FACS) could therefore be employed to separate putative hit cells, i.e. those with increased fluorescent output from the negative background, thus enabling the screening of antiaggregant peptide libraries in a high-throughput manner (Shapiro, 2003).

Given that the dynamic range between the effect of antiaggregants such as Pep2 and background A β 42-GFP fluorescence is low (see above), putative antiaggregant peptide hits in a flow cytometry screen may still be obscured by background noise. Measuring the ratio of cellular GFP output to an internal reference may aid in reducing this; such a ratiometric approach is commonly used in RNA microarray experiments, i.e. between a reference and an experimental sample (Skena et al., 1995). To this end, constitutive expression of the red fluorescent protein mCherry (Shaner et al., 2004) was employed as an internal fluorescent reference. If A β 42-GFP fluorescence increases relative to the mCherry standard, flow cytometry will reveal a shift along the X-axis of a plot of these two parameters (*Figure 5.4*).

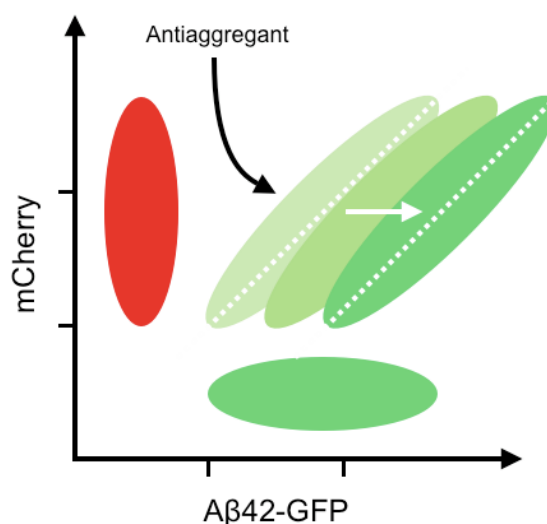


Figure 5.4: Ratiometric approach to identifying antiaggregants using flow cytometry. If a compound exhibits antiaggregant activity, the ratio of A β 42-GFP:mCherry fluorescence will increase; mCherry acts as an internal standard. mCherry (red) and A β 42-GFP (green) alone also plotted.

While absolute A β 42-GFP fluorescence per cell may vary under the same incubation conditions, the relative fluorescence to mCherry should remain constant. In addition, as synthesis rates of GFP have been reported to be relatively sensitive to plasmid copy number (varying 1.5 to 3-fold) (Kelly et al., 2009), mCherry may be incorporated into the same plasmid as A β 42-EGFP to ensure an equal gene copy number between the two components. The only major variable remaining is the expressed DNA library member; therefore an increase in relative A β 42-GFP fluorescence should be observed if a putative antiaggregant peptide is encoded.

5.1.4.3 Random oligonucleotide libraries

While gDNA was used as an input to encode a peptide library in *Chapter 4*, here the use of chemically-synthesised random oligonucleotides was explored. Although the encoded random peptides may be unlikely to adopt a secondary structure (discussed in *Section 1.2.2*), this is not required for short length β -breakers (see *Section 5.1.2.3*). Furthermore, given that a match in overall physiochemical properties (e.g. hydrophobicity) is sufficient as opposed to an exact match of sequence (see *Section 5.1.3.3*), screening of randomly encoded peptides is not inappropriate. Aside from traditional β -breakers, other novel antiaggregant peptides may be identified.

As proof-of-principle, oligonucleotides encoding 12 or 24 random amino acids (excluding the initiating methionine) were used. Random codons (NNN) were utilised, as although includes the possibility for internal stop codons to be incorporated, any restrictions (e.g. NNG/T to eliminate two of the three stop codons) will alter the codon bias for certain amino acids in *E. coli* (Walker et al., 2001). The potential sequence diversity of the aforementioned peptide libraries is huge: for the 12 aa library, $\sim 4.1 \times 10^{15}$ combinations are possible; for the 24 aa library, $\sim 1.7 \times 10^{31}$. In reality, given the similar properties of some residues (e.g. the hydrophobicity of leucine and isoleucine), this diversity is lessened. Still, only a small fraction of this potential sequence space is likely to be explored during screening, and any hits may rather serve as templates for further optimisation.

5.1.5 **Summary**

While *Chapters 3* and *4* dealt with antimicrobial peptides and screening for novel sequences that exhibited microbial toxicity, this chapter serves as further proof-of-principle that other, less readily-observable activities may also be screened for using an *in vivo* system. As an example, aggregation of the peptide A β , which is heavily implicated in the aetiology of Alzheimer's disease, was studied. Peptides possessing antiaggregant properties may have the potential to attenuate clinical progression of the disease, or perhaps to prevent it if used prophylactically. A handful of antiaggregant peptides have already been reported, primarily being homologues to the key hydrophobic β -fold regions of A β that bind and competitively

interrupt aggregation of the same. Additional β -breakers, or novel antiaggregant peptide sequences, could prove useful for further development. As *in vitro* work with chemically-synthesised A β is expensive, using an *in vivo* screen of a peptide library in *E. coli* may allow for hit identification in a cost-effective manner.

A β 42-GFP fusions, when expressed in *E. coli*, aggregate and prevent the correct folding and fluorescence of the GFP domain. Restoration of fluorescence has previously been used as a basis for screening small molecules for antiaggregant properties. This approach was taken and applied to the screening of co-expressed random peptide libraries for putative antiaggregants. Specifically, flow cytometry was employed to ascertain the fluorescence state of the A β 42-GFP fusion in single *E. coli* cells, with this being measured ratiometrically against a constitutively expressed fluorescent mCherry control. The antiaggregant peptides Pep2 and AavLEA1 were assessed for activity in this manner, and FACS initially trialled as a means of facilitating high-throughput screening. Screening libraries on solid medium was also conducted, with several putative antiaggregant hits subsequently assessed using flow cytometry. Further work is ongoing to verify these hits using traditional *in vitro* methodology.

5.2 Results

5.2.1 Construction of vectors for antiaggregant screen

5.2.1.1 Construction of pAG2-A β 42

The pBADm plasmid was chosen as a chassis for the construction of the screening vector pAG2-A β 42 (for details see *Section 2.7.1*). Three main operons formed the basis of the plasmid (*Figure 5.5 A*; full plasmid sequence in *Appendix 7*). Firstly, the coding sequence for the aggregator A β 42-EGFP (enhanced GFP; taken from Baine et al. [2009]) was placed under tight control of the arabinose-inducible *araBAD* promoter (Guzman et al., 1995). Secondly, a “Library” operon was fabricated (see *Appendix 8*), featuring a strong constitutive promoter and *SpeI/AfeI* sites flanking the start codon for DNA library insertion and subsequent expression. Constitutive expression from the Library site should allow for putative antiaggregant peptide to build up in a cell prior to A β 42-EGFP induction, thus maximising the chance that inhibitors of the earliest stages of A β 42-EGFP aggregation may be identified (Kim et

al., 2006; Carter et al., 2010). Finally, mCherry was put under the control of a medium-strength constitutive promoter so that it could be consistently used as an internal standard to ratiometrically compare between the effects of putative antiaggregants on A β 42-EGFP fluorescence output. As the origin of replication utilised by pAG2-A β 42 (pBR322 origin) results in a low copy number of plasmids per cell (15 to 75) (Velappan et al., 2007), integration of genes on a single plasmid is important to minimise spurious ratiometric results due to gene copy variation.

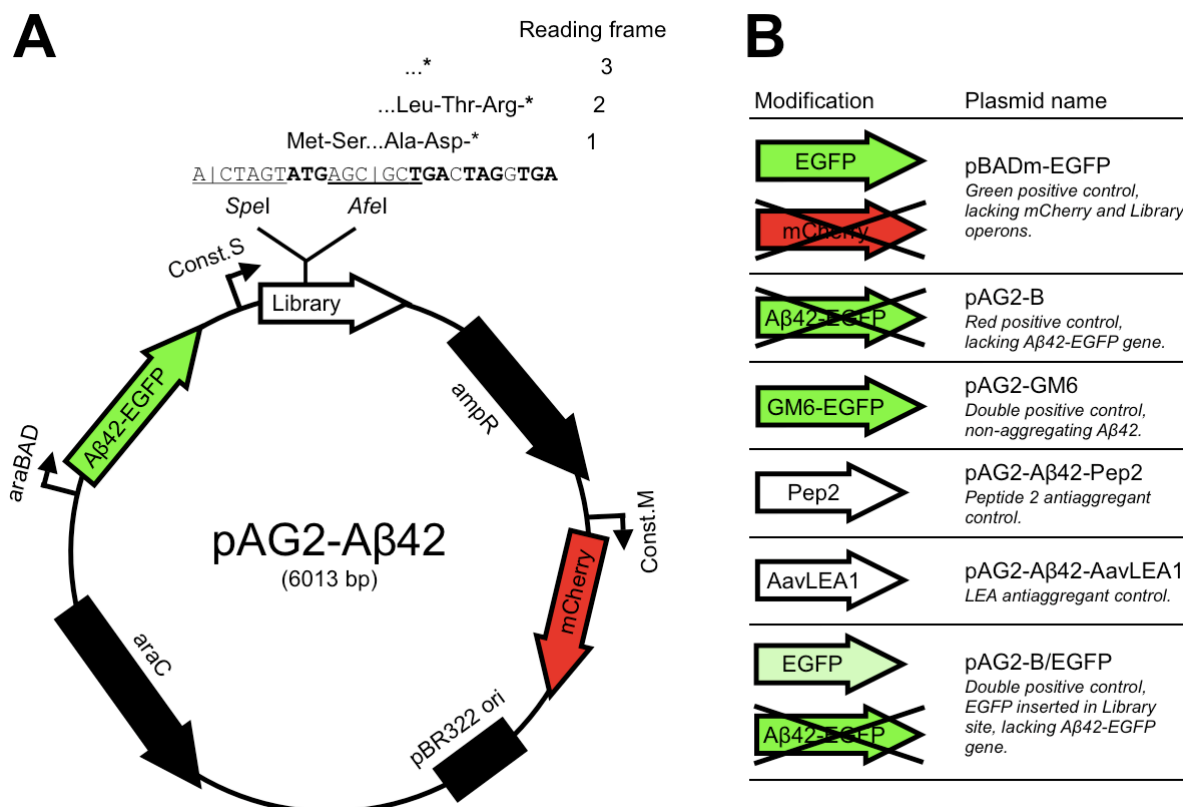


Figure 5.5: The pAG2-A β 42 vector used to screen for putative antiaggregant peptides, and various control modifications. Insert DNA may be ligated into the Library site of pAG2-A β 42 using either the SpeI site (start codon required) or AfeI site. **A:** pAG2-A β 42, with A β 42-EGFP under control of the arabinose-inducible araBAD promoter, and the Library and mCherry operons constitutively expressed. Start and stop codons for all three Library insert reading frames are shown in bold (possible encoded C-terminal residues also indicated). Const.S, strong constitutive promoter; Const.M, medium constitutive promoter. **B:** Modifications made to pAG2-A β 42 to create controls. Green indicates replacement at the araBAD site, white the Library site and red at mCherry. Crosses indicate removal of an operon.

5.2.1.2 Construction of pAG2-A β 42 controls

Modifications were made to the pAG2-A β 42 plasmid to generate a number of controls (Figure 5.5 B; see Section 2.7.2 for details). For a single, arabinose-

inducible green fluorescence control, EGFP was inserted into the *araBAD* site of pBADm (pBADm-EGFP). pAG2-B, a precursor plasmid generated during the construction of pAG2-A β 42, contained constitutively expressed mCherry and served as a single, red fluorescence control. The activity of the Library site could be assessed by the insertion of EGFP into pAG2-B (pAG2-B/EGFP).

In addition, the A β 42 domain of pAG2-A β 42 was modified to incorporate the GM6 mutations (F19S, L34P; Wurth et al. [2002]) to act as a non-aggregating, arabinose-inducible double positive control, i.e. both green and red fluorescence (pAG2-GM6). Lastly, Pep2 and AavLEA1 were inserted into the Library site of pAG2-A β 42 so that their effect on A β 42-EGFP fluorescence/aggregation during co-expression could be assessed (pAG2-A β 42-Pep2 and pAG2-A β 42-AavLEA1 respectively). Some of the amino acid sequences of the constructs mentioned above are shown in *Table 5.1*.

| Product name | Amino acid sequence (# aa) | MW (kDa) |
|--------------|--|----------|
| A β 42 | DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIA (42) | 4.4 |
| GM6 | DAEFRHDSGYEVHHQKLV S FAEDVGSNKGAIIG P MVGGVVIA (42) | 4.4 |
| Pep2 | QK L D V V A ED A GSNK (14) | 1.5 |
| AavLEA1 | <u>SSQQNQNRQGEQQEQGYMEA</u> AKEKVVNAWESTKETLSSTAQAAA E KTA E F RDSAG E T I RDLT G Q A Q E K G O E F K ERAGEKA E ET K ORAGE K MD E T K ORAGE MRENAGQKM E EY K Q Q G K GA E ELRDTAA E KL H QAGEK V KGRD (142) | 15.9 |

Table 5.1: The sequence of A β 42 and the controls GM6, Pep2 and AavLEA1. The two hydrophobic regions of A β 42 are underlined, with comparable mutations in GM6 and Pep2 indicated in bold. The 11-mer repeat regions of AavLEA1 are also underlined (Goyal et al., 2005b). MW, molecular mass.

While homology between A β 42 and GM6/Pep2 is obvious, any similarity to AavLEA1 is less so. Attempted alignment of the AavLEA1 amino acid sequence with A β 42 revealed a small region of homology between the N-terminus of A β 42 and the second 11-mer repeat region of AavLEA1 (*Figure 5.6*), but not with the β -sheet critical hydrophobic regions (see *Section 5.1.2.1*).

```

AB42          -----DAEFRHDSGYEVH- 13
AavLEA1      SSQQNQNRQGEQQEQGYMEAAKEKVVNAWESTKETLSSTAQAAA EKTAEFRDSAGETIRD 60
                                     ****..:*  ::

AB42          -----HQLVFFAEDVGS-----NKGAI I-----GLMVG--GVVIA----- 42
AavLEA1      LTGQAQEKGQEFKERAGEKAEETKQRAGEKMDETKQRAGEMRENAGQKMEEYKQQGKGA 120
                :*  * * .* . * : * * * * :

```

```

AB42          -----
AavLEA1      EELRD TAAEKLHQAGEKVKGRD 142

```

Figure 5.6: The N-terminus of A β 42 shares some homology with the second 11-mer repeat region of AavLEA1. Residue positions indicated. Star, indicates side-chain match; colon, indicates strong similarity between side-chains; period, indicates weak similarity between side-chains. ClustalW2 used to generate alignment (see Section 2.7.7).

5.2.2 Profiling of pAG2-A β 42 and controls

5.2.2.1 Expression of pAG2-A β 42 and controls

The plasmid pAG2-A β 42 and the controls pBADm-EGFP, pAG2-B, pAG2-B/EGFP, pAG2-A β 42-Pep2, pAG2-A β 42-AavLEA1 and pAG2-GM6 were transformed into *E. coli* DH10B cells. Expression from selected constructs was induced at 37°C for 6 h (see Section 2.3.1), and protein products analysed using SDS-PAGE (Figure 5.7).

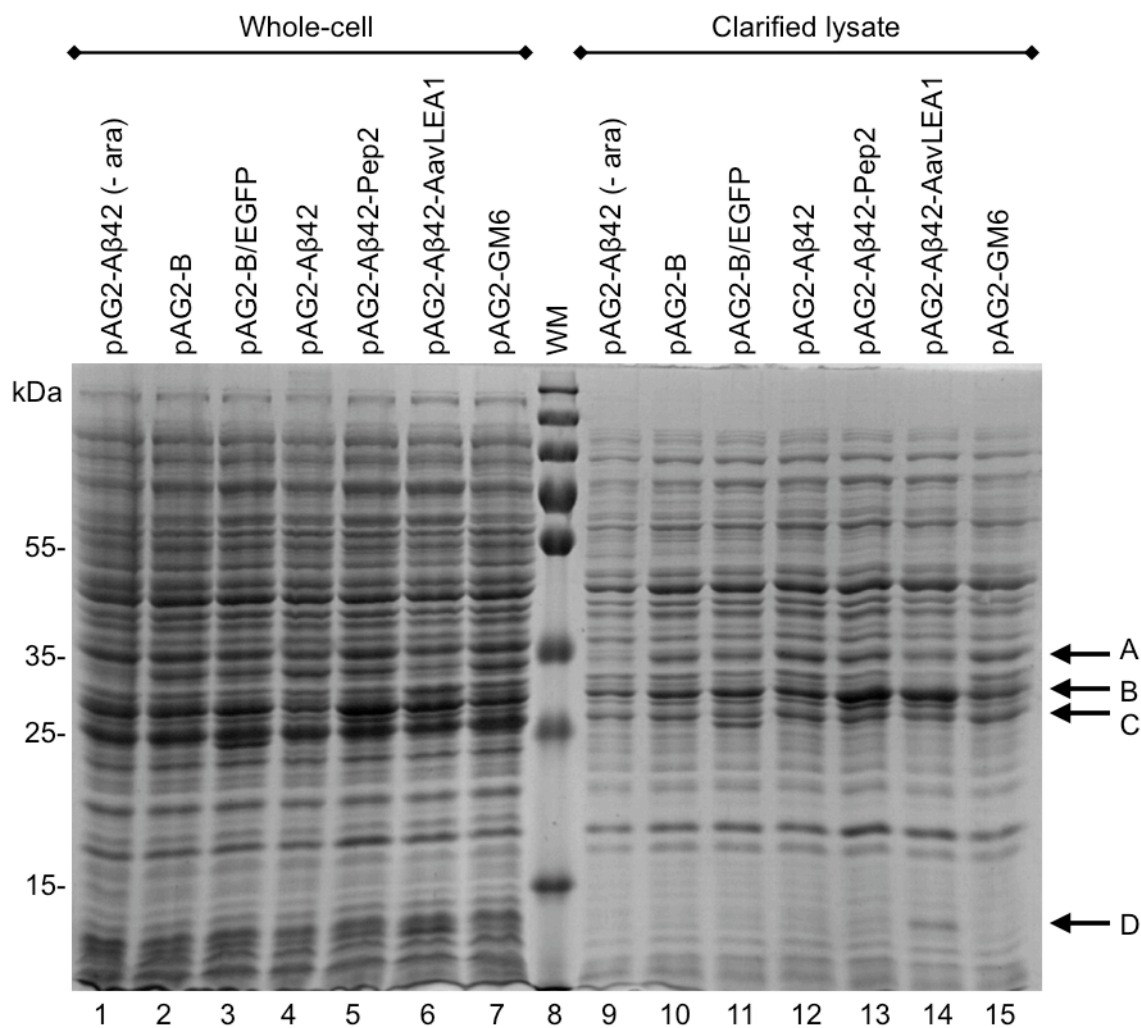


Figure 5.7: Expression from pAG2-A β 42 and selected controls. SDS-PAGE (glycine 12%) of 6 h induction in *E. coli* DH10B at 37°C, ~225 rpm. Expected band sizes: A β 42-EGFP and GM6-EGFP, 32.6 kDa; mCherry/EGFP, 27 kDa; AavLEA1, 15.9 kDa, Pep2, 1.5 kDa. Equal volumes of cell culture were loaded (normalised to the same density via OD₆₀₀). Whole-cell fractions represent both insoluble and soluble protein, clarified lysate soluble protein only. WM, molecular mass marker.

It was hoped that there would be an obvious intensity difference between the A β 42-EGFP band in the whole-cell fraction (a mix of soluble and insoluble protein; lane 4) and its equivalent in the clarified lysate fraction (soluble protein only; lane 12). If A β 42-EGFP aggregates, its band is only expected to be seen in lane 4. However, aside from AavLEA1 (lanes 6 & 14, arrow D), it is difficult to determine which, if any, of the bands in the 25 to 35 kDa region (arrows A to C) correspond to the expected sizes stated in the legend of *Figure 5.7*. Band C may be that of A β 42-EGFP, as it is less intense in the soluble fraction of induced pAG2-A β 42 (compare lanes 4 & 12); however, this is unlikely, as the same band is equally present in samples not containing A β 42-EGFP (e.g. lanes 10 & 11), and

furthermore is running at a lower weight than expected. In addition, there is no obvious intensity difference in the bands arrowed A to C between A β 42-EGFP (lane 12) and the non-aggregating GM6-EGFP control (lane 15); GM6-EGFP is expected to be more soluble.

Although interpretation of *Figure 5.7* is unclear, Wurth et al. (2002) successfully demonstrated by SDS-PAGE that their A β 42-GFP protein product was insoluble, while their non-aggregating mutant GM6-GFP appeared predominantly in soluble fractions. Furthermore, A β 42-GFP was not observed to fluoresce, while GM6-GFP did (confirmed in *Section 5.2.2.2* below). However, Wurth and co-workers' A β 42-GFP differed from the A β 42-EGFP used in the work presented here (sourced from Baine et al. [2009]). The linker region between the A β 42 sequence and the chromophore domain is 15 aa instead of 12 aa, and the GFP variant utilised is also different: Baine used EGFP, while Wurth used a GFP version containing mutations at five positions (Crameri et al., 1996). However, these differences are unlikely to explain any difference in solubility; further work utilising Western blotting with A β 42-specific antibodies is required to resolve this problem.

5.2.2.2 Fluorescence microscopy of pAG2-A β 42 and controls

Although the exact expression and solubility properties of pAG2-A β 42 and the control constructs could not be determined by SDS-PAGE, the next step was to assess their capability to produce fluorescence in *E. coli*. EGFP and mCherry fluorescence in induced cells was observed using fluorescence microscopy (see *Section 2.7.4*) after a 6 h induction (*Figure 5.8*).

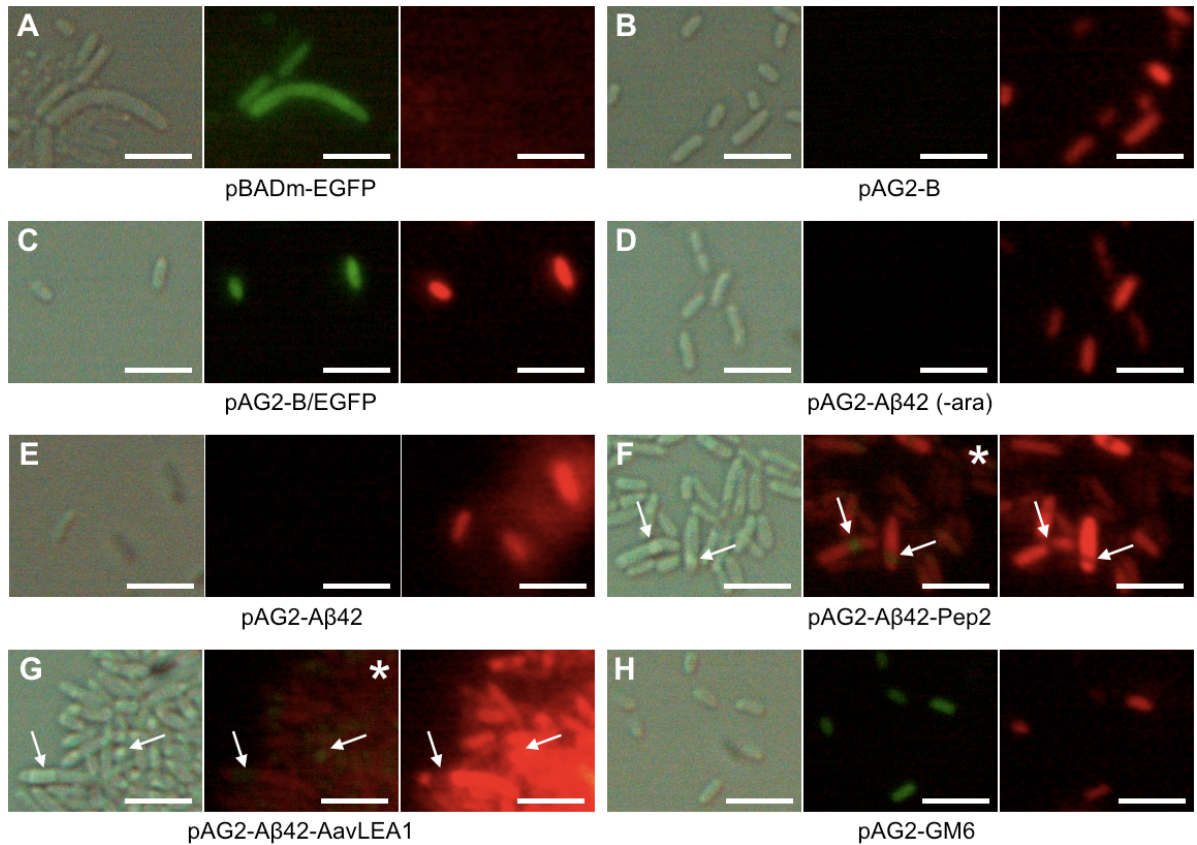


Figure 5.8: Fluorescence microscopy of *E. coli* expressing pAG2-A β 42 and controls. For each panel, from left to right: bright field, green fluorescence, red fluorescence. Photos taken after 6 h arabinose induction (0.05%) in LB medium at 37°C at ~225 rpm. Both uninduced (- ara) and induced pAG2-A β 42 samples shown (panels D & E). Green aggregates arrowed (panels F & G); EGFP photos overexposed to visualise (*), which leads to bleed-through of mCherry fluorescence. Scale bar, 5 μ m.

For the single green and single red controls (panels A & B), fluorescence was as expected as per *Figure 5.5*. However, green fluorescence from pBADm-EGFP cells was seen to bleach quickly; this was also the case for red fluorescence in all samples examined. The pAG2-A β 42 vector, with A β 42-EGFP both uninduced (panel D) and induced (panel E), exhibited no green fluorescence as desired. In contrast, the non-aggregating mutant GM6-EGFP exhibited diffuse green fluorescence when induced (panel H), indicating its solubility and confirming the work of Wurth et al. (2002). Importantly, the Library site of pAG2-A β 42 was shown to be transcriptionally active, as inserted EGFP resulted in green fluorescence (panel C). In addition, similar observations were made of all samples after 23 h of induction (data not shown; however, for a 15 h induction, see *Figure 5.18*), indicating that longer incubation times did not affect the fluorescent state of the cells.

Interestingly, several cells containing pAG2-A β 42-Pep2 and pAG2-A β 42-AavLEA1 (panels F & G) were seen under the microscope to possess single green fluorescent aggregates by eye; however, for photography, these images required overexposure and consequently led to bleed-through of mCherry fluorescence (for explanation see *Section 2.7.4*). While the majority of cells did not exhibit these aggregates, indicating against a homogenous effect, it may be that Pep2 and AavLEA1 are able to interfere with A β 42-EGFP aggregation to a limited extent *in vivo*. This may provide enough time for EGFP to fold before subsequent A β 42 aggregation occurs. Furthermore, red fluorescence from pAG2-A β 42-Pep2 and pAG2-A β 42-AavLEA1 cells (panels F & G) appeared to be brighter than that from other samples, but still bleached as quickly. Related to this, centrifugation of these cultures also resulted in cell pellets with a brighter red than those of their counterparts (data not shown).

Lastly, an additional observation of interest was that induced cells containing pAG2-A β 42-AavLEA1 clumped together (see *Section 5.2.2.3* below, and *Figures 5.27 & 5.31* for photos). This was also observed in pAG2-A β 42-Pep2 cells, but not to the same extent.

5.2.2.3 Growth rates of pAG2-A β 42 and controls

The effect of induction of pAG2-A β 42 and the various controls on the growth rate of *E. coli* in liquid LB was also ascertained. Cultures were induced as per *Section 2.3.1* and OD₆₀₀ monitored over time. *Figure 5.9* shows the resulting growth curves.

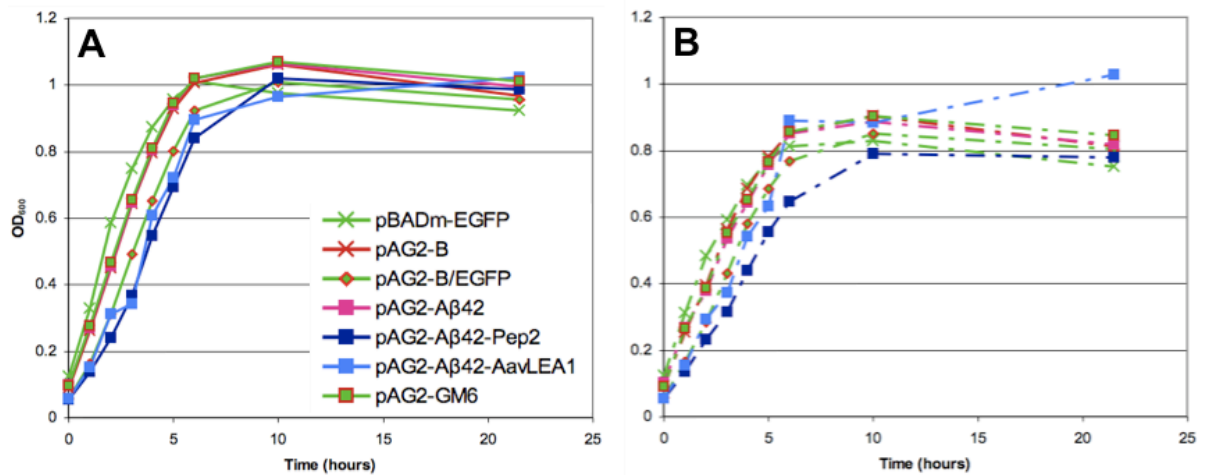


Figure 5.9: pAG2-Aβ42 constructs containing inserts in the Library site exhibit a slight lag in growth. A: Uninduced. B: Arabinose (0.05%) added to induce cultures in LB medium at 0 h (dashed lines), incubated at 37°C at ~225 rpm.

All constructs seemed to grow normally and without any inhibition of growth rate. However, constructs containing inserts in the Library site, i.e. pAG2-B/EGFP (green line, red diamond), pAG2-Aβ42-Pep2 (dark blue line) and pAG2-Aβ42-AavLEA1 (light blue line), showed a slight initial lag in entering exponential-phase growth. The effect was seen in both uninduced and induced cultures. This may be due to the increased metabolic load imposed on these cells due to constitutive expression of a Library site insert.

The OD₆₀₀ of the pAG2-Aβ42-AavLEA1 culture was hard to accurately ascertain due to cell clumping, which was more apparent in the induced sample as opposed to the uninduced sample (*Figure 5.10*). Despite this, cell viability appeared unaffected. While cells were observed to settle out initially, turbidity was restored after several hours continued incubation. The pAG2-Aβ42-Pep2 culture also exhibited slight sedimentation, but nowhere near to the same extent as pAG2-Aβ42-AavLEA1.

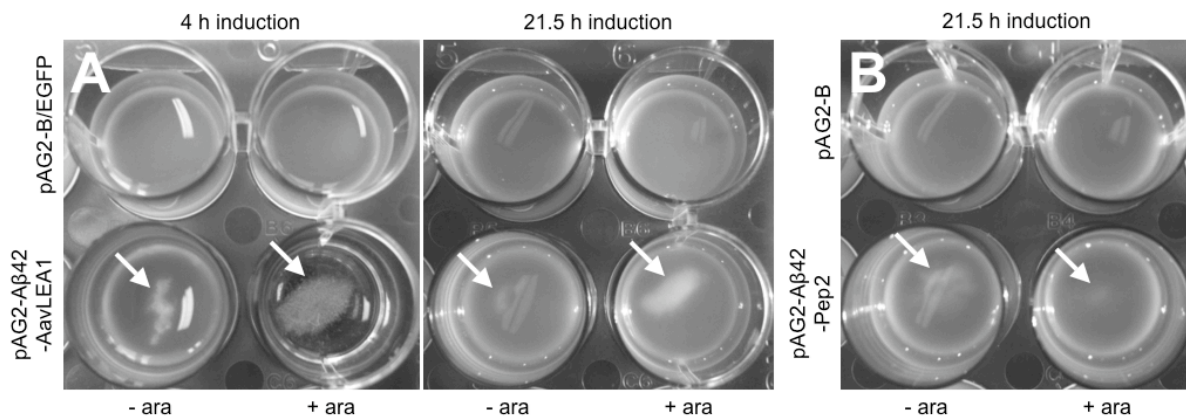


Figure 5.10: Clumping of cells is visible in cultures containing pAG2-A β 42-AavLEA1. Cultures grown as per Figure 5.9, photos taken after 4 h and 21.5 h induction. Cell clumping arrowed, prominent in pAG2-A β 42-AavLEA1 (**A**) but not in pAG2-A β 42-Pep2 (**B**). Ara, arabinose.

5.2.2.4 Flow cytometry of pAG2-A β 42 and controls

To quantitatively assess EGFP and mCherry fluorescence levels, *E. coli* DH10B cells containing the various pAG2-A β 42 constructs outlined in Figure 5.5 were induced for 22 h at 37°C and analysed via flow cytometry (see Section 2.7.5.3). Figure 5.11 shows the resulting dot plots and observed median fluorescence intensities (MFI) for both EGFP and mCherry.

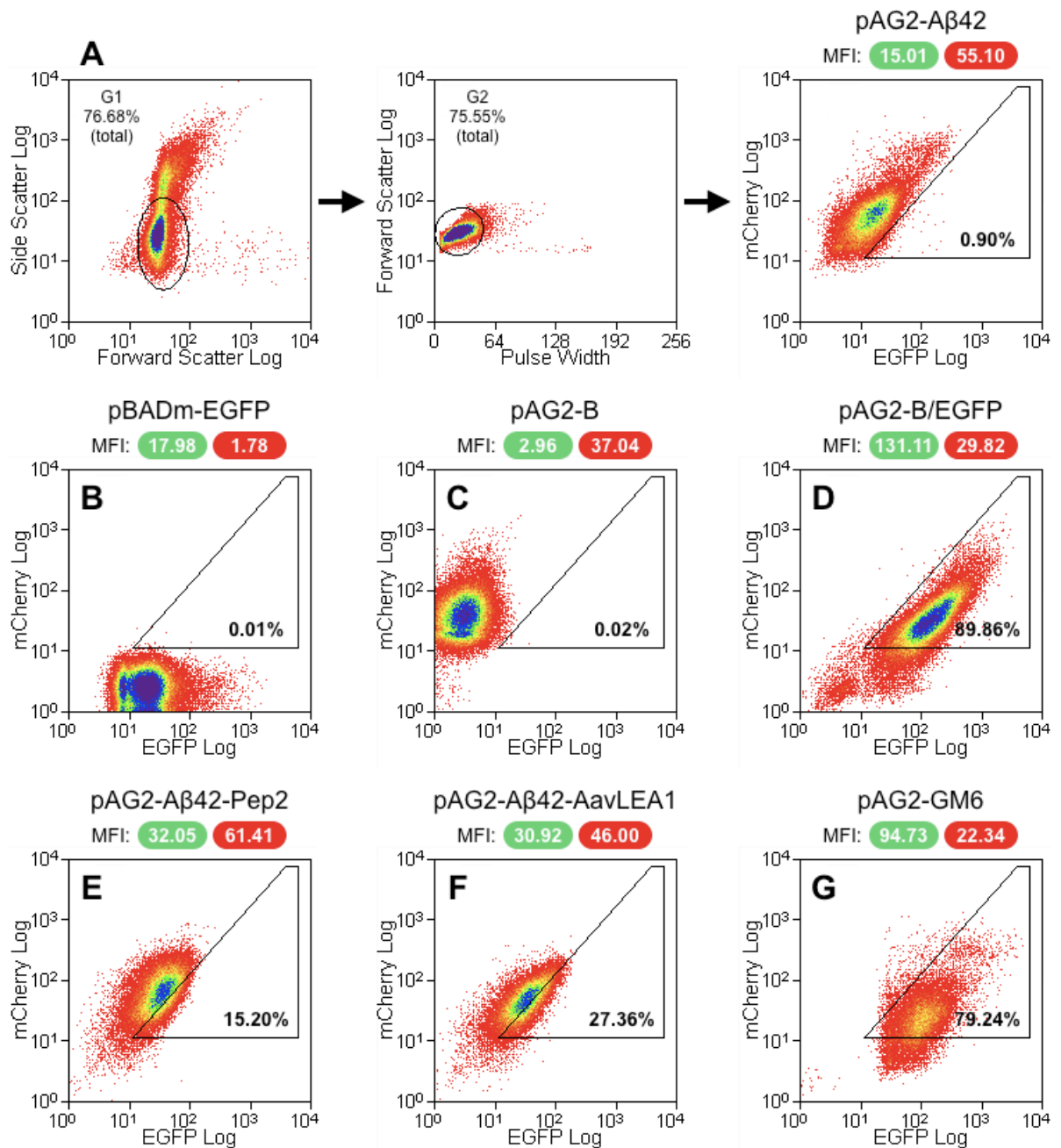


Figure 5.11: Flow cytometry dot plots from *E. coli* expressing pAG2-A β 42 and controls. Cultures induced for 22 h at 37°C at ~225 rpm before analysis of EGFP and mCherry fluorescence. **A:** Initial gating scheme used on all samples (**B** to **G**) to doublet discriminate (G1 & G2; pAG2-A β 42 cells shown as a representative; see Section 2.7.5.1). **B** to **G:** Dot plots for each sample with the same RoI (triangle, with % cells indicated) as the right-most plot of **A**. At least 35,000 total events were recorded for each dot plot (blue highest density, red lowest), representatives shown ($n = 3$). MFI, median fluorescence intensity (arbitrary units) for whole plot: EGFP (green bullet), mCherry (red bullet).

A region of interest (RoI), with a left-most border running parallel to the axis of the pAG2-A β 42 plot (panel A; see also Section 5.1.4.2), was set on the right-most border of the observed pAG2-A β 42 cell cluster and adjusted so that cells from the

single-fluorescence controls (panels B & C) did not fall within it. Approximately 1% of pAG2-A β 42 cells were allowed to enter this RoI, providing a comparison between this sample (pAG2-A β 42) and the various controls. For the non-aggregating pAG2-B/EGFP (panel D) and pAG2-GM6 (panel G) samples, the majority of cells exhibited greater EGFP fluorescence than pAG2-A β 42, thus moving along the X-axis and falling into the RoI. This also occurred for the cells co-expressing Pep2 and AavLEA1 (panels E & F), but to a lesser extent. *Figure 5.12* plots the observed MFIs for each sample, as well as showing the EGFP:mCherry ratios for comparison.

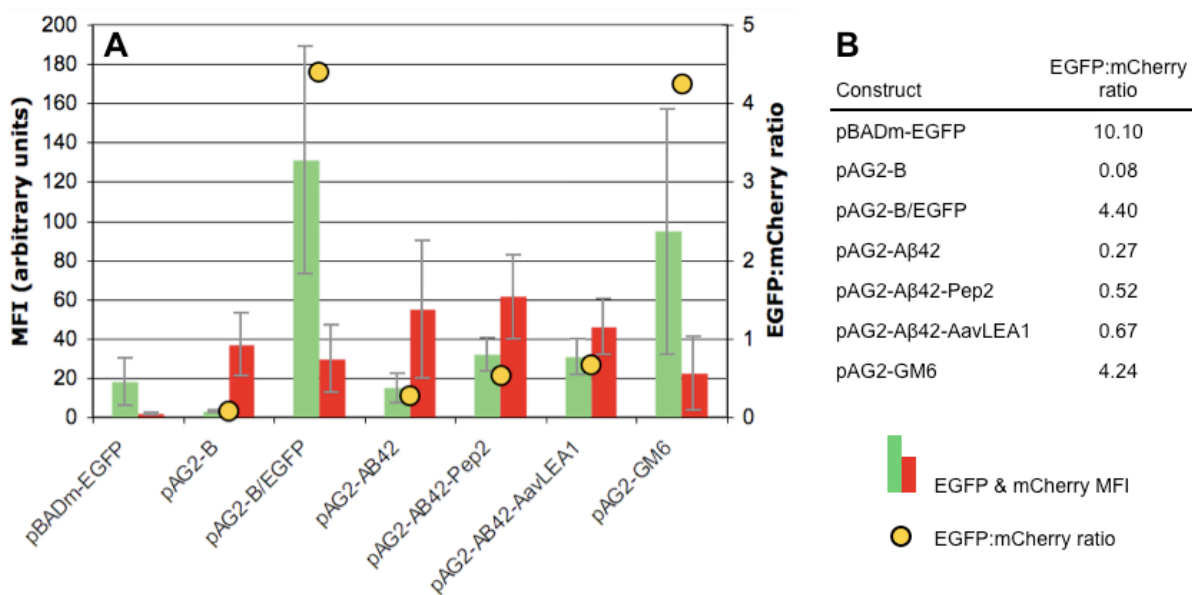


Figure 5.12: MFIs from *E. coli* expressing pAG2-A β 42 and controls. Data taken from representative plots shown in Figure 5.11 ($n = 1$). MFI, median fluorescence intensity; error bars represent coefficient of variation (see Section 2.7.5.4).

The median A β 42-EGFP fluorescence output from pAG2-A β 42-Pep2 was seen to be twice that of pAG2-A β 42 alone (panel A), with a tight coefficient of variation (CV), i.e. spread of data. This finding is in line with that observed by Baine et al. (2009) during their original work on Pep2. Furthermore, AavLEA1 seems to have a similar effect, signalling its potential as an A β 42 antiaggregant. However, neither of these constructs comes close to matching the non-aggregating pAG2-GM6 control, which shows a ~6-fold increase in EGFP output. This serves as a further indicator that Pep2 and AavLEA1 may only serve as rate, rather than absolute, inhibitors of A β 42 aggregation *in vivo* (see Section 5.2.2.2). The green positive control,

pBADm-EGFP, did not express well in this work, but given that pAG2-B/EGFP and pAG2-GM6 performed as expected (*Figure 5.5*), this was not remedied.

With regards to comparison of the A β 42-EGFP:mCherry ratios (panel B), they serve to confirm the median A β 42-EGFP readings. It is worth noting, however, that mCherry levels are quite varied over the samples. For instance, the mCherry MFI for pAG2-A β 42 alone is approximately twice that observed for pAG2-GM6, with a CV value indicating a large spread of data. This indicates that the use of mCherry as an internal *in vivo* standard is not as robust as initially envisaged.

Next, in order to see if cells containing the positive controls could easily be detected amongst a negative background prior to FACS, cultures of “empty” pAG2-A β 42 bacteria were spiked with cultures of pAG2-A β 42-Pep2, pAG2-A β 42-AavLEA1 or pAG2-GM6. Samples were created such that 100% (i.e. the positive control alone), 50%, 10% or 1% of the culture was derived from a particular spike source. *Figure 5.13* shows dot plots of the observed fluorescence outputs for each spiked sample.

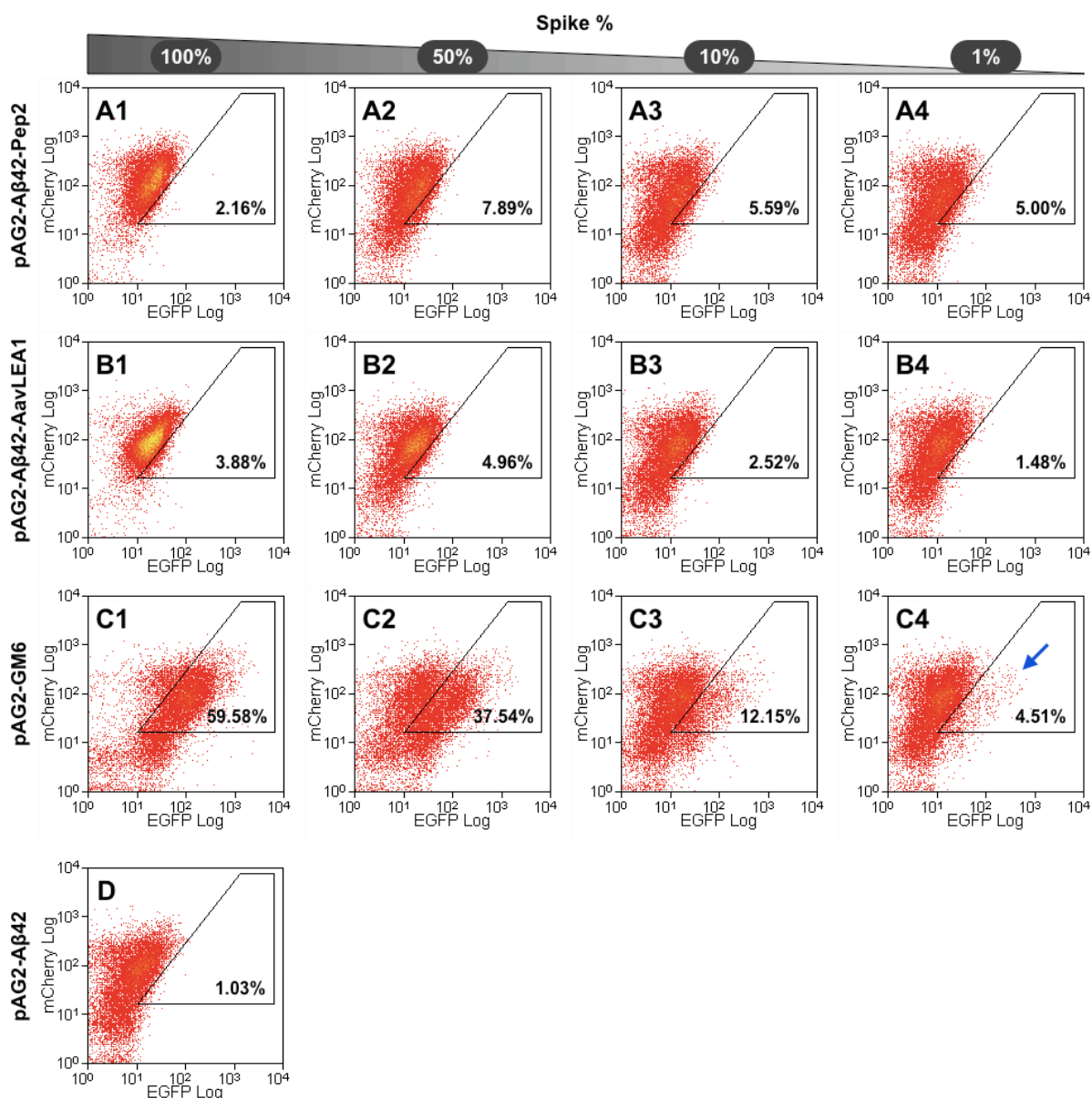


Figure 5.13: Flow cytometry dot plots from *E. coli* expressing pAG2-Aβ42 spiked with pAG2-GM6, pAG2-Aβ42-Pep2 or pAG2-Aβ42-AavLEA1 cultures. Cultures induced separately for 17 h at 37°C at ~225 rpm before spiking. Samples gated and analysed as per Figure 5.11. For pAG2-Aβ42 alone, 1.03% of cells fell into the Rol (panel D). *n* = 1.

The percentage of cells falling into the Rol for the positive controls alone (panels A1, B1 & C1) was observed to be less than that seen in Figure 5.11. The two experiments were conducted in the same manner but on different days; such a difference has been commonly observed due to biological and machine variance between measurement times (Kelly et al., 2009). Furthermore, all spikes with pAG2-Aβ42-Pep2 indicate a higher percentage of cells in the Rol than pAG2-Aβ42-Pep2 alone (panels A1-A4). The same applies for

pAG2-A β 42-AavLEA1 (panels B1-B4). Taken together, this noise indicates that isolating novel peptide hits with similar activity levels to Pep2/AavLEA1 in a library screen through a single gating step would be difficult to achieve using FACS. Instead, multiple rounds of FACS would be required to try to enrich for hit cells with the brightest A β 42-EGFP output. For any hits exhibiting greater EGFP fluorescence, however, a single sort may prove adequate: pAG2-GM6 cells can clearly be seen in panel C4 (arrowed), separate from the “empty” pAG2-A β 42 background (cf. panel D).

The next step was to construct random DNA library inserts for insertion into pAG2-A β 42, and use FACS to attempt to enrich and isolate putative antiaggregant hits.

5.2.3 Using FACS to screen for novel antiaggregant peptides

5.2.3.1 Creation of random DNA library

Oligonucleotides consisting of 74 or 110 bp were purchased as templates for the creation of random DNA libraries for insertion into pAG2-A β 42. These single-stranded DNA sequences contained a large run of random bases (36 or 72 bp; see *Table 5.2*) encoding either 12 or 24 amino acids if translated (excluding the initiating methionine and ignoring internal stop codons). To facilitate the insertion process, the 5' and 3' terminals of the oligonucleotides were conserved, containing a *SpeI* or *AfeI* site respectively. In-frame start and stop codons were also included (*Table 5.2*). However, synthesis of these oligonucleotides was difficult (personal communication, Sigma-Aldrich Company Ltd.), and consequently the number of random bases incorporated could not be absolutely guaranteed. Although translation frame-shifts could occur, the pAG2-A β 42 plasmid was designed to include stop codons in all possible reading frames downstream of the Library insertion point. Only two to five additional amino acids may be incorporated at the C-terminus if a frame-shift occurs (see *Figure 5.5*).

| Name | Sequence (5' to 3') | Notes |
|------|--|--|
| 74 | GCACGCTGCG <u>ACTAGTATG</u> (NNN) ₁₂ TGA <u>AAGCGCT</u> GATCCGATGC | 13 aa peptide; <i>SpeI</i> / <i>AfeI</i> underlined, start/stop codons in bold |
| 110 | GCACGCTGCG <u>ACTAGTATG</u> (NNN) ₂₄ TGA <u>AAGCGCT</u> GATCCGATGC | 25 aa peptide; <i>SpeI</i> / <i>AfeI</i> underlined, start/stop codons in bold |

Table 5.2: Random oligonucleotides used to create libraries encoding random peptides up to 13 or 25 amino acids in length. The name of the oligonucleotide reflects its total length in base pairs. N, random base.

Limited PCR amplification from the oligonucleotides was performed (see *Section 2.7.3*) to try to ensure library diversity by minimising any reaction bias towards a particular sequence. The 10 pmol of oligonucleotide used as a template, when multiplied by Avogadro's number, equates to a potential $\sim 6 \times 10^{12}$ different sequences. Given that only 1×10^4 to 1×10^5 different library members were to be screened during any one FACS run, a large buffer exists against any potential reduction in sequence diversity. Aliquots from each generated library were subsequently ligated into the pAG2-A β 42 vector via the *SpeI* and *AfeI* sites of the Library site and transformed into *E. coli* DH10B.

5.2.3.2 Plasmid stability and assessment of library insertion

During attempts to insert the random DNA libraries into pAG2-A β 42, large numbers of unwanted negative background colonies were obtained after each transformation. Colony numbers (thousands) were indistinguishable between insert ligation plates and those of negative controls, i.e. no insert, transformed with *SpeI*/*AfeI* digested pAG2-A β 42 plasmid. It was hypothesised that the pAG2-A β 42 vector may suffer from instability due to the presence of identical sequence regions, such as the *rrnB* terminator used in all three of the *araBAD*, Library and mCherry operons (see *Appendix 7*). This could allow for homologous recombination and therefore potential plasmid truncation to occur, possibly resulting in the removal of the Library site and hence loss of sensitivity to *SpeI*/*AfeI* digestion. Although the DH10B strain used in this work is a *recA1* mutant (inactivating *E. coli*'s major DNA recombinase), unwanted recombination events may still occur (Kawashima et al., 1984).

To explore this possibility, *SpeI/AfeI* digested pAG2-A β 42 was gel-extracted (see Section 2.2.2) to separate it from any putative recombinant/undigested plasmid, and then directly transformed (without ligation) into DH10B cells. Colonies still resulted, but in their hundreds rather than thousands as noted earlier. Two of these resultant negative control colonies were cultured so that their plasmids could be subjected to restriction digestion analysis (Figure 5.14).

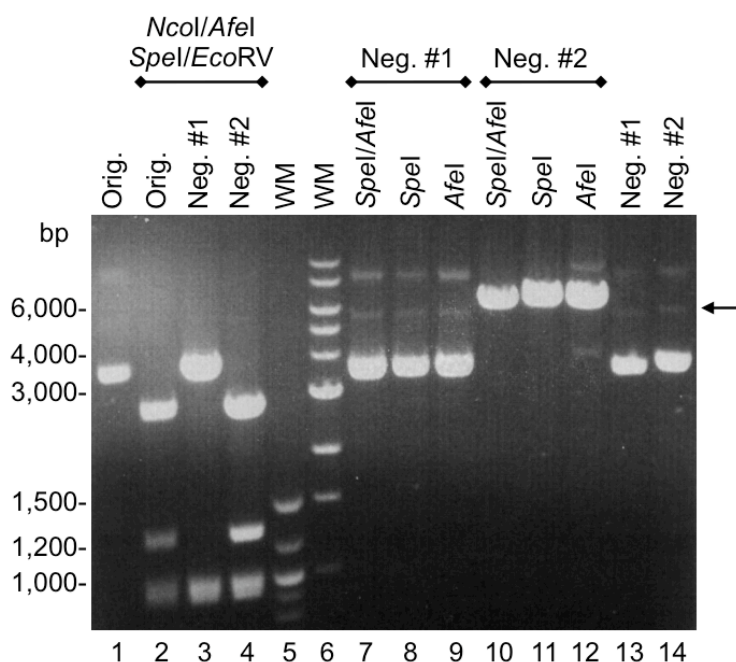


Figure 5.14: Restriction digestion analysis of negative control plasmids indicates pAG2-A β 42 suffers from sequence instability. *Orig.*, original pAG2-A β 42 plasmid; *Neg.*, plasmid isolated from a negative control colony. Desired band sizes for wild-type pAG2-A β 42: *NcoI/AfeI/SpeI/EcoRV*, 2,809, 1,279, 984, 931 & 11 bp; *AfeI*, *SpeI* or both, ~6,010 bp. *WM*, DNA ladder.

Wild-type pAG2-A β 42 was expected to possess bands that summed to ~6 kbp in size. While the original plasmid (*Orig.*; lane 2) and the second negative control plasmid (*Neg. #2*; lane 4) met this criteria when multiply digested, the first negative control plasmid (*Neg. #1*; lane 3) was seen to lack the *SpeI* and *AfeI* sites (confirmed lanes 7 to 14) between the two *NcoI* sites of pAG2-A β 42. If full-length, the largest *NcoI*-related band should be ~4.1 kbp; instead it was observed to be ~3.7 kbp (lane 3). These observations indicate that *Neg. #1* lacks a ~400 bp region that encompasses the Library site. Analysis of the sequence of pAG2-A β 42 (*Appendix 7*) shows that the 5' ends of the homologous *rrnB* terminator regions of

the *araBAD* site and the Library site are ~330 bp apart, indicating that these are the probable recombination sites.

Furthermore, it should be remembered that plasmids Neg. #1 and Neg. #2 both originated from the same transformation using the Orig. plasmid (lane 1). The restriction digest profile of Orig. (lane 2), however, appears identical to that of Neg. #2 (lane 4) in *Figure 5.14*; no trace of the Neg. #1 (lane 3) profile is seen. This indicates that any truncated plasmid is only present in minute (picogram) amounts, as nanogram quantities of DNA are required for detection using ethidium bromide (*Section 2.2.7*; Sambrook & Russell [2001]). While gel-extraction of *SpeI/AfeI* digested pAG2-A β 42 led to a reduction in negative background colonies, it seems impossible to completely remove the undesired truncated plasmid. Upon close examination of *Figure 5.14*, a faint band of relaxed Neg. #1 plasmid can be seen to run just below the 6 kbp marker (lanes 7 to 9; arrowed). As this runs at a similar size to *SpeI/AfeI* digested pAG2-A β 42 (lane 10), contamination and co-purification may occur during the gel-extraction process. New truncation events may also occur. Nevertheless, ligation of the random DNA libraries into gel-extracted pAG2-A β 42 was trialled again to determine the success rate of insertion. Approximately 75% of colonies were found to contain a desired insert in the Library site, as revealed by colony PCR (*Figure 5.15*). Given that negative background colonies containing truncation products should not exhibit any increase in A β 42-EGFP fluorescence, this situation was deemed acceptable enough to proceed with screening.

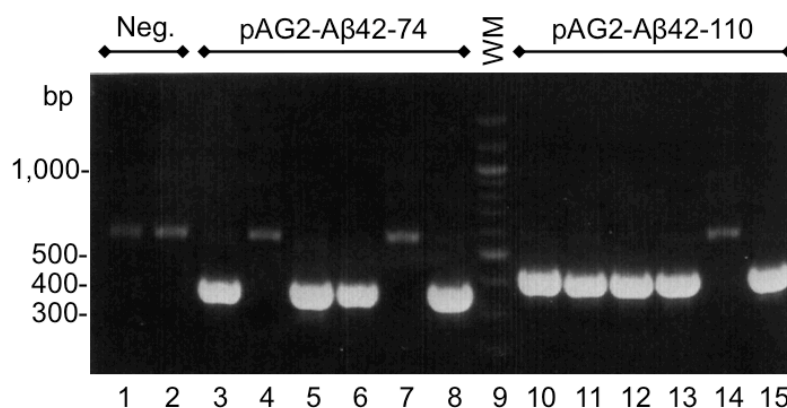


Figure 5.15: Profile of random DNA library inserts ligated into the Library site of gel-extracted pAG2-A β 42. Colony PCR (Section 2.2.5.2) shows the presence of inserts in 9 out of 12 colonies assayed. Empty plasmid should result in a 333 bp band; 74 library insert, 372 bp; 110 library insert, 408 bp. Neg., colonies from a negative control transformation plate; WM, DNA ladder.

Of additional note, colonies picked from a negative control transformation plate (lanes 1 & 2) exhibited a faint band running at ~575 bp. This is an artefact most likely endowed by truncated plasmid: if a Library site was present, a band of 333 bp was expected. Other negative colonies (lanes 4, 7 & 14) gave a similar profile.

5.2.3.3 FACS of pAG2-A β 42-74 and pAG2-A β 42-110 libraries

Two libraries (approximately 25,000 colonies each) from pAG2-A β 42-74 (encoding a 12 aa random peptide) or pAG2-A β 42-110 (encoding a 24 aa random peptide) ligations were prepared, with plate incubation taking place at 30°C to ensure a small colony size and hence high plate density. Colonies were pooled and resuspended as per Section 2.7.5.2. In addition, a similar resuspension of “empty” pAG2-A β 42 colonies was prepared, and subsequently spiked with several colonies of either pAG2-A β 42-Pep2, pAG2-A β 42-AavLEA1 or pAG2-GM6 to act as positive controls (a $\sim 1 \times 10^{-4}$ dilution/0.01% spike; see Table 5.3). It was noted that colonies of pAG2-A β 42-Pep2 and pAG2-A β 42-AavLEA1 alone were smaller than those of the 74 and 110 library colonies (see Section 5.2.2.3), so incubation was continued until approximate size parity was achieved.

Inoculums from the above preparations were induced for ~15 h at 30°C and subjected to FACS (see Section 2.7.5.3). Approximately 1×10^6 cells were analysed per sample, with the top 1% of A β 42-EGFP fluorescence from each sample being

gated and adjusted individually (as per Section 5.2.2.4) prior to sorting into fresh medium (approximately 1×10^4 cells captured). To enrich for high A β 42-EGFP fluorescence, these sorted cultures were directly induced and sorted again for a total of three sorts. Figure 5.16 outlines the workflow and shows the resulting dot plots.

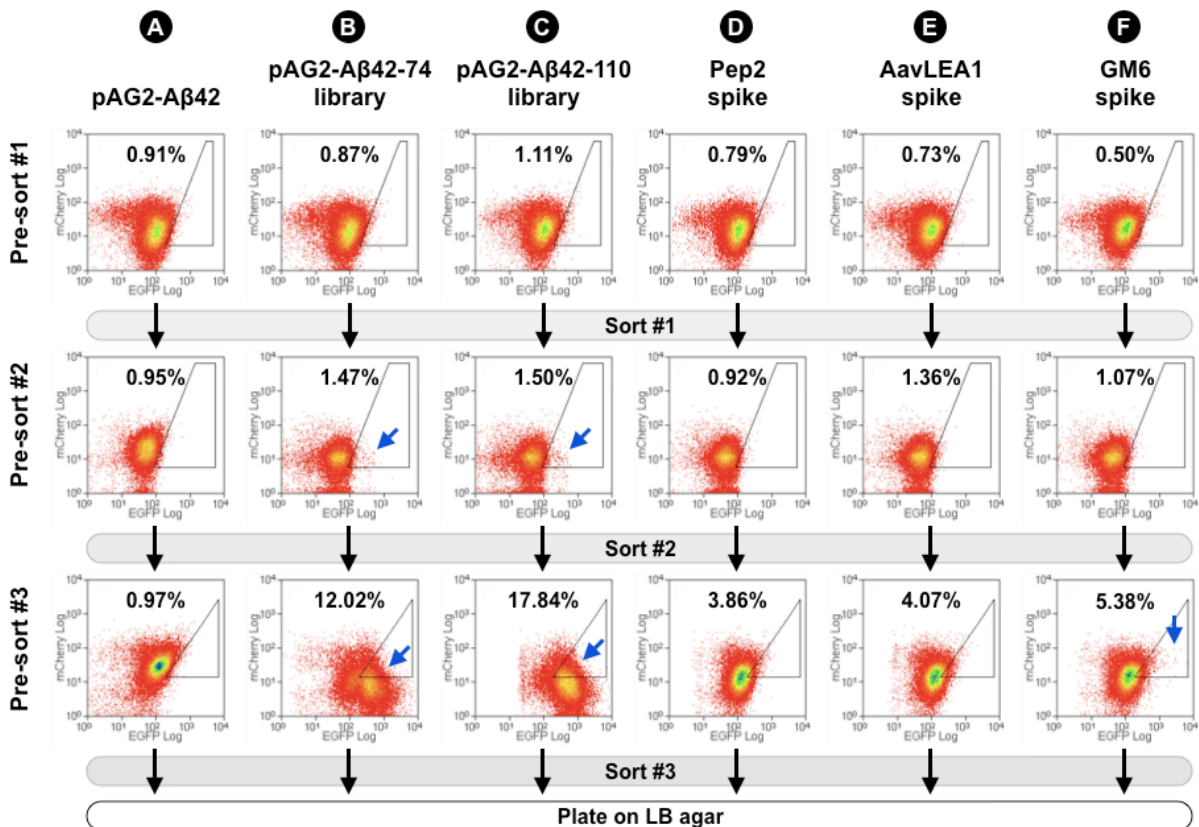


Figure 5.16: Flow cytometry dot plots from FACS of *E. coli* expressing pAG2-A β 42 random peptide libraries and spiked controls. Cultures induced for between 11-15 h at 30°C at ~225 rpm. Gating was as per Figure 5.11; cells with the top 1% of A β 42-EGFP fluorescence that exceeded mCherry background fluorescence were sorted into fresh medium ($\sim 1 \times 10^6$ cells screened per sample). Although this top 1% gate was adjusted for each sample sorted, the same RoI (calibrated in each row to pAG2-A β 42 alone [column A]) is shown for the purposes of comparison between FACS rounds. $n = 1$.

To easily detect enrichment and allow for comparison, the RoI for each round of FACS in Figure 5.16 is shown as calibrated to the pAG2-A β 42 control (column A; shifted to top 1% for each sort). For the spiked positive control cultures, there are indications of enrichment occurring after Sort #2 for the Pep2 and AavLEA1 samples (columns D & E), as the percentage of cells falling into the RoI increases above 1%. This is also seen for the brighter EGFP output of GM6 (column F, arrowed). This experiment was not repeated, however, therefore any possible

enrichment effects are indicative only. With regards to the random peptide libraries, after Sort #1 a small sub-population of cells with brighter A β 42-EGFP fluorescence was seen in the pAG2-A β 42-74 and pAG2-A β 42-110 library samples (columns B & C, arrowed). After Sort #2, these populations were predominant, but appeared to have low levels of mCherry fluorescence (MFIs of 7.84 and 8.12 respectively, compared to 28.76 for the pAG2-A β 42 control). After Sort #3, work was halted to investigate this problem further (see *Section 5.2.3.4* below). Furthermore, the incubation temperature of 30°C used during the rounds of FACS was seen to have an effect on A β 42-EGFP output: the difference between pAG2-GM6 and pAG2-A β 42 fluorescence shrank from ~6-fold (*Section 5.2.2.4*) to ~2-fold (e.g. in Pre-sort #3, EGFP MFIs of 290.23 and 113.48 respectively). This reduction in dynamic range could be explained by a decrease in the rate of A β 42 aggregation *in vivo* at this lower temperature, thus allowing sufficient time for some EGFP to fold.

Before continuing to employ FACS as a screening methodology, it was worth considering the expected enrichment after each round of sorting. If each positive spike is gated for sorting as per the RoI used in *Figure 5.16*, i.e. ~1% of background cells falling into the RoI, then ~10,000 negative cells will be obtained from the sorting of 1×10^6 cells (see *Section 2.7.5.3*). If it is conservatively estimated that only ~5% of pAG2-A β 42-Pep2/AavLEA1 cells meet the same RoI criteria, i.e. a 5-fold enrichment, then only 5 positive cells would be selected for from a 1×10^4 starting dilution (e.g. 5% of 1×10^6 cells sorted multiplied by 1×10^{-4}). This number runs perilously close to being lost in background noise, and is similar for the pAG2-A β 42-74/110 libraries. Sorting ten-fold more cells would be desirable, i.e. 1×10^7 , but time-consuming (~45 min per run). In addition, given that cells containing inserts were seen to be slow to enter exponential phase growth (e.g. pAG2-A β 42-Pep2/AavLEA1; see *Section 5.2.2.3*), any positive cells may be outcompeted at the very first round of sorting by “empty” pAG2-A β 42. For GM6, ~35% of cells will conservatively fall in the RoI, resulting in 35 positive cells being successfully selected. Using the above assumptions of enrichment, i.e. 5-fold for Pep2/AavLEA1 (as well as for a single putative hit from the ~25,000 colony pAG2-A β 42-74/110 libraries, i.e. a 4×10^{-5} starting dilution) and 35-fold for GM6, the percentage of true

positive cells in a population after each round of FACS can be estimated (Table 5.3).

| FACS Round | Positive cell % of total cell population | | | | |
|----------------|--|-------------|------------|---------------|-----------|
| | 74 library | 110 library | Pep2 spike | AavLEA1 spike | GM6 spike |
| <i>Initial</i> | 0.004% | 0.004% | 0.01% | 0.01% | 0.01% |
| #1 | 0.02% | 0.02% | 0.05% | 0.05% | 0.35% |
| #2 | 0.10% | 0.10% | 0.25% | 0.25% | 12.25% |
| #3 | 0.50% | 0.50% | 1.25% | 1.25% | ~100% |
| #4 | 2.50% | 2.50% | 6.25% | 6.25% | |
| #5 | 12.50% | 12.50% | 31.25% | 31.25% | |
| #6 | 62.50% | 62.50% | ~100% | ~100% | |
| #7 | ~100% | ~100% | | | |

Table 5.3: Estimated levels of enrichment after rounds of FACS. Positive cell % of total cell population calculated as the enrichment factor to the power of n sorts, multiplied by the initial cell dilution expressed. Assumptions: % of positive cells that fall into sorting gate: pAG2-A β 42-74/110/Pep2/AavLEA1, 5%; pAG2-GM6, 35%. Starting dilutions: pAG2-A β 42-74/110 libraries, 4×10^{-5} ; pAG2-A β 42-Pep2/AavLEA1 and pAG2-GM6, 1×10^{-4} . The consequent reduction of negative cells in the total population during subsequent FACS rounds is ignored, as is any growth disadvantage incurred by positive cells (see Section 5.2.2.3).

It can be seen that at least six rounds of FACS would be required to enrich the Pep2 and AavLEA1 spikes to near homogeneity; at three rounds, only ~1.25% of cells in the post-sort culture are expected to be truly positive. With regards to Figure 5.16, such numbers would be lost in the pAG2-A β 42 negative background (cf. Figure 5.13). The situation is similar for the pAG2-A β 42-74/110 libraries, where theoretically 1 potential hit could be present amongst the ~25,000 library colonies screened (a starting dilution of 4×10^{-5}). For the GM6 spike, three rounds would be sufficient: only a small fraction is seen in Figure 5.16 after Sort #2 (column F, arrowed), as pAG2-GM6 cells are only expected to make up ~12% of the total population at this stage (multiplied by the 35% expected RoI event rate, this equates to ~4.3% of total cells).

5.2.3.4 Further analysis of FACS data

As *Section 5.2.3.3* indicated, *E. coli* cells containing members of the pAG2-A β 42-74 and pAG2-A β 42-110 libraries were enriched for increased green fluorescence over the course of three rounds of FACS, with a correlated decline in red fluorescence. It appeared that plasmid instability was not just confined to the Library region (see *Section 5.2.3.2*). Upon fluorescence microscopy examination (*Section 2.7.4*), cells with bright green fluorescence were observed (*Figure 5.17 A*), but always lacked mCherry fluorescence. Those with dim green fluorescence were found to be mCherry positive. In addition, bright green fluorescence was seen to be constitutive, as samples plated on solid medium after the final round of FACS gave colonies that continued to fluoresce, despite the lack of inductive agent (*Figure 5.17 B*). It seems that a recombination event between the EGFP and mCherry coding regions of pAG2-A β 42 occurred and was enriched for; *Figure 5.17 C* shows a rare split colony event arising on solid medium. However, a lack of mCherry fluorescence from such cells should have resulted in their rejection by the gating criteria used in *Section 5.2.3.3*. Re-examination of the pAG2-A β 42-74/110 cell populations in pre-sort #3 of *Figure 5.16* (columns B & C, arrowed), however, shows this not to be the case, as a large number of EGFP-bright cells breach the mCherry cut-off and fall into the RoI. If FACS were to be performed again, this gating criterion must be more stringent.

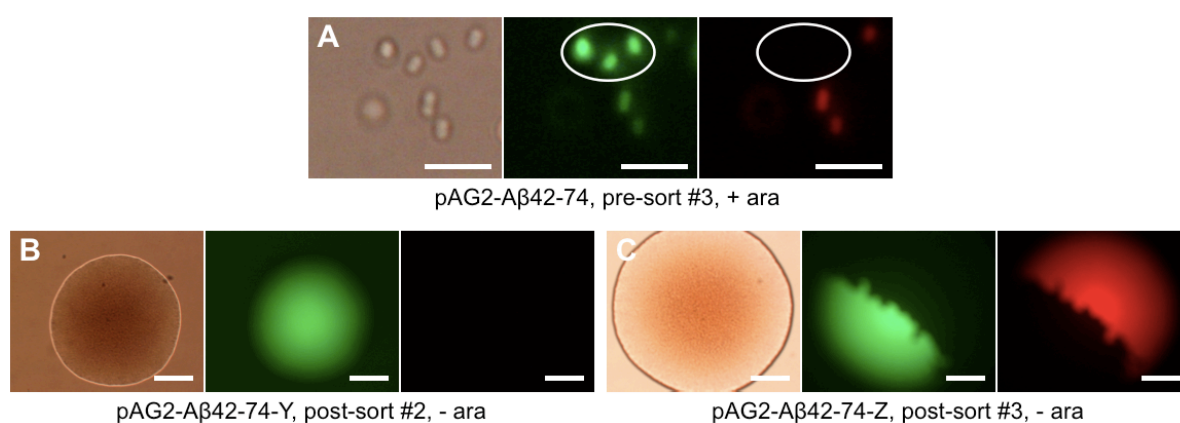


Figure 5.17: FACS of *E. coli* expressing pAG2-A β 42 libraries enriches for mutant recombinants. For each panel, from left to right: bright field, green fluorescence, red fluorescence. **A:** Cells with bright green fluorescence lack red fluorescence (circled). **B & C:** Cells plated on non-inducing medium give rise to colonies exhibiting plasmid recombination. Photos taken after 11 h arabinose induction (0.05%; + ara) in LB medium at 37°C at ~225 rpm (panel A) or 16 h at 37°C without arabinose (panels B & C; - ara). Scale bar, 5 μ m (panel A) or 250 μ m (panels B & C).

Section 5.2.3.3 also indicated an effect of temperature on the aggregation kinetics of A β 42-EGFP. To briefly explore this, several control cultures were induced at either 30°C or 37°C for 15 h and examined using fluorescence microscopy (see Section 2.7.4). At 30°C, pAG2-A β 42 cells showed dim fluorescence, whereas at 37°C no fluorescence was seen (Figure 5.18 A & B). Despite this, the *in vivo* effect of Pep2 and AavLEA1 on A β 42-EGFP aggregation can again be seen (panels C & E; cf. Figure 5.8): the slower aggregation kinetics at 30°C simply makes their effects more pronounced.

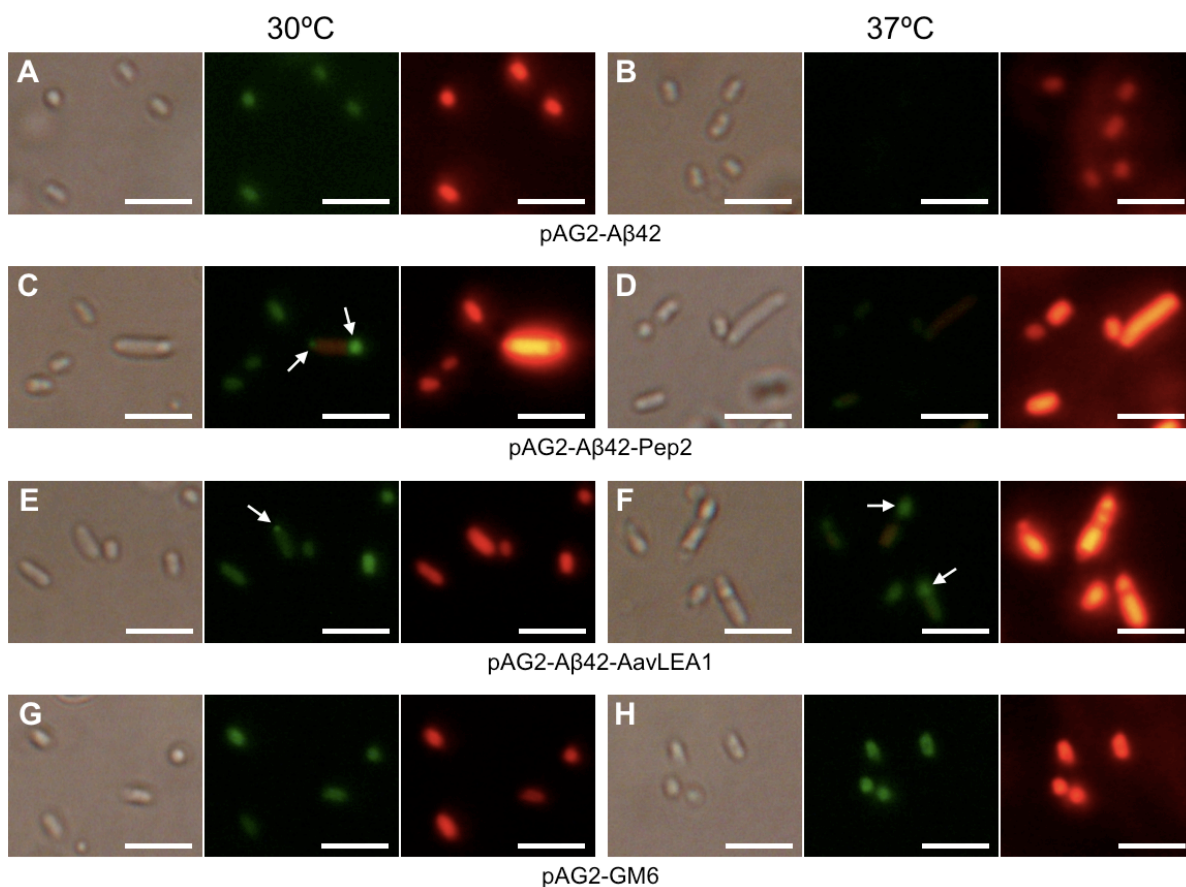


Figure 5.18: E. coli expressing pAG2-A β 42 and controls fluoresce brighter green at 30°C than at 37°C. For each panel, from left to right: bright field, green fluorescence, red fluorescence. Photos taken after 15 h arabinose induction (0.05%) in LB medium at 30°C (left column) or 37°C (right column) at ~225 rpm. Green aggregates arrowed (panels C, E & F). Scale bar, 5 μ m.

Lastly, colonies that resulted from plating out the pAG2-A β 42-74/110 and spiked libraries after the third round of FACS in Section 5.2.3.3 were analysed for the presence of insert in their Library site (Figure 5.19; pAG2-GM6 samples contained no Library insert, so were not examined). Before selection, colonies were examined

via microscopy to confirm that they were fluorescently “normal”, i.e. did not exhibit constitutive green fluorescence. In addition, two such constitutive green mutants were also analysed.

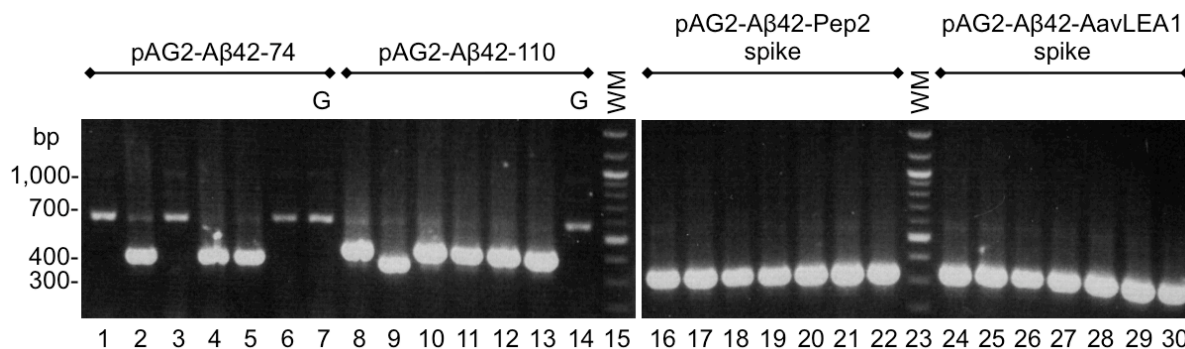


Figure 5.19: Profile of Library site inserts after three rounds of FACS. Colony PCR (Section 2.2.5.2) shows the presence of inserts in 8 out of 12 pAG2-A β 42-74/110 colonies assayed, and in 0 out of 14 pAG2-A β 42-Pep2/AavLEA1 spiked samples. Empty plasmid should result in a 333 bp band; 74 library insert, 372 bp; 110 library insert, 408 bp; Pep2, 370 bp; AavLEA1, 754 bp. G, constitutive green fluorescent mutant; WM, DNA ladder.

All isolated colonies from the pAG2-A β 42-Pep2 and pAG2-A β 42-AavLEA1 spiked samples were found to contain the “empty” pAG2-A β 42 plasmid (lanes 16 to 30); as expected, no Pep2 or AavLEA1 insertions were detected at such an early stage of enrichment (see Section 5.4.3.3). In addition, the constitutive green mutants (lanes 7 & 14) exhibited the ~575 bp artefact band, previously seen to be typical of plasmids lacking the Library site (Section 5.2.3.2). Further analysis of these mutants was not conducted.

Finally, inserts from the sorted libraries of pAG2-A β 42-74 (lanes 4 & 5) and pAG2-A β 42-110 (lanes 10 to 13) were sequenced to check for any possibility of homogeneity at this early stage. Every sequence was found to vary (data not shown). This, together with the lack of positive control isolation from the positive control pAG2-A β 42-Pep2/AavLEA1 spikes, confirmed that the FACS selection criteria used were not stringent enough to resolve true hits from background noise over the number of sorts conducted. Given that a much larger number of sorting rounds are required (see Section 5.2.3.3), and that potential positive hits may be outcompeted during culture growth (Section 5.2.2.3), FACS seems ill-suited to this particular screen. A simpler approach of screening library members directly on solid medium was instead investigated.

5.2.4 Using solid medium to screen for novel antiaggregant peptides

To solve the twin problems of multiple FACS rounds being required to enrich a hit while contending with potential out-competition during growth by negative background cells (*Section 5.2.3.4*), pAG2-A β 42 library members may instead be spatially isolated and induced as colonies on solid medium. Those containing successful Library site inserts would merely grow to a smaller size in a given time (as mentioned in *Section 5.2.3.3*), and colonies with the highest levels of EGFP fluorescence may be selected for further analysis using more sensitive techniques such as flow cytometry.

5.2.4.1 Fluorescence of colonies expressing pAG2-A β 42 and controls

To test the solid medium screening approach, cultures of pAG2-A β 42 and the various controls were plated on inductive LB agar (0.05% arabinose) and incubated for ~16 h at 37°C. In addition, cultures of “empty” pAG2-A β 42 were spiked at a ratio of 9:1 with cultures of pAG2-A β 42-Pep2, pAG2-A β 42-AavLEA1 or pAG2-GM6 to ascertain whether or not these positive control colonies could be readily identified from a negative background. Resultant colonies were examined by fluorescence microscopy (*Figure 5.20*).

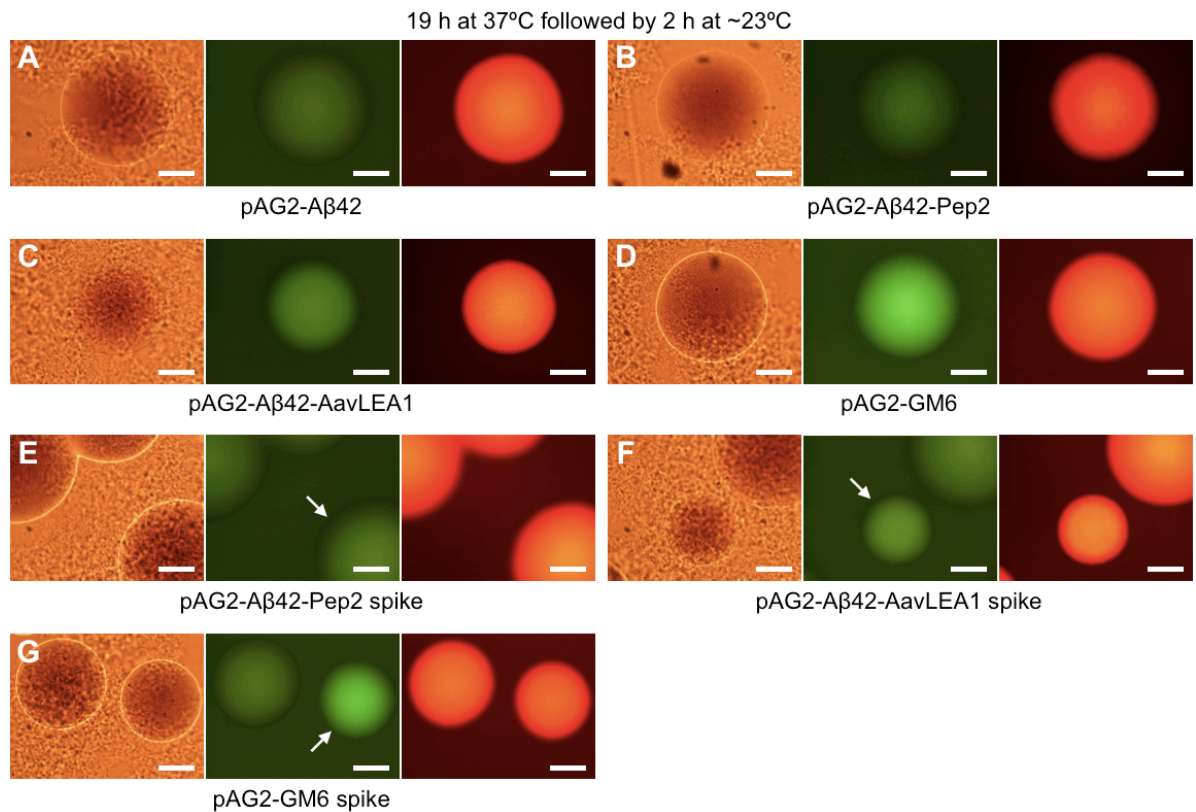


Figure 5.20: Fluorescence microscopy of *E. coli* colonies expressing pAG2-A β 42 and controls. For each panel, from left to right: bright field, green fluorescence, red fluorescence. Arrows indicate putative hit colonies in pAG2-A β 42 samples spiked with indicated controls at a ratio of 9:1. Photos taken after 19 h arabinose induction (0.05%) on LB agar at 37°C, then 2 h at ~23°C. Scale bar, 250 μ m.

While pAG2-A β 42 exhibited faint green fluorescence (panel A), encouragingly the pAG2-A β 42-AavLEA1 colonies appeared noticeably brighter (panel C). It was, however, difficult to interpret an effect given by pAG2-A β 42-Pep2 (panel B). With regards to the sample spikes, putative hit colonies were also detected amongst the background of pAG2-A β 42 colonies, even for Pep2 (panels E to G, arrowed). Furthermore, all colonies examined also expressed mCherry, indicating against undesired plasmid recombination events.

It should be noted that the time delay between removing the plates from incubation at 37°C and inspecting them at room temperature (~23°C) improved the contrast between negative and positive colonies (Figure 5.21). Leaving the plates at ~23°C for 5 h before imaging, rather than 2 h, allowed for a green fluorescence output difference to be more clearly seen between pAG2-A β 42 and pAG2-A β 42-Pep2 (panel E). This effect was especially pronounced for the pAG2-A β 42-AavLEA1

construct (panel C), which became as bright as the non-aggregating pAG2-GM6 control (panel D). This increased amount of green fluorescence was likely due not only to a reduced frequency of encounter between A β 42-EGFP molecules (hence a decrease in A β 42 aggregation; see *Section 5.2.3.4*), but also because of a reduced stability of hydrophobic forces at lower temperatures (Chandler, 2005). The same effect was also seen when plates were incubated at 30°C instead of 37°C (data not shown).

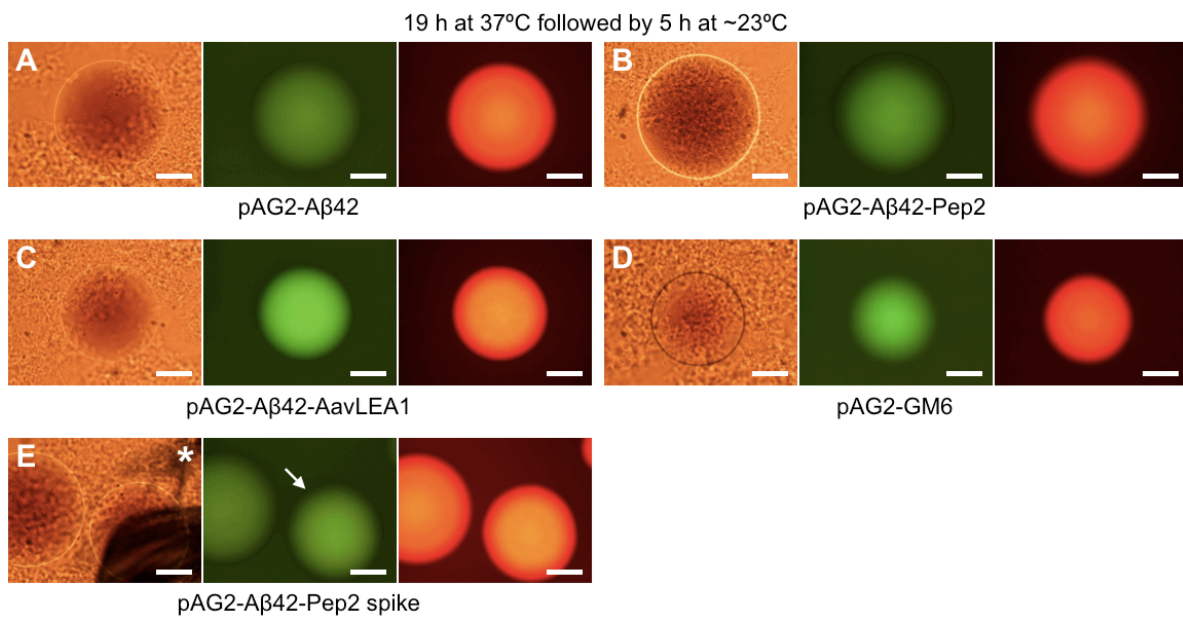


Figure 5.21: Fluorescence microscopy of *E. coli* colonies expressing pAG2-A β 42 and controls: 5 h at ~23°C. Same as Figure 5.20, but photos taken after 5 h at ~23°C post-37°C incubation. *, marker pen artefact. Scale bar, 250 μ m.

Next, the overall green fluorescence from each spiked plate was photographed and the images processed as per *Section 2.7.6* to emphasise colonies with the brightest output (*Figure 5.22*). At least ten hits identified in this manner were manually verified using fluorescence microscopy as above; all showed a level of green fluorescence greater than that of adjacent negative colonies. Note that colonies are required to be of a reasonable size (>0.5 mm in diameter), as otherwise the fluorescence signal was lost to the camera. Panel C illustrates this, with colony size being reduced due to a high cell density.

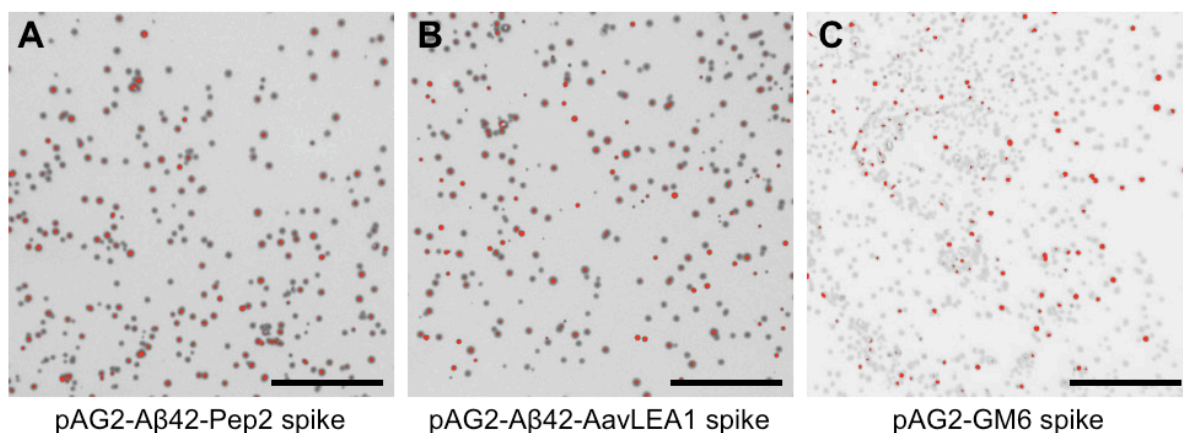


Figure 5.22: Fluorescence image processing of induced pAG2-Aβ42 E. coli colonies spiked with indicated controls. Colonies highlighted in red indicate putative hits in pAG2-Aβ42 samples spiked with indicated controls at a ratio of 9:1. Note that the dynamic range between background and hit is greater for the pAG2-GM6 spike (panel C), hence a greater contrast. Photos taken after 19 h arabinose induction (0.05%) on LB agar at 37°C, then ~4 h at ~23°C. Scale bar, 10 mm.

To confirm correct colony identification, selected hits from the spiked plates shown in *Figure 5.22* were analysed for the presence of the appropriate Pep2 or AavLEA1 Library site insert by colony PCR. (pAG2-GM6 hits could not be identified in this manner, as the construct contains no insert.) Negative colonies (i.e. those with background levels of green fluorescence) were not analysed. *Figure 5.23* shows that 4 from 5 of the selected pAG2-Aβ42-Pep2 hits contained the correct insert (lanes 3 to 7), while for pAG2-Aβ42-AavLEA1 (lanes 11 to 15) the success rate was 5 from 5. The fluorescence imaging and colony PCR results together confirm and extend the work of Wurth et al. (2002) and Baine et al. (2009), indicating that screening *in vivo* peptide libraries using solid medium is a pursuable strategy.

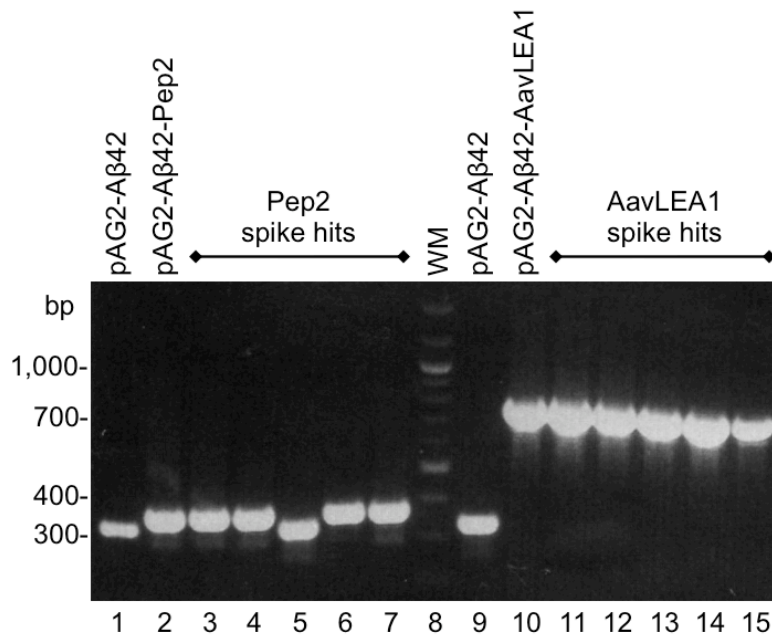


Figure 5.23: Profile of hit colonies identified by fluorescence image analysis. Colony PCR (Section 2.2.5.2) shows the presence of inserts in 4 out of 5 pAG2-A β 42-Pep2 spiked hits, and in 5 out of 5 pAG2-A β 42-AavLEA1 spiked hits. Empty plasmid should result in a 333 bp band; Pep2, 370 bp; AavLEA1, 754 bp. WM, DNA ladder.

5.2.4.2 Screening for novel antiaggregant peptides

The ligated pAG2-A β 42-74 and pAG2-A β 42-110 libraries generated in Section 5.2.3.1 were transformed again into *E. coli* DHB10 and induced on agar as per Section 2.7.3. Approximately 38,000 pAG2-A β 42-74 and 36,000 pAG2-A β 42-110 colonies were generated and imaged as per Section 2.7.6 to try to identify colonies with an above-background green fluorescence output (Figure 5.24 A & B). A number of putative hits were identified, and of these 114 (~0.15% of total) were manually verified using fluorescence microscopy (Section 2.7.4), with mCherry expression also being confirmed (Figure 5.24 C & D). This represented an initial hit rate of approximately 0.21%, as only ~75% of the ~74,000 total colonies plated were expected to contain library inserts (due to contaminating recombinant plasmid; see Section 5.2.3.2 for explanation). In addition, some mutant recombinants were observed at a frequency of one or two colonies per plate, i.e. bright green fluorescence without red fluorescence (see Section 5.2.3.4). While these colonies upset the contrast of the image acquisition process (see panel B), they did not ultimately interfere with the screen. The hit colonies visualised in panels C & D were not pursued further due to their adjacency

to negative colonies; rather, these images serve as further confirmation that the screening methodology is able to isolate putative antiaggregation events.

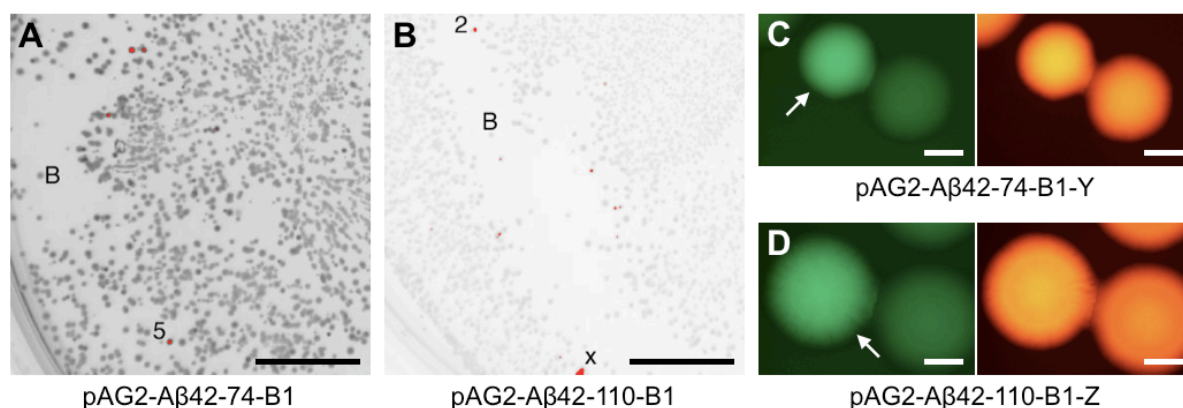


Figure 5.24: Fluorescence image processing of induced pAG2-A β 42 *E. coli* colonies reveals those that may express a novel antiaggregant. Representative experiment shown. Colonies highlighted in red indicate putative hits; those numerically labelled were investigated further (negative controls are lettered). In panel B, x denotes a bright green hit with no red fluorescence (an undesired recombinant), which increased the contrast of the image. Arrows indicate examples of hit colonies under fluorescence microscopy. Photos taken after 16 h arabinose induction (0.05%) on LB agar at 37°C, then ~3.5 h at ~23°C. Representative photos shown ($n = 20$ plates per library); scale bar, 10 mm (panels A & B) or 250 μ m (panels C & D).

5.2.4.3 Validation through flow cytometry

Of the putative antiaggregant hits identified in Section 5.2.4.2, nine pAG2-A β 42-74 and nine pAG2-A β 42-110 colonies were selected for further analysis using flow cytometry. In addition, two pAG2-A β 42-74 and three pAG2-A β 42-110 colonies that showed no signs of increased A β 42-EGFP fluorescence were selected as negative controls (e.g. labelled “B” in Figure 5.24 A & B). Cultures were induced for 14 h at 37°C and analysed as per Section 2.7.5.1, with a delay of ~2 h at ~23°C before flow cytometry to try to maximise green fluorescence output (see Section 5.2.4.1). The resulting dot plots for several samples are shown in Figure 5.25 (see Appendix 10 for all dot plots obtained).

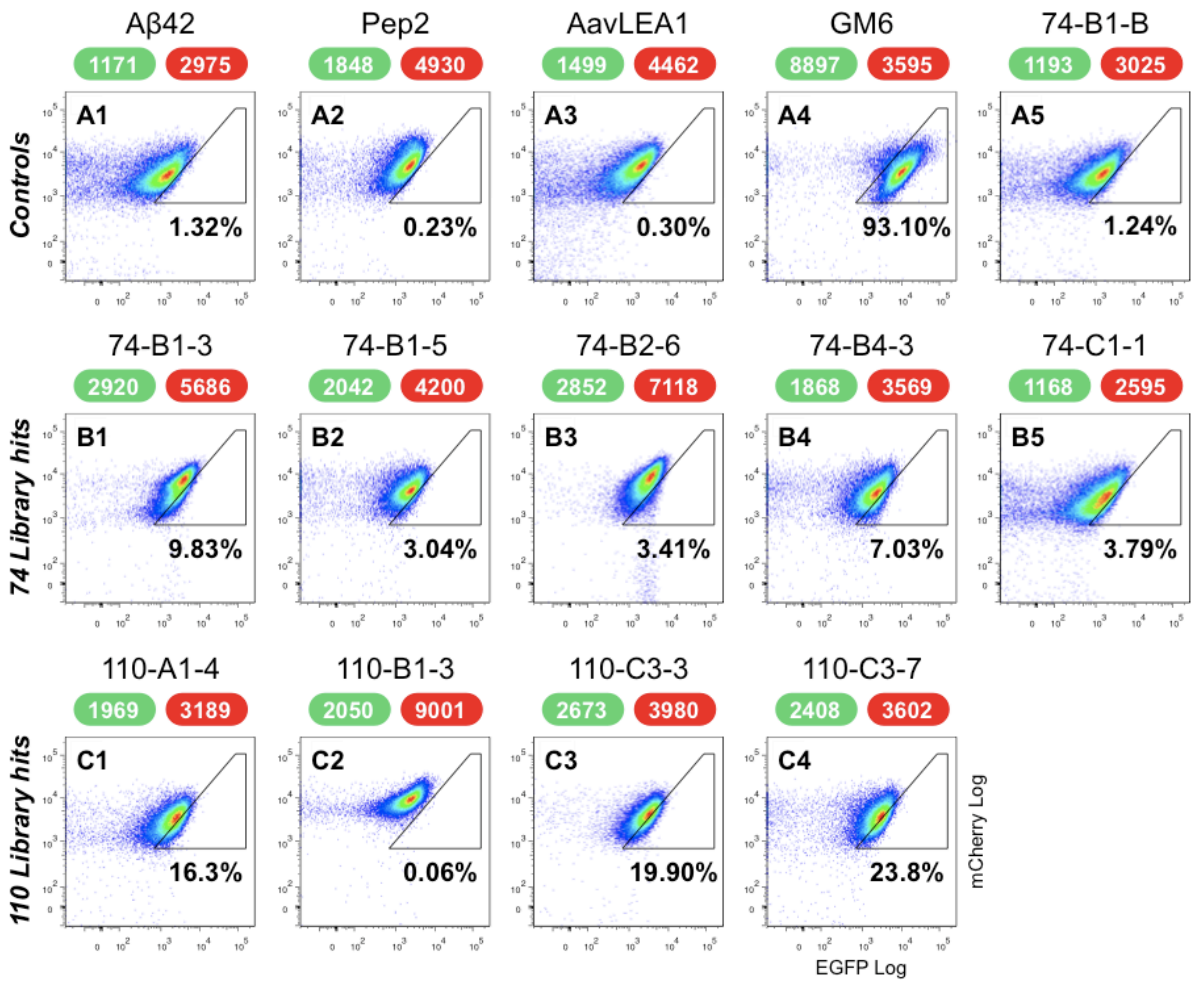


Figure 5.25: Flow cytometry dot plots from *E. coli* expressing pAG2-Aβ42-74 and pAG2-Aβ42-110 library hits and controls. Cultures induced for 14 h at 37°C at ~225 rpm before analysis of EGFP and mCherry fluorescence. 100,000 total events recorded, initial doublet discrimination (data not shown) and Rol covering top ~1% of pAG2-Aβ42 EGFP:mCherry ratio performed as per Figure 5.11. Dot plot densities: blue (highest), red (lowest). Median fluorescence intensity (arbitrary units) shown for whole plot: EGFP (green bullet), mCherry (red bullet). Representative plots shown (n = 3).

A Rol was set for pAG2-Aβ42 as per Section 5.2.2.4 so that the top ~1% of its EGFP:mCherry ratio was selected (panel A1). However, the positive controls pAG2-Aβ42-Pep2/AavLEA1 (panels A2 & A3) did not show any shift along the X-axis as per Figure 5.11. While their raw green MFI values were above that of pAG2-Aβ42, they were less than the two-fold extent that was expected. However, the negative library controls (74-B1-B in panel A5 shown as a representative example) performed similarly to pAG2-Aβ42, although 74-A1-A exhibited a positive shift with 9.39% of cells falling into the Rol (see Appendix 10). With regards to the pAG2-Aβ42-74/110 library hits, all exhibited greater than background Aβ42-EGFP fluorescence, showing that their behaviour in liquid culture is similar to that

observed during their selection on solid medium. An exception to this was 74-C1-1 (panel B5; although panels B2 & B3 show no increase in RoI %, they do exhibit an increase in EGFP output), which was akin to the pAG2-A β 42 control. While no hits exhibited an effect as dramatic as that of the non-aggregating pAG2-GM6 control, 110-A1-4, 110-C3-3 and 110-C3-7 appeared relatively strong (row C). *Figure 5.26* plots the observed MFIs for all of the samples, as well as the EGFP:mCherry ratios for comparison.

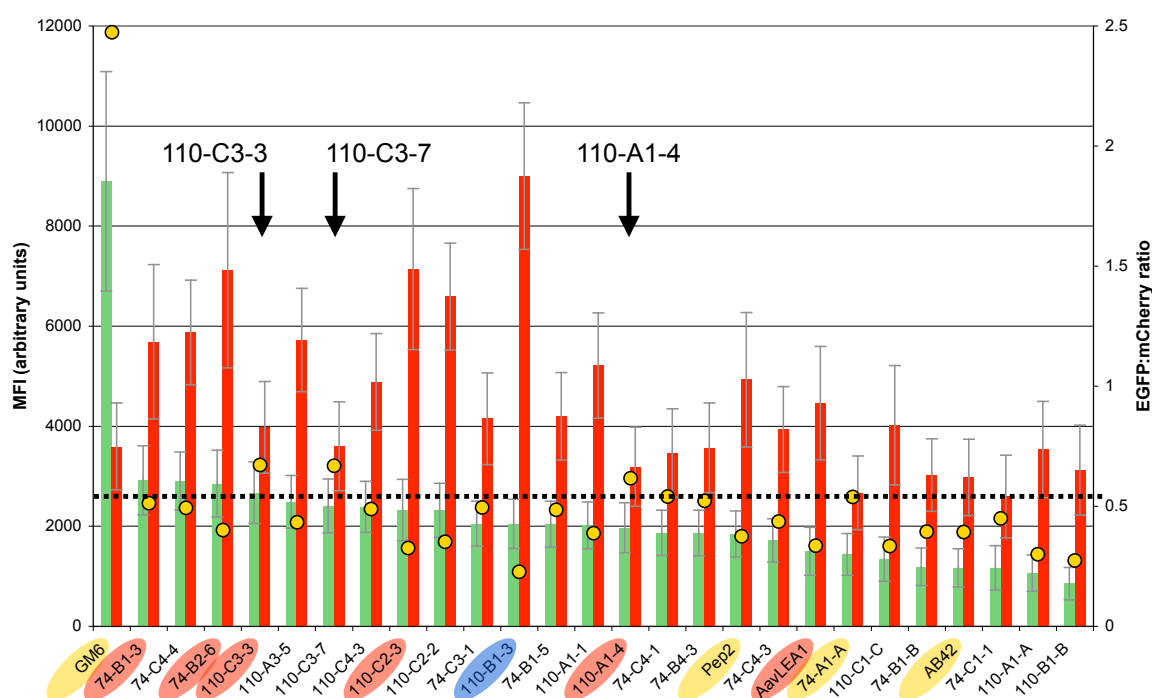


Figure 5.26: MFIs from *E. coli* expressing pAG2-A β 42-74 and pAG2-A β 42-110 library hits and controls. Ranked by EGFP fluorescence, data taken from representative plots shown in Figure 5.25 ($n = 1$). Median fluorescence intensities: EGFP (green bars), mCherry (red bars); EGFP:mCherry ratio (yellow dots). Highlighted labels: selected controls (yellow), clumping cultures (red), example of bright red fluorescence (blue). Error bars represent robust coefficient of variation (see Section 2.7.5.4).

The dotted line in *Figure 5.26* represents a stringency cut-off for potentially significant EGFP:mCherry ratios: while most of the negative controls fall below this, 74-A1-A (highlighted yellow) rivals many of the putative hits. Aside from the non-aggregating pAG2-GM6 controls, only three constructs were seen to exhibit ratios exceeding this high threshold: 110-C3-3, 110-C3-7 and 110-A1-4 (arrowed). Furthermore, as in *Section 5.2.2.4*, mCherry fluorescence was seen to vary widely; for example, there was an approximate 3-fold difference between 110-B1-3

(highlighted blue; *Figure 5.25 C2*) and the pAG2-A β 42 control. In many similar cases, this increase in mCherry fluorescence nullified any gain in A β 42-EGFP output with respect to the ratiometric approach.

A number of constructs (highlighted red) were observed to consistently exhibit cell clumping in liquid culture under both inducing and non-inducing conditions. While similar to that seen with pAG2-A β 42-AavLEA1 (*Section 5.2.2.3*), clumping in these samples was more severe: in the case of 74-B2-6, cells sedimented to the bottom of the incubation vessel, with little turbidity visible in the strata above even during shaking incubation (*Figure 5.27*). However, turbidity in the upper strata was restored after several hours of additional incubation, and cell viability in all cases was unaffected. Immediately before flow cytometry, clumping cultures were thoroughly vortexed to aid dispersion, and the initial gating criteria ensured that only single cells were analysed (see *Section 2.7.5.1*).

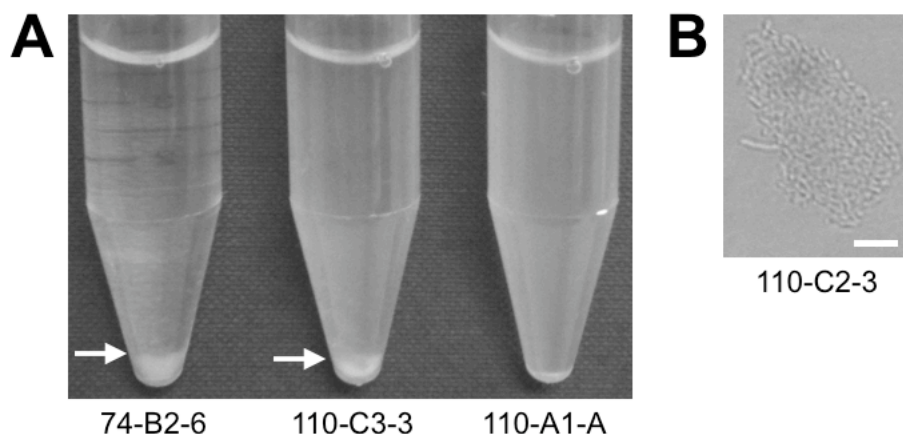


Figure 5.27: Examples of clumping in pAG2-A β 42-74 and pAG2-A β 42-110 cultures. **A:** Photo taken after ~16 h incubation in LB medium (without arabinose) at 37°C at ~225 rpm. Sedimentation arrowed; note the greater clarity of the 74-B2-6 sample. **B:** Photo of clumping cells grown as above, but after 3 h incubation. Scale bar, 5 μ m.

After several repeats, the severity of clumping was ranked as follows: 110-C3-3 > 74-B1-3 > 110-A1-4 > 74-B2-6 > 110-C2-3 > AavLEA1. Interestingly, some of these “clumpers” also exhibited high levels of mCherry fluorescence (see *Figure 5.26*). In an attempt to further normalise the observed mCherry levels, a gate encompassing cells with a narrow range of mCherry MFI was created, and the corresponding EGFP fluorescence levels analysed (*Figure 5.28*).

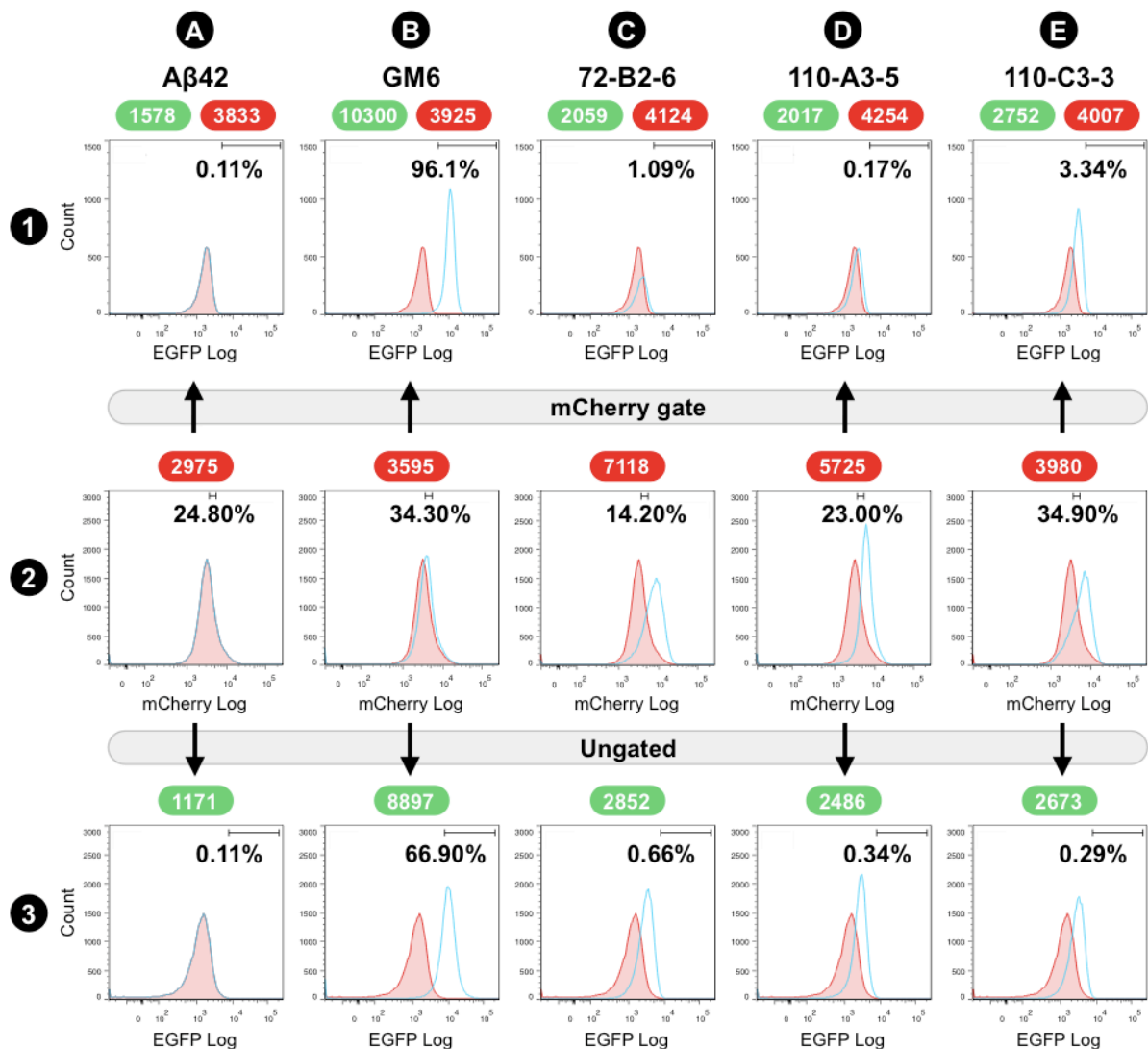


Figure 5.28: Histograms of EGFP fluorescence with/without gating a narrow segment of mCherry fluorescence. Data taken from representative plots shown in Figure 5.25 ($n = 1$), with rows 2 & 3 giving the mCherry (red bullet) and EGFP (green bullet) median fluorescence intensities (arbitrary units) respectively. mCherry fluorescence gated as per narrow band indicated in row 2, corresponding EGFP fluorescence shown in row 1. Histograms (blue) overlaid on that of pAG2-A β 42 (pink) for comparison.

For comparison, the selected histograms are shown overlaid on that of pAG2-A β 42; in rows 1 (mCherry gate) and 3 (ungated), a RoI is also indicated to highlight any comparative increase in EGFP fluorescence. While the non-aggregating pAG2-GM6 control exhibited a positive shift in EGFP fluorescence no matter if it was gated (B1) or not (B3), putative hits such as 72-B2-6 and 110-A3-5 revealed a commensurate drop in EGFP output when their increased mCherry levels were normalised to pAG2-A β 42 (C1 & D1). 110-C3-3, on the other hand, retained a slight increase in EGFP output (E1). Figure 5.29 plots the mCherry-normalised MFIs for all of the putative hits and controls, as well as their adjusted EGFP:mCherry ratios.

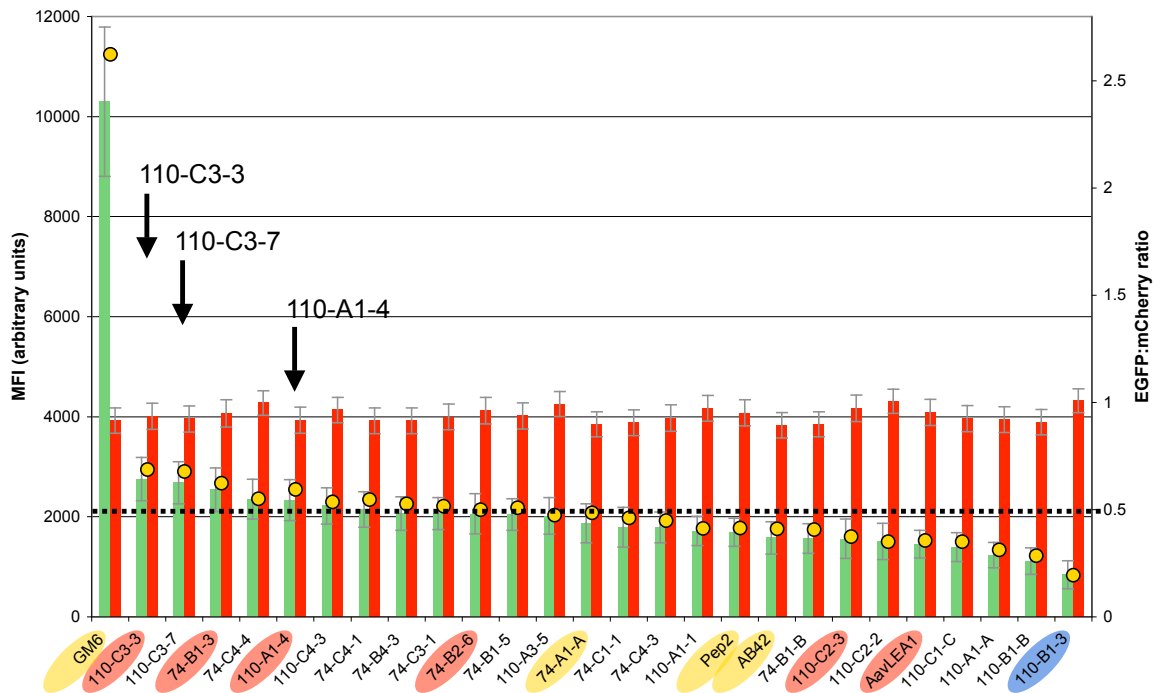


Figure 5.29: MFIs from *E. coli* expressing pAG2-A β 42-74 and pAG2-A β 42-110 library hits and controls: gated to a small segment of mCherry fluorescence. Ranked by EGFP fluorescence, data taken from representative plots shown in Figure 5.25 ($n = 1$). Median fluorescence intensities: EGFP (green bars), mCherry (red bars); EGFP:mCherry ratio (yellow dots). Highlighted labels: selected controls (yellow), clumping cultures (red), example of bright red fluorescence from Figure 5.26 (blue). Error bars represent robust coefficient of variation (see Section 2.7.5.4).

Using the same stringency cut-off as Figure 5.26 (dotted line; the EGFP:mCherry ratio of the negative control 74-A1-A), the same three hits of 110-C3-3, 110-C3-7 and 110-A1-4 were again seen to outperform EGFP fluorescence from the pAG2-A β 42 control. This indicates that assessing the EGFP:mCherry ratio in the first instance is adequate to judge the potential of a hit without resorting to additional gating criteria. Interestingly, constructs that exhibited cell clumping (highlighted in red) showed a relative increase in their normalised fluorescence ratio that correlated with an increased severity of clumping (ranked beneath Figure 5.27). pAG2-A β 42-Pep2/AavLEA1, on the other hand, showed negligible improvement in EGFP output, with their ratios remaining similar to that of pAG2-A β 42 alone. As alluded to earlier in Section 5.2.2.4, this variance between different experiments may make isolating true antiaggregant hits difficult. To try to achieve a consensus, the plasmids from each hit were isolated, retransformed into fresh *E. coli* DH10B, and subsequently induced and analysed by flow cytometry as per Section 2.7.5.1

separately on three consecutive days using the same parameters. *Figure 5.30* plots the mean of the MFIs observed, as well as their EGFP:mCherry ratios.

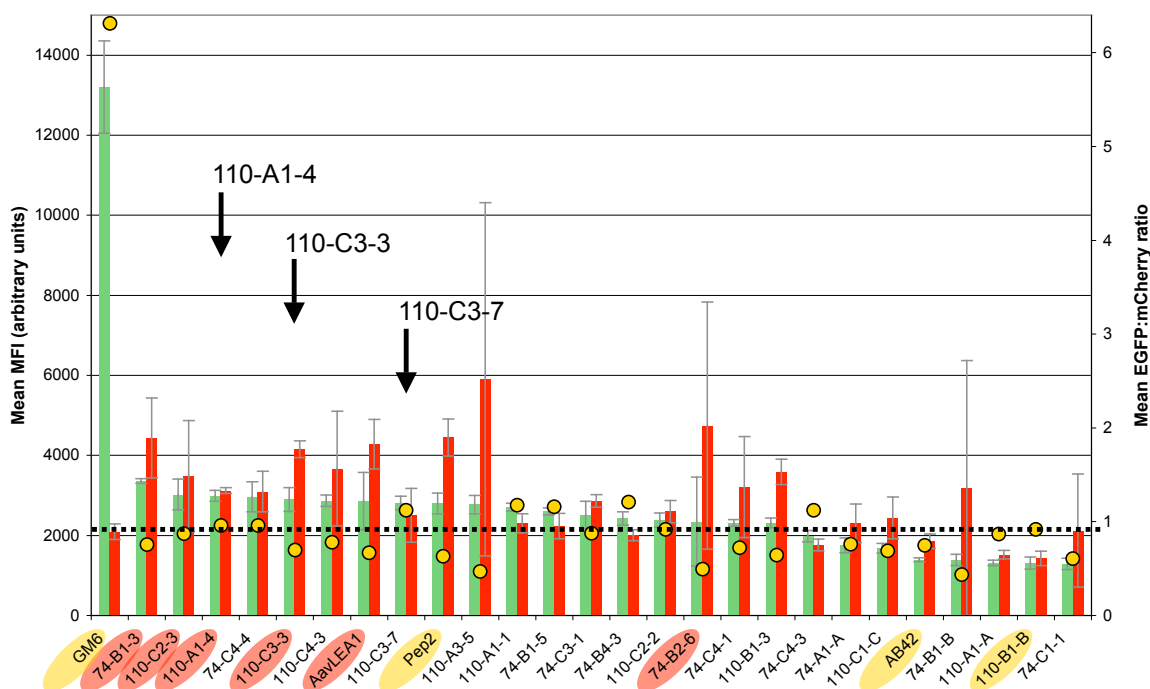


Figure 5.30: Mean MFIs from three repeats of *E. coli* expressing pAG2-A β 42-74 and pAG2-A β 42-110 library hits and controls. Cultures induced and analysed as per Figure 5.25, ranked by EGFP fluorescence. Mean fluorescence intensities: EGFP (green bars), mCherry (red bars); mean EGFP:mCherry ratio (yellow dots). Highlighted labels: selected controls (yellow), clumping cultures (red). Error bars represent 1 standard deviation ($n = 3$).

Despite being carried out under the same conditions, mCherry levels were seen to fluctuate greatly in some samples (such as 110-A3-5, 74-B2-6 and 74-B1-B) over the three repeats. When the stringency cut-off (dotted line) is set to the negative control exhibiting the highest mean EGFP:mCherry ratio (110-B1-B, highlighted yellow), several samples were seen to outperform, including two of those identified initially in *Figure 5.26*. While reproducibility seems to be an issue with this assay, hit trends can be detected.

5.2.4.4 Fluorescence microscopy of library hits

During the conduction of flow cytometry for *Figure 5.30*, *E. coli* expressing pAG2-A β 42-74/110 library hits were inspected using fluorescence microscopy (see *Section 2.7.4*). A selection of these are shown in *Figure 5.31*.

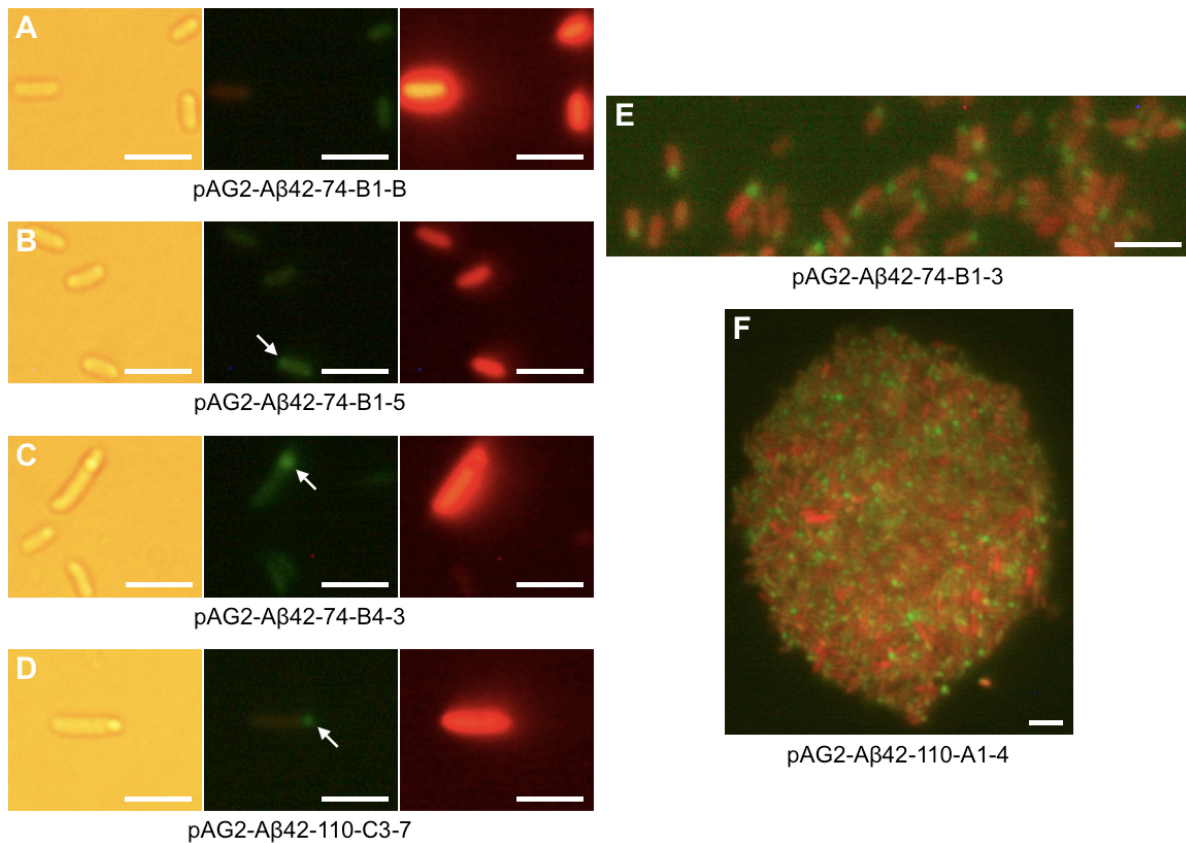


Figure 5.31: *E. coli* expressing pAG2-A β 42 library hits possess green fluorescent aggregates. For panels A to D, from left to right: bright field, green fluorescence, red fluorescence. Panels E & F show green fluorescence filter only, exposure adjusted to allow mCherry bleed-through (see Section 2.7.4). Photos taken after 14 h arabinose induction (0.05%) in LB medium at 37°C at ~225 rpm, then several hours at ~23°C. Green aggregates arrowed (panels B, C & D). Representative photos shown ($n = 3$); scale bar, 5 μ m.

Many cells (but by no means the majority) in each hit culture were observed to possess a single, punctate green fluorescent aggregate at one pole, as observed earlier during induction of the pAG2-A β 42-Pep2 and pAG2-A β 42-AavLEA1 controls in *Figures 5.8 & 5.18*. This is best demonstrated in panel E, while panel F serves to show the same in an example of clumping cells (110-A1-4). Furthermore, cells exhibiting brighter mCherry fluorescence appeared more likely to contain these aggregates (e.g. panels C & D). For the negative controls, however, this was not seen to be the case (e.g. panel A).

5.2.4.5 Bioinformatic analysis of library hits

The putative antiaggregant hits from colony screening (*Section 5.2.4.2*) had their DNA insert sequenced and analysed as per *Section 2.7.7*. *Table 5.4* summarises the findings; for full DNA insert sequences, see *Appendix 9*.

| Name | bp | GC | Peptide sequence (# aa) | Composition | Q | H | Hyphb. | Culture notes |
|---------|-----|-----|---|-------------------------------------|----|-------|--------|--|
| Aβ42 | 126 | 50% | DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIA (42) | Gly/Val (14%) Ala (10%) | 0 | -0.01 | 38% | |
| GM6 | 126 | 52% | DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGPMVGGVVIA (42) | Gly/Val (14%) Ala (10%) | 0 | -0.05 | 33% | |
| Pep2 | 42 | 46% | QKLDVVAEDAGSNK (14) | Asp/Gly/Lys/Val (14%) | -1 | -0.25 | 21% | Bright red pellet |
| AavLEA1 | 426 | 63% | SSQQNQNRQGEQQEQGYMEAAKEKVVNAWESTKETLSSTAQAAAETAEF RDSAGETIRDLTGQAQEKGEFKERAGEKAEETKQRAGEKMDETKQRAGE MRENAGQKMEEYKQQGGKAEELRDTAAEKLHQAGEKVKGRD (142) | Glu (18%) Ala (13%) Lys (12%) | -3 | -0.40 | 12% | Minor cell clumping, bright red pellet |
| 74-B1-3 | 36 | 57% | PGKRAKRTGRW (11) | Arg (27%) Gly/Lys (18%) | +5 | -0.61 | 9% | Cell clumping, bright red pellet |
| 74-B1-5 | 36 | 64% | TPGGWIPLQRNG (12) | Gly (25%) Pro (17%) | +1 | -0.10 | 25% | |
| 74-B2-6 | 36 | 71% | AQGRGRDCSDDR (12) | Gly (33%) Arg (25%) | +2 | -0.50 | 0% | Cell clumping, bright red pellet |
| 74-B4-3 | 36 | 69% | RVGGGPAG (8) <i>No other potential ORFs</i> | Gly (44%) | +1 | -0.05 | 12% | |
| 74-C1-1 | 36 | 62% | GLH (3) <i>No other potential ORFs</i> | - | +1 | 0.09 | 33% | |
| 74-C3-1 | 53 | 55% | EPGKGPNDFGKH (12) | Gly (25%) Lys/Pro (17%) | +1 | -0.30 | 8% | |
| 74-C4-1 | 36 | 69% | ARKPVGRRATRG (12) | Arg (33%) Ala/Gly (17%) | +5 | -0.58 | 8% | |
| 74-C4-3 | 36 | 67% | VAGAGGNSQAP (12) | Gly (33%) Ala (25%) | 0 | 0.02 | 8% | |
| 74-C4-4 | 52 | 60% | CRRDQARHQRG (12) | Arg (42%) Gln (16%) | +5 | -0.90 | 0% | |
| 74-A1-A | 97 | 62% | SRNVGGRGLRRR (12) | Arg (42%) Gly (25%) | +5 | -0.67 | 17% | Negative control |
| 74-B1-B | 36 | 67% | VHLPRGLVVRER (12) | Arg/Val (25%) Leu (17%) | +3 | -0.29 | 42% | Negative control |

| Name | bp | GC | Peptide sequence (# aa) | Composition | Q | H | Hyphb. | Culture notes |
|----------|-----|-----|--|--|------------------------------|------------------------------------|--------------------------------|----------------------------------|
| 110-A1-1 | 71 | 57% | LACGRCR (7) <i>Alternative ORF: ATG 15 bp from RBS</i> WSVQVMDRMCNGNTDPQNGG (21) | Arg/Cys (29%) <i>Alt. ORF:</i> Gly (19%) | +2 <i>Alt. ORF:</i> -1 | -0.35 <i>Alt. ORF:</i> -0.18 | 14% <i>Alt. ORF:</i> 24% | |
| 110-A1-4 | 88 | 60% | RMCNSVGRSVVWVSGRVAQRTSGR (24) | Arg (21%) Gly/Ser/Val (17%) | +5 | -0.31 | 25% | Cell clumping, bright red pellet |
| 110-A3-5 | 167 | 62% | AV (2) <i>Alternative ORF: ATG 8 bp from putative RBS</i> GHGAGRKM <u>MKR</u> (11) | - <i>Alt. ORF:</i> Gly (27%) | 0 <i>Alt. ORF:</i> +5 | 0.39 <i>Alt. ORF:</i> -0.43 | 50% <i>Alt. ORF:</i> 18% | Bright red pellet |
| 110-B1-3 | 78 | 69% | DAKRG (5) <i>Alternative ORF: ATG 11 bp from RBS</i> LSVDSSPAGVGMGGRRAGARHGRGE <u>A</u> (25) | <i>Alt. ORF:</i> Gly (27%) Arg/Lys (18%) | +1 <i>Alt. ORF:</i> +2 | -0.63 <i>Alt. ORF:</i> -0.15 | 0% <i>Alt. ORF:</i> 16% | Bright red pellet |
| 110-C2-2 | 72 | 51% | RDTELWQYKDGCRCCVVGWDRYS (24) | Arg/Asp/Gly/Val (13%) | 0 | -0.28 | 33% | |
| 110-C2-3 | 72 | 62% | (0) <i>Alternative ORF: GTG 9 bp from RBS</i> RRFGVVSLRGREKRALRGGRGDV <u>MKR</u> (26) | <i>Alt. ORF:</i> Arg (31%) Gly (19%) | <i>Alt. ORF:</i> +8 | <i>Alt. ORF:</i> -0.51 | <i>Alt. ORF:</i> 27% | Cell clumping, bright red pellet |
| 110-C3-3 | 72 | 55% | LDRGLVDREYSRRGQAFGQLCARG (24) | Arg (21%) Gly (17%) Leu (13%) | +2 | -0.35 | 25% | Cell clumping, bright red pellet |
| 110-C3-7 | 72 | 65% | GQGCRMVNDPHKRARGRRRGSQG (24) | Arg/Gly (25%) | +7 | -0.54 | 8% | |
| 110-C4-3 | 72 | 46% | RPGSVDSSEQTSVLLKSSFRVYKDS (24) | Ser (25%) Val (13%) | +1 | -0.28 | 29% | |
| 110-A1-A | 72 | 49% | GE (2) <i>No other potential ORFs</i> | - | -1 | -0.23 | 0% | Negative control |
| 110-B1-B | 72 | 55% | GETTSVRTA (9) | Thr (33%) | 0 | -0.24 | 11% | Negative control |
| 110-C1-C | 72 | 54% | SAPCIMSGRCGLVDSVNFRRSRST (24) | Ser (21%) Arg/Gly (13%) | +2 | -0.17 | 25% | Negative control |

Table 5.4: Peptide sequence and properties of pAG2-A β 42 random DNA library hits. Name, 74 (36 bp random DNA library) and 110 (72 bp random DNA library); bp, actual length of DNA insert (excluding start and stop codons); GC, content % of random DNA insert; ORF, open reading frame; RBS, ribosome binding site; Composition, % of predominant amino acids; Q, charge; H, mean hydrophobicity derived from Eisenberg's consensus hydrophobicity scale (1984); Hyphb., % of hydrophobic residues (F, I, L, N, V, W & Y); Culture notes, anything significant noted during the screening process. Amino acids encoded by the plasmid backbone, as opposed to the random DNA insert, are underlined for clarity.

Insert DNA lengths between the start and stop codons were mostly as expected: 36 bp for the 74 library, and 72 bp for the 110 library (see *Section 5.2.3.1*). Other lengths were also seen, e.g. double insertion events in 74-A1-A (97 bp) and 110-A3-5 (167 bp). The GC bias of the DNA of the 18 sequenced putative hits, at an average of 61%, was relatively high. This may have been a consequence of the amino acids favoured for antiaggregant properties by the screen. For example, arginine and glycine were particularly prominent in all hit peptides, and the corresponding codons for these amino acids are GC rich. However, inserts from the negative controls revealed a similar average GC content (~62%). PCR amplification during library generation may have introduced such a bias; or, given the difficulty of synthesising the random oligonucleotides (see *Section 5.2.3.1*), a high GC content could have been favoured from the beginning.

The desired peptide lengths of 12 aa and 24 aa (for the 74 and 110 libraries respectively, discounting the initiating methionine [Ben-Bassat et al., 1987]) were not always achieved. As the length of random base pairs increased, so did the probability of an internal stop codon being incorporated: for the 74 library, ~44% of peptides were predicted to possess an internal stop codon, while for the 110 library the probability rose to ~68% (see *Section 2.7.7* for calculation). However, the actual number of peptides incorporating internal stop codons was seen to be less than predicted: ~27% for the 74 library and ~50% for the 110 library respectively, most likely due to the high GC sequence bias observed (stop codons being TA rich). As a result of internal stop codons, a number of peptides were predicted to encode very short peptides, potentially representing false positives. For instance, 74-C1-1 only encoded 3 aa; this most likely explains why its fluorescence output was similar to that of the pAG2-A β 42 control when analysed by flow cytometry (see *Section 5.2.4.3*). However, for a number of truncated sequences (e.g. 110-C2-3), an alternative ORF could be identified (see *Appendix 9*). This potentially allows for a longer peptide product to be produced.

In general, the putative hit peptides were cationic (average charge of +3) and largely hydrophilic (mean hydrophobicity of -0.33; average hydrophobic residue content of 17%). This differed little from the negative controls (average charge of +3, mean hydrophobicity of -0.34; average hydrophobic residue content of 24%).

None of these properties correlated with an increased effect of a peptide on A β 42-EGFP fluorescence output *in vivo* (see *Section 5.2.4.3*). Additionally, while all culture samples had pink coloured pellets when centrifuged (due to mCherry expression), a number exhibited a noticeably brighter shade of red. These observations unsurprisingly correlated with high mCherry fluorescence levels observed during flow cytometry (e.g. 74-B2-6 & 110-A3-5).

In order to try to identify potential β -breaker hits (see *Section 5.1.3.3*), the putative hit peptides and negative controls were aligned pairwise with A β 42 (*Figure 5.32*). Although average hit peptide hydrophobicity was low (see above), sequences that might potentially complement the hydrophobic β -fold regions of A β 42 (17-21 and 30-42) were focused on.

| <i>Putative antiaggregants</i> | | | | <i>Positive control</i> | | | |
|--------------------------------|----|---------------|----|-------------------------|----|----------------|----|
| AB42 | 21 | AEDVG | 25 | AB42 | 15 | QKLVFFAEDVGSNK | 28 |
| | | :.. | | ● | | | |
| 74-B1-3 | 5 | AKRTG | 9 | Pep2 | 1 | QKLDVVAEDAGSNK | 14 |
| AB42 | 29 | GAIIGLMVGG | 38 | AB42 | 34 | LMVG | 37 |
| | | | | | | .. | |
| 74-B1-5 | 3 | GGWIPLQRNG | 12 | 110-A1-1 | 1 | LACG | 4 |
| AB42 | 21 | AEDVG | 25 | AB42 | 29 | GAIIGLMV | 36 |
| | | :.. | | | | | |
| 74-B2-6 | 1 | AQGRG | 5 | 110-A1-4 | 4 | GNSVGRSV | 11 |
| AB42 | 36 | VGG | 38 | AB42 | 31 | IIGLMVG | 37 |
| | | | | ● | | :... . | |
| 74-B4-3 | 2 | VGG | 4 | 110-A1-1-altO | 5 | VMDRMCG | 11 |
| AB42 | 23 | DVGSN | 27 | AB42 | 6 | HDSG | 9 |
| | | .. : | | | | .. | |
| 74-C3-1 | 8 | DFGKH | 12 | 110-A3-5-altO | 2 | HGAG | 5 |
| AB42 | 36 | VG | 37 | AB42 | 26 | SNKG | 29 |
| | | | | | | :... | |
| 74-C4-1 | 5 | VG | 6 | 110-B1-3 | 2 | AKRG | 5 |
| AB42 | 25 | GSNKGGA | 30 | AB42 | 33 | GLMVGG | 38 |
| | | .. . | | ● | | .. . | |
| 74-C4-3 | 6 | GGNSQA | 11 | 110-B1-3-altO | 9 | GVGMGG | 14 |
| AB42 | 1 | DAEFRHDSG | 9 | AB42 | 31 | IIGL | 34 |
| | | | | ● | | :... | |
| 74-C4-4 | 4 | DQRARHQRG | 12 | 110-C2-2 | 15 | VVGV | 18 |
| AB42 | 36 | VGG | 38 | AB42 | 25 | GSNKGAIIG | 33 |
| | | | | ● | | | |
| 74-A1-A | 4 | VGG | 6 | 110-C2-3-altO | 10 | GREKRALRG | 18 |
| AB42 | 33 | GLMV | 36 | AB42 | 26 | SNKGAIIG | 33 |
| | | : | | ● | | | |
| 74-B1-B | 6 | GLVV | 9 | 110-C3-3 | 11 | SRRGQAFG | 18 |
| AB42 | 29 | G | 29 | AB42 | 14 | HQK | 16 |
| | | | | | | :.. | |
| 110-B1-B | 1 | G | 1 | 110-C3-7 | 11 | HKR | 13 |
| AB42 | 24 | VGSNKGAIIG | 32 | AB42 | 8 | SGYEVH | 13 |
| | | : :..:..:..: | | ● | | .. .: | |
| 110-C1-C | 6 | MGSGRGCVL | 14 | 110-C4-3 | 16 | SSFRVY | 21 |
| | | | | | | .. .: | |

Negative controls

Figure 5.32: Pairwise alignment of putative antiaggregants to A β 42 reveals little homology. Residue positions indicated. Green dots represent hydrophobic region hits, while blue dots represent putative β -turn region hits. Colon, indicates similarity between side-chains; altO, alternative ORF. EMBOSS Matcher used to generate alignments (see Section 2.7.7).

While the putative antiaggregants contained plenty of cationic residues that may act as β -fold disrupter elements (Soto & Estrada, 2005), there were few obvious hydrophobic regions of homology (green dots) to act as recognition elements between these hits and the A β 42 sequence. Indeed, the negative control 74-B1-B showed as much homology to the C-terminal hydrophobic region of A β 42 as did the hit 110-C2-2. Other peptides (blue dots), such as 110-C3-3, showed homology with

the putative β -turn region between residues 25-29 (DaSilva et al., 2010). These could potentially interfere with intermolecular β -sheet hydrogen bonding and thus aggregation, as disruption of a salt bridge formed between Asp23 and Lys28 of adjacent A β peptides has been shown to ameliorate fibril formation (Lührs et al., 2005). Alternatively to acting as traditional β -breakers, it may be that the hit peptides interfere with cross- β -sheet interactions between A β 42-EGFP protofilaments (see *Section 5.1.2.1*). Sato et al. (2006b) showed that the IxGxMxG motif, found at the C-terminus of A β 42, formed “glycine grooves” running along the axis of one outer face of a protofilament. The deployment of peptides incorporating a complementary GxFxGxF motif bound these grooves, and in combination with hydrophilic residues at the “x” positions, inhibited mature fibril formation. However, examination of the putative antiaggregant hits in *Table 5.4* reveals that only 110-A1-4 and 110-C3-3 possess something similar to this motif ($_{10}\text{SxWxSxR}_{16}$ and $_{18}\text{GxLxAxG}_{24}$ respectively). Although these hits exhibited an increase in A β 42-EGFP output during flow cytometry trials (*Section 5.2.4.3*), as well as exhibiting green fluorescent aggregates (*Section 5.2.4.4*), it is unlikely that the above may be their mode of operation; if A β 42-EGFP is able to aggregate into protofibrils before any therapeutic effect takes place, EGFP would not be expected to fold correctly and hence fluoresce.

Next, in order to try to identify any shared regions of homology between the hit peptides themselves, their sequences were aligned (*Table 5.5*).

| | | |
|-----------------|--|----|
| 74-B1-3 | -----PG-----KRAKRTGRW----- | 11 |
| 110-C2-3-altORF | --RRFGVVSLRGRE--KRALRGGRGDVMKR----- | 26 |
| 74-C4-1 | -----ARKPVG-----RRATR-G----- | 12 |
| 110-B1-3 | -----DAKR-G----- | 5 |
| 74-C4-4 | --CRRDQ-----RA--RHQRG----- | 12 |
| 110-B1-3-altORF | --LSVDSSPAGVGMGGRAGARHGRGEA----- | 25 |
| 110-A1-4 | -----RMCG-----NSVGRSVVWVSGRVAQRTSGR | 24 |
| 110-A3-5-altORF | -----G-----HGAGR-KWMK-R----- | 11 |
| 74-B2-6 | -----AQQGRGRDCS GGR----- | 12 |
| 110-C3-7 | GQGCRMVNDPHK-----RARGRGRRGSQG----- | 24 |
| 110-A1-1 | -----LACGRCR----- | 7 |
| 74-B4-3 | -----R--VGGGPAG----- | 8 |
| 110-C2-2 | -----RDTEWQYKDGCR CVVGVGWD RYS----- | 24 |
| 74-C4-3 | -----VAGAGGGNSQAP----- | 12 |
| 74-C3-1 | -----EPG-----KGNDFGKH----- | 12 |
| 110-A1-1-altORF | --WSVQVM DRMCG-----NTTDPQGNGG----- | 21 |
| 110-C3-3 | --LDRGLVDREYS-----RRGQAFGQLCARG----- | 24 |
| 74-B1-5 | -----TPGG-----WIPLQRNG----- | 12 |
| 110-C4-3 | --RPGSV DSEQTS-----VLLKSSF RVYKDS----- | 24 |

Table 5.5: Sequence alignment of putative antiaggregant peptides. Sequence length indicated. Amino acids are colour coded: cationic (pink), hydrophobic (red), anionic (blue) and uncharged hydrophilic (green). altORF, alternative ORF. ClustalW2 used to generate alignment (see Section 2.7.7).

While no obvious motif appears common between all of the hit peptides, some more local alignments can be discerned. At the top of *Table 5.5*, a good level of homology can be seen between the peptides 74-B1-3 and 110-C2-3-altORF; these two constructs (the second harnessing a putative alternative ORF; see *Table 5.4*) show similar levels of fluorescence in *E. coli* (*Figure 5.30*, left-most on X-axis), and both exhibit the cell clumping phenotype. 74-C4-1 is also somewhat homologous to these peptides, but does not display the clumping phenomenon. Peptides 74-B2-6, 110-C3-7 and 110-A1-1 (middle of table) also share a region of homology, but do not have similar fluorescence or phenotypic profiles.

Lastly, hit constructs seen to possess the cell clumping phenotype (*Table 5.4*) were aligned in isolation in a further attempt to find a common motif (*Table 5.6*).

| | | |
|-----------------|--|----|
| 74-B2-6 | -----AQQR--GRDCS GGR----- | 12 |
| 110-C3-3 | LDRG-LVDREYSRRGQAFGQLCARG----- | 24 |
| 74-B1-3 | -----PG-----KRAKRTGRW----- | 11 |
| 110-C2-3-altORF | RRFG-VV---SLRGRE--KRALRGGRGDVMKR----- | 26 |
| 110-A1-4 | RMCGNSVGRSVVWVSGR--VAQRTSGR----- | 24 |
| | * * | |

Table 5.6: Sequence alignment of peptides from clumping constructs. Sequence length indicated. Amino acids are colour coded: cationic (pink), hydrophobic (red), anionic (blue) and uncharged hydrophilic (green). Star, indicates side-chain match; altORF, alternative ORF. ClustalW2 used to generate alignment (see Section 2.7.7).

Aside from 74-B1-3 and the alternative ORF of 110-C2-3 (mentioned above), no other regions of obvious homology were found. While *Table 5.4* revealed that all “clumping” peptides contained a high proportion of arginine and glycine, e.g. 74-B1-3 (Arg 27%, Gly 18%) and 110-C3-3 (Arg 21%, Gly 17%), so do many other “non-clumping” peptides, e.g. 74-C4-1 (Arg 33%, Gly 17%) and 110-C3-7 (Arg 25%, Gly 25%). It is thus hard to find a common denominator that links these peptides.

5.2.4.6 Selection of putative antiaggregant peptides for *in vitro* analysis

Several putative antiaggregant peptides were selected to be chemically-synthesised based on their performance *in vivo* as assessed by flow cytometry (*Section 5.2.4.3*). A range of peptides was chosen (*Table 5.7*), including those whose constructs exhibited cell clumping. It was envisaged that these peptides would be assessed *in vitro* for their ability to influence A β 42 aggregation via the ThT assay (LeVine, 1993) as per Baine et al. (2009), with Pep2 serving as a positive control. However, time constraints did not permit this.

| Name | Peptide (# aa) | Composition | Q | H | Hyphb. | Culture notes |
|----------|-------------------------------|--------------------------------|----|-------|--------|-------------------------------------|
| Pep2 | QKLDVVAEDAGSNK (14) | Asp/Gly/Lys/Val (14%) | -1 | -0.25 | 21% | Bright red pellet |
| 74-B1-3 | PGKRAKRTGRW (11) | Arg (27%) Gly/Lys (18%) | +5 | -0.61 | 9% | Cell clumping, bright red pellet |
| 74-B1-5 | TPGGWIPLQRNG (12) | Gly (25%) Pro (17%) | +1 | -0.10 | 25% | |
| 74-B2-6 | AQGRGRDCSGGR (12) | Gly (33%) Arg (25%) | +2 | -0.50 | 0% | Cell clumping, bright red pellet |
| 74-B4-3 | RVGGGPAG (8) | Gly (44%) | +1 | -0.05 | 12% | |
| 110-A1-4 | RMCGNSVGRSVWVSGRVAQRTSGR (24) | Arg (21%) Gly/Ser/Val (17%) | +5 | -0.31 | 25% | Cell clumping, bright red pellet |
| 110-C3-7 | GQGCRMVNDPHKRARGRRGSQG (24) | Arg/Gly (25%) | +7 | -0.54 | 8% | |
| 74-B1-B | VHLPRGLVVRER (12) | Arg/Val (25%) Leu (17%) | +3 | -0.29 | 42% | Negative control |

Table 5.7: Putative antiaggregant peptides selected for *in vitro* characterisation. Composition, % of predominant amino acids; Q, charge; H, mean hydrophobicity derived from Eisenberg’s consensus hydrophobicity scale (1984); Hyphb., % of hydrophobic residues (F, I, L, N, V, W & Y); Culture notes, anything significant noted during the screening process.

Collaborators in the laboratory of Prof Yizhi Zheng (College of Life Science, Shenzhen University, China) are currently pursuing these experiments. Preliminary results indicate that the peptides 110-A1-4 and 110-C3-7 have a moderate effect on

the rate of A β 42 aggregation, similar to that of the positive control epigallocatechin-3-gallate (EGCG; a polyphenol from green tea) (Ehrnhoefer et al., 2008). The remaining peptides, including Pep2, appear to have no effect.

5.3 Discussion

High-throughput screens that dispense with the need for expensive and cumbersome synthetic A β peptide are desired to aid in the isolation of novel antiaggregants (Lansbury, 2001). The use of an *in vivo* recombinant fluorescent reporter for the aggregation state of A β , first outlined by Wurth et al. (2002), has been coupled in this work to the concurrent constitutive production of peptides to screen for those possessing novel antiaggregant activity. In an attempt to introduce greater robustness, mCherry was also co-expressed as an internal fluorescence standard to allow for ratiometric comparison between samples.

One of the key advantages of the pAG2-A β 42 screen is that, as A β expression is inducible, inhibitors of the earliest events of A β aggregation can potentially be identified: this renders moot what particular oligomeric state is most toxic (Carter et al., 2010). However, given that a heterologous system to the human brain is being used, the clinical activity of any peptide hits identified is putative. Although no absolute prevention of A β 42-EGFP aggregation was observed in *E. coli* (cf. the non-aggregating GM6-EGFP control), this was in line with previous screens that used a similar reporter system (Kim et al., 2006; Baine et al., 2009). A ~2-fold increase in A β 42-EGFP fluorescence is the best previously reported effect for a peptide-based inhibitor, i.e. Pep2 (Baine et al., 2009), as opposed to a ~6-fold increase observed with GM6-EGFP. In this work, several putative antiaggregant hits, such as 110-A1-4 and 110-C3-7, were seen to exhibit similar if not greater increases in A β 42-EGFP output in comparison to Pep2. Despite the internal mCherry standard output fluctuating in some samples, these peptides exhibited consistently higher EGFP:mCherry ratios than the negative controls. In addition, the desiccation-implicated “molecular shield” AavLEA1 (Goyal et al., 2005a) also showed a similar effect to Pep2. While not a peptide, the *in vitro* activity of AavLEA1 against A β 42 would be interesting to ascertain.

Together with the concurrent observation of punctate green fluorescent aggregates in *E. coli* host cells, modest increases in A β 42-EGFP output indicated that at least a reduction in the rate of A β 42 aggregation was achievable *in vivo*. It was initially hoped that diffuse green fluorescence could be restored to the *E. coli* host, i.e. like that observed for the non-aggregating GM6-EGFP control. However, the observation of green fluorescent aggregates, both in the Pep2 and AavLEA1 controls and several hits, indicates that A β 42-EGFP aggregation is only inhibited sufficiently enough to allow EGFP to fold in some cells. Once folded, the chromophore is stable; subsequent aggregation still results in fluorescent particles (Waldo et al., 1999). Similar patterns of discrete aggregation to the work presented here are also observed with stress-denatured proteins (Zietkiewicz & Liberek, 2010), which often localise at the cell poles of a bacterium. An explanation for this localisation is that it partitions dead-end and potentially toxic products away from the critical nucleoid region of a bacterium (Winkler et al., 2010), as well as allowing for asymmetric inheritance of an aggregate burden during cell division to a single daughter cell (Treusch et al., 2009). Overall, this increases the fitness of the bacterial population as a whole. Non-inhibited A β 42-EGFP is also expected to aggregate and localise in the same manner; however, a lack of fluorescence prevents direct visualisation. It would be interesting to assess the structure of both inhibited/uninhibited A β 42-EGFP at the nanometre scale using electron microscopy, as observations on the actual physical state (rather than implied) of such A β fusions have yet to be reported.

The use of FACS was originally envisaged to conduct the antiaggregant peptide library screen, but unfortunately proved unmanageable. Three factors were primarily responsible for this. Firstly, the low dynamic range between background negative cells and the positive control (\sim 2-fold at best for Pep2) led to poor gating of rare positive events; multiple rounds of FACS (six-plus) would be required to enrich a single hit from a small (\sim 1x10⁴) library. Secondly, while the above is an inconvenient but not insurmountable obstacle, the pAG2-A β 42 plasmid exhibited partial instability, resulting in a minute but incurable background population of negative background constructs, as well as rare instances of constitutive green fluorescence that could be inappropriately enriched during FACS. To remediate

this, redesign of the pAG2-A β 42 vector is required to ensure that the operational sequences used are sufficiently heterologous to discourage recombination. Lastly, *E. coli* containing inserts in the pAG2-A β 42 Library site exhibited a slower entry into exponential growth compared to “empty” pAG2-A β 42, potentially leading to any positive hits being out-competed in culture by negative background cells. If all of these problems these could be eliminated the FACS approach may become viable. Instead, the approach of directly inducing and screening pAG2-A β 42 library colonies was utilised (Wurth et al., 2002; Baine et al., 2009). While this mitigated the problems of undesired plasmid recombination and background cells, it could not solve the lack of dynamic range between the A β 42-EGFP output from negative and positive cells.

While flow cytometry of *E. coli* expressing the putative hits was used to further validate the primary colony screen, it also revealed that mCherry fluorescence levels, rather than being uniformly stable, varied between different library members. For example, a ~2-fold difference in mCherry output was observed between 110-B1-3 and the “empty” pAG2-A β 42 control. This put the reliability of using constitutively expressed mCherry as a ratiometric standard in doubt. Why such variance occurred is unclear, but some general comments can be made. Firstly, vector construction is often assumed to be a “plug and play” affair, where discrete encoding sequences can be inserted or removed in isolation. However, as synthetic biologists are realising, biological systems are inherently noisy (Kelly et al., 2009). “Parts” are not insulated from each other, and as a consequence may work unpredictably under different configurations and conditions (Kwok, 2010). Although all variables aside from the random DNA library insert were closely controlled to ensure experimental homogeneity, variability between even the same samples run on different days was observed (e.g. 110-A3-5). Despite such variation, trends can still be discerned between sample runs: Endy and co-workers concluded the same while assessing promoter activity using a fluorescent reporter, as measurements varied by ~2-fold when made by separate research groups despite following the same protocol (Kelly et al., 2009).

Curiously, several putative antiaggregant constructs were seen to exhibit cell clumping during growth in liquid medium. Ranging from severe (110-C3-3) to minor (AavLEA1), this condition led to partial sedimentation, with small (~50 to ~250 μm) bacterial clumps being visible under the microscope. Although this did not affect cell viability, it was less than ideal for the dispersal of single cells during flow cytometry analysis. A tentative explanation of this phenomenon is that clumping was induced by soluble A β 42-EGFP, i.e. the by-product of an active antiaggregant effect. As A β 42 has previously been shown to exhibit an antimicrobial effect against *E. coli* (Soscia et al., 2010), perhaps the A β 42-EGFP fusion possessed a similar activity: *E. coli* cell clumping has previously been observed as a stress response to antibiotics (e.g. tobramycin) (Hoffman et al., 2005). However, there were no other similarities seen between this work and that conducted with AMPs in *Chapter 4*, i.e. a reduced growth rate during intracellular production. And although an increase in A β 42-EGFP fluorescence was observed in most of the clumping constructs during induction, lending some support to the theory of a soluble A β 42-EGFP effect, the clumping phenomenon was also observed when A β 42-EGFP expression was not induced. Given the tightness of *araBAD* repression in the absence of arabinose (Guzman et al., 1995), it is unlikely that soluble A β 42-EGFP is the cause of cell clumping. Instead, this phenotype is most likely to be a stress response to the constitutive expression of particular peptides from the Library site; however, no common homology was found between the five implicated putative peptides. Aside from antibiotics, cell clumping has also been observed as a response to general stresses such as changes in pH and heat shock (Zhang et al., 2007); unfavourable peptide expression may thus be another effector. A few membrane proteins, such as antigen 43 (Kjaergaard et al., 2000) and curli CsgA (Barnhart & Chapman, 2006), are implicated as agents of cell clumping and biofilm formation. Perhaps investigating any up-regulation of these products in the clumping constructs could prove informative. However, with regards to antigen 43 expression, *E. coli* colony morphology is described as being “frizzy” (a distinctly ruffled surface) (Kjaergaard et al., 2000): in the work presented here, colony morphologies were indistinguishable between normal and clumping constructs (data not shown). Lastly, while the AavLEA1 control was seen to endow minor levels of clumping, other members of

the Tunnacliffe laboratory have not reported such an effect when recombinantly expressing this protein in other *E. coli* cell lines.

The sequences of the eighteen putative antiaggregant peptide hits profiled in this work were found to have little homology with the A β 42 peptide, indicating that their mode of action is probably not that of a traditional homologous β -blocker (Soto & Estrada, 2005). Furthermore, runs of three or more hydrophobic residues were rare, indicating against any putative hydrophobic interactions with the β -strand regions of A β 42. An explanation of how these putative antiaggregant peptides interact with A β 42 is not immediately obvious, especially given the lack of overall homology between them. It may be that they have other effects on fluorescent output, such as by facilitating chromophore folding without preventing A β 42 aggregation (Baine et al., 2009). While this theory may explain the observed increase in mCherry fluorescence seen in several constructs, it does not quite fit. Pep2, which exhibited a concurrent *in vivo* increase in mCherry fluorescence alongside that of A β 42-EGFP, also displayed green fluorescent aggregates in a minority of cells. Baine et al. (2009) previously showed that, while lacking a large influence *in vivo* (the fluorescence morphology was not mentioned), Pep2 had an impressive antiaggregant effect on A β 42 alone *in vitro* during ThT assays (LeVine, 1993). As green fluorescent aggregates were also seen in a number of cells expressing putative hits, similar antiaggregation behaviour of these to Pep2 *in vitro* was predicted.

Although preliminary *in vitro* results from collaborators indicated that 110-A1-4 and 110-C3-7 moderately retard the rate of A β 42 aggregation, surprisingly the positive control Pep2 was not seen to exhibit any antiaggregant activity (Y. Zheng, personal communication). The use of electron microscopy to visualise any aggregate morphology affect may be more informative (Sato et al., 2006b). Interestingly, for some antiaggregants, the use of ThT to assess a compound's activity against A β may not be truly indicative of function (Pallitto et al., 1999). For example, the small molecule inositol, despite being shown to induce β -sheet structure in A β 42 by binding ThT, is still able to inhibit overall fibril formation (DaSilva et al., 2010). Electron microscopy revealed that small and apparently non-toxic oligomers are

instead formed in the presence of *scyllo*-inositol (DaSilva et al., 2010). If the formation of such oligomers prevents EGFP from folding, the *in vivo* screen described in this work may miss peptides that act in a similar manner. If some of the hit peptides can be satisfactorily validated *in vitro*, future work may involve testing putative antiaggregant peptides in neuronal cell models to ascertain if they prevent A β toxicity, and then perhaps in animal models such as APP transgenic mice (Walsh & Selkoe, 2007). However, any antiaggregant peptides also need to be capable of uptake across the blood-brain barrier and potentially into neurons themselves (Hartmann et al., 1997; Walsh et al., 2002). Such bioavailability problems (reviewed in *Section 1.1.2* and *1.1.3.1*) remain to be surmounted.

A number of putative hits were predicted to only encode extremely short peptides (<6 aa), and consequently were not expected to be able to exhibit any antiaggregant activity *in vivo* due to short product half-life (Maurizi, 1992). However, closer examination of these putative false-positives revealed that alternative reading frames were present. For example, in 110-C2-3 a stop codon was predicted to immediately halt translation after initiation at the original start codon; however, an alternative start codon for a 26 aa product was present in close proximity to the RBS (see *Appendix 9*). Although peptide production from this alternative ORF (or indeed any ORF) was not directly confirmed, this explanation is not without merit given that 110-C2-3 exhibited a homologous alternative peptide sequence to 74-B1-3 as well as a similar *in vivo* performance. Further investigation *in vitro* using such predicted alternative peptides could confirm whether or not they are true hits. 74-C1-1, on the other hand, looked to be a true false-positive from the colony screen. Encoding 3 aa before an in-frame stop codon was reached, no other ORFs could be identified. Unsurprisingly, this construct performed no differently from the “empty” pAG2-A β 42 control during flow cytometry.

While the work presented here has focused on the internal production of antiaggregant peptides *in vivo*, the system is also amenable to use in a liquid culture well-plate format to assay exogenously added compounds as per Kim et al. (2006) and Baine et al. (2009). In addition, the A β 42-EGFP aggregator operon may be replaced with other aggregating proteins, such as those implicated in

Huntington's or Parkinson's disease (see *Section 5.1.1*). Preliminary data from fusing EGFP to the N-terminus of exon I of the huntingtin gene found fluorescent aggregates form over several hours post-induction; however, total fluorescence output per cell was equivalent to that of EGFP alone (data not shown). It is thus important to note that, if the aggregation rate of the target is slower than that of the folding of EGFP (approximately 0.5 to 1.5 h [Waldo et al., 1999]), the screen loses its sensitivity as fluorescent aggregates are capable of producing similar total amounts of fluorescence per cell as diffuse EGFP.

Further improvements to the screen could involve tweaking either the aggregation rate of the A β component or the sensitivity of the chromophore domain to aggregation events. For instance, using a less-aggressively aggregating form of A β , such as A β 40 or A β 41, may allow for an antiaggregant effect to be better observed. Similarly, Wurth et al. (2002) identified a number of other A β 42 mutants alongside GM6 that still aggregate, but at slower rates than that of the wild-type. With regards to the fluorescent reporter, Waldo and co-workers, who initially described the GFP folding reporter assay (Waldo et al., 1999), have generated a new range of split-GFP variants with a decreased sensitivity to aggregation events (Cabantous et al., 2008). Either of these two approaches may potentially increase the dynamic range between a positive and negative event.

Finally, in a therapeutic sense, absolute prevention of A β 42 aggregation may be unnecessary: if aggregation can be sufficiently slowed, perhaps toxic products can be more easily cleared by the brain's natural degradative machinery (Walker et al., 2001). For similar reasons an ultimate cure, while desired, may too be redundant. As Alzheimer's disease appears to be an inevitable consequence of increased life expectancy (Schnabel, 2011), the production of a viable antiaggregant drug that delays the onset of symptoms for ~30 years may render the problem moot. Unfortunately, most patients who are recruited into clinical trials of potential antiaggregants already exhibit mild-to-moderate progression of the disease (Sheridan, 2009), thus making true assessment of efficacy difficult. A focus of using antiaggregants prophylactically, rather than curatively, is then the next-best strategy. Choosing when to start such a treatment, however, is problematic, as

plaques are observed to form 10 to 15 years prior to symptom onset (Gravitz, 2011). If the presence of plaques were used as a biomarker and somehow identified at the earliest possible stage (histological analysis not being practical), perhaps a prophylactic strategy may be able to prolong symptom onset past probable life expectancy.

In conclusion, an A β antiaggregant peptide screening vector was constructed, with the fluorescent output of A β 42-EGFP linked to its aggregation state. As a novelty, constitutively expressed mCherry was also encoded to act as a constant internal control for ratiometric comparison. DNA libraries encoding a maximum of either 12 or 24 random amino acids were co-expressed and screened for members that increased A β 42-EGFP fluorescence. Although FACS was originally envisaged as the method for isolating putative antiaggregant hits, this proved to be infeasible mainly due to a low dynamic range between positive cells and background noise. Instead, screening was conducted via imaging colonies on solid medium and subsequent validation through flow cytometry. While several peptide hits exceeded or were comparable to the effect of the positive control Pep2 *in vivo*, restoration of EGFP fluorescence to the same level as the non-aggregating GM6 control was not achieved. Surprisingly, some peptides (including Pep2) were also found to influence mCherry output, suggesting alternative modes of action. Further *in vitro* validation of the putative antiaggregant peptide hits is being carried out by collaborators, with 110-A1-4 and 110-C3-7 showing initial signs of promise in slowing the rate of A β 42 aggregation.

CHAPTER 6 – FINAL DISCUSSION

This dissertation has been concerned with exploring the use of *in vivo* DNA library screens to identify sequences that encode novel bioactive peptides. Specifically, the activities of antimicrobials and antiaggregants were focused on, with the emphasis being primarily on the former. In *Chapter 3*, the recently described CPD tag (Shen et al., 2009) was used in *E. coli* to recombinantly produce the model linear cathelicidin K2C18 (Shin et al., 2000) in a simple and straightforward manner. Several variants were also generated and confirmed to exhibit antimicrobial activity against a panel of microbial species. The worth of the CPD production system was thus established for AMP optimisation and structure/function studies.

Chapter 4 employed *E. coli* in an *in vivo cis*-based whole-cell screen to isolate novel AMPs from sheared genomic DNA libraries. Utilising a replica-plating methodology, a number of endogenously expressed hits were identified that required secretion to the periplasm for activity to occur, suggesting a membrane-specific site of action similar to that of K2C18. Chemical synthesis of one such novel AMP (S-H4) revealed the presence of moderate antimicrobial activity, establishing proof-of-principle for this screen.

Finally, as a second *in vivo* screen example, *Chapter 5* investigated the use of an EGFP fusion for reporting the aggregation status of A β (Wurth et al., 2002), which is heavily implicated in Alzheimer's disease (Walsh & Selkoe, 2007). Using a colony screen, random DNA libraries were assayed for sequences that encoded peptides possessing putative antiaggregant properties, and the hits were subsequently assessed using flow cytometry. Furthermore, the use of an internal ratiometric standard (mCherry) was investigated to allow for more robust comparison between the effect of each putative antiaggregant. Although the ratiometric approach did not perform as desired, a number of positive hits were identified. Further *in vitro* validation of several chemically-synthesised hits is ongoing with collaborators.

During this dissertation, a two-step workflow for cost-effective and high-throughput discovery of bioactive peptides has become apparent. Firstly, recombinant production of peptides should be conducted in a suitable model organism, utilising

DNA libraries as the input encoding source. Importantly for the purpose of screening, peptide production is coupled *in vivo* to a measurable output for the desired activity. Thus, each host cell represents an individual test-tube for the assessment of a specific library member; and as long as positive cells can be subsequently isolated, the encoded peptide sequence can easily be deduced from the DNA insert. In addition, the ability for this peptide-producing “living laboratory” to self-replicate may allow for the recovery of sufficient recombinant peptide for purification and further *in vitro* validation. While this was achieved for several peptides using the CPD system in *Chapter 3*, a number of putative AMP hits in *Chapter 4* were unable to be produced in this manner.

The above served as a reminder that successful recombinant peptide production to a meaningful yield is context-specific, and cannot be accurately predicted *a priori*. Therefore, the second step of the workflow is to chemically synthesise selected hit peptides for further validation. Solid-phase peptide synthesis is a mature technology that can be performed generically, and for relatively short runs of amino acids such as peptides, remains the gold-standard (Lax, 2010). While *in vitro* validation using synthetic peptides is required to confirm the accuracy of a screen, chemical synthesis is also likely to be the production method of choice for any hits that proceed to clinical trials. Manufacturing robustness, quality assurance, regulatory compliance and the fact that such production can be easily outsourced are all factors in favour of this approach (Pichereau & Allary, 2006; Lax, 2010).

Rather than being employed directly in a screen from the beginning, however, chemically-synthesised peptides are more suited towards optimising known sequences at a small scale. This is due to the prohibitive cost of individual peptide synthesis: libraries consisting of up to $\sim 1 \times 10^4$ members are the current practical limit, even with the aid of miniaturisation and robotics (Hilpert et al., 2005; Rathinakumar et al., 2009). Recombinant systems can easily and cheaply exceed this limit, with 5×10^6 members being theoretically feasible from a single standard transformation of *E. coli* (assuming 10^9 cfu/ μ g DNA, with 5 ng transformed). As with any screen, the power is in the numbers. The larger a library analysed, the higher the likelihood that a potential candidate peptide will be identified. While the antimicrobial and antiaggregant assays presented here utilised small libraries

($\sim 1 \times 10^4$) as proof-of-principle, they are compatible with off-the-shelf equipment such as colony pickers and automated handling systems, the use of which is ultimately required for the realisation of high-throughput screening (Raventós et al., 2005).

Other recombinant production systems are available for use as hosts for *in vivo* screens, but do not match the robustness and ease of use ascribed to *E. coli* (Baneyx, 1999). Yeasts are capable of the generation of similarly large libraries (in excess of 1×10^9 members), as well as stable protein secretion (Raventós et al., 2005). However, experiments with *S. cerevisiae* and *P. pastoris* to produce K2C18 in this dissertation failed to produce significant yields as opposed to the use of *E. coli*. Mammalian cell systems, while commonly used for the production of many proteinaceous therapeutics (Wurm, 2004), are unwieldy and time-consuming with regards to library transfection and screening (Kaykas & Moon, 2004; Li et al., 2010).

The use of other hosts does hold appeal for certain screens, however. For instance, in the AMP screen outlined in *Chapter 4*, using bacterial strains other than *E. coli* as a production chassis could allow peptides with specificity towards important pathogens to be discovered. Genetically tractable and attenuated relatives of antibiotic-resistant “superbugs” exist, such as *Pseudomonas aeruginosa* (West et al., 1994) and *Salmonella enterica* serovar Typhimurium (Widmaier et al., 2009; Medina et al., 2011). Furthermore, the use of Gram-positives such as *Bacillus subtilis* (Brockmeier et al., 2006) or *Bacillus megatarium* (Malten et al., 2006) could be better suited towards the search for novel AMPs that act exogenously. With the transit of only a single membrane required for protein secretion, the use of the Sec pathway as described in *Chapter 4* should be sufficient to export peptides into the extracellular milieu (Fu et al., 2007). However, secretion of heterologous proteins can be unpredictable due to the presence of quality control and “feeding” proteases co-secreted by the bacterium (Harwood & Cranenburgh, 2008).

If *E. coli* is to remain the host of choice for the AMP screen, further improvements could be made with regards to generating exogenously acting hits. Given that the Sec system, although well described (Saier, 2006), does not guarantee peptide secretion past the periplasm (Ni & Chen, 2009), alternative but lesser proven secretion pathways represent further avenues for investigation. One such candidate

is the ATP-binding cassette Type I system used by pathogenic *E. coli* to secrete the toxin haemolysin (110 kDa; Gentschev et al. [2002]); another is the Type V system, specifically the serine protease autotransporters of the *Enterobacteriaceae* class (SPATEs, ~110 kDa; Yen et al. [2008]). Both are able to secrete protein through the inner and outer membranes of *E. coli* using dedicated molecular machinery: in the case of SPATEs, this is encoded within the protein itself (hence the autotransporter moniker). However, these systems (especially the SPATEs [Jong et al., 2010]) are not well described for heterologous protein production. Furthermore, they both require conserved C-terminal secretion signals: any insertion of DNA upstream for screening may introduce disrupting frame-shifts or internal stop codons, preventing proper translation and secretion. A known exception to this is the colicin V toxin (9 kDa, Type I system); its secretion signal is N-terminal (Hwang et al., 1997). By incorporating the dedicated secretion machinery (ClvA and ClvB, in combination with the endogenous outer membrane protein TolC [Hwang et al., 1997]) into a screening vector, such a system could be trialled for AMP secretion; again, however, its ability to cope with heterologous protein remains to be seen.

Regarding future screen development, other lessons learnt from this dissertation should be applied. Drawing from the AMP screen again, it has been postulated that virtually every peptide that possesses a net positive charge and a few hydrophobic residues will display antimicrobial activity if assayed in common buffers or dilute media (Hancock & Sahl, 2006). Although the novel sequence of S-H4 exhibited antimicrobial activity in high-salt LB media in a similar manner to the known AMP K2C18, it is pertinent to note that not every hit identified can be expected to show clinical efficacy (even discounting stability and bioavailability issues [Mlieghe et al., 2010]). To better facilitate this, the next generation of screens should try to replicate as closely as possible the physiochemical environment in which the peptides will ultimately be deployed. For AMPs, if the end use is topical (e.g. treatment of cystic fibrosis infections), this means modelling high-salt epithelial environments (Knowles et al., 1997); or for systemic use, the use of appropriately cation- and serum-adjusted broths to simulate conditions in the blood (Wiegand et al., 2008). For A β antiaggregants, the employment of media that replicates the properties of cerebrospinal fluid would be desirable (Wishart et al., 2008), although as the screen takes place in the *E. coli* host's cytoplasm it is difficult to optimise this further. At the

very least, if a screen cannot be made physiologically relevant, it should contain strong positive controls that show a good correlation between screen activity and clinical efficacy (if these exist for peptides; if not, small molecule drugs may be employed instead). In this manner novel peptide candidates with similar *in vivo* profiles can be identified and proceed to be validated piecemeal in other systems.

A greater emphasis on *trans*-acting systems, in which the production host does not double as the test organism, has been suggested as a further advance towards the goal of creating more robust screens (Raventós et al., 2005). In this approach, each library clone acts as a miniature peptide production system, with aliquots subsequently taken and added to the assay proper. To simplify peptide recovery, however, secretion is required: and even if reliably achieved, the peptide may be diluted below the concentration required for activity. Additional processing techniques are thus necessary to recover and concentrate a peptide prior to testing, which is not particularly amenable to high-throughput screening. Complicating a screen with such additional processing steps negates the main advantage of *in vivo* screening: self-contained peptide production combined with activity assessment in a discrete self-replicating entity. If the two are separated, the recombinant approach is of comparable utility to that of harnessing chemically-synthesised peptides at the initial stages of a screen.

Looking ahead, it is clear that therapeutic peptides have a promising future. Although they currently suffer from the twin problems of poor bioavailability and stability (Vlieghe et al., 2010), this has not prevented them from filling valuable clinical niches worth approximately US\$13 billion in 2010 (Reichert et al., 2010). While strategies for protecting peptides from breakdown continue to be developed (McGregor, 2008), it seems reasonable to continue the search for novel bioactive peptides to develop the art for eventual deployment into clinical trials. At the very least, “low hanging” targets can continue to be focused upon: profitable hormonal analogues (Loffet, 2002), antimicrobial peptides for use as topical or antifouling agents (Hancock & Sahl, 2006), and even cosmetic product supplements (Pichereau & Allary, 2006) all represent areas of significant interest. It is hoped that the work contained in this dissertation has shown that simple microbial systems can be used towards these developmental ends.

REFERENCES

- Altschul, S., Madden, T., Schäffer, A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25** (17): 3389-402.
- An, L.-L., Yang, Y.-H., Ma, X.-T., Lin, Y.-M., Li, G., Song, Y.-H. and Wu, K. F. (2005). LL-37 enhances adaptive antitumor immune response in a murine model when genetically fused with M-CSFR (J6-1) DNA vaccine. *Leuk Res* **29** (5): 535-43.
- Amiel, J., Trochet, D., Clément-Ziza, M., Munnich, A. and Lyonnet, S. (2004). Polyalanine expansions in human. *Hum Mol Genet* **13** (Suppl 2), R235-43.
- Awtry, E. and Loscalzo, J. (2000). Aspirin. *Circulation* **101** (10): 1206-18.
- Baine, M., Georgie, D., Shiferraw, E., Nguyen, T., Nogaj, L. and Moffet, D. (2009). Inhibition of A β 42 aggregation using peptides selected from combinatorial libraries. *J Pept Sci* **15** (8): 499-503.
- Bals, R. and Wilson, J. (2003). Cathelicidins - a family of multifunctional antimicrobial peptides. *Cell Mol Life Sci* **60** (4): 711-20.
- Baneyx, F. (1999). Recombinant protein expression in Escherichia coli. *Curr Opin Biotechnol* **10** (5): 411-21.
- Barnhart, M. and Chapman, M. (2006). Curli biogenesis and function. *Annu Rev Microbiol* **60**: 131-47.
- Barrow, C. and Zagorski, M. (1991). Solution structures of beta peptide and its constituent fragments: relation to amyloid deposition. *Science* **253** (5016): 179-82.
- Bej, A., Perlin, M. and Atlas, R. (1988). Model suicide vector for containment of genetically engineered microorganisms. *Appl Environ Microbiol* **54** (10): 2472-7.
- Bell, A. (2011). Antimalarial Peptides: The Long and the Short of it. *Curr Pharm Des* **17** (25): 2719-31.
- Ben-Bassat, A., Bauer, K., Chang, S., Myambo, K., Boosman, A. and Chang, S. (1987). Processing of the initiation methionine from proteins: properties of the Escherichia coli methionine aminopeptidase and its gene structure. *J Bacteriol* **169** (2): 751-7.
- Berman, H., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T., Weissig, H., Shindyalov, I. and Bourne, P. (2000). The Protein Data Bank. *Nucleic Acids Res* **28** (1): 235-42.
- Best, C. (1962). The internal secretion of the pancreas. The internal secretion of the pancreas. *Can Med Assoc J* **87**: 1046-51.
- Binz, H., Amstutz, P. and Plückthun, A. (2005). Engineering novel binding proteins from nonimmunoglobulin domains. *Nat Biotechnol* **23** (10): 1257-68.
- Bitan, G. and Teplow, D. (2005). Preparation of aggregate-free, low molecular weight amyloid-beta for assembly and toxicity assays. *Methods Mol Biol* **299**: 3-9.
- Bitan, G., Fradinger, E., Spring, S. and Teplow, D. (2005). Neurotoxic protein oligomers - what you see is not always what you get. *Amyloid* **12** (2): 88-95.
- Blattner, F., Plunkett, G., Bloch, C., Perna, N., Burland, V., Riley, M. et al. and Shao, Y. (1997). The complete genome sequence of Escherichia coli K-12. *Science* **277** (5331): 1453-62.
- Bleicher, K., Böhm, H.-J., Müller, K. and Alanine, A. (2003). Hit and lead generation: beyond high-throughput screening. *Nat Rev Drug Discov* **2** (5): 369-78.
- Blondelle, S. and Lohner, K. (2010). Optimization and High-Throughput Screening of Antimicrobial Peptides. *Curr Pharm Des* **16** (28): 3204-11.
- Boman, H., Agerberth, B. and Boman, A. (1993). Mechanisms of action on Escherichia coli of cecropin P1 and PR-39, two antibacterial peptides from pig intestine. *Infect Immun* **61** (7): 2978-84.
- Bommarius, B., Jenssen, H., Elliott, M., Kindrachuk, J., Pasupuleti, M., Gieren, H., Jaeger, K.-E., Hancock, R. and Kalman, D. (2010). Cost-effective expression and purification of antimicrobial and host defense peptides in Escherichia coli. *Peptides* **31** (11): 1957-65.
- Bond, P. and Khalid, S. (2010). Antimicrobial and cell-penetrating peptides: structure, assembly and mechanisms of membrane lysis via atomistic and coarse-grained molecular dynamics simulations. *Protein Pept Lett* **17** (11): 1313-27.
- Bosch, F. and Rosich, L. (2008). The contributions of Paul Ehrlich to pharmacology: a tribute on the occasion of the centenary of his Nobel Prize. *Pharmacology* **82** (3): 171-9.

- Bowdish, D., Davidson, D. and Hancock, R. (2005a). A re-evaluation of the role of host defence peptides in mammalian immunity. *Curr Protein Pept Sci* **6** (1): 35-51.
- Bowdish, D., Davidson, D., Lau, Y., Lee, K., Scott, M. and Hancock, R. (2005b). Impact of LL-37 on anti-infective immunity. *J Leukoc Biol* **77** (4): 451-9.
- Bowdish, D., Davidson, D., Scott, M. and Hancock, R. (2005c). Immunomodulatory activities of small host defense peptides. *Antimicrob Agents Chemother* **49** (5): 1727-32.
- Bray, B. (2003). Large-scale manufacture of peptide therapeutics by chemical synthesis. *Nat Rev Drug Discov* **2** (7): 587-93.
- Brockmeier, U., Wendorff, M. and Eggert, T. (2006). Versatile expression and secretion vectors for *Bacillus subtilis*. *Curr Microbiol* **52** (2): 143-8.
- Brogden, K. (2005). Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? *Nat Rev Microbiol* **3** (3): 238-50.
- Browne, J., Tunnacliffe, A. and Burnell, A. (2002). Anhydrobiosis: plant desiccation gene found in a nematode. *Nature* **416** (6876): 38.
- Brückner, A., Polge, C., Lentze, N., Auerbach, D. and Schlattner, U. (2009). Yeast two-hybrid, a powerful tool for systems biology. *Int J Mol Sci* **10** (6): 2763-88.
- Brunken, K., Trojanowski, J. and Lee, V.-Y. (2009). Advances in tau-focused drug discovery for Alzheimer's disease and related tauopathies. *Nat Rev Drug Discov* **8** (10): 783-93.
- Brüser, T. (2007). The twin-arginine translocation system and its capability for protein secretion in biotechnological protein production. *Appl Microbiol Biotechnol* **76** (1): 35-45.
- Cabantous, S., Rogers, Y., Terwilliger, T. and Waldo, G. (2008). New molecular reporters for rapid protein folding assays. *PLoS ONE* **3** (6): e2387.
- Canton, B., Labno, A. and Endy, D. (2008). Refinement and standardization of synthetic biological parts and devices. *Nature Biotechnol* **26** (7): 787-93.
- Carter, M., Simms, G. and Weaver, D. (2010). The development of new therapeutics for Alzheimer's disease. *Clin Pharmacol Ther* **88** (4): 475-86.
- Cattaneo, E., Zuccato, C. and Tartari, M. (2005). Normal huntingtin function: an alternative approach to Huntington's disease. *Nat Rev Neurosci* **6** (12): 919-30.
- Chakrabortee, S., Boschetti, C., Walton, L., Sarkar, S., Rubinsztein, D. and Tunnacliffe, A. (2007). Hydrophilic protein associated with desiccation tolerance exhibits broad protein stabilization function. *Proc Natl Acad Sci USA* **104** (46): 18073-8.
- Chakrabortee, S., Tripathi, R., Watson, M., Kaminski-Schierle, G., Kurniawan, D., Kaminski, C., Wise, M. and Tunnacliffe, A. (2012). Intrinsically disordered proteins as molecular shields. *Mol Biosyst* **8** (1): 210-9.
- Chandler, D. (2005). Interfaces and the driving of hydrophobic assembly. *Nature* **437** (7059): 640-7.
- Chen, H., Bjerknes, M., Kumar, R. and Jay, E. (1994). Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of *Escherichia coli* mRNAs. *Nucleic Acids Res* **22** (23): 4953-7.
- Chen, Y., Guarnieri, M., Vasil, A., Vasil, M., Mant, C. and Hodges, R. (2007). Role of peptide hydrophobicity in the mechanism of action of alpha-helical antimicrobial peptides. *Antimicrob Agents Chemother* **51** (4): 1398-406.
- Cheng, X., Liu, G., Ye, G., Wang, H., Shen, X., Wu, K., Xie, J. and Altosaar, I. (2009). Screening and cloning of antimicrobial DNA sequences using a vital staining method. *Gene* **430** (1-2): 132-9.
- Cherkasov, A., Hilpert, K., Jenssen, H., Fjell, C., Waldbrook, M., Mullaly, S., Volkmer, R. and Hancock, R. (2009). Use of artificial intelligence in the design of small peptide antibiotics effective against a broad spectrum of highly antibiotic-resistant superbugs. *ACS Chem Biol* **4** (1): 65-74.
- Choi, J. and Lee, S. (2004). Secretory and extracellular production of recombinant proteins using *Escherichia coli*. *Appl Microbiol Biotechnol* **64** (5): 625-35.
- Cotter, P., Hill, C. and Ross, R. (2005). Bacteriocins: developing innate immunity for food. *Nat Rev Microbiol* **3** (10): 777-88.
- Cramer, A., Whitehorn, E., Tate, E. and Stemmer, W. (1996). Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat Biotechnol* **14** (3): 315-9.
- Crooks, G., Hon, G., Chandonia, J.-M. and Brenner, S. (2004). WebLogo: a sequence logo generator. *Genome Res* **14** (6): 1188-90.

- Daly, R. and Hearn, M. (2005). Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production. *J Mol Recognit* **18** (2): 119-38.
- DaSilva, K. A., Shaw, J. E. and McLaurin, J. (2010). Amyloid-beta fibrillogenesis: structural insight and therapeutic intervention. *Exp Neurol* **223** (2): 311-21.
- Dathe, M. and Wieprecht, T. (1999). Structural features of helical antimicrobial peptides: their potential to modulate activity on model membranes and biological cells. *Biochim Biophys Acta* **1462** (1-2): 71-87.
- Davidson, D., Currie, A., Reid, G., Bowdish, D., MacDonald, K., Ma, R., Hancock, R. and Speert, D. (2004). The cationic antimicrobial peptide LL-37 modulates dendritic cell differentiation and dendritic cell-induced T cell polarization. *J Immunol* **172** (2): 1146-56.
- Davies, J. and Jacob, F. (1968). Genetic mapping of the regulator and operator genes of the lac operon. *J Mol Biol* **36** (3): 413-7.
- De Strooper, B., Vassar, R. and Golde, T. (2010). The secretases: enzymes with therapeutic potential in Alzheimer disease. *Nat Rev Neurol* **6** (2): 99-107.
- Devasahayam, G., Scheld, W. and Hoffman, P. (2010). Newer antibacterial drugs for a new century. *Expert Opin Investig Drugs* **19** (2): 215-34.
- Di Segni, G., Gastaldi, S. and Tocchini-Valentini, G. (2008). Cis- and trans-splicing of mRNAs mediated by tRNA sequences in eukaryotic cells. *Proc Natl Acad Sci USA* **105** (19): 6864-9.
- Dobson, C. (1999). Protein misfolding, evolution and disease. *Trends Biochem Sci* **24** (9): 329-32.
- Dodart, J.-C., Bales, K., Gannon, K., Greene, S., DeMattos, R., Mathis, C., DeLong, C., Wu, S., Wu, X., Holtzman, D. and Paul, S. (2002). Immunization reverses memory deficits without reducing brain A β burden in Alzheimer's disease model. *Nat Neurosci* **5** (5): 452-7.
- Domínguez, M., de La Rosa, M. and Borobio, M. (2001). Application of a spectrophotometric method for the determination of post-antibiotic effect and comparison with viable counts in agar. *J Antimicrob Chemother* **47** (4): 391-8.
- Don, R., Cox, P., Wainwright, B., Baker, K. and Mattick, J. (1991). 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res* **19** (14): 4008.
- Dorschner, R., Pestonjamas, V., Tamakuwala, S., Ohtake, T., Rudisill, J., Nizet, V., Agerberth, B., Gudmundsson, G. and Gallo, R. (2001). Cutaneous injury induces the release of cathelicidin anti-microbial peptides active against group A *Streptococcus*. *J Invest Dermatol* **117** (1): 91-7.
- Drews, J. (1996). Genomic sciences and the medicine of tomorrow. *Nat Biotechnol* **14** (11): 1516-8.
- Drews, J. (2000). Drug discovery: a historical perspective. *Science* **287** (5460): 1960-4.
- Durfee, T., Nelson, R., Baldwin, S., Plunkett, G., Burland, V., Mau, B. et al. and Blattner, F. (2008). The complete genome sequence of *Escherichia coli* DH10B: insights into the biology of a laboratory workhorse. *J Bacteriol* **190** (7): 2597-606.
- Ehrnhoefer, D., Bieschke, J., Boeddrich, A., Herbst, M., Masino, L., Lurz, R., Engemann, S., Pastore, A. and Wanker, E. (2008). EGCG redirects amyloidogenic polypeptides into unstructured, off-pathway oligomers. *Nat Struct Mol Biol* **15** (6): 558-66.
- Eisele, Y., Obermüller, U., Heilbronner, G., Baumann, F., Kaeser, S., Wolburg, H., Walker, L., Staufenbiel, M., Heikenwalder, M. and Jucker, M. (2010). Peripherally applied A β -containing inoculates induce cerebral beta-amyloidosis. *Science* **330** (6006): 980-2.
- Eisenberg, D. (1984). Three-dimensional structure of membrane and surface proteins. *Annu Rev Biochem* **53**: 595-623.
- El Zoeiby, A., Sanschagrin, F., Darveau, A., Brisson, J.-R. and Levesque, R. (2003). Identification of novel inhibitors of *Pseudomonas aeruginosa* MurC enzyme derived from phage-displayed peptide libraries. *J Antimicrob Chemother* **51** (3): 531-43.
- Eriksson, M., Nielsen, P. and Good, L. (2002). Cell permeabilization and uptake of antisense peptide-peptide nucleic acid (PNA) into *Escherichia coli*. *J Biol Chem* **277** (9): 7144-7.
- Fairlamb, A. and Cole, S. (2011). Antimicrobial drug discovery. *Future Microbiol* **6** (6): 601-2.
- Fantner, G., Barbero, R., Gray, D. and Belcher, A. (2010). Kinetics of antimicrobial peptide activity measured on individual bacterial cells using high-speed atomic force microscopy. *Nat Nanotech* **5**: 280-5.
- Fay, D., Fluet, A., Johnson, C. and Link, C. (1998). In vivo aggregation of beta-amyloid peptide variants. *J Neurochem* **71** (4): 1616-25.

- Fields, S. and Song, O. (1989). A novel genetic system to detect protein-protein interactions. *Nature* **340** (6230): 245-6.
- Findeis, M., Musso, G., Arico-Muendel, C., Benjamin, H., Hundal, A., Lee, J., Chin, J., Kelley, M., Wakefield, J., Hayward, N. and Molineaux, S. (1999). Modified-peptide inhibitors of amyloid beta-peptide polymerization. *Biochemistry* **38** (21): 6791-800.
- Fischbach, M., and Walsh, C. (2009). Antibiotics for emerging pathogens. *Science* **325** (5944): 1089-93.
- Fisher, A., Kim, W. and DeLisa, M. (2006). Genetic selection for protein solubility enabled by the folding quality control feature of the twin-arginine translocation pathway. *Protein Sci* **15** (3): 449-58.
- Fjell, C., Hancock, R. and Cherkasov, A. (2007). AMPer: a database and an automated discovery tool for antimicrobial peptides. *Bioinformatics* **23** (9): 1148-55.
- Fjell, C., Jenssen, H., Hilpert, K., Cheung, W., Panté, N., Hancock, R. and Cherkasov, A. (2009). Identification of novel antibacterial peptides by chemoinformatics and machine learning. *J Med Chem* **52** (7): 2006-15.
- Fleming, A. (1929). On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae (2001 reprint). *Bull World Health Organ* **79** (8): 780-90.
- Foerg, C. and Merkle, H. (2008). On the biomedical promise of cell penetrating peptides: limits versus prospects. *J Pharm Sci* **97** (1): 144-62.
- Fraile, S., Roncal, F., Fernández, L. and de Lorenzo, V. (2001). Monitoring intracellular levels of XylR in *Pseudomonas putida* with a single-chain antibody specific for aromatic-responsive enhancer-binding proteins. *J Bacteriol* **183** (19): 5571-9.
- Freceer, V., Ho, B. and Ding, J. (2004). De novo design of potent antimicrobial peptides. *Antimicrob Agents Chemother* **48** (9): 3349-57.
- Fricke, B., Parchmann, O., Kruse, K., Rücknagel, P., Schierhorn, A. and Menge, S. (1999). Characterization and purification of an outer membrane metalloproteinase from *Pseudomonas aeruginosa* with fibrinogenolytic activity. *Biochim Biophys Acta* **1454** (3): 236-50.
- Fu, L., Xu, Z., Li, W., Shuai, J., Lu, P. and Hu, C. (2007). Protein secretion pathways in *Bacillus subtilis*: implication for optimization of heterologous protein secretion. *Biotechnol Adv* **25** (1): 1-12.
- Fukuda, I., Kojoh, K., Tabata, N., Doi, N., Takashima, H., Miyamoto-Sato, E. and Yanagawa, H. (2006). In vitro evolution of single-chain antibodies using mRNA display. *Nucleic Acids Res* **34** (19): e127.
- Fulmer, P. and Wynne, J. (2011). Development of Broad-Spectrum Antimicrobial Latex Paint Surfaces Employing Active Amphiphilic Compounds. *ACS Appl Mater Interfaces* **3** (8): 2878-84.
- Gallo, R., Kim, K., Bernfield, M., Kozak, C., Zanetti, M., Merluzzi, L. and Gennaro, R. (1997). Identification of CRAMP, a cathelin-related antimicrobial peptide expressed in the embryonic and adult mouse. *J Biol Chem* **272** (20): 13088-93.
- Gallo, R. and Huttner, K. (1998). Antimicrobial peptides: an emerging concept in cutaneous biology. *J Invest Dermatol* **111** (5): 739-43.
- Galvin, J., Uryu, K., Lee, V. and Trojanowski, J. (1999). Axon pathology in Parkinson's disease and Lewy body dementia hippocampus contains alpha-, beta-, and gamma-synuclein. *Proc Natl Acad Sci* **96** (23): 13450-5.
- Garbisu, C., Alkorta, I., Llama, M. and Serra, J. (1998). Aerobic chromate reduction by *Bacillus subtilis*. *Biodegradation* **9** (2): 133-41.
- Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. and Bairoch, A. (2003). ExpASY: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* **31** (13): 3784-8.
- Gavathiotis, E., Suzuki, M., Davis, M., Pitter, K., Bird, G., Katz, S., Tu, H.-C., Kim, H., Cheng, E., Tjandra, N. and Walensky, L. (2008). BAX activation is initiated at a novel interaction site. *Nature* **455** (7216): 1076-81.
- Genentech. (1978, September 6). *First successful laboratory production of human insulin announced*. Retrieved July 4, 2011, from <http://www.gene.com/gene/news/press-releases/display.do?method=detail&id=4160>
- Gennaro, R. and Zanetti, M. (2000). Structural features and biological activities of the cathelicidin-derived antimicrobial peptides. *Biopolymers* **55** (1): 31-49.
- Gentschev, I., Dietrich, G. and Goebel, W. (2002). The E. coli alpha-hemolysin secretion system and its use in vaccine development. *Trends Microbiol* **10** (1): 39-45.
- Gifford, J., Hunter, H. and Vogel, H. (2005). Lactoferricin: a lactoferrin-derived peptide with antimicrobial, antiviral, antitumor and immunological properties. *Cell Mol Life Sci* **62** (22): 2588-98.
- Gladyshev, E., Meselson, M. and Arkhipova, I. (2008). Massive horizontal gene transfer in bdelloid rotifers. *Science* **320** (5880): 1210-3.

- Good, L., Sandberg, R., Larsson, O., Nielsen, P. and Wahlestedt, C. (2000). Antisense PNA effects in *Escherichia coli* are limited by the outer-membrane LPS layer. *Microbiology* **146** (10): 2665-70.
- Gordon, D., Sciarretta, K. and Meredith, S. (2001). Inhibition of beta-amyloid(40) fibrillogenesis and disassembly of beta-amyloid(40) fibrils by short beta-amyloid congeners containing N-methyl amino acids at alternate residues. *Biochemistry* **40** (28): 8237-45.
- Gottesman, S. (2004). The small RNA regulators of *Escherichia coli*: roles and mechanisms. *Annu Rev Microbiol* **58**: 303-28.
- Goyal, K., Tisi, L., Basran, A., Browne, J., Burnell, A., Zurdo, J. and Tunnacliffe, A. (2003). Transition from natively unfolded to folded state induced by desiccation in an anhydrobiotic nematode protein. *J Biol Chem* **278** (15): 12977-84.
- Goyal, K., Walton, L. and Tunnacliffe, A. (2005a). LEA proteins prevent protein aggregation due to water stress. *Biochem J* **388** (1), 151-7.
- Goyal, K., Pinelli, C., Maslen, S., Rastogi, R., Stephens, E. and Tunnacliffe, A. (2005b). Dehydration-regulated processing of late embryogenesis abundant protein in a desiccation-tolerant nematode. *FEBS Lett* **579** (19): 4093-8.
- Gravitz, L. (2011). Drugs: A tangled web of targets. *Nature* **475** (7355): S9-S11.
- Green, M. and Loewenstein, P. (1988). Autonomous functional domains of chemically synthesized human immunodeficiency virus tat trans-activator protein. *Cell* **55** (6): 1179-88.
- Guermeur, Y., Geourjon, C., Gallinari, P. and Deléage, G. (1999). Improved performance in protein secondary structure prediction by inhomogeneous score combination. *Bioinformatics* **15** (5): 413-21.
- Gumpert, J. and Hoischen, C. (1998). Use of cell wall-less bacteria (L-forms) for efficient expression and secretion of heterologous gene products. *Curr Opin Biotechnol* **9** (5): 506-9.
- Guzman, L., Belin, D., Carson, M. and Beckwith, J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose pBAD promoter. *J Bacteriol* **177** (14): 4121-30.
- Guzmán, F., Barberis, S. and Illanes, A. (2007). Peptide synthesis: chemical or enzymatic. *Electronic Journal of Biotechnology* **10** (2): 279-314.
- Hale, J. and Hancock, R. (2007). Alternative mechanisms of action of cationic antimicrobial peptides on bacteria. *Expert Review of Anti-infective Therapy* **5** (6): 951-9.
- Hancock, R. (1997). Peptide antibiotics. *Lancet* **349** (9049): 418-22.
- Hancock, R. and Sahl, H.-G. (2006). Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies. *Nat Biotechnol* **24** (12): 1551-7.
- Hanes, J. and Plückthun, A. (1997). In vitro selection and evolution of functional proteins by using ribosome display. *Proc Natl Acad Sci USA* **94** (10): 4937-42.
- Hardy, J. and Selkoe, D. (2002). The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science* **297** (5580): 353-6.
- Harris, A. (1994). Somatostatin and somatostatin analogues: pharmacokinetics and pharmacodynamic effects. *Gut* **35** (Suppl 3), S1-4.
- Hartmann, T., Bieger, S., Brühl, B., Tienari, P., Ida, N., Allsop, D., Roberts, G., Masters, C., Dotti, C., Unsicker, K. and Beyreuther, K. (1997). Distinct sites of intracellular production for Alzheimer's disease Abeta40/42 amyloid peptides. *Nat Med* **3** (9): 1016-20.
- Harwood, C. and Cranenburgh, R. (2008). Bacillus protein secretion: an unfolding story. *Trends Microbiol* **16** (2): 73-9.
- Haynie, S., Crum, G. and Doele, B. (1995). Antimicrobial activities of amphiphilic peptides covalently bonded to a water-insoluble resin. *Antimicrob Agents Chemother* **39** (2): 301-7.
- Heinis, C., Rutherford, T., Freund, S. and Winter, G. (2009). Phage-encoded combinatorial chemical libraries based on bicyclic peptides. *Nat Chem Biol* **5** (7): 502-7.
- Heipieper, H. and de Bont, J. (1994). Adaptation of *Pseudomonas putida* S12 to ethanol and toluene at the level of fatty acid composition of membranes. *Appl Environ Microbiol* **60** (12): 4440-4.
- Hilbich, C., Kisters-Woike, B., Reed, J., Masters, C. and Beyreuther, K. (1991). Aggregation and secondary structure of synthetic amyloid beta A4 peptides of Alzheimer's disease. *J Mol Biol* **218** (1): 149-63.
- Hilpert, K., Volkmer-Engert, R., Walter, T. and Hancock, R. (2005). High-throughput generation of small antibacterial peptides with improved activity. *Nat Biotechnol* **23** (8): 1008-12.

- Hilpert, K., Elliott, M., Volkmer-Engert, R., Henklein, P., Donini, O., Zhou, Q., Winkler, D. and Hancock, R. (2006). Sequence requirements and an optimization strategy for short antimicrobial peptides. *Chem Biol* **13** (10): 1101-7.
- Hilpert, K., Winkler, D. and Hancock, R. (2007). Peptide arrays on cellulose support: SPOT synthesis, a time and cost efficient method for synthesis of large numbers of peptides in a parallel and addressable fashion. *Nat Protoc* **2** (6): 1333-49.
- Hilpert, K., Elliott, M., Jenssen, H., Kindrachuk, J., Fjell, C., Körner, J. et al. and Hancock, R. (2009). Screening and characterization of surface-tethered cationic peptides for antimicrobial activity. *Chem Biol* **16** (1): 58-69.
- Hirata, Y. and Uemura, D. (1986). Halichondrins - antitumor polyether macrolides from a marine sponge. *Pure & Appl Chem* **58** (5): 701-710.
- Hirel, P., Schmitter, M., Dessen, P., Fayat, G. and Blanquet, S. (1989). Extent of N-terminal methionine excision from *Escherichia coli* proteins is governed by the side-chain length of the penultimate amino acid. *Proc Natl Acad Sci USA* **86** (21): 8247-51.
- Hofacker, I. (2003). Vienna RNA secondary structure server. *Nucleic Acids Res* **31** (13): 3429-31.
- Hoffman, L., D'Argenio, D., MacCoss, M., Zhang, Z., Jones, R. and Miller, S. (2005). Aminoglycoside antibiotics induce bacterial biofilm formation. *Nature* **436** (7054): 1171-5.
- Hong, I.-P., Lee, S.-J., Kim, Y.-S. and Choi, S.-G. (2007). Recombinant expression of human cathelicidin (hCAP18/LL-37) in *Pichia pastoris*. *Biotechnol Lett* **29** (1): 73-8.
- Hong, I.-P., Kim, Y.-S. and Choi, S.-G. (2010). Simple purification of human antimicrobial peptide dermcidin (MDCD-1L) by intein-mediated expression in *E. coli*. *J Microbiol Biotechnol* **20** (2): 350-355.
- Hsu, K.-H., Pei, C., Yeh, J.-Y., Shih, C.-H., Chung, Y.-C., Hung, L.-T. and Ou, B.-R. (2009). Production of bioactive human alpha-defensin 5 in *Pichia pastoris*. *J Gen Appl Microbiol* **55** (5): 395-401.
- Huber, P. (2005). *Robust statistics*. Chichester: John Wiley & Sons.
- Huntington's Disease Collaborative Research Group. (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* **72** (6): 971-83.
- Hwang, J., Zhong, X. and Tai, P. (1997). Interactions of dedicated export membrane proteins of the colicin V secretion system: CvaA, a member of the membrane fusion protein family, interacts with CvaB and TolC. *J Bacteriol* **179** (20): 6264-70.
- Ingham, A. and Moore, R. (2007). Recombinant production of antimicrobial peptides in heterologous microbial systems. *Biotechnol Appl Biochem* **47** (1): 1-9.
- International Human Genome Sequencing Consortium. (2004). Finishing the euchromatic sequence of the human genome. *Nature* **431** (7011): 931-45.
- Isaacs, F., Dwyer, D. and Collins, J. (2006). RNA synthetic biology. *Nat Biotechnol* **24** (5): 545-54.
- Ittner, L. and Götz, J. (2011). Amyloid- β and tau - a toxic pas de deux in Alzheimer's disease. *Nat Rev Neurosci* **12** (2): 65-72.
- Jackson, B., Wilhelmus, K. and Mitchell, B. (2007). Genetically regulated filamentation contributes to *Candida albicans* virulence during corneal infection. *Microb Pathog* **42** (2-3): 88-93.
- Jan, A., Gokce, O., Luthi-Carter, R. and Lashuel, H. (2008). The ratio of monomeric to aggregated forms of A β 40 and A β 42 is an important determinant of amyloid-beta aggregation, fibrillogenesis, and toxicity. *J Biol Chem* **283** (42): 28176-89.
- Jarrett, J., Berger, E. and Lansbury, P. (1993). The carboxy terminus of the beta amyloid protein is critical for the seeding of amyloid formation: implications for the pathogenesis of Alzheimer's disease. *Biochemistry* **32** (18): 4693-7.
- Jiang, Z., Vasil, A., Hale, J., Hancock, R., Vasil, M. and Hodges, R. (2008). Effects of net charge and the number of positively charged residues on the biological activity of amphipathic alpha-helical cationic antimicrobial peptides. *Biopolymers* **90** (3): 369-83.
- Johnson, I. (1983). Human insulin from recombinant DNA technology. *Science* **219** (4585): 632-7.
- Joliot, A., Pernelle, C., Deagostini-Bazin, H. and Prochiantz, A. (1991). Antennapedia homeobox peptide regulates neural morphogenesis. *Proc Natl Acad Sci USA* **88** (5): 1864-8.
- Jones, C., Dexter, P., Evans, A., Liu, C., Hultgren, S. and Hruby, D. (2002). *Escherichia coli* DegP protease cleaves between paired hydrophobic residues in a natural substrate: the PapA pilin. *J Bacteriol* **184** (20): 5762-71.

- Jong, W., Saurí, A. and Luirink, J. (2010). Extracellular production of recombinant proteins using bacterial autotransporters. *Curr Opin Biotechnol* **21** (5): 646-52.
- Joung, J., Ramm, E. and Pabo, C. (2000). A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions. *Proc Natl Acad Sci USA* **97** (13): 7382-7.
- Justice, S., Hunstad, D., Cegelski, L. and Hultgren, S. (2008). Morphological plasticity as a bacterial survival strategy. *Nat Rev Microbiol* **6** (2): 162-8.
- Katsoyannis, P. and Tometsko, A. (1966). Insulin synthesis by recombination of A and B chains: a highly efficient method. *Proc Natl Acad Sci USA* **55** (6): 1554-61.
- Katzen, F., Chang, G. and Kudlicki, W. (2005). The past, present and future of cell-free protein synthesis. *Trends Biotechnol* **23** (3): 150-6.
- Kawashima, H., Horii, T., Ogawa, T. and Ogawa, H. (1984). Functional domains of Escherichia coli recA protein deduced from the mutational sites in the gene. *Mol Gen Genet* **193** (2): 288-92.
- Kayed, R., Head, E., Thompson, J., McIntire, T., Milton, S., Cotman, C. and Glabe, C. (2003). Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. *Science* **300** (5618): 486-9.
- Kaykas, A. and Moon, R. (2004). A plasmid-based system for expressing small interfering RNA libraries in mammalian cells. *BMC Cell Biol* **5**: 16.
- Kelly, J., Rubin, A., Davis, J., Ajo-Franklin, C., Cumbers, J., Czar, M., de Mora, K., Gliberman, A., Monie, D. and Endy, D. (2009). Measuring the activity of BioBrick promoters using an in vivo reference standard. *J Biol Eng* **3** (4): 1-13.
- Kilby, J., Hopkins, S., Venetta, T., DiMassimo, B., Cloud, G., Lee, J., Alldredge, L., Hunter, E., Lambert, D., Bolognesi, D., Matthews, T., Johnson, M., Nowak, M., Shaw, G. and Saag, M. (1998). Potent suppression of HIV-1 replication in humans by T-20, a peptide inhibitor of gp41-mediated virus entry. *Nat Med* **4** (11): 1302-7.
- Kim, W., Kim, Y., Min, J., Kim, D., Chang, Y.-T. and Hecht, M. (2006). A high-throughput screen for compounds that inhibit aggregation of the Alzheimer's peptide. *ACS Chem Biol* **1** (7): 461-9.
- Kim, Y. and Cha, H. (2010). Disperse distribution of cationic amino acids on hydrophilic surface of helical wheel enhances antimicrobial peptide activity. *Biotechnol Bioeng* **107** (2): 216-23.
- Kjaergaard, K., Schembri, M., Hasman, H. and Klemm, P. (2000). Antigen 43 from Escherichia coli induces inter- and intraspecies cell aggregation and changes in colony morphology of Pseudomonas fluorescens. *J Bacteriol* **182** (17): 4789-96.
- Klocke, M., Mundt, K., Idler, F., Jung, S. and Backhausen, J. (2005). Heterologous expression of enterocin A, a bacteriocin from Enterococcus faecium, fused to a cellulose-binding domain in Escherichia coli results in a functional protein with inhibitory activity against Listeria. *Appl Microbiol Biotechnol* **67** (4): 532-8.
- Knowles, M., Robinson, J., Wood, R., Pue, C., Mentz, W., Wager, G., Gatzky, J. and Boucher, R. (1997). Ion composition of airway surface liquid of patients with cystic fibrosis as compared with normal and disease-control subjects. *J Clin Invest* **100** (10): 2588-95.
- Kokkoni, N., Stott, K., Amijee, H., Mason, J. and Doig, A. (2006). N-Methylated peptide inhibitors of beta-amyloid aggregation and toxicity. Optimization of the inhibitor structure. *Biochemistry* **45** (32): 9906-18.
- Krahulec, J., Hyřšová, M., Pepeliaev, S., Jílková, J., Cerný, Z. and Macháľková, J. (2010). High level expression and purification of antimicrobial human cathelicidin LL-37 in Escherichia coli. *Appl Microbiol Biotechnol* **88** (1): 167-75.
- Kubinyi, H. (1999). Chance favors the prepared mind - from serendipity to rational drug design. *J Recept Signal Transduct Res* **19** (1-4): 15-39.
- Kubinyi, H. (2003). Drug research: myths, hype and reality. *Nat Rev Drug Discov* **2** (8): 665-8.
- Kwok, R. (2010). Five hard truths for synthetic biology. *Nature* **463** (7279): 288-90.
- Laemmli, U. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227** (5259): 680-5.
- Lande, R., Gregorio, J., Facchinetti, V., Chatterjee, B., Wang, Y.-H., Homey, B., Cao, W. et al. and Gilliet, M. (2007). Plasmacytoid dendritic cells sense self-DNA coupled with antimicrobial peptide. *Nature* **449** (7162): 564-9.
- Lansbury, P. (2001). Following nature's anti-amyloid strategy. *Nat Biotechnol* **19** (2): 112-3.
- Larkin, M., Blackshields, G., Brown, N., Chenna, R., McGettigan, P., McWilliam, H. et al. and Higgins, D. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23** (21): 2947-8.

- Lau, Y., Rozek, A., Scott, M., Goosney, D., Davidson, D. and Hancock, R. (2005). Interaction and cellular localization of the human host defense peptide LL-37 with lung epithelial cells. *Infect Immun* **73** (1): 583-91.
- Lax, R. (2010). The future of peptide development in the pharmaceutical industry. *PharManufacturing: The International Peptide Review* (Sept): 10-15.
- Leader, B., Baca, Q. and Golan, D. (2008). Protein therapeutics: A summary and pharmacological classification. *Nat Rev Drug Discov* **7** (1): 21-39.
- Lederberg, J. and Lederberg, E. (1952). Replica plating and indirect selection of bacterial mutants. *J Bacteriol* **63** (3): 399-406.
- Lee, S.-H., Lee, S., Youn, Y., Na, D., Chae, S., Byun, Y. and Lee, K. (2005a). Synthesis, characterization, and pharmacokinetic studies of PEGylated glucagon-like peptide-1. *Bioconjug Chem* **16** (2): 377-82.
- Lee, C., Williams, T., Wong, D. and Robertson, G. (2005b). An episomal expression vector for screening mutant gene libraries in *Pichia pastoris*. *Plasmid* **54** (1): 80-5.
- Lee, L., Ha, H., Chang, Y.-T. and DeLisa, M. (2009). Discovery of amyloid-beta aggregation inhibitors using an engineered assay for intracellular protein folding and solubility. *Protein Sci* **18** (2): 277-86.
- Leptihn, S., Har, J., Chen, J., Ho, B., Wohland, T. and Ding, J. (2009). Single molecule resolution of the antimicrobial action of quantum dot-labeled sushi peptide on live bacteria. *BMC Biol* **7** (22).
- Lesné, S., Koh, M., Kotilinek, L., Kaye, R., Glabe, C., Yang, A., Gallagher, M. and Ashe, K. (2006). A specific amyloid-beta protein assembly in the brain impairs memory. *Nature* **440** (7082): 352-7.
- LeVine, H. (1993). Thioflavine T interaction with synthetic Alzheimer's disease beta-amyloid peptides: detection of amyloid aggregation in solution. *Protein Sci* **2** (3): 404-10.
- Lewis, K. (2007). Persister cells, dormancy and infectious disease. *Nat Rev Microbiol* **5** (1): 48-56.
- Li, C., Peters, A., Meredith, E., Allman, G. and Savage, P. (1998). Design and synthesis of potent sensitizers of Gram-negative bacteria based on a cholic acid scaffolding. *J Am Chem Soc* **120** (12): 2961-2.
- Li, X., Li, Y., Han, H., Miller, D., Wang, G. (2006). Solution structures of human LL-37 fragments and NMR-based identification of a minimal membrane-targeting antimicrobial and anticancer region. *J Am Chem Soc* **128** (17): 5776-85.
- Li, J.-H. and Vederas, J. (2009). Drug discovery and natural products: end of an era or an endless frontier? *Science* **325** (5937): 161-5.
- Li, C., Blencke, H.-M., Paulsen, V., Haug, T. and Stensvåg, K. (2010). Powerful workhorses for antimicrobial peptide expression and characterization. *Bioengineered Bugs* **1** (3): 217-220.
- Lien, S. and Lowman, H. (2003). Therapeutic peptides. *Trends Biotechnol* **21** (12): 556-62.
- Lin, W., Fullner, K., Clayton, R., Sexton, J., Rogers, M., Calia, K., Calderwood, S., Fraser, C. and Mekalanos, J. (1999). Identification of a vibrio cholerae RTX toxin gene cluster that is tightly linked to the cholera toxin prophage. *Proc Natl Acad Sci USA* **96** (3): 1071-6.
- Lindgren, M., Hällbrink, M., Prochiantz, A. and Langel, U. (2000). Cell-penetrating peptides. *Trends Pharmacol Sci* **21** (3): 99-103.
- Lipinski, C., Lombardo, F., Dominy, B. and Feeney, P. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* **46** (1-3): 3-26.
- Lipsky, B., Holroyd, K. and Zasloff, M. (2008). Topical versus systemic antimicrobial therapy for treating mildly infected diabetic foot ulcers: a randomized, controlled, double-blinded, multicenter trial of pexiganan cream. *Clin Infect Dis* **47** (12): 1537-45.
- Loffet, A. (2002). Peptides as drugs: is there a market? *J Pept Sci* **8** (1): 1-7.
- Loit, E., Wu, K., Cheng, X., Hincke, M. and Altosaar, I. (2008). Functional whole-colony screening method to identify antimicrobial peptides. *J Microbiol Methods* **75** (3): 425-31.
- Loose, C., Jensen, K., Rigoutsos, I. and Stephanopoulos, G. (2006). A linguistic model for the rational design of antimicrobial peptides. *Nature* **443** (7113): 867-9.
- Luheshi, L., Tartaglia, G., Brorsson, A.-C., Pawar, A., Watson, I., Chiti, F., Vendruscolo, M., Lomas, D., Dobson, C. and Crowther, D. (2007). Systematic in vivo analysis of the intrinsic determinants of amyloid Beta pathogenicity. *PLoS Biol* **5** (11): e290.
- Lührs, T., Ritter, C., Adrian, M., Riek-Loher, D., Bohrmann, B., Döbeli, H., Schubert, D. and Riek, R. (2005). 3D structure of Alzheimer's amyloid-beta(1-42) fibrils. *Proc Natl Acad Sci USA* **102** (48): 17342-7.

- Luque-Ortega, J., van't Hof, W., Veerman, E., Saugar, J. and Rivas, L. (2008). Human antimicrobial peptide histatin 5 is a cell-penetrating peptide targeting mitochondrial ATP synthesis in *Leishmania*. *FASEB J* **22** (6): 1817-28.
- Macilwain, C. (1998). When rhetoric hits reality in debate on bioprospecting. *Nature* **392** (6676): 535-40.
- Mahenthiralingam, E., Urban, T. and Goldberg, J. (2005). The multifarious, multireplicon *Burkholderia cepacia* complex. *Nat Rev Microbiol* **3** (2): 144-56.
- Malten, M., Biedendieck, R., Gamer, M., Drews, A.-C., Stammen, S., Buchholz, K., Dijkhuizen, L. and Jahn, D. (2006). A *Bacillus megaterium* plasmid system for the production, export, and one-step purification of affinity-tagged heterologous levansucrase from growth medium. *Appl Environ Microbiol* **72** (2): 1677-9.
- Margulies, M., Egholm, M., Altman, W., Attiya, S., Bader, J., Bemben, L. et al. and Rothberg, J. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437** (7057): 376-80.
- Marr, A., Gooderham, W. and Hancock, R. (2006). Antibacterial peptides for therapeutic use: obstacles and realistic outlook. *Curr Opin Pharmacol* **6** (5): 468-72.
- Matsuzaki, K. (2009). Control of cell selectivity of antimicrobial peptides. *Biochim Biophys Acta* **1788** (8): 1687-92.
- Maurizi, M. (1992). Proteases and protein degradation in *Escherichia coli*. *Experientia* **48** (2): 178-201.
- McGregor, D. (2008). Discovering and improving novel peptide therapeutics. *Curr Opin Pharmacol* **8** (5): 616-9.
- McInnes, C. (2007). Virtual screening strategies in drug discovery. *Curr Opin Chem Biol* **11** (5): 494-502.
- McLaurin, J., Kierstead, M., Brown, M., Hawkes, C., Lambermon, M., Phinney, A. et al. and St George-Hyslop, P. (2006). Cyclohexanehexol inhibitors of Aβ aggregation prevent and reverse Alzheimer phenotype in a mouse model. *Nat Med* **12** (7): 801-8.
- Medina, C., Camacho, E., Flores, A., Mesa-Pereira, B. and Santero, E. (2011). Improved expression systems for regulated expression in *Salmonella* infecting eukaryotic cells. *PLoS ONE* **6** (8): e23055.
- Melo, M., Ferre, R. and Castanho, M. (2009). Antimicrobial peptides: linking partition, activity and high membrane-bound concentrations. *Nat Rev Microbiol* **7** (3): 245-50.
- Merrifield, R. (1963). Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J Am Chem* **85** (14): 2149-2154.
- Miller, O., Bernath, K., Agresti, J., Amitai, G., Kelly, B., Mastrobattista, E., Taly V., Magdassi, S., Tawfik, D. and Griffiths, D. (2006). Directed evolution by in vitro compartmentalization. *Nat Methods* **3** (7): 561-70.
- Moellering, R., Cornejo, M., Davis, T., Del Bianco, C., Aster, J., Blacklow, S., Kung, A., Gilliland D., Verdine, G. and Bradner, J. (2009). Direct inhibition of the NOTCH transcription factor complex. *Nature* **462** (7270): 182-8.
- Monaghan, R. and Barrett, J. (2006). Antibacterial drug discovery - then, now and the genomics future. *Biochem Pharmacol* **71** (7): 901-9.
- Moon, J.-Y., Henzler-Wildman, K. and Ramamoorthy, A. (2006). Expression and purification of a recombinant LL-37 from *Escherichia coli*. *Biochim Biophys Acta* **1758** (9): 1351-8.
- Morita, T., Mochizuki, Y. and Aiba, H. (2006). Translational repression is sufficient for gene silencing by bacterial small noncoding RNAs in the absence of mRNA destruction. *Proc Natl Acad Sci USA* **103** (13): 4858-63.
- Mulvey, M. and Simor, A. (2009). Antimicrobial resistance in hospitals: how concerned should we be? *CMAJ* **180** (4): 408-15.
- Muratovska, A. and Eccles, M. (2004). Conjugate for efficient delivery of short interfering RNA (siRNA) into mammalian cells. *FEBS Lett* **558** (1-3): 63-8.
- Mygind, P., Fischer, R., Schnorr, K., Hansen, M., Sönksen, C., Ludvigsen, S. et al. and Kristensen, H.-H. (2005). Plectasin is a peptide antibiotic with therapeutic potential from a saprophytic fungus. *Nature* **437** (7061): 975-80.
- Nagai, Y., Tucker, T., Ren, H., Kenan, D., Henderson, B., Keene, J., Strittmatter, W. and Burke, J. (2000). Inhibition of polyglutamine protein aggregation and cell death by novel peptides identified by phage display screening. *J Biol Chem* **275** (14): 10437-42.
- Nakamura, Y., Gojobori, T. and Ikemura, T. (2000). Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res* **28** (1): 292.
- Natale, P., Brüser, T. and Driessen, A. (2008). Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane - distinct translocases and mechanisms. *Biochim Biophys Acta* **1778** (9): 1735-56.

- Nguyen, L., Haney, E. and Vogel, H. (2011). The expanding scope of antimicrobial peptide structures and their modes of action. *Trends Biotechnol* **29** (9): 464-72.
- Ni, Y. and Chen, R. (2009). Extracellular recombinant protein production from *Escherichia coli*. *Biotechnol Lett* **31** (11): 1661-70.
- Nizet, V., Ohtake, T., Lauth, X., Trowbridge, J., Rudisill, J., Dorschner, R., Psetonjamas, V., Piraino, J., Huttner, K. and Gallo, R. (2001). Innate antimicrobial peptide protects the skin from invasive bacterial infection. *Nature* **414** (6862): 454-7.
- Odegrip, R., Coomber, D., Eldridge, B., Hederer, R., Kuhlman, P., Ullman, C., FitzGerald and McGregor, D. (2004). CIS display: In vitro selection of peptides from libraries of protein-DNA complexes. *Proc Natl Acad Sci USA* **101** (9): 2806-10.
- Oren, Z. and Shai, Y. (1998). Mode of action of linear amphipathic alpha-helical antimicrobial peptides. *Biopolymers* **47** (6): 451-63.
- Palasek, S., Cox, Z. and Collins, J. (2007). Limiting racemization and aspartimide formation in microwave-enhanced Fmoc solid phase peptide synthesis. *J Pept Sci* **13** (3): 143-8.
- Pallitto, M., Ghanta, J., Heinzelman, P., Kiessling, L. and Murphy, R. (1999). Recognition sequence design for peptidyl modulators of beta-amyloid aggregation and toxicity. *Biochemistry* **38** (12): 3570-8.
- Park, C., Kim, H. and Kim, S. (1998). Mechanism of action of the antimicrobial peptide buforin II: buforin II kills microorganisms by penetrating the cell membrane and inhibiting cellular functions. *Biochem Biophys Res Commun* **244** (1): 253-7.
- Park, S. and Raines, R. (2000). Genetic selection for dissociative inhibitors of designated protein-protein interactions. *Nat Biotechnol* **18** (8): 847-51.
- Park, K., Shin, S., Hahm, K. and Kim, Y. (2003). Structural and functional characterization of CRAMP-18 derived from a cathelicidin-related antimicrobial peptide CRAMP. *B Kor Chem Soc* **24** (10): 1478-84.
- Patrzykat, A., Friedrich, C., Zhang, L., Mendoza, V. and Hancock, R. (2002). Sublethal concentrations of pleurocidin-derived antimicrobial peptides inhibit macromolecular synthesis in *Escherichia coli*. *Antimicrob Agents Chemother* **46** (3): 605-14.
- Pecota, D., Osapay, G., Selsted, M. and Wood, T. (2003). Antimicrobial properties of the *Escherichia coli* R1 plasmid host killing peptide. *J Biotechnol* **100** (1): 1-12.
- Perron, G., Zasloff, M. and Bell, G. (2006). Experimental evolution of resistance to an antimicrobial peptide. *Proc Biol Sci* **273** (1583): 251-6.
- Peschel, A., Otto, M., Jack, R., Kalbacher, H., Jung, G. and Götz, F. (1999). Inactivation of the *dlt* operon in *Staphylococcus aureus* confers sensitivity to defensins, protegrins, and other antimicrobial peptides. *J Biol Chem* **274** (13): 8405-10.
- Peschel, A. and Sahl, H.-G. (2006). The co-evolution of host cationic antimicrobial peptides and microbial resistance. *Nat Rev Microbiol* **4** (7): 529-36.
- Pichereau, C. and Allary, C. (2006). Therapeutic peptides under the spotlight. *European Biopharmaceutical Review* (Winter): 88-91.
- Pierce, M., Raman, C. and Nall, B. (1999). Isothermal titration calorimetry of protein-protein interactions. *Methods* **19** (2): 213-21.
- Pike, C., Burdick, D., Walencewicz, A., Glabe, C. and Cotman, C. (1993). Neurodegeneration induced by beta-amyloid peptides in vitro: the role of peptide assembly state. *J Neurosci* **13** (4): 1676-87.
- Pinkart, H. and White, D. (1997). Phospholipid biosynthesis and solvent tolerance in *Pseudomonas putida* strains. *J Bacteriol* **179** (13): 4219-26.
- Plempner, R. and Wolf, D. (1999). Retrograde protein translocation: ERADication of secretory proteins in health and disease. *Trends Biochem Sci* **24** (7): 266-70.
- Prochazkova, K. and Satchell, K. (2008). Structure-function analysis of inositol hexakisphosphate-induced autoprocessing of the *Vibrio cholerae* multifunctional autoprocessing RTX toxin. *J Biol Chem* **283** (35): 23656-64.
- Prusiner, S. (1998). Prions. *Proc Natl Acad Sci USA* **95** (23): 13363-83.
- Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M. and Séraphin, B. (2001). The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* **24** (3): 218-29.
- Quail, M. (2005). DNA: mechanical breakage. In *Encyclopedia of Life Sciences* (pp. 1-4). Chichester: John Wiley & Sons. Retrieved January 12, 2009, from <http://www.els.net>, doi:10.1038/npg.els.0005333

- Quan, J. and Tian, J. (2009). Circular polymerase extension cloning of complex gene libraries and pathways. *PLoS ONE* **4** (7): e6441.
- Ramanathan, B., Davis, E., Ross, C. and Blecha, F. (2002). Cathelicidins: microbicidal activity, mechanisms of action, and roles in innate immunity. *Microbes Infect* **4** (3): 361-72.
- Ramon, J., Saez, V., Baez, R., Aldana, R. and Hardy, E. (2005). PEGylated interferon- α 2b: a branched 40K polyethylene glycol derivative. *Pharm Res* **22** (8): 1374-86.
- Rapoza, M. and Webster, R. (1993). The filamentous bacteriophage assembly proteins require the bacterial SecA protein for correct localization to the membrane. *J Bacteriol* **175** (6): 1856-9.
- Rathinakumar, R., Walkenhorst, W. and Wimley, W. (2009). Broad-spectrum antimicrobial peptides by rational combinatorial design and high-throughput screening: the importance of interfacial activity. *J Am Chem Soc* **131** (22): 7609-17.
- Raventós, D., Taboureau, O., Mygind, P., Nielsen, J., Sonksen, C. and Kristensen, H.-H. (2005). Improving on nature's defenses: optimization & high throughput screening of antimicrobial peptides. *Comb Chem High Throughput Screen* **8** (3): 219-33.
- Reichert, J. and Wenger, J. (2008). Development trends for new cancer therapeutics and vaccines. *Drug Discov Today* **13** (1-2): 30-7.
- Reichert, J., Saladin, P., Seckler, P., Riviere, P., Tartar, A. and Dunn, K. (2008). *Development trends for peptide therapeutics*. San Diego: Peptide Therapeutics Foundation. Retrieved June 29, 2011, from <http://www.peptidetherapeutics.org/>
- Reichert, J., Pechon, P., Tartar, A. and Dunn, M. (2010). *Development trends for peptide therapeutics*. San Diego: Peptide Therapeutics Foundation. Retrieved June 1, 2011, from <http://www.peptidetherapeutics.org/>
- Rice, P., Longden, I. and Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* **16** (6): 276-7.
- Rich, R. and Myszkka, D. (2000). Advances in surface plasmon resonance biosensor analysis. *Curr Opin Biotechnol* **11** (1): 54-61.
- Riechmann, L. and Winter, G. (2000). Novel folded protein domains generated by combinatorial shuffling of polypeptide segments. *Proc Natl Acad Sci USA* **97** (18): 10068-73.
- Ringquist, S., Shinedling, S., Barrick, D., Green, L., Binkley, J., Stormo, G. and Gold, L. (1992). Translation initiation in Escherichia coli: sequences within the ribosome-binding site. *Mol Microbiol* **6** (9): 1219-29.
- Ro, D.-K., Paradise, E., Ouellet, M., Fisher, K., Newman, K., Ndungu, J., Ho, K., Eachus, R., Ham, T., Kirby, J., Chang, M., Withers, S., Shiba, Y., Sarpong, R. and Keasling, J. (2006). Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440** (7086): 940-3.
- Roberts, R. and Szostak, J. (1997). RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proc Natl Acad Sci USA* **94** (23): 12297-302.
- Roopenian, D. and Akilesh, S. (2007). FcRn: the neonatal Fc receptor comes of age. *Nat Rev Immunol* **7** (9): 715-25.
- Rosenbloom, A. (2006). Is there a role for recombinant insulin-like growth factor-I in the treatment of idiopathic short stature? *Lancet* **368** (9535): 612-6.
- Rossi, L., Rangasamy, P., Zhang, J., Qiu, X. and Wu, G. (2007). Research advances in the development of peptide antibiotics. *J Pharm Sci* **97** (3): 1060-70.
- Roth, M., Tomlinson, B. and Blessed, G. (1966). Correlation between scores for dementia and counts of 'senile plaques' in cerebral grey matter of elderly subjects. *Nature* **209** (5018): 109-10.
- Ruiz, N., Kahne, D. and Silhavy, T. (2006). Advances in understanding bacterial outer-membrane biogenesis. *Nat Rev Microbiol* **4** (1): 57-66.
- Saag, K., Shane, E., Boonen, S., Marín, F., Donley, D., Taylor, K., Dalsky, G. and Marcus, R. (2007). Teriparatide or alendronate in glucocorticoid-induced osteoporosis. *N Engl J Med* **357** (20): 2028-39.
- Saier, M. (2006). Protein secretion and membrane insertion systems in gram-negative bacteria. *J Membr Biol* **214** (2): 75-90.
- Saiman, L., Tabibi, S., Starner, T., San Gabriel, P., Winokur, P., Jia, H., McCray, P. and Tack, B. (2001). Cathelicidin peptides inhibit multiply antibiotic-resistant pathogens from patients with cystic fibrosis. *Antimicrob Agents Chemother* **45** (10): 2838-44.
- Sandgren, S., Wittrup, A., Cheng, F., Jönsson, M., Eklund, E., Busch, S. and Belting, M. (2004). The human antimicrobial peptide LL-37 transfers extracellular DNA plasmid to the nuclear compartments of mammalian cells via lipid rafts and proteoglycan-dependent endocytosis. *J Biol Chem* **279** (17): 17951-6.

- Sanger, F. (1959). Chemistry of insulin. *Science* **159** (3359): 1340-4.
- Sambrook, J. and Russell, D. (2001). *Molecular Cloning: A Laboratory Manual* (3rd ed.). Cold Spring Harbor: Cold Spring Harbor Laboratory Press.
- Samuelsen, O., Haukland, H., Jenssen, H., Krämer, M., Sandvik, K., Ulvatne, H. and Vorland, L. (2005). Induced resistance to the antimicrobial peptide lactoferricin B in *Staphylococcus aureus*. *FEBS Lett* **579** (16): 3421-6.
- Satchell, K. (2007). MARTX, multifunctional autoprocessing repeats-in-toxin toxins. *Infect Immun* **75** (11): 5079-84.
- Sato, A., Viswanathan, M., Kent, R. and Wood, C. (2006a). Therapeutic peptides: technological advances driving peptides into development. *Curr Opin Biotechnol* **17** (6): 638-42.
- Sato, T., Kienlen-Campard, P., Ahmed, M., Liu, W., Li, H., Elliott, J., Aimoto, S., Constantinescu, S., Octave, J.-N. and Smith, S. (2006b). Inhibitors of amyloid toxicity based on beta-sheet packing of A β 40 and A β 42. *Biochemistry* **45** (17): 5503-16.
- Sawyer, J., Martin, N. and Hancock, R. (1988). Interaction of macrophage cationic proteins with the outer membrane of *Pseudomonas aeruginosa*. *Infect Immun* **56** (3), 693-8.
- Schafmeister, C. E., Po, J. and Verdine, G. L. (2000). An all-hydrocarbon crosslinking system for enhancing the helicity and metabolic stability of peptides. *J Am Chem Soc* **122** (24): 5891-2.
- Schägger, H. (2006). Tricine-SDS-PAGE. *Nat Protoc* **1** (1): 16-22.
- Schena, M., Shalon, D., Davis, R. and Brown, P. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270** (5235): 467-70.
- Schiffer, M. and Edmundson, A. (1967). Use of helical wheels to represent the structures of proteins and to identify segments with helical potential. *Biophys J* **7** (2): 121-35.
- Schlieker, C., Bukau, B. and Mogk, A. (2002). Prevention and reversion of protein aggregation by molecular chaperones in the *E. coli* cytosol: implications for their applicability in biotechnology. *J Biotechnol* **96** (1): 13-21.
- Schnabel, J. (2011). Amyloid: Little proteins, big clues. *Nature* **475** (7355): S12-4.
- Schoeman, H., Vivier, M., Du Toit, M., Dicks, L. and Pretorius, I. (1999). The development of bactericidal yeast strains by expressing the *Pediococcus acidilactici* pediocin gene (*pedA*) in *Saccharomyces cerevisiae*. *Yeast* **15** (8): 647-56.
- Schreiber, S. (2000). Target-oriented and diversity-oriented organic synthesis in drug discovery. *Science* **287** (5460): 1964-9.
- Schwarze, S., Ho, A., Vocero-Akbani, A. and Dowdy, S. (1999). In vivo protein transduction: delivery of a biologically active protein into the mouse. *Science* **285** (5433): 1569-72.
- Scott, M., Dullaghan, E., Mookherjee, N., Glavas, N., Waldbrook, M. et al. and Hancock, R. (2007). An anti-infective peptide that selectively modulates the innate immune response. *Nat Biotechnol* **25** (4): 465-72.
- Sebbage, V. (2009). Cell-penetrating peptides and their therapeutic applications. *Bioscience Horizons* **2** (1): 64-72.
- Selkoe, D. (2002). Alzheimer's disease is a synaptic failure. *Science* **298** (5594): 789-91.
- Sewald, N. and Jakubke, H.-D. (2009). *Peptides: Chemistry and Biology* (2nd ed.). Weinheim: Wiley-VCH.
- Shaner, N., Campbell, R., Steinbach, P., Giepmans, B., Palmer, A. and Tsien, R. (2004). Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nat Biotechnol* **22** (12): 1567-72.
- Shankar, G., Li, S., Mehta, T., Garcia-Munoz, A., Shepardson, N., Smith, I. et al. and Selkoe, D. (2008). Amyloid-beta protein dimers isolated directly from Alzheimer's brains impair synaptic plasticity and memory. *Nat Med* **14** (8): 837-42.
- Shapiro, H. M. (2003). *Practical Flow Cytometry* (4th ed.). Chichester: John Wiley & Sons.
- Shen, A., Lupardus, P., Morell, M., Ponder, E., Sadaghiani, A., Garcia, K. and Bogoy, M. (2009a). Simplified, enhanced protein purification using an inducible, autoprocessing enzyme tag. *PLoS ONE* **4** (12): e8119.
- Shen, A., Lupardus, P., Albrow, V., Guzzetta, A., Powers, J., Garcia, K. and Bogoy, M. (2009b). Mechanistic and structural insights into the proteolytic activation of *Vibrio cholerae* MARTX toxin. *Nat Chem Biol* **5** (7): 469-78.
- Sheridan, C. (2009). J&J's billion dollar punt on anti-amyloid antibody. *Nat Biotechnol* **27** (8): 679-81.

- Shin, S., Kang, S., Lee, D., Eom, S., Song, W. and Kim, J. (2000). CRAMP analogues having potent antibiotic activity against bacterial, fungal, and tumor cells without hemolytic activity. *Biochem Biophys Res Commun* **275** (3): 904-9.
- Shoichet, B. (2004). Virtual screening of chemical libraries. *Nature* **432** (7019): 862-5.
- Simon, R., Kania, R., Zuckermann, R., Huebner, V., Jewell, D., Banville, S., Ng, S., Wang, L., Rosenberg, S. and Marlowe, C. (1992). Peptoids: a modular approach to drug discovery. *Proc Natl Acad Sci USA* **89** (20): 9367-71.
- Simonetti, G., Baffa, S. and Simonetti, N. (2001). Contact imidazole activity against resistant bacteria and fungi. *Int J Antimicrob Agents* **17** (5): 389-93.
- Singh, J., Whitwill, S., Lacroix, G., Douglas, J., Dubuc, E., Allard, G., Keller W. and Scherthaner, J. (2009). The use of Group 3 LEA proteins as fusion partners in facilitating recombinant expression of recalcitrant proteins in *E. coli*. *Protein Expr Purif* **67** (1): 15-22.
- Skarnes, R. C. and Watson, D. W. (1957). Antimicrobial factors of normal tissues and fluids. *Bacteriol Rev* **21** (4): 273-94.
- Smith, G. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228** (4705): 1315-7.
- Sneader, W. (2001). History of Insulin. In *Encyclopedia of Life Sciences*. Chichester: John Wiley & Sons. Retrieved July 5, 2011, from <http://www.els.net>, doi:10.1038/npg.els.0003623
- Sneader, W. (2005). *Drug discovery: a history*. Chichester: John Wiley & Sons.
- Sochacki, K., Barns, K., Bucki, R. and Weisshaar, J. (2011). Real-time attack on single *Escherichia coli* cells by the human antimicrobial peptide LL-37. *Proc Natl Acad Sci USA* **108** (16): e77-81.
- Sogin, M., Morrison, H., Huber, J., Welch, D., Huse, S., Neal, P., Arrieta, J. and Herndl, G. (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc Natl Acad Sci USA* **103** (32): 12115-20.
- Soscia, S., Kirby, J., Washicosky, K., Tucker, S., Ingelsson, M., Hyman, B., Burton, A., Goldstein, L., Duong, S., Tanzi, R. and Moir, R. (2010). The Alzheimer's Disease-Associated Amyloid beta-Protein Is an Antimicrobial Peptide. *PLoS ONE* **5** (3): e9505.
- Soto, C. and Estrada, L. (2005). Amyloid inhibitors and beta-sheet breakers. *Subcell Biochem* **38**: 351-64.
- Steiner, D. and Oyer, P. (1967). The biosynthesis of insulin and a probable precursor of insulin by a human islet cell adenoma. *Proc Natl Acad Sci USA* **57** (2): 473-80.
- Steiner, D., Park, S.-Y., Støy, J., Philipson, L. and Bell, G. (2009). A brief perspective on insulin production. *Diabetes Obes Metab* **11** (S4): 189-96.
- Steinstraesser, L., Tack, B., Waring, A., Hong, T., Boo, L., Fan, M.-H., Remick, D., Su, G., Lehrer, R. and Wang, S. (2002). Activity of novispirin G10 against *Pseudomonas aeruginosa* in vitro and in infected burns. *Antimicrob Agents Chemother* **46** (6): 1837-44.
- Stemmer, W., Cramer, A., Ha, K., Brennan, T. and Heyneker, H. (1995). Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* **164** (1): 49-53.
- Stewart, K., Horton, K. and Kelley, S. (2008). Cell-penetrating peptides as delivery vehicles for biology and medicine. *Org Biomol Chem* **6** (13): 2242-55.
- Structural Genomics Consortium, China Structural Genomics Consortium, Northeast Structural Genomics Consortium, Gräslund, S., Nordlund, P., Weigelt, J. et al. and Gunsalus, K. (2008). Protein production and purification. *Nat Methods* **5** (2): 135-46.
- Stumpe, S., Schmid, R., Stephens, D., Georgiou, G. and Bakker, E. (1998). Identification of OmpT as the protease that hydrolyzes the antimicrobial peptide protamine before it enters growing cells of *Escherichia coli*. *J Bacteriol* **180** (15): 4002-6.
- Subbalakshmi, C. and Sitaram, N. (1998). Mechanism of antimicrobial action of indolicidin. *FEMS Microbiol Lett* **160** (1): 91-6.
- Taguchi, S., Nakagawa, K., Maeno, M. and Momose, H. (1994). In vivo monitoring system for structure-function relationship analysis of the antibacterial peptide apidaecin. *Appl Environ Microbiol* **60** (10): 3566-72.
- Taguchi, S., Kuwasako, K., Suenaga, A., Okada, M. and Momose, H. (2000). Functional mapping against *Escherichia coli* for the broad-spectrum antimicrobial peptide, thanatin, based on an in vivo monitoring assay system. *J Biochem* **128** (5): 745-54.
- Tanaka, T., Kokuryu, Y. and Matsunaga, T. (2008). Novel method for selection of antimicrobial peptides from a phage display library by use of bacterial magnetic particles. *Appl Environ Microbiol* **74** (24): 7600-6.

- Tang, Y., Yuan, J., Osapay, G., Osapay, K., Tran, D., Miller, C., Ouellette, A. and Selsted, M. (1999). A cyclic antimicrobial peptide produced in primate leukocytes by the ligation of two truncated alpha-defensins. *Science* **286** (5439): 498-502.
- Thie, H., Schirrmann, T., Paschke, M., Dübel, S. and Hust, M. (2008). SRP and Sec pathway leader peptides for antibody phage display and antibody fragment production in *E. coli*. *N Biotechnol* **25** (1): 49-54.
- Thomas, P., Ko, Y. and Pedersen, P. (1992). Altered protein folding may be the molecular basis of most cases of cystic fibrosis. *FEBS Lett* **312** (1): 7-9.
- Tiozzo, E., Rocco, G., Tossi, A. and Romeo, D. (1998). Wide-spectrum antibiotic activity of synthetic, amphipathic peptides. *Biochem Biophys Res Commun* **249** (1): 202-6.
- Toke, O. (2005). Antimicrobial peptides: new candidates in the fight against bacterial infections. *Biopolymers* **80** (6): 717-35.
- Towle, M., Salvato, K., Budrow, J., Wels, B., Kuznetsov, G., Aalfs, K. et al. and Littlefield, B. (2001). In vitro and in vivo anticancer activities of synthetic macrocyclic ketone analogues of halichondrin B. *Cancer Res* **61** (3): 1013-21.
- Travis, S., Anderson, N., Forsyth, W., Espiritu, C., Conway, B., Greenberg, E., McCray, P., Lehrer, R., Welsh, M. and Tack, B. (2000). Bactericidal activity of mammalian cathelicidin-derived peptides. *Infect Immun* **68** (5): 2748-55.
- Treusch, S., Cyr, D. and Lindquist, S. (2009). Amyloid deposits: protection against toxic protein species? *Cell Cycle* **8** (11): 1668-74.
- Tunnacliffe, A. and Wise, M. (2007). The continuing conundrum of the LEA proteins. *Naturwissenschaften* **94** (10): 791-812.
- Urbanc, B., Cruz, L., Yun, S., Buldyrev, S., Bitan, G., Teplow, D. and Stanley, H. (2004). In silico study of amyloid beta-protein folding and oligomerization. *Proc Natl Acad Sci USA* **101** (50): 17345-50.
- Vaara, M. (1992). Agents that increase the permeability of the outer membrane. *Microbiol Rev* **56** (3): 395-411.
- van't Hof, W., Veerman, E., Helmerhorst, E. and Amerongen, A. (2001). Antimicrobial peptides: properties and applicability. *Biol Chem* **382** (4): 597-619.
- Velappan, N., Sblattero, D., Chasteen, L., Pavlik, P. and Bradbury, A. (2007). Plasmid incompatibility: more compatible than previously thought? *Protein Eng Des Sel* **20** (7): 309-13.
- Visser, L., Hiemstra, P., van den Barselaar, M., Ballieux, P. and van Furth, R. (1996). Role of YadA in resistance to killing of *Yersinia enterocolitica* by antimicrobial polypeptides of human granulocytes. *Infect Immun* **64** (5): 1653-8.
- Vlieghe, P., Lisowski, V., Martinez, J. and Khrestchatsky, M. (2010). Synthetic therapeutic peptides: science and market. *Drug Discov Today* **15** (1-2), 40-56.
- Vogel, G. (2008). The inner lives of sponges. *Science* **320** (5879), 1028-30.
- Wadia, J., Stan, R. and Dowdy, S. (2004). Transducible TAT-HA fusogenic peptide enhances escape of TAT-fusion proteins after lipid raft macropinocytosis. *Nat Med* **10** (3): 310-5.
- Waldo, G., Standish, B., Berendzen, J. and Terwilliger, T. (1999). Rapid protein-folding assay using green fluorescent protein. *Nat Biotechnol* **17** (7): 691-5.
- Walker, J., Roth, J. and Altman, E. (2001). An in vivo study of novel bioactive peptides that inhibit the growth of *Escherichia coli*. *J Pept Res* **58** (5): 380-8.
- Walsh, D., Klyubin, I., Fadeeva, J., Rowan, M. and Selkoe, D. (2002). Amyloid-beta oligomers: their production, toxicity and therapeutic inhibition. *Biochem Soc T* **30** (4): 552-7.
- Walsh, D. and Selkoe, D. (2007). Abeta oligomers - a decade of discovery. *J Neurochem* **101** (5): 1172-84.
- Wang, G., Li, X. and Wang, Z. (2009). APD2: the updated antimicrobial peptide database and its application in peptide design. *Nucleic Acids Res* **37** (S1): D933-7.
- Warkentin, T., Greinacher, A. and Koster, A. (2008). Bivalirudin. *Thromb Haemost* **99** (5): 830-9.
- Watt, P. (2006). Screening for peptide drugs from the natural repertoire of biodiverse protein folds. *Nat Biotechnol* **24** (2): 177-83.
- Watt, P. (2008). Engineered antibody fragments and alternative protein scaffolds. *BioSpectrum Asia Edition* **3** (15): 12-4.
- Watt, P. (2009). Phenotypic screening of phylomer peptide libraries derived from genome fragments to identify and validate new targets and therapeutics. *Future Med Chem* **1** (2): 257-65.

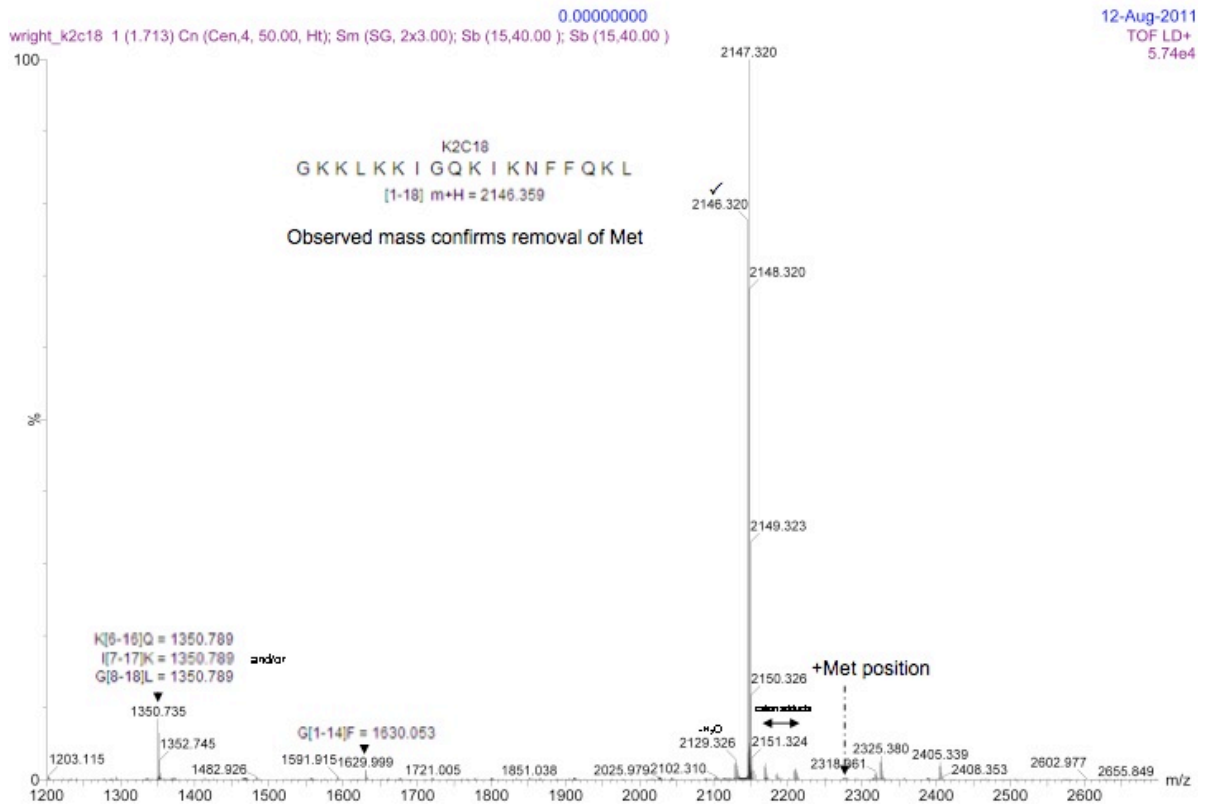
- Werle, M. and Bernkop-Schnürch, A. (2006). Strategies to improve plasma half life time of peptide and protein drugs. *Amino Acids* **30** (4): 351-67.
- West, R., Yocum, R. and Ptashne, M. (1984). *Saccharomyces cerevisiae* GAL1-GAL10 divergent promoter region: location and function of the upstream activating sequence UASG. *Mol Cell Biol* **4** (11): 2467-78.
- West, S., Schweizer, H., Dall, C., Sample, A. and Runyen-Janecky, L. (1994). Construction of improved *Escherichia-Pseudomonas* shuttle vectors derived from pUC18/19 and sequence of the region required for their replication in *Pseudomonas aeruginosa*. *Gene* **148** (1): 81-6.
- Widmaier, D., Tullman-Ercek, D., Mirsky, E., Hill, R., Govindarajan, S., Minshull, J., Voigt, C. (2009). Engineering the *Salmonella* type III secretion system to export spider silk monomers. *Mol Syst Biol* **5** (309): 1-9.
- Wiegand, I., Hilpert, K. and Hancock, R. (2008). Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. *Nat Protoc* **3** (2): 163-75.
- Wigley, W., Stidham, R., Smith, N., Hunt, J. and Thomas, P. (2001). Protein solubility and folding monitored in vivo by structural complementation of a genetic marker protein. *Nat Biotechnol* **19** (2): 131-6.
- Wilks, J. and Slonczewski, J. (2007). pH of the cytoplasm and periplasm of *Escherichia coli*: rapid measurement by green fluorescent protein fluorimetry. *J Bacteriol* **189** (15): 5601-7.
- Winkler, J., Seybert, A., König, L., Pruggnaller, S., Haselmann, U., Sourjik, V., Weiss, M., Frangakis, A., Mogk, A. and Bukau, B. (2010). Quantitative and spatio-temporal features of protein aggregation in *Escherichia coli* and consequences on protein quality control and cellular ageing. *EMBO J* **29** (5): 910-23.
- Wishart, D., Lewis, M., Morrissey, J., Flegel, M., Jeroncic, K., Xiong, Y. et al. and Li, L. (2008). The human cerebrospinal fluid metabolome. *J Chromatogr B Analyt Technol Biomed Life Sci* **871** (2): 164-73.
- Wolf, Y., Grishin, N. and Koonin, E. (2000). Estimating the number of protein folds and families from complete genome data. *J Mol Biol* **299** (4): 897-905.
- Wrighton, N., Farrell, F., Chang, R., Kashyap, A., Barbone, F., Mulcahy, L., Johnson, D., Barrett, R., Jolliffe, L. and Dower, W. (1996). Small peptides as potent mimetics of the protein hormone erythropoietin. *Science* **273** (5274): 458-64.
- Wurth, C., Guimard, N. and Hecht, M. (2002). Mutations that reduce aggregation of the Alzheimer's A β 42 peptide: an unbiased search for the sequence determinants of A β amyloidogenesis. *J Mol Biol* **319** (5): 1279-90.
- Yang, S.-T., Shin, S., Lee, C., Kim, Y.-C., Hahm, K.-S. and Kim, J. (2003). Selective cytotoxicity following Arg-to-Lys substitution in tritrypticin adopting a unique amphipathic turn structure. *FEBS Lett* **540** (1-3): 229-33.
- Yen, Y., Kostakioti, M., Henderson, I. and Stathopoulos, C. (2008). Common themes and variations in serine protease autotransporters. *Trends Microbiol* **16** (8): 370-9.
- Yeung, A., Gellatly, S. and Hancock, R. (2011). Multifunctional cationic host defence peptides and their clinical applications. *Cell Mol Life Sci* **68** (13): 2161-76.
- Young, T. and Schultz, P. (2010). Beyond the canonical 20 amino acids: expanding the genetic lexicon. *J Biol Chem* **285** (15): 11039-44.
- Yu, K., Park, K., Kang, S.-W., Shin, S., Hahm, K.-S. and Kim, Y. (2002). Solution structure of a cathelicidin-derived antimicrobial peptide, CRAMP as determined by NMR spectroscopy. *J Pept Res* **60** (1): 1-9.
- Zahnd, C., Amstutz, P. and Plückthun, A. (2007). Ribosome display: selecting and evolving proteins in vitro that specifically bind to a target. *Nat Methods* **4** (3): 269-79.
- Zanetti, M. (2005). The role of cathelicidins in the innate host defenses of mammals. *Curr Issues Mol Biol* **7** (2): 179-96.
- Zasloff, M. (1987). Magainins, a class of antimicrobial peptides from *Xenopus* skin: isolation, characterization of two active forms, and partial cDNA sequence of a precursor. *Proc Natl Acad Sci USA* **84** (15): 5449-53.
- Zasloff, M. (2002). Antimicrobial peptides of multicellular organisms. *Nature* **415** (6870): 389-95.
- Zhang, X.-S., García-Contreras, R. and Wood, T. (2007). YcfR (BhsA) influences *Escherichia coli* biofilm formation through stress response and surface hydrophobicity. *J Bacteriol* **189** (8): 3051-62.
- Zhao, L., O'Reilly, M., Shultz, M. and Chmielewski, J. (2003). Interfacial peptide inhibitors of HIV-1 integrase activity and dimerization. *Bioorg Med Chem Lett* **13** (6): 1175-7.
- Zhou, J. and Erdman, J. (1995). Phytic acid in health and disease. *Crit Rev Food Sci Nutr* **35** (6): 495-508.
- Zietkiewicz, S. and Liberek, K. (2010). Dispose to the pole-protein aggregation control in bacteria. *EMBO J* **29** (5): 869-70.

APPENDICES

Appendix 1

Mass spectrometry result for recombinant K2C18 (Section 3.2.2.3). The expected mass of K2C18 without initiating methionine was 2,146.4 Da; observed was 2,146.3 Da. This confirms that K2C18 lacked the initiating methionine.

Maldi analysis. 0.2ul sample + 2ul CHCA matrix. 120lsr, 2000pv, extcal PEG



Appendix 2

Amino acid analysis results for recombinant K2C18 and MCC18 analogues (see Section 3.2.4.2). Note that asparagine and glutamine are hydrolysed to aspartic acid and glutamic acid respectively, so are counted in this manner in the output. The indole ring of tryptophan is destroyed during hydrolysis, so this amino acid is not identified. Some amino acids are excluded from the analysis of a particular run (e.g. MCC18) as contaminants in the peptide solution result in poor fits to the true expected values. Exclusion of such contaminating amino acids allows a more accurate concentration of the desired peptide to be measured.

| AAA Results | | | |
|--|----------------|-----------------|------------------------------------|
| Sample: K2C18 | | | |
| Integer fit of measured mole ratios to expected values | | | |
| Amino acid | Expected value | Observed value | Closeness of fit to expected value |
| Cys | 0 | 0.00 | - |
| Asp | 1 | 1.05 | better than 5% |
| Thr | 0 | 0.00 | - |
| Ser | 0 | 0.00 | - |
| Glu | 2 | 2.01 | better than 5% |
| Gly | 2 | 2.00 | better than 5% |
| Ala | 0 | 0.00 | - |
| Val | 0 | 0.00 | - |
| Met | 0 | 0.00 | - |
| Ile | 2 | 1.93 | better than 5% |
| Leu | 2 | 1.95 | better than 5% |
| Norleu std | 0 | 0.00 | - |
| Tyr | 0 | 0.00 | - |
| Phe | 2 | 1.98 | better than 5% |
| His | 0 | 0.00 | - |
| Lys | 7 | 7.08 | better than 5% |
| Arg | 0 | 0.00 | - |
| Pro | 0 | 0.00 | - |
| Trp | 0 | 0.00 | (not determined) |
| Total | 18 | residues | |

| AAA Results | | | |
|--|----------------|-----------------|------------------------------------|
| Sample: syn | | | |
| Integer fit of measured mole ratios to expected values | | | |
| Amino acid | Expected value | Observed value | Closeness of fit to expected value |
| Cys | 0 | 0.00 | - |
| Asp | 1 | 1.06 | within 5-10% |
| Thr | 0 | 0.00 | - |
| Ser | 0 | 0.00 | - |
| Glu | 2 | 2.05 | better than 5% |
| Gly | 2 | 2.13 | within 5-10% |
| Ala | 0 | 0.00 | - |
| Val | 0 | 0.00 | - |
| Met | 0 | 0.00 | - |
| Ile | 2 | 1.90 | within 5-10% |
| Leu | 2 | 1.93 | better than 5% |
| Norleu std | 0 | 0.00 | - |
| Tyr | 0 | 0.00 | - |
| Phe | 2 | 1.88 | within 5-10% |
| His | 0 | 0.00 | - |
| Lys | 7 | 7.05 | better than 5% |
| Arg | 0 | 0.00 | - |
| Pro | 0 | 0.00 | - |
| Trp | 0 | 0.00 | (not determined) |
| Total | 18 | residues | |

| AAA Results | | | |
|--|----------------|-----------------|------------------------------------|
| Sample: K2C18 (F14K) | | | |
| Integer fit of measured mole ratios to expected values | | | |
| Amino acid | Expected value | Observed value | Closeness of fit to expected value |
| Cys | 0 | 0.00 | - |
| Asp | 1 | 1.03 | better than 5% |
| Thr | 0 | 0.00 | - |
| Ser | 0 | 0.00 | - |
| Glu | 2 | 2.06 | better than 5% |
| Gly | 2 | 2.12 | within 5-10% |
| Ala | 0 | 0.00 | - |
| Val | 0 | 0.00 | - |
| Met | 0 | 0.00 | - |
| Ile | 2 | 1.85 | within 5-10% |
| Leu | 2 | 1.97 | better than 5% |
| Norleu std | 0 | 0.00 | - |
| Tyr | 0 | 0.00 | - |
| Phe | 1 | 0.84 | >10% out |
| His | 0 | 0.00 | - |
| Lys | 8 | 8.14 | better than 5% |
| Arg | 0 | 0.00 | - |
| Pro | 0 | 0.00 | - |
| Trp | 0 | 0.00 | (not determined) |
| Total | 18 | residues | |

| AAA Results | | | |
|--|----------------|-----------------|------------------------------------|
| Sample: K2C18 (Q9K) | | | |
| Integer fit of measured mole ratios to expected values | | | |
| Amino acid | Expected value | Observed value | Closeness of fit to expected value |
| Cys | 0 | 0.00 | - |
| Asp | 1 | 1.13 | >10% out |
| Thr | 0 | 0.00 | - |
| Ser | 0 | 0.00 | - |
| Glu | 1 | 1.17 | >10% out |
| Gly | 2 | 2.06 | better than 5% |
| Ala | 0 | 0.00 | - |
| Val | 0 | 0.00 | - |
| Met | 0 | 0.00 | - |
| Ile | 2 | 1.84 | within 5-10% |
| Leu | 2 | 1.94 | better than 5% |
| Norleu std | 0 | 0.00 | - |
| Tyr | 0 | 0.00 | - |
| Phe | 2 | 1.78 | >10% out |
| His | 0 | 0.00 | - |
| Lys | 8 | 8.08 | better than 5% |
| Arg | 0 | 0.00 | - |
| Pro | 0 | 0.00 | - |
| Trp | 0 | 0.00 | (not determined) |
| Total | 18 | residues | |

| Average of 2 | | | |
|---------------|-----------------|-------|--|
| Total sample | 0.421 nmoles | 0.44 | |
| Concentration | 0.91 ug | 0.95 | |
| | 21.05 nmoles/ml | 21.99 | |
| | 45.25 ug/ml | 47.27 | |

| Average of 2 | | | |
|---------------|-----------------|-----------------|--|
| Total sample | 0.225 nmoles | 0.47 | |
| Concentration | 0.47 ug | 11.26 nmoles/ml | |
| | 11.26 nmoles/ml | 23.65 ug/ml | |

| Average of 2 | | | |
|---------------|-----------------|-------|--|
| Total sample | 0.361 nmoles | 0.19 | |
| Concentration | 0.77 ug | 0.41 | |
| | 18.04 nmoles/ml | 9.52 | |
| | 38.46 ug/ml | 20.46 | |

| Average of 2 | | | |
|---------------|----------------|-------|--|
| Total sample | 0.151 nmoles | 0.19 | |
| Concentration | 0.32 ug | 0.41 | |
| | 7.54 nmoles/ml | 9.52 | |
| | 16.20 ug/ml | 20.46 | |

| AAA Results | | | |
|--|----------------|-----------------|------------------------------------|
| Sample: K2C18-2 | | | |
| Integer fit of measured mole ratios to expected values | | | |
| Amino acid | Expected value | Observed value | Closeness of fit to expected value |
| Cys | 0 | 0.00 | - |
| Asp | 1 | 1.08 | within 5-10% |
| Thr | 0 | 0.00 | - |
| Ser | 0 | 0.00 | - |
| Glu | 2 | 2.03 | better than 5% |
| Gly | 2 | 2.16 | within 5-10% |
| Ala | 0 | 0.00 | - |
| Val | 0 | 0.00 | - |
| Met | 0 | 0.00 | - |
| Ile | 2 | 1.89 | within 5-10% |
| Leu | 2 | 1.94 | better than 5% |
| Norleu std | 0 | 0.00 | - |
| Tyr | 0 | 0.00 | - |
| Phe | 2 | 1.81 | >10% out |
| His | 0 | 0.00 | - |
| Lys | 7 | 7.09 | better than 5% |
| Arg | 0 | 0.00 | - |
| Pro | 0 | 0.00 | - |
| Trp | 0 | 0.00 | (not determined) |
| Total | 18 | residues | |

| | | | Average of 2 |
|---------------|-----------------|--|--------------|
| Total sample | 0.370 nmoles | | 0.37 |
| | 0.80 ug | | 0.79 |
| Concentration | 18.51 nmoles/ml | | 18.35 |
| | 39.78 ug/ml | | 39.45 |

| AAA Results | | | |
|--|----------------|-----------------|------------------------------------|
| Sample: MCC18 | | | |
| Integer fit of measured mole ratios to expected values | | | |
| Amino acid | Expected value | Observed value | Closeness of fit to expected value |
| Cys | 0 | 0.00 | - |
| Asp | 1 | 1.18 | >10% out |
| Thr | 0 | 0.00 | - |
| Ser | 0 | 0.00 | - |
| Glu | 2 | 2.14 | within 5-10% |
| Gly | 2 | 2.14 | within 5-10% |
| Ala | 0 | 0.00 | - |
| Val | 0 | 0.00 | - |
| Met | 0 | 0.00 | - |
| Ile | 0 | excluded | - |
| Leu | 0 | excluded | - |
| Norleu std | 0 | 0.00 | - |
| Tyr | 0 | 0.00 | - |
| Phe | 1 | 0.80 | >10% out |
| His | 0 | 0.00 | - |
| Lys | 6 | 5.66 | within 5-10% |
| Arg | 1 | 1.08 | within 5-10% |
| Pro | 0 | 0.00 | - |
| Trp | 0 | 0.00 | (not determined) |
| Total | 13 | residues | |

| | | | Average of 2 |
|---------------|-----------------|--|--------------|
| Total sample | 1.147 nmoles | | 1.01 |
| | 2.46 ug | | 2.16 |
| Concentration | 54.02 nmoles/ml | | 50.41 |
| | 122.88 ug/ml | | 108.06 |

| AAA Results | | | |
|--|----------------|-----------------|------------------------------------|
| Sample: MCC18 (E2K) | | | |
| Integer fit of measured mole ratios to expected values | | | |
| Amino acid | Expected value | Observed value | Closeness of fit to expected value |
| Cys | 0 | 0.00 | - |
| Asp | 1 | 1.11 | within 5-10% |
| Thr | 0 | 0.00 | - |
| Ser | 0 | 0.00 | - |
| Glu | 1 | 1.15 | >10% out |
| Gly | 0 | excluded | - |
| Ala | 0 | 0.00 | - |
| Val | 0 | 0.00 | - |
| Met | 0 | 0.00 | - |
| Ile | 2 | 1.91 | better than 5% |
| Leu | 3 | 3.02 | better than 5% |
| Norleu std | 0 | 0.00 | - |
| Tyr | 0 | 0.00 | - |
| Phe | 1 | 0.67 | >10% out |
| His | 0 | 0.00 | - |
| Lys | 7 | 7.21 | better than 5% |
| Arg | 1 | 0.93 | within 5-10% |
| Pro | 0 | 0.00 | - |
| Trp | 0 | 0.00 | (not determined) |
| Total | 16 | residues | |

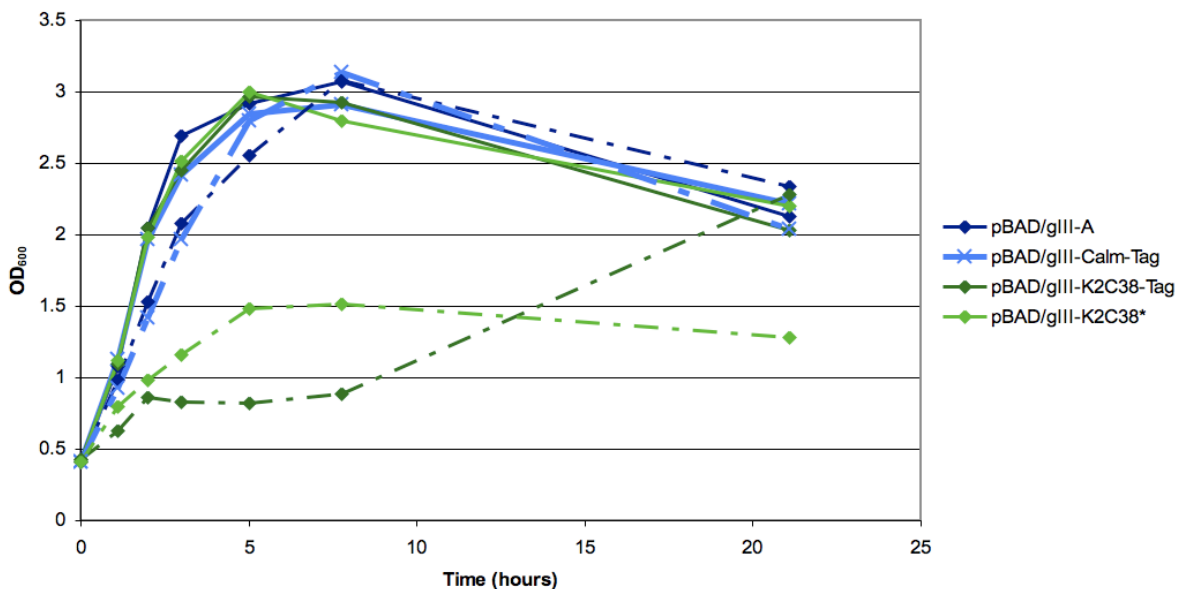
| | | | Average of 2 |
|---------------|-----------------|--|--------------|
| Total sample | 0.215 nmoles | | 0.25 |
| | 0.46 ug | | 0.54 |
| Concentration | 10.74 nmoles/ml | | 12.51 |
| | 23.01 ug/ml | | 26.81 |

| AAA Results | | | |
|--|----------------|-----------------|------------------------------------|
| Sample: MCC18 (E2K Q9K) | | | |
| Integer fit of measured mole ratios to expected values | | | |
| Amino acid | Expected value | Observed value | Closeness of fit to expected value |
| Cys | 0 | 0.00 | - |
| Asp | 1 | 1.03 | better than 5% |
| Thr | 0 | 0.00 | - |
| Ser | 0 | 0.00 | - |
| Glu | 0 | 0.00 | - |
| Gly | 2 | 2.10 | better than 5% |
| Ala | 0 | 0.00 | - |
| Val | 0 | 0.00 | - |
| Met | 0 | 0.00 | - |
| Ile | 2 | 1.98 | better than 5% |
| Leu | 3 | 2.85 | within 5-10% |
| Norleu std | 0 | 0.00 | - |
| Tyr | 0 | 0.00 | - |
| Phe | 0 | excluded | - |
| His | 0 | 0.00 | - |
| Lys | 8 | 7.96 | better than 5% |
| Arg | 1 | 1.09 | within 5-10% |
| Pro | 0 | 0.00 | - |
| Trp | 0 | 0.00 | (not determined) |
| Total | 17 | residues | |

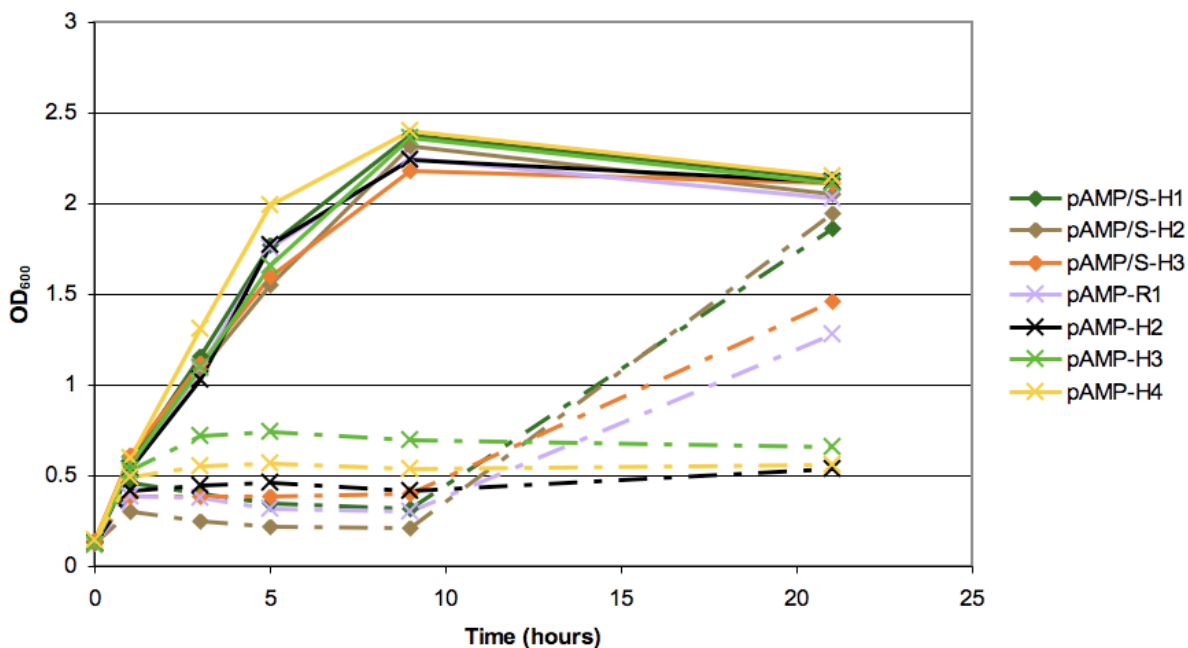
| | | | Average of 2 |
|---------------|----------------|--|--------------|
| Total sample | 0.123 nmoles | | 0.12 |
| | 0.26 ug | | 0.26 |
| Concentration | 6.15 nmoles/ml | | 6.14 |
| | 13.16 ug/ml | | 13.14 |

Appendix 3

Growth curve of pBAD/gIII-Calm-Tag (see Section 4.2.2.3), performed as per Figure 4.4 except with 50 mL culture volume in 250 mL shake flasks. pIII-tagged calmodulin (pBAD/gIII-Calm-Tag), which is targeted to the periplasm, shows the same level of growth inhibition when induced as the pBAD/gIII-A control.



Growth curves of retransformed hits from Figure 4.14 A (see Section 4.2.4.3), performed as per Figure 4.14 except with 25 mL culture volume in 50 mL shake tubes. Similar growth profiles to Figure 4.1.4 A are observed, confirming that the effector agent is plasmid encoded.



Appendix 4

DNA sequences inserted into the pAMP/S or pAMP *AfeI* site are shown, as well as the predicted translation product (*Section 4.2.4.4*). For pAMP/S peptides, only residues downstream of the pIII secretion signal cleavage site are shown, i.e. TMS onwards. For pAMP peptides, the initiating methionine is assumed to be removed (Ben-Bassat et al., 1987); the same N-terminal residues of TMS result. Plasmid-contributed residues, i.e. those outside the insert DNA, are underlined.

pAMP/S-H1

389 bp

CGGAATGCAGTGGAGAGTGGGAAGAGAATCGAATGGCGAAGCGGATGGAATGCAGTGGTGAACAGAATG
GAATGGAGGATGGTATGGAGAATGAAATGGAGTGGAGAATGGAATGGAAGGGAGCATGGAATGGAATG
GAGAATGGAATGGAATGGAGAATGGAATGAAATGGAATGGAGAATGGAATGGAATAGAGAATGGAATG
GAATAGAGCATGGAATGGAATGGAGAATAGAATGGAAGGGAGAATGGAATGGAATGGAAAGGATGATG
GTATGGAGAATGGAATGGACAGTGGAAATGAATGGAATGGAGAATGGAATGGAAGATGGATTGGAACGG
AGAATGGAATAGAGTGGAGAATGGAAAAGAGCGGAATGGAATGGAGAAT

22 aa

TMSRNAVESGRESNGEADGMQW

pAMP/S-H2

130 bp

TTACTTGTCTTATTTCAAATAAATATGATTTAAAAATTGAAGCCATTGTACAAATAGAGTTTTGGAA
TAGAACTCGTTTATAAATCAGAGACCACTCACAGGAATGTTCTTCAGCCCCTTTGAAAGGTG

30 aa

TMSLLVLFQNKYDLKIEAIVQIEFWRTRL

pAMP/S-H3

218 bp

GTTTTAAATGAATTCAAAAGTCAGTGTCTGTTTTGTTTCATGCATTCATCCATCAAACATTACCTGA
GTGTTTACCATGGGCCATATACTGCACTGAACAAGGGAGAAACGGTCATTCTCAGCTCCCACATTCAA
GGATTTTCATCCTCTAGTGGAGGAGTCAGGAAAGGAAATAAGCATATGAAGTTCTCTGGGATACTTCGT
ATGCCTAGTATTCC

79 aa

TMSVLNEFKSQCHLFFVHAFIHQTLPECLPWAIYCTEQGRNGHSQLPHSRISSSSGVVRKGNKHKMFSG
ILRMPSPISK

pAMP/S-H4

136 bp

AAAAGGTACCTTGGTACCCGATGCCTGTCCCTTATGTGGGCCAAGGGAGACCTGAGACAAATAAAAAAT
GGGAAAAGAGGGGACAAGAAGGAGTGATTGTTAGAGAGGCTGTCTGAATAATGCTTTATCAAATAAA

34 aa

TMSKRYLGTRCLSLMWAKGDLRQIKMGKRGDKKE

pAMP/S-H5

179 bp

TTCTCTAAATTTCTGTTCTTGCTTCATTTCAATTCATTTGATCTTCATTCACTGATACCCCTTTCTTCCA
GTTGATCAAATTGGCTACTGAAGCTTGTGCATTTCGTACGTTAGTTCTTGTGCCATGGTTTTTCAGCTCC
ATCAGGTCATTTAAGGACTTCTCTACACTGGTTATTCTAGTTA

20aa

TMSFSKFLFLLHFIHLIFIH

pAMP/S-H7

105 bp

GAAAGAAGCAGACATGAAAGGTCATATATTGTAGAATTCATTTATATGCAATGTTTCAGACTAAGCCA
ATCGACAGAGATAGAAAGTAGATGAGAGGTTTTCCAAC

33 aa

TMSERSRHERSYIVEFHLYAMFRLSQSTEIESR

pAMP/S-H8

384 bp

GTGATTGTACTAATTTACATTTCTACTAAGAAGGTCCAAGGGTTTATTTTTCTTCACATTCTTGCTAG
CATTTGTGTTGCCCGTCATTCCGATATAAGCCATTTGAAGTGGGGTGATGCTATATCTCATTGCAGTT
TTGACTTGCAATTTCTCTATTGGTGATATTGAGCACCTTTTCATATGCCTGTTTGCCATTTTTATGTCT
TCTTTTGAGAAATGTCGATTCAGATCTTTTGCCCATTTTTTCAATGAGGTTATTAGATTTTTTTTTCT
ACAGAGTTGATTGAGCTACTTATGTATTCTGATTATGAATCCCTTGTCAGATGGGTGGTTTGCAAATA
TTTTCTCTTATTCTGTGGGTTGTCTCTCCCTGAGTTGATTGTT

96 aa

TMSVIVLIYISTKKVQGFIFLHILASICVARHSDISHLNWGDAISHCSFDLHFSIGDIEHLFICLFAI
FMSSFEEKRFRSFAHFFNEVIRFFFLQS

pAMP/S-H10

1186 bp

CTGTGGTCTGATAGCATGGCTGATGTGATTTTTGATACTTTTGAACCTACTGAGACTTGCTTTATGGCA
GAGCATGTGGTTAATCTTGGAGTATGGTCCATGTACAGATGAGAAGGATGCATATTCTATGGTTGAGG
GGTATTCTGTAGATGTGTATTAGATACAATGGGTCAAAGTGTCAAATTTACATCCAGAATATTTTTGTT
AGTATTCTGCCTCAATGATCTGTCTAATGTTGTCAGTGGGGTGTGAAGTTCCTCACTATTATTGTGT
GGCTACCTATGCTTTTTGTCAGGTTGAGAATTACTTGTTTTATGAATCTGGATGCTCCATTGTTAGGT
GTGTATACATTTAGCATAGTTAAGTCTTCTTTTTGAATTGAACCTTTTATCATTATATAATGCCCTTC
TTTGTTCTTTTTCTACTGATTTTCATTGGCTATGTGCCTTATTATATTGGCTAGGACTTCCAGTATTATT
TTGAATAGGTGTGGTGACAGAGATCATCTTGCCTTGTTCCCTGATCTTAAGGAGATAGCATTACCAT
TAAGCGTAAAATCTATTTACATTTAATGTGATTATTGATATGGTTTGACTGACTGCAACCATTTTTCT
ATTTGTTTTCTATATATCCCATGCTCTTTTTGTTTCTCTGTTCCCTCCTCTTTTTGTGTTATTTTTCTC
GCATATCATTTTAATTCCTTTGTTGGTTTCTTTAAACTACATTTTTTAGTTACTTCATGTTTCCCTCAA
GAGATCTTAATTCATTATAATCTAATTCAGATTAATAAATAACTTAATTCATGATAATATGAAAATTT
GCTTCTATATAGCTCCATTTCCCTCCCTCTACATTTGTGGTATTATTGACATATATGTTTTAAGTGCAT
CACAGCAATGTTTATGAATATTGTTTTATGTAATTTTTAAAAAATCAGTTTGGAACACACACACACA
TATAGAGAGAGAGAGATAGATAATATACTTAGGTAATTACCTTTATCTGTGAAATGTAATTTCTTTGTG
TGGACTCAAGTTACCTCTTGATTTTCCCTTACTTGAAGGTTCTTTCTTTAGCATTTCTTATAAGACA
GATCTGTTAGCAATGCACCTTTTGTTTACCTGGAATAACTTTATTTCTATTTCATCATCAAAGGATAG
TTTTATTGAATGCAATAATCTTGATTGACA

55 aa

TMSLWSDSMADVILILLNLLRLALWQSMWLIILEYGPCTDEKDAYSMVEGYSDVY

pAMP/S-R14

171 bp

AAATTAAATATGCAAATAATCATTGGCCTGGTGCTAATAGACATTGCTTTGTTCAAATCACTGGAGAT
TTGCAAGAACATTTGTTGTAAATCAAACCTATTGAGATCGAAGGCACTTTTAGACATTTGTCTATTTT
TAGCAGCCATGATTTGTATTAGCGCAACATAGACT

58 aa

TMSKLNMQIIIGLVLIDIALFKSLEICKNICCKSNSLRKALLDICLFLAAMICISAT

pAMP/S-R15

175 bp

CGGAACCTCATTTTCAATGCTTCTACTCGCTATCAACTTGATCTTTCTAATTGTTGGACTGATCTCGAT
TATCAACAATGGAAAGATGGAAAATATTTTCTGCAAGTAAACCCAAAGGTAATATTGTTTACATGGT
ACAAACAGGCGAGAAGAATCTACGAACTCGTTGTGATAT

62 aa

TMSRNSFSMLLLAINLI FLIVGLISIIINNGKMENIFLQVNPKVILFTWYKQARRIYELVVIC

pAMP/S-R17

235 bp

GAAGATTATTTTCGACGAACGTTGAATATCGAACAAGTGGACATCCGTTTGATTGATGAAAATGAAAT
TGAAGCAGCGAGTATTTCCCAATCTGGAAGAGATTTTACCTGGAAAGCCATTGATTCAATTTTCGACATG
AACCAATGATTTCCATTCGTTTGATTAATCGACAACCTTATTCTGGTCATTTTGAATGGACTATTCCG
GTGATCAATGGTGATACCGTGGAGAAATTGG

82 aa

TMSDYFRRTLNI EQVDIRLIDENEIEAASIPNLEEILPGKPLIHFRHEPMISIRLINRQPYSGHFEW
TIPVINGDTVEKLG

pAMP/S-R20

164 bp

CTGCATCAAACCATTGTTATGTGTAGAAAATTGCTGTTGAATCTTTTTGGTTTCTTTATCTTTGGGTG
TATCAGCGAATTTTACTACCAAAGGAGATGAACAGCCCTAATAGATAGGAATATTAATTATGTAATAT
GATGAATAATTCAGAATTTAACCTCCAT

35 aa

TMSLHQTIVMCRKLLLNLF~~GGFF~~I~~FGC~~ISEFY~~YQRR~~

pAMP/S-R21

367 bp

TTTGAAGTTATCTGATTAAGATATAATGGGAAGCCATATGATATGGGATAGATGTATGCATTTGGTCT
AACAGTAAAGTGAGAAAGAATGTAAACGTCGAATTGAATAATAA~~ACTT~~GATGTCTGTTTCGAATATGAT
TGATGTGTACGTGAGAATATGTTGTATATGTTTGAAGAGAAAATTTCAAGAGGA~~ACTTTT~~CGGACTCTA
AGAAGTGCAGTTTGC~~TTGGC~~ATTGTCTTTCTGTTCAATTGAAAATATGTTGCCTAGTGTCTTTGCAA
GAGAATTGTGAGA~~ACCAA~~AGGAATAGATGACGCTACTTATTGTATATAACCATTGTACTGTTTCGTTT
ATCATTCCATATTACTATTCTTTTGAA

7 aa

TMSFEVI

pAMP/S-R22

116 bp

TACAGCAAAACAATCACTGTAAAGACAAACATAATCATCGGATTGACTTACCAGTAACAGCATCTTGG
TGTTTCTTACGGAGTTGAGCAGCTGTGGCTTCATGTTGAAGATTAGCT

21 aa

TMSYSKTITVKTNIIIGLTYQ

pAMP/S-R25

175 bp

TTATTATATTCATTGAGACGTCGTTGCTGGCCGTTTCGTGATATGTTTTCAAAGGGCGATGCTATTACT
CACTTCAGCTGTTGGAGTTTTCCGAATAATCATTTCGACTTTTCTTCGATCTTTTTAGTCAAACCTCGT
GGATCTTTTTGGAATGAATTAGCAATCTTCATGGATTTCG

53 aa

TMSLLYSLRRRCWPFVICFQRAMLLLTSAVGVFRIIIRLFFDLFSQTSWIFLE

pAMP/S-R28

347 bp

GAAATGGAAATGTACTTACTCGGAGTATCTTGCGGTTGAGAGAAGTGTGTTGAAACTCAGATGAACAA
TACTAGGAAAGTATTGCCTTTTTATACGAGAGGCATTGTTGACCAACA~~ACTATCTCTTTTTCTTTCTC~~
TCTCCCTCTTCC~~TGAGCGATAGCGAAAGGGCGTCATTTAAATGAATTATATATTGTCGAAGGATTATC~~
ATTGTTGGGCACGAGTCTGTCCGCCCTGTTTTAACGATTGGATCATTACGTTCAATCTTTTCGATCCT
TGCTCTGTGCATACATATATATATATCTCTATCTAAACAGAATAAATAAAGGGCGAATGACACGTGTA
ATTAGTT

62 aa

TMSEWKCTYSEYLAVERSVFETQMN~~NTRKVL~~PFYTRGIVDQQLS~~LFLSLSLFLS~~DSERASFK

pAMP/S-R30

145 bp

CTTGCAACTATGTTGCAAATAGGTGCAATGTTAATCTTCTTCACTGGACATATCATAGATACGATTGG
ACGTCGTCGATCGATTCA~~TTAATAACTGCTCTGCTCTTAATAACCTCGTTGATTACACAAGCATGCT~~
TACAATTTCG

52 aa

TMSLATMLQIGAMLIFFTGHIIDTIGRRRSIHLITALLLITSLITQACLQFG

pAMP/S-R31

373 bp

ATAGAAATAGATTTAGTTTTACTTATCCATATACACACCCACACCCCGTGTGAGTTT TAGGAATAGAAAT
AGATTTAGTTTTACTTATACATATACACACCCACACCCCTGTGAGTTT TAGGAATAGAAATAGATTTAG
TTTTACTTATCCATATACACACCCACACCCCTGTGAGTTT TAGGAATAGAAATAGATTTAGTTCTACTT
ATACATACATACCCACACCCCTGTGAGTTT TAGGAATAGAAATAGATTTAGTTTTACTTATACATAC
ATACCCACACCCCTGTGAGTTT TAGGAATAGAAATAGATTTAGTTTTACTTATACATACATACCCAC
ACACCCCTGTGAGTTT TAGGAATAGAAATAGATT

128 aa

TMSIEIDLVLIIHIHTHTRVSLGIEIDLVLIIHIHTHTPVSLGIEIDLVLIIHIHTHTPVSLGIEIDL
VLLIHTYPHTPVSLGIEIDLVLLIHTYPHTPVSLGIEIDLVLLIHTYPHTPVSLGIEIDC

pAMP/S-R32

238 bp

CTTTTTCCCTTCTCTCTCTCTCTCTCTTTTTCTTTTTATTTTTCTTTTACGAACGAAACAAGAAATAATA
ATATGCGTTGATGTGTCCGAGACTAGCATAGAGAAAGAGAGAAAGAAGAGCGATTAGAAAAGGGGAGA
AAGAAAAGGAGACCTTTCGTTTACATACATTACCATGCCATCACTCAAGATGCCGTTTTCTTTTTCT
TTCGAAACTATCTATTCTTTTTTTTTTCGAAAACAT

24 aa

TMSLFSLLSLSLSLLFFFYERNKK

pAMP/S-R33

108 bp

TGTTGGATTTTTTGCATTCCCTTTTATATTCCATT CAGGTAAGTTTTTTAAGGCGCTTTTTGCTTTT
CCTTATACCAGTGCCAGACAAATCCGATCAGCGAAAATGT

41 aa

TMSCWIFLHSLLYSIQVSFLRRFLLFLIPVPDKSDQRKCAK

pAMP/S-R34

168 bp

TATTTCGAATTGAATATTCGACCTTGCCGTTCTTGTTCTGCGCGACGACGTCGTTCAATCGCAGCTTT
CTCTTTCAGATCGATGGGAAGATCGAGTTTATACATATCTGTCTCCGGATATAAGAATGAAAGAATGT
TGTGTATATTA AAAACTTTCTTCTCCTTCAA

61 aa

TMSYFELNIRPCRSARRRRSIAAFSFRSMGRSSLYISVSGYKNERMLCILKTFFSFKAK

pAMP/S-R36

199 bp

ATAGCCACTTTTTGCATGCTTTGCTGATCTTTGTTACCTTT CACAAGCCATTAGTAAGAAACATTTG
GGGAACGATAAAAATCATTATGAATTTTATCTTTTATATGCGTAAAAAGAAGTGAATTAAGTAAGTA
TGATGCGTAGTATTGATTTATGTGATCACATTCTTATCTGATCCAAGGTATTTTTATGACTGCA

43 aa

TMSIATFCMLSLIFVTFHKPLVRNIWGTIKSLLNFIFYMRKKK

pAMP/S-R37

132 bp

TCACATATTATCACTTCTACTCAATGGAGTTATTCCACCTCTGCTCATGTTTATTTTTGGATGTTGG
ACAGTACAGAATATTCGTAAGTCCGTCACGCAACACGACAATCGGGCTCGACACATTCCGCTG

49 aa

TMSSHIIHFYSMELFHLCSCLFLDVGQYRIFVKSVTQHDNRARHIPLAK

pAMP/S-R39

123 bp

GACCGTGTTACTGCGTTTTGAGTTT CACAACGTCTACGTTTTAGTAAGCAATTTTCGATATAGACTAGA
TGTTCAAGAGTATATTCGATCGCCAAGTTATTCAA ACTA ACTAGAATGTCATTTA

38 aa

TMSDRVTAFEFHNVYVLVSNFRYRLDVQEYIRSPSYSN

pAMP/S-R40

222 bp

GAGTTGATTTCAAACATTTTCATTTTGTATAGATCGTCTGAACTTTTGATTTCACTGAGTATAACGTC
TAAATCTTCCGCTTCGTTAGCTTGCCTTAGTGTTTGAATTTGAACCTTTTGTGCCAAGTAACTATCAA
ACTTGGTGCTCAAGTTCGAAAGTTGTTTCGCTCATATAAAGCAGTTGCTGTTGAAATTTGGTGCTCATT
GATGGCGAGAAAGACGAA

60 aa

TMSELISKLFILYRSELLISLSITSNSSASLACLVEIEPFVPSKLSNLVLFESCSLI

pAMP/S-R41

161 bp

AACTTGTTGCTGTTCCAGTTGTGCCTAAACAAGAAATTTAACATCGTGTTCTCGAATAGCAATGTGAA
TATGAATTTAGTTTTACCAGTTGTAGTTGTGCGACGACAAAGTCGTATCACTTCTACCAGCAAAAAAAT
AGATAGGCACAGTGATCGCTCCGAT

48 aa

TMSNLLLFQLCLNKKFNIVFSNSNVNMNLVLPVVVVDDKVVSLPAKK

pAMP-R1

337 bp

TTCTATGCTCTTTTTTACTTTCTTTCTTTCGCTTTGGTAACAGAGCAGAATATATACCAGTGTTTCGTT
CGTTTGTCTGTCTCCTCTTTTTCGCGCTCTTTATCCCTTTCGTTGACATTTGACATCTGCTGTTGCTTG
ATGCATACGTATGTAGAGATGTGTGCAGATATATACATGTACATTTACTTACTTTTTCGAACAAAAAAA
GAAGTAACTGGGTACACTATAACAATCGAAGGACAATGACGCAGCACATAATCTTTTCGTTTTCTCTTT
TCTTTACTCCATCAATTTTCAGATTTATCTTCCAATGAACTAAGACGAAAGGACGTTATTATCAAA

42 aa

TMSFYVSFLLSFFALVTEQNIYQCSFVCLSPLLRLFIPFVDI

pAMP-H2

383 bp

CTTTTTCTTTTTCTTTTTTTTTTTTTTTGAGACAGGGCCTCACTCTGTCTCACAGGCTGGAGTGCAATGGC
ACATCTTGGCTCACTGCAGCCTCTGCCTCCCGGGTTCTAGCAATTCTCATGCCTCAGCCTCCCAAGT
AGCTGAGATTACAGGCATGCACCACAATGCCCGGCTAAGTTCTGTATTTTTAGTAGAGACAGGGTTTT
GCCATGTTGGTTGGCTGGCCTCGAATTCCTGACTTCAAGTGATCTGTCCATCTCGGCCTCCCAAGT
GCTGGGATTACAGGCATGAGCCACCACGCATGGCCCTGCCTTCTTGAGATTCTTGCTGAGACTGTACA
AGGCGTTGACTCTTGTCTAATCTTTTGCCTGCTCCTCTGTCT

11 aa

TMSLFLFFFF

pAMP-H3

334 bp

AGTTTAGCGGAACATTACATACAATTGAAAGAGAATGTGGTGAGCTGGAACACAGATTTTGGAGATGA
TGCCCAACCGAGAAACAGAATATGAGAGGTGAAGTGACATGGAATAGATCTAACTTAGGTCTAGCTAA
ATTCCTAGTGGAAGGAGAAAGAGAATGGGGGAAAGTCAATATAAGAGCAGCTAATAACCAAGAACTT
TTCAGAACTGATAAAAGGTAGAAACAAAAATCCACATATTCAGGGATCCAGATGAATTCCAAGCAGTG
AAAATAAAAATCAATCTAACTTGAAGCACCATAGTGATAAATGCAGAACACCAAGACAA

33 aa

TMSSLAEHYIQLKENVVSWNDFGDDAQPRNRI

pAMP-H4

268 bp

ACCTGGCCAATATTGGCAGCTTTAGGCTTCAGAAGTGTTTTTATCTTTTTTGGTGAGGAGGTGATGTA
ATGTAAGCAGGGGAGGTGTTTTGTGTTTCCACCTTGAGAGTAAGGTGGCTTGTATCAGAAGAAAA
TGAATTTCTACATTAATAAAAGGGCAGAGAAGAAAGTTCCTTGGATTAGATTGATTTAGTTATTCC
AAAATGTATACATATATCAAAATATCGGCCTGGCATGGTGGCTCACACCTGTCATCCCAACT

25 aa

TMSTWPI LLAALGFRTVFIFFGEEVM

pAMP-H5

273 bp

AACAAAAATGGCTGTTATCAAAAGCCAGTGCCTCTCCTGGCTGTTCACTGTTACCTCATTACAACCAT
GATGTTATCTGCTGGGTATCTGGGTTTACCCTGACCACAGCAGAGGCTTGGACATTATTTGCTCAAA
TGATTTTCTTCTCCTCTTCTTTTGTAAAAATTATGTGCATGAGGGATCAGGCTGGCGCAGGCCGGGT
CGGGGTAGGAGTGGTTATTCCACAGAGCACAATGTGTACTTTTTTCAGCATTCTACTTACTTCCATAAA
G

96 aa

TMSNKNGCYQKPVPLLAHVHCYLITMMLSAGYLGLPLTTAEAWTLFAQMIFSSSFCFNYVHEGSGWR
RPGRGRSGYSTEHNVYFFSILLTSIKAK

pAMP-H6

362 bp

GGAGGGCAGGGAAGGGGCAGGTAGAGTTGGGAAAGGGAGGGAGGGGCAGGGAGGGGTAAGGGACGGGG
TGAATGGAGGAAGGGAGGACAGGGAGCAGGGAGAGAAGAGAAGTCAGCAAGGACCAAGGAGTGAACCC
AGAGGAAGTAAAAGGGTAGAAGAGGGGCCCTGGCAGGGGCTCCAGTAACAAGGAGGAGGGAGGGTGGC
AGGGGAGTGACACATGGGAGATGGGGAGGCAAATAAAGAGGAGAAGGCAAAGACAAGCCGAGGGAGGA
GTGGAGAGAAGGGTGGAGAAGGAGTGAGGTGTGTGACTGCACAGGGAGGCGAGAGGAGACGCAGGTGG
TCTTGGGAGGGAAGCGAGAGGG

10 aa

TMSGGQGRGR

pAMP-H8

404 bp

GTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGT
TAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTA
GGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGG
GTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGT
TAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTA
GGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGT

141 aa

TMSVR
VR
RVLSK

pAMP-H9

304 bp

AATAAAAAATAAAAAGTGTCCCTTTTCTCTGCTTCCTTGCCAGCATTTGTTATGTTTTTTGGTCTTTT
TGACAACAGTCACCTAACAAAGTAAGATGACACCTCATTGTGGTTTTTCATTTGCATTTTCTGATAGTG
ATGTTGAGCATTTCTTCGTATATGTGTTAGCCATTTGTATGTCTTCTTTTGGAAAATGTCTATTTCAGA
GAATTTGCTCCCCTCACTTTTTTTTTTTTTGGAGACAGGGTCTCACTCTGTCAACCAGGCTGGAGTACAG
TGGTGGGATCTCAGCTCACTGCAACCTCTGCC

105 aa

TMSNKNKKCSLFSASLPAFVMFFGLFDNSHLTSKMTPHCGFHLHFPDSDVEHFFVYVLAICMSSFEKC
LFREFAPLTFFFLRQGLTLPRLYSSGGISAHCNLCR

pAMP-H11

176 bp

CAGAGAGAAAAAGAGAGTTGCTTAAAGAACATATTTAAAACATTTGTGGTATGATATGGAATGTGAC
CACCTTCTCCCACCTCCACCAAGAAATATTCAGGAAAAGCTACAAAACCCATCAACGCTTTGGGAGT
AAAGGATGTGGAGAAATTGGAACTTTCATGCACCGTGGGT

48 aa

TMSQREKESCLKNIFKNICGMIWNVTTFLLPPPRNIQEKLQNPSTLWE

pAMP-H12

212 bp

CTAATTAATAAGTTTTATTTATTTTGGCAAATACACAATTTAATTACTTTACTAGCTGTTAGCTGCAA
ACTGCCTTAGCTCCTTTGGAGTTTTTCAAATGTGCAGAAAACCTCAGGTGGAAGAGCCCAAAGCAAAC
ACACCGCTTTTTCTGATAACGGCATTTTCCAGACGTCATTTTCCCAGCTGTCAGAGTGTGATAAAAG
GCTTGGGT

53 aa

TMSLINKFYLFWQIHNLITLLAVSCKLLSSFQVGFQMRKPQVEEPKAKHTAFS

pAMP-R13

690 bp

ATACTTCAAATTCCTTTATCGTGTTATATTCATATTTAAATGTAATCGTGGTCCATATCCATTATGTTTA
GATTGGACTGAAGTTTGTGATGGAAAATTTGATTGTCTTGATGGTCCATTTGATGAAGAACATTGTAC
ACAAATTGATAATGATTATGAGATTCATAAGAGTACATTAATAATCGATGGATTTCCCGATAATCTTC
TTGGTTATCCACCTATTGTTTATCTTGAAGATGTTAAATGTCATGAATCACCTTTAACAAGTTCTTGT
ACGAAATCACGTCAAAAACAATTATTCGAAGCAATGTTTTCTATCAAAGATAATTCAACAACAGATAA
TTGTTGGTCAGCATTTAAATGTATTTTACCAATATCAATATCATTAGGTTCAATTTGTAATAGTTTTT
GTTCAAATGATAATTGTCTTGAATTTATGAAGATGAATGTCCATCAATGTTTTATATACCAATTATT
CCTATTTCTATTTACTAATATATATTTTGGCTTATGAAAACTCGATTCAAAAATGTTCAATCATGGACT
ATTTCAACATCCTTATATTTGTTATAATGATTCTCGTTACGATAAGTATTTTACTAATGAATCAGTAC
TATTATTTTCATTCGAGAAAATGTTTTCGTTACAAGACTTCAAATACAAATCATATTTTTGATTTGAAA
TATTTACAAT

25 aa

TMSILQLYRVIFILNVIVVHIHYV

pAMP-R14

362 bp

TTCAAATGAAAAACCACTGTATATCGATGTTTTTCCGGTATGAAATCTTTTTGAAGAAAAAAATTTTC
ATGGAAGAATCATTCATGTGTTTTTGAATACGTAGAATCGTGTGGGCTCGACAAATATTACTTTTATAC
AACGGTTTTTGATCAGCCAACAGATGTGACGCGTATGGAATTTCCAACATACGGTAATTGTTTTTTCATT
TATTTGGAATACAACATCACCAACAAATTAATCATTGAATCGAAGAAAAATCATCAGGTATGTACTA
ATGTTTCGATCGAATTTTCGGAATTTTCTGTTTTATCATAGATCGGAAAGCTAATGTTTCGATATGTT
ACCATTATGTAATAGTATTCAG

51 aa

TMSFKMKNHCISMFFRYEIFLKKKISWKNYSVFEYVESGLDKYYFIQRF

pAMP-R17

221 bp

TACATTTGTAAATATATGACACGAATCGTTTTTCAGGATTCAATGGTCAATTCTGGCAGGAATGGCTTT
TCAAACGGTGTATTGTTTCGATTAGTGTGGATTGACTACAGGTATTTGAATTCCTTCATCTCTTATCA
AATCACACGTTGTTTCTTCTTTAGTTGGGATATTATGGAACCTTTCACATACTTCATTTTCGTATTCA
ACTGTATTTATGGCTTA

76 aa

TMSYICKYMTRIVFRIQWSILAGMAFQTVLFLRWIDYRYLNSSSLIKSHVVSFSWDIMEPFTYFI
SYSTVFMA

pAMP-R18

129 bp

TAGTTAATAGAATGAAAACGGTGGGTGTGAGGGTGGGTGGGTGTGGGTGAGGGTGTGGGTGAGTTTGA
AAATAGGAAAGAAAGAATCTCAAGATTAGAAATCTTTTCGTCTGTTACAAAAATCTCTAT

3 aa

TMS

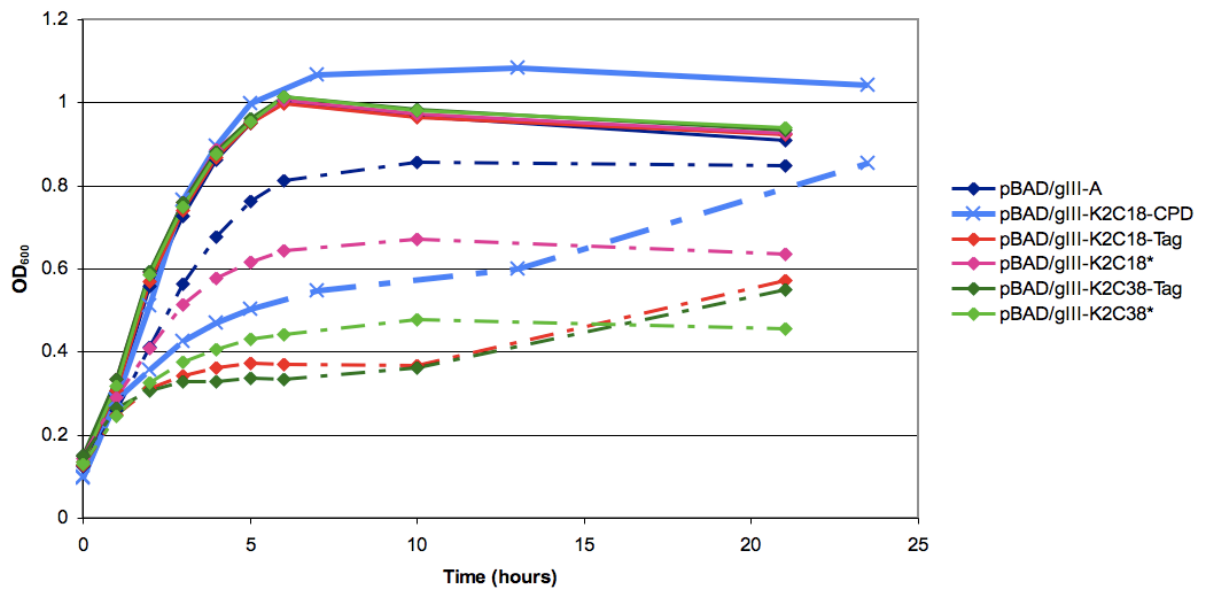
Appendix 5

Amino acid analysis result for chemically-synthesised S-H4 (Section 4.2.6.2). Note that asparagine and glutamine are hydrolysed to aspartic acid and glutamic acid respectively, so are counted in this manner in the output. The indole ring of tryptophan is destroyed during hydrolysis, so this amino acid is not identified.

| AAA Results | | | |
|---|----------------|-----------------|------------------------------------|
| Sample: | S-H4 | | |
| <i>Integer fit of measured mole ratios to expected values</i> | | | |
| Amino acid | Expected value | Observed value | Closeness of fit to expected value |
| Cys | 1 | excluded | - |
| Asp | 2 | 2.03 | better than 5% |
| Thr | 1 | 0.95 | better than 5% |
| Ser | 1 | 0.97 | better than 5% |
| Glu | 2 | 2.02 | better than 5% |
| Gly | 4 | 4.16 | better than 5% |
| Ala | 1 | 1.10 | within 5-10% |
| Val | 0 | 0.00 | - |
| Met | 2 | 2.01 | better than 5% |
| Ile | 1 | 0.98 | better than 5% |
| Leu | 4 | 3.86 | better than 5% |
| Norleu std | 0 | 0.00 | - |
| Tyr | 1 | excluded | - |
| Phe | 0 | 0.00 | - |
| His | 0 | 0.00 | - |
| Lys | 6 | 6.18 | better than 5% |
| Arg | 4 | 3.73 | within 5-10% |
| Pro | 0 | 0.00 | - |
| Trp | 1 | 1.00 | (not determined) |
| Total | 31 | residues | |
| Total sample | 0.310 | nmoles | |
| | 1.14 | ug | |
| Concentration | 15.51 | nmoles/ml | |
| | 56.89 | ug/ml | |

Appendix 6

Growth curve of pBAD/gIII-K2C18-CPD (see Section 4.3), performed as per Section 4.2.2.3 (n = 2, representative shown) and overlaid on Figure 4.4. The growth inhibitory effect of K2C18 tagged at the C-terminus with CPD falls between that of pBAD/gIII-K2C38* and pBAD/gIII-K2C18*.



Appendix 7

The 6,013 bp sequence of plasmid pAG2-A β 42 (Section 5.2.1.1). Includes mutations to vector (backbone is pBAD/gIII-A) acquired during construction (G1603T, deletion of T at position 1905); these are in regions of the plasmid that are not predicted to have any function. The A β 42-EGFP gene (Baine et al., 2009) contains the arabinose-inducible *araBAD* promoter; the Library gene contains a strong constitutive promoter (BBa_J23119); and the mCherry gene contains a medium constitutive promoter (BBa_J23101). All genes share similar terminator regions, i.e. *E. coli rrmB* operon T1/T2 terminators (underlined).

Key for coding regions: A β 42 Linker EGFP mCherry Library ATG

```
AAGAAACCAATTGTCCATATTGCATCAGACATTGCCGTCACCTGCGTCTTTTACTGGCTCTTCTCGCTAACCAAA
CCGGTAACCCCGCTTATTTAAAAGCATTCTGTAACAAAGCGGGACCAAGCCATGACAAAAACGCGTAACAAAAG
TGCTATAAATCACGGCAGAAAAGTCCACATTGATTATTTGCACGGCGTCACACTTTGCTATGCCATAGCATTTT
TATCCATAAGATTAGCGGATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTT
GGGCTAACAGGAGGAATTAACCATGGATGCGGAATTTTCGCCATGATTCTGGCTATGAAGTGCATCATCAGAAAC
TGGTGTTTTTTCGGGAAGATGTGGGCTCTAACAAAGGCGCGATTATTGGCCTGATGGTGGGCGGGCTGGTGATT
GCGGGATCCGCTGGCTCCGCTGCTGGTCTGGCGAATCCCATATGGTGAGCAAGGGCGAGGAGCTGTTACCCGG
GGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAACGGCCACAAGTTTCAGCGTGTCCGGCGAGGGCGAGG
GCGATGCCACCTACGGCAAGCTGACCCCTGAAGTTCATCTGCACCACCGCAAGCTGCCCGTGGCCCTGGCCACC
CTCGTGACCACCTGACCTACGGCGTGCAGTGTTCAGCCGCTACCCCGACCACATGAAGCAGCAGCAGCTTCTT
CAAGTCCGCCATGCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCC
GCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCCTCGACTTCAAGGAGGAC
GGCAACATCTCGGGCACAAGCTGGAGTACAACATAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAA
GAACGGCATCAAGGTGAACCTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCCTACC
AGCAGAACACCCCATCGGCGACGGCCCCGTGCTGCTGCCGACAACCACTACCTGAGCACCCAGTCCGCCCTG
AGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCTGCTGGAGTTCGTGACCGCCCGGGGATCACTCTCGG
CATGGACGAGCTGTACAAGTAAACGGTCTCCAGCTTGGCTGTTTTGGCGGATGAGAGAAGATTTTCAGCCTGAT
ACAGATTAAATCAGAACGCAGAAGCGGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCAC
CTGACCCCATGCCGAACCTCAGAAGTGAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTA
GGGAACCTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTTCGTTTTATCTGTTGTTTGT
CGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAGCTTGACAGCTAGCTC
AGTCTTAGGTATAATGCTAGCTACTCCGCAAAGAGGAGAAAACCTAGTATGAGCGCTGACTAGGTGAGACGTCAG
GCGACGTGCCCTGGCTGCAGTAGCGGGTGGTCCCACCTGACCCCATGCCGAACCTCAGAAGTGAACGCCGTAGC
GCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTAGGGAACCTGCCAGGCATCAAATAAAACGAAAGGCTCAGT
CGAAAGACTGGGCCTTTTCGTTTTATCTGTTGTTTGTGCGGTGAACGCTCTCCTCCGCCATAAACTGCCAGGCATC
AAATTAAGCAGAAGGCCATCCTGACGGATGGCCTTTTTGCGTTTTCTACAAACTCTTTTGTATTATTTTCTAAAT
ACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTA
TGAGTATTCAACATTTCCGTGTCGCCCTTATCCCTTTTTGCGGCATTTTGCCTTCTGTTTTTGTCTACCCA
GAAACGCTGGTGAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAA
CAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAAGCTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTAT
GTGGCGCGGTATTATCCCGTGTGACGCGGGCAAGCAACTCGGTGCGCCGATACACTATTCTCAGAATGAC
TTGGTTGAGTACTACCAGTACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGC
CATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTT
TTTTGCACAACATGGGGGATCATGTAACCTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATAACCAAC
GACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACCTATTAAGTGGCGAACTACTTAC
TCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCC
TTCCGGCTGGCTGGTTTTATTGCTGATAAATCTGGAGCCGGTGGAGCTGGGTCTCGCGGTATCATTGCAGCACTG
GGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAA
TAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAAGTGTGACACCAAGTTTACTCATATATAC
TTTAGATTGATTTAAAACCTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACC
AAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGA
TCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAACCACCGCTACCAGCGGTGGTTTTGTTTGGCCG
ATCAAGAGCTACCAACTCTTTTTCCGAAGGTAAGTGGCTTCAGCAGAGCGCAGATACCAATACTGTCTTCTTA
GTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTT
ACCAGTGGCTGCTGCCAGTGGCGATAAGTCTGTCTTACCAGGTTGGACTCAAGACGATAGTTACCAGGATAAGG
```

CGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGA
TACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGGAGAAAGGCGGACAGGTATCCGGTAAGCGG
CAGGGTCGGAACAGGAGAGCGCACGAGGGGAGCTTCCAGGGGAAACGCCTGGTATCTTTATAGTCCTGTCCGGT
TTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGC
AACCGGGCCTTTTTACGGTTCCTGGCCTTTTGTGCTGGCCTTTTGTGCACATGTTCTTTTCTGCGTTATCCCCTGA
TTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGACGCCGAACGACCGAGCGCAGCG
AGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTACAC
CGCATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATAACTCCGCTATCGCTT
TACAGCTAGCTCAGTCCTAGGTATTATGCTAGCTACTAGAGAAAAGAGGAGAAATACTAGATGGTGAGCAAGGGC
GAGGAGGATAACATGGCCATCATCAAGGAGTTTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCA
CGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCA
AGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAG
CACCCCGCCGACATCCCCGACTACTTGAAGCTGTCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTT
CGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCTTGCAGGACGGCGAGTTCATCTACAAGGTGAAGC
TGC CGCGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAG
CGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTA
CGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCCTGCAGCTGCCCGCGCCCTACAACGTCAACATCA
AGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACCGCGCCGAGGGCCGCCACTCC
ACCGGCGGCATGGACGAGCTGTACAAGTAATAATACTAGAGCCAGGCATCAAATAAAACGAAAGGCTCAGTCCGA
AAGACTGGGCCCTTTCGTTTTATCTGTTGTTTGTGCGGTGAACGCTCTCTACTAGAGTCACACTGGCTCACCTTCG
GGTGGGCCCTTCTGCGTTTTATAGGTTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGG
GCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTGAGAGGTTTTTC
ACCGTCATCACCGAAACGCGCGAGGCAGCAGATCAATTCGCGCGCGAAGGCGAAGCGGCATGCATAATGTGCCT
GTCAAATGGACGAAGCAGGGATTCTGCAAACCCTATGCTACTCCGTCAAGCCGTCAATTGTCTGATTGCTTACC
AATTATGACAACTTGACGGCTACATCATTCACTTTTTCTTACAACCGGCACGGAACCTCGCTCGGGCTGGCCCC
GGTGCATTTTTTAAATACCCGCGAGAAATAGAGTTGATCGTCAAACCAACATTGCGACCCGACGGTGGCGATAG
GCATCCGGGTGGTGCTCAAAGCAGCTTCGCCCTGGCTGATACGTTGGTTCCTCGCGCAGCTTAAGACGCTAATC
CCTAAGTGTGGCGGAAAAGATGTGACAGACGCGACGGCGACAAGCAAACATGCTGTGCGACGCTGGCGATATC
AAAATTGCTGTCTGCCAGGTGATCGCTGATGTACTGACAAGCCTCGCGTACCCGATTATCCATCGGTGGATGGA
GCGACTCGTTAATCGCTTCCATGCGCCGAGTAACAATTGCTCAAGCAGATTTATCGCCAGCAGCTCCGAATAG
CGCCCTTCCCCCTGCCCCGGCGTTAATGATTTGCCCAAACAGGTGCGTGAATGCGGCTGGTGCCTTCCATCCGG
GCGAAAAGAACCCCGTATTGGCAAATATTGACGGCCAGTTAAGCCATTTCATGCCAGTAGGCGCGCGGACGAAAGT
AAACCCACTGGTGATACCATTGCGGAGCCTCCGGATGACGACCGTAGTGATGAATCTCTCCTGGCGGGAACAGC
AAAATATCACCCGGTCGGCAAACAAATTTCTCGTCCCTGATTTTTTACCACCCCTGACCGCGAATGGTGAGATT
GAGAATATAACCTTTTCAATCCCAGCGGTCCGGTCGATAAAAAAATCGAGATAACCGTTGGCCTCAATCGGCCTTA
AACCCGCCACCAGATGGGCATTAACGAGTATCCCGGCAGCAGGGGATCATTTTGCCTTCAGCCATACTTTTC
ATACTCCCGCCATTCAGAG

Translated (295 aa):

MDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIAGSAGSAAGSGESHMVSKGEELFTGVVPII
LDGDVNGHKFSVSGEGEGDATYGLTLKFICTTGKLPVPWPTLVTTLYGVQCFSRYPDHMKQHDFFKSAMPEG
YVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGI DFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKVNE
KIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMVLEFVTAAGITLGMDELYK*

The linker is nearly the same as that used by Waldo et al. (1999) in their GFP folding reporter, except GSAGSAAGSGEF has been modified to GSAGSAAGSGESHMV, making it 3 aa longer.

Appendix 8

The 228 bp sequence of the Library site of plasmid pAG2-A β 42 (Section 5.2.1.1). Regions in order of notation: BBa_J23119 constitutive promoter (highlighted), BBa_B0034 ribosome binding site (highlighted), *SpeI* site (underlined), start codon (bold), *AfeI* site (underlined), stop codons for all three reading frames (bold), rrnB T1 & T2 transcriptional terminators (highlighted).

```
TTGACAGCTAGCTCAGTCCTAGGTATAATGCTAGCTACTCCGCAAAGAGGAGAAAACTAGT  
ATGAGCGCTGACTAGGTGAGACGTCAGGCGACGTGCCTGGCTGCAGTAGCGCGGTGGTCCC  
ACCTGACCCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCT  
CCCCATGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGAC  
TGGGCCTTTCGTTTTATCTGTTGTTGTCGGTGAACGCTCTCCT
```

Appendix 9

Random DNA sequences inserted into the pAG2-A β 42 Library *SpeI/AfeI* site are shown (between the fixed start and stop codons), as well as the predicted translation product(s). A number of constructs may contain additional open reading frames (ORFs), including the use of alternative start codons such as GTG (Blattner et al., 1997); these are underlined where applicable. In the case of pAG2-A β 42-110-A3-5, an additional putative RBS is highlighted (Ringquist et al., 1992; Chen et al., 1994). For the original Library site RBS and distance from the start codon (7 bp), see *Appendix 8*. Some peptides contain plasmid-contributed residues, i.e. those outside the insert DNA sequence (underlined). The initiating methionine is also assumed to be removed (Ben-Bassat et al., 1987).

For properties of the peptides, see *Section 5.2.4.5*.

A β 42

126 bp

GATGCGGAATTTTCGCCATGATTCTGGCTATGAAGTGCATCATCAGAACTGGTGTTTTTTTCGCGAAGA
TGTGGGCTCTAACAAAGGCGCGATTATTGGCCTGATGGTGGGCGGCGTGGTGATTGCG

42 aa

DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIA

GM6

126 bp

GATGCGGAATTTTCGCCATGATTCTGGCTATGAAGTGCATCATCAGAACTGGTGTCTTTTTTCGCGAAGA
TGTGGGCTCTAACAAAGGCGCGATTATTGGCCCGATGGTGGGCGGCGTGGTGATTGCG

42 aa

DAEFRHDSGYEVHHQKLV**S**FAEDVGSNKGAIIG**PM**VGGVVIA

Mutations in comparison to A β 42 indicated in bold.

pAG2-A β 42-Pep2

42 bp

CAGAACTGGATGTGGTAGCGGAAGATGCTGGCTCTAACAAA

14 aa

QKLDVVAEDAGSNK

pAG2-A β 42-AavLEA1

426 bp

TCCTCTCAGCAGAACCAGAACCAGACAGGGTGAGCAGCAGGAGCAGGGCTACATGGAGGCGGCCAAGGA
GAAGGTCGTCAACGCATGGGAGAGCACGAAGGAAACCTCTCGAGCACGGCTCAAGCGGCCGCCGAGA
AGACGGCTGAGTTTCGCGATTCCGCCGGTGAGACCATCCGTGACCTGACCGGACAGGCGCAGGAGAAG
GGTCAGGAGTTCAAGGAGCGCGCTGGCGAGAAGGCAGAGGAGACGAAGCAGCGTGCCGGGGAGAAGAT
GGATGAGACCAAGCAGCGGGCTGGCGAAATGCGCGAGAACCGGGCCAGAAGATGGAGGAGTACAAGC
AGCAGGGCAAGGGCAAGGCCGAGGAGCTTCGCGACACTGCCGCCGAGAAGCTCCACCAGGCTGGCGAG
AAGGTCAAGGGCCGCGAC

142 aa

SSQQNQNRQGEQQEQGYMEAAKEKVVNAWESTKETLSSTAQAAAETAEFRDSAGETIRDLTGQAQEK
GQEFKERAGEKAEEKQRAGEKMDKQRAGEMRENAGQKMEYKQQGKGAELRDTAAEKLNHAGE
KVKGRD

pAG2-A β 42-74-B1-3

36 bp

CCAGGGAAGAGGGCGAAACGAACAGGGAGGTGGTAG

11 aa

PGKRAKRTGRW

pAG2-A β 42-74-B1-5

36 bp

ACTCCGGGGGATGGATACCCCTGCAGCGCAACGGG

12 aa

TPGGWIPLQRNG

pAG2-A β 42-74-B2-6

36 bp

GCCCAGGGACGGGGGCGCGATTGCAGCGGAGGCAGG

12 aa

AQGRGRDCS GGR

pAG2-A β 42-74-B4-3

36 bp

AGGGTCGGGGGAGGCCAGCGGGGTAGGACGCCTCT

8 aa

RVGGGPAG

No other potential ORFs.

pAG2-A β 42-74-C1-1

36 bp

GGACTCCACTAAGTCGGCGGCTGCGGGGCAGTCAGA

3 aa

GLH

No other potential ORFs.

pAG2-A β 42-74-C3-1

53 bp

GAGCCAGGGAAGGGTCCCAACGACTTCGGCAAACATTGAAGCGCATCGGATCA

12 aa

EPGKGPNDFGKH

pAG2-A β 42-74-C4-1

36 bp

GCACGGAAGCCAGTGGGCCGACGAGCGACACGGGGG

12 aa

ARKPVGRRATRG

pAG2-A β 42-74-C4-3

36 bp

GTGGCGGGGGCGGGAGGAGGCAATCCCAGGCTCCA

12 aa

VAGAGGNSQAP

pAG2-A β 42-74-C4-4

52 bp

TGCAGGAGGGATCAAAGAGCCCGGCACCAGAGAGGCTGAAGCGCATCGGATC

12 aa

CRRDQRARHQRG

pAG2-A β 42-74-A1-A

97 bp

AGCAGAAACGTTGGCGGGCGTGGCCTGCGGCGCCGTTGAAGCGCACGCTGCGACTAGTATGTGAGGGG

GGTTTTGGGCTGCTTGTGGGGGGTGGT

12 aa

SRNVGGRGLRRR

pAG2-Aβ42-74-B1-B

36 bp

GTGCACCTGCCACGGGGGCTGGTGGTCCGGGAACGT

12 aa

VHLPRGLVVRER

pAG2-Aβ42-110-A1-1

71 bp

TTGGCATTGGTTCGGTGCAGGTGATGGATAGGATGTGTGGTAACACCACAGACCCCCAGGGAAACGGC
GGG

8 aa

LACGRCR

Alternative ORF, using ATG 15 bp downstream of RBS:

21 aa

WSVQVMDRMCNTTDPQNGG

pAG2-Aβ42-110-A1-4

88 bp

CGGATGTGTGGGAATAGTGTCCGGCGGTTCGGTGTGGGTTTCGGGTAGGGTAGCTCAAAGGACGAGCGG
ACGGTGAAGCGCATCGGATC

24 aa

RMCGNSVGRSVVWVSGRVAQRTSGR

pAG2-Aβ42-110-A3-5

167 bp

GCCGTGTAGTTTCGTGGAGAGGGACGAAGACAGCTAGCGCCGAGGGATGCTCTGCCCCAGCCGCCCGT
GTTTTGAAGCGCACGCTGCGACTAGTATGCGCAACAATTGGAGCCAAGCAGACCAGAGAAGGGACGAC
GCGATGGGGCACGGGGCGGGCAGGAAGTGA

2 aa

AV

Alternative ORF, using ATG 8 bp downstream of a putative RBS (highlighted above):

11 aa

GHGAGRKWMKR

pAG2-Aβ42-110-B1-3

78 bp

GATGCTAAGCGTGGATAGTTCGCCGGCAGGGGTTGGCATGGGCGGGCGAGCAGGGGCCCGACACGGGC
GGGTGAAGC

5 aa

DAKRG

Alternative ORF, using ATG 11 bp downstream of RBS:

25 aa

LSVDSSPAGVGMGGRAGARHGRGEA

pAG2-Aβ42-110-C2-2

72 bp

CGGGACACGGAGCTGTGGCAGTACAAGGATGGTTGTAGATGTGTAGTAGGGGTTGGATGGGATCGCTA
TAGT

24 aa

RDTELWQYKDGCRCVVGVGWD RYS

pAG2-Aβ42-110-C2-3

72 bp

GTGAGGCGCTTCGGGGTCGTCTCTCTCAGAGGACGAGAGAAGAGAGCATTGCGGGGGGGCAGGGGCGA
TGTCA

0 aa

*

Alternative ORF, using GTG 9 bp downstream of RBS:

26 aa

RRFGVVSRLRGREKRALRGGRGDVMKR

pAG2-Aβ42-110-C3-3

72 bp

CTAGACAGGGGTTTGGTGGACAGGGAGTATAGTCGCAGGGGTCAAGCTTTTGGGCAGTTATGCGCAAG
GGGG

24 aa

LDRGLVDREYSRRGQAFGQLCARG

pAG2-Aβ42-110-C3-7

72 bp

GGGCAGGGTTGCCGCATGGTCAATGACCCTCATAAGAGGGCGCGGGGAGAGGTAGGCGAGGGAGCCA
GGGG

24 aa

GQGCRMVNDPHKRARGRRRSQG

pAG2-Aβ42-110-C4-3

72 bp

CGGCCTGGGTCGGTTGATTCTGAACAGACGAGTGTGGTTGAAATCTTCCTTCGGGTATACAAGGA
CAGT

24 aa

RPGSVDSEQTSVLLKSSFRVYKDS

pAG2-Aβ42-110-A1-A

72 bp

GGCGAATAGATCGCTTGGCGGGAGCGGGAGCGAGTGGGAAAGAACGATTTTCATAGACAATTTTTAGTG
TGGA

2 aa

GE

No other potential ORFs.

pAG2-Aβ42-110-B1-B

72 bp

GGTGAAACGACTAGTGTGCGAACGGCGTAGTGGATTTCAGGAGTCGGAGTCGATGGGTGTGGTGATCCG
CTTG

9 aa

GETTSVRTA

pAG2-Aβ42-110-C1-C

72 bp

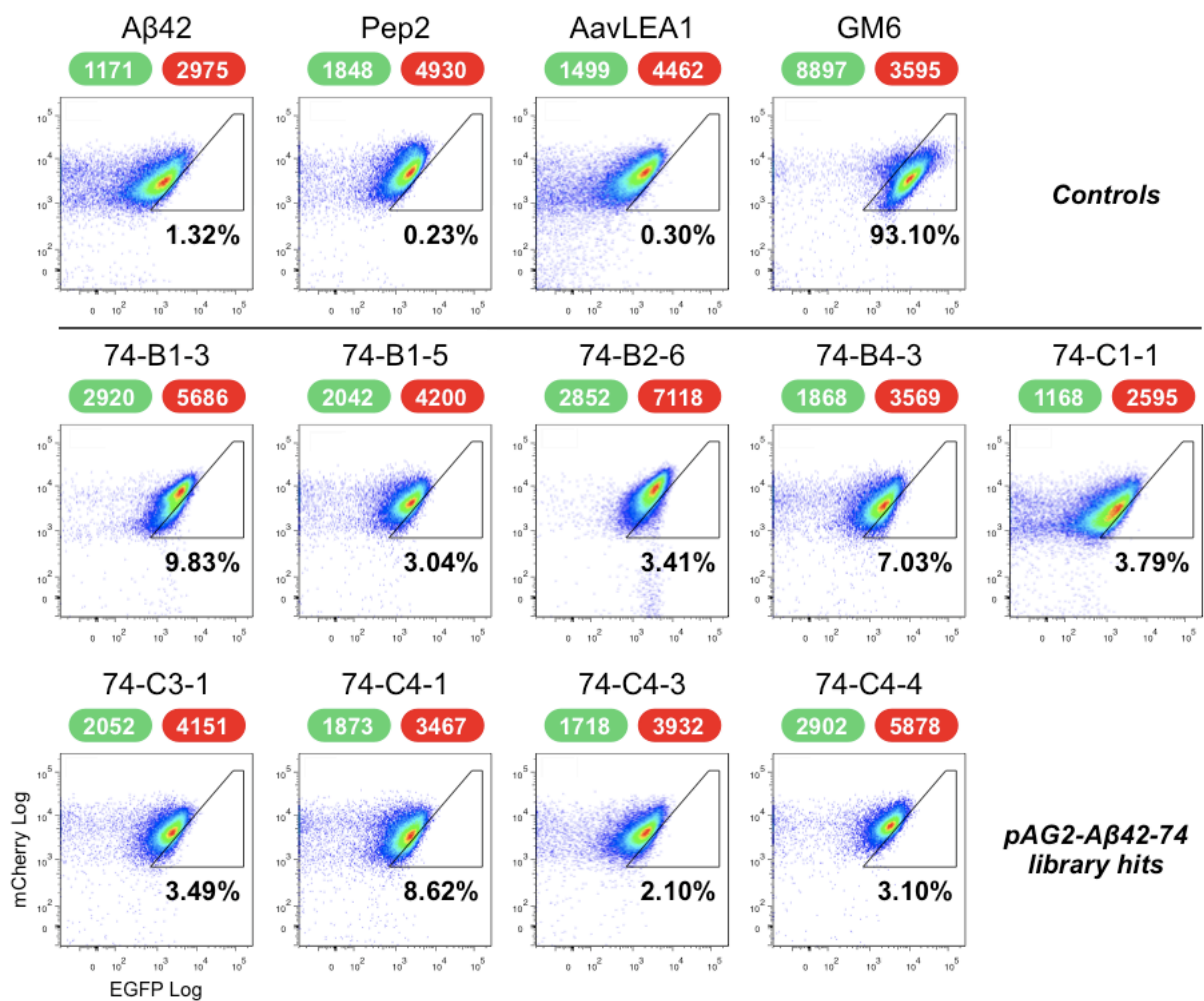
AGTGCGCCTTGTATTATGGGATCGGGCGTTGTGGGGTTTTAGACTCGGTCAACTTTCGCTCCCGCAG
TACC

24 aa

SAPCIMSGRRCVLDVNFRRSRST

Appendix 10

Flow cytometry dot plots of analysed putative antiaggregant hits from the pAG2-A β 42-74 and pAG2-A β 42-110 libraries (Section 5.2.4.3). This analysis was repeated a total of 3 times; one representative run is shown. Cultures induced for 14 h at 37°C at ~225 rpm before analysis of EGFP and mCherry fluorescence. 100,000 total events recorded per plot (red highest density, blue lowest), then an initial gating scheme was used on all samples to doublet discriminate (see Section 2.7.5.1 and Section 5.2.2.4). A region of interest was calibrated as per Section 5.2.2.4. Median fluorescence intensity (arbitrary units) shown for whole plot: EGFP (green bullet), mCherry (red bullet).



Continued over page.

