

University of Southampton Research Repository  
ePrints Soton

Copyright © and Moral Rights for this thesis are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given e.g.

AUTHOR (year of submission) "Full thesis title", University of Southampton, name of the University School or Department, PhD Thesis, pagination

**UNIVERSITY OF SOUTHAMPTON**

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES

School of Chemistry

**IDENTIFICATION OF INHIBITORS OF PROTEIN-PROTEIN  
INTERACTIONS ESSENTIAL FOR VIRULENCE OF PATHOGENIC  
BACTERIA**

by

**Abigail Male**

Thesis for the degree of Doctor of Philosophy

April 2014



**UNIVERSITY OF SOUTHAMPTON**

**Abstract**

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES

School of Chemistry

Thesis for the degree of Doctor of Philosophy

**IDENTIFICATION OF INHIBITORS OF PROTEIN-PROTEIN INTERACTIONS  
ESSENTIAL FOR VIRULENCE OF PATHOGENIC BACTERIA**

Abigail Male

There is a continuous requirement for broad-spectrum post-exposure antibiotic therapeutics. Meeting this challenge relies on the production of compounds that successfully disrupt bacterial systems identified as both conserved and essential. Here, inhibitors of protein-protein interactions involved in the Phage shock protein response and toxin internalisation, within *Burkholderia pseudomallei* and *Bacillus anthracis*, respectively have been identified. This was achieved using a high-throughput screen that combines a bacterial reverse two-hybrid system and an intein-mediated method for the generation of cyclic peptide libraries.

A reverse two-hybrid system for *Burkholderia pseudomallei* PspA homodimerisation was constructed, alongside a heterodimeric system for the interaction between *Bacillus anthracis* protective antigen and the mammalian receptor, capillary morphogenesis gene-2. From both systems a series of peptide sequences were identified with potential inhibitory activity within the reverse two-hybrid system. These compounds were synthesised and their activity assessed using a selection of *in vitro* assays.

This study identified two cyclic peptides sequences active in the reverse two-hybrid system against PspA oligomerisation; which were not active *in vitro*. In contrast, three linear peptides were isolated that demonstrated the ability to disrupt the interaction between protective antigen and the mammalian receptor, with one binding specifically to the receptor. This linear inhibitor provides the foundation for the development of a more potent antimicrobial for the arsenal against *Bacillus anthracis*.



<b>Abstract.....</b>	<b>i</b>
<b>Table of figures.....</b>	<b>xi</b>
<b>Table of tables .....</b>	<b>xvii</b>
<b>Declaration of Authorship.....</b>	<b>xix</b>
<b>Acknowledgements .....</b>	<b>xxi</b>
<b>Abbreviations .....</b>	<b>xxiii</b>
<b>Amino acids .....</b>	<b>xxvi</b>
<b>1 Introduction.....</b>	<b>1</b>
1.1 Bacterial antimicrobial drug resistance .....	5
1.2 Virulence factors as new targets for antimicrobial development .....	6
1.3 The potential of targeting protein-protein interactions.....	9
1.4 Reverse two-hybrid systems .....	11
1.5 Split-intein circular ligation of peptides and proteins .....	17
1.6 Overview of the project.....	25
<b>2 Materials and methods.....</b>	<b>27</b>
2.1 Bacterial strains.....	28
2.2 Primers.....	28
2.3 General preparation of materials.....	31
2.3.1 Luria Broth (LB) medium.....	31
2.3.2 Luria Bertani (LB) agar plates.....	31
2.3.3 Agar plates.....	31
2.3.4 Antibiotic concentrations.....	31
2.3.5 Super optimal catabolite medium .....	31
2.4 General molecular biology procedures.....	32
2.4.1 Plasmid isolation.....	32
2.4.2 Polymerase chain reaction .....	32

2.4.3 Site-directed mutagenesis.....	33
2.4.4 Restriction digestion .....	34
2.4.5 Gel purification of digested fragments.....	35
2.4.6 Ligation .....	35
2.4.7 Preparation and transformation of chemically-competent cells .....	35
2.4.8 Preparation and transformation of electro-competent cells.....	36
2.4.9 Transformation of electro-competent cells containing the helper plasmid	
37	
2.4.10 Preparation of minimal media agar plates.....	37
2.4.11 Ortho-nitrophenyl $\beta$ -galactosidase assay .....	38
<b>2.5 General protein procedures .....</b>	<b>39</b>
2.5.1 Preparing and running an SDS-PAGE gel.....	39
<b>2.6 General peptide synthesis .....</b>	<b>41</b>
2.6.1 Synthesis of a general cyclic peptide using solid-phase synthesis.....	41
2.6.2 Kaiser Test.....	42
<b>3 Investigating PspA oligomerisation as a potential antimicrobial target</b>	<b>43</b>
<b>3.1 Introduction: The Phage shock protein response.....</b>	<b>43</b>
3.1.1 Identification of the Psp response .....	45
3.1.2 Regulation of the Phage shock protein response .....	48
3.1.3 Induction of the Psp response .....	51
3.1.4 Phage shock protein A and its role within the Psp response.....	55
3.1.5 Other proteins involved in the response .....	57
3.1.6 The Psp response within <i>Burkholderia pseudomallei</i> .....	59
<b>3.2 Results: Inhibitor screening of PspA oligomerisation.....</b>	<b>63</b>
3.2.1 Construction of the PspA Reverse-two-hybrid system .....	63
3.2.1.1 Construction of pTHCP16 PspA.....	63
3.2.1.2 Construction of pAH68 PspA and integration onto the chromosome.....	64

---

3.2.1.3	Verifying the integration of PspA .....	68
3.2.1.4	Further verification of PspA integration.....	70
3.2.2	Verification of the PspA PPI within the RTHS.....	72
3.2.2.1	Verification of the PspA PPI using a drop-spotting assay.....	72
3.2.2.2	Quantification of the PspA dimerisation by o-nitrophenyl $\beta$ -galactosidase assays	74
3.2.3	SICLOPPS screening against PspA homodimerisation.....	76
3.2.3.1	SGWX <sub>5</sub> Library construction.....	76
3.2.3.2	SX <sub>5</sub> G and SX <sub>6</sub> G library construction.....	79
3.2.3.3	CX <sub>5</sub> library construction .....	79
3.2.3.4	Screening of the four libraries .....	80
3.2.3.5	Secondary screening and screening for specificity .....	82
3.2.3.6	Sequencing and ranking of active SICLOPPS peptides.....	85
3.2.4	Verification of intein expression in top ranking peptides .....	85
<b>3.3</b>	<b>Results: Testing the PspA inhibitors <i>in vitro</i> .....</b>	<b>89</b>
3.3.1	Synthesis of the cyclic peptide inhibitors.....	89
3.3.1.1	Cyclo-SGWLVYWF (compound 2) .....	92
3.3.1.2	Cyclo-SGWSIVFI (compound 4).....	94
3.3.1.3	Improving the solubility of cyclo-SGWLVYWF (2) by PEGylation .....	95
3.3.2	PspA protein expression.....	100
3.3.2.1	Purification of the insoluble and soluble fractions.....	100
3.3.3	Probing the structure of PspA .....	104
3.3.4	Development of an <i>in vitro</i> assay for the validation of selected inhibitors	
	111	
3.3.4.1	Size exclusion chromatography .....	111
3.3.4.2	Isothermal titration calorimetry of PspA inhibitor binding.....	122
3.3.4.3	Microscale Thermophoresis of PspA inhibitor binding.....	126
3.3.4.4	Testing cyclo-SGWLVYWF in <i>B. pseudomallei</i> using SICLOPPS.....	129
<b>3.4</b>	<b>Summary and discussion for Chapter 3.....</b>	<b>131</b>

---

<b>3.5 Materials and methods: Identification of inhibitors of PspA oligomerisation .....</b>	<b>141</b>
3.5.1 Construction of the PspA RTHS and SICLOPPS libraries.....	141
3.5.1.1 Constructing pTHCP16 PspA.....	141
3.5.1.2 Constructing pAH68 PspA .....	141
3.5.1.3 Integration of pAH68 PspA onto the chromosome of SNS118.....	142
3.5.1.4 Identification of integration .....	142
3.5.1.5 Confirming integration by PCR.....	143
3.5.1.6 Further PCR conditions for verifying Integration .....	144
3.5.1.7 CX <sub>5</sub> SICLOPPS library construction for PspA dimerisation .....	145
3.5.1.8 SGWX <sub>5</sub> SICLOPPS library construction for PspA dimerisation.....	145
3.5.1.9 SX <sub>5</sub> G SICLOPPS library construction for PspA dimerisation .....	146
3.5.1.10 SX <sub>6</sub> G SICLOPPS library construction for PspA dimerisation.....	146
3.5.2 Genetic selection and inhibitor screening of PspA .....	146
3.5.2.1 Drop-spotting to verify PspA dimerisation .....	146
3.5.2.2 Screening of inhibitors against PspA dimerisation.....	146
3.5.3 Construction of inhibitor splicing mutants and splicing verification .....	146
3.5.3.1 Construction of I <sub>C</sub> mutant pARCB.....	146
3.5.3.2 Construction of I <sub>N</sub> mutant pARCB.....	147
3.5.3.3 Construction of SGWIYWNV plasmid.....	147
3.5.3.4 Construction of SGWLVYWF, SGWSIVFI and SGWIYWNV I <sub>C</sub> mutant plasmid	
147	
3.5.3.5 Construction of SGWLVYWF, SGWSIVFI and SGWIYWNVF I <sub>N</sub> mutant plasmid	
148	
3.5.3.6 Expression of cyclic peptide inhibitors.....	148
3.5.4 PspA protein purification .....	149
3.5.4.1 Construction of pET28 PspA expression plasmid .....	149
3.5.4.2 Native PspA Protein Purification from the soluble fraction .....	149
3.5.4.3 Native PspA Protein Purification from the insoluble fraction .....	150

---

3.5.4.4 Denaturing purification of PspA .....	150
3.5.4.5 Size exclusion chromatography .....	152
3.5.4.6 Western blot analysis of His <sub>6</sub> -tagged PspA .....	152
3.5.4.7 Construction of pMAL-tagged PspA .....	153
3.5.4.8 Expression and purification of MBP-tagged PspA .....	153
3.5.4.9 Western blot analysis of MBP-tagged PspA.....	153
3.5.5 Synthesis of PspA inhibitors .....	154
3.5.5.1 Linear-WLVYWFSG.....	154
3.5.5.2 Cyclo-SGWLVYWF .....	155
3.5.5.3 Linear-WSIVFISG .....	156
3.5.5.4 Cyclo-SGWSIVFI .....	157
3.5.5.5 Crude linear-WLVYWF <sub>CG</sub> .....	158
3.5.5.6 Linear-WLVYWF <sub>C(PEG<sub>750</sub>)G</sub> .....	158
3.5.5.7 Cyclo-C(PEG <sub>750</sub> )GWLVYWF .....	159
3.5.5.8 Linear-WLVYWF <sub>C(PEG<sub>5000</sub>)G</sub> .....	160
3.5.6 In vitro assays for PspA inhibitor screening .....	160
3.5.6.1 Electron microscopy .....	160
3.5.6.2 Isothermal titration calorimetry.....	160
3.5.6.3 Microscale thermophoresis .....	161
<b>4 Investigating anthrax toxin internalisation as a potential antimicrobial target .....</b>	<b>163</b>
<b>4.1 Introduction: <i>Bacillus anthracis</i> and anthrax .....</b>	<b>164</b>
4.1.1 Pathogenesis and clinical manifestations.....	164
4.1.2 The <i>Bacillus anthracis</i> toxins and their entry into the host cell.....	165
4.1.3 Structural properties of protective antigen and the heptameric prepore	
168	
4.1.4 Critical residues of protective antigen essential for binding and	
translocation of the anthrax toxins .....	170

---

4.1.5	Identification of receptors for toxin internalisation .....	172
4.1.6	The manipulation of host signalling pathways by anthrax toxins.....	176
4.1.7	Current treatments and prophylactics .....	177
4.1.8	Protein-protein interactions within anthrax progression.....	180
<b>4.2</b>	<b>Results: Inhibitor screening of the dimerisation of the anthrax toxin PA and the mouse receptor CMG2 .....</b>	<b>183</b>
4.2.1	Design and construction of the Anthrax RTHS .....	183
4.2.2	Verification of the interaction between PA <sub>488-735</sub> and CMG2 <sub>38-218</sub> proteins in the RTHS .....	184
4.2.2.1	Verification of the PPI using a drop-spotting assay.....	184
4.2.2.2	Quantification of the PPI by ortho-nitrophenyl β-galactosidase assay .....	186
4.2.3	Screening of the CX <sub>5</sub> library.....	187
4.2.4	Sequencing and ranking of active SICLOPPS peptides .....	190
4.2.5	Investigating intein toxicity.....	193
4.2.6	Screening of the SGWX <sub>5</sub> library.....	195
<b>4.3</b>	<b>Results: Testing the anthrax inhibitors in vitro .....</b>	<b>199</b>
4.3.1	Synthesis of anthrax inhibitors .....	199
4.3.2	Protein expression of GST-tagged CMG2 <sub>38-218</sub> and His6-tagged PA <sub>488-735</sub> .201	201
<b>4.4</b>	<b>Results: Development of an in vitro assay to verify activity of the anthrax toxin internalisation inhibitors.....</b>	<b>205</b>
4.4.1	Development of an ELISA to quantify the inhibitors activity .....	205
4.4.2	Investigating the importance of the amino acid sequence .....	210
4.4.3	Probing the importance of the phenylalanine residue.....	212
4.4.4	Using the ELISA to ascertain which domain of PA the inhibitors disrupt the interaction with CMG2.....	214
4.4.5	Identification of the binding partner of CLRFT .....	217
4.4.5.1	Identification of the binding partner of CLRFT by MST .....	217

4.4.5.2 Identification of the binding partner of CLRFT by ITC .....	218
4.4.6 Optimisation of the binding of CLRFT.....	221
4.4.6.1 Synthesis of the CLRFT phenylalanine analogues.....	221
4.4.6.2 Comparing the binding of the CLRFT analogues .....	222
<b>4.5 Summary and discussion for Chapter 4.....</b>	<b>225</b>
<b>4.6 Materials and methods: Identification of inhibitors of anthrax toxin internalisation.....</b>	<b>233</b>
4.6.1 Construction of the Anthrax RTHS.....	233
4.6.2 Screening of inhibitors .....	233
4.6.3 Anthrax protein purification.....	233
4.6.3.1 Construction of pGEX CMG2 <sub>38-218</sub> expression plasmid.....	233
4.6.3.2 Construction of pET28 CMG2 <sub>38-218</sub> expression plasmid .....	233
4.6.3.3 Construction of pET28 PA <sub>488-735</sub> expression plasmid.....	234
4.6.3.4 Expression and purification of GST-tagged CMG2 <sub>38-218</sub> .....	234
4.6.3.5 Expression and purification of His <sub>6</sub> -tagged CMG2 <sub>38-218</sub> .....	235
4.6.3.6 Expression and purification of His <sub>6</sub> -tagged PA <sub>488-735</sub> .....	236
4.6.3.7 Western blot analysis of His <sub>6</sub> -tagged PA <sub>488-735</sub> and CMG2 <sub>38-218</sub> .....	236
4.6.3.8 Western blot analysis of GST-tagged CMG2 <sub>38-218</sub> .....	236
4.6.4 Synthesis of anthrax inhibitors .....	237
4.6.4.1 Synthesis of linear CLRFT .....	237
4.6.4.2 Synthesis of linear CLRPT .....	238
4.6.4.3 Synthesis of linear CMNHFP.....	239
4.6.4.4 Synthesis of linear scrambled CLRFT.....	240
4.6.4.5 Synthesis of linear scrambled CMNHFP .....	241
4.6.4.6 Synthesis of linear CMNHAPA.....	242
4.6.4.7 Synthesis of linear CLRF(4-NO <sub>2</sub> )T .....	243
4.6.4.8 Synthesis of linear CLRYT .....	244
4.6.4.9 Synthesis of linear CLRY(3,5-Br <sub>2</sub> )T .....	245
4.6.4.10 Synthesis of linear CLRY(2-NO <sub>2</sub> )T .....	246

---

4.6.4.11	Synthesis of linear CLRF(4-Cl)T .....	247
4.6.4.12	Synthesis of linear CLRF(4-F)T .....	248
4.6.4.13	Synthesis of linear CLRF(4-CN)T.....	249
4.6.4.14	Synthesis of linear CLRF(4-Bz)T .....	250
4.6.4.15	Synthesis of linear CLR(Phg)T.....	251
4.6.4.16	Synthesis of linear CLR(hPhe)T .....	252
4.6.4.17	Synthesis of linear CLRdFT .....	253
4.6.5	In vitro assays for Anthrax inhibitor screening .....	254
4.6.5.1	Sandwich ELISA with GST-tagged CMG2 <sub>38-218</sub> and His <sub>6</sub> -tagged PA <sub>488-735</sub> .....	254
4.6.5.2	Sandwich ELISA with His <sub>6</sub> -tagged CMG2 <sub>38-218</sub> and PA <sub>596-735</sub> .....	254
4.6.5.3	Isothermal titration calorimetry.....	254
4.6.5.4	Microscale thermophoresis .....	255
<b>5</b>	<b>Final conclusions and future work .....</b>	<b>257</b>
<b>6</b>	<b>Appendices .....</b>	<b>263</b>
6.1	Appendix 1 .....	263
6.2	Appendix 2 .....	264
6.3	Appendix 3 .....	265
6.4	Appendix 4 .....	266
6.5	Appendix 5 .....	267
6.6	Appendix 6 .....	268
<b>7</b>	<b>References .....</b>	<b>271</b>

**Table of figures**

Figure 1: Timeline of the deployment of antibiotics against the development of resistance observed in a clinical setting.....	4
Figure 2: Chemical structure of the lead compound identified for T3SS inhibition (a salicylidene acylhydrazide).....	8
Figure 3: Diagram showing the first yeast two-hybrid system.....	11
Figure 4: Diagram showing different operator sequences.....	14
Figure 5: The RTHS utilised by Horswill <i>et al.</i> (2004) to identify inhibitors of ribonucleotide reductase heterodimerisation and the homodimerisation of HIV protease.....	15
Figure 6: Illustration of the four types of inteins.....	18
Figure 7: Scheme showing the proposed mechanism for intein processing.....	19
Figure 8: Split-inverted intein used for SICLOPPS .....	21
Figure 9: Scheme for the mechanism of the split inverted intein using SICLOPPS .....	21
Figure 10: Working model of the Psp response in <i>E. coli</i> .....	47
Figure 11: Domain organisation of <i>E. coli</i> PspF compared to a typical bEBP .....	50
Figure 12: Alignment of the predicted secondary structure and domains .....	55
Figure 13: <i>B. pseudomallei</i> psp operon .....	60
Figure 14: Representation of plasmid pTHCP16 constructed by Horswill <i>et al.</i> (2004) .....	63
Figure 15: Representation of plasmid pAH68 constructed by Haldimann <i>et al.</i> (2001).....	65
Figure 16: Diagram illustrating how PCR is used to identify successful integration.....	68
Figure 17: 1% agarose gel to verify successful integration .....	69
Figure 18: 1% agarose gel to identify successful integrants with multiple integration.....	69
Figure 19: Split primer PCR to confirm the integration. ....	70
Figure 20: PCR using P1 and P4 primers.....	71
Figure 21: Drop-spotting of PspA RTHS strains onto minimal media.....	73
Figure 22: Drop-spotting of PspA RTHS strains onto LB agar media containing varying levels of IPTG to show toxicity.....	74
Figure 23: Breakdown of ONPG into ONP for ONPG assay.....	74
Figure 24: Graph of the average results from the ONPG assay.....	75

Figure 25: Gel investigating restriction digestion with <i>Hind</i> III and <i>Bgl</i> I in different buffers .....	77
Figure 26: LB agar plates showing transformation efficiency of different ratios of backbone and insert used in a ligation.....	78
Figure 27: Drop-spotting ranking of selected peptides alongside their reference number .....	84
Figure 28: Sequences of active peptides identified through SICLOPPS screening.....	85
Figure 29: SDS-PAGE gel showing intein processing and the absence of intein processing in mutated inteins .....	87
Figure 30: Drop-spotting of the active inteins and the mutant, non-splicing inteins.....	88
Figure 31: Reaction scheme for the standard synthesis of an SGWX <sub>5</sub> cyclic peptide, with a glycine derived Wang resin.....	91
Figure 32: (A) Linear-WLVYWFSG (1). (B) <i>Cyclo</i> -SGWLVYWF (2). ....	92
Figure 33: Scheme of Boc removal from a tryptophan residue.....	93
Figure 34: (A) Linear-WSIVFISG (3). (B) <i>Cyclo</i> -SGWSIVFI (4) .....	94
Figure 35: Reaction scheme for the PEGylation of <i>cyclo</i> -CGWLVYWF peptide (7).....	96
Figure 36: Mass spectrum analysis of linear-WLVYWFC(PEG <sub>750</sub> )G (6) .....	97
Figure 37: Mass spectrum for <i>cyclo</i> -C(PEG <sub>750</sub> )GWLVYWF (7) .....	98
Figure 38: Mass spectrum for linear-WLVYWFC(PEG <sub>5000</sub> )G (8) .....	99
Figure 39: Purification of His <sub>6</sub> -tagged PspA from the soluble and insoluble fraction.....	101
Figure 40: SEC purification of His <sub>6</sub> -tagged PspA from the soluble and insoluble fraction.....	103
Figure 41: Electron micrographs of negatively stained <i>B. pseudomallei</i> His <sub>6</sub> -tagged PspA complexes.....	105
Figure 42: Electron micrographs of negatively stained PspA aggregate.....	106
Figure 43: Representative images taken of <i>B. pseudomallei</i> His <sub>6</sub> -tagged PspA purified in the presence of CHAPS from the insoluble fraction. (A) Mixture of rod-like structures and ring-like structures. (B) Rod-like structure has tapered ends. (C) A solitary ring-like species.	107
Figure 44: Purification of <i>E. coli</i> His <sub>6</sub> -tagged PspA protein .....	108
Figure 45: Electron micrographs of negatively stained <i>E. coli</i> His <sub>6</sub> -tagged PspA complexes.....	110
Figure 46: Chromatogram of His <sub>6</sub> -tagged PspA (1 mg/ml) incubated with inhibitor (2 mM) for 12 hours.....	112

---

Figure 47: Chromatogram of His <sub>6</sub> -tagged PspA (1 mg/ml) purified and then heated to 100°C in the presence or absence of 8 M urea .....	113
Figure 48: Denaturing studies of His <sub>6</sub> -tagged PspA .....	114
Figure 49: Denaturing and refolding studies.....	115
Figure 50: <i>In vitro</i> cell-free expression of His <sub>6</sub> -tagged PspA .....	117
Figure 51: Analysis of His <sub>6</sub> -tagged PspA on a Superose 6 size exclusion column.....	118
Figure 52: Optimisation of His <sub>6</sub> -tagged PspA expression and purification.....	119
Figure 53: His <sub>6</sub> -tagged PspA expressed by <i>in vitro</i> cell free expression.....	120
Figure 54: Co-expression of His <sub>6</sub> -tagged PspA and inhibitors.....	121
Figure 55: Purification of MBP-tagged PspA .....	124
Figure 56: Graphs of ITC of MBP-tagged PspA with inhibitors .....	125
Figure 57: MST of inhibitors against MBP-tagged PspA .....	128
Figure 58: Comparison of growth of <i>B. pseudomallei</i> K96243 wild-type, a <i>pspA</i> null mutant and wild-type conjugated with pBRH4-SGWLVYWF .....	130
Figure 59: Structure of linear-WL VYWF SG (1).....	154
Figure 60: Structure of cyclo-SGWLVYWF (2).....	155
Figure 61: Structure of linear-WSIVFISG (3).....	156
Figure 62: Structure of cyclo-SGWSIVFI (4). ....	157
Figure 63: Structure of linear-WL VYWF CG (5). ....	158
Figure 64: Structure of linear-WL VYWF(PEG <sub>750</sub> )G (6).....	158
Figure 65: Structure of cyclo-C(PEG <sub>750</sub> )GWLVYWF (7).....	159
Figure 66: Structure of linear-WL VYWF(PEG <sub>5000</sub> )CG (8).....	160
Figure 67: Diagram illustrating the internalisation of anthrax toxins LF and EF utilising the interaction between PA and a cell surface receptor .....	167
Figure 68: Crystal structure of PA <sub>83</sub> .....	169
Figure 69: Crystal structure of (PA <sub>63</sub> ) <sub>7</sub> .....	170
Figure 70: Crystal structure of PA <sub>63</sub> and CMG2 .....	174
Figure 71: Crystal structure of hexameric TEM-8.....	175
Figure 72: Representation of plasmid pTHCP14 constructed by Horswill <i>et al.</i> (2004) .....	183
Figure 73: Drop-spotting of Anthrax RTHS strains onto minimal media .....	185

---

Figure 74: Graph of the average results from the ONPG assay.....	187
Figure 75: Representative secondary screen of the CX <sub>5</sub> library in the Anthrax RTHS and p6-UEV RTHS.....	190
Figure 76: Secondary screen of the active sequences from the CX <sub>5</sub> library .....	191
Figure 77: Sequences of active peptides identified through SICLOPPS screening on less stringent conditions.....	194
Figure 78: Drop-spotting of the SGWX <sub>5</sub> library in Anthrax RTHS .....	196
Figure 79: Structure of CLRFT (9), CLRPT (10) and CMNHFPA (11).....	200
Figure 80: Affinity purification of GST-tagged CMG2 <sub>38-218</sub> and His <sub>6</sub> -tagged PA <sub>488-735</sub> .....	202
Figure 81: Purification of His <sub>6</sub> -tagged PA <sub>488-735</sub> and GST-tagged CMG2 <sub>38-218</sub> .....	203
Figure 82: Comparison of different concentrations of His <sub>6</sub> -tagged PA <sub>488-735</sub> and GST-tagged CMG2 <sub>38-218</sub> .....	206
Figure 83: Comparison of different concentrations of MgCl <sub>2</sub> .....	207
Figure 84: Optimisation of primary and secondary antibody concentration.....	208
Figure 85: ELISA of CLRFT, CLRPT and CMNHFPA against the dimerisation of GST-tagged CMG2 <sub>38-218</sub> and His <sub>6</sub> -tagged PA <sub>488-735</sub> . .....	209
Figure 86: Structure of scrambled analogues.....	211
Figure 87: Representative plot of the ELISA of CLRFT, scrambled CLRFT and CMNHFPA against the dimerisation of His <sub>6</sub> -tagged PA <sub>488-735</sub> and GST-tagged CMG2 <sub>38-218</sub> . ..	212
Figure 88: Structure of CMNHAPA (14) the alanine analogue of CMNHFPA (11). .....	213
Figure 89: Representative plot of ELISA comparing the activity of CMNHFPA and CMNHAPA, against the dimerisation of His <sub>6</sub> -tagged PA <sub>488-735</sub> and GST-tagged CMG2 <sub>38-218</sub> . solution...	213
Figure 90: Purification of His <sub>6</sub> -tagged CMG2 <sub>38-218</sub> .....	215
Figure 91: ELISA of CLRFT, CLRPT and CMNHFPA against the dimerisation of His <sub>6</sub> -tagged CMG2 <sub>38-218</sub> and PA <sub>596-735</sub> .....	216
Figure 92: MST graph of CLRFT and CLRPT against His <sub>6</sub> -tagged CMG2 <sub>38-218</sub> and His <sub>6</sub> -tagged PA <sub>488-735</sub> .....	218
Figure 93: Graph of ITC of His <sub>6</sub> -tagged PA <sub>488-735</sub> and His <sub>6</sub> -tagged CMG2 <sub>38-218</sub> with CLRFT .....	220
Figure 94: Structure of linear CLRFT (9). .....	237
Figure 95: Structure of linear CLRPT (10).....	238

---

Figure 96: Structure of linear CMNHFPA (11).....	239
Figure 97: Structure of linear scrambled CLRFT (12).....	240
Figure 98: Structure of linear scrambled CMNHFPA (13).....	241
Figure 99: Structure of linear CMNHAPA (14) .....	242
Figure 100: Structure of linear CLRF(4-NO <sub>2</sub> )T (15).....	243
Figure 101: Structure of linear CLRYT (16).....	244
Figure 102: Structure of linear CLRY(3,5-Br <sub>2</sub> )T (17).....	245
Figure 103: Structure of linear CLRY(3-NO <sub>2</sub> )T (18).....	246
Figure 104: Structure of linear CLRF(4-Cl)T (19) .....	247
Figure 105: Structure of linear CLRF(4-F)T (20).....	248
Figure 106: Structure of linear CLRF(4-CN)T (21).....	249
Figure 107: Structure of linear CLRF(4-Bz)T (22) .....	250
Figure 108: Structure of linear CLR(Phg)T (23).....	251
Figure 109: Structure of linear CLR(hPhe)T (24).....	252
Figure 110: Structure of linear CLRdFT (25).....	253
Figure 111: Plates showing the screening of the SGWX <sub>5</sub> library .....	264
Figure 112: Optimisation of DMSO concentration for PspA and inhibitor binding studies by ITC. .....	265
Figure 113: Representative screen of the CX <sub>5</sub> library in the Anthrax RTHS.....	266
Figure 114: Optimisation of the concentration of the primary PA <sub>596-735</sub> antibody and secondary antibody, anti-rabbit-HRP .....	267
Figure 115: MST of inhibitors against His <sub>6</sub> -tagged CMG2 <sub>38-218</sub> .....	269



**Table of tables**

Table 1: Bacterial strains.....	28
Table 2: Primer sequences, restriction digest sites underlined.....	28
Table 3: Typical antibiotic concentrations used during cloning.....	31
Table 4: Composition of GoTaq PCR reaction mixture.....	32
Table 5: PCR Program for GoTaq polymerase.....	32
Table 6: Composition of 50 x tris-acetate-EDTA buffer.....	33
Table 7: Composition of site-directed mutagenesis reaction mixture.....	33
Table 8: PCR program for site-directed mutagenesis with Pfu polymerase.....	34
Table 9: Composition of restriction digestions without BSA.....	34
Table 10: Composition of ligation reaction mixture.....	35
Table 11: Composition of TBFI buffer.....	36
Table 12: Composition of TBFII buffer.....	36
Table 13: Composition of minimal media agar plates.....	38
Table 14: Composition of 10 x minimal media salt solution.....	38
Table 15: Composition of Z-buffer.....	39
Table 16: Composition of resolving gel.....	39
Table 17: Composition of stacking gel.....	40
Table 18: Composition of the 2 x loading dye.....	40
Table 19: Composition of the 5 x running buffer.....	40
Table 20: Composition of the Coomassie Blue staining solution.....	41
Table 21: Composition of the destain solution.....	41
Table 22: HPLC program for the general purification of linear and cyclic peptides.....	42
Table 23: Table containing all the conditions used to optimise integration of the pAH68 PspA into the chromosome of the SNS118 bacterial strain.....	68
Table 24: Table describing all the supplements added to the minimal media agar for the drop- spotting and library selection plates.....	81
Table 25: Table outlining the different libraries used for screening.....	82

Table 26: The intein mutations used during the construction of non-splicing inteins to verify intein processing. <sup>246</sup> .....	86
Table 27: Composition of GoTaq PCR reaction mixture with P1, P2, P3 and P4 primers.....	142
Table 28: PCR program for GoTaq polymerase with P1, P2, P3 and P4 primers.....	143
Table 29: Composition of LongAmp PCR reaction mixture with P1 and P4 primers. ....	143
Table 30: PCR program for LongAmp polymerase PCR with P1 and P4 primers. ....	144
Table 31: Conditions for PCR for the verification of integration. ....	144
Table 32: Composition of chitin-binding buffer. ....	149
Table 33: Composition of PspA lysis buffer.....	150
Table 34: Composition of PspA extraction buffer.....	150
Table 35: Composition of PspA denaturing lysis buffer.....	151
Table 36: Composition of PspA denaturing wash buffer.....	151
Table 37: Composition of PspA refolding buffer. ....	151
Table 38: Composition of PspA refolding elution buffer. ....	152
Table 39: Composition of PspA gel filtration buffer. ....	152
Table 40: Composition of MST-optimised buffer.....	161
Table 41: Sequences of the 25 active and specific sequences from the SGWX <sub>5</sub> SICLOPPS screen in the Anthrax and p6-UEV RTHS.....	198
Table 42: IC <sub>50</sub> data for each repeat of the ELISA and the average and standard deviation. ....	210
Table 43: Modifications of the phenylalanine residue in each analogue.....	222
Table 44: K <sub>d</sub> of each phenylalanine analogue tested by MST. ....	223
Table 45: Composition of ECPM1 media.....	235
Table 46: Composition of trace elements solution.....	235
Table 47: PspF homologues in <i>B. pseudomallei</i> K96243.....	263

### **Declaration of Authorship**

I, Abigail Male declare that the thesis entitled:

Identification of inhibitors of protein-protein interactions essential for virulence in pathogenic bacteria

and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work is under review to be published:

Self-assembly of *Escherichia coli* Phage shock protein A, Abigail Male, Petra Oyston, Ali Tavassoli, Advances in Microbiology

Evaluating the Role of Phage-Shock Protein A in *Burkholderia pseudomallei* Stephanie J. Southern, Abigail Male, Timothy Milne, Mitali Sarkar-Tyson, Ali Tavassoli, Petra C. F. Oyston PLOSOne

Signed: .....

Date: .....



## **Acknowledgements**

Firstly, I would like to thank Dr. Ali Tavassoli for the opportunity to carry out this research and the countless advice and guidance given to me throughout my PhD. I would also like to thank the past and present members of the Tavassoli group; especially Dr. Fedor Foranov whose work on the Anthrax project has been built upon in this thesis and Charlie Saunders, for drop-spotting without complaint.

I would like to extend my gratitude to our collaborators at DSTL, Professor Petra Oyston, Dr. Mitali Sarkar-Tyson and Stephanie Southern. Their help and guidance on the project has been invaluable.

I would also like to thank Dr. Stuart Findlow and Dr. Julie Herniman for their help and advice with NMR and MS, respectively. Additionally, my thanks go to Dr. Emilia Danilowicz-Luebert (Nanotemper Technologies) and Peter Gimeson (GE Healthcare) for their advice with MST and ITC and the analysis of the data.

Special thanks must go to Dr. Geraldine Price and Dr. Warren Duffy for their time in proof-reading my thesis.

Finally, I would like to thank my family and friends, in particular Tom Griscti-Perry for his continuous support and patience over the course of my studies.



**Abbreviations**

3-AT	3-amino triazole
AAA <sup>+</sup>	ATPases associated with diverse cellular activities
bEBP	Bacterial enhancer-binding proteins
bp	Base pairs
BSA	Bovine serum albumin
cAMP	Cyclic adenosine monophosphate
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate
CMG2	Capillary morphogenesis gene-2
CRIM	Conditional-Replication, Integration, and Modular
CtBP	C-terminal binding protein
DCM	Dichloromethane
DHFR	Dihydrofolate reductase
DIC	Diisopropylcarbodiimide
DMF	Dimethylformamide
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
ECL	Enhanced chemiluminescence
EDC	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
EF	Edema factor
ELISA	Enzyme-linked immunosorbent assay
ET	Edema toxin
Fmoc	Fluorenylmethyloxycarbonyl
FRET	Fluorescence resonance energy transfer
GFP	Green fluorescent protein

GST	Glutathione S-transferase
HIF	Hypoxia inducible factor
His <sub>6</sub>	Hexa-histidine
HIV	Human immunodeficiency virus
HOBr	Hydroxybenzotriazole
HSP	Heat-shock proteins
IC <sub>50</sub>	Half maximal inhibitory concentration
IHF	Integration host factor
Int	Phage integrase
IR	Infrared
IPTG	Isopropyl-β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
LF	Lethal factor
LT	Lethal toxin
MAPKK	Mitogen-activated protein kinase kinases-2
MBP	Maltose-binding protein
MCS	Multiple cloning site
MIDAS	Metal ion-dependent adhesion site
MKCs	Mitogen-activated protein kinases
MS	Mass spectrometry
MST	Microscale thermophoresis
NMR	Nuclear magnetic resonance
ONP	<i>o</i> -nitrophenyl
ONPG	<i>o</i> -nitrophenyl β-galactosidase
PBS	Phosphate buffered saline

PA	Protective antigen
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PMF	Proton motive force
PPI	Protein-protein interaction
Psp	Phage shock protein
RP-HPLC	Reverse-phase High performance liquid chromatography
RTHS	Reverse two-hybrid system
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SICLOPPS	Split-intein circular ligation of peptides and proteins
SOC	Super optimal catabolite
T3SS	Type-three secretion system
TAE	Tris-acetate-EDTA
Tat	Twin-arginine translocation
TBF	Transformation buffer
TEM-8	Tumour endothelial marker-8
TEMED	Tetramethylethylenediamine
TFA	Trifluoroacetic acid
THS	Two-hybrid system
TIS	Triisopropylsilane
UAS	Upstream activation sequence
Vipp1	Vesicle-inducing protein in plastids 1
VWA	von Willebrand factor A domain

**Amino acids**

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

## 1 Introduction

An antimicrobial substance has the ability to kill or inhibit the growth of a pathogenic microorganism. The emergence of two synthetic drugs: Salvarsan available in 1910 as a treatment for syphilis<sup>1</sup> and sulfonamides synthesised in 1932,<sup>2,3</sup> arguably marked the start of the modern era of antimicrobials. The isolation of penicillin, in the 1940s, can be considered as the first clinically significant group of antibiotics. Discovered in 1928, Alexander Fleming observed that the growth of *Staphylococcus aureus* was inhibited in a region surrounding a fungus, identified from the *Penicillium* genus.<sup>4</sup> The discovery of penicillin signalled the start of the development of antimicrobials from natural substances. The golden age of antimicrobials followed, with the discovery of streptomycin in 1944 and the isolation of chloramphenicol, tetracycline, vancomycin and macrolides from soil bacteria.<sup>5,6</sup> In addition to antimicrobials from natural sources, synthetic drugs were also developed during this period including nalidixic acid, a quinolone derivative.<sup>7</sup> Antimicrobials have constantly undergone structural optimisation in order to provide improved antimicrobial activity against a broader spectrum of targets. This development of antimicrobials is exemplified by the β-lactams, whereby the natural antibiotic penicillin was modified to produce the semi-synthetic ampicillin and piperacillin. Ampicillin is an antimicrobial effective against Gram-negative and Gram-positive bacteria, which are also covered by penicillin and piperacillin. Piperacillin has a bactericidal effect against 84% of *Pseudomonas aeruginosa* at a dose of 100 µg/ml; this is greater than other antibiotics available (i.e. carbenicillin 60%).<sup>8,9</sup>

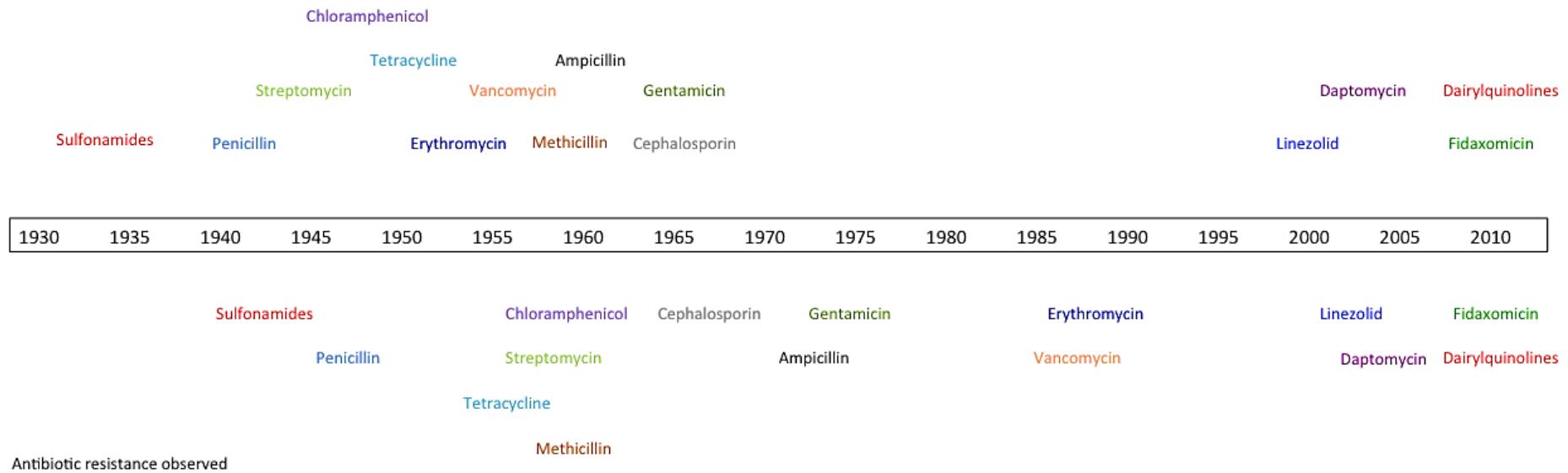
Since their initial discovery there has been a continued requirement for a second-generation of antimicrobials, as a result of bacterial drug resistance. The development of resistance to current antimicrobials is due to their mechanism of action. All current antimicrobials are either bactericidal or bacteriostatic.<sup>10</sup> Bactericidal antimicrobials directly cause bacterial death: for example β-lactams target cell wall biosynthesis, while ciprofloxacin targets gyrases and topoisomerases in deoxyribonucleic acid (DNA) replication. In contrast, bacteriostatic antimicrobials

limit bacterial growth by targeting metabolism (i.e. protein production) and allow the host's defensive machinery to eradicate the bacteria. Examples of bacteriostatic inhibitors include doxycycline, an inhibitor of protein synthesis, which functions by binding to the 30S subunit of the ribosome, preventing the binding of aminoacyl-tRNA, and inhibiting mRNA translation. The mechanism of action for both bacteriostatic and bactericidal antibiotics generates stringent selection pressures on bacterial growth, leading to a propensity for developing drug resistance.

Multi-drug resistant *S. aueru*s is the best-known example of a bacterium acquiring antimicrobial resistance, originally to sulfonamides then to penicillin in the 1950s.<sup>11</sup> Methicillin was developed in response to the increase in resistance, and is not susceptible to bacterial β-lactamases and therefore circumnavigated the acquired drug resistance. However, methicillin-resistant *S. aueru*s was isolated in the UK only one year after the implementation of the clinical use of methicillin.<sup>12,13</sup> In addition, in 2003 approximately 63% of *S. aueru*s infections in the USA were methicillin-resistant; a 10% increase from 1995, and a dramatic increase compared to 1974 statistics, when the figure was only 2%.<sup>14-16</sup> Characteristically, all antimicrobials have shown clinically relevant resistance within 15 years of their release (Figure 1).<sup>17</sup> Over the last 30 years the only two clinically relevant antibiotics with new modes of action developed have been linezolid, an oxazolidinone that targets protein synthesis, and daptomycin, which targets multiple aspects of bacterial cellular membrane function. Both are narrow-spectrum antibiotics that are effective against severe infections from Gram-positive bacteria, for example methicillin-resistant *S. aueru*s. They were licensed in 2000 and 2003, respectively, and showed evidence of resistance within a year of their release.<sup>18-20</sup> Reinforcing the premise that only traditional targets are being probed for the development of new antimicrobials, in 2011 a new drug was licensed to treat *Clostridium difficile* called lipiarmycin. Lipiarmycin is a fidaxomicin, a macrocyclic antibiotic that targets the RNA polymerase.<sup>21</sup> In order to develop resistance to lipiarmycin a specific mutation in the RNA polymerase is required, which is unlikely; this has led to a low occurrence of resistance observed *in vitro* and in a clinical setting. Consequently, with the proper distribution and use of the drug, resistance could be controlled.<sup>21</sup> An example of a drug in development is Cadazolid, a member of the oxazolidinone class of drugs. It is currently in advanced clinical trials as a treatment for *C. difficile*, working by

inhibiting protein synthesis, in turn, mitigating toxin and spore formation.<sup>22</sup> But, protein synthesis is a typical target for a bacteriostatic drug, falling into the same pitfalls of resistance development.

Antibiotic licensing



**Figure 1: Timeline of the deployment of antibiotics against the development of resistance observed in a clinical setting.** Each antibiotic is located on the timeline where they were clinically available and when resistance was observed.

## **1.1 Bacterial antimicrobial drug resistance**

Bacterial resistance has always been evident in the environment, yet increasing prevalence of drug resistance in bacteria in a clinical setting has occurred due to the widespread use and misuse of antibiotics.<sup>23</sup> There are three main routes by which resistance has evolved:

- 1) **Poor management** of antibiotic use in the clinical setting, including over-prescription, misuse and poor patient compliance. The misuse of antibiotics is exemplified by the treatment of tuberculosis, whereby due to the low number of cases, over 15% of antimicrobial treatments were incorrectly prescribed (i.e. dose and course length).<sup>24</sup> Furthermore, a recent study in Hong Kong suggested that a lack of knowledge about antibiotics led to a 37% non-compliance rate.<sup>25</sup> A similar study conducted in Ireland showed 59% of patients did not complete the course of treatment, while 39% did not adhere to the guidelines regarding administration.<sup>26</sup>
- 2) **Extensive use in veterinary medicine and agriculture** has been shown by mathematical models to accelerate antibiotic resistance in a clinical setting.<sup>27</sup> For example, vancomycin is used in humans, but there is also an animal equivalent. Its use in animals vastly outweighs the use in humans, which has implications for humans as cross-species resistance can develop.<sup>28</sup>
- 3) **Natural occurrence of resistance.** Drug resistance occurs due to selection by inheritable resistance, due to mutations caused by errors in DNA replication or acquisition of a resistance gene by horizontal gene transfer. These genetic differences often instruct a fitness cost, which typically selects against them, yet in the presence of an antibiotic this small population have a fitness advantage and are able to dominate the culture.

There are a variety of mechanisms by which resistance can occur, including three basic strategies: inactivation of the antibiotic via chemical modifications or degradation; modifications of the antibiotic target site; and limiting the intracellular concentration of the drug via increased efflux pumps and decreased porin activity.

Resistance to tetracycline can be triggered by a variety of routes, including its removal from the cell via an efflux pump<sup>29</sup> and by impeding the uptake of tetracycline.<sup>30</sup> Furthermore, protection of the ribosome to prevent tetracycline from binding to the ribosome inhibits its activity. The variety of mechanisms by which resistance occurs makes predicting its emergence difficult; nevertheless these mechanisms are well studied and understood.

Conversely, bacterial populations also exhibit resistance by a more sophisticated less studied tolerance mechanism, caused by the presence of persisters.<sup>31,32</sup> These cells are dormant in the presence of a bactericidal antibiotic, in a multi-drug resistant state; they are not mutants, instead they are stochastic phenotypic variants of the active cells.<sup>33</sup> Studies have demonstrated that it is a noninherited, epigenetic trait, rendering the descendants of persisters susceptible to the antibiotics.<sup>34,35</sup> Little is known about the mechanism by which persisters form and function. However, as antibiotic concentration tapers off, their growth is activated causing a relapse in infection.

Regardless of the cause of resistance, there is an increasing requirement for the development of new approaches to antimicrobial drug discovery to overcome the problem of resistance. An emerging area involves targeting *in vivo* essential virulence factors.<sup>36</sup>

## **1.2 Virulence factors as new targets for antimicrobial development**

Over the last decade virulence factors have been investigated as potential targets to overcome the development of resistance to current antibiotics. The underlying mechanisms of bacterial pathogenicity are the bacteria's ability to invade a host (i.e. adherence, production of proteins and enzymes that promote invasion and their ability to evade the host's defensive machinery) and to cause disease.<sup>37</sup> As well as the hypothesised diminished resistance development in a clinical environment, there are other advantages to targeting virulence factors.<sup>36,38</sup> One advantage is their inherent specificity as they only affect a pathogen's ability to cause disease; they should have little effect on most of the human natural microbiota. Anti-virulence therapies can be narrow-spectrum: this would inherently limit their clinical

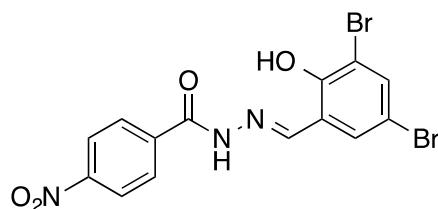
application yet conceivably limit development of resistance. Nevertheless, their side effects are difficult to predict definitively, although metagenomic analysis of human flora is assisting in this.<sup>39</sup> Another drawback of a virulence antibiotic is the need to have an understanding of virulence factors and how the host and bacterium interact. Nonetheless, next generation sequencing is making this more feasible in a shorter time frame.<sup>40</sup>

Several virulence factors have been identified as potential targets and are currently in the research stage of development. Target factors include those relating to adhesion,<sup>41-45</sup> secretion systems,<sup>46-49</sup> quorum sensing<sup>50-52</sup> and the host's defensive survival mechanisms.<sup>53</sup>

Pilicides are an example of a class of drugs that target adhesion and biofilm formation, often the first step in the infection process. Adhesion can involve the formation of fimbria or pili on the outside of the cell that adhere to the host. Pilicides mimic normal pili subunits and are incorporated as the pilus assembles, preventing elongation and the formation of a functional pilus.<sup>54</sup> In mouse models the pilicides have demonstrated the ability to decrease bacterial adhesion to the bladder and prevent infection.<sup>54</sup> The expression of components of the pilus has also been targeted. *Vibrio cholera*, the causative agent of cholera, has two virulence factors: cholera toxin (CT) and the toxin co-regulated pilus (TCP). Virstatin has been identified as a compound that inhibits the expression of the transcriptional regulator, ToxT; this in turn prevents expression of the TCP and CT, both essential for colonisation. In a mouse model, colonisation was reduced by four-logs in the presence of the inhibitor, demonstrating attenuation of *V. cholerae* in the presence of virstatin.<sup>45</sup>

Several studies have targeted bacteria's ability to deliver toxins and proteins, essential for virulence, across the membrane.<sup>55-57</sup> Some pathogens such as *Yersinia pestis*, produce specialised secretory systems, known as 'type three secretion systems' (T3SS). The T3SS forms a needle-like structure consisting of a basal body with a needle that allows improved invasion or attachment. Kauppi *et al.* (2003) screened a library of 9,400 compounds against the T3SS in *Yersinia pseudotuberculosis* and identified a class of lead compounds based on salicylidene acylhydrazides (Figure 2). These compounds function by blocking the export of

*Yersinia* outer proteins (Yops) that, once injected in the host, modulate phagocytosis and cell signalling pathways involved in the host's innate immune response.<sup>58</sup> Due to the homology of secretion systems in Gram-negative bacteria, targeting these systems gives rise to a potentially broader-spectrum antibiotic.<sup>59,60</sup> With *Chlamydia trachomatis*, the lead compound reversibly inhibited the cell cycle and infectivity through the T3SS.<sup>61</sup> In addition, when exposed to the compound, *Shigella flexineri* showed a shorter needle length by electron microscopy. This demonstrates the broad-spectrum potential of targeting virulence factors, while providing a strong indication that salicylidene acylhydrazides functioned by altering needle assembly.<sup>47,62</sup>



**Figure 2: Chemical structure of the lead compound identified for T3SS inhibition (a salicylidene acylhydrazide).<sup>63</sup>**

Other examples of antimicrobials that target secretion include a thiazolidinone compound that inhibits the assembly of the needle complex in *Salmonella enterica* serovar Typhimurium.<sup>64</sup> Thiazolidinone has again highlighted the potential broad-spectrum nature of targeting the secretion systems as it subsequently was used to block the secretion of virulence proteins via the type IV secretion system in *Francisella*. This was proposed to be by targeting the outer membrane ring proteins.<sup>64</sup> These findings support the premise that targeting virulence factors could be a viable option for antimicrobial development.

Targeting toxin production is another potential mechanism for the development of antimicrobials. The shigatoxin group of *E. coli* expresses the shiga toxin, which causes diarrhoea in infected individuals. The shiga toxin (subunit A) binds to a pentameric component (subunit B); allowing binding to the globosylceramide 3 ganglioside receptor, which in turn causes internalisation of subunit A. Once internalised, the shiga toxin cleaves a glycosidic bond on 28S RNA, inhibiting

protein synthesis, resulting in cell death. The antimicrobial Synsorb-pk, binds to the B subunit and prevents receptor-binding, internalisation and, in turn, cell death.<sup>65</sup> Similarly, methionyl tRNA (REP3123) inhibits the transcription of *C. difficile* toxin A and B subunits with 20-times greater efficacy than vancomycin, and reduces spore formation by ten-fold.<sup>66,67</sup> These mechanisms are not essential for the pathogen's ability to survive, but by inhibiting these processes the pathogen's ability to cause disease is attenuated, reducing the selection pressure on the bacteria theoretically reducing the bacteria's susceptibility to develop resistance.

Protein-protein interactions (PPIs) involved in virulence provide a potential novel target. Secretion systems provide a good example of how PPIs involved in virulence can be targeted. The homodimerisation of VirB8, which is essential for the type IV secretion system of *Brucella abortus* 2308, has already been probed using a bacterial reverse two-hybrid system and inhibitors have been identified.<sup>68</sup> The most potent inhibitor, B8I-2, reduces replication of *B. abortus* in J774 macrophages at 12.5 µM.<sup>68</sup> In a similar manner, the formation of the needle tip complex could also be targeted. The tip of the needle consists of a pentamer of a single protein (i.e. LcrV within the *Yersinia* genus) and it forms at the interface between the pathogen and the host cell, playing a crucial role in the infection process. Use of size exclusion chromatography, along with transmission electron microscopy, in studies of *Yersinia* LcrV and its homologue PcrV from *P. aeruginosa*, have confirmed the formation of discrete pentamers *in vitro*. The disruption of this pentameric PPI would putatively inhibit secretion.<sup>69</sup>

Rational design of an inhibitor that disrupts protein oligomerisation essential for virulence could provide a novel therapeutic drug for the treatment of these pathogenic diseases. Described above are examples of how PPIs involved in virulence can be probed to provide new antimicrobial drugs.

### **1.3 The potential of targeting protein-protein interactions**

Due to the frequency of PPIs involved in important biological functions, it is surprising that currently there are so few inhibitors that target them. However, when considering the properties of PPIs, it is clear why they have been previously described as 'high hanging fruits', referring to the prevailing difficulties surrounding

their discovery.<sup>70</sup> A key difficulty in inhibiting PPIs arises from the large featureless interface found between the proteins involved (typically ~1,500-3,000 Å); compared to those involved in protein-small molecule interactions (generally ~300-1,000 Å).<sup>70</sup> Also, the interface between the two proteins is characteristically flat, with no distinguishable grooves or pockets present to allow the binding of small molecules. Due to this, many current inhibitors of PPIs are large, not bioavailable or drug-like molecules and disobey Lipinski's rule of five for oral-drugability.<sup>71</sup> For example, Navitoclax, a B-cell lymphoma-XL inhibitor, violates three of Lipinski's rules in the rule of five (i.e. a molecular mass of 975 Da, a cLogP of 12, and 11 hydrogen-bond acceptors).<sup>72</sup> Despite this, mutational studies have shown that for most PPIs only a small subset of residues contribute to the majority of the free energy of binding,<sup>73</sup> allowing a small molecule to potentially disrupt the interaction demonstrating the potential for small molecule inhibitor identification.<sup>72</sup>

Although challenging, several successful attempts have been made in identifying inhibitors of PPIs, prior to association of the subunits. The most widely studied PPI is p53 and its negative regulator MDM2 involved in tumour initiation. Several inhibitors have been identified i.e. spirooxindoles and 1,4-benzodiazepine-2,5-diones.<sup>74-76</sup> Similarly, extensive efforts have been made to develop an inhibitor of human immunodeficiency virus (HIV) infection, by inhibiting the interaction between gp120 on the viral surface and CD4, with varying degrees of success. A naphthalene sulfonic acid dye was demonstrated to be active on a micromolar scale *in vitro*; however, *in vivo* it degraded into toxic by-products and its study was discontinued.<sup>77</sup> A second inhibitor, FP21399 is currently in phase II clinical trials.<sup>78</sup>

Significant effort has been made in utilising genetic selection to identify PPI inhibitors. This has frequently involved coupling either a protein complementation assay or two-hybrid system (THS),<sup>79</sup> to a compound library that is screened.<sup>80</sup> Whilst considerable progress has been made in the identification of PPI inhibitors in the last decade, in light of the inherent challenges this entails, novel strategies for tackling protein interfaces are still required.

## 1.4 Reverse two-hybrid systems

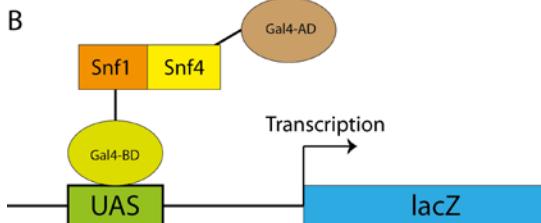
There are many *in vitro* assays available to study PPIs including: isothermal titration calorimetry (ITC), fluorescence polarisation/anisotropy, surface plasmon resonance (SPR), thermophoresis, enzyme-linked immunosorbent assays (ELISA) and fluorescence resonance energy transfer (FRET). These assays identify PPIs and can be adapted to screen for inhibitors, however, the use of an *in vivo* assay – for example a bacterial THS or a protein complementation assay – provides a better mimic for the native conditions of the proteins. THS were initially developed in yeast to identify and characterise PPIs, utilising a PPI to activate the transcription of a gene that can be monitored by a chromogenic assay or via host survival.<sup>79</sup>

Classically, two chimeric proteins are expressed: one containing the N-terminal of the *Saccharomyces cerevisiae* Gal4 (1-147) DNA-binding domain attached to the bait protein, and the other the C-terminal of the Gal4 (768-881) activation domain attached to the prey protein. When the N-terminal Gal4 binds to the Gal4 upstream activating sequence (UAS), the bait protein recruits the prey protein, stimulating gene expression downstream. An overview of this type of system is shown in Figure 3, where the PPI between Snf1 and Snf4 was investigated.<sup>79</sup>

A



B



**Figure 3: Diagram showing the first yeast two-hybrid system.<sup>79</sup>** (A) Snf1 and Snf4 are recombinantly expressed with the Gal4 DNA-binding domain (BD) and Gal4 activation domain (AD), respectively. Gal4-BD-Snf1 binds upstream to the UAS, upstream of the reporter gene, *lacZ*. (B) Gal4-AD-Snf4 interacts with Gal4-BD-Snf1 and recruits the transcription machinery of *lacZ*.

Subsequently, the yeast THS was successfully used to identify homodimeric and heterodimeric interactions between Bcl-2 and other Bcl proteins.<sup>81</sup> The proteins were recombinantly expressed with either the LexA DNA-binding domain or the B42 activation domain, and dimerisation was linked to the expression of the chromogenic reporter, *lacZ*. This study provided evidence for Bcl-2 homodimerisation and its interaction with Mcl-1, Bcl-X-L and Bcl-S. Furthermore, the THS was used to demonstrate how homodimerisation of Bcl-2 required two specific domains using truncated forms of the protein. Residues 1-82 are required for interaction with residues 83-218.<sup>81</sup> This demonstrates the capacity for the yeast THS to be used for the discovery and characterisation of PPIs.

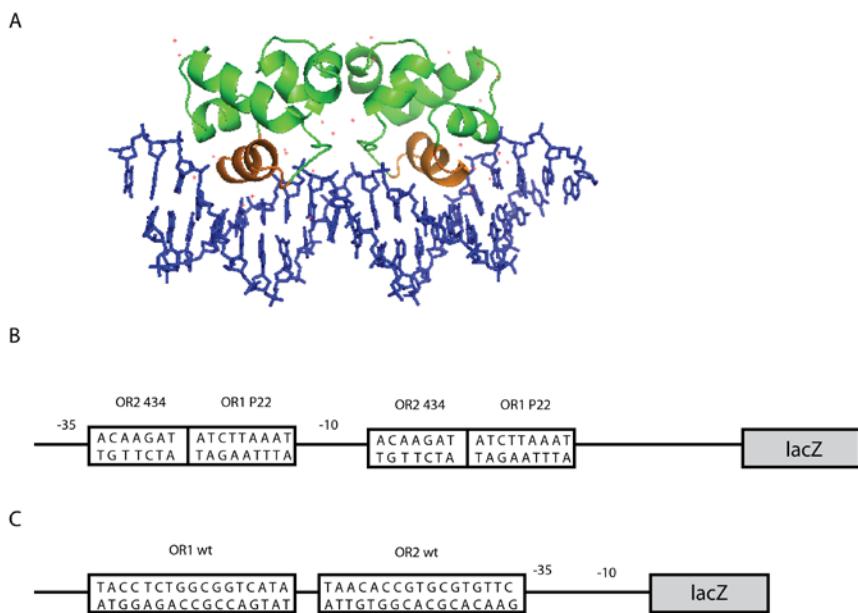
Nevertheless, the use of THS within yeast has disadvantages. Yeast growth is slow compared to bacterial growth. Similarly, transformation efficiency into bacteria is approximately 10,000-times more efficient than the transformation into yeast, thus using bacteria would allow larger genetic libraries to be screened. Consequently, the use of THS was expanded into bacteria and further developed to use a reporter construct that leads to cell survival on selective media.<sup>82</sup>

The bacterial reverse two-hybrid system (RTHS), used in this study, is based on a bacteriophage regulatory system that controls the switch from the lysogenic cycle to the lytic cycle within λ-phages.<sup>83</sup> The repressor, a homodimer of two cI proteins, binds to the right operator on the chromosome of the phage. It is located between two promoters; this region contains three operator regions (O<sub>R</sub>1, 2 and 3). The repressors bind to the DNA-binding domain by inserting two identical α-helices into the major grooves (similar to that observed for the 434-DNA-binding domain in Figure 4A).<sup>84</sup> In the lysogen mode, binding of the cI protein to O<sub>R</sub>1 and O<sub>R</sub>2 blocks transcription of the rightward promoter and activates the leftward promoter, in turn initiating transcription of the *cI* gene. Damage to host DNA or an environment which promotes the lytic cycle, leads to self-cleavage of the cI proteins, resulting in an inability to bind to the O<sub>R</sub>1 and O<sub>R</sub>2 sites as they are not dimerised. This causes transcription of the rightward promoter initiating the expression of early genes

involved in the lytic cycle (e.g. *cro* gene), consequently switching the phage from its dormant state (lysogenic) to an inductive state.<sup>85,86</sup>

The bacterial RTHS uses the cI proteins and the rightward promoter system to link the interaction of chimeric proteins to the expression of downstream reporter genes. The  $\lambda$  cI proteins recognise the corresponding operator DNA, a palindromic sequence, and as a dimeric species it inhibits transcription.<sup>84,87</sup> The  $\lambda$  cI-repressor contains two domains: the N-terminal domain that controls the contact with the operator; and the C-terminal domain, which permits the dimerisation required for binding.<sup>88</sup> The initial development of RTHS utilised the  $\lambda$  repressor, whereby the cI N-terminal domain forms a chimeric protein with one of the dimerising proteins of interest.<sup>89-91</sup> Upon dimerisation the proteins are able to bind to the operator and confer immunity to  $\lambda$  infection.<sup>89-91</sup>

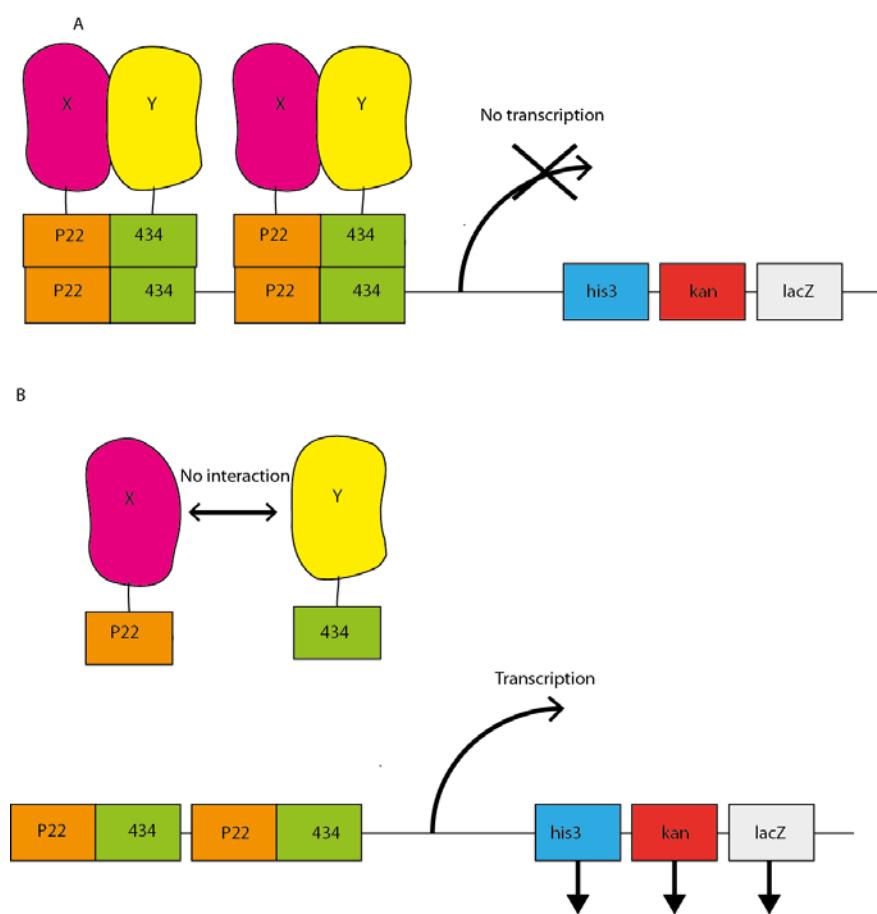
For a homodimeric system the coliphage 434-repressor was utilised. In order to generate a heterodimeric system, the recognition helix of the coliphage 434-repressor was replaced by that of *Salmonella* phage P22 to allow the formation of mixed dimers that recognise a hybrid, non-symmetric 434/P22 operator (Figure 4B).<sup>92,93</sup> In both systems there is a reporter gene construct downstream of a transcription promoter. The systems have been successfully used with a *lacZ* as the reporter gene to identify PPIs involved in bacterial septosome differentiation and the oligomerisation of the histone-like, nucleoid-structuring protein (H-NS), which is a transcriptional repressor.<sup>93-95</sup> The latter has been modified further to show tetramerisation of the H-NS protein using the wild-type operator for the H-NS repression (Figure 4C). Dimeric H-NS binds to the high-affinity operator (OR1<sub>wt</sub>). Once bound the tetramer can form, binding to the low-affinity operator (OR2<sub>wt</sub>), inhibiting transcription by blocking the RNA polymerase from binding to the -35 element.<sup>95</sup>



**Figure 4: Diagram showing different operator sequences.** (A) Shows how the 434-coliphage operator binds with the 434-repressor using two identical  $\alpha$ -helices in the C-terminal (highlighted in orange), structure adapted from PDB 2OR1.<sup>84,92</sup> (B) The heterodimer hybrid operator used by Di Lallo *et al.* (2001)<sup>93</sup> (C) Demonstrates how the traditional RTHS can be modified to utilise wild-type operators, OR1/2<sub>wt</sub> to show tetramerisation.<sup>95</sup>

Along with modifications to the operators and the DNA-binding domains, the reporter construct can also be altered. This permits detection of PPIs through monitoring host survival, and combined with the chromogenic reporter, *lacZ*, allowing quantification. Joung *et al.* (2000) screened libraries of a variety of Cys<sub>2</sub>His<sub>2</sub> zinc finger proteins that bind to TATA and p53 DNA-binding sites using the reporter genes *his3* (imidazole glycerol phosphate dehydratase) and *aadA*.<sup>96</sup> *his3* is the yeast analogue of the bacterial *hisB* gene.<sup>97</sup> In both yeast and bacteria its expression produces a bifunctional enzyme required in histidine biosynthesis; the C-terminal domain is required to catalyse the sixth step, where imidazole glycerol-phosphate is dehydrated to imidazole acetol-phosphate, and the N-terminal is involved in the eighth-step where L-histidinol-phosphate is dephosphorylated to L-histidinol.<sup>98</sup> *aadA* confers spectinomycin resistance. By using reporter genes that affect host survival, the library size that can be screened is increased 100-1000-fold.<sup>96</sup> Also, *his3* reportedly demonstrates a very low false positive rate (approximately  $3 \times 10^{-8}$ ) on histidine-selective media.<sup>96</sup>

Horswill *et al.* (2004) further developed the system to produce a homodimeric and heterodimeric system whereby the reporter gene construct was integrated onto the chromosome of *E. coli* BW27786, a strain with a *hisB* null mutation (residues 165–355).<sup>80</sup> Genes within the reporter construct were *his3*, *kan<sup>R</sup>* and *lacZ*. *kan<sup>R</sup>* (aminoglycoside 3'-phosphotransferase) provides kanamycin resistance; both *kan<sup>R</sup>* and *his3* are chemically tuneable phenotypes that are required for cell viability on selective media that is histidine-deficient and contains kanamycin. The presence of the *lacZ* gene expressing β-galactosidase facilitates quantification of the PPI through the use of an *o*-nitrophenyl β-galactosidase (ONPG) assay (as illustrated in Figure 5).<sup>99</sup>



**Figure 5: The RTHS utilised by Horswill *et al.* (2004) to identify inhibitors of ribonucleotide reductase heterodimerisation and the homodimerisation of HIV protease.<sup>80</sup>** (A) Proteins (X and Y) interact and the DNA-binding domains can bind to the operator and inhibit transcription of reporter genes. (B) Shows how when the proteins (X and Y) do not interact transcription can occur allowing expression of the reporter genes.

In general when the chimeric proteins interact they bind to the DNA operator region inhibiting the transcription of essential genes, causing cell death on selective media. However, if the proteins do not interact they cannot bind to the DNA operator and so transcription of the reporter gene construct is permitted allowing growth on selective media. Horswill *et al.* (2004) subsequently combined this RTHS with a genetic screening method that produced cyclic peptides *in vivo* using an intein-mediated process known as SICLOPPS (split-intein circular ligation of peptides and proteins; described further in Section 1.5).<sup>80,100</sup> Successful integration of the two novel platforms was demonstrated using the homodimeric PPI of the HIV protease and the heterodimeric PPI of ribonucleotide reductase.<sup>80</sup> The linear peptides TVSYEL and FTLDADF (a C-terminal mimic of ribonucleotide reductase), were initially tested to confirm the chemical tuneability of the RTHS and the ability to use the RTHS to screen for modulators of PPIs. A SICLOPPS library (described further in Section 1.5) was then screened against the ribonucleotide reductase PPI. Of the  $10^8$  cyclic peptides screened, 262 potential inhibitors were isolated, 24 of which were shown to be either false positives or non-specific. These were then ranked based on activity, revealing eight key sequences with inhibitory effects; each included the recurring motif: Ar-X-F (where Ar is an aromatic amino acid; X is any amino acid; and F is phenylalanine).<sup>80</sup> This active motif was consistent with the literature, where the tripeptide W/FXF was found to be as active as the C-terminal mimic of ribonucleotide reductase FTLDADF.<sup>80,101</sup>

Bacterial RTHS have also been utilised to identify inhibitors using commercially available libraries. For example, a commercially available library was screened against the DNA-binding domain of Ic1R and a green fluorescent protein (GFP) reporter gene, the inhibitor I-8-15 disrupted the homodimerisation of histidine kinase.<sup>102</sup> This PPI is essential in the bacterial signal transduction pathway. There are several benefits to targeting homologous pathways within bacteria, which include the potential for broad-spectrum activity of the inhibitors. For example, I-8-15 was shown to have dual activity against *S. auerus* and vancomycin-resistant *Enterococcus faecalis*.<sup>102</sup> Nevertheless, with commercial libraries the number of compounds tested is limited. An alternative approach uses split-inteins to generate a library of more than 3.2 million cyclic peptides that can be tested in tandem,

providing a higher-throughput method of inhibitor screening than traditional methods.<sup>80</sup>

## **1.5 Split-intein circular ligation of peptides and proteins**

Inteins are internal protein sequences that post-translationally self-excise from the host protein and catalyse the ligation of the exteins producing a stable protein. Two groups independently discovered the first inteins in 1990; both of which isolated the *S. cerevisiae* vacuolar H<sup>+</sup>-ATPase subunit.<sup>103,104</sup> Although initially presented as two different genes (i.e. *VMA1* and *TFPI*), it was confirmed that both were in fact the same gene, each conferring trifluoperazine resistance and calcium ion dependent growth.<sup>105</sup> During the characterisation of H<sup>+</sup>-ATPase, the gene that was sequenced predicted a 118.6 kDa protein, but a protein with a molecular weight of 69 kDa was isolated instead. This 69 kDa species showed 70% homology with the H<sup>+</sup>-ATPase from other organisms of the C- and N-terminus (residues 1-284 and 739-1,071, respectively), although there was no homology observed within the central region of the gene sequence.<sup>103,104</sup> This region was subsequently shown to have 30% identity with the yeast HO endonuclease and 34% identity with homothallic switching endonuclease.<sup>104,106</sup> Intron splicing would explain the apparent loss of the middle portion of the protein, however, the detected mRNA was in the non-spliced form and intron sequences observed in yeast were absent, intimating it should be a post-translational modification.<sup>103</sup>

There are four main types of inteins: maxi-, mini-, *trans*-splicing and alanine-inteins, as illustrated in Figure 6A-D. VMA1 is an example of a maxi-intein, whereby the N-terminal domain is fused to a homing endonuclease sequence followed by the C-terminal domain.<sup>103,104,106</sup> In contrast, mini-inteins do not contain the homing endonuclease domain. Initially these were only produced artificially and not found in nature.<sup>107</sup> However, more recently, naturally occurring mini-inteins have been observed in the *Mycobacterium xenopi gryA* gene.<sup>108</sup> *Trans*-splicing inteins are two separate fragments each encoding a portion of the intein fragment and a portion of the extein. The first naturally-occurring *trans*-acting intein capable of protein *trans*-splicing was isolated from a cyanobacterium, *Synechocystis sp.* PCC6803, where the N- and C-terminal halves formed the catalytic α-subunit of DNA polymerase III

DnaE and are encoded by *dnaE-n* and *dnaE-c* genes.<sup>109,110</sup> Finally, the alanine-inteins contain an N-terminal alanine residue, where characteristically a cysteine is observed.<sup>111</sup>

A



B



C

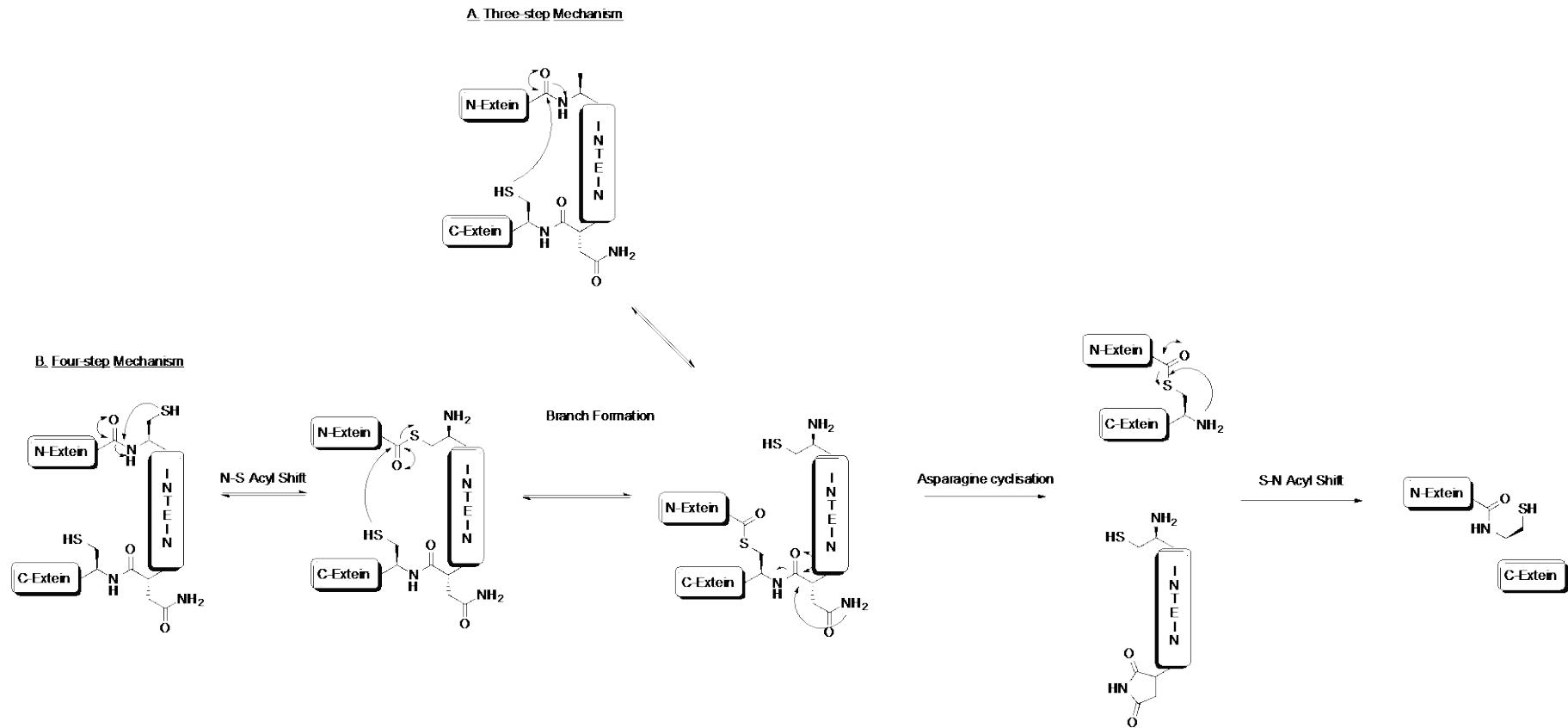


D



**Figure 6: Illustration of the four types of inteins.** (A) The maxi-intein containing the intein, extein and homing endonuclease sequence. (B) The mini-intein where the homing endonuclease domain is absent. (C) The *trans*-acting inteins where the proteins are split. (D) The alanine-inteins, which have an alanine in the position typically occupied by a cysteine.

The discovery of alanine-inteins prompted further studies into the mechanism by which inteins splice. Typically, splicing occurs via a four-step mechanism.<sup>112</sup> Splicing is initiated by an N-S acyl shift (when the N-terminal residue is a cysteine) then a transesterification to form the lariat intermediate. The asparagine then cyclises releasing the exteins with the newly formed peptide bond, which can then undergo a spontaneous acyl rearrangement to form the thermodynamically favoured native peptide bond (Figure 7A).<sup>112-115</sup> In the absence of the catalytic cysteine (i.e. in alanine-inteins) the N-S acyl shift cannot occur, therefore, it proceeds via a three-step mechanism, whereby the nucleophile directly attacks the peptide bond of the N-extein (Figure 7B).

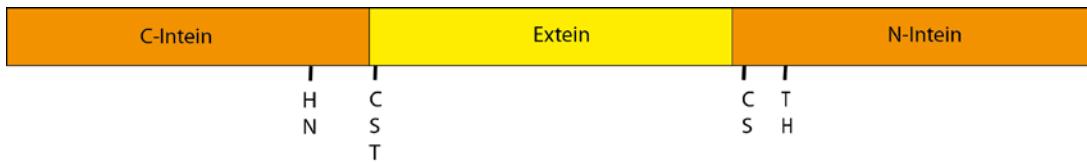


**Figure 7: Scheme showing the proposed mechanism for intein processing.** (A) Three-step mechanism of alanine-inteins. (B) Maxi- and mini-intein splicing.

Inteins are a valuable tool in chemical biology; their applications range from protein purification to PPI modulators.<sup>116,117</sup> Inteins have also been used for the production of cyclic peptides. The desire for peptides that modulate PPIs has greatly increased in the last 20 years, as they represent a valuable tool for mediating biological processes, especially acting as PPI inhibitors. The use of short cationic amphiphilic molecules as antibiotics has been extensively reviewed by Hancock *et al.* (1999 and 2006).<sup>118,119</sup> The progress of the peptides in clinical trials and development is explored in each review. An example of a clinically available topical peptide antibiotic is gramicidin S, a cyclic decapeptide consisting of two penta-peptides joined head-to-tail.<sup>120</sup>

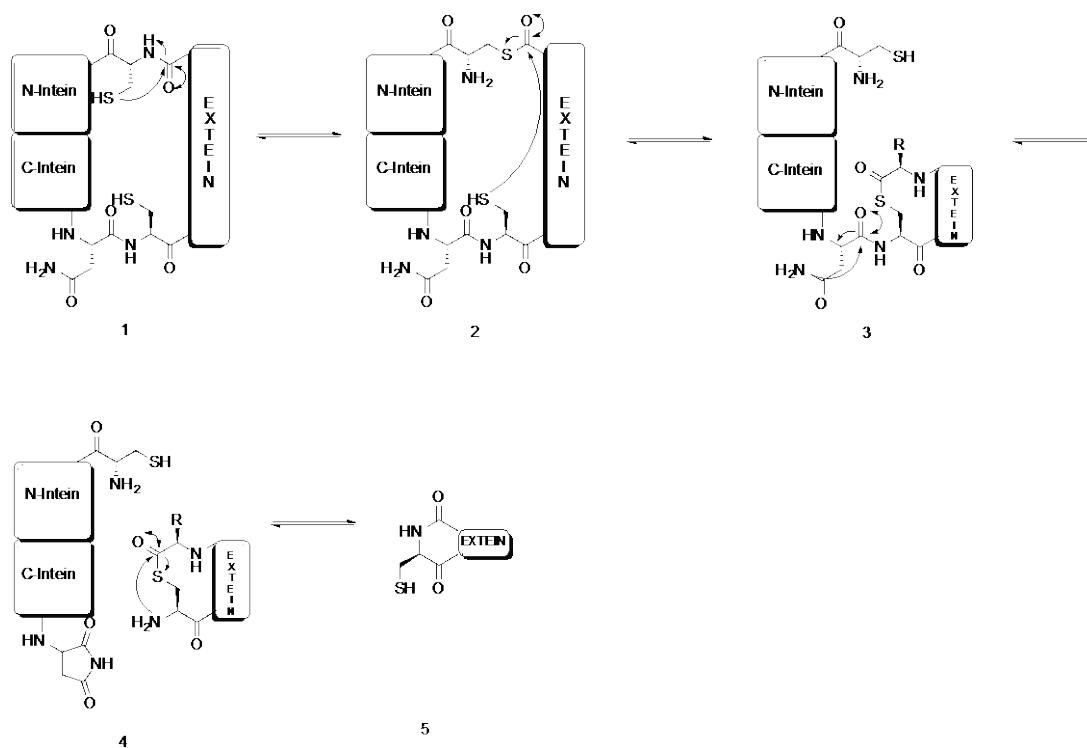
Despite the commercial success of gramicidin S many peptides have had limited success in clinical trials. Many peptides are toxic systemically and to ameliorate this, treatment is classically administered via a topical route. Another driver for topical rather than systemic delivery relates to the membrane permeability of these peptides. Many have high permeability in relation to bacterial cell membranes, but are unable to easily pass into eukaryotic cells owing to the more complex nature of their plasma membranes. The stability of peptides *in vivo* is also a disadvantage, with many being highly susceptible to degradation by naturally occurring proteases. Nevertheless this can be overcome using cyclic peptide inhibitors.<sup>121</sup> The largest disadvantage is the cost of production, yet there is potential to optimise a cyclic peptide into a small molecule, which might avoid the pitfalls classically observed by cyclic peptides.<sup>122</sup> Cyclic peptides can be used as a platform to identify active motifs and binding pockets, and then further developed into a more bioavailable molecule.

SICLOPPS is a robust method for the biosynthesis of cyclic peptides using the *S. sp.* PCC6803 DnaE split intein to facilitate cyclisation. The system is able to produce the cyclic peptides rapidly *in vivo*.<sup>117</sup> Both *cis*- and *trans*-acting types of inteins, described previously, produce linear proteins.<sup>123</sup> Cyclic peptides provide significant stability against cellular degradation by the cells catabolic machinery, as there are no exposed termini so they are not susceptible to cellular exoproteases.<sup>124</sup> Altering the intein sequence described above can produce these cyclic peptides; the extein (the desired cyclic peptide sequence) was inserted between the split-inverted inteins (C-intein: extein: N-intein, as illustrated by Figure 8).



**Figure 8: Split-inverted intein used for SICLOPPS.** The extein or degenerate region is flanked by two inteins.<sup>100</sup>

As before the inteins undergo an autocatalytic process; however, the initiating step involves the association of the two inteins, forming an active intein.<sup>117</sup>



**Figure 9: Scheme for the mechanism of the split inverted intein using SICLOPPS.** First there is an N-S acyl shift, followed by a transesterification. The asparagine side-chain then undergoes cyclisation, which liberates the mature protein that can then rearrange to form the thermodynamic product.<sup>117</sup>

The potential of SICLOPPS was first demonstrated by Scott *et al.* (1999) when it was used to express cyclic dihydrofolate reductase (DHFR) and pseudostellarin F.<sup>100</sup> Cyclic DHFR produced using SICLOPPS was active in enzyme assays, producing identical steady-state kinetic parameters and substrate, cofactor and methotrexate dissociation constants to wild-type DHFR. Similarly, the pseudostellarin F was successful in inhibiting tyrosinase *in vivo*. Furthermore, this study investigated the structural requirements for the biosynthesis of a cyclic peptide library. For example, using pseudostellarin F the importance of the first residue was probed, the serine residue in the first position was mutated to either a cysteine or threonine, as this residue is essential for the nucleophilic step in intein-processing, with the cysteine being the most efficient *in vivo*. All three amino acids permitted full intein processing. Similarly, the final residue of the extein sequence was probed, with variable processing results observed, dependent on the new residue, with glutamic acid, asparagine and proline failing to process fully. The failure of proline to process was not unexpected as it is the third consecutive proline in the sequence, and therefore processing would likely be encumbered by steric hindrance. The length of the extein sequence was subsequently analysed, and the processing of peptides as small as four amino acids was observed by mass spectrometry. This study was important as it showed how the potential for systematic bias in SICLOPPS library generation was limited.<sup>100</sup>

SICLOPPS was further developed to generate libraries of cyclic peptides that could be used *in vivo* as PPI modulators.<sup>117</sup> The expansion of SICLOPPS to generate cyclic peptide libraries was achieved at the oligonucleotide level.<sup>117</sup> When each codon in the extein region is in the form NNS (where N codes for any of the four DNA bases, and S represents cytosine, guanine or thymine), each insert generated by polymerase chain reaction (PCR) is unique. By eliminating adenine in the third position, this allows all twenty of the naturally-occurring amino acids to be produced, but excludes the stop codons UAA (ochre) and UGA (opal). UAG (amber), however, will still occur, but only in 1 in every 32 codons. This prevents the majority of the termination of the transcription and translation of the sequence.<sup>117</sup> For library generation, the extein region was limited to either a pentameric or hexameric degenerate sequence. This is because the theoretical number of unique cyclic peptides in a pentameric

library is 3.2 million, therefore the transformation efficiency of electro-competent *E. coli* exceeds the maximum library size by ten-fold.<sup>117</sup>

Scott *et al.* (2001) generated SX<sub>5</sub> and SGX<sub>5</sub>PL libraries.<sup>125</sup> These libraries showed a minor bias whereby arginine and serine were under-represented whilst leucine and glutamic acid were over-represented. Additionally, only 68% of the examined clones yielded cyclic peptides. Yet, it is difficult to determine if this is statistically significant as only a small number of library members were characterised (19 in total).<sup>100</sup> These biases limit the SICLOPPS technology; nevertheless, as a tool to identify PPI inhibitors, SICLOPPS provides a starting point, where the inhibitors can be further characterised to identify the active motif and protein-binding pocket, and can be optimised to produce a more potent inhibitor. Spurr *et al.* (2012) took a cyclic peptide inhibitor identified by Tavassoli and Benkovic (2005) targeted against the aminoimidazole carboxamide ribonucleotide transformylase (ATIC) homodimeric enzyme and identified the active motif as RY by alanine scanning.<sup>122, 126</sup> Non-natural analogues were synthesised and tested *in vitro*. The original peptide identified by SICLOPPS screening had a K<sub>i</sub> of only 17 μM, whereas the most potent analogue showed a 25-fold improvement in efficacy (K<sub>i</sub> of 685 nM).<sup>122</sup>

SICLOPPS has been used to generate many peptide inhibitor libraries and screened for PPI inhibitors within a variety of assays.<sup>127-130</sup> As described previously, Horswill *et al.* (2004) combined the SICLOPPS technology with the bacterial RTHS to screen for inhibitors of PPIs including homodimeric PPI of HIV protease and the heterodimeric ribonucleotide reductase interaction.<sup>80</sup> The combination of the systems has also been utilised to identify modulators of the interaction between HIV Gag protein and the human host protein TSG101,<sup>131</sup> the homodimer ATIC,<sup>126</sup> and the interaction of Hdmx or Hdm2 with p53.<sup>132</sup> More recently, the systems have been used to identify inhibitors of the dimerisation of the C-terminal binding proteins (CtBP)<sup>133</sup> and the heterodimeric interaction between hypoxia inducible factor (HIF)-1α and HIF-1β.<sup>134</sup>

CtBP1 and 2 are transcriptional repressors of genes that promote cancer proliferation and survival. The homodimerisation of CtBP1 and 2 respond to increased NADH

levels, however, the role of the dimer compared to the monomer has not been fully characterised. Therefore a modulator of dimer formation could be used to identify which pathways CtBP1 and 2 regulate. The RTHS and SICLOPPS screening yielded *cyclo-SGWTVVVRMY*.<sup>133</sup> This peptide disrupts the homodimerisation of CtBP1 and CtBP2 with an half maximal inhibitory concentration ( $IC_{50}$ ) of 24.9  $\mu$ M and 19.7  $\mu$ M, respectively. *Cyclo-SGWTVVVRMY* was utilised *in vivo* to show how CtBP dimerisation is required for maintaining mitosis of cancer cells.<sup>133</sup>

An inhibitor of HIF dimerisation (*cyclo-CLLFVY*) is a further example of how SICLOPPS and the bacterial RTHS can be used as a high-throughput platform for the identification of PPI inhibitors. HIF is a heterodimeric transcription factor (composed of HIF-1 $\alpha$  and HIF-1 $\beta$ ) that regulates cellular responses to low oxygen (hypoxia). This response is essential in tumour survival and proliferation. *Cyclo-CLLFVY* has been shown to function both *in vitro* and *in vivo*, with the potential for further optimisation and development as a novel cancer therapeutic. Importantly, it was the first example of a specific HIF-1 inhibitor that does not affect the HIF-2 isoform.<sup>134,135</sup> Firstly, the activity of *cyclo-CLLFVY* was tested by ELISA to show disruption of the PPI *in vitro* ( $IC_{50}$  of 1.3  $\mu$ M).<sup>134</sup> Secondly, the binding of *cyclo-CLLFVY* to the Per-Arnst-Sims-B domain of HIF-1 $\alpha$  was shown by ITC which gave a binding affinity of  $124 \pm 23$  nM. This finding was supported by a fluorescent binding assay using a Megastoke 673 derivative of *cyclo-CLLFVY* and a pull-down assay using streptavidin beads coated with biotin-polyethylene glycol (PEG)-triazole-*cyclo-ALLFVY* compound. Furthermore, the effect of the inhibitor on the hypoxia response in MCF-7 and U2OS mammalian cancer cell lines was investigated. By studying the expression of vascular endothelial growth factor, a downstream target regulated by HIF-1 dimerisation at the transcriptional level, changes in hypoxia signalling and HIF dimerisation could be monitored. A three-fold increase in transcription of vascular endothelial growth factor was observed by quantitative PCR, which was attenuated in the presence of *cyclo-CLLFVY*.<sup>134</sup>

CtBP and HIF are good examples of where the dual RTHS-SICLOPPS platform has been used against PPIs essential for tumourgenesis.<sup>133,134</sup> The rapid nature of the SICLOPPS technology combined with the RTHS platform provides a high-

throughput screen of millions of cyclic peptides for inhibitory effect on desired PPIs unparalleled by traditional synthetic libraries. Here lies a potential methodology for providing novel inhibitors of virulence factors for further development into antimicrobials.

## **1.6 Overview of the project**

There is a growing need for novel targets for antimicrobials. Current literature suggests that targeting virulence factors will slow or negate the development of resistance.<sup>38, 36</sup> Here, two virulence factors from two bacterial species have been selected to demonstrate that the bacterial RTHS and SICLOPPS can be used in tandem to identify antimicrobials targeted against bacterial virulence factors. In Chapter 3 the phage shock protein response from *Burkholderia pseudomallei* is targeted by identifying inhibitors of phage shock protein A (PspA) oligomerisation. Chapter 4 then utilises the screening platform to identify inhibitors of *Bacillus anthracis* toxin internalisation. These represent examples of both homodimeric and heterodimeric systems. The most potent inhibitors identified, in both cases, were then tested *in vitro* to establish their efficacy.



## 2 Materials and methods

Reagents used for synthesis were purchased from Sigma Aldrich (UK) or Novabiochem (Merck), and used unmodified. PCR was carried out either with an Eppendorf Mastercycler or BioRad MyCycler Thermal Cycler. DNA purification was carried out using QIAGEN DNA mini- and maxi-prep kits (QIAGEN, Germany). DNA concentration was measured using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, USA). Agarose and sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) gels were run using BioRad Power Pack Basic and imaged on BioRad Universal Hood II Imager using Quantity One software from BioRad (BioRad Laboratories Ltd., UK).

Transformation by electroporation was carried out on the BioRad MicroPulser electroporator. Optical density of cell culture was measured using a Cary 100Bio UV-Visible spectrometer (Agilent Technologies, UK). Enzyme X, pDraw and Serial Cloner were used for selecting restriction endonucleases, calculating annealing temperatures and primer design. Sequencing was done by Source Bioscience (Oxford, UK) or Eurofins MWG (Germany). Molecular biology reagents were purchased from New England Biolabs (NEB) or Promega. Size exclusion chromatography (SEC) was carried out with a HiLoad<sup>TM</sup> 26/600 Superdex<sup>TM</sup> 200 prep grade column or a Superose 6 gel filtration column (GE Healthcare, UK) and affinity purification was carried out on 1 or 5 ml HisTrap HP columns/ GSTTrap HP columns (GE Healthcare, UK) using an AKTAPrime<sup>TM</sup> Plus FPLC (GE Healthcare, UK). HPLC was carried out on a HPLC Waters system (Waters, UK). Mass spectrometry was carried out on a Waters 2MD Single Quadrupole mass spectrometry (MS) by the University of Southampton service. <sup>1</sup>H, COSY and NOSEY data was obtained using Bruker DPX400/1 nuclear magnetic resonance (NMR) spectrometer at the University of Southampton (UK). ELISA results were measured using a Tecan Infinite M200 Pro micro-plate reader.

## 2.1 Bacterial strains

**Table 1:** Bacterial strains.

STRAIN	GENOTYPE
DH5α	<i>fhuA2 lac(del)U169 phoA glnV44 Φ80' lacZ(del)M15</i> <i>gyrA96 recA1 relA1 endA1 thi-1 hsdR17</i>
NEB-5-α	<i>fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80Δ</i> <i>(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17</i>
NEB-10-β	<i>F endA1 recA1 galE15 galK16 nupG rpsL ΔlacX74</i> <i>Φ80lacZΔM15 araD139 Δ(ara,leu)7697 mcrA Δ(mrr-hsdRMS-mcrBC) λ</i>
NEB Express	<i>fhuA2 [lon] ompT gal sulA11 R(mcr-73::miniTn10--Tet<sup>S</sup>)2 [dcm] R(zgb-210::Tn10--Tet<sup>S</sup>) endA1 Δ(mcrC-mrr)114::IS10</i>
BL21 (DE3)	<i>fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS</i> <i>λ DE3 = λ sBamHIo ΔEcoRI-B</i> <i>int::(lacI::PlacUV5::T7 gene1) i21 Δnin5</i>

## 2.2 Primers

**Table 2:** Primer sequences, restriction digest sites underlined.

PRIMER	SEQUENCE
GENERAL PRIMERS	
M13-49	GAG CGG ATA ACA ATT TCA CAC AGG
M13-43	AGG GTT TTC CCA GTC ACG ACG TT
T7-F	TAA TAC GAC TCA CTA TAG GG
T7-R	GCT AGT TAT TGC TCA GCG G
pGEX-F	GGG CTG GCA AGC CAC GTT TGG TG
pGEX-R	CCG GGA GCT GCA TGT GTC AGA GG
P1	GGA ATC AAT GCC TGA GTG
P2	ACT TAA CGG CTG ACA TGG
P3	ACG AGT ATC GAG ATG GCA

P4	GGC ATC AAC AGC ACA TTC
<hr/>	
SICLOPPS PRIMERS	
FLC+5	GGA ATT <u>CGC CAA TGG GGC</u> GAT CGC CCA CAA TGC NNS NNS NNS NNS NNS TGC TTA AGT TTT GGC ACC G
InR-BgII-R	ACG AAT TCA AAA ACC GTG GCT <u>CTA GAA TTG</u> GCA A
SGWX <sub>5</sub> -F	GGA ATT <u>CGC CAA TGG GGC</u> GAT CGC CCA CAA TAG CGG CTG GNN SNN SNN SNN SNN STG CTT AAG TTT TGG C
CX <sub>5</sub> -F	GGA ATT <u>CGC CAA TGG GGC</u> GAT CGC CCA CAA TTG CNN SNN SNN SNN SNN STG CTT AAG TTT TGG C
SX <sub>5</sub> G-F	GGA ATT <u>CGC CAA TGG GGC</u> GAT CGC CCA CAA TAG CNN SNN SNN SNN SNN SGG CTG CTT AAG TTT TGG C
SX <sub>6</sub> G-F	GGA ATT <u>CGC CAA TGG GGC</u> GAT CGC CCA CAA TAG CNN SNN SNN SNN SNN SNN SGG CTG CTT AAG TTT TGG
CBD-R	GGA ATT CAA GCT TTC ATT GAA GCT GCC ACA AGG
Zipper	GGA ATT CGC CAA TGG GGC GAT CGC C
<hr/>	
PSPA REVERSE-TWO HYBRID PRIMERS	
PspA( <i>Bam</i> HI)-F	CGG CGG <u>GGA TCC</u> ATG TCG CTT TTC GAC TCT ATT TCG CGC
PspA( <i>Sac</i> I)-R	GGC GGC <u>GAG CTC</u> TTA CTG CGC CGG CGT GTT CAG TTG C
SDM PspA-F	GAT CGT GCG GGA GCT TGA CGA CAG
SDM PspA-R	CTG TCG TCA AGC TCC CGC ACG ATC
<hr/>	
<i>E. coli</i> PSPA EXPRESSION PRIMERS	
ECPspA( <i>Bam</i> HI)-F	CGG CGG <u>GGA TCC</u> ATG GGT ATT TTT TCT CGC TTT GCC
ECPspA( <i>Sac</i> I)-R	GGC GGC <u>GAG CTC</u> TTA TTGA TTG TCT TGC

**B. PSEUDOMALLEI PSPA EXPRESSION PRIMERS**

PspA(*Bam*HI)-F CGG CGG GGA TCC ATG TCG CTT TTC GAC TCT  
ATT TCG CGC

PspA(*Sac*I)-R GGC GGC GAG CTC TTA CTG CGC CGG CGT GTT  
CAG TTG C

**B. PSEUDOMALLEI PSPA INHIBITOR MUTANTS**

I<sub>C</sub> (H24A, F26A)-F CGG CGG CCA TGG TTA AAG TTA TCG G  
I<sub>C</sub> (H24A, F26A)-R GGG CGA TCG CCC CAT TGG CTA GCA GAG CAT

TAG CGT CTT GGG G

I<sub>N</sub> (T69A, H72A)-F TAA CTG CTT AAG TTT TGG CAC CGA G

I<sub>N</sub> (T69A, H72A)-R GTT TAA TGG TAC CAG CGT CAA GTA ATG G

SGW1-F CGG CGG CGC CAA TGG GGC ATC GCC CAC  
AAT AGC GGC TGG CTT GTC TAT TGG TTC TGC  
TTA AGT TTT GGC

SGW11-F CGG CGG CGC CAA TGG GGC ATC GAC CAC  
AAT AGC GGC TGG AGC ATT GTT TTC ATC TGC  
TTA AGT TTT GGC

SGWIYWNV-F CGG CGG CGC CAA TGG GGC ATC GCC CAC  
AAT AGC GGC TGG ATT TAT TGG AAC GTG TGC  
TTA AGT TTT GGC

**MALTOSE-BINDING TAGGED PSPA PRIMERS**

PspA(*Xmn*I)-F CGG CGG GGA AGG ATT TCA TGT CGC TTT TCG  
ACT CTA TTT CG

PspA(*Bam*HI)-R GGC GGC GGA TCC TTA CTG CGC CGG CGT GTT  
CAG

**PROTECTIVE ANTIGEN AND RECEPTOR EXPRESSION PRIMERS**

CMG2(*Bam*HI)-F CGG CGG GGA TCC ATG TCT TGC AAA AAA GCC  
TTC G

CMG2(*Xma*I)-R GGC GGC CCC GGG TTA TCC TAT CTC ATA GC

PA(*Bam*HI)-F CGG CGG GGA TCC ATG ACA ACT CA CGT ATC ATT  
TTT AAT GG

PA(*Sac*I)-R GGC GGC GAG CTC TTA TCC TAT CTC ATA GCC

CMG2(*Sac*I)-R GGC GGC GAG CTC TTA GTT TAA CAT GAT TGA GC

## 2.3 General preparation of materials

### 2.3.1 *Luria Broth (LB) medium*

6.25 g of LB powder (Fisher Scientific, UK) was suspended in 250 ml of sterilised distilled water and autoclaved for 20 minutes at 115°C on the liquid cycle.

### 2.3.2 *Luria Bertani (LB) agar plates*

10 g LB agar mixture (Fisher Scientific, UK) was dissolved in 250 ml of sterilised distilled water and autoclaved for 20 minutes at 115°C on the liquid cycle.

### 2.3.3 *Agar plates*

3.8 g agar mixture (Fisher Scientific, UK) was dissolved in 200 ml of sterilised distilled water and autoclaved for 20 minutes at 115°C on the liquid cycle.

### 2.3.4 *Antibiotic concentrations*

**Table 3: Typical antibiotic concentrations used during cloning.**

---

ANTIBIOTIC	STOCK CONCENTRATION MG/ML	WORKING CONCENTRATION µG/ML	
		PLASMID	CHROMOSOME
Ampicillin	20	100	50
Chloramphenicol	10	35	N/A
Kanamycin	10	50	25
Spectinomycin	10	50	25

---

### 2.3.5 *Super optimal catabolite medium*

To 50 ml LB media, 1 M MgCl<sub>2</sub> (500 µl), 2 M MgSO<sub>4</sub> (500 µl) and of 20% (w/v) glucose (500 µl) were added. Filter sterilised (0.45 µM Millipore filter) and stored at 4°C.

## 2.4 General molecular biology procedures

### 2.4.1 *Plasmid isolation*

Plasmids were purified using QIAGEN mini- or maxi-prep purification kit as per manufacturer's instructions.

### 2.4.2 *Polymerase chain reaction*

Reaction mixture was prepared as in Table 4.

**Table 4: Composition of GoTaq PCR reaction mixture.**

REAGENT	VOLUME	FINAL CONCENTRATION
GoTaq polymerase	0.5 µl	1.25 U
GoTaq buffer	10 µl	5 x concentration
10 mM dNTPs	5 µl	0.2 mM each dNTP
Forward and reverse primers	1 µl each	100 pmol/µl
Template DNA	1 µl	100 ng
Sterile deionised water	31.5 µl	Up to 50 µl

The PCR program was set up as shown in Table 5.

**Table 5: PCR Program for GoTaq polymerase.** Where X is the annealing temperature of the primers and Y is the extension time, the extension time with GoTaq polymerase is 1,000 base pairs (bp) per 1 minute.

TEMPERATURE	TIME	NUMBER OF CYCLES
95°C	2.5 minutes	
95°C	1.5 minutes	- \
X°C	30 seconds	30 cycles
72°C	Y minutes	- /
72°C	10 minutes	
4°C	∞	

PCR products were analysed by gel electrophoresis on a 1% agarose (w/v, agarose in tris-acetate- ethylenediaminetetraacetic acid (EDTA) (TAE) buffer (Table 6).

**Table 6: Composition of 50 x tris-acetate-EDTA buffer.**

REAGENT	AMOUNT	CONCENTRATION	FINAL CONCENTRATION
Tris-base	242 g		2 M
Glacial acetic acid	57.1 ml		1 M
EDTA	100 ml	0.5 M	50 mM
Sterile deionised water	To 1000 ml		
Adjusted to pH 8.0 with dilute NaOH			

#### **2.4.3 Site-directed mutagenesis**

Reaction mixture was prepared as in Table 7.

**Table 7: Composition of site-directed mutagenesis reaction mixture.**

REAGENT	VOLUME	FINAL CONCENTRATION
Pfu polymerase buffer	5 µl	10 x concentration
10mM dNTPs	4 µl	10 mM
Forward and reverse primer	0.2 µl each	100 pmol/µl
Template DNA	1 µl	10 ng
Pfu polymerase	1 µl	1.25 U
Sterile deionised water	37.6 µl	Up to 50µl

The PCR program was set up as shown in Table 8.

**Table 8: PCR program for site-directed mutagenesis with Pfu polymerase.**

Where X is the annealing temperature and Y is the extension time.

TEMPERATURE	TIME	NUMBER OF CYCLES
94°C	1 minutes	
94°C	30 seconds	- \
X°C	30 seconds	12 cycles
68°C	Y minutes	_ /
4°C	∞	

The PCR mix was incubated with 1 µl *Dpn*I and 1 µl buffer 4 for one hour at 37°C and then purified on a 1% agarose gel and transform into chemically-competent cells.

#### **2.4.4 *Restriction digestion***

Restriction digestion of plasmids and PCR products was achieved by mixing plasmid with the appropriate buffer for the chosen restriction endonuclease (Table 9). Bovine serum albumin (BSA) was added, if required replacing water. The reaction was incubated at 37°C for 2-16 hours, then analysed by gel electrophoresis. Analytical restriction digests were carried out as described above on a five-fold scale.

**Table 9: Composition of restriction digestions without BSA.** If required the water can be omitted and BSA used in its place.

REAGENTS	VOLUME
Plasmid	38 µl
Buffer	5 µl
Restriction endonuclease 1	1 µl
Restriction endonuclease 2	1 µl
Sterile deionised water	5 µl

#### **2.4.5 Gel purification of digested fragments**

Completed restriction digestion were loaded onto a 1% agarose gel to separate the fragments of interest. The required band was excised and the DNA recovered from the agarose using QIAGEN gel extraction kit as per manufacturer's instructions.

#### **2.4.6 Ligation**

Ligation of digested plasmid backbone and gene insert was set up as detailed in Table 10. Reactions were carried out over degenerate at 4°C and then heat inactivated at 70°C for 10 minutes. The volume of insert and vector for a 1:6 ratio was calculated by Equation 1.

$$\text{Insert mass (ng)} = 6 \times \left[ \frac{\text{Insert length in bp}}{\text{Backbone length in bp}} \right] \times \text{backbone mass (ng)}$$

**Equation 1: Ligation calculation**

**Table 10: Composition of ligation reaction mixture.**

REAGENTS	VOLUME
Ligase buffer	1 µl
Linear backbone	4-7.5 µl
Insert	0.5-4 µl
T4 DNA ligase	1 µl
Sterile deionised water	Up to 9 µl

#### **2.4.7 Preparation and transformation of chemically-competent cells**

A sample of *E. coli* was incubated overnight in 5 ml of LB media. 50 ml LB media was inoculated with 1% (50 µl) of the culture and incubated until the OD<sub>600</sub> reaches between 0.5-0.7. The pellet was obtained by centrifugation (3,100 rpm, 15 minutes, 4°C) and was resuspended in ice-cold transformation buffer (TBF) I buffer (10 ml) (Table 11) and again centrifuged (3,100 rpm, 15 minutes, 4°C). The supernatant was discarded and the cell pellet was resuspended in ice-cold TBFII buffer (2 ml) (Table 12), aliquoted, snap frozen and stored at -80°C.

**Table 11: Composition of TBFI buffer.**

REAGENT	QUANTITY	FINAL CONCENTRATION
Potassium acetate	0.59 g	30 mM
Rubidium chloride	2.42 g	100 mM
Calcium chloride	0.29 g	10 mM
Manganese chloride	2 g	50 mM
Glycerol	30 ml	15% v/v
Sterile deionised water	Up to 200 ml	
Adjusted to pH 5.8 with 1% acetic acid		

**Table 12: Composition of TBFII buffer.**

REAGENTS	QUANTITY	FINAL CONCENTRATION
MOPS	0.21 g	10 mM
Rubidium chloride	0.10 g	10 mM
Calcium chloride	1.10 g	75 mM
Glycerol	15 ml	15% v/v
Sterile deionised water	Up to 100 ml	
Adjusted to pH 6.5 with dilute NaOH		

When required for transformation the cells were thawed on ice and incubated on ice with 5 µl of DNA (100 ng/µl) for 30 minutes. The sample was typically heat shocked for 45 seconds at 42°C. The cells were immediately incubated on ice for 2 minutes then incubated in 895 µl super optimal catabolite (SOC) media at 37°C for one hour before plating onto LB agar plates containing antibiotics and incubated overnight at 37°C.

#### **2.4.8 Preparation and transformation of electro-competent cells**

A sample of *E. coli* was incubated overnight in 5 ml of LB media, and then used to inoculate 250 ml LB media. This culture was incubated until the OD<sub>600</sub> reached between 0.5-0.7. The culture was then centrifuged (3,100 rpm, 15 minutes, 4°C); the cell pellet was resuspended in ice-cold 10% glycerol (250 ml) centrifuged (4,000

rpm, 15 minutes, 4°C). Supernatant discarded and cell pellet was resuspended in ice cold 10% glycerol (125 ml). Repeat cycle of centrifugation with each re-suspension reducing the volume of 10% glycerol each time, until 2 ml 10% glycerol was left. Sample was aliquoted, snap frozen, and then stored at -80°C.

When required for transformation, the cells were thawed on ice. 5 µl of ligation mixture was added and transferred into a 1 mm electroporation cuvette (Biorad, UK). The mixture was electroporated (Ec1 setting, 1.8 kV), mixed immediately with 895 µl SOC and incubated at 37°C for one hour, before plating onto LB agar plates containing antibiotics and incubated overnight at 37°C.

#### ***2.4.9 Transformation of electro-competent cells containing the helper plasmid***

For the transformation of a heat-sensitive helper plasmid, the electroporation was carried out as 2.4.8. However, the recovery mixture was incubated at 37°C for one hour and then at 42°C for 30 minutes before plating onto LB agar plates containing antibiotics and incubated overnight at 37°C.

#### ***2.4.10 Preparation of minimal media agar plates***

Agar was prepared as 2.3.3 and supplemented with reagents in Table 13, with varying concentrations of 3-amino triazole (3-AT) (2.5 mM and 5.0 mM), isopropyl-β-D-1-thiogalactopyranoside (IPTG) (0-50 µM) and kanamycin (25 µg/ml and 50 µg/ml), as required.

**Table 13: Composition of minimal media agar plates.**

REAGENTS	VOLUME	CONCENTRATION
Agar	200 ml	
Minimal media salts (Table 14)	25 ml	10 x solution
50% glycerol	10 ml	4%
1 M MgSO <sub>4</sub>	0.25 ml	1 mM
Ampicillin	0.625 ml	50 µg/ml
Spectinomycin	0.625 ml	25 µg/ml
Sterile deionised water	Up to 250 ml	

**Table 14: Composition of 10 x minimal media salt solution.**

REAGENTS	QUANTITY
Ammonium sulphate	5.0 g
Potassium phosphate monobasic	22.5 g
Potassium phosphate dibasic	52.5 g
Sodium citrate	2.5 g
Sterile deionised water	Up to 500 ml

#### 2.4.11 *Ortho-nitrophenyl β-galactosidase assay*

3 ml LB media was mixed with 1% (30 µl) of each overnight culture, with antibiotics (replicated 6 times per sample) and incubated until OD<sub>600</sub> reached 0.4. Each sample was subsequently incubated with IPTG at varying concentration (i.e. 0, 10, 25, 50, 100, 250 µM) for a further 1.5 hours after which the OD<sub>600</sub> was recorded. 0.3 ml of this culture was added to 0.4 ml Z buffer (Table 15), 1 drop of chloroform and 1 drop 0.1% SDS, and vortex for 10 seconds. Each sample was incubated at 37°C for one minute and then 200 µl of ONPG (4 mg/ml) was added and incubated further at 37°C. The time (in seconds) was recorded for the solution to go yellow, after which time the solution was quenched with 500 µl 1 M Na<sub>2</sub>CO<sub>3</sub> and the OD<sub>420</sub> recorded.

$$\text{Activity} = \frac{\text{OD}_{420}}{\text{time/sec}} \quad \text{Equation 2}$$

$$\text{Standardised activity} = \frac{\text{Activity}}{\text{Volume}} \quad \text{Equation 3}$$

$$\text{Specific activity} = \frac{\text{Standardised activity}}{\text{OD}_{600}} \quad \text{Equation 4}$$

**Table 15: Composition of Z-buffer.**

REAGENTS	QUANTITY	FINAL CONCENTRATION
Na <sub>2</sub> HPO <sub>4</sub> .7H <sub>2</sub> O	0.80 g	60 mM
NaH <sub>2</sub> PO <sub>4</sub> .H <sub>2</sub> O	0.28 g	40 mM
1 M KCl	0.50 ml	10 mM
1 M MgSO <sub>4</sub>	0.05 ml	1 mM
β-Mercaptoethanol	0.14 ml	40 mM
Sterile deionised water	Up to 50 ml	
Adjust to pH 7.0 with dilute NaOH		

## 2.5 General protein procedures

### 2.5.1 Preparing and running an SDS-PAGE gel

Prepare the gel using the recipe in Table 16 and Table 17

**Table 16: Composition of resolving gel.**

REAGENTS	AMOUNTS		
	10%	12%	15%
Sterile deionised water	9.9 ml	10.2 ml	7.0 ml
1.5 M Tris-base (pH 8.8)	6.3 ml	7.5 ml	7.5 ml
10% (w/v) SDS	0.25 ml	0.30 ml	0.30 ml
Acrylamide/bisacrylamide (30%/0.8%) (w/v)	8.3 ml	12.0 ml	15.0 ml
10% (w/v) ammonium persulphate	0.25 ml	0.25 ml	0.15 ml
Tetramethyleneethylenediamine (TEMED)	0.02 ml	0.02 ml	0.02 ml

**Table 17: Composition of stacking gel.**

REAGENTS	CONCENTRATION	AMOUNTS
Sterile deionised water		3.075 ml
1.5 M Tris-base (pH 8.8)	1.5 M	1.25 ml
20% (w/v) SDS	20% (w/v)	0.025 ml
Acrylamide/bisacrylamide (30%/0.8%) (w/v)		0.67 ml
Ammonium persulphate	10% (w/v)	0.025 ml
TEMED		0.005 ml

---

The samples were mixed with 2 x loading dye (Table 18) and incubated at 100°C for 10 minutes to denature the protein. The sample was then loaded and run on an SDS-PAGE gel in running buffer (Table 19).

**Table 18: Composition of the 2 x loading dye.**

REAGENTS	CONCENTRATION	AMOUNTS	FINAL CONCENTRATION
Tris-HCl	1M	2.5 ml	100 mM
SDS	20% (w/v)	5 ml	4% (w/v)
Bromophenol blue		Trace	0.2%
Glycerol	50%	10 ml	20% (v/v)
Dithioltreitol	1M	1.25 ml	50 mM
Sterile deionised water		Up to 25 ml	

---

**Table 19: Composition of the 5 x running buffer.**

REAGENTS	AMOUNTS	FINAL CONCENTRATION
Tris-base	15.1 g	0.12 M
Glycine	94 g	1.25 M
20% (w/v) SDS solution	25 ml	0.5%
Sterile deionised water	Up to 1000 ml	

---

The gel was analysed by staining in Coomassie Brilliant Blue (Table 20) and destained (Table 21).

**Table 20: Composition of the Coomassie Blue staining solution.**

REAGENTS	AMOUNTS
Methanol	400 ml
Glacial acetic acid	100 ml
Coomassie	1 g
Sterile deionised water	500 ml

**Table 21: Composition of the destain solution.**

REAGENTS	AMOUNTS
Methanol	100 ml
Glacial acetic acid	200 ml
Sterile deionised water	800 ml

## **2.6 General peptide synthesis**

### **2.6.1 *Synthesis of a general cyclic peptide using solid-phase synthesis***

Firstly, the resin (0.5 mmol) was swelled in dimethylformamide (DMF) and the solvent removed *in vacuo*. The resin was deprotected using 20% piperidine (v/v) in DMF (10 ml) and washed with DMF, dichloromethane (DCM) and diethyl ether; a Kaiser test confirmed successful deprotection (Section 2.6.2). A solution of the fluorenylmethyloxycarbonyl (Fmoc)-protected amino acid (1.5 mmol) in DMF (3 ml) was mixed with hydroxybenzotriazole (HOBr) (203 mg, 3 eq, 1.5 mmol) and diisopropylcarbodiimide (DIC) (232 µl, 3 eq, 1.5 mmol) for 15 minutes, producing a white solid, and then added to the deprotected resin and agitated with bubbling argon for 2 hours. The resin was then washed with DMF, DCM and diethyl ether and the coupling was verified using the Kaiser test (Section 2.6.2). The deprotection and coupling steps were repeated for each amino acid, producing the protected linear peptide on the resin. The Fmoc-deprotected linear peptide was cleaved from the resin

and R-group deprotection achieved using a trifluoroacetic acid (TFA) cleavage mix with scavengers and stirred under argon for 3 hours. The TFA was removed *in vacuo* and the sample was washed with DCM and methanol. The residue was precipitated in diethyl ether, purified by reverse-phase High performance liquid chromatography (RP-HPLC) (Table 22), lyophilised and analysed by mass spectrometry (MS), nuclear magnetic resonance (NMR) spectroscopy, infrared spectroscopy (IR), melting point analysis and analytical HPLC.

**Table 22: HPLC program for the general purification of linear and cyclic peptides.** Flow rate: 17 ml/min.

TIME	WATER (%)	ACETONITRILE (%)
	95	5
1 minute	95	5
9 minutes	95	5
10 minutes	40	60
10 minutes 50 seconds	95	5
15 minutes	95	5
20 minutes	95	5
20 minutes 5 seconds	95	5

Linear peptides were cyclised by stirring in a solution of HOBr (6 eq), 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) (3 eq) and DMF (1 mg/ml) for 24 hours. The solvent removed *in vacuo* and the residue was purified by ether precipitation and RP-HPLC (Table 22), solvents removed by lyophilisation, and analysed by MS, melting point analysis, NMR, infrared (IR) and analytical HPLC.

### **2.6.2 Kaiser Test**

To the beads, solution A (0.4 ml of 0.001 M potassium cyanide (aq) in 20 ml pyridine) (100 µl) and solution B (0.5 g ninhydrin in 10 ml ethanol) (25 µl) were added and heated for 2-5 minutes at 145°C. If deprotected amine was present the beads changed colour from yellow to blue. If protected amine was present the beads and solution stayed yellow.

### 3 Investigating PspA oligomerisation as a potential antimicrobial target

The initial aim of this chapter was to identify cyclic peptide inhibitors of *Burkholderia pseudomallei* PspA oligomerisation using a combination of a bacterial RTHS and inhibitor compounds generated by a cyclic peptide technology. The activity of the inhibitors identified was investigated in a selection of *in vitro* assays.

#### 3.1 Introduction: The Phage shock protein response

Bacterial stress responses are a prospective target for the development of antimicrobials. These systems provide bacteria with mechanisms to survive fluctuations to their environmental conditions. Unless bacteria adapt, these changes can impair growth and viability. Stresses include changes in pH, the presence of detergent, infection and the host's antibacterial armament. These changes can alter the cell membrane integrity and specifically stimulate extracytoplasmic stress responses.<sup>136</sup>

There are four main adaptive systems that function to maintain the cell membrane integrity: the sigma factor ( $\sigma^E$ ) system,<sup>137,138</sup> two signal transduction pathways<sup>139-141</sup> and the Phage shock protein (Psp) response.<sup>142</sup> The  $\sigma^E$  and Psp pathways are regulated by sequestration and release of transcription factors in response to signals: these associate with the RNA polymerase to permit  $\sigma^E$ -mediated transcription.<sup>143,144</sup> The  $\sigma^E$  is predicted to be the major stress regulator affecting at least 69 transcription units containing 114 genes with various functions, including the folding and degradation of polypeptides in the periplasm and  $\sigma^{32}$  involved in the heat-shock response.<sup>138,145-147</sup> Bae (bacterial adaptive response) and Cpx (conjugative pilus expression) two-component regulatory systems are two examples of signal transduction pathways involved in stress survival.<sup>139</sup> Both systems contain a sensor kinase (BaeS and CpxA, respectively) that is autophosphorylated after induction and activates the response regulator (BaeR and CpxR, respectively). This, in turn, initiates the expression of hundreds of proteins involved in protein folding and degradation.<sup>136,147</sup>

The bacterial stress responses described above have all demonstrated a stress-dependent synergy. For example, in a *rpoE* mutant of *S. Typhimurium* the Psp response can compensate for the absence of  $\sigma^E$  during stationary phase, a known inducer of the Psp response (described in Section 3.1.3).<sup>148,149</sup> However, typically there is little overlap observed between the above systems: each response has different cues and fulfils a distinct physiological role. Stresses that induce the Psp response (i.e. secretin mislocalisation or overexpression) have not been observed inducing the  $\sigma^E$  or Cpx responses and conversely inducers of  $\sigma^E$  and Cpx (i.e. oxidative stress and overexpression of P Pilus subunits) do not induce the Psp response.<sup>150</sup> The responses appear to have a specialised function during stress; consequently the function of all pathways in tandem may be required to elicit a full response under extracytoplasmic stress. Therefore, inhibiting one of the responses may prevent pathogenic bacteria from surviving when exposed to extracytoplasmic stresses, specific to that response, indicating that they could be viable targets for antimicrobial drug discovery.<sup>147</sup>

Within a variety of Gram-negative pathogens the Psp response has been shown to be essential for virulence, as many of the inducing stresses observed *in vitro* mimic those encountered during infection. Using signature-tagged transposon mutagenesis, Darwin *et al.* (1999) demonstrated that in mouse models of infection, a *Yersinia enterocolitica* *pspC* null mutant showed equivalent attenuation to a virulence-plasmid deficient strain.<sup>151</sup> The Psp response counters the extracytoplasmic stresses caused during secretion (the induction of the Psp response and its links with secretion are discussed in more details in Section 3.1.3). T3SS encode a needle-like structure that is involved in cell-cell contact. It permits direct translocation of effectors from the bacterial cytoplasm to the host cell cytoplasm, resulting in changes to the cell membrane (secretion systems are described in more detail in Section 3.1.3). The Psp response counters membrane damage caused during the secretion of effectors (i.e. virulence factors), linking the response to virulence. In a similar manner, PspA expression is induced when the twin-arginine translocation (Tat) export system, responsible for secretion across lipid membranes, becomes saturated.<sup>152</sup> Similarly, mutations in Tat export components that compromise protein export lead to elevated PspA expression. Due to the Psp responses' key role in

virulence for many pathogenic bacteria it provides a promising target for the development of new antimicrobials.

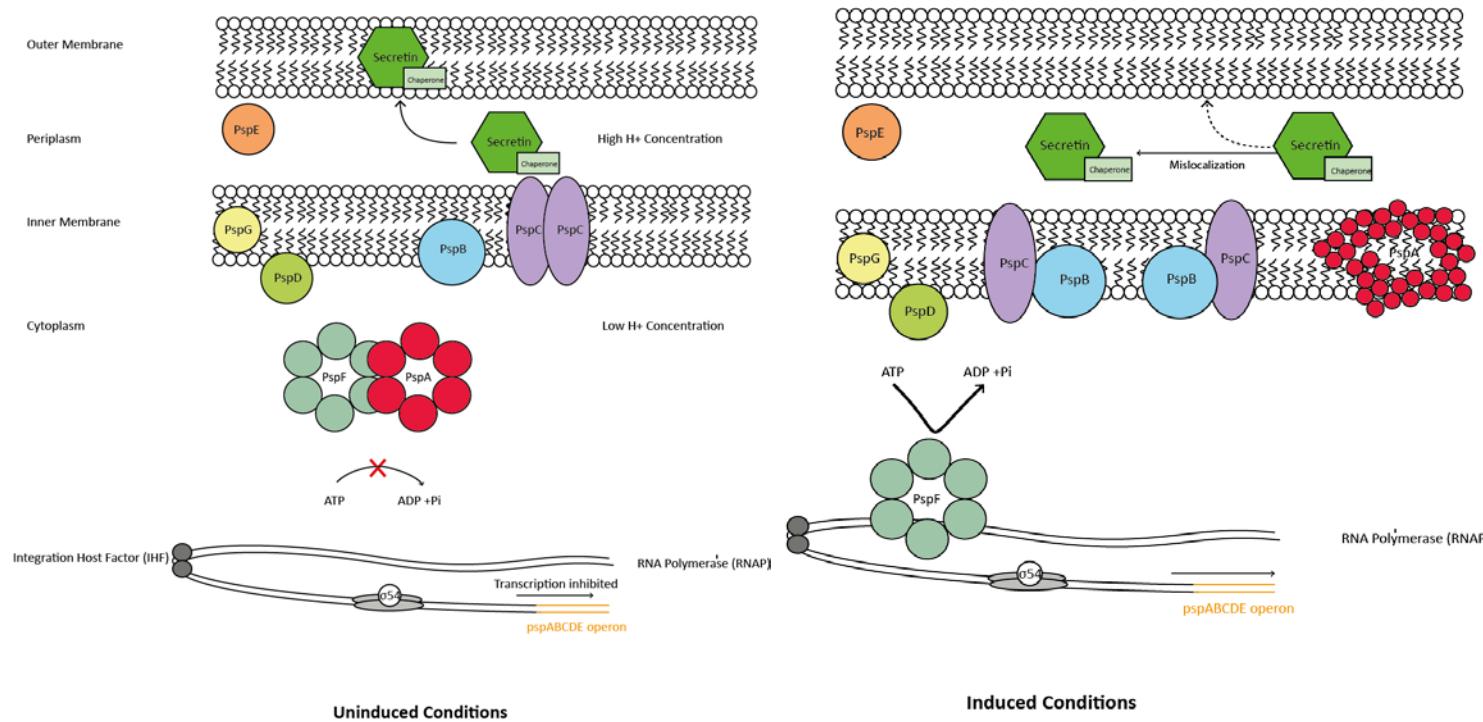
Nevertheless, a cautionary note should be added as under certain stresses multiple stress responses are activated. For example, the maturation of biofilms in *E. coli* an induction of other extracytoplasmic stress responses is observed.<sup>153</sup> After *cpxP*, *pspA* is the second most strongly overexpressed gene during biofilm formation, with an 8.42-fold increase in expression observed. Despite this the expression of the *psp* operon was not essential for biofilm formation, however, a mutant defective for expression of PspF affected biofilm formation; this could be due to an unknown overlap between the stress responses under certain stresses.<sup>153</sup>

### **3.1.1 Identification of the Psp response**

The Psp response maintains bacterial membrane integrity during extracytoplasmic stress.<sup>142</sup> Brisette *et al.* (1990) initially discovered the response in *E. coli* during filamentous phage infection. Cells were transformed with a plasmid encoding the phage gene IV (encoding pIV) and the expression of a 25 kDa protein was observed, that was absent in uninfected cells. This protein was subsequently named the Phage shock protein A (PspA).<sup>142</sup> pIV is an integral membrane protein involved in phage assembly by forming an outer membrane component of the exit channel for mature phage particle exportation.<sup>154,155</sup> It is this multimeric channel that is believed to have induced the Psp response by destabilising the membrane.<sup>156</sup>

The Psp system in *E. coli* contains two divergent transcriptional units. The first is regulated by an upstream  $\sigma^{54}$ -promoter and contains five open-reading frames (referred to, hereafter, as the *psp* operon) encoding: PspA, PspB, PspC, PspD and PspE proteins. The second transcriptional unit contains *pspF*, encoding the bacterial enhancer-binding protein (bEBP) PspF, a protein that acts as a transcriptional activator of the *psp* operon.<sup>143</sup> Later, a third gene was identified that is physically unlinked to the *psp* operon and *pspF*, but is upregulated in tandem with the *psp* operon genes by  $\sigma^{54}$ , PspF, the integration host factor (IHF) and PspA (this mechanism will be discussed later in Section 3.1.2): this gene was subsequently named *pspG*.<sup>157</sup>

Figure 10 depicts a working model of how these proteins interact to maintain the cell membrane integrity during stress. The function of each protein is not fully characterised, however, a general mechanism has been widely accepted.<sup>142</sup> Simply, in a cell where the membrane integrity is intact PspA interacts with PspF inhibiting PspF from activating the transcription of the *psp* operon and *pspG* gene; in this way PspA self-regulates its own transcription.<sup>158</sup> In contrast, under inducing conditions (i.e. mislocalisation of secretin), a stress is detected by a putative mechanism involving the membrane-bound proteins PspB and PspC.<sup>143</sup> However, the method by which the stress is detected depends on the bacterial species and the stress. The PspA-PspF complex dissociates permitting PspF to activate the transcription of the *psp* operon and *pspG*.<sup>144</sup> This elevates the levels of PspA and allows it to fulfil its physiological role in maintaining the cell membrane integrity.



**Figure 10: Working model of the Psp response in *E. coli*.** In a non-induced state PspA is bound to PspF, as a dodecameric complex blocking the transcription of the *psp* operon. When PspB and PspC detect a stress, for example secretin mislocalisation, PspA releases PspF. PspF isomerises the  $\sigma^{54}$ -factor, allowing it to confer polymerase specificity, and in turn, allows the transcription of the *psp* operon. PspA consequently accumulates at the inner membrane, restoring the cell membrane integrity; however, the precise method by which this occurs is not fully understood. Furthermore the specific roles of PspD, PspE and PspG have yet to be fully elucidated.

The Psp system is highly conserved, making it an attractive target for potential broad-spectrum antibiotics. The proteins PspA and PspF are highly conserved, whereas the other five Psp proteins (i.e. PspB, PspC, PspD, PspE and PspG) are less conserved.<sup>159</sup> For example, the Psp system of *E. coli*, *S. flexneri* and *S. Typhimurium* consists of all seven proteins,<sup>142</sup> whereas, *V. cholerae* and *Desulfovibrio vulgaris* only contain PspA, PspB, PspC and PspF.<sup>144</sup> Within *Y. enterocolitica* the operon encodes PspA, PspB, PspC, PspD and PspF, with 78%, 76%, 52%, 37% and 75% sequence homology to *E. coli*, respectively.<sup>160</sup> *Streptomyces lividans* (a Gram-positive bacteria) has a further reduced *psp* operon containing only PspA.<sup>161</sup>

In addition to the complete and reduced systems evident in prokaryotes, PspA homologues have also been identified in archaea and cyanobacteria. *Haloferax volcanii* is an example of a halophilic archaea with a homologous response that is induced by high-salt conditions suggesting that the system is important for survival.<sup>162</sup> Similarly, a homologue within cyanobacteria and plants has also been identified. The Vesicle-inducing protein in plastids 1 (Vipp1) is essential in plant thylakoid membrane formation (within both chloroplasts and cyanobacteria) and photosynthesis.<sup>163</sup> *S. sp. PCC 6803* Vipp1 and *E. coli* PspA share 27% sequence identity and a sequence similarity of 51%. Vipp1 and PspA are functionally analogous, with Vipp1 able to complement the export defect conferred by a *pspA* null mutant strain during protein export through the TAT export machinery;<sup>152</sup> the reverse, however, was not possible within cyanobacteria that contain both genes.<sup>164</sup> The retention of PspA and its homologues, even in the reduced system, implies that it has an important role in maintaining the membrane integrity. This indicates that the Psp response could be a viable antimicrobial target that has broad-spectrum activity like inhibitors of the secretion systems.<sup>159</sup>

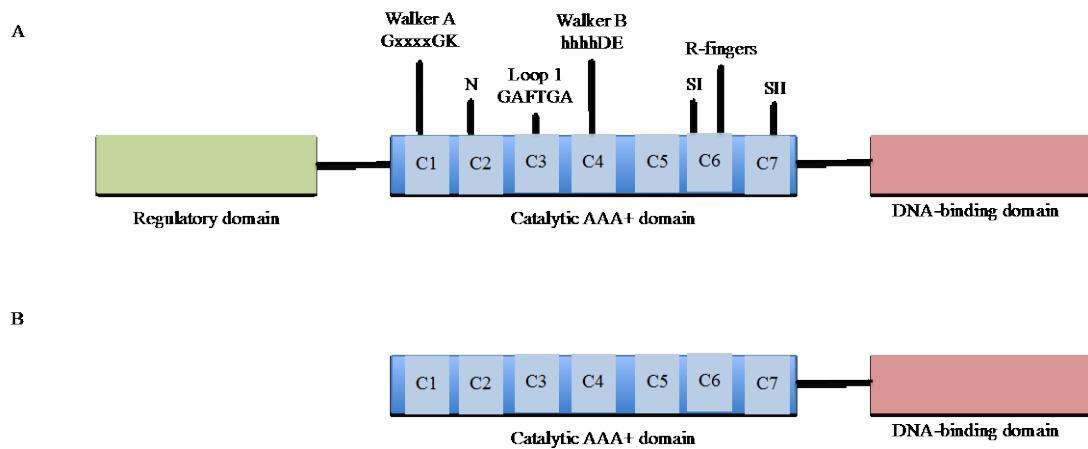
### **3.1.2 Regulation of the Phage shock protein response**

As previously described the Psp response is a regulatory system that controls the transcription of the *psp* operon. In *E. coli*, transcription of genes involves the use of an RNA polymerase containing six subunits. These include the  $\alpha$ ,  $\beta$ ,  $\beta'$ , and  $\omega$  subunits that form the catalytic core enzyme and a dissociable sigma ( $\sigma$ ) subunit that confers recognition specificity. The most commonly used sigma factor is the primary  $\sigma$ -factor ( $\sigma^{70}$ ) that recognises the -35 (TTGACA) and -10 (TATAAT) consensus

sequences. In contrast, in the Psp system a  $\sigma^{54}$ -dependent promoter (encoded for by *rpoN* and located upstream of *pspA*) controls the expression of the *psp* operon.

Under a range of extracytoplasmic stresses *rpoN* null mutants abolish the expression of the *psp* operon.<sup>165</sup> Unlike  $\sigma^{70}$ , the  $\sigma^{54}$  recognises -24 (GG) and -12 (GC) consensus sequences.<sup>146,165-167</sup> The  $\sigma^{54}$ -dependent promoter sites typically have a cytosine in the second position of the -12 sequence (96%). The *pspA* promoter from *E. coli* and *Y. enterocoililica* are two of only four examples without a cytosine at this position.<sup>160,167</sup> The  $\sigma^{54}$ -dependent promoters require specific activators called bEBPs, e.g. PspF.

Bacterial EBPs are ATPases Associated with diverse cellular Activities (AAA+) proteins: these proteins utilise adenosine triphosphate (ATP) hydrolysis to isomerise the transcriptionally inactive  $\sigma^{54}$ -promoter DNA complex into a transcriptionally active open complex.<sup>168,169</sup> Classically, bEBPs contain three domains: an N-terminal regulatory domain; a central AAA+ domain containing the ATP activity; and a C-terminal DNA-binding domain. The central AAA+ domain of PspF is a conserved region of about 200 amino acids (Figure 11), with seven conserved motifs called C1-C7,<sup>170</sup> including a Walker A (residues 37-45), also known as the P-loop or phosphate-binding loop and Walker B (residues 103-108). Both are implicated in ATP binding and hydrolysis. The nucleotide-binding site forms at the interface between two adjacent subunits, with an arginine-finger contributing to hydrolysis.<sup>171</sup> While recognition and remodelling of the closed promoter complex and the  $\sigma^{54}$  requires two motifs in the AAA+ domain called Loop 1 (residues 79-94) and Loop 2 (131-139). It is the GAFTGA motif in Loop 1 that directly contacts the  $\sigma^{54}$  in the closed-promoter state, with the alanine substitution of the threonine residue resulting in the loss of binding to the  $\sigma^{54}$ .<sup>172</sup> This binding brings the  $\gamma$ -phosphate of the ATP into close proximity with N64, a residue critical for activity.<sup>173</sup>



**Figure 11: Domain organisation of *E. coli* PspF compared to a typical bEBP.**

Where there is an additional regulatory domain. Both share a common AAA+ domain with a Loop 1, Walker A (where x is any amino acid) and B domain (where h is hydrophobic amino acid), arginine fingers and Sensors I and II. Figure adapted from Joly *et al.* (2009).<sup>174</sup>

There are, however, examples of truncated bEBPs containing no N-terminal regulatory domain. These utilise *trans*-acting regulatory proteins (such as HrpS and HrpR from *Pseudomonas syringae*).<sup>175</sup> PspF is an example of a truncated bEBP that lacks an N-terminal regulatory domain. In the Psp response, PspA functions as the *trans*-acting regulatory domain, in turn, regulating its own transcription.<sup>165</sup> AAA+ characteristically form distinct hexamer structures.<sup>176</sup> These hexamers have been observed in the Psp response, whereby PspF forms a hexameric species that interacts with hexameric PspA forming a dodecameric inhibitory complex identified by size exclusion chromatography (SEC).<sup>177,178</sup> PspA was shown to bind to the ATP binding site on the AAA+ domain of PspF.<sup>158</sup> This is located at the interface between two adjacent PspF units, involving a critical tryptophan residue (W56), revealed by alanine scanning.<sup>158</sup> The inhibitory mechanism of PspA could therefore be either by blocking ATP binding or by restricting PspF rearrangement into the correct conformation or oligomeric form.<sup>158</sup>

Upon stress induction, the PspA-PspF inhibitory complex dissociates and the DNA-binding domain of PspF binds to UASI/II activating  $\sigma^{54}$ -dependent transcription from these promoters; they are located between -80 and -126 relative to the *pspA* promoter

within *E. coli*.<sup>178</sup> In addition, the IHF improves binding specificity by allowing DNA looping, improving the binding of PspF to the UAS sites, as well as eliminating cross-activation with heterologous  $\sigma^{54}$ -RNAP activators.<sup>179</sup> The effect of the IHF on specificity and expression is dependent on the method of induction. When the IHF was mutated, rendering it inactive, a reduction in expression of the *psp* operon in response to ethanol and heat-shock was observed.<sup>180</sup> In contrast, when the system was induced by osmotic shock, no difference in expression levels was observed between the wild-type and the mutant.<sup>180</sup> This is just one example whereby there are variations in the Psp response depending on the species or stress. Understanding how the Psp response is induced is critical to identifying how it can be targeted for antimicrobial target development, however, the understanding of the mechanisms is currently limited.

### **3.1.3 Induction of the Psp response**

When the response was first discovered during filamentous phage infection, heat-shock proteins (HSPs) were also induced.<sup>142</sup> Therefore, *E. coli* PspA was originally thought to be a HSP. However, unlike HSPs that are strongly induced by a temperature shift from 37°C to 42°C, activation of the Psp response required a higher temperature shift from 37°C to 50°C to show significant induction. While expression of *pspA* was induced at lower temperature shifts, the levels of protein were only one-tenth of those observed after induction at 50°C. A similar result was observed with ethanol induction: higher concentrations of ethanol (i.e. 10% volume/volume) were required to induce *pspA* expression compared to the expression of classical HSPs (i.e. 4% volume/volume). This information, in conjunction with evidence that induction of the *psp* operon does not require the heat-shock sigma factor ( $\sigma^{32}$ ) suggests that PspA and other Psp proteins are not typical HSPs.<sup>142</sup>

Other conditions have been evaluated for their ability to induce the Psp response. All inducers are stresses that jeopardise the cell membrane integrity. For example water-immiscible hydrophobic solvents (i.e. hexane and cyclooctane) induce *psp* operon expression because they intercalate into biological membranes, resulting in a weakening of the lipid interactions, disrupting the membrane integrity.<sup>181</sup> A similar effect is observed when fatty acid biosynthesis is disrupted, using mutagenesis or the addition of diazaborine and cerulenin.<sup>182</sup> These impair phospholipid biosynthesis that

may interfere with protein translocation or affect the cell membrane. In addition, hyperosmotic shock produced by the addition of either 0.75 M NaCl or 0.6 M sucrose also stimulates the Psp response; the addition of these species causes an intracellular accumulation of potassium, which induces the expression of several genes, as well as the *psp* operon, to restore turgor pressure.<sup>142,183</sup> Vrancken *et al.* (2008) described the effect of each stress inducer outlined above on bacterial growth rates of *S. lividan* *pspA* null mutants and the expression of the *pspA* promoter by linking it to a promoterless eGFP gene and monitoring the fluorescence. PspA expression was clearly upregulated under all these stress conditions.<sup>161</sup> The inducing conditions stated above mimic those encountered when bacteria invade a host. The host's innate immune mechanisms assault the bacteria with changes in temperature, osmolarity, and pH. These jeopardise the cell membrane integrity by altering the integral cell membrane properties, causing misfolding or mislocalisation of cell membrane proteins and disruption of the membrane permeability barrier. In order to maximise the chance of survival when presented with these changing conditions, bacteria must respond quickly and modify gene expression to optimise survival.

As well as the physical changes on the cell membrane upon host infection, pathogenic bacteria must release virulence factors into the host. This is achieved using secretion systems that can impact the cell membrane integrity and induce the Psp response, as discussed in Section 3.1. Much of the research into the induction of the Psp response has focused on secretion systems. Secretion systems are formed by an organisation of secretins, which are a family of oligomeric pore-forming proteins. The secretins are typically located in the outer membrane and are involved in several important processes including the export of virulence proteins, crucial for bacterial pathogenesis, from the cytoplasm to the cell exterior.<sup>184-186</sup> There are seven main secretion systems (Type 1-7 Secretion Systems, T1-7SS) present in bacteria, as well as, other transporter systems; for example, the ATP binding cassette (ABC) type transporters<sup>187</sup> and the Tat export system.<sup>188</sup> The T3SS is commonly used by pathogenic bacteria to transport virulence proteins into eukaryotic cells.<sup>189</sup>

Microarray analysis has demonstrated how Psp proteins are expressed in response to secretin-mediated stress, with a 28-fold increase in expression observed for PspA.<sup>190</sup> The Psp response is initiated when the T3SS becomes blocked or transported

proteins are over-expressed or mislocalised. Three examples of how protein mislocalisation and overexpression of proteins can induce the Psp response are:

1. **Expression of mutant forms of the membrane protein, PhoE,** the secretion system becomes blocked.<sup>191,192</sup>
2. **Absence of YidC,** which is an inner membrane protein that transports proteins across the membrane, bypassing the secretion system, by aiding membrane protein insertion.<sup>193-195</sup>
3. **Absence of chaperone proteins,** for example, within *Y. enterocolitica* mislocalisation of the *Klebsiella oxytoca* pIV homologue, PulD in the absence of its chaperone PulS induces expression of the *psp* operon; however, this is abrogated when PulS is expressed.<sup>196,197</sup>

Currently it is unclear how secretin mislocalisation triggers the induction of the Psp response. During filamentous phage (pIV) secretion the periplasmic contents do not leak,<sup>198</sup> therefore the Psp response is not prompted by a leaky outer membrane. This is supported by evidence that filamentous phage pIII, that makes the membrane permeable, does not induce the Psp response.<sup>142</sup> Furthermore, in mutant strains whereby pIV cannot translocate across the membrane due to a defective signal sequence or failure to oligomerise, the Psp response fails to be induced.<sup>199</sup> In view of the observations described above, it is hypothesised that the physical process of inserting the secretin into the membrane stimulates induction of the Psp response.

In addition to the premise that physical disruption triggers the Psp response, some of the inducing conditions described (including secretion) have been linked with a dissipation of the proton motive force (PMF). PMF can be described as the storing of energy as a combination of a proton and electrochemical gradients across a membrane. It is integral in membrane potential, flagellar motor rotation, protein translocation across a membrane and the transmembrane pH gradient and is therefore susceptible to alterations in the membrane permeability. During PhoE mislocalisation there is a proton dissipation, which is restored by the Psp response inducing overexpression of PspA in *E. coli*.<sup>191,192</sup> Similarly, the mislocalisation of PulD induces a dissipation of PMF that is restored by the Psp response. This link has been exemplified further in *Y. enterocolitica* whereby in a *psp* null mutant a decrease in PMF was observed when mislocalisation of the secretin component of the *Ysc*-

Yop type III secretin system occurs.<sup>190</sup> Kobayashi *et al.* (2007) supported the premise that the Psp response was induced by a dissipation of PMF, by demonstrating how *in vitro* recombinantly expressed PspA binds to membrane lipids and blocks proton leakage from membrane vesicles.<sup>200</sup> The response of the Psp system to dissipation of PMF has also been investigated using proton ionophores (i.e. carbonyl cyanide m-chlorophenyl hydrazine); these are agents that uncouple the proton gradient that is established during normal activity of electron carriers in the electron transport chain. They can be used to study PMF dissipation by disrupting the electrochemical gradient across the inner membrane, causing an influx of protons across the membrane, in turn, inducing the Psp response.<sup>148</sup>

The link between the Psp response and PMF has been best exemplified under alkaline conditions at the stationary phase. A *pspABC* null mutant grows at approximately the same rate as the wild-type strain at pH 9.0 during the logarithmic phase. Conversely, a significant difference is observed between the mutant and wild-type strains during stationary phase growth at pH 9.0.<sup>148</sup> As the external pH increases the cell attempts to maintain the internal pH between 7.5-7.9, increasing the membrane potential. At pH 9.0 the pH gradient shifts out of the cell and opposes the PMF. The PMF must be maintained which costs the cell energy. The Psp response is induced, as energy is dissipated trying to reach the stationary phase.

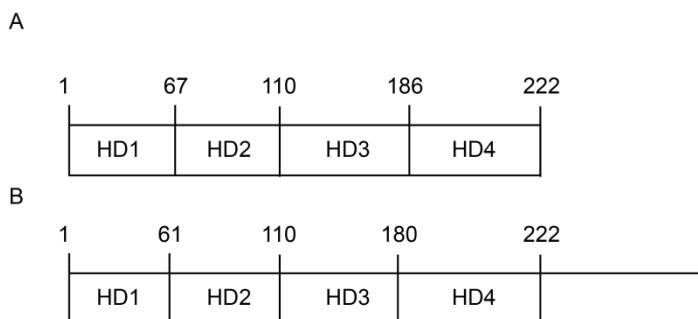
Increasing evidence suggests that mechanisms other than the dissipation of PMF or physical disruption of the membrane may also activate the Psp response. Cells are able to survive in the absence of PMF, after treatment with carbonyl cyanide m-chlorophenyl hydrazine, on glucose rich media<sup>201</sup> and some inducers of the Psp response do not cause PMF dissipation (i.e. YidC depletion).<sup>202</sup> Recently the two-component ArcAB system was implicated in the induction of the Psp response.<sup>203</sup> The ArcAB proteins are electron carriers in the electron transport chain; ArcB is a redox sensor that phosphorylates ArcA causing the repression of genes involved in aerobic respiration and upregulation of genes required in anaerobic respiration. The regulation of genes by ArcA and a direct interaction between ArcB and the Psp proteins is required to propagate the signal stress induced by pIV-secretin under microaerobiosis.<sup>204</sup> The effect of ArcB has been linked with PMF because when

there is a reduction in ubiquinone levels, there is an influx of protons into the cytosol dissipating PMF; this activates ArcB and the expression of the *psp* operon.<sup>203</sup>

Section 3.1.3 has described some of the research over the last decade demonstrating how the Psp response has many inducing signals and how further investigation is required to elucidate how the Psp response is activated.<sup>204</sup>

### 3.1.4 Phage shock protein A and its role within the Psp response

PspA is the most conserved protein in the Psp response.<sup>159</sup> Located within the cytoplasm and peripheral inner membrane,<sup>205</sup> *E. coli* PspA forms a helical coiled-coil structure with four  $\alpha$ -helical domains (e.g. HD1-HD4).<sup>177</sup> HD1-3 encourages dimer formation, while the absence of HD1, leaving HD2-3, produces a hexameric species that forms the inhibitory complex with PspF (Figure 12A). The presence of all four helical domains allows the formation of higher-order structures; it is these structures that permit PspA to fulfil its physiological role in maintaining cell membrane integrity.<sup>158,177</sup> Within *E. coli* the dominant species is hexameric PspA.<sup>206</sup>



**Figure 12: Alignment of the predicted secondary structure and domains. (A) *E. coli* PspA. (B) PspA homologue *Synechocystis* Vipp1.** Domains 1-4 are all required for oligomerisation, the absence of domain 4 in PspA limits the quaternary structure to dimers only, while the removal of domain 1 and 4 permits hexamer formation. Vipp1 contains an additional C-terminal domain not required for complex formation. Domains are assigned according to regions defined previously for *E. coli* PspA.<sup>158,163,207,208</sup>

PspA fulfils two roles: one role is to self-regulate its own transcription as well as the rest of the *psp* operon and the second is a physiological role to restore cell membrane

integrity after stress.<sup>158 200</sup> It has been discussed in Section 3.1.2 that PspA and PspF form a dodecameric inhibitory complex preventing PspF from activating the transcription of the *psp* operon.<sup>158</sup> Induction causes the inhibitory complex to dissociate causing overexpression of PspA. After induction PspA accumulates at the cell membrane as large oligomeric species; in this form PspA is able to bind to the membrane and suppress proton leakage of stress-damaged membranes. This has been demonstrated *in vitro* where oligomeric PspA bound to ethanol-damaged liposomes and membrane vesicles containing phosphatidylserine and phosphatidylglycerol suppresses proton leakage.<sup>200</sup>

Using negative staining techniques, Hankamer *et al.* (2004) visualised these higher-order species, from *E. coli* by electron microscopy and single particle analysis. The species were characterised as ‘doughnut’-shaped, with a diameter of 8.5 nm, nine-fold rotational symmetry and each unit containing thirty-six individual PspA subunits.<sup>209</sup> Similar ring-like higher-order structures have been observed with the PspA homologue *Arabidopsis thaliana* Vipp1; with a 22 nm diameter, 12-17-fold rotational symmetry, consisting of 48-68 subunits in total. The formation of the higher-order structures of Vipp1 is dependent on the α-helical PspA-like domain.<sup>163,207,208</sup>

In addition to the ‘doughnut’-shaped species, Vipp1 has been shown to form higher-order super-complexes, in the form of rods and microtubules.<sup>210,211</sup> The function of these higher-order rod-like structures has yet to be determined. The rod-shaped oligomers of Vipp1 were purified in the absence of detergent, whereas the discrete rings observed by Hankamer *et al.* (2004) were purified in the presence of 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS).<sup>209-211</sup> Although *E. coli* PspA has not been shown to form rod-like structures, larger clathrin-like scaffolds have been observed when PspA was purified in the absence of the detergent, CHAPS. Standar *et al.* (2008) proposed that the structural differences between PspA species previously observed could be due to the presence of CHAPS detergent during the purification.<sup>212</sup> It is hypothesised that the scaffolds have a function in restoring cell membrane integrity; supposedly achieved by stabilising any leaks or ‘holes’ caused by stress.<sup>212</sup>

The subcellular localisation of PspA has been investigated, and again shows that PspA forms a variety of oligomeric species.<sup>216</sup> Prior to stress induction PspA is mobile in the cytoplasm.<sup>205,213,214</sup> However, after stress induction (i.e. with YscC and YsaC secretins) PspA, PspB and PspC localised predominately at the poles in highly focused clusters independent of PspF activity. In contrast, PspF remained evenly distributed throughout the cytoplasm.<sup>205,213,214</sup> The localisation of PspA was independent of the method of induction.<sup>213</sup> Calculation of the fluorescence of these accumulations of GFP-PspA, indicated that different higher-order structures were forming, with 59% showing a two- to three-fold increase in intensity above background GFP and a further 20% showing more than six-fold increase. The latter could be the ‘doughnut’-shaped species and larger clathrin-like species.<sup>205,209,212</sup>

Currently, only *E. coli* PspA has been visualised by electron microscopy, and there is much debate surrounding the form PspA takes *in vivo* to fulfil its physiological role. The literature indicates that only in an oligomeric state can it sustain cell membrane integrity. Additionally, PspA is highly conserved amongst bacteria, intimating that it has an important role generally in maintaining cell viability. Therefore, inhibiting the oligomerisation of PspA could be a suitable PPI target for antimicrobial development.

### **3.1.5 Other proteins involved in the response**

There are other PPIs involved in the Psp response, beyond PspA-PspA and PspA-PspF. However, the role of the other protein interactions are still being investigated, and it is possible other PPIs have yet to be determined. In order to initiate the Psp response membrane stress must be detected; this is stress inducer-dependent. For example, in *E. coli* heat-shock, a known inducer of the Psp response does not require any membrane-located receptors, whereas osmotic shock, phage infection and ethanol treatment are detected by the membrane-bound proteins, PspB and PspC.<sup>165</sup> Detection is also species-dependent as *pspB* and *pspC* are not conserved; exemplified by *S. lividans* that contains the reduced system with only *pspA*.<sup>161</sup>

PspB is an inner membrane protein with a hydrophobic N-terminal domain and a C-terminal cytoplasmic domain with coiled-coil properties. In contrast, PspC has an N-

terminal cytoplasmic domain, a single transmembrane segment and a periplasmic C-terminal with a leucine zipper motif.<sup>143,192</sup> Random mutagenesis undertaken on both PspB and PspC demonstrated that mutations in the cytoplasmic and transmembrane domain of PspC upregulated expression of  $\phi pspA-lacZ$ , indicating a negative regulatory role, while mutations in the periplasmic domain suggested a positive regulatory role. In contrast, mutations in the cytoplasmic domain of PspB, elicited a null mutant phenotype.<sup>215</sup> PspB and PspC play an important role in regulating the Psp response in both *E. coli* and *Y. enterocolitica*.<sup>216</sup>

Other genes present on the *psp* operon are PspD, PspE and PspG. PspE is localised in the periplasm and is transcribed from both the *psp* operon and its own  $\sigma^{70}$ -dependent promoter.<sup>143,217</sup> During infection of eukaryotic cells with *S. Typhimurium* its expression was highly induced, intimating that PspE plays a critical role during infection.<sup>218</sup> PspE is a rhodanese-related enzyme that acts as a detoxification enzyme for cyanide and heavy metals, contributing 85% of the total rhodanese activity.<sup>219</sup> Rhodanese activity has been linked with the Psp response within the cell. It is hypothesised that the cell membrane integrity is threatened by cyanide presence and so the Psp response is activated to eliminate the damaging cyanide.<sup>217</sup> But it was identified as a non-essential component of the Psp response, as it is independently expressed but not by typical inducing conditions of the Psp response (i.e. osmotic and heat-shock).<sup>219</sup> The proposition that PspE is not an essential component of the Psp response is supported by its lack of conservation amongst species.<sup>144</sup>

Another protein involved in the Psp response of *E. coli* is PspD. Very little is known about PspD, as its expression has never been demonstrated *in vivo*. However, it is known to be a peripherally-bound inner-membrane protein that does not form any cross-linked complexes when studied by *in vivo* cross-linking.<sup>220</sup>

The final protein PspG was identified in *Y. enterocolitica* and *E. coli*; again, little is known of its role in the Psp response.<sup>157,221</sup> It is positively regulated by PspF and negatively regulated by PspA, from a promoter like *pspA* that does not contain the conserved cytosine in the second position of the consensus sequence.<sup>160</sup> Similar to PspA, PspG has been visualised by fluorescence microscopy to localise at the poles and the cell membrane after stress induction giving PspG a hypothetical role as an

effector of the Psp response. In addition, PspG self-associates as a hexamer, trimer and dimer, and interact with PspC, suggesting that stress signalling could be mediated by a PspC and PspG interaction.<sup>174</sup>

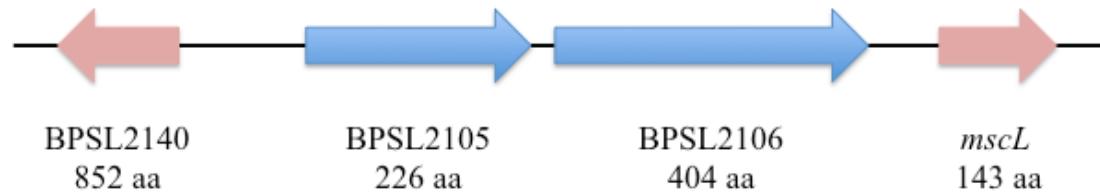
Over the last twenty years research efforts have focused on elucidating the role of the Psp response for virulence, how the response is induced and the role of each protein on the *psp* operon. It is apparent that the function of the proteins is stress and species dependent, with PspA being the most conserved protein, playing a key role in restoring cell membrane integrity. The same statements cannot be made about the other proteins in the Psp response, where further investigation is required to clarify their role and interactions.

### **3.1.6 The Psp response within *Burkholderia pseudomallei***

This research involved the identification of cyclic peptide inhibitors of PPIs involved in the Psp response of *B. pseudomallei*. *B. pseudomallei* is a Gram-negative motile bacillus and the causative agent of melioidosis.<sup>222</sup> Melioidosis is endemic to both Northern Australia, where it causes bacteraemic pneumonia<sup>223</sup> and Northern Thailand, where it is the leading cause of community-acquired septicemia.<sup>224</sup> Transmission of the disease can occur by inhalation, ingestion and inoculation; this leads to the development of different symptoms dependent on the mode of exposure. Typically, in Northern Thailand infection is via a cutaneous route, due to the prevalence of the bacteria in the soil of rice paddy fields. In addition, the number of cases increases dramatically during the monsoon season, causing an increase in transmission via inhalation.<sup>225</sup> This transmissibility of *B. pseudomallei* via the aerosol route has led to *B. pseudomallei* being classed as a category B agent by the US Centres for Disease Control and Prevention.

There are limited therapeutic options for the treatment of *B. pseudomallei*, due to the high rate of relapse and resistance to diverse groups of antibiotics. The recommended therapeutic regime is a prolonged and aggressive course of intravenous and oral antibiotics, because of inherent resistance.<sup>226</sup> Eradication of *B. pseudomallei* infection is a significant challenge, especially due to its resistance to antibiotics, latent infections and that the current available medicines are associated with side effects and poor compliance.

The aim of this project was to identify cyclic peptide inhibitors of the Psp response, which may be suitable as novel therapeutics for melioidosis. This project ran in parallel with a study by collaborators at DSTL to characterise the Psp response within *B. pseudomallei* K96243 and to establish it was essential, and if it was, under which conditions this phenotype was observed.<sup>227</sup> Initial bioinformatic studies identified an operon containing BPSL2105, a *pspA* homologue, and BPSL2106, a putative membrane protein (Figure 13). BPSL2105 is 226 amino acids long with no significant identity to *E. coli* PspA but 22% identity to *Y. enterocolitica*.<sup>227</sup> This information suggested that *B. pseudomallei* has a reduced Psp system, similar to *S. lividans*.<sup>161</sup>



**Figure 13: *B. pseudomallei* psp operon.** BPSL2105 represents the putative *B. pseudomallei* PspA homologue and BSL2106 is the putative membrane-bound protein. Based on a bioinformatic study by S. Southern (DSTL).<sup>227</sup>

A bioinformatic study of PspF was also carried out and 14 putative proteins were identified: all were described as  $\sigma^{54}$ -transcriptional response regulators, of a similar size to *E. coli* PspF (See Appendix 1 for a table describing PspF homologues in *B. pseudomallei* K96243, replicated with permission from S. Southern, DSTL).<sup>227</sup> All are of a similar size to reported homologues and have between 33-48% sequence identities with *Y. enterocolitica*. By studying the domains of each putative protein, in comparison to *Y. enterocolitica* PspF, BPSS2250 was selected for further study, as it has the AAA+ domain at its N-terminus and a C-terminal DNA-binding domain. The only caveat to the identification of BPSS2250 as PspF is it was not adjacent to the putative *pspA* in the chromosome, as observed for other species.

Within the Psp system there are several PPIs that could be targeted using the RTHS and SICLOPPS screening technology. For this project the oligomerisation of PspA

was chosen, as the *psp* operon within *B. pseudomallei* has not been fully characterised and appears to have a reduced system containing only PspA, PspF and a hypothetical third gene. Although PspF could be targeted, the identification of PspF was more tentative due to its location on the chromosome.<sup>227</sup> Still, by disrupting the oligomerisation of PspA the bacteria's ability to survive under extracytoplasmic stress is potentially inhibited. Hexameric PspA also represses expression of the *psp* operon by forming a dodecameric species with PspF, which could also be disrupted by any potential inhibitor identified during the screen. Nevertheless, by targeting PspA the inhibitor could have dual-functionality, the latter of which is undesired but would be counteracted later in the pathway. The other option is to target the oligomerisation of PspF, as the interface between two PspF sub-units is required for  $\sigma^{54}$  isomerisation. In a system where the *psp* operon has been fully characterised (i.e. *E. coli*) PspF might be a more suitable target, however, within *B. pseudomallei*, where the identification of PspF is speculative, the oligomerisation of PspA was thought to be a better target.



### 3.2 Results: Inhibitor screening of PspA oligomerisation

The Psp response is involved in maintaining cell membrane integrity during extracytoplasmic stress. The oligomerisation of PspA is crucial in the response, assembling at the cell membrane and restoring cell membrane integrity. This physiological role can only occur when PspA is in an oligomeric state. Therefore, an inhibitor of the homodimerisation of PspA may hinder the ability of *B. pseudomallei* to survive cellular stress, providing a suitable antimicrobial target. This part of the project uses a combination of a bacterial RTHS and SICLOPPS library screening to identify cyclic peptide inhibitors capable of disrupting the homodimerisation of PspA.

#### 3.2.1 *Construction of the PspA Reverse-two-hybrid system*

##### 3.2.1.1 *Construction of pTHCP16 PspA*

A homodimeric RTHS was constructed to screen for cyclic peptide inhibitors of PspA dimerisation. To construct the RTHS, the *pspA* gene from *B. pseudomallei* K96243 was cloned into pTHCP16- a plasmid that contains the DNA-binding domain of the wild-type 434-repressor, downstream of an IPTG-inducible pTAC promoter (Figure 14). The pTHCP16 vector backbone was constructed from pMAL-c2x (New England Biolabs),<sup>80</sup> which includes a multiple cloning site (MCS) containing the following restriction sites: *SalI*, *BamHI*, *NruI* and *SacI*.



**Figure 14:** Representation of plasmid pTHCP16 constructed by Horswill *et al.* (2004).<sup>80</sup> (A) 434-repressor is under the control of the IPTG inducible pTAC promoter with MCS (*SalI*, *BamHI*, *NruI* and *SacI*). (B) Design to allow expression of C-terminal PspA recombinant proteins with 434.

The target gene, in this case *pspA*, was cloned into the MCS to encode a PspA fused to the 434-DNA-binding domain, the expression and dimerisation of which would

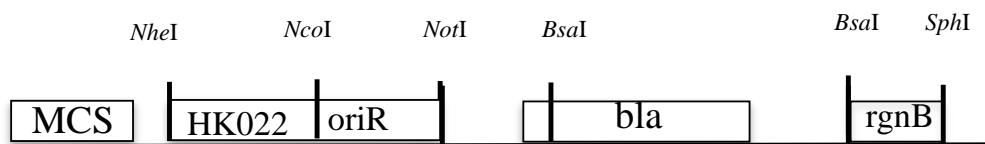
regulate downstream transcription of a reporter construct.<sup>80</sup> *pspA* was cloned into pTHCP16 via the *Bam*HI and *Sac*I restriction sites. *Nru*I was not used because the restriction endonuclease for *Nru*I is methylation sensitive, and the *Sal*II restriction site was deemed to be too close to *Bam*HI, reducing cloning efficiency. The *Sac*I restriction digest site present in *pspA*, was first removed by site-directed mutagenesis, which replaced the terminal cytosine of the *Sac*I restriction site with thymidine, creating a silent mutation. Subsequent incubation with the restriction enzyme *Dpn*I removed the original template DNA, as this endonuclease only recognises methylated DNA through the methylated sequence G<sup>M</sup>A/TC, and *E. coli* DNA is methylated.

The mutated *pspA* gene was amplified by PCR, which incorporated the restriction sites *Bam*HI and *Sac*I, as well as an overhang region of six base pairs (bp) designed to promote efficient restriction digestion. This fragment was then cloned into pTHCP16, enabling the expression of the 434-PspA recombinant proteins under the control of IPTG.<sup>80</sup> Successful ligation was confirmed by analysis of products by PCR, analytical DNA restriction digestions via gel electrophoresis, and ultimately by DNA sequencing.

### *3.2.1.2 Construction of pAH68 PspA and integration onto the chromosome*

For inhibitor screening the 434-*pspA* cassette must be integrated into the chromosome. This is because during screening the SICLOPPS plasmid pARCB is transformed into the RTHS; this plasmid has a p15a origin of replication and chloramphenicol resistance. Despite these features being compatible with pTHCP16, one plasmid must be integrated because recombinant plasmids are often unstable and are subsequently lost in the absence of continual selection pressures. Furthermore plasmid loss or rearrangement can occur if the product of an artificially expressed gene is detrimental to cell growth, which in this case would lead to false positives.<sup>228</sup> In addition, the use of an integrated plasmid eliminates stochastic variation in protein expression, as plasmid copy number can vary depending on specific growth conditions.<sup>229,230</sup>

Conditional-Replication, Integration, and Modular (CRIM) plasmids were utilised for integration.<sup>231</sup> This system exploits plasmids carrying a conditional-replication origin and a phage attachment (*attP*) site; facilitating site-specific integration at the bacterial attachment site (*attB*) in the presence of a phage integrase (Int). CRIM plasmids have a  $\gamma$  replication origin of R6K, which requires the *trans*-acting  $\Pi$  protein (encoded by *pir*) for replication. The plasmids also contain the bacterial (*rgnB*) and phage  $\lambda$  (t0, tL3) terminators flanking the cloning region to protect other segments from transcriptional read-through. Integration can be achieved by electroporation of the CRIM plasmid into the chromosome of the host strain carrying a helper plasmid. This is a thermosensitive replicon allowing synthesis of the phage integrase (Int) protein, under control of the temperature-sensitive repressor *cI857*. The host strain is cured of the helper plasmid by prolonged incubation at 42°C (Figure 15 depicts the pAH68 specific plasmid map).<sup>231,232</sup>



**Figure 15: Representation of plasmid pAH68 constructed by Haldimann *et al.* (2001).**<sup>231</sup> The plasmid contains the MCS, the phage integrase site (HK022), the origin of replication (*oriR*) and the ampicillin resistance gene (*bla*).<sup>231</sup>

The lambda integrase site (the most efficient) was not available as it had already been used for the incorporation of the reporter gene construct; in the case of the PspA RTHS, the second most efficient integrase site, HK022, was utilised.<sup>233,234</sup> Hence, the *434-pspA* cassette was cloned into pAH68- a CRIM plasmid containing the HK022 phage attachment site- using the restriction sites *MscI* and *SacI*. This allowed the inclusion of the *lacI* gene, providing controlled expression during screening.<sup>233</sup> The *lacI* gene was included in the pAH68 construct due to the lack of *lacI<sup>Q</sup>* in BW27786. Initially, Khlebnikov *et al.* (2001) reported that the BW27786 strain contained the *lacI<sup>Q</sup>* gene. This allowed mediation of background expression of the recombinant proteins from pAH68.<sup>235</sup> However, Baba *et al.* (2006) later discovered that the strain BW25113 from which BW27786 was derived was in fact *lacI<sup>+</sup>* not *lacI<sup>Q</sup>*.<sup>236</sup> This

phenotype infers decreased production of *lacI* compared to the *lacI<sup>Q</sup>* through the absence of a mutation of its –35 promoter. This can result in a decrease in *lacI* production and results in basal expression of the recombinant protein.<sup>237</sup> Successful ligation was confirmed by PCR, analytical DNA restriction digestions analysed via gel electrophoresis and finally DNA sequencing.

The pAH68 PspA plasmid was integrated into the chromosome of BW27786; the strain used for genetic selections.<sup>235</sup> This strain is derived from *E. coli* K12 where residues 165–355 of the *hisB* gene corresponding to the imidazole glycerol-phosphate dehydratase activity were previously deleted.<sup>238</sup> The reporter-construct (*his3-Kan<sup>R</sup>-lacZ* operon) and 434-repressor were also integrated into the chromosome.<sup>229,231</sup> Strain BW27786Δ*hisB* containing the homodimeric reporter will be referred to as SNS118.<sup>80</sup>

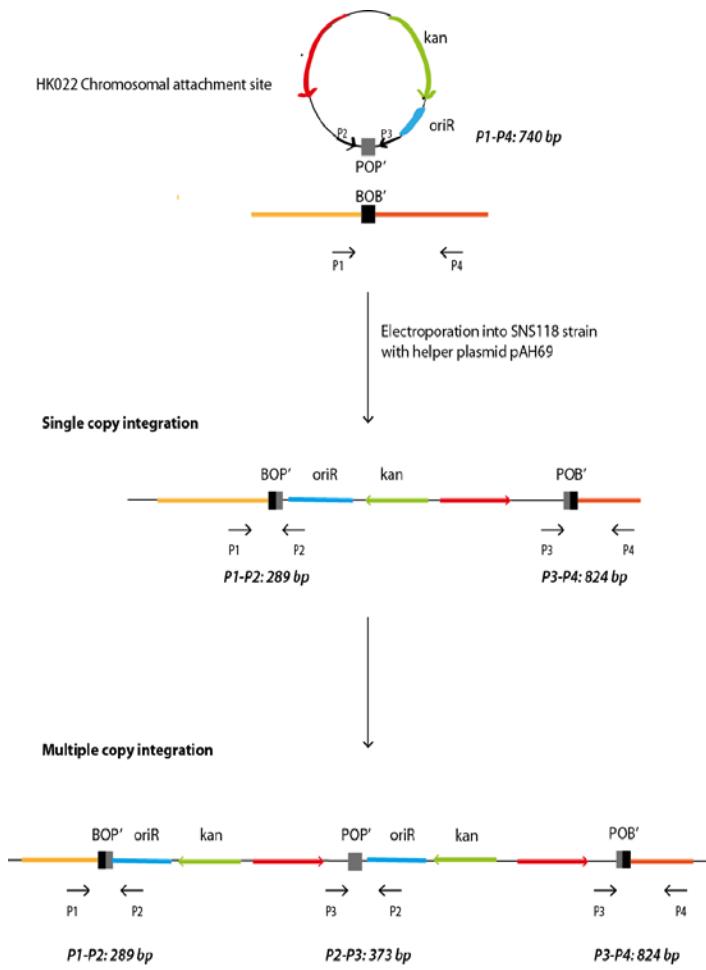
The integration of the CRIM plasmid was carried out by transforming pAH68 PspA into SNS118 containing the helper plasmid, pAH69. This step required optimisation of the standard protocol described in Haldimann *et al.* (2001).<sup>231</sup> Table 23 outlines all the strategies employed to achieve integration; only condition 12 resulted in integration. The key difference for condition 12 was the method by which the competent cells were prepared. It employed a temperature shift from 30°C to 42°C. It is likely that this step aided integration, via expression of the heat sensitive Int protein prior to transformation, providing higher background levels.

**Table 23: Table containing all the conditions used to optimise integration of the pAH68 PspA into the chromosome of the SNS118 bacterial strain.** Successful conditions are highlighted in bold

CONDITIONS	PREPARATION OF COMPETENT CELLS	SOC TEMP	INITIAL TEMP. AND TIME	42°C WATER BATH	FINAL TEMP. AND TIME
1 <sup>231</sup>		On ice	30°C for 60 min	No	42°C for 30 min
2		On ice	30°C for 60 min	Yes for 10 min	42°C for 30 min
3		On ice	30°C for 45 min	No	42°C for 45 min
4	Electro-competent cells: Culture grown at 30°C to OD <sub>600</sub> of 0.6 then made competent by standard procedure in 10% glycerol	37°C	30°C for 60 min	No	42°C for 30 min
5		On ice	30°C for 60 min	Yes for 10 min	42°C for 60 min
6		On ice	30°C for 10 min 37°C for 50 min	No	42°C for 30 min
7		On ice	37°C for 60 min	No	42°C for 30 min
8		On ice	30°C for 60 min	No	42°C for 30 min
9	<b>Electro-competent cells: Culture grown at 30°C to OD<sub>600</sub> of 0.4 then incubated at 42°C till OD<sub>600</sub> of 0.6 then made competent by standard procedure in 10% glycerol</b>	On ice	30°C for 60 min	Yes for 10 min	42°C for 30 min
10		On ice	30°C for 45 min	No	42°C for 45 min
11		On ice	30°C for 10 min 37°C for 50 min	No	42°C for 30 min
12		On ice	<b>37°C for 60 min</b>	No	<b>42°C for 30 min</b>
13	Chemically-competent cells: Culture grown at 30°C to OD <sub>600</sub> of 0.4 then incubated at 42°C till OD <sub>600</sub> of 0.6 then competent by standard procedure in 10% glycerol	On ice	30°C for 60 min	No	42°C for 30 min
14		On ice	37°C for 60 min	No	42°C for 30 min

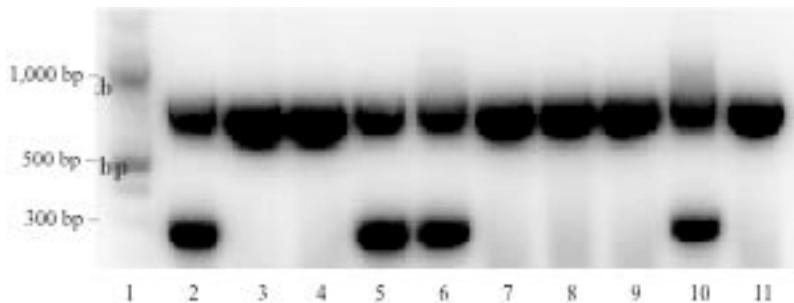
### 3.2.1.3 Verifying the integration of *PspA*

PCR was used to identify integrants of the 434-*pspA* cassette into the chromosome. This entailed the use of four primers: P1, P2, P3 and P4 (see Figure 23). The presence of the desired product, the single integrant which occurs after Holliday junction rearrangement, gave rise to hybrid bacterial and phage attachment sites which when amplified by PCR and visualised by gel electrophoresis, resulted in the loss of a band at 740 bp and the gain of two bands at 824 bp and 289 bp. The two newly generated PCR products represent the amplification of the newly-formed phage and bacterial attachment sites.<sup>231</sup>



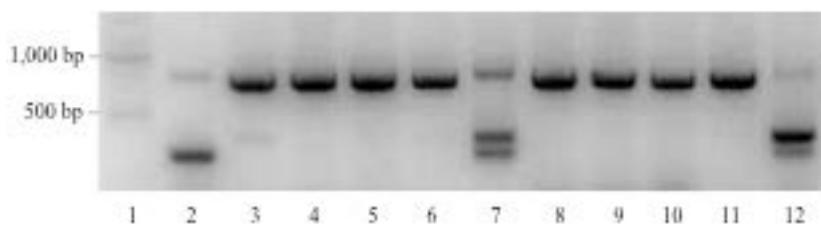
**Figure 16: Diagram illustrating how PCR is used to identify successful integration.** When there is no integration a band at 740 bp is observed, when single integration occurs two bands (824 bp and 289 bp) are amplified, and then with multiple integration three bands are observed at 824 bp, 373 bp and 289 bp.<sup>231</sup>

Figure 17 shows three integrants, signified by the bands observed at 289 bp and 824 bp; while the unsuccessful attempts have a solitary band at 740 bp.



**Figure 17: 1% agarose gel to verify successful integration.** *Lane 1:* 2Log ladder (NEB); *Lane 2:* Positive control showing two bands (824 bp and 289 bp); *Lane 3:* Negative control showing one band (740 bp); *Lane 5, 6 and 10:* Positive PspA integrants; *Lane 4, 7, 8, 9 and 11:* Negative PspA integrants.

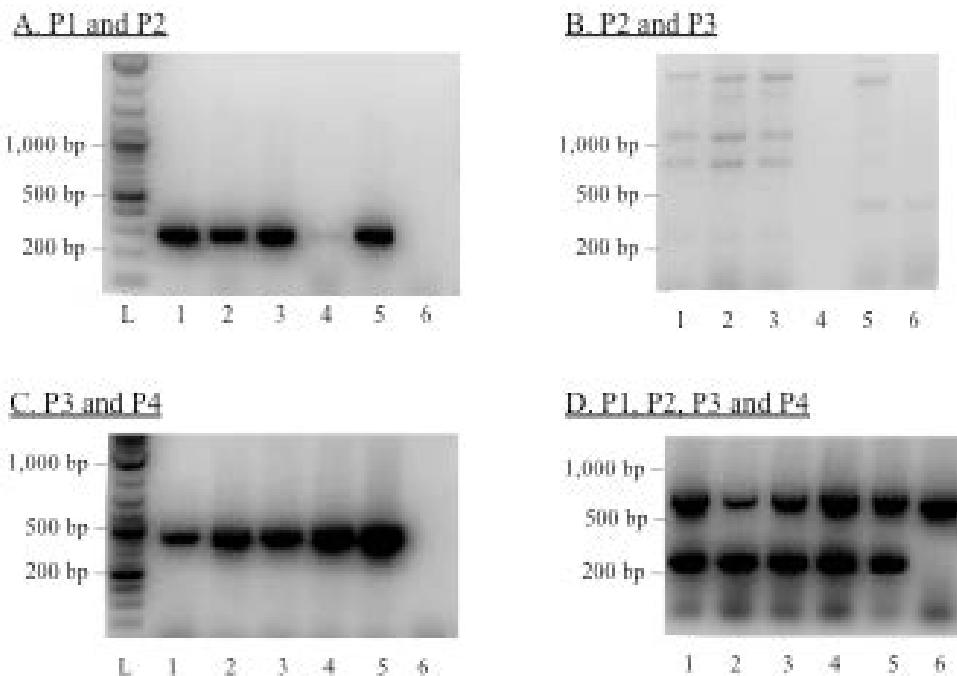
During the attempts at integration, on a number of occasions an additional band at 373 bp was observed (Figure 18). This may arise either from multiple integration, or the continued presence of pAH68 PspA plasmid. Multiple integration was frequently observed under conditions where the initial incubation temperature was 30°C, although not exclusively; it also occurred regardless of plasmid concentration. Three successful integrants were taken forward and used in the integration and interaction verification experiments (see Chapter 2).



**Figure 18: 1% agarose gel to identify successful integrants with multiple integration.** *Lane 1:* 2Log ladder (NEB); *Lane 2:* Positive control showing two bands (824 bp and 289 bp); *Lane 3:* Negative control showing one band (740 bp); *Lane 4, 5, 6 and 8-11:* Negative PspA integrants; *Lane 7 and 12:* Multiple PspA integrants (824 bp, 373 bp and 289 bp).

### 3.2.1.4 Further verification of *PspA* integration

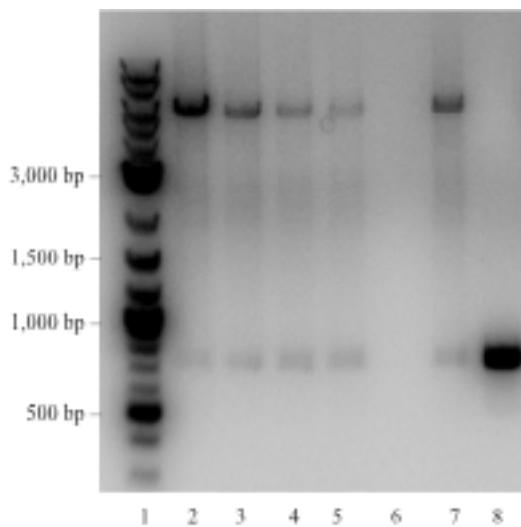
More PCR experiments were carried out in order to further confirm successful integration of pAH68 PspA into the chromosome of SNS118 strain. Figure 19 shows separation of the PCR done in tandem in Figure 17 for clarity, whereby the presence of bands at 289 bp and 840 bp indicate integration, while the absence of a band at 373 bp with P2 and P3 primers confirms single integration.



**Figure 19: Split primer PCR to confirm the integration.** Primers 1 and 2 give a band at 289 bp if integration has occurred; Primers 2 and 3 should give no band, band present at 373 bp indicates plasmid present or multiple integration; Primers 3 and 4 a band at 824 bp should be observed indicating integration. *Lane L:* 2Log ladder (NEB); *Lane 1:* Integration of pAH68 PspA 1; *Lane 2:* Integration of pAH68 PspA 2; *Lane 3:* Integration of pAH68 PspA 3; *Lane 4:* Integration of pAH68 PspA 4; *Lane 5:* Positive control (p6-UEV); *Lane 6:* Negative control (SNS118).

PCR with P1 and P4 amplified the integrated plasmid, yielding a band at approximately 6000 bp, which was isolated, purified and DNA sequenced to confirm the integrity of the *pspA* gene (Figure 20). Other PCR experiments were also carried out, using primers internal to the plasmid: these included M13-49, M13-43, PspA-F

and PspA-R and primers on the chromosome (Table 27). All confirmed the presence of *434-pspA* in the chromosome.



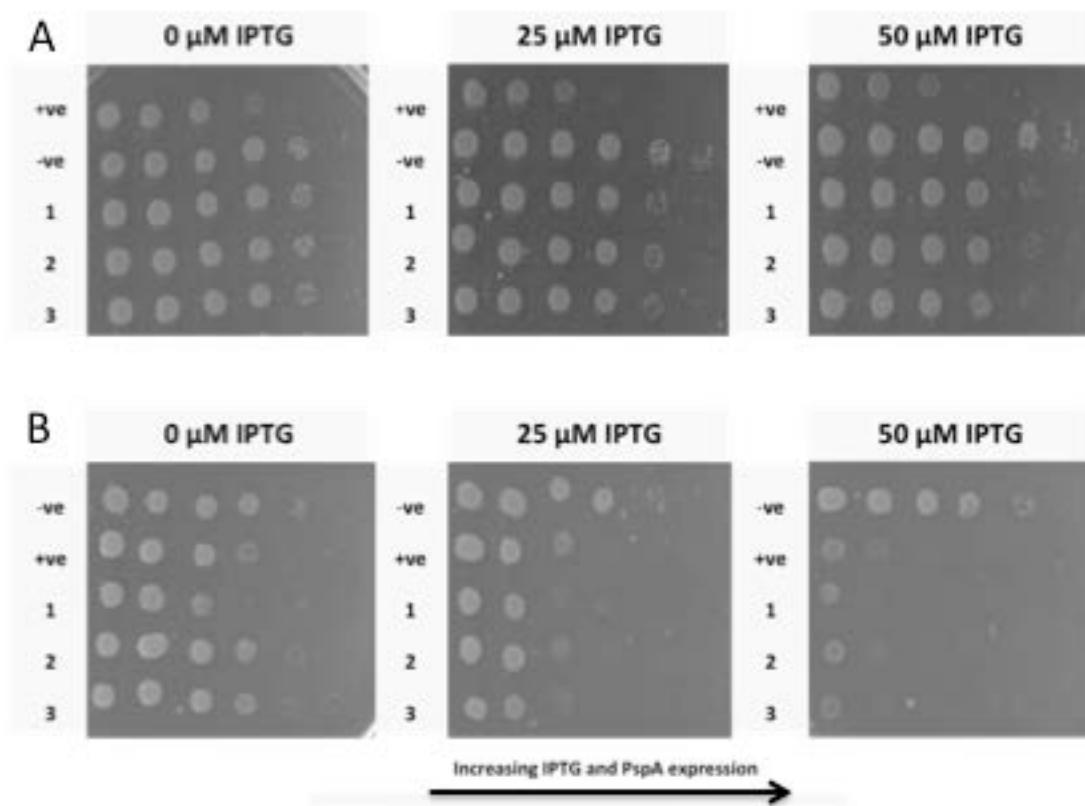
**Figure 20: PCR using P1 and P4 primers.** If integration was successful there would be a band around 6000 bp, whereas if it was not integrated then a band at 740 bp was observed. *Lane 1:* 2Log ladder (NEB); *Lanes 2-5:* PspA integrant 1-4; *Lane 7:* Positive control (p6-UEV RTHS previously built in the lab); *Lane 8:* Negative (SNS118).

The plasmid, pAH68 PspA, contains a *pir* origin of replication and so cannot replicate in the absence of the *trans*-acting  $\Pi$  protein. A *pir*+ strain is therefore essential for replication: SNS118 is not a *pir*+ strain and so the plasmid would not be expected to replicate in this strain. Despite this, to eliminate the possibility of the presence of either pAH68 PspA or the helper plasmid, all plasmids were isolated and analysed by gel electrophoresis. No bands were observed, confirming the absence of either plasmid. In addition, a PCR was carried out using the primers PspA(*Bam*HI)-F and P2 (Table 2), which further confirmed the absence of pAH68 PspA, via the absence of a band at approximately 2000 bp that would arise if the plasmid was present. This is because once integration has occurred the P2 annealing site moves upstream of the *pspA* gene preventing amplification with these primers.

### **3.2.2 Verification of the PspA PPI within the RTHS**

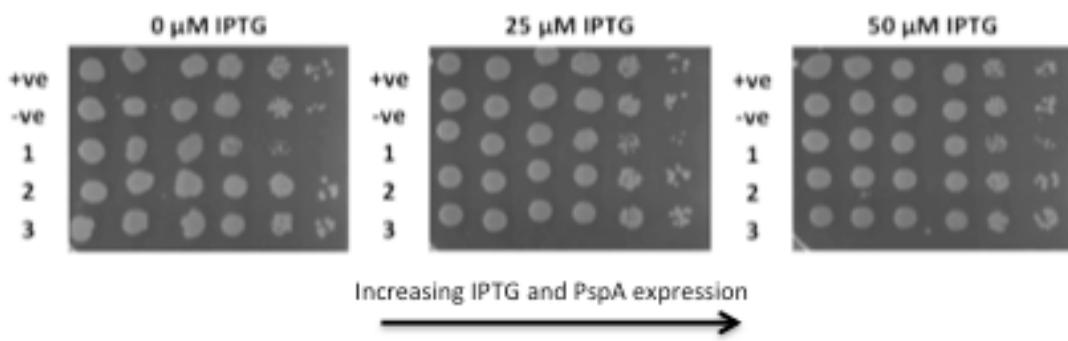
#### *3.2.2.1 Verification of the PspA PPI using a drop-spotting assay*

Having constructed the PspA RTHS, the next step was to confirm the PPI, which was achieved by drop-spotting ten-fold serial dilutions of the PspA RTHS onto selective media. The selective media contained IPTG at concentrations ranging from 0-100 µM, as well as 3-AT and kanamycin, which were used at two distinct concentrations, the first representing high stringency growth conditions (i.e. 5.0 mM 3-AT and 50 µg/ml kanamycin), while the second represented less stringent growth conditions (i.e. 2.5 mM 3-AT and 25 µg/ml kanamycin). As anticipated, the capacity for colonies to grow decreased as IPTG concentration increased, and was attenuated under the more stringent conditions. Specifically, there was a 1000-fold reduction in the number of colonies at 50 µM IPTG under the more stringent conditions, whereas under the less stringent conditions, this was reduced to 100-fold (Figure 21). The function of the PspA RTHS was compared to the heterodimeric Gag-TSG101 RTHS produced previously by the Tavassoli group (hereafter referred to as the p6-UEV RTHS after the subdomains used within the system).<sup>131</sup> This RTHS demonstrated an inhibition of growth at less stringent conditions compared to the PspA RTHS.



**Figure 21: Drop-spotting of PspA RTHS strains onto minimal media.** (A) Less stringent conditions (2.5 mM 3-AT and 25 µg/ml kanamycin). (B) More stringent conditions (5.0 mM 3-AT and 50 µg/ml kanamycin). The positive control is p6-UEV RTHS and the negative control is 434 alone integrated onto the chromosome. 1-3 are PspA RTHS samples. This showed decreased growth as IPTG was increased from 0 µM to 50 µM, 1000-fold on more stringent conditions and 10-fold shutdown on less stringent conditions.

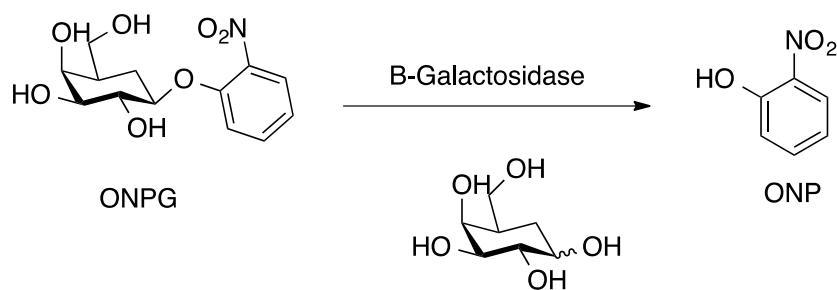
To assess if the growth inhibition observed was due to toxicity from the 434-PspA recombinant proteins, rather than protein dimerisation inhibiting expression of essential genes, the PspA RTHS was drop-spotted onto LB agar containing varying concentrations of IPTG. As the concentration of IPTG increased, the level of protein expression increased, and if the presence of these proteins were toxic, then decreased growth would be observed even in the absence of selection pressure (i.e. histidine-containing media and kanamycin). No growth decrease was observed on LB agar suggesting that the effect observed in Figure 21 was due to the PPI and not toxicity (Figure 22).



**Figure 22: Drop-spotting of PspA RTHS strains onto LB agar media containing varying levels of IPTG to show toxicity.** The positive control is p6-UEV RTHS and the negative control is 434 alone integrated into the chromosome and 1-3 are PspA RTHS samples. This showed that the recombinant protein was not toxic as IPTG was increased from 0  $\mu$ M to 50  $\mu$ M, increasing the levels of protein expressed.

### 3.2.2.2 Quantification of the PspA dimerisation by *o*-nitrophenyl $\beta$ -galactosidase assays

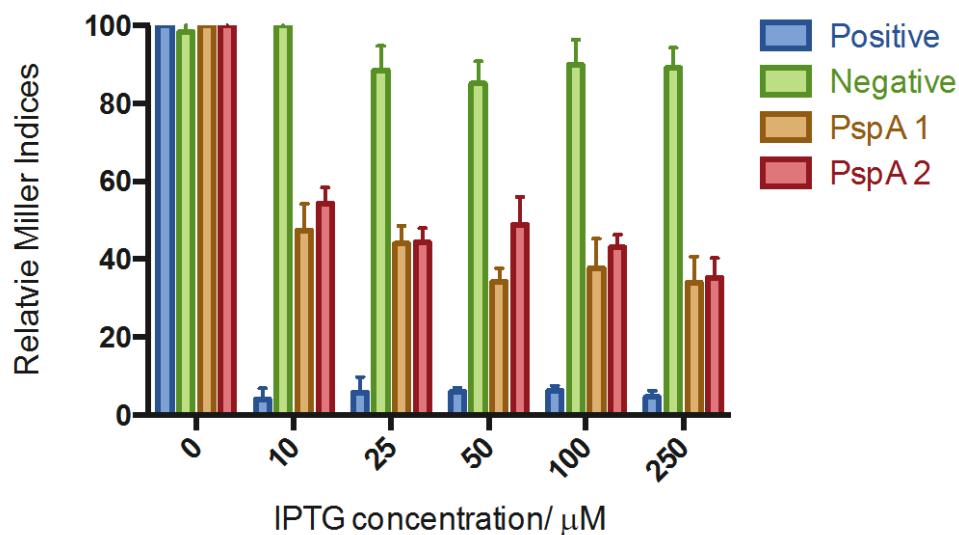
Having established that PspA dimerises within the RTHS, ONPG assays were used to quantify the PPI, using the third reporter gene (*lacZ*), which encodes  $\beta$ -galactosidase. ONPG is an artificial chromogenic compound that is colourless and can be hydrolysed using  $\beta$ -galactosidase. The product of this hydrolysis is *o*-nitrophenyl (ONP), which is yellow and detectable at  $\lambda_{\text{max}} = 420$  nm. This colour change allows the rate of conversion to be quantified using UV spectrometry (Figure 23). If ONPG is added in excess the amount of  $\beta$ -galactosidase produced can be monitored, allowing quantification of the PPI.<sup>239 240</sup>



**Figure 23: Breakdown of ONPG into ONP for ONPG assay.** ONPG (colourless) breaks down using  $\beta$ -galactosidase to ONP (yellow in colour) that can be monitored using UV spectrometry.<sup>239</sup>

If protein dimerisation does not occur, transcription of the reporter-construct, including the *lacZ* gene, is permitted, expressing β-galactosidase. If protein dimerisation occurs, then transcription of the *lacZ* gene is inhibited, preventing the breakdown of ONPG. Therefore, as IPTG concentration increases, more protein is expressed and thus less ONPG is hydrolysed. This can be visualised by a decrease in the intensity of the yellow colouration produced.

ONPG assays were carried out using two of the integrants identified (Figure 24). The data indicates that PspA follows a similar pattern to the positive control (CtBP homodimerisation RTHS).<sup>133</sup> As IPTG levels are increased, there was a decrease in the relative miller index and at higher concentrations of IPTG the relative miller index plateaued as the RTHS became saturated. The negative control remained constant, independent of the concentration of IPTG. In order to define when a PPI has occurred, Di Lallo *et al.* (2001) set the limit of interaction at 50% activity repression.<sup>93</sup> Both PspA RTHS tested showed a decrease of more than 50% confirming that the protein was interacting in the RTHS and forming a functional repressor.



**Figure 24: Graph of the average results from the ONPG assay.** Carried out in triplicate with two of the PspA RTHS and a positive control of CtBP RTHS, a previously constructed RTHS and a negative control of SNS118 with only 434-binding domain expressed. Strains with PspA present showed a decrease in ONPG breakdown as IPTG increased.

### **3.2.3 SICLOPPS screening against PspA homodimerisation**

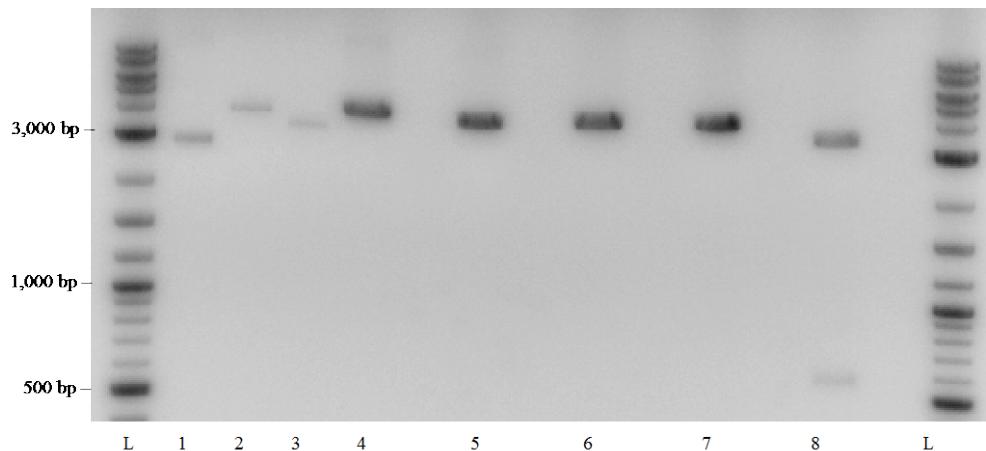
Four SICLOPPS libraries were constructed: CX<sub>5</sub>, SGWX<sub>5</sub>, SX<sub>5</sub>G and SX<sub>6</sub>G. The aim was to transform these into the PspA RTHS and to utilise drop-spotting to show inhibition of the dimerisation. The CX<sub>5</sub> was the smallest library, with a very active nucleophile involved in the cyclisation step during intein processing. Issues arise when trying to synthesise these cyclic peptides, due to the cysteine requiring additional protection to prevent side-chain cyclisation and the potential for the formation of disulphide bonds. Moreover, if the degenerate sequence contains no chromophore, then purification of the resulting peptide by HPLC is difficult as there is no signal at 280 nm and only the backbone at 220 nm is detected. The SGW library eliminates these problems by including a tryptophan and exchanging the cysteine with a serine, which is a less powerful nucleophile so whilst intein processing may be slower the selected peptides will be easier to synthesise. Furthermore, the inclusion of glycine eliminates racemisation during the cyclisation step of the chemical synthesis.<sup>241</sup> The final two library sequences, SX<sub>5</sub>G and SX<sub>6</sub>G libraries ensured that the *in vivo* cyclisation junction is constant for all sequences, theoretically reducing intein-processing variation. This library was used to identify inhibitors of β-clamp formation, initially an SIIDSAGX<sub>6</sub>GASTSESG library was screened and then the top sequences were tested further using the reduced SX<sub>6</sub>G library, however, the sequence bias of the library has not been investigated.<sup>242</sup>

#### **3.2.3.1 SGWX<sub>5</sub> Library construction**

The SGWX<sub>5</sub> library was constructed as outlined in Tavassoli and Benkovic (2007) whereby the degenerate region was amplified by PCR with primers, SGW-F and CBD-R (Table 2), which included a repeating NNS motif and incorporated the restriction sites *Hind*III and *Bgl*II. A secondary PCR followed to realign the mismatched strands. This fragment was cloned into pARCBD by restriction digestion with *Hind*III and *Bgl*II.<sup>117</sup> Originally constructed by Scott *et al.* (2001), using the plasmid pAR3 as the backbone, pARCBD allows for the expression of the intein construct.<sup>100,125</sup> pARCBD contains: chloramphenicol resistance; the p15A origin of replication; the AraC repressor; and an AraB promoter (after which the I<sub>C</sub>-peptide-I<sub>N</sub> construct was cloned).<sup>100,243</sup> The p15A origin of replication provides a

low copy number of the plasmid (10-12 copies per cell), restricting the expression level and therefore limiting the toxicity effects produced by the expression of the SICLOPPS construct. Unlike the native protein sequence from *S. sp* PCC6803, an A35H mutation of the C-terminal intein was introduced to improve the yield of cyclic peptide product by more efficiently catalysing the final step of the intein processing (Figure 7). A sequence coding for a chitin-binding domain was introduced as a fusion of the  $I_N$  intein to allow for protein purification.<sup>100</sup>

Optimisation of the restriction digestion was required to gain sufficient backbone and insert concentration to generate the library. The restriction endonucleases (*Bg*II and *Hind*III) used to restriction digest the fragments do not work optimally in the same buffers. Consequently, analytical restriction digestions analysed by gel electrophoresis were carried out to identify whether sequential restriction digestion with each enzyme in turn is more efficient, or if the two enzymes could be used together in sub-optimal buffer conditions. The latter would reduce the number of purification steps required, improving the final yield of the digested product. The restriction endonucleases were shown to undergo complete restriction digestion within 16 hours regardless of the buffer used (Figure 25).

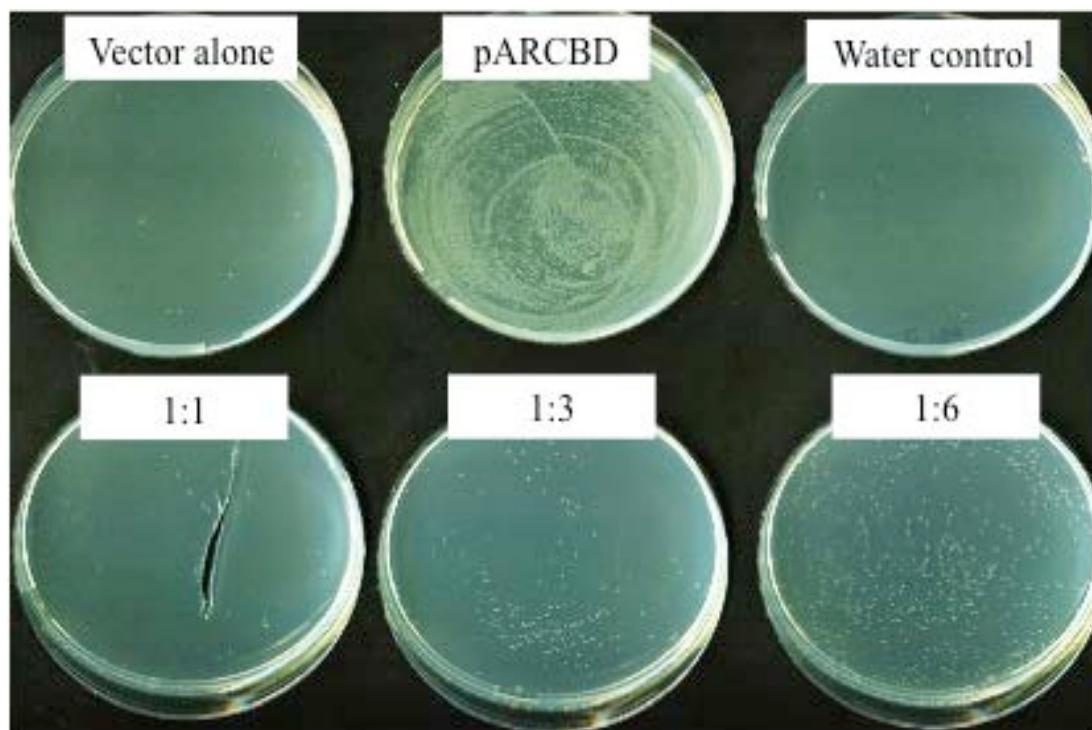


**Figure 25: Gel investigating restriction digestion with *Hind*III and *Bg*II in different buffers.** Lane L: 2Log ladder (NEB); Lane 1: pARCBP uncut; Lane 2: pARCBP digested with *Hind*III; Lane 3: pARCBP digested with *Bg*II and *Hind*III separately; Lane 4: pARCBP digested with *Hind*III in buffer 3; Lane 5: pARCBP digested with *Hind*III in buffer 2; Lane 6: pARCBP digested with *Bg*II in buffer 2;

*Lane 7: pARCBdigested with *Bgl*I in buffer 3; Lane 8: pARCBdigested simultaneously digested with *Hind*III and *Bgl*I in buffer 2.*

Specifically, simultaneous restriction digestion of pARCBdigested was observed following restriction digestion with *Hind*III and *Bgl*I in buffer 2, with two bands visible at 3376 bp and 561 bp, corresponding to the backbone and the unwanted insert, respectively (Lane 8, Figure 25).

Optimisation of the ligation to allow full coverage of the library was required during transformation in *E. coli*. The ligation was carried out for 16 hours at both 4°C and 16°C with no difference observed: however, when the ratio between backbone (100 ng) and insert was varied from 1:1, 1:3 and 1:6 (backbone:insert), transformation efficiency greatly varied. By plating onto LB agar, a 100-fold increase was observed between the 1:1 and 1:6 ratios (Figure 26).



**Figure 26: LB agar plates showing transformation efficiency of different ratios of backbone and insert used in a ligation.** 100 ng backbone was used with either 80 ng (1:1), 240 ng (1:3) or 480 ng (1:6) of insert being used. This showed that for this ligation a ratio of 1:6 provided the highest efficiency.

Unlike the protocol from Tavassoli and Benkovic (2007),<sup>117</sup> in the current study the ligation mixture was transformed directly into the PspA RTHS, rather than transforming initially into a generic strain, this eliminated any sequence bias that may be caused by this step.

Optimisation of the transformation was required to allow complete coverage of the library. Successive optimisation of the transformation efficiency has been outlined below; using the protocol from Tavassoli and Benkovic (2007) the efficiency was  $4.5 \times 10^6$ .<sup>117</sup> Small modifications to the protocol, for example the pre-incubation time and the temperature of the SOC media, had little effect on the efficiency. Similarly, doubling the backbone concentration and dialysing four ligations had little effect on the efficiency.<sup>244</sup> A ten-fold improvement was observed, however, when cells were cultured at 18°C for 16 hours which causes an increase in the production of saturated lipids in the cell membrane: improving the number of cells surviving electroporation and thus increasing transformation efficiency. By culturing at 18°C and improving experimental technique (e.g. keeping cells cold throughout and suspending cells gently in order to prevent lysing of cells) transformation efficiencies of greater than  $10^7$  were obtained for each library.<sup>245</sup> This allowed SGWX<sub>5</sub> library transformation with an efficiency of  $1.8 \times 10^7$ , providing full coverage of the library.

### *3.2.3.2 SX<sub>5</sub>G and SX<sub>6</sub>G library construction*

Similar to the SGWX<sub>5</sub> library, the SX<sub>5</sub>G and SX<sub>6</sub>G libraries were generated according to the procedure outlined in the Tavassoli and Benkovic (2007),<sup>117</sup> and the optimised procedures outlined in Section 3.2.3.1. The optimised transformation protocol led to efficiencies of  $8.5 \times 10^7$  and  $7.0 \times 10^7$  for SX<sub>5</sub>G and SX<sub>6</sub>G, respectively, which would lead to full coverage of both libraries.

### *3.2.3.3 CX<sub>5</sub> library construction*

In contrast to the SGWX<sub>5</sub> and SX<sub>5/6</sub>G libraries, the CX<sub>5</sub> library was not constructed according to the Tavassoli and Benkovic protocol (2007).<sup>117</sup> Instead the CX<sub>5</sub> method involved two primers (Table 2) that annealed to each other, eliminating the need for a template, creating a smaller library insert that may improve ligation efficiency. The

primers also introduced the restriction sites *Bgl*I and *Bgl*II to allow cloning into pARCB. The transformations were initially attempted using chemically-competent NEB-10- $\beta$  cells, but this only yielded an efficiency of  $10^6$ . Therefore, commercial NEB-10- $\beta$  electro-competent cells were used instead; this resulted in an efficiency of  $2.0 \times 10^8$ . This library was then extracted and re-transformed into the PspA RTHS with an efficiency of  $8.8 \times 10^7$ .

#### *3.2.3.4 Screening of the four libraries*

After transformation of the libraries into the PspA RTHS, the recovery solution was plated onto selective minimal media plates; the supplements, quantities and reason for each component are outlined in Table 24. From these plates, colonies were cultured in LB media and then drop-spotted onto four different plates, each containing antibiotics, 3-AT and growth supplements. Of these four plates, one also contained IPTG (50  $\mu$ M); another contained arabinose (6.5  $\mu$ M); and one had both IPTG and arabinose added, while the fourth plate had no additional supplements

**Table 24:** Table describing all the supplements added to the minimal media agar for the drop-spotting and library selection plates.

SUPPLEMENT	AMOUNT	REASON FOR ADDITION
Ampicillin	50 µg/ml	Selective pressure to ensure maintenance of pAH68 PspA construct in the chromosome
Spectinomycin	25 µg/ml	
3-AT	5 mM	Conditions identified by drop-spotting to show inhibition of growth when the reporter gene construct expression is inhibited
Kanamycin	50 µg/ml	
Chloramphenicol	35 µg/ml	Selective pressure to maintain pARCB
IPTG	50 µM	To induce the expression of the 434-PspA recombinant protein
Arabinose	6.5 µM	To induce the expression of the cyclic peptide
MgSO <sub>4</sub>	1 mM	Essential nutrients required for cell survival
Minimal media salts	1 x	
Glycerol	4%	Carbon and energy source

On drop-spotting plates supplemented with no arabinose or IPTG, full growth was observed: this was because neither the cyclic peptide nor PspA were expressed. Upon addition of IPTG a 1000-fold decrease in growth was observed due to the expression of PspA, its subsequent dimerisation and inhibition of transcription of

essential genes. On plates supplemented with only arabinose, if the cyclic peptide was toxic, cell death was observed. In the presence of both IPTG and arabinose, however, the expression of the cyclic peptide and PspA was induced; restoration of growth for some samples was observed because the dimerisation of PspA was disrupted by the cyclic peptide allowing expression of the reporter gene construct.

Table 25 outlines the colonies screened from each library. Appendix 2 shows sample plates from an SGW library screen whereby five colonies show inhibition of dimerisation by showing a growth advantage on the arabinose and IPTG selective media.

**Table 25: Table outlining the different libraries used for screening.**

LIBRARY	NUMBER OF COLONIES PICKED	INHIBITORS IDENTIFIED
CX <sub>5</sub>	288	C1-6
CX <sub>5</sub>	192	C7-12
CX <sub>5</sub>	288	C13-31
CX <sub>5</sub>	96	C32
CX <sub>5</sub>	96	C33-50
SGWX <sub>5</sub>	288	S1-30
SX <sub>5</sub> G	96	SX <sub>5</sub> G1-17
SX <sub>6</sub> G	96	SX <sub>6</sub> G1
<b>TOTAL</b>	<b>1440</b>	<b>98</b>

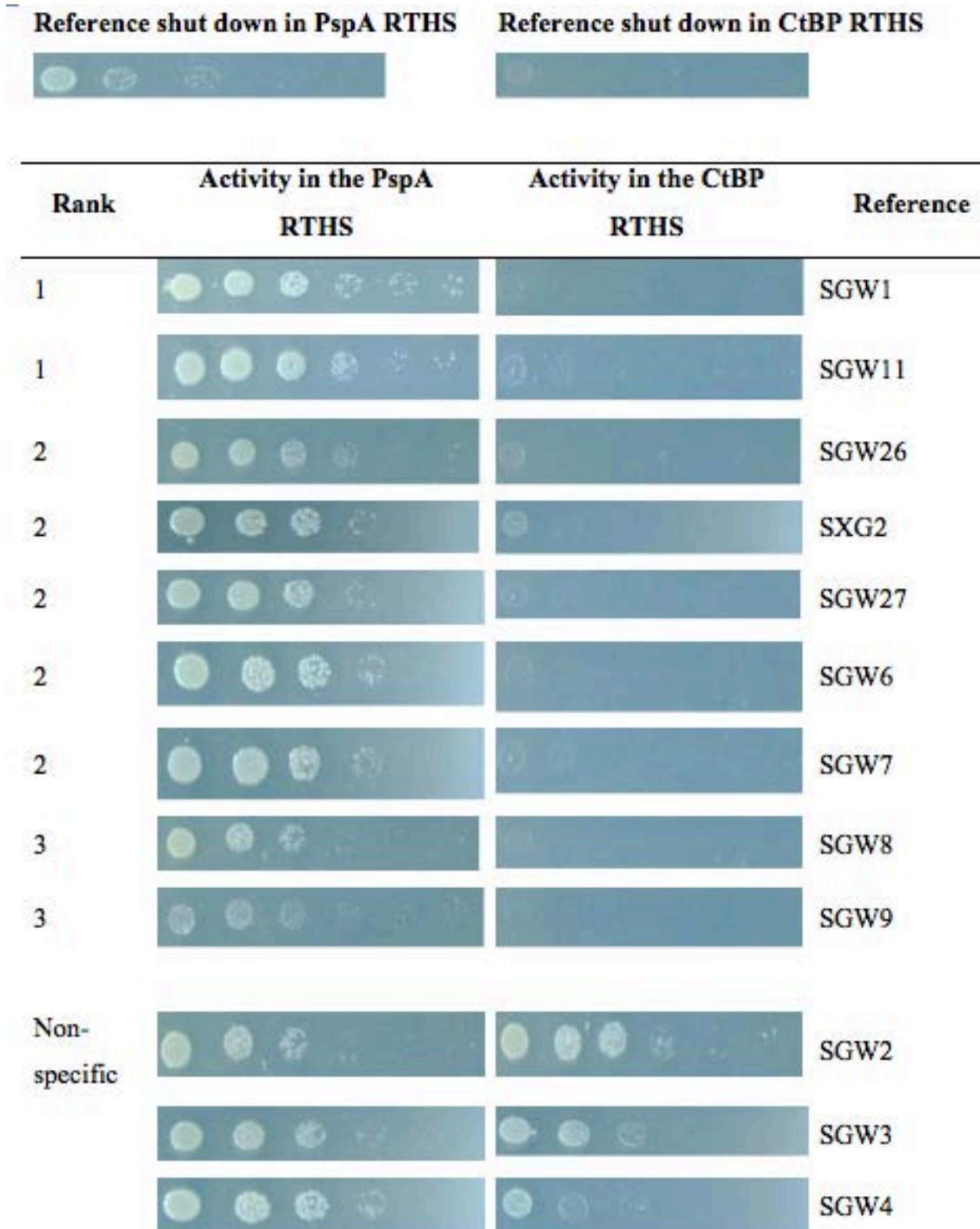
---

Of the 1440 colonies isolated and drop-spotted, 98 showed potential inhibitory activity, putatively, by disrupting PspA dimerisation.

### *3.2.3.5 Secondary screening and screening for specificity*

Following each screen, any potential inhibitors identified (7% of the total samples drop-spotted) were cultured, and the active SICLOPPS plasmid was isolated and re-

transformed back into the PspA RTHS strain (in order to check for retention of phenotype) and also transformed into the CtBP RTHS strain.<sup>133</sup> CtBP is an unrelated mammalian homodimeric protein interaction. Specific inhibitors of PspA will have no effect on the CtBP RTHS.<sup>133</sup> With the CX<sub>5</sub> library, all 50 inhibitors failed to either maintain their phenotype or showed signs of non-specific inhibition. The reason for the loss of phenotype could be due to false positives. It may also be considered that as PspA forms an oligomeric species the proteins may interact by multiple orientations to form the oligomer. It may be that the inhibitor did inhibit one orientation of the PspA interaction but that there are other possible interactions that mean the transcription of the reporter genes is still inhibited, and that the concentration of the inhibitor was insufficient to disrupt the multiple interactions. Despite the lack of success with the CX<sub>5</sub> library, 12 inhibitors in the SGWX<sub>5</sub> and SX<sub>5/6</sub>G screens showed retention of phenotype, and 9 of these were specific for the PspA PPI (Figure 27).



**Figure 27: Drop-spotting ranking of selected peptides alongside their reference number.** The reference shut-down was drop-spotting on minimal media containing no arabinose, only supplemented with IPTG.

### 3.2.3.6 Sequencing and ranking of active SICLOPPS peptides

The peptides that retained the phenotype and showed specificity towards the PspA PPI were sequenced (Figure 28). Although there are no repeating motifs, the top two peptides are both hydrophobic peptides and contain aromatic residues. Despite eliminating the ochre and opal stop codons at the nucleotide level, one-third of the active plasmids sequenced contained amber stop codons, meaning that cyclisation would not have been possible.

Rank 1	S	G	W	L	V	Y	W	F
	S	G	W	S	I	V	F	I
Rank 2	S	G	W	S	M			
	S	G	W	I	F	P		
	S	G	W	W	H	A	M	M
	S	G	W	R	S	F	C	
	S	S	A	S	L	P	G	

**Figure 28: Sequences of active peptides identified through SICLOPPS screening.** Ranking of the successful cyclic peptides according to their growth in the presence of arabinose and IPTG. The motif from the library design are in grey, yellow represents non-polar aliphatic residues, green polar uncharged residues, blue aromatic residues, black stop codons and purple represents charged residues.

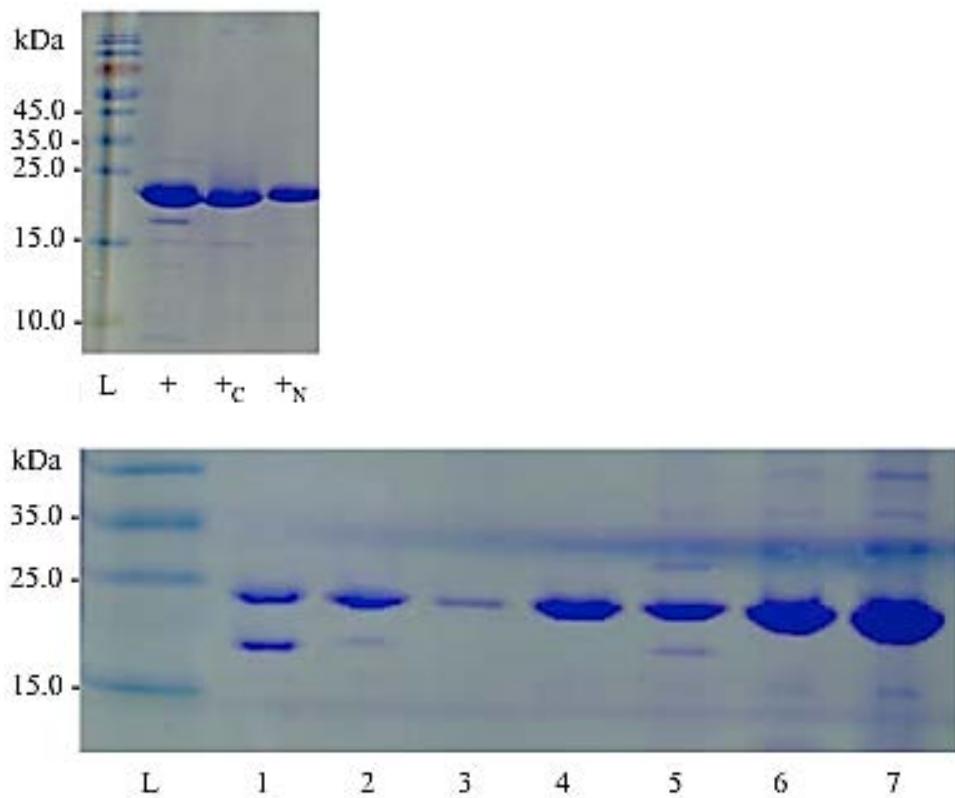
### 3.2.4 Verification of intein expression in top ranking peptides

In order to verify intein processing during screening and improved activity of the cyclic peptide compared to their linear counterparts, non-splicing variants were constructed. Briefly, nine plasmids were constructed (outlined in Table 26) whereby a C- and N-terminal intein mutant of each peptide was cloned. The positive control was the top ranking cyclic peptide inhibitor identified against the p6-UEV heterodimerisation.<sup>131</sup>

**Table 26: The intein mutations used during the construction of non-splicing inteins to verify intein processing.<sup>246</sup>**

SEQUENCE	MUTATIONS	REFERENCE
SGWLVWYF	None	1
SGWLVWYF	H24A and F26A	1 I <sub>C</sub>
SGWLVWYF	T69A and H72A	1 I <sub>N</sub>
SGWSIVFI	None	11
SGWSIVFI	H24A and F26A	11 I <sub>C</sub>
SGWSIVFI	T69A and H72A	11 I <sub>N</sub>
SGWIYWNF	None	+
SGWIYWNF	H24A and F26A	+ I <sub>C</sub>
SGWIYWNF	T69A and H72A	+ I <sub>N</sub>

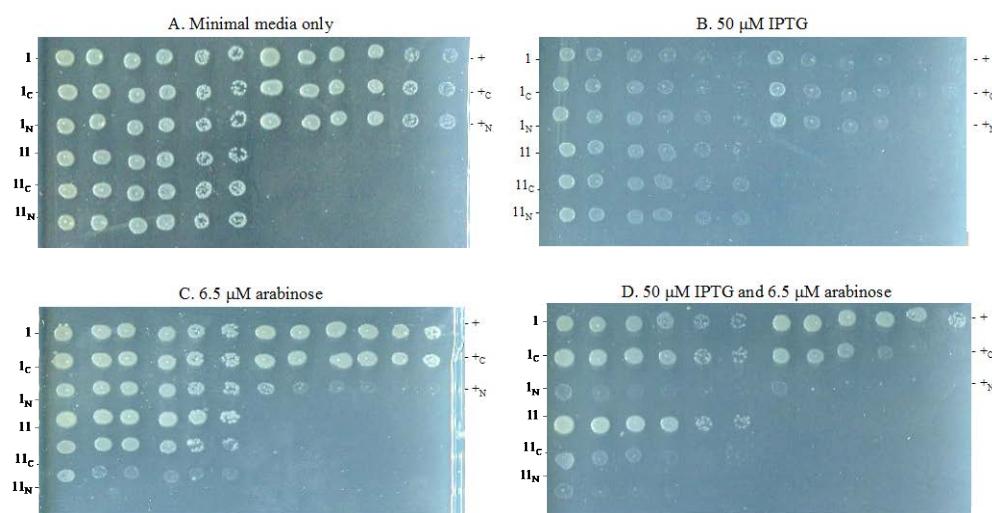
To confirm splicing, the plasmids were transformed into BL21 (DE3) and the inteins were expressed, bound to chitin affinity beads and incubated overnight at room temperature. The beads were then analysed by SDS-PAGE gel, alongside the positive control (Figure 29). For the non-mutated constructs bands at 25.6 kDa and 21.7 kDa were observed corresponding to the unprocessed I<sub>C</sub>-Peptide-I<sub>N</sub> construct and I<sub>N</sub>-CBD sequences, respectively. The latter intermediate indicates intein processing. The band corresponding to intein processing was absent in the mutant strains, suggesting that the inteins did not process correctly; this result was confirmed in the positive control.



**Figure 29: SDS-PAGE gel showing intein processing and the absence of intein processing in mutated inteins.** (A) Positive control. *Lane L:* Ladder; *Lane +:* SGWIYWNF showing splicing; *Lane +<sub>C</sub>:* C-terminal mutant SGWIYNWF showing no splicing; *Lane +<sub>N</sub>:* N-terminal mutant SGWIYNWF showing no splicing. (B) PspA peptides; *Lane L:* Ladder; *Lane 1:* pARCBP; *Lane 2:* SGWLVYWF showing splicing; *Lane 3:* C-terminal mutant SGWLVWYF showing no splicing; *Lane 4:* N-terminal mutant SGWSIVFI showing no splicing; *Lane 5:* SGWSIVFI showing splicing; *Lane 6:* C-terminal mutant SGWSIVFI showing no splicing; *Lane 7:* N-terminal SGWSIVFI showing no splicing.

The plasmids were subsequently transformed into the corresponding RTHS (i.e. the PspA RTHS and for the positive control the p6-UEV RTHS). These were then drop-spotted onto conditions described in Section 3.2.2.1. On plates supplemented with no arabinose or IPTG, full growth was observed: this was because neither the cyclic peptide nor PspA were expressed. Upon addition of IPTG a 1000-fold decrease in growth was observed due to the expression of PspA, its subsequent dimerisation and inhibition of transcription of essential genes. On plates supplemented with only arabinose, if the cyclic peptide was toxic, cell death was observed, this was the case

for the  $I_N$  mutants for peptide 11 and the positive control (Figure 30). In the presence of both IPTG and arabinose, under which the expression of the cyclic peptide and PspA were induced, restoration of growth was observed for all processing inteins. For peptide 11 (*cyclo*-SGWSIVFI) and the positive control no restoration of growth was observed with both mutants suggesting that the linear peptide was not active. However, for peptide 1 (*cyclo*-SGWLVYWF) growth was observed for the C-terminal mutant, suggesting that the linear peptide may be able to disrupt the PspA dimerisation. These results therefore indicate that for the linear and cyclic variants of SGWLVYWF could be tested *in vitro*.



**Figure 30: Drop-spotting of the active inteins and the mutant, non-splicing inteins.** (A) Drop-spotting onto minimal media no peptides or recombinant 434-PspA are being expressed, full growth observed. (B) Drop spotting onto minimal media containing IPTG, the recombinant 434-PspA protein was expressed repressing the transcription of the reporter genes causing cell death. (C) Drop-spotting onto minimal media containing arabinose, the expression of the cyclic peptide was activated; full growth was observed when the inteins or peptides are not toxic. The N-terminal mutant does not grow fully indicating toxicity. (D) Drop-spotting onto minimal media containing arabinose and IPTG. Cell death was observed where the processed peptide was not active. Peptide 1 is *cyclo*-SGWLVYWF and peptide 11 was *cyclo*-SGWSIVFI, the positive control was SGWIYWNF.

### **3.3 Results: Testing the PspA inhibitors *in vitro***

Following screening, the active sequences were elucidated and ranked. The two most active cyclic peptides against PspA oligomerisation were synthesised using Fmoc solid-phase peptide synthesis to allow further validation of inhibitor activity by *in vitro* and *in vivo* assays.

#### ***3.3.1 Synthesis of the cyclic peptide inhibitors***

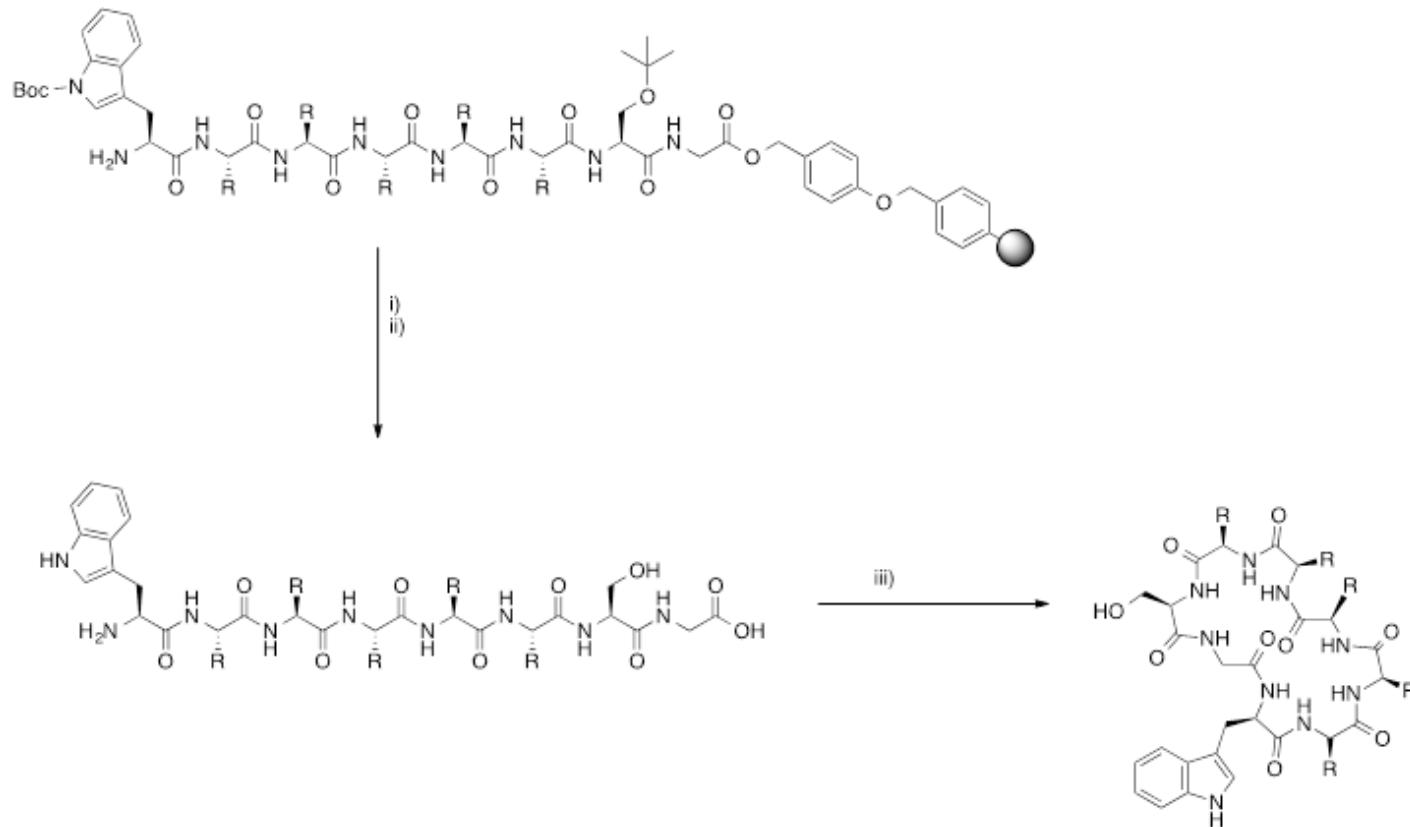
The inhibitors *cyclo*-SGWLVYWF and *cyclo*-SGWSIVFI were synthesised using standard Fmoc solid-phase peptide synthesis. Pioneered by Robert Merrifield in the early 1960s, solid-phase synthesis allows the stepwise addition of protected amino acids to a growing peptide chain, which is bound by a covalent bond to a solid polystyrene resin.<sup>247</sup> It is more efficient, overcomes the solubility issues common to solution-phase synthesis and provides a convenient work-up and purification procedure. Standard solid-phase synthesis requires protection of the free amine by either the tert-butyloxycarbonyl (Boc) or an Fmoc protecting group. Fmoc-protection provides an orthogonal protection methodology, allowing the specific deprotection under basic conditions of the amine group, leaving the acid-labile side-chain protecting groups intact. In contrast, cleavage of the Boc protecting group would deprotect any acid-labile protecting groups and cleave the peptide from the resin simultaneously.<sup>248,249</sup> In the case of the peptides synthesised in this project an acid-labile Wang linker resin was used alongside acid-labile protecting groups for the side-chains. Therefore, Fmoc-protection of the free amine was used to allow orthogonal deprotection.

The protocol of the solid-phase synthesis of a peptide used here involved an initial deprotection of an Fmoc-group using 20% piperidine in DMF. Due to the non-spontaneous nature of the amide bond formation, with the necessary elimination of water taking place only at high temperature (>200°C), it was critical to activate the carboxylic acid.<sup>250</sup> Several coupling reagents could have been used. Here initial activation was carried out using a carbodiimide to produce an *O*-acylurea.<sup>250</sup> There are two main examples of carbodiimides available: dicyclohexycarbodiimide (DCC), which produces an insoluble dicyclohexylurea that would not be removed during the

filtration/purification steps, and DIC, which generates a by-product soluble in DMF, allowing easy purification by filtration. Consequently, DIC was used to synthesize the peptides described in this chapter. In addition, the additive HOBt, was used in conjunction with the DIC, as HOBt suppresses racemisation at non-chiral centres in the peptides.<sup>251,252</sup> The mechanism involves the protonation of the DIC by the peptide carboxylic acid and then attack by the carboxylate anion forming the *O*-acylurea, which provides a good leaving group. This species can then react with HOBt to give the active OBT ester, which in turn encourages the approach of the amine via hydrogen bonding.

Despite the highly efficient coupling techniques, efficiencies can be decreased due to sequence, steric hindrance or  $\beta$ -sheet formation. Therefore, the deprotection and coupling steps were monitored using ninhydrin (2,2-dihydroxyindane-1,3-dione). The test, known as the Kaiser test, monitors the presence of the free amine. When ninhydrin reacts with free primary amines it forms a violet colour compound, known as Ruhemann's purple. The Kaiser test can be quantitative by monitoring at 570 nm. However, the test is not always accurate as the pyridine present in Kaiser solution A can partially deprotect the Fmoc-protected amino acid residue, allowing reaction with the ninhydrin to give an elevated reading than the true coupling efficiency warrants. However, under these conditions as a qualitative test it was sufficient to demonstrate successful coupling and deprotection.<sup>253</sup> If the coupling was successful then Fmoc-deprotection and coupling were repeated until the peptide chain was complete. Upon completion, the peptide was simultaneously deprotected and cleaved from the resin using a cleavage cocktail of TFA and scavengers that reduce side reactions (i.e. water and triisopropylsilane (TIS)).<sup>254</sup> The crude linear peptide was purified by ether precipitation and RP-HPLC and lyophilised prior to cyclisation. This step typically yielded between 10-20% yields after purification.

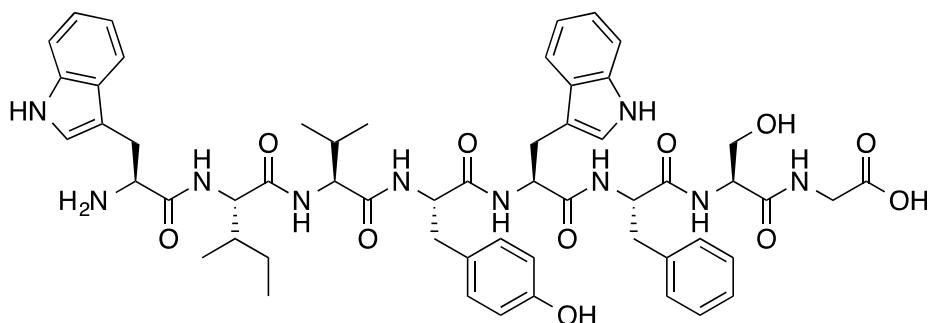
The cyclisation was carried out in solution-phase using EDC and HOBt in high dilution to allow head-to-tail peptide bond formation between the N- and the C-terminal and discourage polymerisation. Again the cyclic peptide was purified by RP-HPLC and lyophilised: this step typically yielded between 10-15% (Figure 31). All linear and cyclic peptides were analysed by low-resolution and high-resolution MS, NMR, IR, melting point and analytical HPLC.



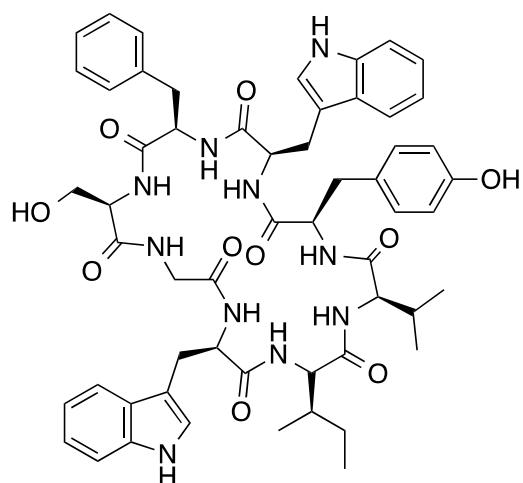
**Figure 31: Reaction scheme for the standard synthesis of an SGWX<sub>5</sub> cyclic peptide, with a glycine derived Wang resin.** The linear peptide was synthesised on resin by Fmoc deprotection followed by coupling with HOBr (3 eq), DIC (3 eq) and Fmoc-protected amino acid (3 eq) in DMF. Deprotection and coupling was verified by Kaiser test and repeated until the complete linear peptide had been synthesised. i) 20% piperidine in DMF (10 ml), 30 minutes; ii) 95% TFA, 2.5% water, 2.5% TIS, 2 hours; iii) 6 eq HOBr, 6 eq EDC in DMF (1 ml/mg), 24 hours.

### 3.3.1.1 Cyclo-SGWLVYWF (compound 2)

A (1)

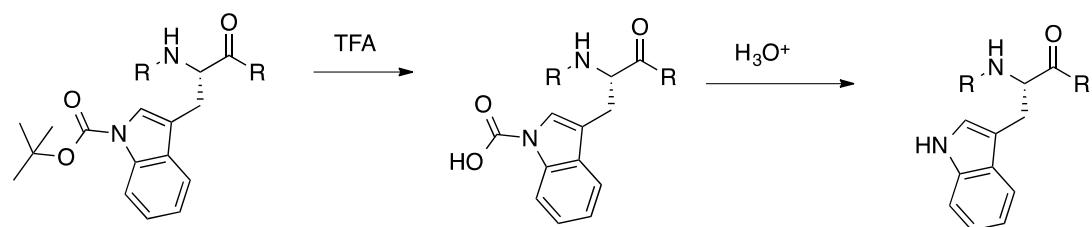


B (2)



**Figure 32:** (A) Linear-WL VYWFSG (1). (B) Cyclo-SGWLVYWF (2).

Linear-WL VYWFSG (1) was synthesised using the procedure outlined in 2.6.1. The glycine was attached to the Wang linker to prevent racemisation during cyclisation. The peptide was cleaved using the scavengers TIS and water. After purification this method resulted in a yield of 68 mg (26%). By MS a peak at mass +44 was observed, reputed to be a reversible by-product from incomplete deprotection of the tryptophan Boc-protecting group (Figure 33). Deprotection of the Boc protecting group with TFA initially cleaves the t-butyl component of the Boc protecting group, leaving an N-carboxy group attached to the nitrogen of the tryptophan group.<sup>255</sup> This carboxyl can be removed by treatment with water and lyophilisation.

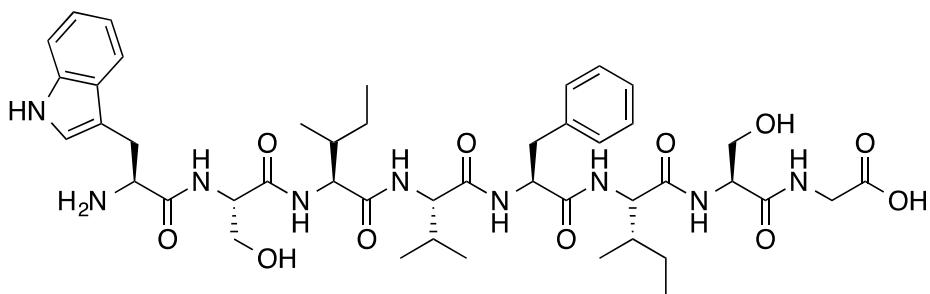


**Figure 33: Scheme of Boc removal from a tryptophan residue.**

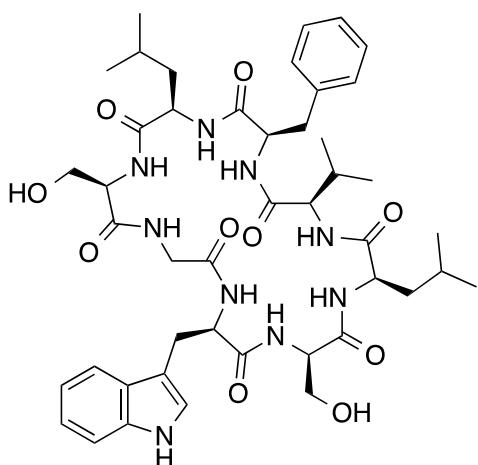
The peptide was treated with a mixture water and DMF (1:1) for six hours followed by lyophilisation. Although the carboxyl-species was still present by MS analysis the linear product was cyclised, because purification by RP-HPLC in aqueous solution removed the remaining by-product. The peptide was cyclised in high dilution and purified by RP-HPLC to yield 15 mg (29%) of compound 4. The low yield was hypothesised to be due to the low solubility of the peptides in water. *Cyclo-SGWLVYWF* (2) contains a high proportion of hydrophobic residue reducing the solubility of the compound in aqueous solution, similar low yields were observed during synthesis of the hydrophobic C-terminus of  $\beta$ -amylose.<sup>256</sup> To overcome low yields during synthesis, Condon *et al.* (2008) demonstrated how RP-HPLC was avoided by precipitation with water then ether.<sup>256</sup> Although *cyclo-SGWLVYWF* (2) was not purified using this method, future synthesis could investigate the merits of precipitation in water for partially soluble peptides.

**3.3.1.2 Cyclo-SGWSIVFI (compound 4)**

A (3)



B (4)



**Figure 34: (A) Linear-WSIVFISG (3). (B) Cyclo-SGWSIVFI (4)**

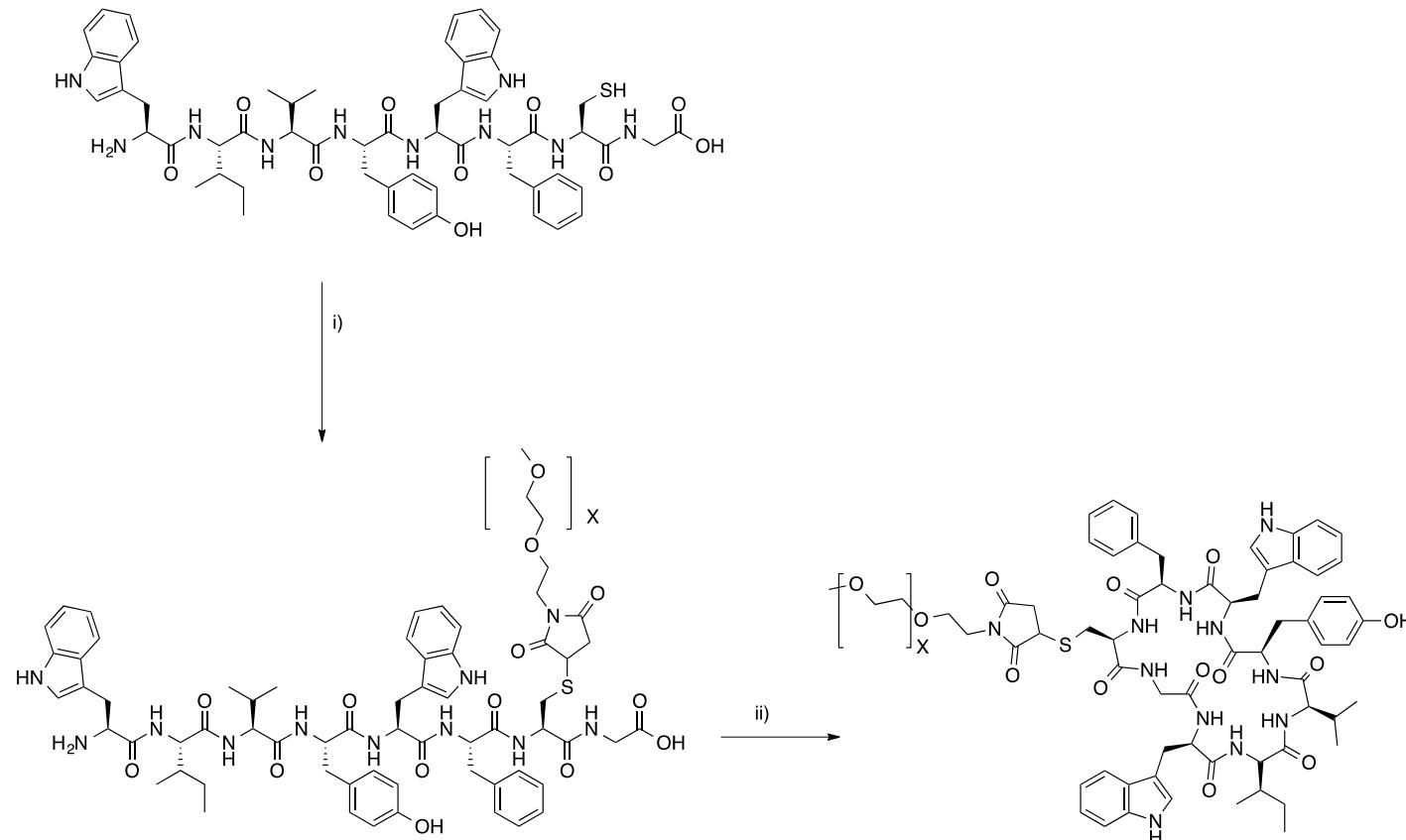
Linear-WSIVFISG (3) was synthesised using the procedure outlined in 2.6.1. As for *cyclo*-SGWLVYWF (2), glycine was attached to the Wang linker to prevent racemisation during cyclisation. The peptide was cleaved under acidic conditions in the presence of the scavengers TIS and water. After RP-HPLC purification a yield of 41.1 mg (9%) was obtained. When linear-WSIVFISG (3) was analysed by MS a mass +44 peak was also observed, similar to the peak observed for linear-SGWLVYWF. The peak was not observed after cyclisation. The peptide was cyclised in high dilution and as before purified by RP-HPLC to yield 4 mg (23%) of compound 4.

Neither *cyclo*-SGWLVYWF (2) nor *cyclo*-SGWSIVFI (4) are completely soluble in water due to their hydrophobic nature: both consist of 62% hydrophobic residues

(i.e. isoleucine, phenylalanine and valine). Probing the isoelectric properties of each peptide, linear-WLVYWFSG (1) has a predicted isoelectric point of pH 5.93 and linear-WSIVFISG (3) is pH 6.03, therefore solubilisation could be achieved in both cases by increasing the pH. Unfortunately, increasing the pH for *in vitro* assays is not always viable due to the stability of the protein, especially in ITC where the pH of the buffer must match exactly and an increase in pH to improve the solubility of the peptide could result in protein denaturation. Consequently, structural modifications to improve the solubility of peptides were investigated instead.

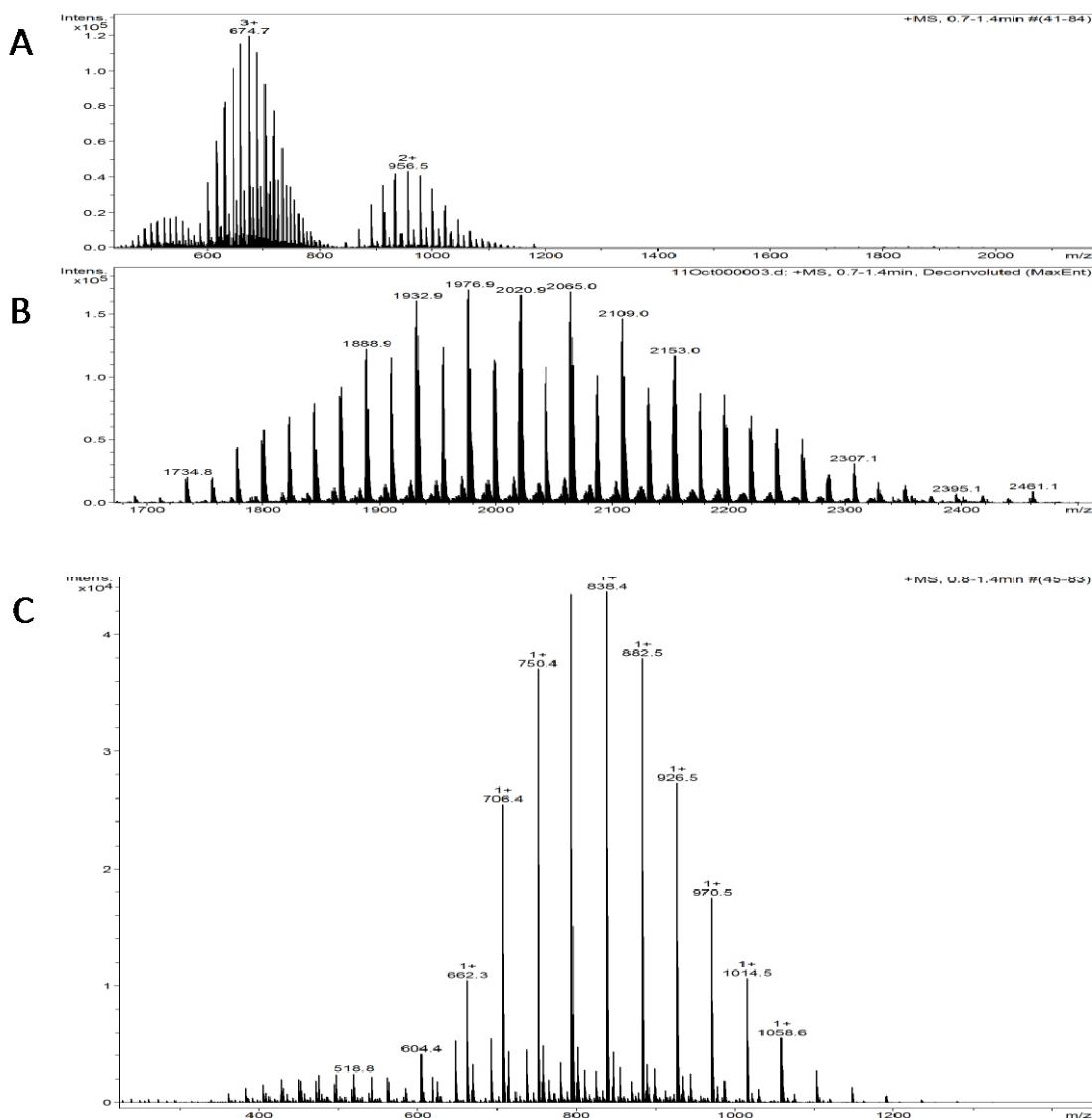
### *3.3.1.3 Improving the solubility of cyclo-SGWLVYWF (2) by PEGylation*

Initially tagging with the solubilising group polyethylene glycol (PEG) was investigated to improve the solubility of *cyclo*-SGWLVYWF (2). The cysteine analogue of linear-WLVYWFSG (1) (linear-WLVYWF<sub>CG</sub> (5)) was synthesised by standard solid-phase synthesis, and the crude peptide was reacted directly in DMF with maleimide-PEG<sub>750</sub>, triethylamine was added to maintain the pH at 7.0, by controlling the pH, undesirable attack from the N-terminus was hindered, making the cysteine the more favourable nucleophile.



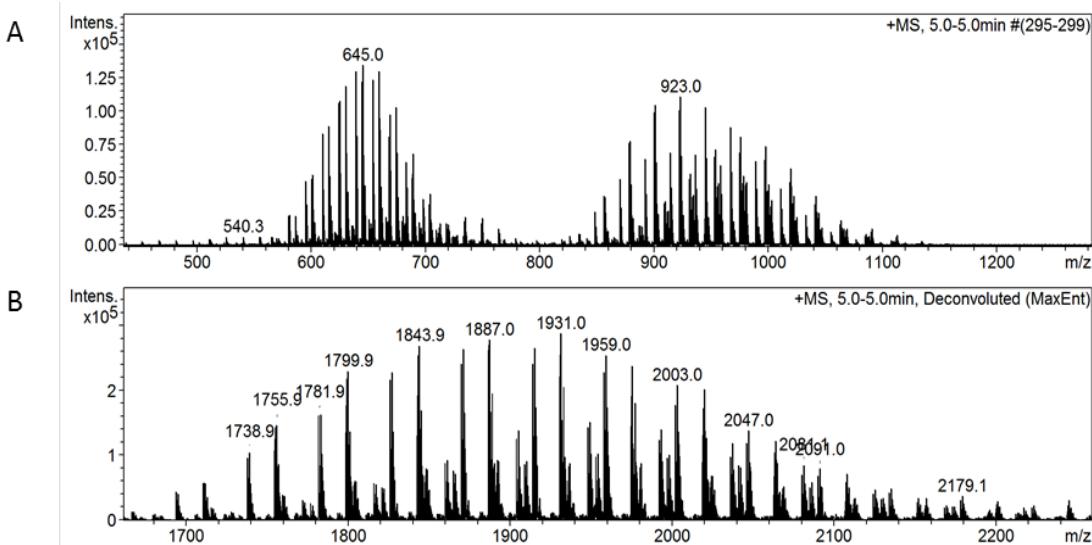
**Figure 35: Reaction scheme for the PEGylation of cyclo-CGWLVYWF peptide (7).** The linear peptide was synthesised on resin by Fmoc-deprotection followed by coupling with HOBr (3 eq), DIC (3 eq) and Fmoc-protected amino acid (3 eq) in DMF. Deprotection and coupling was verified by Kaiser test and repeated until the complete linear peptide had been synthesised. i) Maleimide-PEG<sub>X</sub> (1.1 eq), 10 µl triethylamine, RT, 16 hours; ii) 6 eq HOBr, 6 eq EDC in DMF (1 ml/mg), 24 hours. X= 750 or 5000.

The product was analysed by matrix-assisted laser desorption and ionisation-time-of-flight mass spectrometry (MALDI-TOF MS) and compared to the starting material (i.e. maleimide-PEG<sub>750</sub>). After spectral deconvolution a mass range that corresponded to successful PEGylation was observed (Figure 36A-B) and when compared to the PEG<sub>750</sub> alone (Figure 36C) the reaction was deemed successful. Unfortunately, linear-WLVYWF(PEG<sub>750</sub>)G (6) was still not completely soluble in 10% dimethylsulfoxide (DMSO) and water.



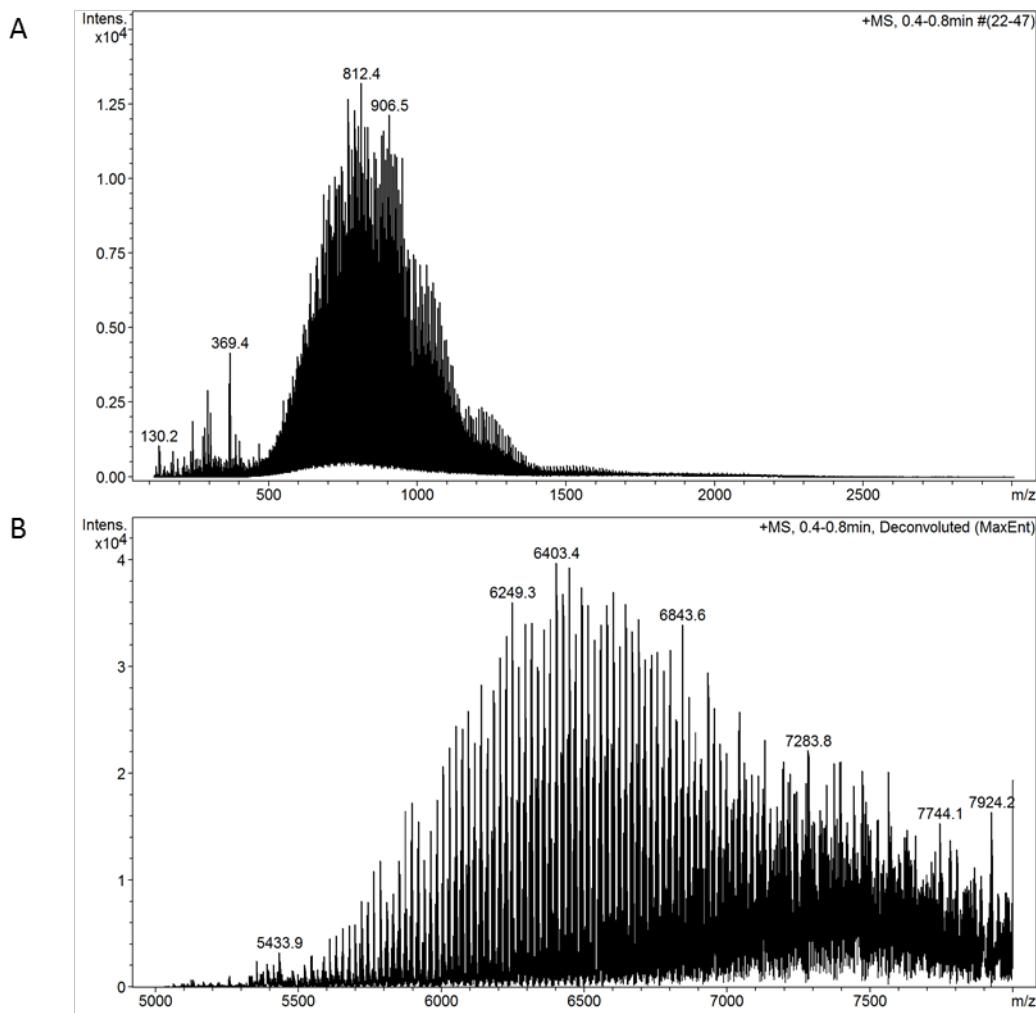
**Figure 36: Mass spectrum analysis of linear-WLVYWF(PEG<sub>750</sub>)G (6).** (A) Mass spectrum of linear-PEG750-CGWLVYWF. (B) Deconvoluted analysis of linear-WLVYWF(PEG<sub>750</sub>)G. (C) Mass spectrum of PEG<sub>750</sub> alone.

Despite the insolubility of the PEGylated peptide, as a proof of the concept the PEGylated peptide was cyclised using HOBt and EDC and yielded a mixture of starting material and cyclic product, as evident by MALDI-TOF MS (Figure 37). With optimisation this step may be viable, however: due to the persistent insolubility it was not optimised.



**Figure 37:** Mass spectrum for *cyclo-C(PEG<sub>750</sub>)GWLVYWF (7)*. (A) Mass spectrum of *cyclo-C(PEG<sub>750</sub>)GWLVYWF (7)*, showing a mixture of starting material and product. (B) Deconvoluted spectra of *cyclo-PEG<sub>750</sub>-CGWLVYWF*.

Due to the continuing solubility issues, PEG<sub>5000</sub> was used to investigate if a larger PEG species would improve solubility. Unlike PEG<sub>750</sub>, PEG<sub>5000</sub> is not soluble in DMF and so the coupling was carried out in a 1:1 ratio of water to DMF. As evident in the mass spectrum presented in Figure 38 confirmation of the correct product or completion of the reaction was difficult; confidence in the purity of the product could not be determined by MS. Notwithstanding, the complicated spectra, when compared to PEG<sub>5000</sub> alone it could be tentatively concluded that the coupling was successful and yet whether the reaction went to completion remains unresolved. The product yielded from the reaction was, however, 100% soluble in water. Despite solubilisation of linear-CGWLVYWF, PEGylation was abandoned as a method of solubilising the peptides. This was due to the lack of confidence in the characterisation of the final product as it could not be elucidated by MS or other techniques available due to the complexity of the molecule.



**Figure 38: Mass spectrum for linear-WLVYWFC(PEG<sub>5000</sub>)G (8).** (A) Mass spectrum of linear-WLVYWFC(PEG<sub>5000</sub>)G (8), showing no clear peaks indicating product or starting material. (B) Deconvolution spectra of linear-WLVYWFC(PEG<sub>5000</sub>)G (8).

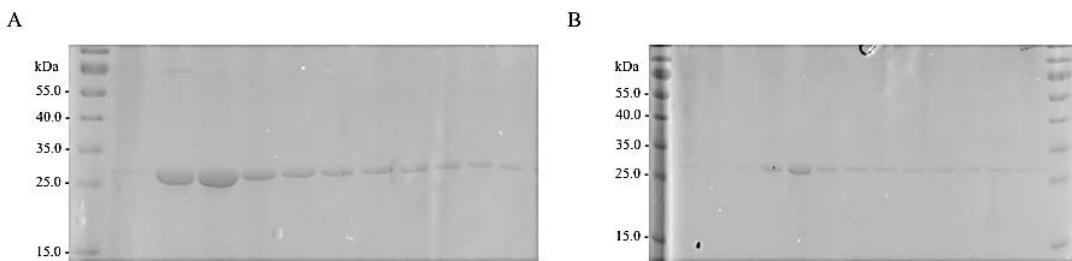
The lack of complete solubility in 10% DMSO did not affect future experiments as a saturated solution was made and the concentration was calculated by measuring the absorbance at 280 nm. Similar poor solubility was observed with *cyclo*-ADLRWF, an inhibitor of Hdmx activity; nevertheless, the inhibitor was tested and optimised to produce a potent analogue.<sup>132</sup>

### **3.3.2 PspA protein expression**

#### *3.3.2.1 Purification of the insoluble and soluble fractions*

In order to study the affinity of the selected peptides for *B. pseudomallei* PspA, the *B. pseudomallei* K96243 *pspA* gene was cloned into the expression plasmid, pET28a. The *pspA* gene was amplified by PCR to include the restriction sites *Bam*HI and *Sac*I and then cloned into the MCS of pET28a using standard cloning techniques. The final plasmid contained a gene encoding an N-terminal hexahistidine-tagged PspA protein under the control of the *lac* operon, and is hereby referred to as pET28a-PspA. The His<sub>6</sub>-tag was selected to allow purification by nickel affinity chromatography.<sup>257,258</sup> Also previous work showed how *E. coli* PspA with an N-terminal His<sub>6</sub>-tag can form 36-meric species with a suspected physiological role in restoring cell membrane integrity.<sup>209</sup>

To establish whether *B. pseudomallei* PspA can be expressed as described in the literature for *E. coli* PspA,<sup>259</sup> an initial small-scale growth of *E. coli* strain BL21 (DE3) harbouring the PspA expression plasmid pET28a-PspA was carried out. The protein was purified from the insoluble fraction after extraction in a high salt buffer containing the detergent CHAPS.<sup>259</sup> Purification by nickel affinity chromatography yielded 2.4 mg of purified recombinant His<sub>6</sub>-tagged PspA from the insoluble fraction of 4.3 g cell paste. In contrast, when His<sub>6</sub>-tagged PspA was purified from the soluble fraction (i.e. eliminating the extraction step from the insoluble fraction with buffer containing CHAPS detergent) 3.1 mg purified His<sub>6</sub>-tagged PspA was obtained from the same mass of cell paste. Both protein samples were deemed pure by SDS-PAGE analysis (Figure 39). Due to the higher yield observed from the soluble fraction and the potential for more misfolded protein from the insoluble fraction, unless otherwise stated protein in future experiments was obtained from the soluble fraction.

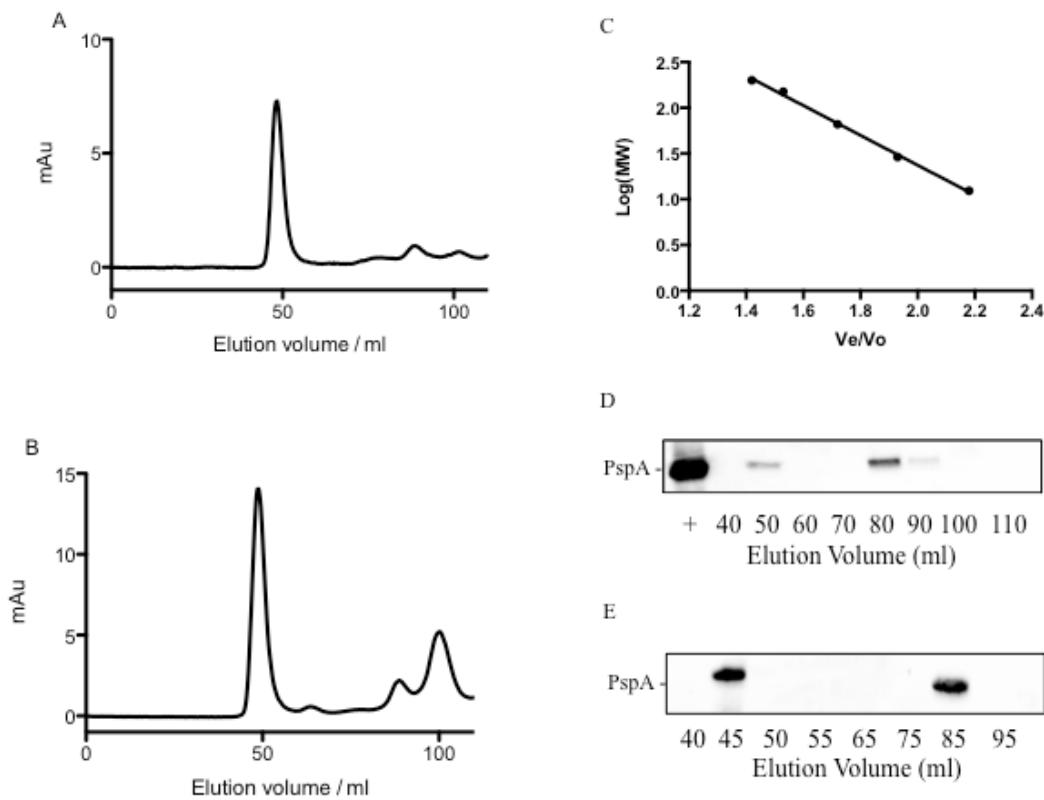


**Figure 39: Purification of His<sub>6</sub>-tagged PspA from the soluble and insoluble fraction.** (A) 10% SDS-PAGE gel of His<sub>6</sub>-tagged PspA purified from the insoluble fraction. Lanes represent an elution profile from a Nickel affinity column. (B) 10% SDS-PAGE gel of His<sub>6</sub>-tagged PspA purified from the soluble fraction. Lanes represent an elution profile from a Nickel affinity column . In both cases the protein was expressed from pET28a-PspA transformed in BL21 (DE3). Induction was carried out after an OD<sub>600</sub> of 0.6 was obtained with 1 mM IPTG and purified by nickel affinity chromatography. (Fisher PageRuler Prestained protein ladder 10-170 kDa).

Purification from the soluble fraction was therefore optimised in order to improve the yield of correctly folded protein. In accordance with the literature the protein was initially induced with 1 mM IPTG and then incubated for a further 3 hours at 30°C.<sup>259</sup> However, longer expression times (i.e. 18 hours) at a lower temperature of 18°C were also investigated. An increased proportion of protein was observed in the soluble fraction after longer incubation at lower temperatures. Protein used in the following experiments was purified from the soluble fraction using longer expression times at lower temperatures, unless otherwise stated. Furthermore, IPTG concentrations for induction and imidazole concentration during elution were investigated, but little effect was observed in the purity or concentration of the protein. Concentrations of 1 mM IPTG and 500 mM imidazole were therefore used in subsequent experiments.

In order to further purify the recombinant protein, the fractions obtained from the soluble and insoluble fraction purification were individually concentrated to 2 ml and loaded onto a Superdex 200 size exclusion column (Figure 40A and Figure 40B, respectively). This column separates proteins according to their hydrodynamic radius giving an indication of their size and molecular weight.<sup>260</sup> Using this technique it

was possible to postulate what higher-order species' of His<sub>6</sub>-tagged PspA were being formed.<sup>261,262</sup> For the protein sample purified from the insoluble fraction, the His<sub>6</sub>-tagged PspA protein eluted at 49.5 ml and between 83.3-109.0 ml, while His<sub>6</sub>-tagged PspA purified from the soluble fraction eluted at 48.6 ml and between 84.0-107.0 ml. The peak eluting from both samples at approximately 49.0 ml corresponded to a His<sub>6</sub>-tagged PspA species with a molecular mass greater than 600 kDa. These have been shown to include the reported 36-meric ring structure and clathrin-like structures and may also include other higher-order His<sub>6</sub>-tagged PspA complexes.<sup>209,212</sup> The elution volumes were predicted for different species of His<sub>6</sub>-tagged PspA using molecular weight calibration standards (Figure 40C). The samples eluting between 83.3-109.0 ml corresponded to either dimeric (calculated at 81.9 ml) or monomeric (calculated at 90.3 ml) species. The presence of His<sub>6</sub>-tagged PspA in the peaks at approximately 49.0 ml and 83-110 ml was confirmed by Western blot analysis (Figure 40D and Figure 40E).



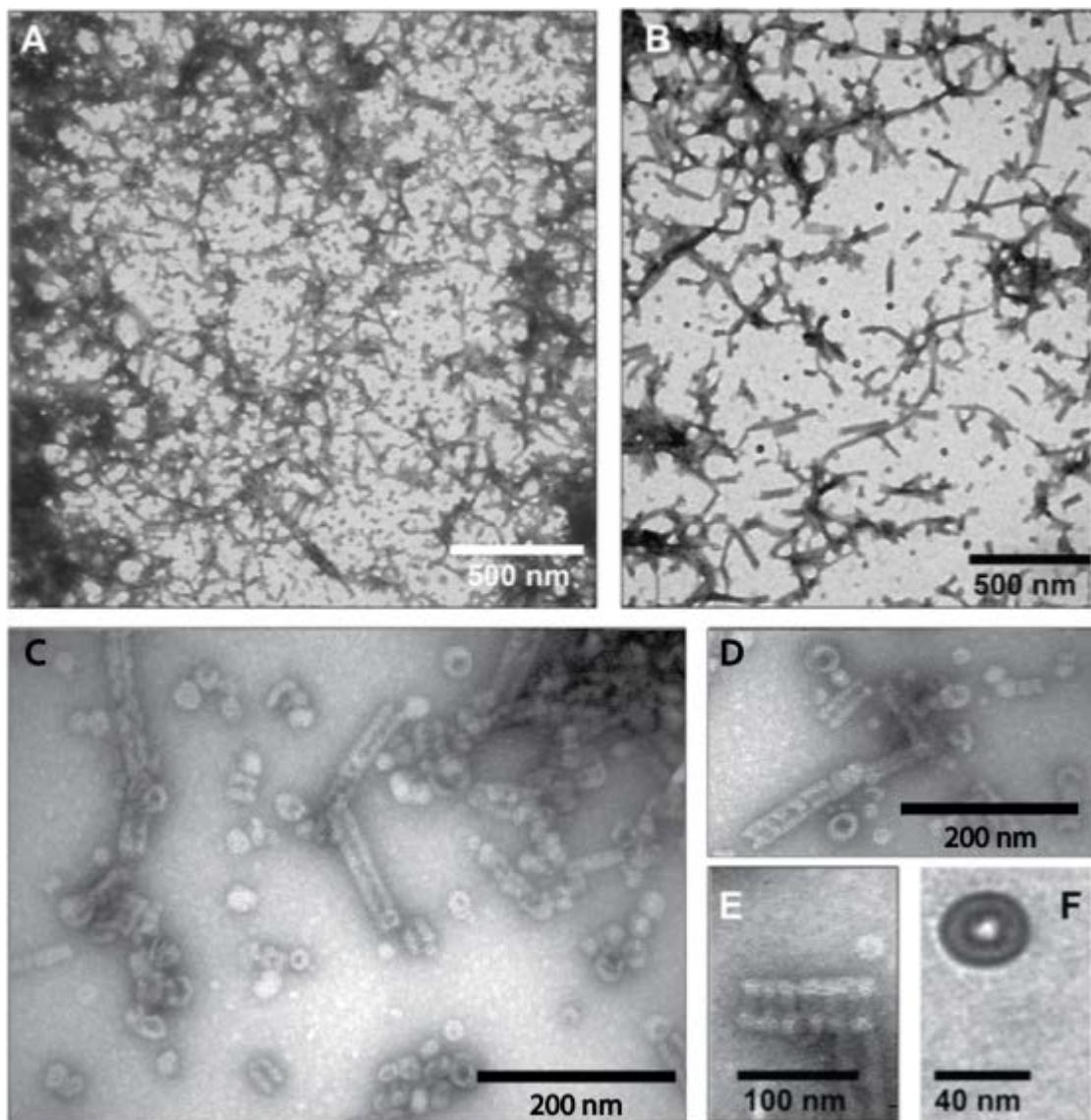
**Figure 40: SEC purification of His<sub>6</sub>-tagged PspA from the soluble and insoluble fraction.** (A) Chromatograms of *B. pseudomallei* His<sub>6</sub>-tagged PspA purified from the soluble. (B) Chromatograms of *B. pseudomallei* His<sub>6</sub>-tagged PspA purified from the insoluble. Fractions run on the Superdex 200 size exclusion column. Both samples have two sets of peaks, the first eluting approximately 49 ml corresponding to the void volume containing aggregated and putative higher-order structures of PspA. The second are clusters of peaks around 83-109 ml corresponding to lower-order structures of PspA (i.e. dimers and monomeric species). Samples were run in gel filtration buffer. (C) Calibration curve for the Superdex 200 size exclusion column (MWGF200, Sigma-Aldrich). (D) Western blots showing His<sub>6</sub>-tagged PspA was present in the peaks outlined above for soluble purifications. (E) Western blots showing His<sub>6</sub>-tagged PspA is present in the peaks outlined above for insoluble purifications His<sub>6</sub>-tagged protein was probed with 1:500 dilution of anti-His primary antibody and 1:10,000 anti-mouse-HRP secondary antibody.

### **3.3.3 Probing the structure of PspA**

Due to the elution of His<sub>6</sub>-tagged PspA in the void volume, it was necessary to verify that the protein was forming a functional structure and was not just aggregated and unfolded protein. The desalted His<sub>6</sub>-tagged PspA sample (from the void volume) was negatively stained with uranyl acetate and imaged by transmission electron microscopy. Figure 41 depicts His<sub>6</sub>-tagged PspA purified from the soluble fraction. Several species were observed: the first species were ring-shaped complexes, previously observed for *E. coli* PspA.<sup>209</sup> Here the outer diameter is approximately 40 nm: this is slightly larger than the ring structures previously observed for *E. coli* PspA (i.e. 20 nm) and is more in line with the size of the *E. coli* clathrin-like species observed (i.e. 30-40 nm). Despite their size they appear to be ring-like structures and not the clathrin-like structures (Figure 41F). This presumed hole in the centre has a diameter of approximately 10-12 nm, which is consistent with previous data.<sup>209</sup>

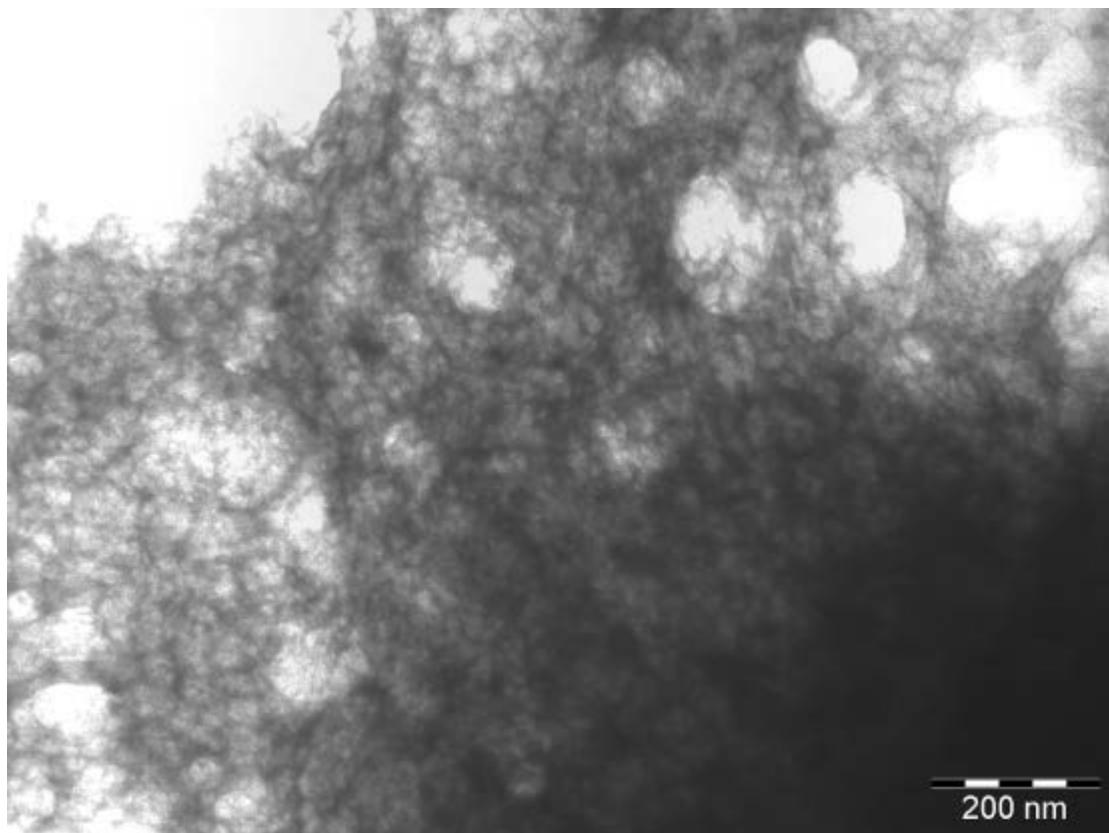
The second type of structure observed was a rod-like complex (Figure 41B-D): this is analogous to the rod-like complexes observed for the PspA homologue Vipp1.<sup>210</sup> A slight variation in the diameter between these rod-like complexes was observed, the diameter averaged between 40-45 nm, in agreement with the diameter observed here for the individual rings. Along each rod-like complex striations and indentations are visible, even at lower magnification (Figure 41E). These are uniform, despite the apparent differences in ring-size and are approximately 15 nm apart; similar striations and indentations are observed with Vipp1.<sup>210</sup> This suggests that it is the ring-like structures stacking on top of one another to form these rod-like species. These rod-like structures can be curved and have tapered ends (Figure 41D), suggesting that the rings do not necessarily stack directly on top of one another towards the end.

Furthermore, the rod-like structures have also been observed forming super-complexes, in the form of an ordered mesh-like structure (Figure 41A and Figure 41B). At lower magnification the large expansive network can be observed; clearly showing the interaction between these rod-like species. This has not previously been observed for PspA or Vipp1. Their physiological relevance needs to be further investigated.



**Figure 41: Electron micrographs of negatively stained *B. pseudomallei* His<sub>6</sub>-tagged PspA complexes.** (A and B) Mixtures of ring-like, rod-like structures and mesh-like structures are readily visible. Further assembly of these structures to a mesh-like scaffold may also be envisaged. (C and D) The ring and rod-like structures are visible in this field of view, with the rod-like structures showing the striations. (E) Close-up of a rod-like structure, clearly showing the indentations and striations that indicate stacking of ring-like structures. This is also an example of the tapered end observed occasionally. (F) A ring-like PspA putative 36-meric structure.

Within the sample unwanted protein aggregation was observed (Figure 42); however, this was minimal (approximately 2-5%) and gave a distinct unordered structure, different from the ordered mesh-like structures previously observed in Figure 41A and Figure 41B. The ordered nature of these rod-like structures indicates that the ring structures are not randomly interacting and these are not unwanted protein aggregation, but a structure that may have a physiological role.

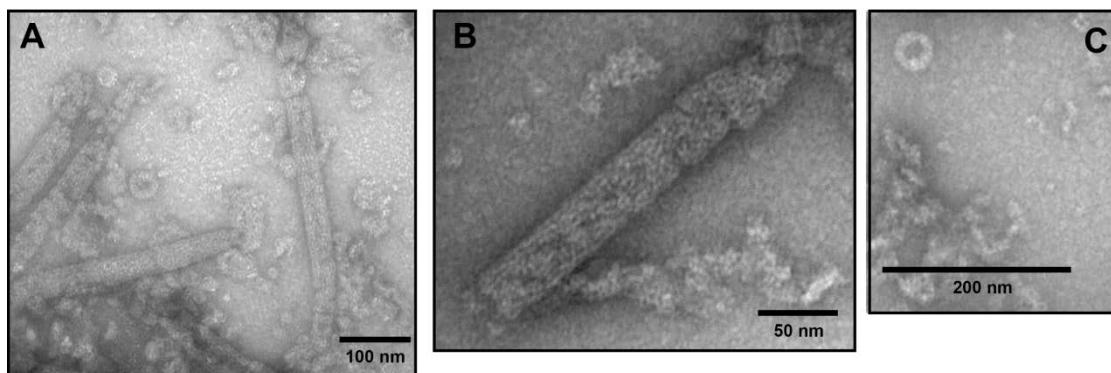


**Figure 42: Electron micrographs of negatively stained PspA aggregate.**

Standar *et al.* (2008) demonstrated how *E. coli* PspA can form higher-order structures (i.e. clathrin-structures) when purified from the soluble fraction in the absence of CHAPS detergent.<sup>212</sup> It was hypothesised that the lack of higher-ordered structures observed previously by Hankamer *et al.* (2004) was due to the presence of CHAPS detergent in the extraction buffer.<sup>209,212</sup> It was therefore necessary to confirm that the rod-shaped and mesh-like super-complexes observed with *B. pseudomallei* PspA were not an artefact of purification and the absence of CHAPS detergent. Therefore *B. pseudomallei* PspA was purified from the insoluble fraction with an extraction buffer supplemented with CHAPS detergent. Additionally

CHAPS was added to the eluted fractions after nickel affinity chromatography purification, as outlined by Elderkin *et al.* (2002).<sup>259</sup>

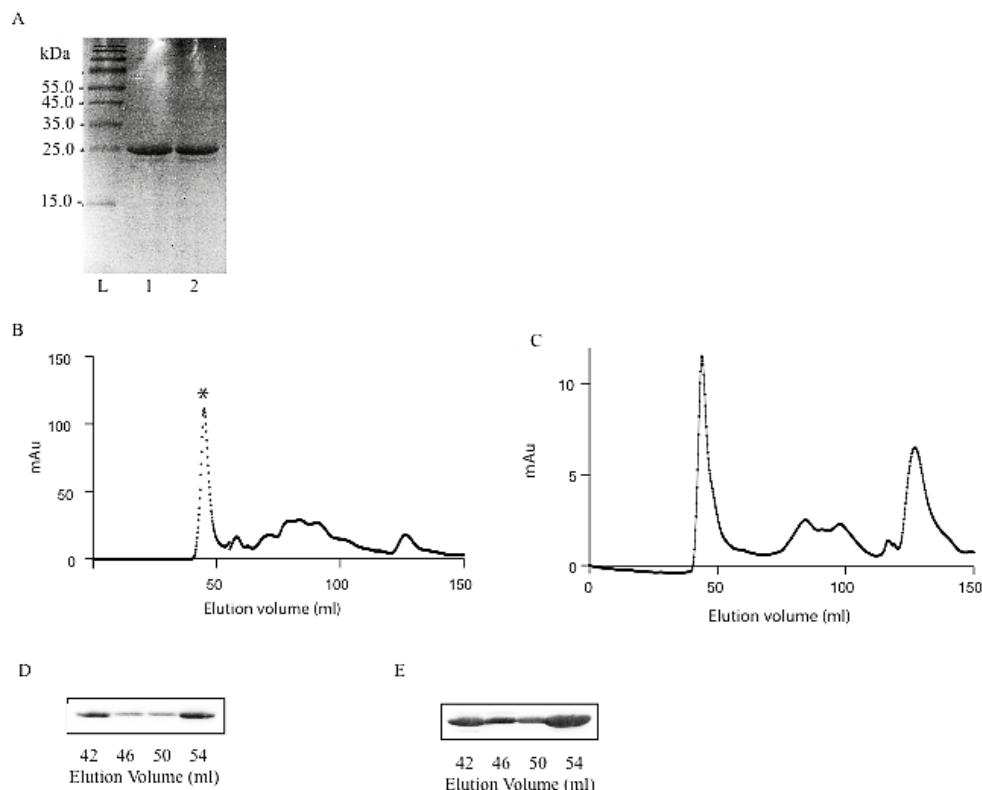
Similar higher-order species were observed for His<sub>6</sub>-tagged PspA purified from the insoluble fraction in the presence of detergent CHAPS (Figure 43). Ring-like structures with a diameter of 40-45 nm were observed (Figure 43C), alongside rod-like species between 100 and 500 nm long (Figure 43A and Figure 43B). However, unlike with His<sub>6</sub>-tagged PspA purified from the soluble fraction in the absence of CHAPS, no mesh-like structures were observed. Their absence does not conclude that in the presence of CHAPS they do not form; it could just be that the mesh-like structures were not in the field of vision of the sample visualised.



**Figure 43: Representative images taken of *B. pseudomallei* His<sub>6</sub>-tagged PspA purified in the presence of CHAPS from the insoluble fraction. (A)** Mixture of rod-like structures and ring-like structures. **(B)** Rod-like structure has tapered ends. **(C)** A solitary ring-like species.

The higher-order structures observed for *B. pseudomallei* PspA purified in the presence and absence of CHAPS from the insoluble and soluble fraction, respectively, were different from the previously observed structures for *E. coli* PspA.<sup>209,212</sup> To assess whether the rod-like and mesh-like structures observed for *B. pseudomallei* were organism specific, *E. coli* K12 PspA was cloned into the MCS of pET28a, using the restriction sites, *Bam*HI and *Sac*I. The *E. coli* His<sub>6</sub>-tagged PspA was then expressed and purified in the presence and absence of CHAPS (Figure 44). In the presence and absence of CHAPS, 4.7 mg and 6.4 mg of protein were purified, respectively. This was shown to be pure by SDS-PAGE analysis (Figure 44A).

Using SEC *E. coli* PspA was further purified showing similar traces to *B. pseudomallei*, whereby the majority of the protein was eluted in the void volume at approximately 44.0 ml (Figure 44B and Figure 44C).

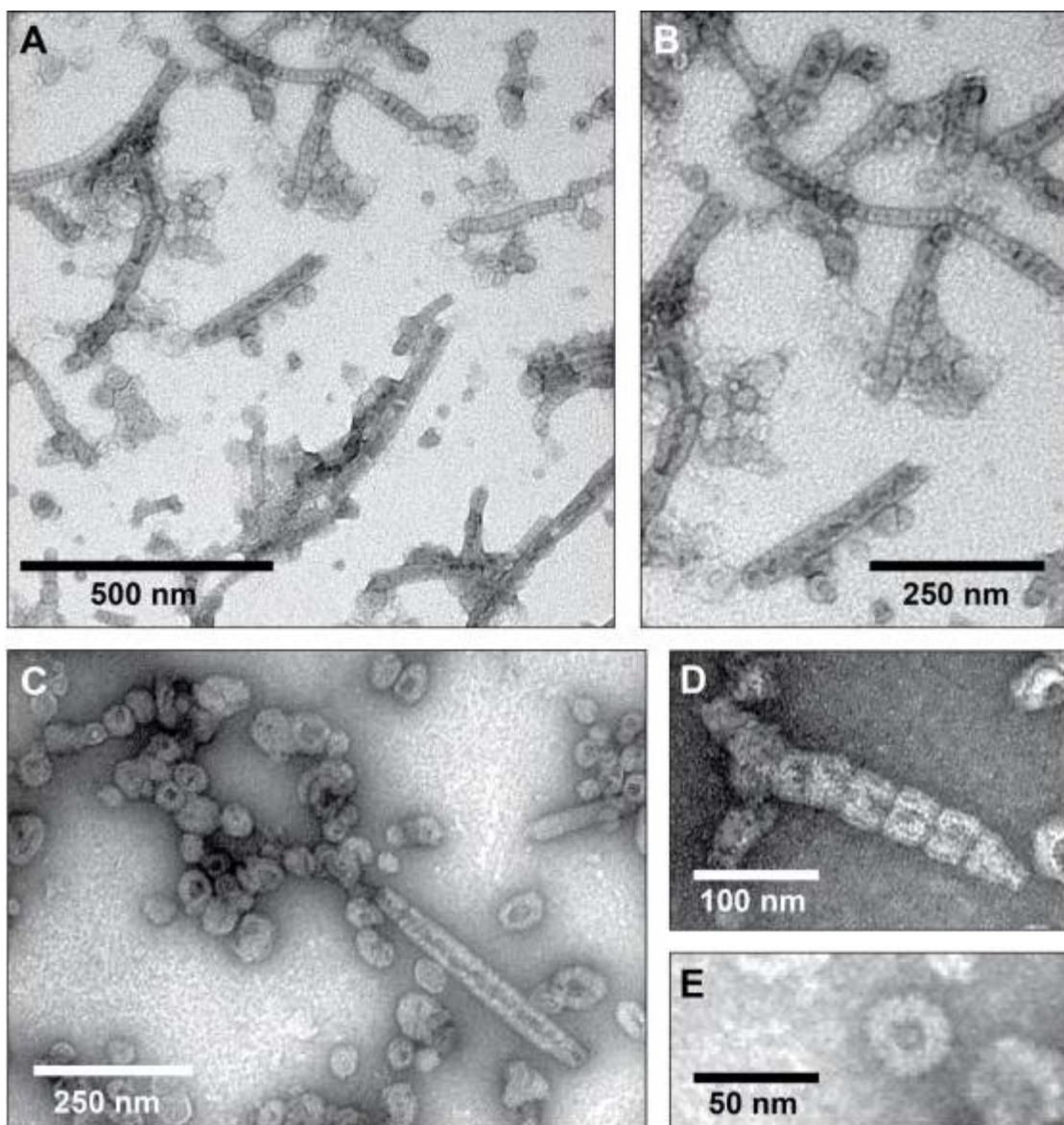


**Figure 44: Purification of *E. coli* His<sub>6</sub>-tagged PspA protein.** (A) 10% SDS-PAGE gel of *E. coli* His<sub>6</sub>-tagged PspA purified in the absence of CHAPS detergent and in the presence of CHAPS detergent. Lane L: Fisher Prestained PageRuler protein ladder 10-170 kDa; Lane 1: Purification in the absence of CHAPS; Lane 2: Purification in the presence of CHAPS. (B) Elution profile for *E. coli* His<sub>6</sub>-tagged PspA purified in the presence of CHAPS detergent. Peaks at 43.4 ml and a range between 67.3-108.7 ml correspond to higher- and lower-order complexes, respectively. (C) Elution profile for *E. coli* His<sub>6</sub>-tagged PspA purified in the absence of CHAPS detergent. Peaks at 44.3 ml and a range between 59.7-109.4 ml correspond to higher- and lower-order complexes, respectively. (D) Western blot of *E. coli* His<sub>6</sub>-tagged PspA in the oligomeric peak (\*) purified in the presence of CHAPS detergent. (E) Western blot of *E. coli* His<sub>6</sub>-tagged PspA in the oligomeric peak (\*) purified in the absence of CHAPS detergent. His<sub>6</sub>-tagged protein was probed with 1:500 dilution of anti-His primary antibody and 1:10,000 anti-mouse-HRP secondary antibody.

The void volume of each sample was visualised by transmission electron microscopy, showing similar rod-like and mesh-like structures observed for *B. pseudomallei* PspA (Figure 45). No difference was observed in the structures adopted by *E. coli* His<sub>6</sub>-tagged PspA purified in the presence or absence of CHAPS. Both samples showed a mixture of two distinct structures (Figure 45A); the first species were the previously observed ring-shaped complexes, which had been shown to correspond to a 36-meric species of PspA with nine-fold rotational symmetry.<sup>209</sup> The outer diameter of these ring-like structures observed was approximately 30 nm.<sup>209</sup> A ring of weak contrast with a black, stain-filled region in the middle was clearly observed, suggesting a low protein concentration and thus a hole in the centre with a diameter of approximately 7-10 nm, consistent with previously reported data.<sup>209</sup>

The second type of structure observed in both samples was rod-like (Figure 45A-Figure 45D), analogous to the rod-like complexes observed for Vipp1 and *B. pseudomallei* PspA.<sup>210</sup> The diameter of the Vipp1 rod-like complexes averaged between 30 and 40 nm, in agreement with the diameter of the rod-like structures observed here. The average length of the rods was around 500 nm, although significantly longer and shorter species were frequently observed (Figure 45B). Along each rod-like complex striations and indentations were visible (Figure 45D), even at lower magnification (Figure 45A). These were uniform and approximately 30 nm apart (Figure 45E), suggesting that they were composed of the super-assembly of the ring-like structures stacked on top of each other; similar striations and indentations are observed with Vipp1 and above for *B. pseudomallei* PspA.<sup>210</sup> These rod-like structures were observed to be both curved and straight and to have open or tapered ends (Figure 45B and Figure 45D).

Again it should be noted that neither of these species were aggregated protein. Within both samples protein aggregation was observed; however, this was minimal (approximately 5-10%) and gave a disordered mass that was distinct from the self-assembled species, similar to the aggregate protein observed for *B. pseudomallei* PspA in Figure 42.



**Figure 45: Electron micrographs of negatively stained *E. coli* His<sub>6</sub>-tagged PspA complexes.** (A) A mixture of ring-like and rod-like structures is readily visible. Further assembly of these structures to a mesh-like scaffold may also be envisaged. (B) A view of the image in (A) at higher magnification. The individual components of the rod-like structures are visible, as are clathrin-like structures previously reported. (C) Both types of structured assembly are visible in this field of view. (D) Close-up of a rod-like structure, clearly showing the indentations and striations that indicate stacking of ring-like structures. This is also an example of the tapered end observed occasionally. (E) A ring-like structure.

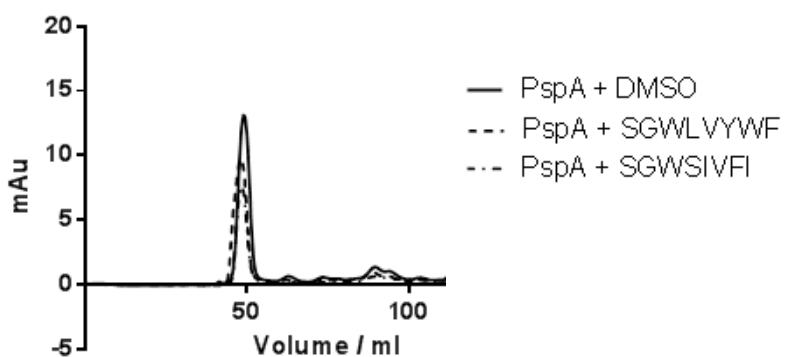
The super-assembly of *B. pseudomallei* and *E. coli* His<sub>6</sub>-tagged PspA protein into these ordered, rod-like structures indicates that the previously reported ring-like

structures are not the only forms of PspA. The data suggests that these ring-like structures further interact to form rod-like species that further assemble into a mesh-like super-complex, however, their physiological role in maintaining cell membrane integrity must be investigated. Having established that the protein being expressed forms an ordered structure described in the literature, studies into establishing an *in vitro* assay to test the inhibitors identified in the SICLOPPS screen were investigated.

### **3.3.4 Development of an *in vitro* assay for the validation of selected inhibitors**

#### **3.3.4.1 Size exclusion chromatography**

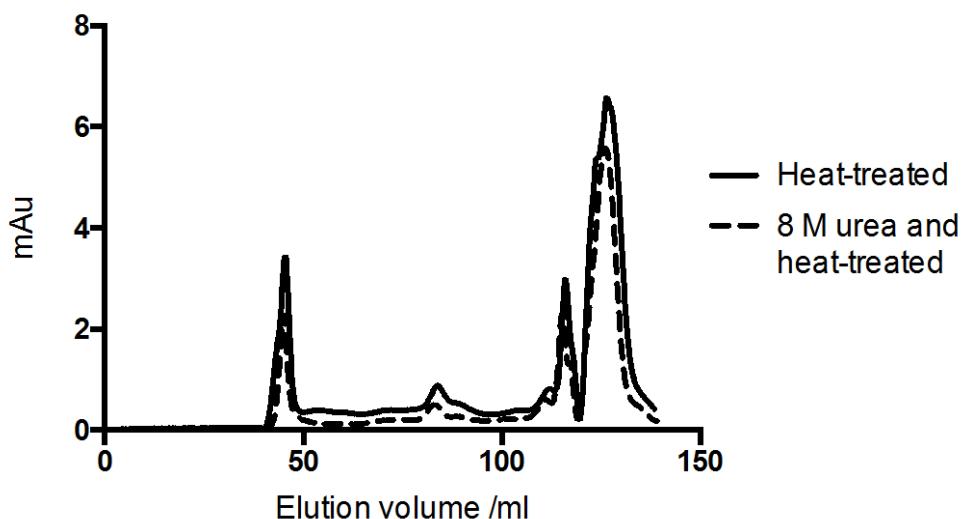
To evaluate the most potent inhibitors identified in the SICLOPPS screen (i.e. *cyclo*-SGWLVYWF and *cyclo*-SGWSIVFI) SEC was investigated for its viability as an assay to test the inhibitors further. An excess of each inhibitor was incubated with purified recombinant protein (1 mg/ml) prior to injection onto the Superdex 200 size exclusion column. A shift in the peak from the void volume (approximately 48 ml) to a higher elution volume (i.e. lower-order molecular weight peak, between 80-100 ml) would indicate a disruption of the PspA PPI. No shift was observed (Figure 46) regardless of the incubation time. The appearance of an additional peak at approximately 125 ml was the inhibitor eluting from the column in the post-column volume front due to the tryptophan absorbing at 280 nm. The presence of inhibitor in this peak was also confirmed by a control of peptide alone (trace not shown). The negative control was due to the low solubility of the inhibitors in water alone, therefore a control with 10% DMSO was analysed to ensure it had no effect on the oligomerisation of PspA.



**Figure 46: Chromatogram of His<sub>6</sub>-tagged PspA (1 mg/ml) incubated with inhibitor (2 mM) for 12 hours.** Analysed on the Superdex 200 size exclusion column equilibrated in gel filtration buffer. The chromatogram of PspA + DMSO showed a peak at 48 ml (the void volume), when the inhibitor was added no change in the chromatogram was observed.

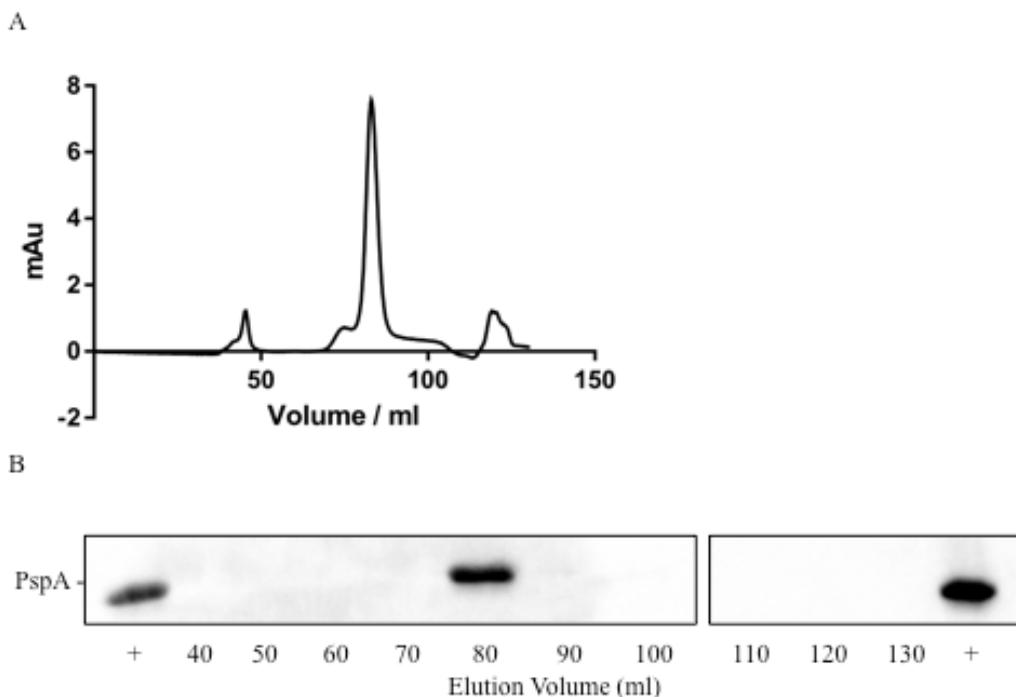
The absence of disruption observed was speculated to be due to the pre-formation of stable large higher-order structures, observed by electron microscopy that was not disrupted by the inhibitor. Therefore, it was necessary to establish experimental conditions where the inhibitor could be present during the assembly of these higher-order PspA structures. Consequently, denaturation and refolding the PspA protein was investigated, so during refolding the inhibitor could be present.

Initially, His<sub>6</sub>-tagged PspA was purified as previously described from the soluble fraction by nickel affinity chromatography. The purified protein was split into two and both samples were heated to 100°C for 10 minutes to initiate denaturation. In one sample urea was also added to a final concentration of 8 M to further denature the protein. Both samples were analysed by SEC. Neither sample showed evidence of PspA denaturation as the protein still eluted in the void volume (Figure 47).



**Figure 47: Chromatogram of His<sub>6</sub>-tagged PspA (1 mg/ml) purified and then heated to 100°C in the presence or absence of 8 M urea.** The proteins were run on a Superdex 200 size exclusion column equilibrated in gel filtration buffer. No significant change in chromatogram was observed.

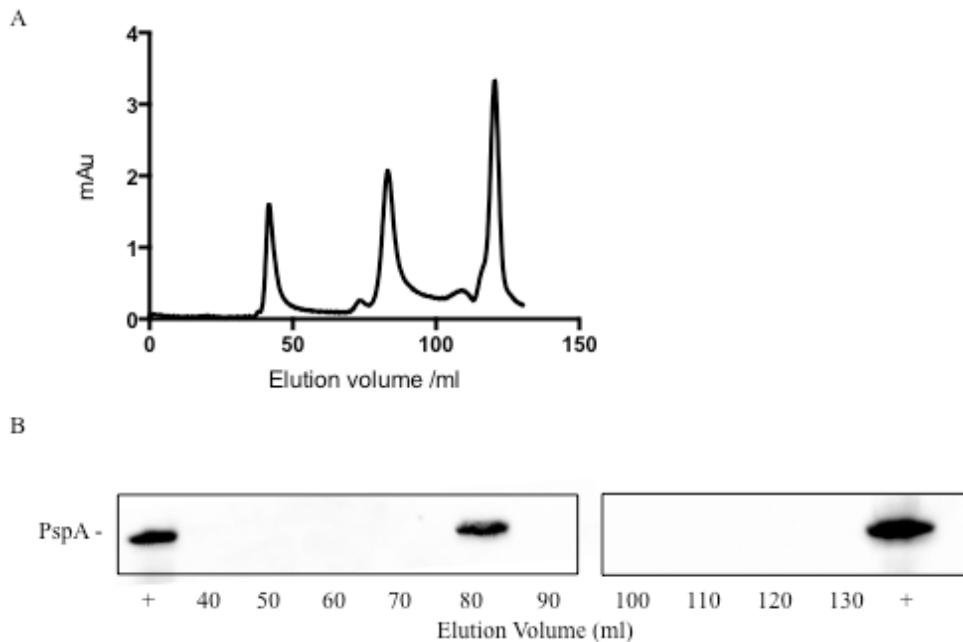
As the post-purification addition of urea did not denature His<sub>6</sub>-tagged PspA, purification in the presence of denaturing agents was attempted, followed by refolding using a urea gradient, with the protein immobilised on a nickel affinity column. After incubation the cells were lysed with a buffer containing 6 M guanidine hydrochloride and β-mercaptoethanol. The soluble fraction was then purified by nickel affinity chromatography with a wash buffer containing 8 M urea and 1 mM β-mercaptoethanol. The denatured His<sub>6</sub>-tagged PspA was isolated by eluting from the column with 500 mM imidazole. The sample was analysed further by SEC where a peak eluted at 83.1 ml, corresponding to monomeric His<sub>6</sub>-tagged PspA (calculated at 90.3 ml), its presence was confirmed by Western blot analysis (Figure 48). However, due to the high-level of denaturing agents the protein will not be in an active conformation.



**Figure 48: Denaturing studies of His<sub>6</sub>-tagged PspA.** (A) Chromatogram of His<sub>6</sub>-tagged PspA (1 mg/ml) purified in the presence of guanidine hydrochloride and β-mercaptoethanol. The proteins were run on a Superdex 200 size exclusion column equilibrated in gel filtration buffer. A shift from the previously observed void volume (48.0 ml) to 83.1 ml was observed. (B) Western blot analysis of the His<sub>6</sub>-tagged PspA detected in the peak corresponding to monomeric PspA (83.1 ml). His<sub>6</sub>-tagged protein was probed with 1:500 dilution of anti-His primary antibody and 1:10,000 anti-mouse-HRP secondary antibody.

A second experiment was carried out whereby the protein was refolded to form the higher-order structures previously observed by electron microscopy in the void volume peak: the aim of this was to investigate whether the technique was suitable to have the inhibitor present during refolding. His<sub>6</sub>-tagged PspA was purified as described above from the soluble fraction, using a lysis buffer containing the denaturing agents guanidine hydrochloride and β-mercaptoethanol. Once the denatured protein was bound to the nickel affinity column, the column was washed with a gradient of urea (8 to 0 M). This process encouraged refolding of the protein that can then be eluted from the column with imidazole. The SEC chromatogram showed refolding of the protein. However, this was incomplete. A peak at 83.0 ml was still observed indicating the presence of monomeric His<sub>6</sub>-tagged PspA,

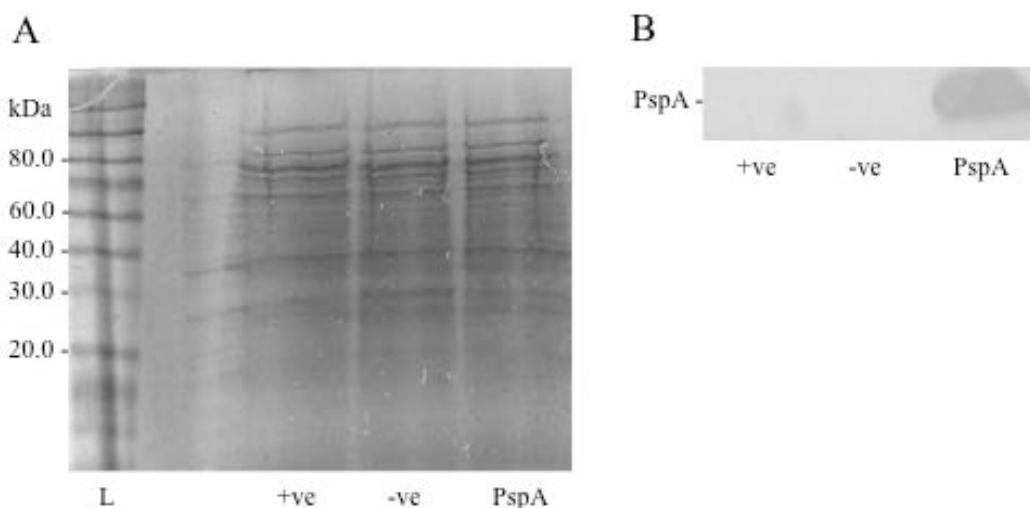
alongside a peak at 48.3 ml corresponding to the void volume, which would contain the higher-order structures of PspA or aggregate (Figure 49). The experiment was repeated using different volumes for the urea gradient ranging from 40-150 ml. However, no complete refolding was observed. The integrity of the void volume (i.e. the presence of ordered structures as opposed to aggregate) would have been investigated by electron microscopy along with circular dichroism to verify the secondary structure. However, due to the high volume of refolding buffer used for the inhibitor screening, the inhibitor would have needed to be present throughout, which at a 1 mM concentration would have been approximately 105 mg, over ten-times the yield from a standard synthesis. This was therefore not a viable assay. Consequently, a different method of testing the inhibitors was required that could be carried out on a smaller scale to reduce the amount of inhibitor required.



**Figure 49: Denaturing and refolding studies.** (A) Representative chromatogram of refolded His<sub>6</sub>-tagged PspA (1 mg/ml) purified in the presence of guanidine hydrochloride and β-mercaptoethanol and then refolded using a 100 ml urea gradient. The proteins were run on a Superdex 200 gel filtration column equilibrated in gel filtration buffer. A shift from the previously observed monomeric peak at 83.1 ml to the void volume at (48.0 ml) was observed. (B) Western blot analysis His<sub>6</sub>-tagged PspA was detected in the peak corresponding to monomeric PspA (83.1 ml). His<sub>6</sub>-tagged protein was probed with 1:500 dilution of anti-His primary antibody and 1:10,000 anti-mouse-HRP secondary antibody.

Cell-free expression of PspA was investigated as it is small-scale.<sup>263</sup> Therefore it could be a viable method of testing the inhibitors during the production of His<sub>6</sub>-tagged PspA before oligomers and aggregates have formed. The S30 T7 High-Yield Protein Expression System (Promega) was selected, as it is an *E. coli* extract-based cell-free protein synthesis system used in high-level expression of recombinant proteins within an hour. The system allows the transcription and translation of genes under the control of a T7 promoter, by the addition of an *E. coli* extract containing T7 RNA polymerase.<sup>264</sup> A vector with the *Renilla reniformis* luciferase gene downstream of a T7 promoter was used as a control. When expression from this control plasmid was detected in conjunction with a Dual Glo Luciferase Assay kit (Promega) it confirmed the activity of the system.

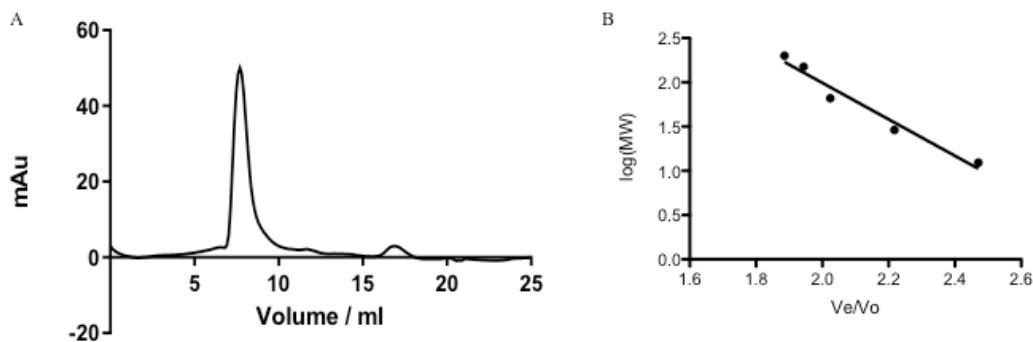
Due to the nature of the technique, analysis by SDS-PAGE was not possible due to the high level of background proteins (Figure 50A). Therefore, analysis by Western blot was required to confirm the expression of His<sub>6</sub>-tagged proteins (Figure 50B). A band with a mass of approximately 28 kDa was observed in samples that contained the pET28a-PspA plasmid suggesting that His<sub>6</sub>-tagged PspA protein was expressed successfully. In contrast, no His<sub>6</sub>-tagged proteins were detected by Western blot analysis in the positive control; instead luciferase activity was detected confirming successful expression.



**Figure 50: *In vitro* cell-free expression of His<sub>6</sub>-tagged PspA.** (A) 10% SDS-PAGE gel of His<sub>6</sub>-tagged PspA expressed using the S30 T7 High Yield Protein Expression System (Promega). *Lane L*: Colour Plus Protein Ladder Broad Range (NEB); *Lane +ve*: Positive control sample containing a control vector with the *R. reniformis* luciferase gene under the control of a T7 promoter; *Lane -ve*: Negative control containing no expression plasmid; *Lane PspA*: Sample containing pET28a-PspA. (B) Western blot of samples analysed in A by SDS-PAGE gel, a His<sub>6</sub>-tagged protein was detected only in the sample containing pET28a-PspA. His<sub>6</sub>-tagged protein was probed with 1:500 dilution of anti-His primary antibody and 1:10,000 anti-mouse-HRP secondary antibody.

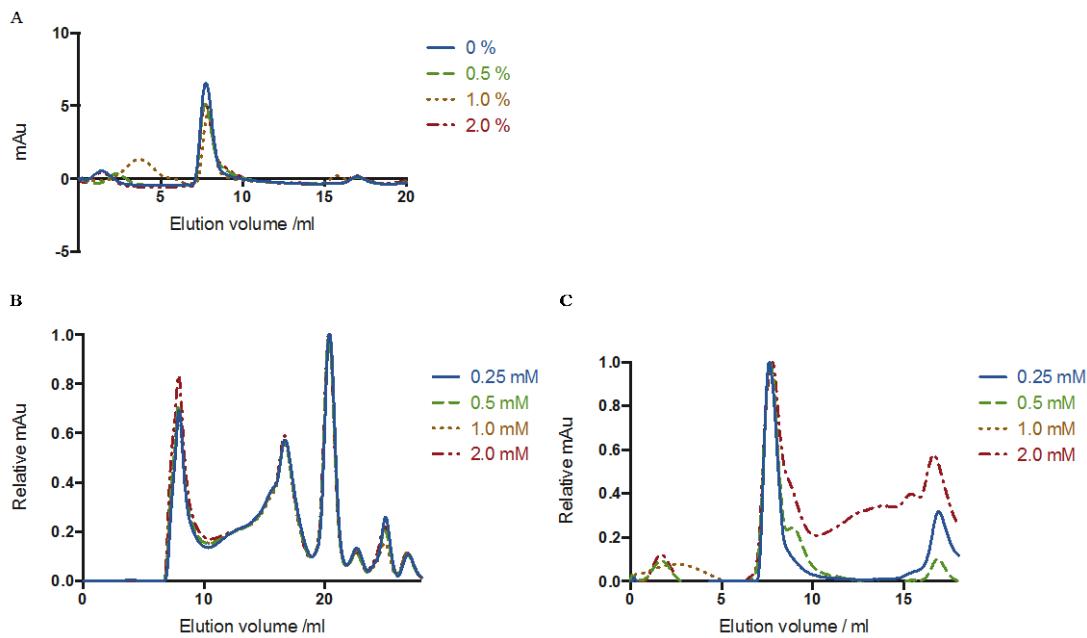
In order to visualise the formation of the different species of His<sub>6</sub>-tagged PspA, the samples were analysed by SEC. For subsequent experiments the samples were analysed on a Superose 6 size exclusion column. This column was selected as it has a greater range for protein fractionation (i.e. 10-5,000 kDa) and a smaller bed volume, allowing lower protein sample volumes to be used. Using this column the ring-like structures and the rod-like structures, meshes and aggregated protein should separate. The void volume for the column was expected to elute at 8 ml, containing the aggregated and rod-like species of PspA, while the putative 36-meric PspA ring-like structures were calculated to elute at 12 ml, estimated using the molecular weight calibration standards (Figure 51B). The hexameric, dimeric and monomeric structures of PspA were calculated to elute at 15.0 ml, 16.2 ml and 16.8 ml, respectively. Initially, to investigate where different species of His<sub>6</sub>-tagged PspA

eluted, a purified sample of bacterially-expressed His<sub>6</sub>-tagged PspA was injected onto the column, this eluted at 7.5 ml, which corresponds to the void volume (Figure 51A), indicating that His<sub>6</sub>-tagged PspA forms a structure with a mass greater than 5,000 kDa (i.e. aggregates and rod-like species). A small peak at 16.3 ml was also observed, predicted to be either dimeric or monomeric His<sub>6</sub>-tagged PspA.



**Figure 51: Analysis of His<sub>6</sub>-tagged PspA on a Superose 6 size exclusion column.** (A) Chromatogram of His<sub>6</sub>-tagged PspA (1 mg/ml), the void volume peak at 7.5 ml is visible along with a small peak at 16.3 ml corresponding to dimer or monomeric His<sub>6</sub>-tagged PspA. Samples run in gel filtration buffer. (B) Molecular weight standard calibration curve for the Superose 6 size exclusion column (MWGF200, Sigma-Aldrich).

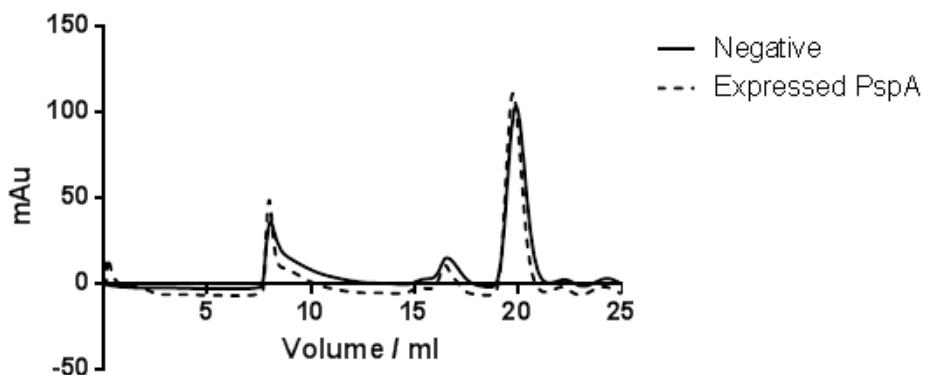
This column had the capacity to separate the ring-like species from the void volume containing the rod- and mesh-like species and aggregate. Therefore, modifications to the bacterial expression and purification were made to reduce the formation of these super-complexes. These included the addition of 10% glycerol, 2 M urea and CHAPS detergent during purification, none of which affected the chromatogram (exemplified using CHAPS in Figure 52A). Induction with varying concentrations of IPTG (i.e. 0.25-2.0 mM), different incubation times (i.e. 2 hours to 16 hours) and temperatures (i.e. 30°C and 18°C) were also tried. But no variation of the conditions altered the chromatogram (represented using different concentrations of IPTG during induction in Figure 52B and Figure 52C).



**Figure 52: Optimisation of His<sub>6</sub>-tagged PspA expression and purification.** (A) Chromatogram demonstrating how the presence of CHAPS detergent at different concentrations during purification had no effect on the high-order formation of His<sub>6</sub>-tagged PspA structure. (B and C) Varying the IPTG concentration used for induction had little influence over the expression of lower-order species. (B) is the crude lysate while (C) is the purified protein. All samples (1 mg/ml, 100 µl) were analysed on the Superose 6 size exclusion column equilibrated in gel filtration buffer. No significant change in chromatogram was observed.

Despite the absence of two peaks corresponding to the undesired aggregated product and the higher-ordered structures, a shift from oligomer to monomer should be evident and so the column was still used for the experiments described below to confirm the activity of the inhibitors.

Firstly, a sample containing expressed His<sub>6</sub>-tagged PspA and a negative control using the *in vitro* cell-free expression system were compared: no differences in the chromatograms were observed. This was hypothesised to be due to the lysate proteins, present in the expression system, masking the His<sub>6</sub>-tagged PspA (Figure 53).



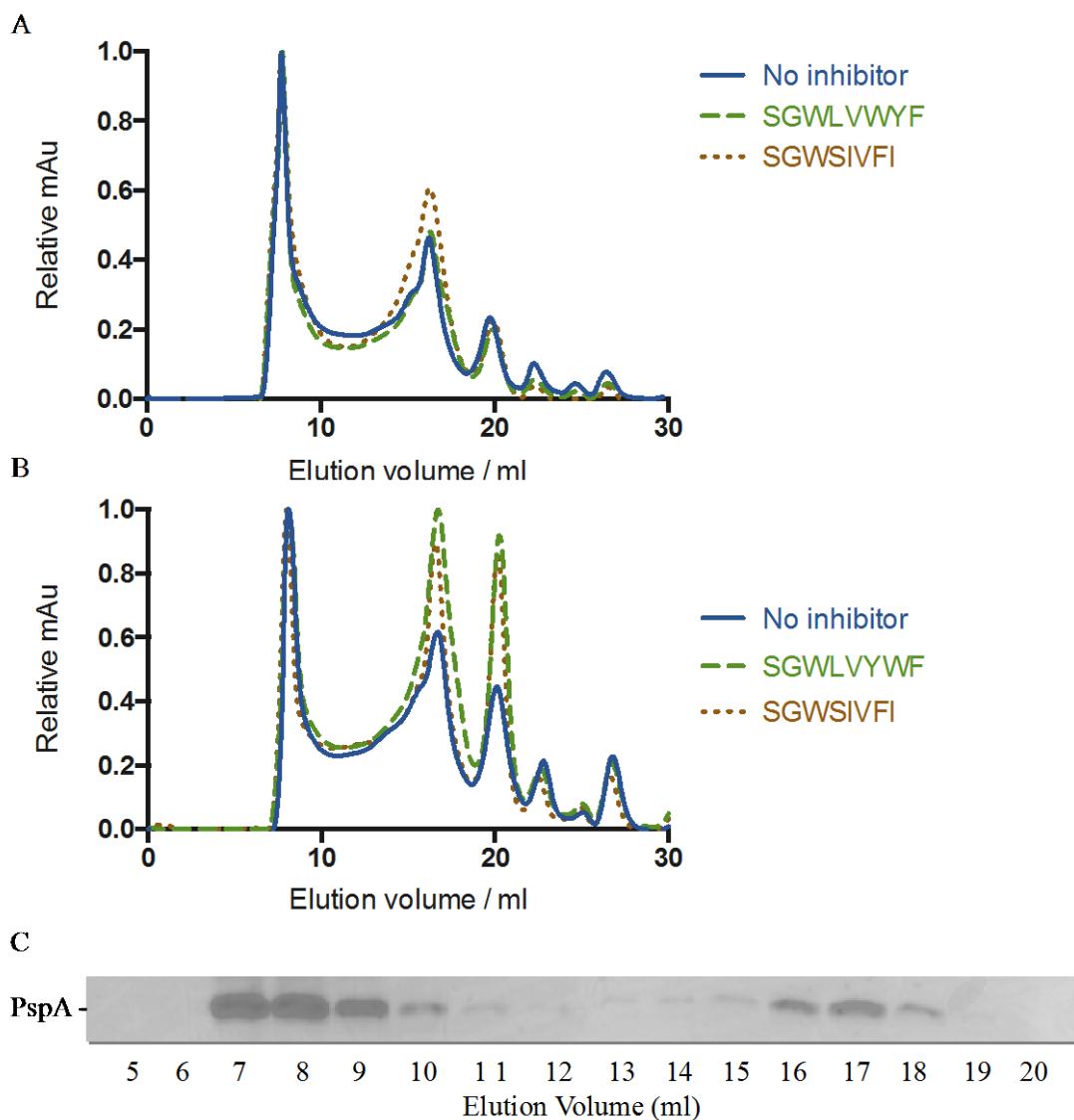
**Figure 53: His<sub>6</sub>-tagged PspA expressed by *in vitro* cell free expression.**

Chromatogram of the crude extracts from the *in vitro* protein expression kits. The negative control contained no plasmid containing a T7 promoter; in contrast the expressed PspA contained pET28a-PspA plasmid that can be used to express His<sub>6</sub>-tagged PspA. 100 µl of each sample was analysed on the Superose 6 size exclusion column equilibrated in gel filtration buffer. No significant change in chromatogram was observed.

In order to analyse the His<sub>6</sub>-tagged PspA protein species instead of the whole cell lysate, the samples were purified using nickel affinity chromatography and then analysed by SEC. Hypothetically, the cell-free purification kits yield 500 µg/ml; the manufacturer of the Superose 6 size exclusion column has demonstrated that carbonic anhydrase, with a molecular weight similar to His<sub>6</sub>-tagged PspA (i.e. 29 kDa), can be detected at a concentration of 3 mg/ml. Consequently, a protein with a concentration of 500 µg/ml should be detectable. However, as His<sub>6</sub>-tagged PspA has only one tyrosine and no tryptophan residues its signal at 280 nm will be low. Accordingly, the proteins were too dilute to be visualised this way, and so cell-free expression was deemed an unsuitable method in which to screen the inhibitors. Ideally, the protein samples could have been concentrated and analysed on a size exclusion column that requires a smaller sample volume, had one been available.

Utilising SEC with purified recombinant protein as a method of assessing the activity of the inhibitors was proving difficult. Consequently, co-expression of the inhibitors from the SICLOPPS plasmid in BL21 (DE3) expressing His<sub>6</sub>-tagged PspA in parallel was investigated. pET28a-PspA and pARCB (with the gene encoding

the relevant inhibitor flanked by the intein genes) were transformed into BL21 (DE3). The expression of the cyclic peptide was induced by arabinose, followed by the induction of His<sub>6</sub>-tagged PspA by IPTG. The culture was incubated overnight and the cells lysed. A sample of the soluble fraction was then analysed directly by SEC on a Superose 6 size exclusion column. The presence of His<sub>6</sub>-tagged PspA in the peaks at 7.9 ml and 16.9 ml was confirmed by Western blot analysis (Figure 54A-C).



**Figure 54: Co-expression of His<sub>6</sub>-tagged PspA and inhibitors. (A and B)**

Chromatograms of the soluble fraction from PspA expressions in the presence of the inhibitors *cyclo*-SGWLVYWF and *cyclo*-SGWSIVFI. The negative was a sample only expressing His<sub>6</sub>-tagged PspA with no inhibitor. The samples (100 µl) were run on a Superose 6 gel filtration column equilibrated in gel filtration buffer. (A) Shows

no significant change in chromatogram. (B) Shows an increase in lower-order species in the presence of inhibitor. (C) Western blot analysis of the fractions eluted in 61B. His<sub>6</sub>-tagged protein was probed with 1:500 dilution of anti-His primary antibody and 1:10,000 anti-mouse-HRP secondary antibody.

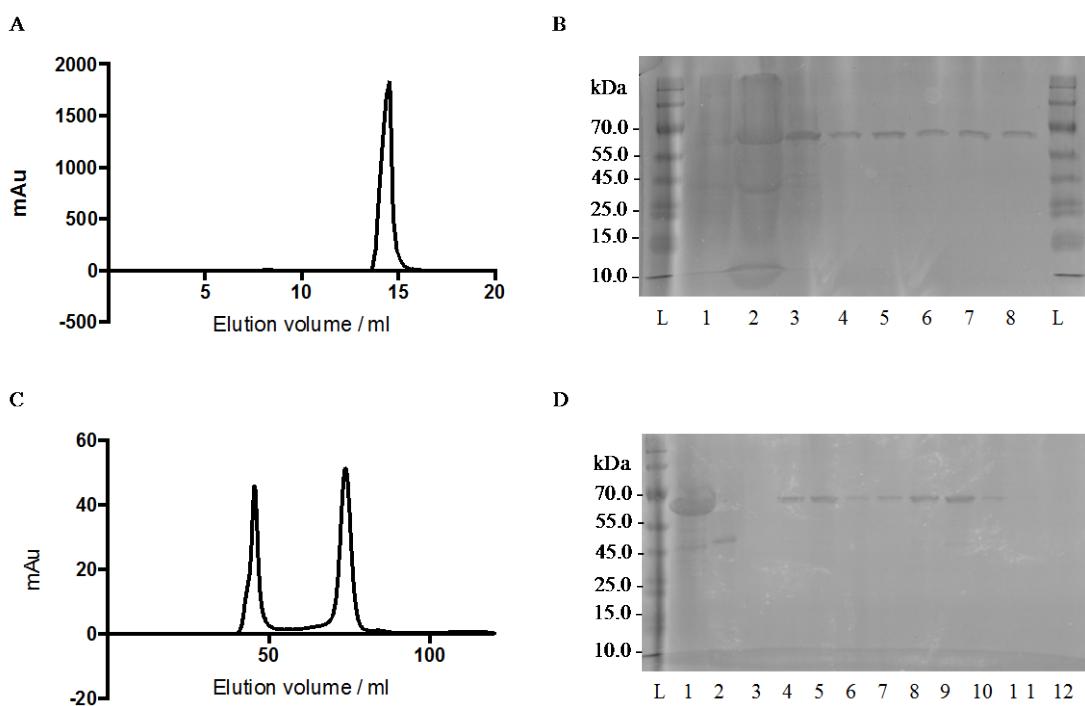
Figure 54A shows a chromatogram that suggests the inhibitor *cyclo*-SGWSIVFI was causing a shift from oligomer to the monomer His<sub>6</sub>-tagged PspA. Figure 54B shows two-fold more oligomeric species than monomeric for the negative control-expressing no inhibitors- compared to the inhibitors *cyclo*-SGWLVYWF (with a 5:4 ratio between oligomer and monomer species) and *cyclo*-SGWSIVFI (with a 1:1 ratio between oligomer and monomer species). Although this data shows a difference in the ratio between the peaks in the control and the samples in the presence of inhibitor, the repeatability of the experiments suggests that the differences in repeats could be due to different expression levels of inhibitor, background proteins and His<sub>6</sub>-tagged PspA. No conclusions can be drawn from this data and further optimisation and investigation would be essential. However, the Superpose 6 column was on loan from GE Healthcare and was returned before comprehensive optimisation of this assay could be completed. Also the nature of the experiment entailed loading crude lysate onto the column, which was not practically ideal. Therefore using SEC to test the efficacy of the inhibitors was deemed unsuitable due to the variability of results, methodology and the availability of suitable size exclusion columns; hence, SEC as a method of analysis was abandoned.

### *3.3.4.2 Isothermal titration calorimetry of PspA inhibitor binding*

Due to the inconclusive results showing how the inhibitors disrupt PspA by SEC, direct binding of the inhibitor to the protein was investigated. Purifying recombinant PspA with a His<sub>6</sub>-tag produced oligomeric PspA, as determined by SEC and electron microscopy. The complex nature of the oligomeric structures would not give an accurate representation of how the inhibitor binds to a PspA monomer. In order to probe the binding to monomeric PspA an MBP purification tag was used for expression. MBP is a commonly used protein expression tag that allows purification using amylose resin and elution with maltose; MBP improves the solubility of recombinant proteins and encourages the expression of monomeric proteins in the cytoplasm.<sup>265,266</sup> The *B. pseudomallei pspA* gene was therefore cloned into the MCS

of pMAL-c5x plasmid downstream from the *malE* gene of *E. coli*, which encodes the MBP. The cloning was carried out using the restriction sites *XmnI* and *BamHI*. *XmnI* is a blunt-ended restriction enzyme; therefore the *pspA* gene was amplified by PCR with a 5'-blunt end and a 3'-cohesive end that could be cleaved with *BamHI*. The construct plasmid will henceforth be referred to as pMAL-PspA.

The pMAL-PspA plasmid was transformed into chemically-competent NEB Express cells, a strain that lacks the T7 RNA polymerase and instead uses an *E. coli* RNA polymerase. MBP-tagged PspA was expressed in media containing glucose to repress the expression of amylase, which can affect binding to the amylose resin. Initially induction was attained with IPTG and expression at 37°C for two hours. The cells were lysed and purified on an amylose affinity column (Figure 55A); the sample was deemed pure by SDS-PAGE analysis (Figure 55B). SEC was used to further purify the sample; two peaks were observed (i.e. 45.3 ml and 73.5 ml) the latter confirmed the presence of monomeric MBP-tagged PspA (predicted to elute at 76.7 ml, Figure 55C) and the presence of MBP-tagged PspA protein was confirmed by SDS-PAGE analysis with a band at 67 kDa (Figure 55D). No MBP alone (predicted to elute at 83.6 ml, 42.5 kDa) was observed. Nonetheless, a large proportion of the protein was eluted in the void volume, suggesting undesired aggregation. In order to reduce aggregation expression at 18°C overnight was tried; however, no significant decrease in aggregation was observed. Consequently, subsequent expressions were carried out at 37°C for two hours.

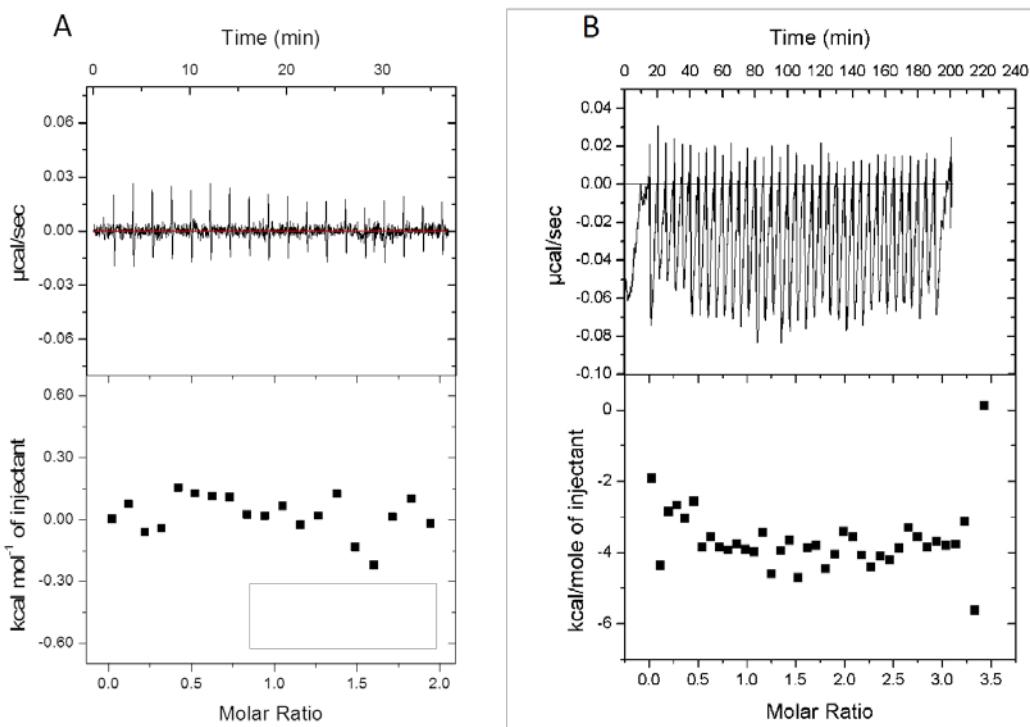


**Figure 55: Purification of MBP-tagged PspA.** (A) Chromatogram of MBP-tagged PspA purified on an amylose column, eluted between 13.5-15.5 ml. (B) 10% SDS-PAGE gel of the fractions of MBP-tagged PspA purified by amylose affinity column. *Lane L:* Fisher Prestained PageRuler Protein ladder (10-170 kDa); *Lane 1:* Insoluble fraction; *Lane 2:* Soluble fraction; *Lane 3:* Soluble purified pooled; *Lane 4:* 13.5-13.9 ml; *Lane 5:* 14.0-14.4 ml; *Lane 6:* 14.5-14.9 ml; *Lane 7:* 15.0-15.4 ml; *Lane 8:* 15.5-16.0 ml. (C) Chromatogram of MBP-tagged PspA purified on the Superdex 200 size exclusion column, two peaks present containing aggregated MBP-tagged PspA (73.5 ml) and monomeric MBP-tagged PspA (83.6 ml). Sample run in gel filtration buffer. (D) 10% SDS-PAGE gel of the fractions of MBP-tagged PspA (67 kDa) purified by SEC. *Lane L:* Fisher Prestained PageRuler Protein ladder (10-170 kDa); *Lane 1:* Purified MBP-tagged PspA; *Lane 2:* MBP protein alone; *Lane 3:* 35-40 ml; *Lane 4:* 40-45 ml; *Lane 5:* 45-50 ml; *Lane 6:* 50-55 ml; *Lane 7:* 55-70 ml; *Lane 8:* 70-75 ml; *Lane 9:* 75-80 ml; *Lane 10:* 80-85 ml; *Lane 11:* 85-90 ml; *Lane 12:* 90 + ml.

The binding of each inhibitor to MBP-tagged PspA was subsequently investigated by ITC. ITC measures the heat changes associated with binding relative to a reference cell, when inhibitor is injected into the protein sample. This technique has been extensively reviewed.<sup>267,268</sup> Using ITC the thermodynamic parameters of an

interaction can be obtained (i.e. binding affinity ( $K_D$ ), stoichiometry (n), enthalpy ( $\Delta H$ ) and entropy ( $\Delta S$ )).

The inhibitors were dissolved in 10% DMSO in aqueous solution due to their poor solubility. In order to account for the variability in DMSO concentration due to its hygroscopic nature, DMSO titrations were carried out to ensure complete buffer match. By titrating in the inhibitors into buffer containing different concentrations of DMSO (i.e. 0.7, 0.9 and 1.0%), a buffer match at 0.9% DMSO was observed and thus used in all subsequent experiments (Appendix 3). For ITC measurements MBP-tagged PspA was loaded into the measurement cell, and the inhibitor (i.e. *cyclo*-SGWLVYWF or *cyclo*-SGWSIVFI) was added in small increments under continuous stirring. The concentration of inhibitor used was 500  $\mu\text{M}$ , a ten-fold excess of inhibitor to protein (50  $\mu\text{M}$ ). No binding for either inhibitor was observed (Figure 56A and Figure 56B).



**Figure 56: Graphs of ITC of MBP-tagged PspA with inhibitors.** (A) *Cyclo*-SGWLVYWF (500  $\mu\text{M}$ ) was titrated with MBP-tagged PspA (50  $\mu\text{M}$ ) at 25°C, no binding was observed. (B) *Cyclo*-SGWSIVFI (500  $\mu\text{M}$ ) was titrated with MBP-tagged PspA (50  $\mu\text{M}$ ) at 25°C, no binding was observed.

### 3.3.4.3 Microscale Thermophoresis of PspA inhibitor binding

No binding was observed for either inhibitor to MBP-tagged PspA by ITC. Binding of the inhibitor could be weak and outside of the range of  $K_d$  obtainable from the experimental parameters used above (i.e. the protein concentration may not have been sufficient). Consequently, a technique whereby the experimental parameters had less of an effect on the accessible  $K_d$  was desired. Two options were available SPR and MST. Both techniques can determine affinities; although, MST does not require immobilisation to a chip like SPR does. In addition, MST can detect a wider range of affinities from pM to mM, not practical by ITC, due to the requirement of protein concentration. MST was therefore used to investigate the binding of the inhibitors *cyclo*-SGWLVYWF and *cyclo*-SGWSIVFI to MBP-tagged PspA.

MST detects a change in the hydration shell, conformation and charge or size of proteins by measuring changes in the protein's mobility in response to microscopic temperature gradients.<sup>269</sup> Thermophoresis is defined as the movement of molecules along a temperature gradient. Protein samples that are incubated with varying concentrations of inhibitor are placed in capillaries; to each capillary a microscopic temperature gradient is generated using an IR laser. This heats an area of water in the buffer solution by 2-6 K whilst the periphery maintains its temperature. In response to this temperature gradient the protein molecules move from the heated area to the cooler periphery, decreasing the density of molecules in the heated zone. Movement occurs by diffusion until a steady state is reached after about 10-30 seconds. This local change in protein concentration in relation to the temperature difference can be quantified using the Soret coefficient ( $S_T$ ) in Equation 5.

$$c_{\text{hot}}/c_{\text{cold}} = \exp(-S_T \Delta T) \approx 1 - S_T \Delta T \quad (\text{Equation 5})$$

The protein is fluorescently labelled and the normalised fluorescence  $F_{\text{norm}}$  is a measure of the Soret coefficient plus the fluorescence change due to the fluorophores temperature dependence ( $\delta F/\delta T$ ). An approximation for this is given in Equation 6.

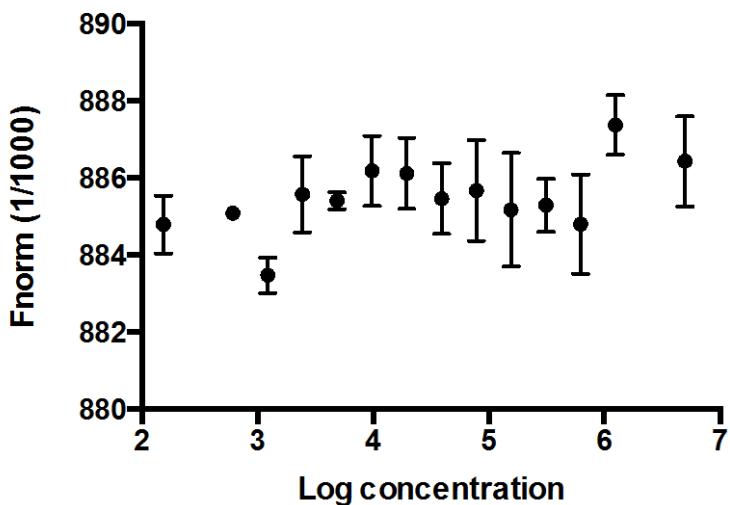
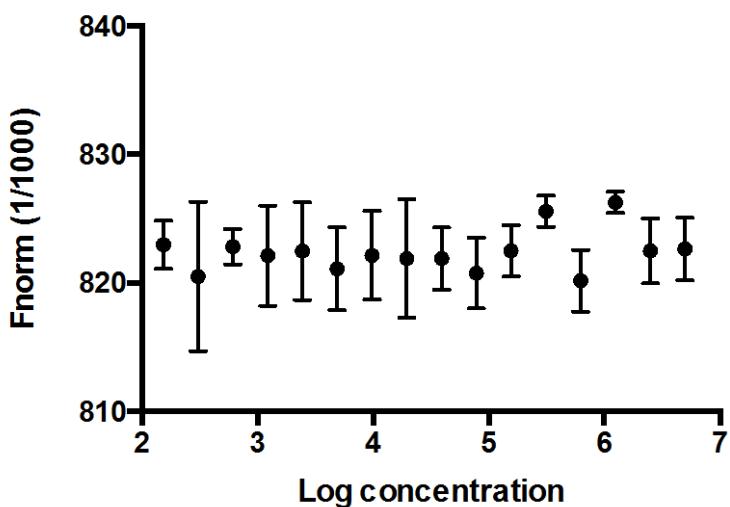
$$F_{\text{norm}} = F_{\text{hot}} / F_{\text{cold}} = 1 - (S_T - \delta F/\delta T) \Delta T \quad (\text{Equation 6})$$

A range of inhibitor concentrations are used, in conjunction with a constant protein concentration.  $F_{\text{norm}}$  changes as a function of the inhibitor concentration, according to Equation 7. Where  $F(A)_{\text{norm}}$  is the contribution of the unbound fluorescent molecule A and  $F(AT)_{\text{norm}}$  is the contribution of the complex of the fluorescent protein (A) and its interacting titrant (T). x is the fraction of fluorescent molecules that formed the complex.

$$F_{\text{norm}} = (1-x)F(A)_{\text{norm}} + xF(AT)_{\text{norm}} \quad (\text{Equation 7})$$

As the concentration of inhibitor is increased and incubated with a constant concentration of the fluorescent protein, the proportion of complexes formed increases until all the fluorescent protein has formed a complex with the inhibitor. From Equation 7 the fraction of bound protein (x) can be calculated from the change in  $F_{\text{norm}}$ . The normalised fluorescence is plotted against time in seconds.

MBP-tagged PspA was coupled to a fluorescent tag (NT-647) via the amine bond of the lysine residues; this was incubated with increasing concentrations of either *cyclo*-SGWLVYWF or *cyclo*-SGWSIVFI. The concentrations of inhibitor titrated into samples of protein ranged from 5 mM to 152.6 nM. No binding was observed for either inhibitor (Figure 57).

**A****B**

**Figure 57: MST of inhibitors against MBP-tagged PspA. (A) Cyclo-SGWLVYWF. (B) Cyclo-SGWSIVFI.** The MBP-tagged PspA was labelled covalently with NT-647 via the amine of the lysine residues. In both MST experiments (A) and (B) the concentration of MBP-tagged PspA remained constant (approximately 50 nM), while the concentration of the respective non-labelled inhibitor; either *cyclo*-SGWLVYWF (A) or *cyclo*-SGWSIVFI (B) was varied between 5 mM and 152 nM. The assay was carried out in MST-optimised buffer. No binding was observed.

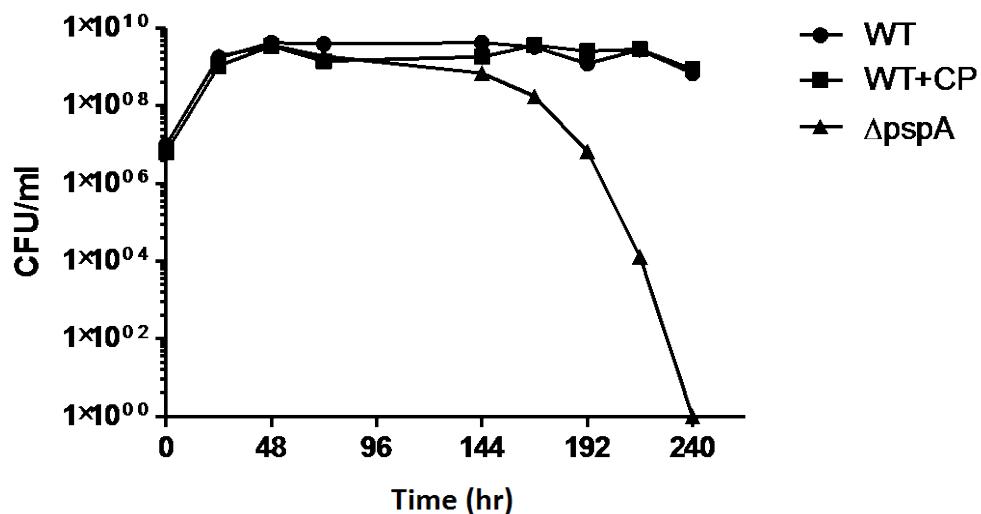
### 3.3.4.4 Testing cyclo-SGWLVYWF in *B. pseudomallei* using SICLOPPS

Using an *in vitro* assay the activity of the inhibitors could not be confirmed; therefore, the inhibitors were tested *in vivo*. This was achieved by constructing a plasmid containing codon-optimised inteins, to allow for optimal expression in *B. pseudomallei*, flanking the inhibitor sequence. Tests were carried out in partnership with collaborators at DSTL, by investigating the effect of the inhibitor on growth in stationary phase.<sup>227</sup>

First, the intein sequences had to be codon-optimised for *B. pseudomallei*. This was because *B. pseudomallei* is a GC rich pathogen (68.5%).<sup>270</sup> Using the JAVA codon optimisation tool, the sequence,<sup>271</sup> which started with a GC content of 45.5% was enhanced to 66.6% more in line with the desired GC content for *B. pseudomallei*. The gene encoding the optimised inteins flanking the peptide sequence (SGWLVYWF) was synthesised by Eurofins to incorporate the restriction sites, *Spe*I and *Eco*RV. These restriction sites allowed cloning of the sequence into pBHR4, a plasmid based on the broad host range plasmid, pBHR1.<sup>272</sup> The plasmid pBHR4-SGWLVYWF was constructed by standard cloning techniques and verified by DNA sequencing.

The pBHR4-SGWLVYWF plasmid was conjugated with *B. pseudomallei* K96243 by S. Southern (DSTL) who then compared the growth of this strain to the growth of the wild-type and a *pspA* null mutant strain in a stationary phase growth assay. It has previously been shown for *E. coli* that PspA is essential for survival in stationary phase in alkaline pH. More recently our collaborators at DSTL have demonstrated that the same phenotype is observed for *B. pseudomallei*.<sup>227</sup> To summarise the cultures were incubated at 37°C for ten days and every 24 hours a sample was analysed for cell viability. The anticipated results were that the absence of PspA would cause an observable growth defect after six days, while the wild-type would survive at stationary phase to the end of the experiment. In the presence of the inhibitor the PspA oligomerisation would be inhibited causing a growth defect similar to that observed for the null mutant, if the inhibitor was active.

As expected the *B. pseudomallei* *pspA* null mutant strain maintained its previously observed phenotype of a growth disadvantage after six days (Figure 58). The wild-type and wild-type in the presence of the inhibitor showed no growth defect, suggesting that the inhibitor was inactive in this assay.



**Figure 58: Comparison of growth of *B. pseudomallei* K96243 wild-type, a *pspA* null mutant and wild-type conjugated with pBRH4-SGWLVYWF. A growth defect was observed after six days in the *pspA* null mutant, whereas the wild-type in the presence and absence of the inhibitor showed no growth defect.**

### **3.4 Summary and discussion for Chapter 3**

The Psp response is involved in maintaining cell membrane integrity during stress conditions.<sup>142</sup> PspA oligomerisation is essential for this in *E. coli* and *Y. enterocolitica*.<sup>214</sup> Chapter 3 describes the construction of a PspA RTHS and its use in identifying cyclic peptide inhibitors of *B. pseudomallei* PspA oligomerisation.

Firstly, a PspA RTHS was constructed by integrating the *434-pspA* gene into the chromosome of SNS118.<sup>80,231</sup> The main issue during construction was the integration into the chromosome of SNS118. The incorporation of a brief heat shock step (42°C) to the protocol by Haldimann *et al.* (2001) did not improve the integration efficiency;<sup>231</sup> similarly varying the length of time the cells were incubated at each temperature did not allow integration. This difficulty in integrating the pAH68 plasmid encoding the 434/P22-recombinant proteins was also observed during the integration of the heterodimer HIF-1α and HIF-1β and the homodimer CtBP. It is hypothesised that the variation in the success of integration is due to the size of the insert.<sup>133,134</sup> In all cases the preparation of the electro-competent cells was targeted by incubating the cells at 42°C for a short period of time (e.g. 20 minutes) during the preparation, a modification described by Baba *et al.* (2006).<sup>232,236</sup> This preparation of competent cells allowed integration of the *434-pspA* gene into the chromosome of SNS118. The expression of int<sub>HK022</sub> phage integrase is mediated by the ci857 repressor, which is denatured at these higher temperatures.<sup>273</sup> This additional heat-shock step, in theory, allowed a greater expression of the heat-sensitive int<sub>HK022</sub> phage integrase protein, owing to the cells being incubated at a temperature closer to the active range of int<sub>HK022</sub> phage integrase proteins.<sup>231</sup>

As observed with other PPIs probed using the RTHS,<sup>126,127,131-134</sup> the verification of the PspA PPI by drop-spotting onto selective minimal media showed that growth retardation of *E. coli* occurred in response to increased concentrations of IPTG, and when cells were grown under more stringent conditions, relative to the less stringent conditions. It is postulated that this was because increasing IPTG concentration further induces 434-PspA expression and increased dimerisation. This leads to an increase in binding of the 434-PspA recombinant proteins to the 434-DNA-binding domains in the chromosome, thereby obstructing the binding of the RNA polymerase and preventing expression of the reporter-construct. The effects observed were

confirmed to be due to the PPI and not protein toxicity by drop-spotting onto LB agar at varying IPTG concentrations where no inhibition of growth was observed.

The ONPG assay was used to quantify the PPI. Similar to the positive control, as IPTG levels increased, so did the levels of protein dimerisation within the PspA RTHS.<sup>131</sup> This dimerisation caused the transcription of the *lacZ* gene to be inhibited leading to a reduction in the levels of β-galactosidase. Consequently, less ONPG was hydrolysed into ONP and the colour intensity was lower, demonstrated by the decrease in the calculated relative Miller indices. Predictably, the system became saturated at higher levels of IPTG so no change in activity was observed. Similar saturation was observed by Miranda *et al.* (2011) at approximately 18% activity, with the heterodimeric PPI of the influenza viral protein NS1 and the N-terminal domain of its binding partner, RIG-1.<sup>274</sup> The negative control remained constant regardless of IPTG concentration; in line with previous findings whereby a residual β-galactosidase synthesis of 5% was observed for the expression of just the heterodimeric repressors, P22 and 434.<sup>94</sup> In order to define when a PPI had occurred, Di Lallo *et al.* (2001) set the limit of interaction at 50% activity repression.<sup>93</sup> Both PspA RTHS samples showed a decrease of 60% confirming that the proteins were interacting. This is greater repression than observed in the CtBP1 and CtBP2 homodimeric RTHS', whereby only 50% activity repression was detected; however, it is less than the repression observed for the HIF-1α and HIF-1β heterodimerisation (i.e. 80%).<sup>134</sup>

Combining the results from the drop-spotting and the ONPG assays, it was concluded that the SICLOPPS screening should be carried out using 50 μM IPTG, 5.0 mM 3-AT and 50 μg/ml kanamycin. These conditions are more stringent than those used in the screening of the heterodimeric p6-UEV system (i.e. 30 μM IPTG, 25 μg/ml kanamycin and 2.5 mM 3-AT) and the homodimeric ATIC system (i.e. 50 μM IPTG, 25 μg/ml kanamycin and 2.5 mM 3-AT).<sup>126,131</sup> The conditions, however, are less stringent than those used for the ribonucleotide reductase selection (i.e. 200 μM IPTG, 50 μg/ml kanamycin and 2.5 mM 3-AT),<sup>80</sup> and HIF-1α and HIF-1β dimerisation (i.e. 100 μM IPTG, 50 μg/ml kanamycin and 7.5 mM 3-AT).<sup>134</sup> These differences observed with the ONPG assay and drop-spotting assay may not be

primarily due to the strength of the binding interaction. The dimerisation of the proteins brings the two repressor subunits into contact, and it may be that the orientation of the p6-UEV and ATIC subunits in dimerisation are more advantageous than those produced with the PspA or the ribonucleotide reductase interaction. Little is known about how PspA subunits interact, although the higher-order species have been visualised by electron microscopy.<sup>209,212</sup> Standar *et al.* (2008) proposed that head-to-tail dimers of PspA form the minimal subunit for the clathrin-like scaffold.<sup>212</sup> Therefore, the PspA interaction could be anti-parallel, suggesting that the N-terminal fused proteins used in the PspA RTHS may not have the optimal orientation, hindering the PPI within the RTHS. To overcome this future work could include the construction of an N- and C-terminal fused-PspA RTHS to probe different orientations.<sup>212</sup> In addition the variations in IPTG concentration used during the ONPG and drop-spotting assays may also be attributed to the levels of protein expression. Here, PspA is codon-optimised for *B. pseudomallei*, therefore this may affect the levels of protein expression occurring in the *E. coli* SNS118. Similarly, the expression of mammalian proteins, CtBP and HIF, may also be affected. This would lead to the same IPTG concentration producing different levels of protein.

Having established that the recombinant PspA proteins dimerise within the RTHS, the system was utilised for inhibitor screening. Four libraries were constructed (i.e. CX<sub>5</sub>, SGWX<sub>5</sub>, SX<sub>5</sub>G and SX<sub>6</sub>G). The libraries each contained 3.2-7.2 million unique sequences, thus, transformation efficiency had to be optimised to transform the entire library. The SGWX<sub>5</sub> library was the most successful library with six of the seven most active peptides being isolated from this library: the last peptide sequence was from the SX<sub>5</sub>G library. This may be due to the size and conformational restraint of the cyclic peptide isolated from each library. For example, the CX<sub>5</sub> peptide will be slightly smaller and more constrained. Due to the large surface areas involved in PPIs a larger peptide may be required.<sup>70</sup> Birts *et al.* (2013) screened three SICLOPPS libraries of different lengths against CtBP dimerisation (i.e. SGWX<sub>4</sub>, SGWX<sub>5</sub> and SGWX<sub>6</sub>). As observed here the top sequences were from a single library, the SGWX<sub>6</sub> library, suggesting that all the inhibitors are targeting the same structural features.<sup>133</sup> Similarly, a combinatorial screen of four libraries of cyclic peptides ranging from 3-6 amino acids for K-Ras function, yielded 20 potential sequences, 18 of which were from the six amino acid library.<sup>275</sup> This dominance of one particular

library could be due to the way in which the PspA subunits interact. Little is known about how PspA oligomerisation occurs; however, it is dependent on the presence of all four domains.<sup>158</sup> The interaction of PspA subunits may be dependent on a binding pocket that the more cumbersome SGWX<sub>5</sub> library may fit with increased efficiency, improving disruption. However, without a crystal structure of PspA and its oligomerisation one can only speculate as to how the proteins interact and what residues are essential for interaction.

The two inhibitors identified were *cyclo*-SGWLVYWF and *cyclo*-SGWSIVFI; these both consisted of aromatic and hydrophobic residues. There was little homology between the two sequences isolated; this may indicate that they disrupt the PPI but bind to different regions on the protein. Typically, some trends are observed between active sequences,<sup>132,133</sup> however, as observed here during the identification of inhibitors that interfere with the β-sliding clamp of the replisome, little homology was evident (i.e. *cyclo*-WAGSWG, *cyclo*-VFLCGC and *cyclo*-SQGLFK), suggesting each inhibitor could be targeting a different binding site.<sup>242</sup> The dominance of aromatic and aliphatic residues in the PspA inhibitors makes them both 62% hydrophobic. This is analogous to the cyclic peptide inhibitors identified for HIF dimerisation (i.e. *cyclo*-CLLFVY) and Hdm2 and p53 binding (i.e. *cyclo*-CLWWYM) using the SICLOPPS technology, both contained 67% hydrophobic residues.<sup>132,134</sup> Although a highly hydrophobic inhibitor, Miranda *et al.* (2013) demonstrated that *cyclo*-CLLFVY was able to disrupt HIF-1 dimerisation specifically over HIF-2 dimerisation, and bind with nanomolar affinity specifically to the Per-Arnt-Sims domain B of HIF-1α, even though in order to solubilise these inhibitors tagging and DMSO were required.<sup>134</sup> The high affinity of binding of this protein and peptide and similar small molecules to large protein interfaces may be surprising:<sup>70</sup> however, mutational studies of 2325 residues have shown that PPIs are often dependent on ‘hotspots’.<sup>276</sup> These typically contain the residues: tryptophan, tyrosine and arginine as they are able to form hydrogen bonds and exclude solvents from the binding pocket, increasing the free energy of binding.<sup>277</sup> The interaction between interleukin (IL)-2 and IL-2Rα has been disrupted with an inhibitor that has been predicted by mapping studies to bind into two ‘hotspots’.<sup>278</sup> The guanido group

binds into a rigid hydrophilic pocket while the dichlorophenyl moiety binds into a hydrophobic pocket.<sup>277</sup>

Hypothetically, it may be these ‘hotspots’ that are targeted by the cyclic peptide inhibitors identified using the SICLOPPS platform. In the case of the PspA, HIF-1 and HdmX-p53 inhibitors due to their hydrophobic nature, they may be binding specifically to hydrophobic pockets.<sup>132,134</sup> Inhibitors of PPIs involved in viral entry of HIV, and eukaryotic transcription provide an insight into how hydrophobic pockets may be good targets. Previously, the hydrophobic pocket in HIV-1 glycoprotein-41 has been targeted, as it is highly conserved and essential for viral entry.<sup>279</sup> Similarly, the interaction between the eukaryotic Initiation Factor-4G and eukaryotic Initiation Factor-4E is critical in gene expression at the transcriptional level. A small molecule inhibitor has been identified and using NMR been shown to bind specifically to eukaryotic Initiation Factor-4E via the residues H37, V69, L131 and I138.<sup>280</sup> Mapping studies support the evidence from NMR that the inhibitor bound to a hydrophobic pocket containing the residues valine, leucine and isoleucine.<sup>277</sup>

Having identified two active peptides (*cyclo*-SGWLVYWF and *cyclo*-SGWSIVFI) *in vitro* assays were developed to test the efficacy of each inhibitor. SEC has successfully been utilised to show disruption of the homodimerisation of CtBP and ATIC with their respective cyclic peptide inhibitors.<sup>122,133</sup> However, for the oligomerisation of PspA, SEC was unsuccessful: there are two possibilities for this lack of disruption. The number of PspA molecules in the native environment involved in the oligomerisation may be too great for the inhibitor to function; moreover, they may interact by different orientations and residues. The inhibitors identified may inhibit one orientation of PspA dimerisation (i.e. the head-to-head orientation) but not the others. Secondly, the RTHS is a highly contrived system: the formation of the higher-order structures observed by electron microscopy are stable large structures,<sup>209,212</sup> which may not be disrupted by the inhibitor or the concentrations of inhibitor provided in each experiment, whereas the inhibitor in the RTHS may be sufficient to disrupt the dimer, due to the restriction in the different PPIs between PspA i.e. head-to-head.

Due to the lack of success using SEC to demonstrate disruption of the PPI, biophysical techniques to show direct binding of the inhibitors to PspA were utilised. Firstly, expression of MBP-tagged PspA was required to reduce the oligomerisation of PspA observed with His<sub>6</sub>-tagged PspA.<sup>209,212</sup> By ITC the inhibitors did not bind to the MBP-tagged PspA. ITC is a highly sensitive technique and the design of the experiment is critical. Typically, other techniques (i.e. SPR and MST) are used to identify a range for the K<sub>d</sub> and then ITC is used to confirm these findings. Successful ITC is dependent on the Wiseman *c* value, which itself is dependent on the binding constant, K<sub>a</sub>, stoichiometry and the protein concentration in the cell (Equation 8).<sup>268</sup>

$$c = nK_a[M]_t \quad (\text{Equation 8})$$

The *c* value for most experiments should be between 50 and 500 for measuring accurate K<sub>a</sub> values, as protein saturation can be achieved by adding as little as 2 equivalents of inhibitor. Using this information in experimental design the protein concentration should be in the range of the K<sub>d</sub> and the ligand (in this case the inhibitors) should be in ten-fold excess. If the incorrect protein concentration was selected for the affinity of the interaction between the inhibitor and MBP-tagged PspA a low *c* value would occur. Although, ITC data with low *c* values have been successfully published, the validity of the data should be questioned.<sup>281</sup> Turnbull *et al.* (2003) demonstrated that as long as the protein and inhibitor concentrations are accurately known, the signal-to-noise ratio is low and a significant proportion of the isotherm is used for analysis, the thermodynamic properties can be obtained with confidence.<sup>282</sup> Had the inhibitor bound to MBP-tagged PspA, using a protein concentration lower than the K<sub>d</sub> an isotherm should have still been obtainable, but with characteristic properties of a low *c* value experiment.<sup>282</sup> No such isotherm was obtained, indicating that no binding was occurring. Although it is unlikely that changing the experimental parameters would reveal binding, poor experimental design cannot be eliminated and at higher protein and inhibitor concentrations binding might be observed. Therefore, due to the lack of information regarding the K<sub>d</sub> of binding of the inhibitors to PspA, ITC experiments were halted until more information was obtained.

Consequently, other biophysical techniques were investigated. MST is a high-throughput method that produces binding data rapidly using small quantities of

protein and inhibitor.<sup>269</sup> Despite using a wide range of inhibitor concentrations, neither *cyclo*-SGWLVYWF nor *cyclo*-SGWSIVFI showed any indication of binding at concentrations up 5 mM. The plots were similar to the ligand binding of olfactory receptor Olfr226 to its ligand 2,4-DNT when the receptor has been denatured, indicating no binding.<sup>283</sup> This supported the preliminary ITC data that no binding was occurring. Nevertheless, the absence of binding observed could be attributed to the inhibitors binding on the interface of two PspA subunits, making very weak affinity when only one PspA subunit is present.<sup>158</sup>

Having investigated the disruption of PspA oligomerisation using SEC and direct binding of the inhibitor to MBP-tagged PspA with no success, an *in vivo* assay was explored. Previous work by S. Southern at DSTL confirmed a phenotype in a *pspA* mutant during stationary phase for *B. pseudomallei*,<sup>227</sup> observed previously for *E. coli* PspA.<sup>148</sup> Building upon this knowledge a SICLOPPS plasmid-encoding *cyclo*-SGWLWYWF, which had been codon-optimised for expression within *B. pseudomallei*, was expressed in *B. pseudomallei* in tandem with activation of the Psp response. No growth defect was observed compared to the wild-type in the presence of the inhibitor. If the inhibitor had been active in the assay then a growth disadvantage would have been observed. Critically this work was not replicated, with no confirmation that the SICLOPPS plasmid was expressing the inhibitor correctly in *B. pseudomallei*. The method of expressing cyclic peptides using inteins has, however, been carried out in hosts other than *E. coli*. Kritzer *et al.* (2009) successfully identified cyclic peptides that reduce the toxicity of  $\alpha$ -synuclein using intein processing in a yeast model;<sup>129</sup> while Kinsella *et al.* (2002) utilised the SICLOPPS platform to identify inhibitors of interleukin-4 signalling in human B cells using a retroviral expression system.<sup>130</sup> The inhibitor expression should be confirmed by extraction from the cell lysate and analysis by HPLC and MS.<sup>100</sup>

Unfortunately, it is difficult to say whether the observed lack of *in vitro* and *in vivo* activity was typical of the RTHS and SICLOPPS screening platform, as only successful screens are published.<sup>126,127,131-134</sup> In contrast, there could be a high occurrence of failed inhibitors after initial screening. The failure of these peptide sequences to demonstrate activity could be due to the large scaffold that PspA forms when recombinantly expressed<sup>209,212</sup> and observed *in vivo* using GFP-fused PspA.<sup>205</sup>

Moreover, the lack of activity of the inhibitors *in vivo* could be, due to the *B. pseudomallei* Psp system itself. This project was in collaboration with DSTL. One aim was to identify cyclic peptide inhibitors of PspA dimerisation and confirm their activity; the other aim was to characterise the *psp* operon and confirm its requirement for *B. pseudomallei* survival. This study identified a PspA homologue in *B. pseudomallei* (BPSL2105), which forms a putative operon with BPSL2106 (a putative membrane protein).<sup>227</sup> This reduced system is identical to the *psp* operon of *S. lividans*.<sup>161</sup> The role of the putative *pspA* was investigated using deletion mutants and shown to influence survival during stationary phase.<sup>227</sup> This, in conjunction with the electron microscopy that demonstrated how the recombinantly expressed PspA formed ring-like species, similar to those observed for *E. coli*, suggested that BPSL2105 was the PspA homologue for *B. pseudomallei*. Nevertheless, it did not appear to be essential for survival in stress assays that typically induce the response in *E. coli*. The non-essential nature of PspA could be due to the presence of other stress responses and their overlap with the Psp response compensating in its absence, or that BPSL2105 is not PspA, but a similar protein.<sup>147,227</sup>

In addition to the inhibitor screening experiments in this chapter, it has been revealed by electron microscopy that *E. coli* and *B. pseudomallei* PspA units assemble to form rod-like complexes analogous to those previously observed for *Chlamydomonas reinhardtii* Vipp1.<sup>211</sup> Both proteins (Vipp1 and PspA) are predicted to be highly  $\alpha$ -helical (97%), with the  $\alpha$ -helical domains in Vipp1 being critical to the formation of higher-order complexes.<sup>207</sup> Here the formation of these rod-like complexes occurred independent of the presence or absence of detergent in the purification buffer (as had been previously hypothesised).<sup>212</sup> Previous studies only report the formation of ring-like<sup>209</sup> and clathrin-like structures<sup>212</sup> for *E. coli* PspA. One possibility for the absence of this species in prior studies could be due to the concentration of the protein used. Studies have shown that when Vipp1 expression is significantly reduced (e.g. due to the use of a mutant strain), these rod-like structures are no longer observed;<sup>284</sup> indicating that protein concentration is a key determinant for the formation of these higher-order PspA structures. Once formed, these higher-order structures may restore membrane integrity by localising to the cytoplasmic face of the inner membrane; PspA is known to interact with the phospholipids phosphatidylglycerol and phosphatidylserine when in an oligomeric form.<sup>200</sup> It may be postulated that

these mesh-like structures are the physiologically relevant form of PspA, functioning as scaffolds to maintain membrane integrity during the Psp response to membrane stress. However, they may be an artefact of purification.

A study by Yamaguchi *et al.* (2013) utilised GFP-fused PspA to monitor the subcellular location before and after induction supports the premise that these rod-like structures have a biological relevance. GFP-fused PspA was evenly distributed within the cytoplasm prior to induction, with highly dense clusters observed at the membrane (primarily around the poles) after PspA induction.<sup>205,214</sup> The large variation in the fluorescence intensity of the GFP-fused PspA clusters is characteristic of the formation of several higher-order species: 59% of the observed clusters showed a two- to three-fold increase in intensity above background GFP, while the intensity of a further 20% of the clusters had increased by more than six-fold.<sup>205</sup> In light of these observations, it is possible that clusters showing a large increase in fluorescence intensity correspond to the rod-like species, while the species showing lower fluorescence intensity may be the 36-meric ring species, clathrin-like scaffolds or the rod-like complexes observed here.<sup>209,212</sup> It should be noted that such localisation has also been observed with Vipp1,<sup>210</sup> with the formation of large membrane based scaffolds restoring the membrane potential under stress.<sup>285</sup> It is interesting to note that the analogous rod-shaped complexes formed by Vipp1 are not permanent, but have been found to be disassembled by HSP70B (DnaK)-CDJ2-CGE1 chaperones *in vitro*; when formed in cell extracts, these structures are also disassembled by the chaperones in an ATP-dependent process.<sup>211</sup> It may be envisaged that the mesh-like scaffolds formed by PspA are dynamic structures that are disassembled by an analogous, currently unknown mechanism. The self-assembly of ring-shaped protein complexes to rod-like structures has also been shown with holins.<sup>286</sup> The oligomerisation of holins forms lethal holes in the cytoplasmic membrane, thus controlling the length of the infection cycle of tailed phages. Both the holin protein S105 and the PspA homologue, LiaH, form a dimeric species with an indentation that could further assemble into rod-shaped structures.<sup>286,287</sup>

*In vitro* higher-ordered species are required for the interaction with liposomes. These species have never been observed in bacteria; therefore, they may be an artefact of

the purification. There can only be speculation as to the physiological role of the observed rod-like complexes; the biological implications and importance of these structures is unknown and further studies are required to better elucidate their purpose and function. In the future this could be achieved by replicating the experiments by Kobayashi *et al.* (2007) with *B. pseudomallei*-like liposomes, whereby recombinantly expressed PspA oligomer was shown to bind to membrane phospholipids *in vitro* by a sucrose gradient.<sup>200</sup> Additionally, His<sub>6</sub>-tag immuno-staining could be used to confirm the putative 36-meric ring-like structure and its presence in the rod-like structures.

Overall the work in this chapter utilised the SICLOPPS and RTHS technology to identify two cyclic peptides with potential inhibitory activity against the oligomerisation of PspA (i.e. *cyclo*-SGWLVYWF and *cyclo*-SGWSIVFI). Neither cyclic peptide demonstrated any ability to disrupt PspA oligomerisation *in vitro* or *in vivo*, as discussed this may have been due to the incorrect assay used or the nature of the PPI being too complex for the inhibitors isolated from the RTHS and SICLOPPS screen. Due to the lack of information about the Psp response within *B. pseudomallei* and the comprehensive study of the inhibitors using a variety of techniques yielding no definitive evidence for activity, the Psp system as a target was abandoned. This chapter has illustrated that, when selecting a target, a comprehensive understanding of the system is required, with a variety of robust assays available to verify the activity of the peptides after screening.

In order to probe the potential of using the combination of a RTHS and SICLOPPS technology to screen for inhibitors of PPIs essential for virulence in pathogenic bacteria, the next target was selected on the basis of established assays in the literature to test the identified inhibitors<sup>288-291</sup> and the number of subunits involved in the PPI.<sup>292-295</sup> The focus of Chapter 4 was anthrax toxin internalisation, principally the interaction between the anthrax toxin, protective antigen and the mammalian receptor.

## 3.5 Materials and methods: Identification of inhibitors of PspA oligomerisation

### 3.5.1 *Construction of the PspA RTHS and SICLOPPS libraries*

Unless otherwise stated the techniques described below were carried out as outlined in the general methods in Chapter 2.

#### 3.5.1.1 *Constructing pTHCP16 PspA*

The *SacI* restriction site was removed from *B. pseudomallei* K96243 *pspA* by site-directed mutagenesis with Pfu polymerase (Promega, UK) using the primers SDM PspA-F and SDM PspA-R (Table 2), using an extension time of 8 minutes and annealing temperature of 56°C. PCR was followed by an hours incubation at 37°C with *DpnI* in buffer 4 and then gel purified. The purified product was transformed into DH5α chemically-competent cells and plated onto LB agar plates (100 µg/ml ampicillin). Successful mutation was confirmed by analytical restriction digestion with *SacI* restriction endonuclease using buffer 4 and BSA and verified by DNA sequencing. Mutated *pspA* was amplified by PCR using GoTaq polymerase (Promega, UK) with the primers PspA(*BamHI*)-F and PspA(*SacI*)-R (Table 2) with an annealing temperature of 55°C and extension time of 45 seconds: the product was purified by gel extraction. The PCR product and pTHCP16 plasmid were restriction digested with *BamHI* and *SacI* using buffer 4 and BSA: the fragments were gel purified and ligated overnight at 4°C, at a 1:3 ratio (backbone:insert). The ligation was transformed into DH5α chemically-competent cells, plated onto LB agar plates (100 µg/ml ampicillin) and incubated overnight at 37°C. Positive colonies were identified by colony PCR, using the M13-49 and M13-43 primers (Table 2, annealing temperature of 57°C and extension time of 1 minute), analytical restriction digestion (using *NdeI* and *HindIII*) and confirmed by DNA sequencing.

#### 3.5.1.2 *Constructing pAH68 PspA*

pTHCP16 PspA and pAH68 were digested with *MscI* and *SacI* overnight at 37°C: the enzymes were heat inactivated at 65°C for 20 minutes, and the pAH68 backbone

was treated with thermosensitive alkaline phosphatase for 1 hour at 37°C. The fragments were gel purified and ligated overnight at 16°C, at a 1:3 ratio (backbone:insert). The ligation mixture was chemically transformed into chemically-competent DH5α-*pir* cells and grown overnight on LB agar plates (100 µg/ml ampicillin) at 37°C. Positive colonies were identified by colony PCR using M13-49 and P2 primers (Table 2, annealing temperature of 52°C and extension time of 1 minute and 45 seconds) and analytical digests (using *SacI* and *NdeI* restriction endonucleases).

#### *3.5.1.3 Integration of pAH68 PspA onto the chromosome of SNS118*

A 5 ml culture of SNS118 was incubated overnight and used to inoculate 250 ml LB media. The culture was incubated at 30°C until the OD<sub>600</sub> reached 0.4. Then incubated further at 42°C until the OD<sub>600</sub> reached 0.6. The electro-competent cells were prepared as described in Section 2.4.8. The electroporated recovery solution was incubated at 37°C for 1 hour and subsequently at 42°C for 30 minutes, to eject the helper plasmid, before plating onto LB agar plates (50 µg/ml ampicillin and 25 µg/ml spectinomycin) and incubated overnight at 37°C.

#### *3.5.1.4 Identification of integration*

Integrants were identified by PCR, as outlined in Table 27 and Table 28.

**Table 27: Composition of GoTaq PCR reaction mixture with P1, P2, P3 and P4 primers.**

REAGENT	VOLUME	FINAL CONCENTRATION
GoTaq polymerase	0.5 µl	1.25 U
GoTaq buffer	10 µl	5 x concentration
10 mM dNTPs	5 µl	0.2 mM each dNTP
P1, P2, P3 and P4	1 µl each	100 pmol/µl
DNA template	1 µl	100 ng/µl
Sterile deionised water	29.5 µl	Up to 50 µl

**Table 28: PCR program for GoTaq polymerase with P1, P2, P3 and P4 primers.**

TEMPERATURE	TIME	NUMBER OF CYCLES
95°C	2.5 minutes	
95°C	1.5 minutes	-\
55°C	30 seconds	30 cycles
72°C	1 minute	_/_
72°C	10 minutes	
4°C	∞	

The products were analysed on a 1% agarose gel by gel electrophoresis, where bands at 824 bp and 289 bp were observed. If there was no integration, a band at 740 bp was observed.

#### 3.5.1.5 Confirming integration by PCR

P1 and P4 primers and LongAmp polymerase (NEB, UK) amplifies pAH68 PspA off the chromosome. The reaction parameters are outlined in Table 29 and Table 30.

**Table 29: Composition of LongAmp PCR reaction mixture with P1 and P4 primers.**

REAGENT	VOLUME	FINAL CONCENTRATION
LongAmp polymerase	1.25 U	10 x concentration
LongAmp buffer	10 µl	5 x concentration
10 mM dNTPs	2 µl	0.2 mM each dNTP
P1 and P4 primers	2 µl each	100 pmol/µl
DNA template	4 µl	100 ng/µl
Sterile deionised water	28.5 µl	Up to 50 µl

**Table 30: PCR program for LongAmp polymerase PCR with P1 and P4 primers.**

TEMPERATURE	TIME	NUMBER OF CYCLES
94°C	2.5 minutes	
94°C	30 seconds	- \
51°C	30 seconds	30 cycles
65°C	4 minutes and 30 seconds	_ /
65°C	10 minutes	
4°C	∞	

The products were analysed by 1% agarose gel electrophoresis, where a band at 6000 bp was observed.

#### 3.5.1.6 Further PCR conditions for verifying Integration

Several additional PCR experiments were conducted to verify integration. The details are outlined below in Table 31 using GoTaq polymerase and parameters outlined in 2.4.2.

**Table 31: Conditions for PCR for the verification of integration.**

PRIMERS	POLYMERASE	ANNEALING TEMP	EXTENSION TIME	BAND SIZE
				TEMP
M13-49/P2	GoTaq	52°C	1 minute and 30 seconds	No band
M13-49/PspAR	GoTaq	55°C	1 minute and 15 seconds	1080 bp
PspAF /P2	GoTaq	52°C	1 minute and 15 seconds	No band
M13-49/P4	GoTaq	52°C	2 minutes	2220 bp
P1/PspA-R	LongAmp	51°C	5 minutes	≈5000 bp
P4/PspA-F	GoTaq	52°C	2 minutes	≈2000 bp
P1/P2	LongAmp	51°C	6 minutes	≈6000 bp

All PCR products were analysed on a 1% agarose gel by gel electrophoresis.

### 3.5.1.7 CX<sub>5</sub> SICLOPPS library construction for PspA dimerisation

The I<sub>N</sub>, degenerate and I<sub>C</sub> sequence were amplified using FLC+5 and InR-Bg/II-R primers using Klenow polymerase. 5 µl of each primer was diluted in 40 µl water and heated to 98°C for 3 minutes and 55°C for 2 minutes. 20 µl of each primer solution was incubated with 4 µl dNTP, 4 µl buffer 2, 10 µl water and 2 µl Klenow polymerase for 1 hour at 25 °C. The mixture was purified by PCR purification. pARCBD-BgII/BgII plasmid and the PCR product were restriction digested with BgII and BgII in buffer 3 overnight at 37°C and purified by gel extraction. The ligation was done at a 1:6 ratio (backbone:insert) and incubated overnight at 16°C. The ligation mixture was transformed into electro-competent NEB-10-β cells (NEB, UK). A 10<sup>8</sup> efficiency was achieved and the remaining SOC recovery solution was incubated in 10 ml LB medium overnight at 37°C, frozen stocks were aliquoted and stored at -80°C. When required the stocks were maxi-prepped and transformed into electro-competent PspA RTHS cells, where it was plated on minimal media plates supplemented with arabinose (6.5 µM), IPTG (50 µM), ampicillin (50 µg/ml), spectinomycin (25 µg/ml), chloramphenicol (35 µg/ml), kanamycin (50 µg/ml) and 3-AT (5.0 mM).

### 3.5.1.8 SGWX<sub>5</sub> SICLOPPS library construction for PspA dimerisation

The I<sub>N</sub>, degenerate sequence and I<sub>C</sub> sequence were amplified using SGWX<sub>5</sub>-F primer and CBD-R primer using GoTaq polymerase, with an annealing temperature of 55°C and an extension time of 45 seconds. A second PCR was carried out using the zipper primer and CBD-R to realign the fragments, at an annealing temperature of 55°C and an extension time of 45 seconds. The PCR product was column-purified and, along with pARCBD plasmid, restriction digested with 2 µl BgII and 1µl HindIII in buffer 2. The products were ligated at a 1:6 ratio (backbone:insert) and electroporated into electro-competent PspA RTHS cells and plated onto LB agar plates (35 µg/ml chloramphenicol) giving an efficiency of 3.6 x 10<sup>6</sup> and minimal media plates supplemented with arabinose (6.5 µM), IPTG (50 µM), ampicillin (50 µg/ml), spectinomycin (25 µg/ml), chloramphenicol (35 µg/ml), kanamycin (50 µg/ml) and 3-AT (5.0 mM).

#### *3.5.1.9 SX<sub>5</sub>G SICLOPPS library construction for PspA dimerisation*

The I<sub>N</sub>, degenerate sequence and I<sub>C</sub> sequence were amplified using SX<sub>5</sub>G-F primer and CBD-R primer and realigned as Section 3.5.1.8. The PCR product was purified and cloned in pARCBP as described in Section 3.5.1.8.

#### *3.5.1.10 SX<sub>6</sub>G SICLOPPS library construction for PspA dimerisation*

The I<sub>N</sub>, degenerate sequence and I<sub>C</sub> sequence were amplified using SX<sub>6</sub>G-F primer and CBD-R and realigned as Section 3.5.1.8. The PCR product was purified and cloned in pARCBP as described in Section 3.5.1.8.

### **3.5.2 Genetic selection and inhibitor screening of PspA**

#### *3.5.2.1 Drop-spotting to verify PspA dimerisation*

As in Section 2.4.10 the integrants were drop-spotted onto minimal media supplemented with selection varying concentrations of IPTG, kanamycin and 3-AT.

#### *3.5.2.2 Screening of inhibitors against PspA dimerisation*

Each library was plated onto minimal media plates supplemented with 50 µg/ml kanamycin, 5.0 mM 3-AT, 50 µM IPTG and 6.5 µM arabinose and incubated for 72 hours at 37°C. Colonies were grown overnight in 1 ml cultures containing 50 µg/ml ampicillin, 25 µg/ml spectinomycin and 35 µg/ml chloramphenicol and drop-spotted (as Section 2.4.10) onto minimal media plates containing selection conditions as described in Section 3.2.2.1.

### **3.5.3 Construction of inhibitor splicing mutants and splicing verification**

#### *3.5.3.1 Construction of I<sub>C</sub> mutant pARCBP*

The intein-extein region of pARCBP was amplified by PCR using primers I<sub>C</sub> (H24A, F26A)-F and I<sub>C</sub> (H24A, F26A)-R from pARCBP, introducing the corresponding mutations into the C-terminal intein (annealing temperature 50°C and

extension time 15 seconds) and then cloned into pARCBP by digesting with *Nco*I and *Bgl*II restriction endonucleases in buffer 3. The fragments were ligated at a 1:6 (backbone:insert) ratio at 16°C and transformed into chemically-competent NEB-5- $\alpha$ . This construct was confirmed by DNA sequencing. The construct was subsequently referred to as pARCBP-I<sub>C</sub> (H24A, F26A).

### *3.5.3.2 Construction of I<sub>N</sub> mutant pARCBP*

The intein-extein region of pARCBP was amplified by PCR using primers I<sub>N</sub> (T69A, H72A)-F and I<sub>N</sub> (T69A, H72A)-R from pRSET-I<sub>N</sub>mut, incorporating the mutations into the N-terminal intein (annealing temperature 55°C and extension time 45 seconds) and then cloned into pARCBP by digesting with *Kpn*I and *Afl*III restriction endonucleases in buffer 4 and BSA. The fragments were ligated at a 1:6 (backbone:insert) ratio at 16°C and transformed into chemically-competent NEB-5- $\alpha$ . This construct was confirmed by DNA sequencing. The construct was subsequently referred to as pARCBP-I<sub>N</sub> (T69A, H72A).

### *3.5.3.3 Construction of SGWIYWNV plasmid*

The intein-extein region of pARCBP was amplified by PCR using primers SGWIYWNV-F, and CBD-R, incorporating the positive control gene sequence, SGWIYWNV, into the extein region (annealing temperature 55°C and extension time 45 seconds) and then cloned into pARCBP by digesting with *Bgl*I and *Hind*III restriction endonucleases in buffer 2 and BSA. The fragments were ligated at a 1:6 (backbone:insert) ratio at 4°C and transformed initially into chemically-competent NEB-5- $\alpha$ . This construct was confirmed by DNA sequencing.

### *3.5.3.4 Construction of SGWLVYWF, SGWSIVFI and SGWIYWNV I<sub>C</sub> mutant plasmid*

The I<sub>C</sub>-mutated intein-extein region of pARCBP was amplified by PCR using primers SGWLVYWF-F, SGWSIVFI-F or SGWIYWNV-F and CBD-R, introducing the corresponding mutations into the degenerate extein sequence (annealing temperature 55°C and extension time 45 seconds) and then cloned into pARCBP-I<sub>C</sub>

(H24A, F26A) by restriction digesting with *Hind*III and *Bgl*II restriction endonucleases in buffer 2. The fragments were ligated at a 1:6 (backbone:insert) ratio at 16°C and transformed initially into chemically-competent NEB-5- $\alpha$ , then for expression into either chemically-competent BL21 (DE3) or the PspA RTHS or p6-UEV RTHS. These constructs were confirmed by DNA sequencing.

### *3.5.3.5 Construction of SGWLVYWF, SGWSIVFI and SGWIYWNVF I<sub>N</sub> mutant plasmid*

The I<sub>N</sub>-mutated intein-extein region of pARCBP-I<sub>N</sub> (T69A, H72A) was amplified by PCR using primers SGWLVYWF-F, SGWSIVFI-F or SGWIYWNV-F and CBD-R from pARCBP, introducing the degenerate extein sequence (annealing temperature 55°C and extension time 45 seconds) and then cloned into pARCBP by restriction digesting with *Hind*III and *Bgl*II restriction endonucleases in buffer 2. The fragments were ligated at a 1:6 (backbone:insert) ratio at 16°C and transformed initially into chemically-competent NEB-5- $\alpha$ , then for expression into either chemically-competent BL21 (DE3) or the PspA RTHS or p6-UEV RTHS. These constructs were confirmed by DNA sequencing.

### *3.5.3.6 Expression of cyclic peptide inhibitors*

Cells were grown in LB media to an OD<sub>600</sub> of 0.6, and protein expression was induced with 0.5% arabinose and incubated for a further 3 hours at 30°C. After incubation cells were harvested by centrifugation and stored at -80°C. Cells were resuspended in chitin-binding buffer (Table 32) and lysed by sonication. The insoluble and soluble fractions were separated by centrifugation at 8,000 rpm for 30 minutes. The soluble fraction was then purified using chitin beads (as per manufacturer's protocols), and chitin-binding buffer (Table 32). Proteins were analysed by 15% SDS-PAGE gels.

**Table 32: Composition of chitin-binding buffer.**

REAGENTS	AMOUNTS	FINAL CONCENTRATION
Tris- HCl	1.21 g	20 mM
Sodium chloride	14.5 g	0.5 mM
TCEP	125 mg	1.0 mM
Sterile deionised water	Up to 500 ml	
Adjust pH to 7.8 with 1 M HCl		

### 3.5.4 *PspA protein purification*

#### 3.5.4.1 *Construction of pET28 PspA expression plasmid*

The gene encoding *B. pseudomallei* K96243 PspA was amplified by PCR using the primers PspA(*Bam*HI)-F and PspA(*Sac*I)-R (annealing temperature of 58°C and an extension time of 45 seconds) and cloned into the corresponding sites of pET28a (Novagen) by digesting with the corresponding restriction endonucleases in buffer 4 and BSA. The fragments were ligated at a 1:3 (backbone:insert) ratio at 16°C and transformed initially into chemically-competent DH5α, then for expression into chemically-competent BL21 (DE3). The resulting pET28-hexa-histidine (*His*<sub>6</sub>)-*pspA* plasmid encoded an N-terminally His<sub>6</sub>-tagged PspA. This construct was confirmed by DNA sequencing.

#### 3.5.4.2 *Native PspA Protein Purification from the soluble fraction*

PspA purification procedure was based on the method described by Elderkin *et al.* (2005).<sup>158</sup> Briefly, cells were grown in LB media to an OD<sub>600</sub> of 0.6, and protein expression was induced with 1 mM IPTG and incubated further overnight at 18°C. After incubation cells were harvested by centrifugation and stored at -80°C. Cells were re-suspended in lysis buffer (Table 33) and lysed by sonication. The insoluble and soluble fractions were separated by centrifugation at 8,000 rpm for 30 minutes. The soluble fraction was purified by nickel affinity chromatography using a lysis buffer with 20 mM imidazole and eluted from the column by the addition of 500 mM

imidazole. Proteins were analysed by Bradford assay (as per manufacturer's protocols) and 10% SDS-PAGE gels. Yields were typically between 0.5-0.95 mg/ml.

**Table 33: Composition of PspA lysis buffer.**

REAGENTS	AMOUNTS	FINAL CONCENTRATION
Tris-base	6. 06 g	100 mM
Sodium chloride	1.46 g	50 mM
Sodium thiocyanate	3.30 g	75 mM
Sterile deionised water	Up to 500 ml	
	Adjust pH to 7.5 with 1 M HCl	

#### *3.5.4.3 Native PspA Protein Purification from the insoluble fraction*

For purification from the insoluble fractions cells were prepared as 3.5.4.2, but after centrifugation the insoluble fraction was resuspended in extraction buffer (Table 34) and incubated at 4°C for 2 hours. Subsequently, the sample was centrifuged at 8,000 rpm for 30 minutes and then purified as previously described. Yields were typically between 0.45-0.80 mg/ml.

**Table 34: Composition of PspA extraction buffer.**

REAGENTS	AMOUNTS	FINAL CONCENTRATION
Tris-base	6.06 g	100 mM
Sodium chloride	17.53 g	600 mM
Sodium thiocyanate	3.30 g	75 mM
CHAPS	5.5 g	1.1%
Sterile deionised water	Up to 500 ml	
	Adjust pH to 7.5 with 1 M HCl	

#### *3.5.4.4 Denaturing purification of PspA*

Under denaturing conditions PspA was purified as Kobayashi *et al.* (2007),<sup>200</sup> using the buffers outlined in Table 35-38. For refolding experiments the urea was removed

using a concentration gradient into refolding buffer at a flow rate of 0.5 min/ml over 40-150 ml (Table 37). Yields were typically between 0.45-0.80 mg/ml.

**Table 35: Composition of PspA denaturing lysis buffer.**

REAGENTS	AMOUNTS	FINAL CONCENTRATION
Tris-base	0.12 g	20 mM
Sodium chloride	1.46 g	500 mM
β-mercaptoethanol	0.004 ml	1 mM
Guanidine hydrochloride	28.65 g	6 M
Sterile deionised water	Up to 50 ml	
Adjust pH to 7.5 with 1 M HCl		

**Table 36: Composition of PspA denaturing wash buffer.**

REAGENTS	AMOUNTS	FINAL CONCENTRATION
Tris-base	0.24 g	20 mM
Sodium chloride	2.92 g	500 mM
β-mercaptoethanol	0.007 ml	1 mM
Imidazole	0.068 g	10 mM
Urea	48.0 g	8 M
Sterile deionised water	Up to 100 ml	
Adjust pH to 7.5 with 1 M HCl		

**Table 37: Composition of PspA refolding buffer.**

REAGENTS	AMOUNTS	FINAL CONCENTRATION
Tris-base	0.48 g	20 mM
Sodium chloride	5.84 g	500 mM
β-mercaptoethanol	0.014 ml	1 mM
Imidazole	0.27 g	20 mM
Sterile deionised water	Up to 200 ml	
Adjust pH to 7.5 with 1 M HCl		

**Table 38: Composition of PspA refolding elution buffer.**

REAGENTS	AMOUNTS	FINAL CONCENTRATION
Tris-base	0.48 g	20 mM
Sodium chloride	5.84 g	500 mM
$\beta$ -mercaptoethanol	0.014 ml	1 mM
Imidazole	6.80 g	500 mM
Sterile deionised water	Up to 200 ml	
Adjust pH to 7.5 with 1 M HCl		

#### 3.5.4.5 Size exclusion chromatography

Purified PspA (1 ml at 0.5 mg/ml) in gel filtration buffer (Table 39) was analysed on a HiLoad 16/600 Superdex 200 prep grade column or Superose 6 column (GE Healthcare, UK) using a flow rate of either 1 or 0.5 ml/min at room temperature. Protein was detected at 280 nm wavelength. Peaks were analysed by Bradford assay, as per manufacturers' instruction and 10% SDS-PAGE gels.

**Table 39: Composition of PspA gel filtration buffer.**

REAGENTS	AMOUNTS	FINAL CONCENTRATION
Tris-base	2.42 g	20 mM
Sodium chloride	2.92-8.76 g	50-150 mM
Sodium thiocyanate	6.08 g	75 mM
Sterile deionised water	Up to 1000 ml	
Adjust pH to 7.5 with 1 M HCl		

#### 3.5.4.6 Western blot analysis of His<sub>6</sub>-tagged PspA

Proteins were separated by 12% SDS-PAGE gel and transferred onto nitrocellulose transfer membranes (Protran, Whatman). The blots were blocked for 1 hour in 5% milk in phosphate buffer saline (PBS) with 0.05% Tween-20 at room temperature. They were then incubated for 1 hour with anti-His (27-4710-01, GE Healthcare) at a 1 in 1000 dilution in 5% milk in PBS with 0.05% Tween-20 followed by incubation with anti-mouse-HRP (NA931, GE Healthcare) at a 1 in 5000 dilution. Secondary

antibody was detected using an enhanced chemiluminescence (ECL) reagent (Western C, Biorad).

#### *3.5.4.7 Construction of pMAL-tagged PspA*

The gene encoding *B. pseudomallei* PspA was amplified by PCR using the primers PspA-F and PspA(*SacI*)-R and Phusion polymerase (annealing temperature of 55°C and an extension time of 10 seconds). It was cloned into the corresponding sites of pMAL (NEB, UK) by digesting the plasmid with *Xma*I and *Bam*HI restriction endonucleases in buffer 4 and the insert with *Bam*HI. The fragments were ligated at a 1:3 (backbone:insert) ratio at 4°C and transformed initially into chemically-competent DH5 $\alpha$ , then for expression into chemically-competent NEB-express. The resulting pMAL-PspA plasmid encodes an N-terminally maltose-binding protein (MBP)-tagged PspA. This construct was confirmed by DNA sequencing.

#### *3.5.4.8 Expression and purification of MBP-tagged PspA*

Cells were grown in LB media with 0.2% glucose to an OD<sub>600</sub> of 0.6, and protein expression was induced with 0.3 mM IPTG and incubated for 2 hours at 37°C. After incubation cells were harvested by centrifugation, re-suspended in lysis buffer 3329) and frozen at -20°C. Cells were defrosted and subsequently lysed by sonication. The insoluble and soluble fractions were separated by centrifugation at 8,000 rpm for 30 minutes. The soluble fraction was purified by amylose affinity chromatography (as per manufacturer's protocols) using a wash buffer (Table 33) and eluted from the column by the addition of 10 mM maltose. Proteins were analysed by Bradford assay (as per manufacturer's protocols) and 10% SDS-PAGE gels. Yields were typically between 1.2-1.5 mg/ml.

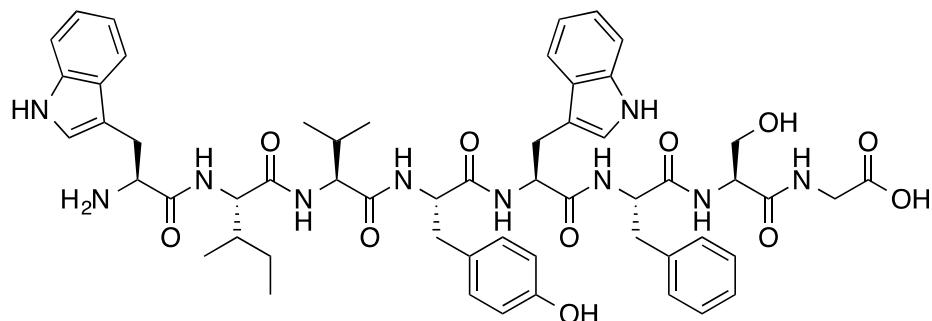
#### *3.5.4.9 Western blot analysis of MBP-tagged PspA*

Proteins were separated by 12% SDS-PAGE GEL and transferred onto nitrocellulose transfer membranes (Protran, Whatman). The blots were blocked for 1 hour in 5% milk in PBS with 0.05% Tween-20 at room temperature. They were then incubated for 1 hour with anti-MBP (E8032S, NEB) at a 1 in 5000 dilution in 5% milk in PBS

with 0.05% Tween-20 followed by incubation with anti-rabbit-HRP (RPN4301, Amersham) at a 1 in 20,000 dilution. Secondary antibody was detected using an ECL reagent (Western C, Biorad).

### **3.5.5 Synthesis of PspA inhibitors**

#### **3.5.5.1 Linear-WLVYWFSG**

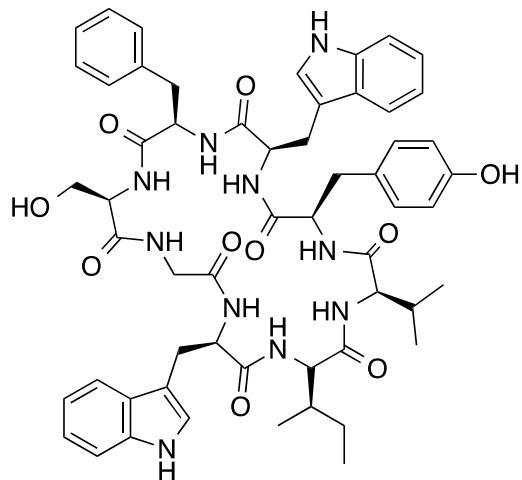


**Figure 59: Structure of linear-WLVYWFSG (1).**

Peptide was synthesised as 2.6.1 yielding 68 mg (26%) of 1 as a white solid. M.p 164.8-181.7°C;  $^1\text{H}$  NMR (600 MHz, DMSO- $d_6$ )  $\delta$  ppm 10.99 (1 H, s, Trp-COOH) 10.71 (1 H, s, Trp-Ar-NH) 9.11 (1 H, s, Tyr-OH) 8.72 (1 H, d,  $J$ =7.93 Hz, Leu-NH) 8.10 (1 H, d,  $J$ =7.93 Hz, Ser-NH) 8.07 (1H, d,  $J$ =7.93 Hz, Tyr-NH) 8.01 (1 H, t,  $J$ =5.65 Hz, Gly-NH) 7.97 (1 H, d,  $J$ =7.63 Hz, Phe-NH) 7.94 (1 H, d,  $J$ =8.85 Hz, Val-NH) 7.90 (1 H, d,  $J$ =8.24 Hz, Trp<sub>2</sub>-NH) 7.73 (1 H, d,  $J$ =7.93 Hz, Trp-ArH) 7.52 (1 H, d,  $J$ =7.63 Hz, Trp-ArH) 7.37 (1 H, d,  $J$ =8.24 Hz, Trp-ArH) 7.30 (1 H, d,  $J$ =8.24 Hz, Trp-ArH) 7.18 - 7.23 (5 H, m, Trp- $\gamma$ H) 6.99 - 7.04 (2 H, m, Trp-ArH) 7.05 (1 H, s, Trp-ArH) 7.10 (1 H, t,  $J$ =7.48 Hz, Trp- $\gamma$ H) 7.12 -7.16 (1 H, m, Trp-ArH) 6.91 - 6.98 (3 H, m, Phe-ArH) 6.58 (2 H, d,  $J$ =8.54 Hz, Phe-ArH) 4.79-5.02 (1 H, m, Trp<sub>1</sub>-NH, Ser-OH or Try-OH) 4.58 - 4.66 (1 H, m, Tyr- $\alpha$ H) 4.52 (1 H, d,  $J$ =6.10 Hz, Phe- $\alpha$ H) 4.39 - 4.49 (2 H, m, Leu- $\alpha$ H and Trp<sub>2</sub>- $\alpha$ H) 4.32 - 4.38 (1H, m, Ser- $\alpha$ H) 4.16 (1 H, t,  $J$ =7.93 Hz, Val- $\alpha$ H) 4.01 (1 H, dd,  $J$ =8.85, 4.58 Hz, Trp<sub>1</sub>- $\alpha$ H) 3.77 (2 H, d,  $J$ =6.41 Hz, Gly- $\alpha$ H) 3.53 - 3.65 (2 H, m, Ser- $\beta$ H) 3.33 (s, solvent-H<sub>2</sub>O) 3.22 (1 H, dd,  $J$ =14.80, 4.12 Hz, Trp<sub>1</sub>- $\alpha$ H) 3.06 (2 H, dd,  $J$ =13.89, 4.12 Hz, Trp<sub>1</sub>- $\beta$ H or Tyr- $\beta$ H and Trp<sub>2</sub>- $\beta$ H) 2.97 (1 H, dd,  $J$ =14.95, 9.16 Hz, Trp<sub>1</sub>- $\beta$ H) 2.86 - 2.94 (1 H, m, Tyr- $\beta$ H) 2.76 - 2.85 (2 H, m, Trp<sub>1</sub>- $\beta$ H) 2.60 (1 H, dd,  $J$ =14.04, 9.46 Hz, Trp<sub>2</sub>- $\beta$ H) 1.90 (1

H, doublet of septet,  $J=6.71$  Hz, Val- $\beta$ H) 1.58 - 1.66 (2 H, m, Leu- $\beta$ H) 1.40 - 1.53 (1 H, m, Leu- $\gamma$ H) 0.89 (3 H, d,  $J=6.41$  Hz, Leu- $\delta$ H) 0.85 (3 H, d,  $J=6.41$  Hz, Leu- $\delta$ H) 0.74 (6 H, dd,  $J=6.26, 4.43$  Hz, Val- $\gamma$ H); Analytical HPLC (280 nm) 20.9 min; IR (neat) 3302, 1649, 1514, 1209 cm<sup>-1</sup>; MS (ESI+)  $m/z$  (%) 1101.5 ((M + H + CO<sub>2</sub>)<sup>+</sup>, 100), 1124.1 ((M + Na + CO<sub>2</sub>)<sup>+</sup>, 74.7).

### 3.5.5.2 Cyclo-SGWLVYWF (2)

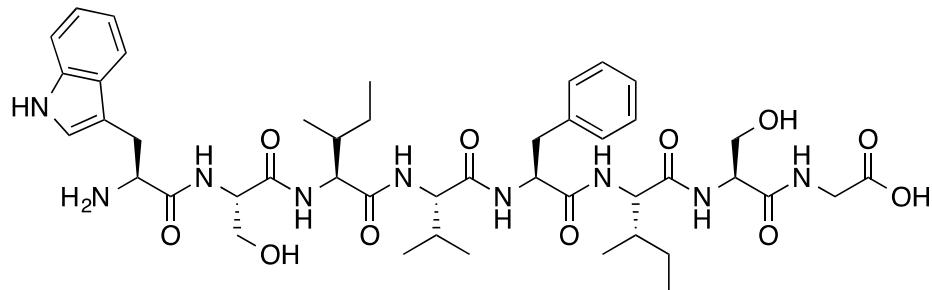


**Figure 60: Structure of cyclo-SGWLVYWF (2).**

Peptide was synthesised as 2.6.1 yielding 15 mg (29%) of 2 as a white solid. M.p 172.4-193.6°C; <sup>1</sup>H NMR (600 MHz, DMSO-*d*<sub>6</sub>) δ ppm 10.78 - 10.86 (2 H, m, Trp<sub>1</sub>-ArNH and Trp<sub>2</sub>-ArNH) 9.14 (1 H, br. s., Tyr-OH) 8.43 (1 H, br. s., Gly-NH) 8.18 (2 H, br. s., Ser-NH and Tyr-NH) 8.00 - 8.09 (1 H, m, Val-NH) 7.94 - 8.00 (1 H, m, Trp<sub>1</sub>-NH) 7.89 - 7.94 (1 H, m, Trp<sub>2</sub>-NH) 7.82 - 7.89 (1 H, m, Leu-NH) 7.47 - 7.57 (2 H, m, Trp<sub>1</sub>-ArH and Trp<sub>2</sub>-ArH) 7.31 (2 H, dd,  $J=18.01, 7.93$  Hz, Trp<sub>1</sub>-ArH and Trp<sub>2</sub>-ArH) 7.18 - 7.23 (2 H, m, Trp<sub>1</sub>-ArH and Trp<sub>2</sub>-ArH) 7.11 - 7.18 (5 H, m, Phe-ArH) 7.03 - 7.09 (2 H, m, Trp<sub>1</sub>-ArH) 6.93 - 7.01 (2 H, m, Trp<sub>2</sub>-ArH) 6.90 (2 H, d,  $J=8.24$  Hz, Tyr-ArH) 6.59 (2 H, d,  $J=8.24$  Hz, Tyr-ArH) 4.52 (1 H, br. s., Trp<sub>2</sub>-αH) 4.41 (2 H, d,  $J=10.07$  Hz, Trp<sub>1</sub>-αH, Tyr-αH) 4.30 (1 H, br. s., Val-αH) 4.16 (1 H, br. s., Ser-αH) 4.11 (1 H, d,  $J=5.49$  Hz, Leu-αH) 3.84 (1 H, br. s., Leu-βH) 3.72 - 3.81 (2 H, m, Gly-αH) 3.62 - 3.69 (5 H, m, Gly-αH, Ser-βH) 3.24 - 3.31 (1 H, m, Trp<sub>2</sub>-βH) 3.05 (2 H, td,  $J=15.41, 4.88$  Hz) 2.94 - 3.01 (2 H, m, Trp<sub>1</sub>-βH) 2.79 - 2.91 (3 H, m, Leu-γH, Tyr-βH) 1.87 - 1.95 (1 H, m, Leu-γH) 1.47 - 1.65 (3 H, m, Val-βH) 0.82 (4 H, d,

$J=6.10$  Hz, Val- $\gamma$ H) 0.78 (4 H, d,  $J=6.10$  Hz, Val- $\delta$ H) 0.72 (8 H, dd,  $J=12.82, 6.71$  Hz, Leu- $\gamma$ H); Analytical HPLC (280 nm) 22.5 min; IR (neat) 3301, 1649, 1514, 1208 cm<sup>-1</sup>; MS (ESI+)  $m/z$  (%) 1061.6 ((M + Na)<sup>+</sup>, 100) HRMS (ESI+) for C<sub>56</sub>H<sub>66</sub>N<sub>10</sub>NaO<sub>10</sub> 1061.5 (M + Na)<sup>+</sup> No accurate mass as >1000 gmol<sup>-1</sup>.

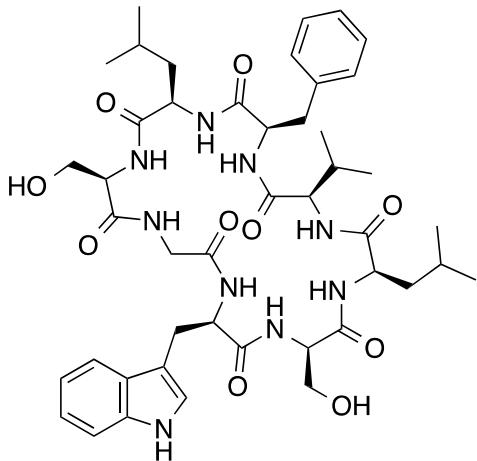
### 3.5.5.3 Linear-WSIVFISG



**Figure 61: Structure of linear-WSIVFISG (3).**

Peptide was synthesised as 2.6.1 yielding 41.1mg (9%) of 3 as a white solid. M.p 182.9-206.8°C; <sup>1</sup>H NMR (600 MHz, DMSO-*d*<sub>6</sub>) δ ppm 12.56 (1 H, br. s., Gly-COOH) 10.80 (1 H, s, Trp-ArNH) 8.59 (1 H, d,  $J=7.32$  Hz, Ser<sub>1</sub>-NH) 8.51 (1 H, d,  $J=7.32$  Hz, Trp-NH) 8.09 (1 H, d,  $J=7.32$  Hz, Phe-NH) 8.04 (1 H, br. s., Ser<sub>2</sub>-NH) 7.84 - 7.96 (5 H, m, Gly-NH) 7.81 (1 H, d,  $J=7.32$  Hz, Val-NH) 7.70 (2 H, dd,  $J=17.09, 7.32$  Hz, Ile<sub>1</sub>-NH, Ile<sub>2</sub>-NH) 7.32 (1 H, d,  $J=9.77$  Hz, ArH) 7.17 - 7.21 (4 H, m, ArH) 7.09 - 7.15 (2 H, m, ArH) 7.02 - 7.06 (1 H, m, ArH) 6.97 (1 H, t,  $J=7.32$  Hz, ArH) 4.96 (2 H, br. s., Ser-OH) 4.71 - 4.81 (1 H, m, Ser<sub>1</sub>-αH) 4.58 - 4.64 (1 H, m, Ser<sub>2</sub>-αH) 4.34 - 4.44 (1 H, m, Trp-αH) 4.23 - 4.30 (3 H, m, Phe-αH, Ile<sub>1</sub>-αH, Ile<sub>2</sub>-αH) 4.15 (1 H, t,  $J=7.32$  Hz, Val-αH) 3.67 - 3.73 (1 H, m) 3.50 - 3.65 (6 H, m, Gly-αH, Trp-βH, Phe-βH) 3.38 (8 H, br. s., Gly-αH) 3.15 (1 H, d,  $J=12.21$  Hz, Ser<sub>1</sub>-βH) 2.96 (1 H, d,  $J=9.77$  Hz, Ser<sub>2</sub>-βH) 2.89 (1 H, dd,  $J=14.65, 9.77$  Hz, Ser<sub>1</sub>-βH) 2.77 (1 H, dd,  $J=14.65, 9.77$  Hz, Ser<sub>2</sub>-βH) 1.85 - 1.93 (1 H, m, Val-βH) 1.66 - 1.76 (2 H, m, Ile-βH) 1.35 - 1.48 (2 H, m, Ile-δ<sub>1</sub>H or Ile-δ<sub>2</sub>H) 1.01 - 1.12 (2 H, m, Ile-δ<sub>1</sub>H or Ile-δ<sub>2</sub>H) 0.72 - 0.84 (24 H, m, Val-γ<sub>1</sub>H, Val-γ<sub>2</sub>H, Ile-γ<sub>1</sub>H, Ile-γ<sub>2</sub>H, Ile-ε<sub>1</sub>H, Ile-ε<sub>2</sub>H); Analytical HPLC (280 nm) 21.65 min; IR (neat) 3268, 2964, 1627, 1537 cm<sup>-1</sup>; MS (ESI+)  $m/z$  (%) 908.6 ((M + H)<sup>+</sup>, 100); 930.6 ((M + Na)<sup>+</sup>, 40); HRMS (ESI+) for C<sub>45</sub>H<sub>65</sub>N<sub>9</sub>O<sub>11</sub> (M + H)<sup>+</sup> calcd 908.4865, found 908.4876 (1.2 ppm error).

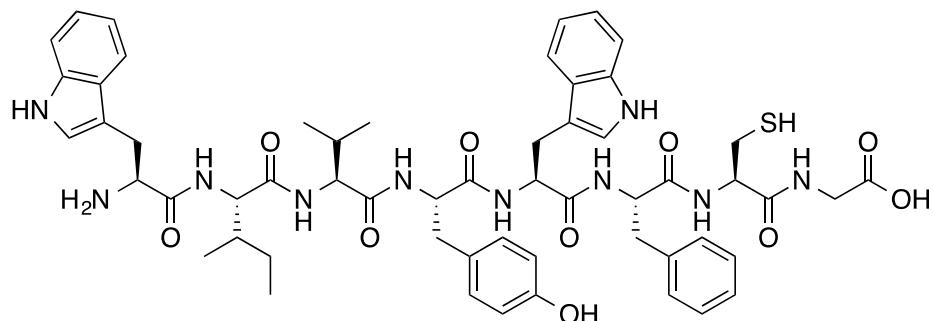
### 3.5.5.4 Cyclo-SGWSIVFI



**Figure 62: Structure of cyclo-SGWSIVFI (4).**

Peptide was synthesised as 2.6.1 yielding 4 mg (23%) of 4 as a white solid. M.p 185.9-204.9°C; <sup>1</sup>H NMR (600 MHz, DMSO-*d*<sub>6</sub>) δ ppm 10.79 (1 H, d, *J*=1.83 Hz, Trp-ArNH) 8.58 (1 H, br. s., Ser<sub>1</sub>-NH) 8.34 (1 H, m, Trp-NH) 8.13 (1 H, d, *J*=8.85 Hz, Phe-NH) 7.91 - 8.08 (3 H, m, Ser<sub>2</sub>-NH, Gly-NH) 7.81 - 7.89 (3 H, m, Val-NH) 7.59 (1 H, d, *J*=7.63 Hz, ArH) 7.53 - 7.55 (1 H, m, Ile<sub>1</sub>-NH) 7.38 - 7.39 (1 H, m, Ile<sub>2</sub>-NH) 7.27 - 7.32 (2 H, m, ArH) 7.23 - 7.26 (5 H, m, ArH) 7.14 - 7.21 (2 H, m, ArH) 7.11 (1 H, d, *J*=2.14 Hz, Trp-indoleH) 7.00 - 7.08 (1 H, m, ArH) 4.42 - 4.55 (4 H, m, Phe-αH, Ser<sub>2</sub>-αH, Gly-αH) 4.37 (1 H, q, *J*=6.41 Hz, ) 4.26 (1 H, dd, *J*=7.63, 5.19 Hz, Val-αH) 4.05 - 4.17 (3 H, m, Ile<sub>1</sub>-αH) 3.92 - 4.02 (3 H, m, Ile<sub>2</sub>-αH) 3.86 (2 H, dd, *J*=16.48, 6.10 Hz, Ser<sub>1</sub>-αH) 3.36 - 3.81 (81 H, m, Ser<sub>1</sub>-βH, Trp-αH, Trp-βH) 3.17 - 3.28 (3 H, m, Phe-βH,) 3.01 - 3.09 (1 H, m, Val-βH) 2.80 - 2.87 (1 H, m, Ser<sub>2</sub>-βH) 1.80 - 1.94 (2 H, m, Ile<sub>1</sub>-βH, Ile<sub>2</sub>-βH) 1.42 - 1.53 (1 H, m, Ile-δ<sub>1</sub>H or Ile-δ<sub>2</sub>H) 1.27 - 1.40 (2 H, m, Ile-δ<sub>1</sub>H or Ile-δ<sub>2</sub>H) 0.70 - 1.10 (17 H, m, Val-γ<sub>1</sub>H, Val-γ<sub>2</sub>H, Ile-γ<sub>1</sub>H, Ile-γ<sub>2</sub>H, Ile-ε<sub>1</sub>H, Ile-ε<sub>2</sub>H) 0.58 - 0.69 (9 H, m, Ile-γ<sub>1</sub>H, Ile-γ<sub>2</sub>H, Ile-ε<sub>1</sub>H, Ile-ε<sub>2</sub>H) 0.56 (3 H, d, *J*=6.71 Hz, Ile-ε<sub>1</sub>H or Ile-ε<sub>2</sub>H); Analytical HPLC (220 nm) 22.51 min; IR (neat) 3293, 2966, 1644, 1518 cm<sup>-1</sup>; MS (ESI+) *m/z* (%) 912.6 ((M + H)<sup>+</sup>, 100); HRMS (ESI+) for C<sub>45</sub>H<sub>63</sub>N<sub>9</sub>NaO<sub>10</sub> (M + Na)<sup>+</sup> calcd 912.4579, found 912.4590 (1.2 ppm error).

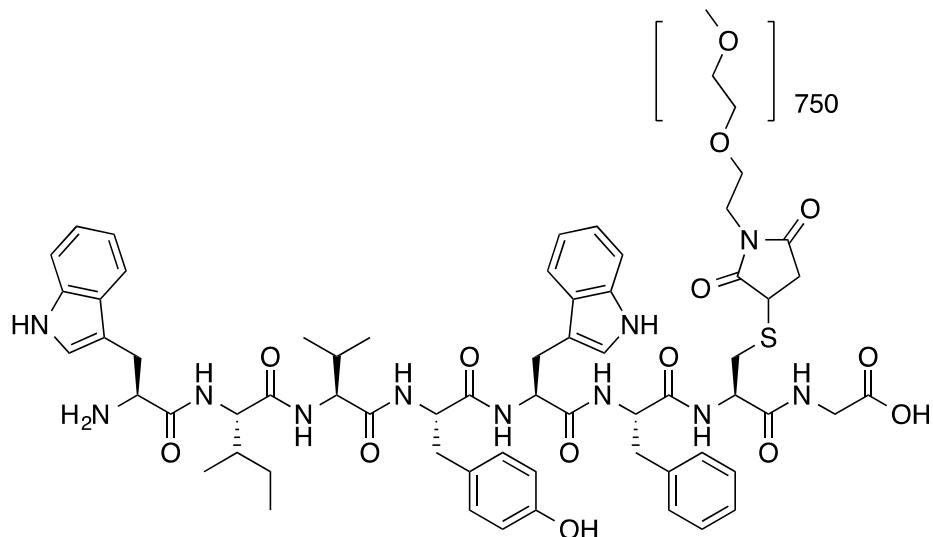
### 3.5.5.5 Crude linear-WLVYWFCG



**Figure 63:** Structure of linear-WLVYWFCG (5).

Peptide was synthesised as 2.6.1. HPLC (280 nm) 15.91 min; MS (ESI+)  $m/z$  (%) 1073.6 ( $(M + H)^+$ , 58.32), 1117.6 ( $(M + 45)^+$ , 100), 1161.6 ( $(M + 45 + 45)^+$ , 35.7).

### 3.5.5.6 Linear-WLVYWFC(PEG<sub>750</sub>)G

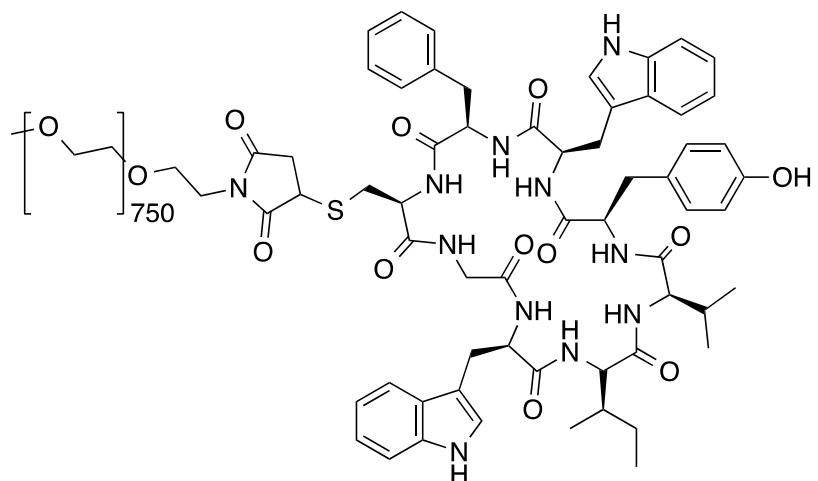


**Figure 64:** Structure of linear-WLVYWFC(PEG<sub>750</sub>)G (6).

Peptide 5 (8.6 mg, 125  $\mu\text{mol}$ ) was dissolved in 2.25 ml anhydrous DMF and added drop-wise into a stirred solution of maleimide-PEG<sub>750</sub> (103 mg, 137.5  $\mu\text{mol}$ ) in 2.0 ml anhydrous DMF and 15  $\mu\text{l}$  triethylamine. The reaction mixture was stirred at 20°C under argon overnight then the DMF removed *in vacuo* and precipitated in diethyl ether. The product was purified by RP-HPLC and lyophilised. HPLC purification provided 25.0 mg (10%) of 6 as a white solid. Analytical HPLC (280

nm) 21.83 min (95% pure); IR (neat) 3275, 3288, 1693, 1625, 1907  $\text{cm}^{-1}$ ; MS (ESI-TOF)  $m/z$  (%) 674.7 ( $\text{M} + 3\text{H}$ ) $^{3+}$ , 956.5 ( $\text{M} + \text{Na} + \text{H}$ ) $^{2+}$ .

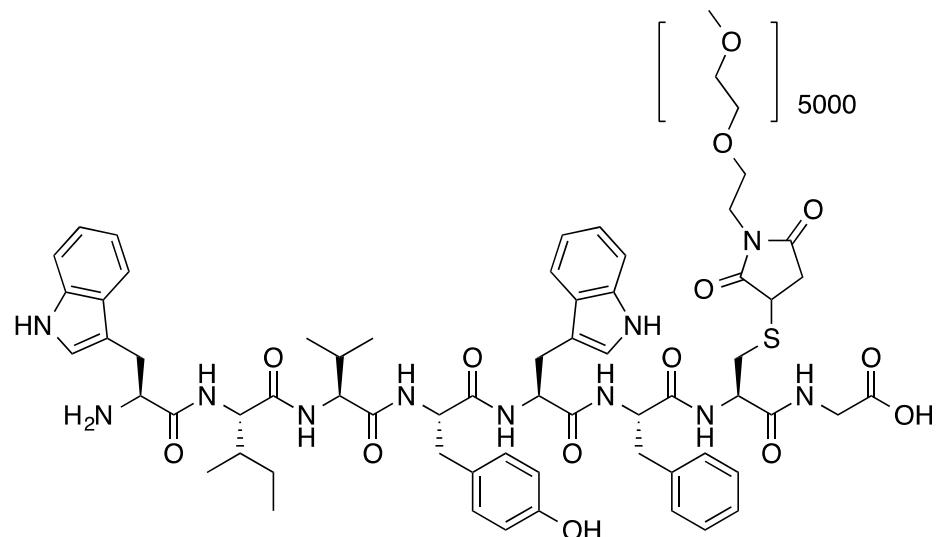
### 3.5.5.7 Cyclo-C(PEG<sub>750</sub>)GWLVYWF



**Figure 65:** Structure of cyclo-C(PEG<sub>750</sub>)GWLVYWF (7).

Peptide 6 (10 mg, 5.1  $\mu\text{mol}$ ) was dissolved in a solution of 10.0 ml anhydrous DMF with HOBr (4.1 mg, 30.6  $\mu\text{mol}$ ) and EDC (2.3 mg, 15.3  $\mu\text{mol}$ ). The reaction mixture was stirred at 20°C under argon overnight then the DMF removed *in vacuo* and precipitated in diethyl ether. The product was purified by RP-HPLC and lyophilised. HPLC purification provided 5 mg (53%) of 7 as a white solid. IR (neat) 3283, 2872, 1657, 1096  $\text{cm}^{-1}$ ; MS (ESI-TOF)  $m/z$  (%) 923.0 ( $\text{M} + 2\text{H}$ ) $^{2+}$ .

### 3.5.5.8 Linear-WLVYWFC(PEG<sub>5000</sub>)G



**Figure 66: Structure of linear-WLVYWFC(PEG<sub>5000</sub>)CG (8).**

Peptide 5 (8.6 mg, 125 µmol) was dissolved in 2.25 ml anhydrous DMF and added drop-wise into a stirred solution of maleimide-PEG<sub>5000</sub> (687.5 mg, 137.5 µmol) in 2.0 ml DMF, 100 µl triethylamine and 100 µl water. The reaction mixture was stirred at 20°C under argon overnight then the DMF removed *in vacuo* and precipitated in diethyl ether. The product was purified by RP-HPLC and lyophilised. HPLC purification provided 8 mg (1%) of 8 as a white solid.

### 3.5.6 *In vitro assays for PspA inhibitor screening*

#### 3.5.6.1 *Electron microscopy*

Samples of purified His<sub>6</sub>-tagged PspA were negatively stained with 2% (w/v) uranyl acetate on 50% carbonyl formvar grids, before imaging on a Hitachi H7000 Transmission Electron microscope operated at 80 kV with magnifications ranging from x 50,000 to x 200,000.

#### 3.5.6.2 *Isothermal titration calorimetry*

Experiments were carried out using a Micro-ITC calorimeter (Microcal) at 25°C. Multiple injections of inhibitor (either *cyclo*-SGWSIVFI or *cyclo*-SGWLVYWF)

were made into the cell containing MBP-tagged PspA (monomeric) under continuous stirring. Data was analysed using Microcal Origin software.

### 3.5.6.3 *Microscale thermophoresis*

Microscale thermophoresis (MST) experiments were performed on a Monolith NT.115 system (NanoTemper Technologies) using 100% LED and 20% IR-laser power. Laser on and off times were set at 30 seconds and 5 seconds, respectively. MBP-tagged PspA was labelled with NT647 (NanoTemper Technologies) and used at a final concentration of 80 nM. A two-fold dilution series was prepared for the unlabelled *cyclo*-SGWLVWYF and *cyclo*-SGWSIVFI in MST-optimised buffer (Table 40). Samples were filled into hydrophilic capillaries (NanoTemper Technologies) for measurement. The final concentrations of the peptides ranged from 5 mM to 152 nM.

**Table 40: Composition of MST-optimised buffer.**

REAGENTS	AMOUNTS	FINAL CONCENTRATION
Tris-base	50 ml of 1 M solution	50 mM
Sodium chloride	8.55 g	150 mM
MgCl <sub>2</sub>	10 ml of 1 M solution	10 mM
Tween-20	0.5 ml	0.05%
BSA	500 mg	0.5 mg/ml
Sterile deionised water	Up to 1000 ml	
Adjust pH to 7.6 with 1 M HCl		



## 4 Investigating anthrax toxin internalisation as a potential antimicrobial target

Chapter 3 describes the use of a bacterial RTHS and the SICLOPPS screening platform to identify two cyclic peptide sequences that potentially disrupt the oligomerisation of PspA. Unfortunately, the Psp response in *B. pseudomallei* was shown not to be essential for virulence by collaborators at DSTL, and the activity of these inhibitors was not successfully confirmed in either an *in vitro* or *in vivo* assay. Therefore, this chapter focuses on targeting *B. anthracis* use of toxins in the infection process causing cell death; a process essential for virulence.

Toxin production occurs within most pathogenic bacteria and is the route by which disease progresses. As briefly described in Section 1.1, there are many examples of pathogens using toxins to negate host function. This was exemplified by *C. difficile* that expresses toxins A and B. These cause diarrhoea in infected patients and so their activity provides a good target for the development of novel therapeutics.<sup>296</sup> Similarly, *Corynebacterium diphtheriae*, the causative agent of diphtheria, produces a toxin that has two subunits (A and B) joined by a disulphide bond. Subunit B binds to the host's surface and allows internalisation of subunit A. Once internalised subunit A inhibits translation by inactivating the eukaryotic elongation factor-2 protein via modification of the modified histidine, diphthamide, with NAD<sup>+</sup>.<sup>297</sup> These mechanisms are not essential for the pathogen's ability to survive, but by inhibiting these processes the pathogen's ability to cause disease is attenuated. Targeting virulence factors has the advantage of reducing the selection pressure on the bacteria from antibiotics, theoretically reducing the susceptibility to develop drug resistance. Consequently, mechanisms involving toxin production or their function provide an avenue for the development of novel broad-spectrum antibiotics.

Specifically, this chapter focuses on the virulence of *B. anthracis* that uses three toxins to hinder the host's ability to initiate the immune response. These toxins are the protective antigen (PA), edema factor (EF) and lethal factor (LF).<sup>298,299</sup> The formation of AB subunits between PA and either EF or LF, similar to those observed

for both *C. difficile* and *C. diphtheriae*, encourages the internalisation of the toxins into the cell where they can mediate the host's native immune processes.

The aim of this project was to identify inhibitors of the heterodimerisation of PA and the mammalian receptor using a combination of a bacterial RTHS and SICLOPPS. Having identified inhibitors of this PPI essential for virulence the inhibitors were tested *in vitro* and modified to identify a more potent inhibitor.

## **4.1 Introduction: *Bacillus anthracis* and anthrax**

Anthrax is a disease of two sides: on the one hand it has been instrumental in the advancement of microbiology medicine; and on the other has been used as a bioweapon. For example, in 1879 it was key in the formulation of Koch's postulate that draws the link between a pathogen and the disease.<sup>300</sup> *B. anthracis* was the first bacterium associated with causing a disease.<sup>300</sup> Additionally, anthrax paved the way for modern immunisation as the use of an attenuated strain was successfully administered in the vaccination of domestic animals against the disease.<sup>301</sup> Nevertheless, until the twentieth century anthrax was a common disease amongst animals and people and has been used to cause atrocities worldwide on several occasions. This was exemplified by the 2001 terror attack via the US postal system, and the deliberate release of anthrax spores in Kameido, Japan in 1993.<sup>302,303</sup> Consequently, the scientific community has consistently ranked the use of anthrax spores as the leading biological threat currently available.<sup>304</sup> It is therefore essential to have a molecular understanding of the mode of action of the disease and its progression, to allow the development of novel antimicrobials.

### ***4.1.1 Pathogenesis and clinical manifestations***

*B. anthracis* is the etiological agent of anthrax.<sup>300</sup> Its name *anthracis* is derived from the Greek word for coal, anthrax, and refers to the large, black skin lesions typical with cutaneous anthrax. *B. anthracis* is a large rod-shaped, spore-forming Gram-positive bacterium and like other members of the *Bacillus* genus can form endospores that are highly resistant, able to persist in environmental extremes for decades. These spores enter the body via inhalation, ingestion or a cutaneous route.

Once in the body they are phagocytised by macrophages, where due to their resistive nature they germinate yielding vegetative bacteria that lyse the cell. The bacteria consequently invade the lymphatic system and blood stream, initiating septicemia and toxemia, ultimately causing the death of the host. Once the bacteria have exhausted the host's supply of nutrients, sporulation is activated and the spores are released into the environment where they lay dormant, waiting for a new host.<sup>305</sup>

Of the three routes of infection, cutaneous anthrax is the most common form of the disease, associated with a high survival rate and characterised by an edematous-necrotic lesion that develops into skin necrosis. In contrast, infection by inhalation has a mortality rate close to 95%, with rapid onset of respiratory symptoms that progress to systemic failure. Ingestion is the typical route of infection in domestic livestock whilst they are grazing. Infections by ingestion and inhalation have high mortality rates and easy distribution for the use of *B. anthracis* as a bioweapon. Accordingly, the development of an antimicrobial that can halt disease progression and will not be susceptible to resistance is essential.

Regardless of the mechanism by which the disease manifests itself, the progression of the disease remains constant. The process is dependent on a tripartite lethal combination, consisting of PA, LF and EF.

#### **4.1.2 *The Bacillus anthracis* toxins and their entry into the host cell**

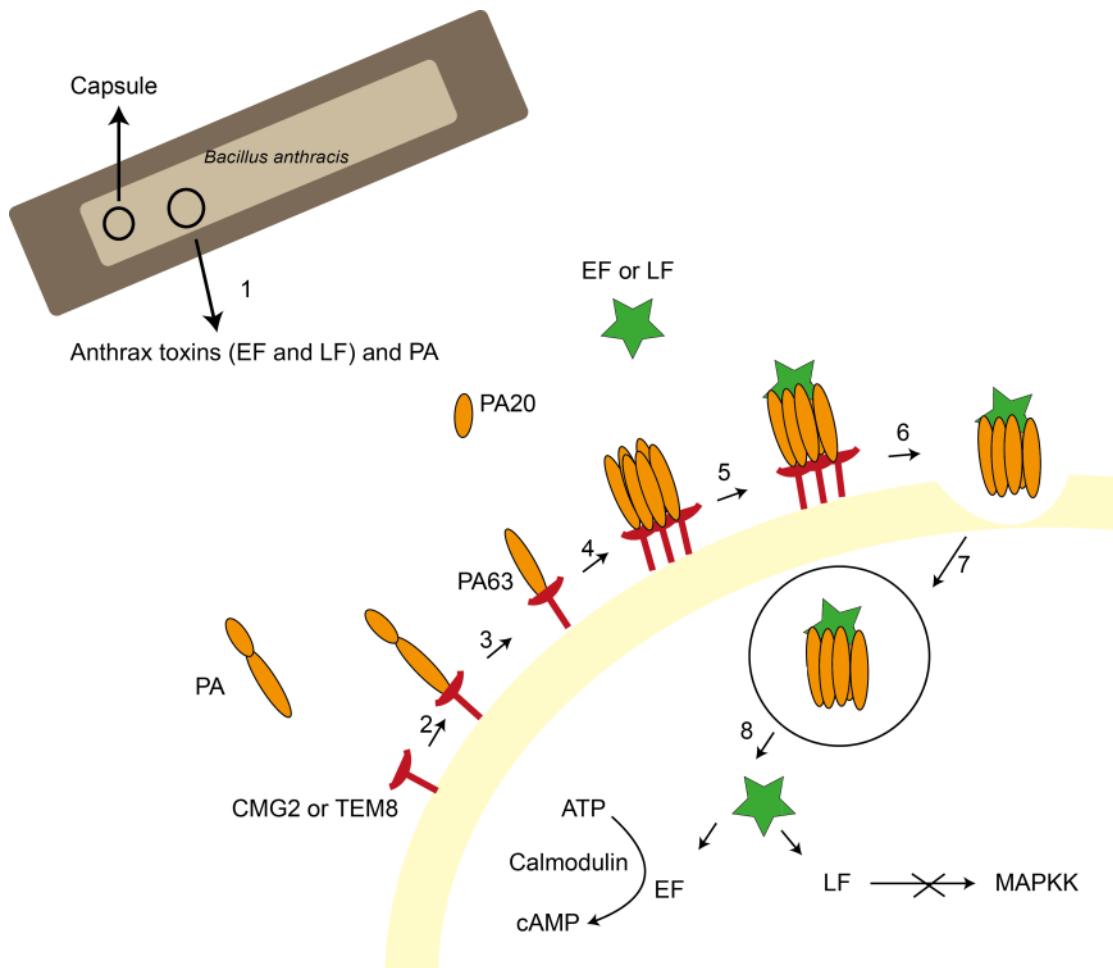
Like many pathogenic bacteria, *B. anthracis* harbours two virulence plasmids, pXO1 and pXO2.<sup>306</sup> These plasmids include the genes *cya*, *lef* and *pag* that encode the three anthrax toxins: EF, LF and PA, respectively.<sup>307-309</sup> The virulence plasmids also encode genes involved in the synthesis of the poly-D-glutamyl capsule, a protein capsule composed of poly-D-glutamic acid. They contribute to the inhibition of phagocytosis of the vegetative bacteria (i.e. *capA*, *capB*, *capC* and *dep*). EF, LF and PA are toxins involved in *B. anthracis* pathogenicity; they form two-component protein complexes comprising of A and B subunits similar to those described in Section 1.1 and 3.1.1 for shiga and diphtheria toxins. Subunit A is typically the active protein involved in stimulating a response, while subunit B is a binding-protein that mediates the translocation of subunit A into the host cytosol. PA is the

central component of the anthrax toxins and represents subunit B, facilitating passage into the cell of subunit A (i.e. either EF or LF).

Toxin challenge begins when full-length PA (PA<sub>83</sub>) binds to a surface receptor, either capillary morphogenesis gene 2 (CMG2) or tumour endothelial marker-8, (TEM-8) that are described in further detail in Section 4.1.5. Bound PA<sub>83</sub> is then proteolytically activated by the pro-protein convertase furin.<sup>310</sup> This cleaves a 20 kDa subunit, leaving truncated PA (PA<sub>63</sub>) that forms a heptamer, (PA<sub>63</sub>)<sub>7</sub>, which remains bound to the receptor. The formation of this PA receptor exposes a large hydrophobic surface in domain 1 of PA, via the β<sub>12</sub>-β<sub>13</sub> hairpin, with a positive charge, which is referred to as the pre pore.<sup>311,312</sup> LF or EF bind to the pre pore with a high affinity ( $K_d = 1 \text{ nM}$ )<sup>313,314</sup> via their conserved N-terminus.<sup>315</sup> This conserved region has seven amino acid residues that form a complimentary flat surface with a net negative charge to bind to the pre pore. Whether EF and LF can bind to only oligomeric PA or monomeric PA has divided opinion. Originally it was proposed that three molecules of LF could only bind PA in a heptameric form binding to the region spanning the intersection between two PA<sub>63</sub> units.<sup>316</sup> However, a subsequent study using the same PA<sub>63</sub> mutants D152K, K199E, R468A and R470D (that prevent oligomerisation binding) shows the ability of LF to bind to monomeric PA using native-PAGE and SEC multi-angle laser scattering to monitor the formation of the complexes.<sup>317,318</sup> Nevertheless, oligomeric PA is required for anthrax toxin internalisation.

After binding, the toxin complexes are translocated into the cytosol via a pH-dependent clathrin-mediated endocytosis. The change in pH causes a transformation of the pre pore into a pore, simultaneously unfolding EF and LF (Figure 67).<sup>319</sup> Specifically during the endocytosis, phosphorylation of the receptor is instigated by a conformational change in the transmembrane and cytosolic domains of the receptor.<sup>320</sup> This causes the receptor to be ubiquitinated triggering the formation of the clathrin coat.<sup>320</sup> The PA<sub>63</sub> pre pore converts to its pore-forming conformation and inserts through clathrin-coated pits into the membrane of intraluminal vesicles within early endosomes.<sup>321</sup> LF (and EF, although not shown experimentally) are then

translocated into the lumen of these vesicles and trafficked via multivesicular bodies into late endosomes, where the toxins are released into the cytosol.<sup>314,320,321</sup>



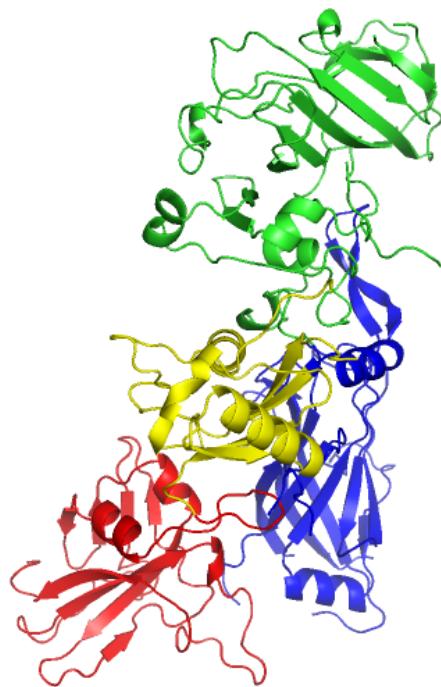
**Figure 67: Diagram illustrating the internalisation of anthrax toxins LF and EF utilising the interaction between PA and a cell surface receptor.** Anthrax toxins are expressed (1). PA binds to the receptor (2) and is cleaved by a furin to produce the PA<sub>63</sub> subunit (3) that subsequently forms a heptamer (4). This complex can then recruit EF and LF (5) and under the control of pH internalise the toxins (6) and (7) where they can modulate cell-signalling pathways involved in the host's immune response (8).

Disease progression requires the internalisation of two toxins, EF and LF, via the interaction between PA and its receptors CMG2 and TEM-8. The structural knowledge about how the toxins interact and pH mediates the entry of anthrax toxins into the host cell, provides a good basis for the rational design of inhibitors. Design

of a specific inhibitor that could target receptor-toxin PPI or AB-subunit association would halt disease progression. In addition, the two toxins, LF and EF, act on a variety of targets once localised in the host cytoplasm, causing a synergistic attack on the host's immune system, described further in Section 4.1.6. Their activity could provide a second line of defence against anthrax infection, and provide viable targets for the development of antimicrobials.

#### ***4.1.3 Structural properties of protective antigen and the heptameric prepore***

PA is the central component of the anthrax toxin internalisation mechanism and its structural properties have been investigated using crystallography, NMR and mutational studies.<sup>311,322-326</sup> A 2.1 Å crystal structure revealed how PA forms a long, flat molecule with dimensions 100 x 50 x 30 Å, consisting of four domains.<sup>322</sup> Domain 1 (residues 1-258) consists of a β-sandwich, with several small helices and a pair of adjacent calcium ions. The conserved cleavage site (RKKR, residues 164-167) is also located in this domain;<sup>310</sup> Domain 2 (residues 259-487) has a β-barrel core that lines the lumen in its heptameric form, with a large flexible loop involved in membrane binding, between 2β<sub>2</sub>-2β<sub>3</sub> (residues 302-323) strands.<sup>323,324</sup> Domain 3 (residues 488-595) has a four-stranded mixed β-sheet, two smaller sheets and four helices; it has been implicated in the formation of the heptameric prepore. Finally, there is domain 4 (residues 596-735) that plays a key role in receptor binding and consists of an initial hairpin and helix, followed by a β-sandwich (Figure 68).<sup>322,327</sup>

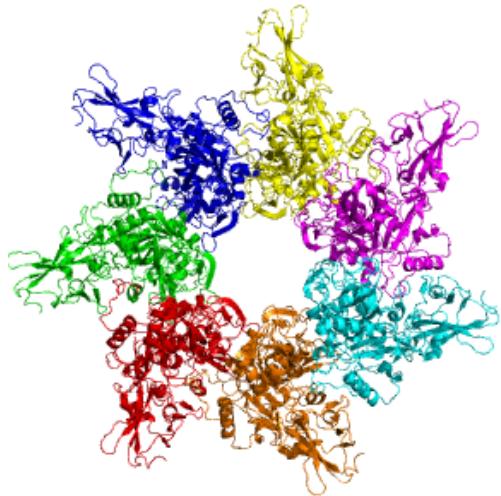


**Figure 68: Crystal structure of PA<sub>83</sub>.** Generated in PyMol from PDB 1ACC, Domain 1 (1-258) is green, domain 2 (259-487) is blue, domain 3 (488-595) is yellow and domain 4 (596-735) is red.<sup>322</sup>

After binding to the receptor, the N-terminus of PA<sub>83</sub> is enzymatically cleaved leading to the association of seven subunits to form the prepore that inserts into the membranes to form the translocation channel.<sup>310,328</sup> The cleavage of the 20 kDa subunit is essential to allow heptamerisation, as when PA<sub>63</sub> is substituted with PA<sub>83</sub> there is evidence of steric hindrance blocking oligomerisation.<sup>322</sup> The function of the 20 kDa subunit has been separately shown to affect human peripheral blood leukocytes and induce apoptosis, even in the absence of other anthrax components. Further studies, however, are required to fully elucidate its function in virulence.<sup>329</sup>

A 4.5 Å crystal structure of the prepore showed a hollow ring, 160 Å in diameter and 85 Å high (Figure 69).<sup>322</sup> The prepore has been observed by electron microscopy forming the described ring-like species.<sup>330</sup> The central lumen has an average diameter of 20-35 Å and is polar and negatively charged, with the exception of a conserved phenylalanine (Phe427) located on a solvent-exposed loop.<sup>322,324</sup> Each PA<sub>63</sub> monomer contributes this phenylalanine residue to produce an  $\varphi$ -clamp, that aids

translocation of EF and LF by recognising a hydrophobic region on the toxin via hydrophobic interactions.<sup>331</sup>



**Figure 69: Crystal structure of (PA<sub>63</sub>)<sub>7</sub>.** Generated in PyMol from PDB 1TZO. Each PA<sub>63</sub> subunit is individually coloured.<sup>322</sup>

More recently, a PA octamer has been resolved by x-ray crystallography and mass spectrometry. This octamer is stable at physiological pH and can bind and translocate EF and LF into the cytosol with similar efficiencies observed for heptameric PA<sub>63</sub>.<sup>332,333</sup> Analysis of Chinese hamster ovary (CHO)-R1 cells, revealed a mixture of both octameric (20-30%) and heptameric PA<sub>63</sub>. The two structures are hypothesised to fulfil the same role under different conditions. Reputedly, the octamer is more stable and so can form prior to binding to the receptor, for example, in the blood plasma, while the heptamer associates once PA<sub>63</sub> is bound to the receptor.<sup>332</sup>

#### **4.1.4 Critical residues of protective antigen essential for binding and translocation of the anthrax toxins**

The interpretation of crystal studies, has aided the understanding of how PA<sub>63</sub> may interact with EF and LF, and form the heptamer to bind to the receptor. Mutational studies, however, have given further insight and confirmed which residues are critical for function. Mapping of PA<sub>63</sub> by introducing dominant-negative mutations via scanning mutagenesis<sup>325,334</sup> or substitution of the loop between 2β<sub>2</sub>-2β<sub>3</sub> of

domain 2 with the residues from the iota- $\beta$  toxin (the *Clostridium perfringens* homologue) confirmed that all four domains are required to mediate toxicity.<sup>334</sup> With dominant-negative mutations, the mutated PA<sub>63</sub> is incorporated into the heptameric structure, still retaining its ability to bind to LF and oligomerise and yet internalisation is inhibited.<sup>325,334-336</sup> Using the domain 2 homologue substitution, no toxicity was observed (even at concentrations of 100 µg/ml) compared to the wild-type PA<sub>63</sub> where 50% of cells were lysed at 0.04 µg/ml. Similarly using a diphtheria toxin-A (DTA) LF recombinant protein (LF-DTA) fused to a protein synthesis inhibitor,<sup>337</sup> no protein synthesis inhibition was observed. Conversely, in the presence of wild-type PA<sub>63</sub> 90% of protein synthesis was inhibited; illustrating how domain 2 is critical in the internalisation of EF and LF. In addition, dominant-negative mutations (for example K397D and D425K) have also been successfully shown to be more immunogenic than PA as a vaccine, highlighting the potential for rationally designed molecules for the inhibition of *B. anthracis* virulence.<sup>338,339</sup>

As well as internalisation, there are many residues critical for binding to the receptor. The loop in the 2β<sub>2</sub>-2β<sub>3</sub> regions in domain 2 is disordered when PA<sub>83</sub> is in its monomeric form yet it inserts into the membrane as a hairpin to generate a 14-stranded β-barrel pore during receptor binding.<sup>311,324</sup> In addition, domain 4 has been implicated with binding to the receptors and to the toxins. Deletions of the last 3, 5, 6, 12 or 14 residues from the C-terminal shows that the presence of the last 12-14 residues are essential for the internalisation of LF into JK774.1 cells and EF binding. Deletions of fewer residues showed a decrease in PA binding and toxin binding; nevertheless, it was not as significant as the effect observed for 12-14 residue deletions.<sup>340</sup> Varughese *et al.* (1999) later proposed that the involvement of the C-terminal in binding was due to instability in conformation in its absence.<sup>327</sup> Furthermore, their mutagenesis studies of a small solvent-exposed loop in domain 4 affected the ability of PA<sub>63</sub> to bind to the receptor and stimulated a toxic effect. Rosovitz *et al.* (2003) contributed to elucidating further the involvement of the small loop in domain 4, by treating the RAW264.7 cell line with mutated PA and LF.<sup>341</sup> If binding was inhibited then the cells were not lysed by LF internalisation, but if the mutant retained its ability to bind then the cells were lysed. This effect was monitored by an MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) assay. The results revealed that Asp683 was critical for binding and

toxicity, with a 1000-fold decrease in activity in its absence. Moreover, Y681, N682, P686 showed a significant decrease in toxicity by 10-100-fold. The same pattern was observed using recombinant LF-DTA and monitoring their effect on CHO-R1 cells.<sup>341</sup>

Structural studies of PA have established that each domain plays a critical role in either binding or toxin internalisation. Using cell-based and molecular genetic studies two cellular receptors for the docking of PA have been identified, each mediating the delivery of the anthrax toxins, LF and EF, under different conditions.

#### ***4.1.5 Identification of receptors for toxin internalisation***

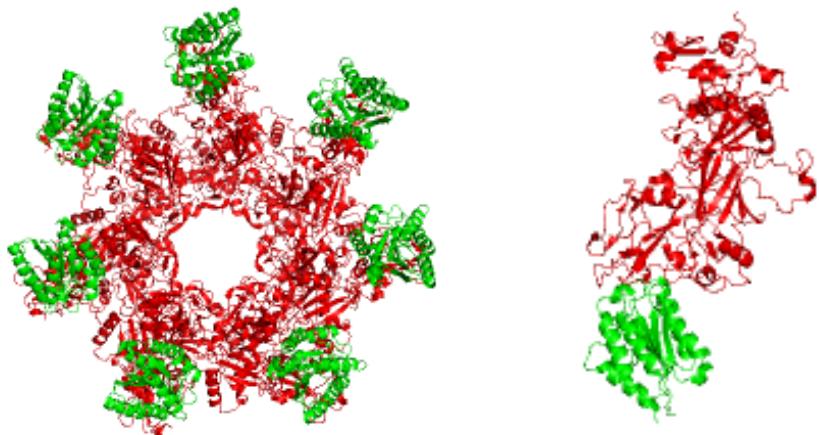
For the internalisation of anthrax toxins, LF and EF, PA binds to a cell surface receptor on the host, converting to a pore and mediating toxin translocation. Two cell surface receptors that bind PA<sub>83</sub> have been identified: ANTXR1 (TEM-8) and ANTXR2 (CMG2).<sup>292,342</sup> TEM-8 and CMG2 are both expressed in vein endothelial cells where they play an important role in angiogenesis.<sup>343,344</sup> TEM-8 influences cell migration and tubule formation and CMG2 promotes endothelial cell proliferation. In addition to the role of CMG2 in angiogenesis, it has also been implicated in juvenile hyaline fibromatosis and infantile systemic hyalinosis, through mutations and insertions in the von Willebrand factor A domain (VWA) and transmembrane domain.<sup>345,346</sup> Similarly, mutations in TEM-8 have been linked with infantile haemangioma, causing localised and rapid growth of angiogenesis.<sup>347</sup> Nevertheless, the physiological roles of TEM-8 and CMG2 in the host are poorly understood. *B. anthracis* has hijacked these receptors to facilitate disease progression; theoretically any inhibitors for anthrax that target and bind CMG2 or TEM-8 may serendipitously provide a novel inhibitor of angiogenesis and *vice versa*.<sup>289,291</sup>

Both CMG2 and TEM-8 share 40% overall sequence homology on the amino acid level and 60% in the extracellular domain referred to as the VWA domain. This domain contains a metal ion dependent adhesion site (MIDAS) motif (DxSxS...T...D, where x is any amino acid) that is required for ligand binding; binding of PA to the VWA is dependent on the presence of a divalent cation.<sup>292</sup> There is much debate in the literature about which divalent cation enhances binding with PA. An ELISA established by Dawson (2009) described how in the absence of

a divalent cation little binding was observed; however, 1 mM Mg<sup>2+</sup> improved binding 9.8-fold, and Zn<sup>2+</sup> enhanced binding 7.4-fold.<sup>290</sup> This was supported by further ELISA experiments, FRET experiments and SPR that showed binding was enhanced four- to five-fold in the presence of 1 mM Mg<sup>2+</sup> compared to Ca<sup>2+</sup> cations.<sup>288,342,348,349</sup> In the initial reporting of CMG2 Ca<sup>2+</sup> ions were proposed to enrich binding, with a greater effect than Mg<sup>2+</sup>.<sup>342</sup> In contrast, crystallographic evidence supports the hypothesis that Mg<sup>2+</sup> is better at coordinating to the MIDAS in the VWA domain of CMG2. Ca<sup>2+</sup> was not favoured due to the presence of a serine (S54) and threonine (T118) in the binding pocket.<sup>350</sup> The choice of divalent ion in any *in vitro* assay used in this study is critical and a variety should be tested to ensure that optimal binding is being achieved.

Originally, it was proposed that anthrax toxin internalisation occurred irrespective of the cell-line or receptor; however, recent research shows that the receptor type has a profound influence on the binding of PA and toxin entry.<sup>317</sup> The affinity of the TEM-8 expressing cells and PA was K<sub>d</sub> = 9.5 nM, whereas the affinity for CMG2-expressing cells and PA was three-fold higher.<sup>351</sup> Similar trends were observed during *in vitro* binding studies, with K<sub>d</sub> values of 170 pM and 1.1 μM for CMG2 and TEM-8, respectively.<sup>288</sup> Overall PA has a higher affinity for CMG2.

Much debate has surrounded the cause of the difference in affinity and the pH dependent prepore to pore formation in the last ten years. The difference in binding was proposed to be due to their structure and the interaction with PA. Two crystal structures of the PA heptamer bound to the VWA of CMG2 were determined at 4.3 Å and 2.5 Å, both showing the interaction between the MIDAS and D683 of PA<sub>63</sub>.<sup>293,352</sup> Additional interactions were observed between four loops in CMG2 (residues 52-56, 87-88, 111-122 and 152-158) and four loops of PA (residues 340-348, 654-662, 681-688 and 714-716) and between the 2β<sub>2</sub>-2β<sub>4</sub> loops of PA previously described (Figure 70).<sup>311</sup>



**Figure 70: Crystal structure of PA<sub>63</sub> and CMG2.** Generated in PyMol from PDB 1TZN, PA<sub>63</sub> is represented in red while the receptor CMG2 is shown in green.<sup>311</sup>

More recently a crystal structure for TEM-8 (Figure 71) was solved and compared to CMG2.<sup>353</sup> The structures were highly conserved based on sequence alignment: importantly the residues that differed were located on the binding interface of CMG2.<sup>353</sup> The residues were divided into four regions of TEM-8: domain 1 (residues 56 and 57) located in helix  $\alpha$ 1 (this corresponded to the region of CMG2 that interacts with domain 2 of PA); domain 2 (residues 87 and 88) located in the  $2\beta_2$ - $2\beta_3$  loop, the region in the crystal structure that was not well aligned; domain 3 (residues 113–117, Tyr119, His121, Glu122, and Glu125) located in the  $\alpha_2$ - $\alpha_3$  loop; and domain 4 (residues 152–156 and Tyr158) re-located in the  $\beta_3$ - $\beta_4$  loop. Domains 2–4 have been shown to interact with domain 4 of PA.<sup>293,295</sup> Therefore it could be hypothesised that these non-conserved residues in the PA-binding interface are responsible for the striking differences in the pH dependence and receptor function of TEM-8 and CMG2.



**Figure 71: Crystal structure of hexameric TEM-8.** Generated in PyMol from PDB 3N2N.<sup>353</sup>

The high affinity for CMG2 has been proposed to be due to its conformation.<sup>352</sup> The receptor/PA binding resembles the  $\alpha$ -integrin: these exist in two conformations, open and closed, having high and low affinities for their ligands, respectively.<sup>354</sup> The difference in binding is attributed to interactions involving the MIDAS, including the aspartic acid and the threonine that allows the switch from an indirect metal coordination in the closed state to a direct in the open state.<sup>354</sup> TEM-8 has been shown to form both the open and closed conformation,<sup>353,355</sup> with a loss of the T118 residue, reducing the receptor activity by 100-1000-fold, and intimating that the open form is essential for PA binding. It is proposed that CMG2 is fixed in the open conformation, leading to its high affinity, as the closed form has not been observed.<sup>311</sup> In contrast, Scobie *et al.* (2006) indicated that the T118 residue was not required for CMG2 binding to PA, showing that it is not required for the formation of an open-like conformation. They propose instead that the higher binding affinity associated with CMG2 overcomes the requirement for an open-like configuration, yet the exact mechanism requires further investigation.<sup>356</sup>

Irrespective of the reason for the difference in binding of PA to TEM-8 and CMG2,

the latter has been shown to be the dominant receptor used in anthrax toxin entry. Mouse studies, using knockout mutants, show how lethal toxin internalisation is predominantly facilitated by CMG2 with TEM-8 playing only a minor role.<sup>357,358</sup> This is likely to be due to the higher affinity of PA for CMG2 than for TEM8 although the precise mechanism behind this remains under debate.<sup>351,357-359</sup>

As described in Section 4.1.2, the current model of toxin entry into a host cell is the binding of PA<sub>83</sub> to a cellular receptor, its cleavage and heptamerisation followed by recruitment of the toxins, LF and EF. Final internalisation of these toxins allows the manipulation of the host's signalling pathways and disease progression. The identification and characterisation of the two receptors, TEM-8 and CMG2, and their interaction with PA has led to a greater understanding of this pathway. The final component to understand is how EF and LF cause lethality once in the cytosol.

#### ***4.1.6 The manipulation of host signalling pathways by anthrax toxins***

*B. anthracis* produces a number of virulence factors that are involved in the pathogenesis and establishment of the disease. Plasma from infected animals produced lethality in uninfected animals.<sup>360,361</sup> This was later attributed to toxins: edema toxin (ET), formed from the association of EF and PA causing localised skin edema and, at higher doses, lethality. For this to occur, however, the involvement of PA is essential, similarly in lethal toxin (LT); the lethality observed is dependent on the association of PA and LF. The synergistic relationship between the toxins has been demonstrated using mutant mice and macrophages, where in the absence of PA, no effects were observed.<sup>362,363</sup> The internalisation of EF and LF via the interaction between PA and its receptor is required to localise the toxins to interfere with cellular response to bacterial infection, which is primarily the host's immune response.

LF is a metalloprotease that cleaves mitogen-activated protein kinase kinases-2 (MAPKK) involved in signalling pathways.<sup>364</sup> LF consists of four domains: domain 1, the PA-binding domain,<sup>365,366</sup> and the C-terminus (domains 2-4) that forms the catalytic domain with a long deep groove (about 40 Å long) containing the active site that binds the N-terminal tail of MAPKK,<sup>364</sup> which once bound is cleaved.<sup>367-370</sup> Mitogen-activated protein kinases (MKKs) are involved in the three-component

phosphorylation cascade and respond to stresses, growth and cytokines. These signals activate MAPKK that phosphorylate and activate MKKs by binding to the same site as LF.<sup>371</sup> MKKs in turn activate the MAPKs signalling pathways like p38 and ERK1/2 (extracellular-signal-regulated kinase 1/2). The inactivation of MAPKK prevents the initial phosphorylation and leads to a shutdown in signalling pathways and gene expression. Host survival is dependent on the MAPK pathways; blockage of which leads to a suppressed immune response, induction of macrophage apoptosis<sup>372</sup> and suppression of dendritic cells function.<sup>370,373</sup>

The second toxin is EF, a calmodulin or calcium-dependent adenylyl cyclase that increases intracellular cyclic adenosine monophosphate (cAMP) levels.<sup>374</sup> Increased levels of cAMP disrupt the homeostasis of the cell, like inactivation of MAPKK, eventually interrupting cell signalling. The activity of EF is dependent on the presence of calcium and calmodulin,<sup>375</sup> with the presence of calmodulin increasing activity 1000-fold.<sup>376</sup> Similar to LF, it has a PA-binding N-terminal domain, and a catalytic domain; however, it has an additional C-terminal helical domain. To accommodate calmodulin, the catalytic domain undergoes a 15 Å translation and a 30° rotation away from the catalytic core, forming a clamp that encompasses calmodulin.<sup>375,376</sup>

*B. anthracis* targets the host's immune response after toxin internalisation; hindering neutrophil function, phagocytosis and superoxide production, as well as dendritic cell function.<sup>363,372</sup> LT and ET interfere with these processes, by increasing the levels of cAMP and inactivating the signalling pathways essential for the expression of pro-inflammatory mediators, oxidative burst and cell migration.<sup>374</sup> ET and LT represent good prophylactic and therapeutic targets as they contribute to the pathogenicity of anthrax disease. Both are virulence factors that suppress the host's immune response.

#### **4.1.7 Current treatments and prophylactics**

The postal anthrax attacks in the US in 2001 heightened the interest in the development of a novel therapeutic for anthrax disease.<sup>302</sup> The development of anthrax inhibitors has been well recorded. Since the use of an attenuated *B. anthracis* strain by Louis Pasteur in the vaccination of live animals in 1881 there has been a

prophylactic treatment available for *B. anthracis* infection.<sup>301</sup> Current vaccinations for livestock use the Sterne strain, a live avirulent vaccine. Similar live vaccines are available for human use in some countries; however, due to the concern of residual virulence, countries like the UK and the US do not approve the use of such vaccines in humans. Consequently, the use of partially purified PA of a non-encapsulated *B. anthracis* strain, whereby the bacteria only harbours the pXO1 plasmid, is approved. These treatments are a prophylactic and require long-term administration making them costly, so they are reserved for patients at repeated occupational exposure risk.

Yet, the use of live vaccines is redundant in cases of post-exposure. For these the vaccine is administered in conjunction with antibiotics like gentamicin, tetracycline, ciprofloxacin, erythromycin and doxycycline. This treatment occasionally results in a dormant disease that once treatment has ceased reverts to a rapidly progressive disease. Furthermore, resistance to some antimicrobials, for example cefuroxime, ceftotaxime, ceftizoxime and ceftazidime, has already been observed.<sup>377</sup> In addition, with the potential use of anthrax as a bioweapon, the isolation of a resistant strain to ciprofloxacin or doxycycline is not unrealistic.<sup>378,379</sup>

The use of passive immunisation or antibodies has also been approved, for example raximacaub, with many of them targeting PA.<sup>380</sup> They function by mitigating different roles of PA including: the ability to bind to the receptor, LF and EF binding and prepore-pore formation.<sup>381-383</sup> They are good prophylactic treatments and post-exposure therapeutics, inhibiting disease progression allowing bacterial clearance when used in conjunction with an antibiotic. There are limitations in their uptake into the body and the production of an antibody therapeutic is very costly due to their size and complexity. Consequently, there are restrictions in their availability and distribution. Hence, diversification into the area of small molecules has been investigated to identify cheaper alternatives, that are commercially viable and more bioavailable.

The last ten years has seen a surge in the research of small molecules for their potential as anthrax disease antimicrobials. There are several examples of clinically approved drugs for other diseases that show anti-anthrax activity, and could be used in an emergency.<sup>289,384,385</sup> Nevertheless, a more potent and specific inhibitor is still

desirable. Current research has primarily focused on the activity of LF, its ability to interact with PA and enter the host cell. Current inhibitors were identified using a generic library in conjunction with either phage display,<sup>386,387</sup> a phenotypic screen of lethal toxin induced macrophage (RAW264.7 or J774A.1 cells) cell death<sup>388-392</sup> or a more novel approach using a peptide cleavage fluorescent assay.<sup>393</sup> In each case the activity of each isolated inhibitor was confirmed and its mode of action identified. Celestrol, a quinone methide triterpene, inhibited LT operation by inhibiting the proteasome pathway,<sup>390</sup> similar to the inhibitor (4-methoxy-2-[2-(5-methoxy-2-nitrosophenyl)ethyl]-1-nitrosobenzene, 4MNB) identified with an IC<sub>50</sub> of 18 μM.<sup>388</sup> In this screen a second, more potent, inhibitor (IC<sub>50</sub> = 56 nM) was also identified: the natural product toosendanin. Unlike 4MNB, it inhibited uptake of LF into the cytoplasm via an unknown mechanism.<sup>388</sup> Analysing the crystal structures of the inhibitor bound to LF, and confirmation by enzymatic assays probing the activity has also identified inhibitors that inhibit the proteolysis activity of MAPKK1 by LF; two of which had nanomolar activity.<sup>392,393</sup> One was based on a hydroxamate group and the other was a phenylfuran-2-ylmethylene rhodanine acetic acid derivative. Both bound to the active site of LF, blocking its ability to cleave MAPKK.<sup>392,393</sup>

The potency of the small molecule inhibitors that target binding of either PA to the receptor or EF/LF to PA has been greatly improved by making the inhibitor polyvalent.<sup>386,392,394</sup> A 7,500-fold increase in activity was observed for the peptide sequence HTSTYWWLDGAP, identified by phage display when it was attached to a scaffold like β-cyclodextrin (IC<sub>50</sub> of 150 μM to 20 nM). Furthermore, no adverse effects were observed in Fischer 344 rats challenged with ten-times the minimal lethal dose of PA and LF.<sup>386</sup> Interestingly, the active motif TYWWLD was again identified in a second phage display screen by Gujrati *et al.* (2005),<sup>395</sup> which was further developed into a polyvalent inhibitor using a β-cyclodextrin scaffold and shown to function by competitively binding to PA and blocking the binding of LF.<sup>391,396</sup> The polyvalent inhibitor showed more than 100,000-fold improvement in activity than the parent peptide. Treated Fischer 344 rats challenged with LF and PA showed prevention in the animals becoming moribund.<sup>391</sup>

The polyvalent approach was applied to targeting the interaction between PA and the receptor CMG2. Using phage display an inhibitor (AWPLSQLDSYN) was identified

---

that when bound to a liposome scaffold demonstrated more than a 50,000-fold increase in potency, to yield an inhibitor with an IC<sub>50</sub> of 40 nM.<sup>387</sup> The polyvalent inhibitor had the ability to bind to CHO-R1 cells expressing both CMG2 and TEM-8, suggesting that it can be used as a more generic inhibitor. Additionally, as PA is the only well characterised ligand of CMG2, the interaction between CMG2 and PA could be used to identify inhibitors of angiogenesis. Rogers *et al.* (2012) using a FRET-based high-throughput inhibitor screen identified tannic acid and *cis*-platin as inhibitors of the CMG2 and PA interaction. Instead of testing for activity in an anthrax-based *in vitro* assay, they confirmed binding to CMG2 by SPR and activity in an endothelial cell migration assay. Concerns over the toxicity of the molecules in a corneal pocket mouse study and the impurity of the compound were raised suggesting that caution should be taken when using this as definitive evidence that an inhibitor of CMG2 and PA interaction would inhibit angiogenesis, further research is required.<sup>289</sup> A similar study was carried out using TEM-8.<sup>291</sup> There are few examples of attempts to either rationally design or screen a small molecule library to disrupt the interaction between PA and either TEM-8 or CMG2, making it a good avenue for the future development of novel therapeutics.

#### **4.1.8 Protein-protein interactions within anthrax progression**

Individually the toxin components (i.e. PA, LF and EF) have no effect on their host. In contrast when they are combined to form the lethal toxin (PA and LF) and edema toxin (EF and PA) they cause cell death and tissue edema, respectively. Within the anthrax internalisation system there are several PPIs that could be targeted using the RTHS and SICLOPPS screening platform. Mourez *et al.* (2001) have already demonstrated that PPI can be successfully targeted using phage display to identify and isolate inhibitors of the formation of toxin complexes.<sup>386</sup>

Targeting the heterodimerisation of CMG2 and PA hinders disease progression, as this interaction mediates the entry of the toxins EF and LF. For this study, CMG2 was selected over TEM-8 as it has been shown to facilitate toxin internalisation more than TEM-8; CMG2 is also present in a variety of tissue-types and has a higher affinity for PA, as shown by biophysical binding studies and via cell-based assays.<sup>288,357</sup> Furthermore, another advantage to this target is related to the development of resistance. Resistance can occur naturally, but it can be introduced

intentionally for use as a more virulent strain as a bioweapon. Inhibiting the interaction of a ligand with host receptors used by pathogens represents a powerful strategy to overcome this problem because extensive alterations to the pathogen will be required to allow it to adapt to the change to a new receptor.

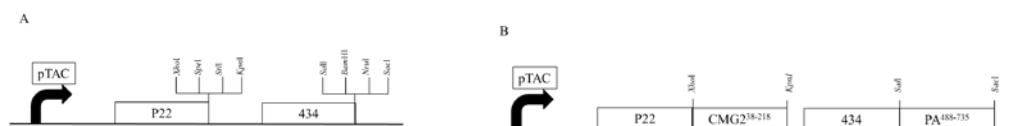


## 4.2 Results: Inhibitor screening of the dimerisation of the anthrax toxin PA and the mouse receptor CMG2

The PA of *B. anthracis* is an integral component in the mechanism of anthrax toxicity. It binds to the human receptors TEM8 and CMG2 and then internalises anthrax toxins EF and LF where they play a key role in the development of the disease. Therefore, an inhibitor of the heterodimerisation of CMG2 and PA would be expected to disarm *B. anthracis* preventing the progression of the disease. This Chapter uses a combination of a bacterial RTHS and SICLOPPS library screening to identify cyclic peptide inhibitors capable of disrupting the heterodimerisation of the human receptor CMG2 and the anthrax toxin PA.

### 4.2.1 Design and construction of the Anthrax RTHS

The Anthrax RTHS was constructed by Dr. F. Foranov, using the protocol described for the construction of the PspA RTHS (Section 3.2.1). The plasmid pTHCP14 designed by Horswill *et al.* (2004) for a heterodimeric interaction was used.<sup>80</sup> Like pTHCP16, pTHCP14 contains the DNA-binding domain of the wild-type 434-repressor, as well as an IPTG-inducible pTAC promoter. Additionally upstream of the wild-type 434-DNA-binding domain the mutant 434-DNA-binding domain possessing P22 binding specificity (referred to as P22) was located. Like pTHCP16 the pTHCP14 vector backbone was sourced from pMAL-c2x (New England Biolabs) ensuring ampicillin antibiotic resistance,<sup>80</sup> and contained two MCS downstream of the DNA-binding domain genes, as illustrated in Figure 72A.



**Figure 72: Representation of plasmid pTHCP14 constructed by Horswill *et al.* (2004).**<sup>80</sup> (A) P22 and 434-repressors are under the control of the IPTG-inducible pTAC promoter with downstream MCS. (B) Design to allow cloning and expression of C-terminal CMG2<sub>38-218</sub> and PA<sub>488-735</sub> recombinant proteins with the 434- and P22-repressors, respectively.

In order to identify inhibitors of the interaction between CMG2 and PA, the respective genes were cloned into the MCS downstream of each DNA-binding domain gene. The plasmid was designed to allow simultaneous expression of both P22-CMG2<sub>38-218</sub> and 434-PA<sub>488-735</sub>, upon addition of IPTG. As described in Section 4.1, mutagenesis, crystallographic and biophysical studies have focused on probing the interaction between CMG2 and PA and identifying key residues involved.

Crystal studies have shown that the interaction is dependent on the VWA domain of CMG2 and the second and fourth domain of PA.<sup>295</sup> The binding of the VWA of CMG2 with full length PA occurs with a 170 pM affinity, and has a slow dissociation rate constant of approximately  $10^{-5} \text{ s}^{-1}$ .<sup>288</sup> Therefore, for the Anthrax RTHS, the VWA domain of CMG2 (residues 38-218) was used alongside the third and fourth domains of PA (residues 488-735).

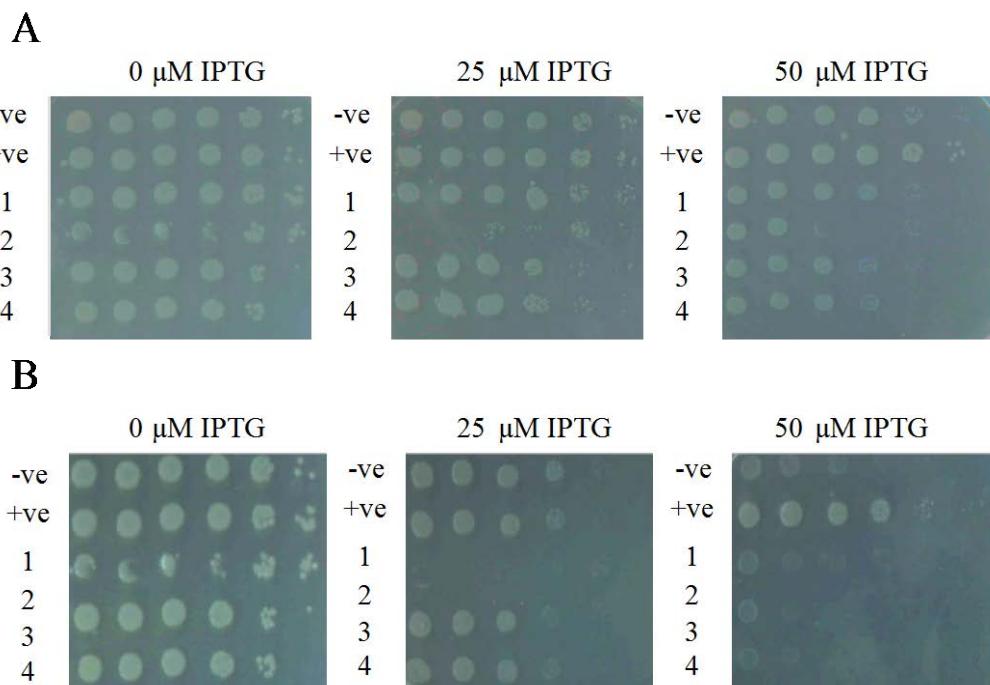
As for the PspA RTHS, *P22-CMG2*<sub>38-218</sub> and *434-PA*<sub>488-735</sub> were sub-cloned into the CRIM plasmid, pAH68, which was subsequently integrated into the chromosome of SNS126. Henceforth, the system will be referred to as the Anthrax RTHS.

#### **4.2.2 Verification of the interaction between PA<sub>488-735</sub> and CMG2<sub>38-218</sub> proteins in the RTHS**

##### **4.2.2.1 Verification of the PPI using a drop-spotting assay**

As the Anthrax RTHS was already constructed, the next step was to verify that the two recombinant proteins interacted within the RTHS to form a functional repressor. The PPI was probed by drop-spotting ten-fold serial dilutions onto selective media containing IPTG at various concentrations (i.e. 0-100 µM), 3-AT and kanamycin. Two concentrations of 3-AT and kanamycin were used: the first concentration for more stringent growth conditions (5.0 mM 3-AT and 50 µg/ml kanamycin); and the second for less stringent growth conditions (2.5 mM 3-AT and 25 µg/ml kanamycin). As anticipated the colonies' capacity to grow decreased as IPTG increased in concentration and between the milder and more stringent conditions. With the latter the loss of survival was 1,000-fold at 25 µM IPTG, whereas on the less stringent conditions it was ten-fold (Figure 73). The Anthrax RTHS was compared with the p6-UEV RTHS.<sup>131</sup> This system demonstrated an inhibition of

growth on similar conditions to the Anthrax RTHS. The Anthrax RTHS was also compared to a negative control containing only the P22- and 434-DNA-binding domains, with no dimerisation domain. For this no decrease in growth was observed at elevated IPTG concentrations.



**Figure 73: Drop-spotting of Anthrax RTHS strains onto minimal media.** (A) Less stringent conditions (2.5 mM 3-AT and 25 µg/ml kanamycin). (B) More stringent conditions (5.0 mM 3-AT and 50 µg/ml kanamycin). The positive control is p6-UEV RTHS and the negative control is a system only expressing the 434- and P22-DNA-binding domains alone integrated into the chromosome and 1-3 are the Anthrax RTHS samples. This showed decreased growth as IPTG was increased from 0 µM to 50 µM, 1,000-fold on more stringent conditions and ten-fold shutdown on milder conditions.

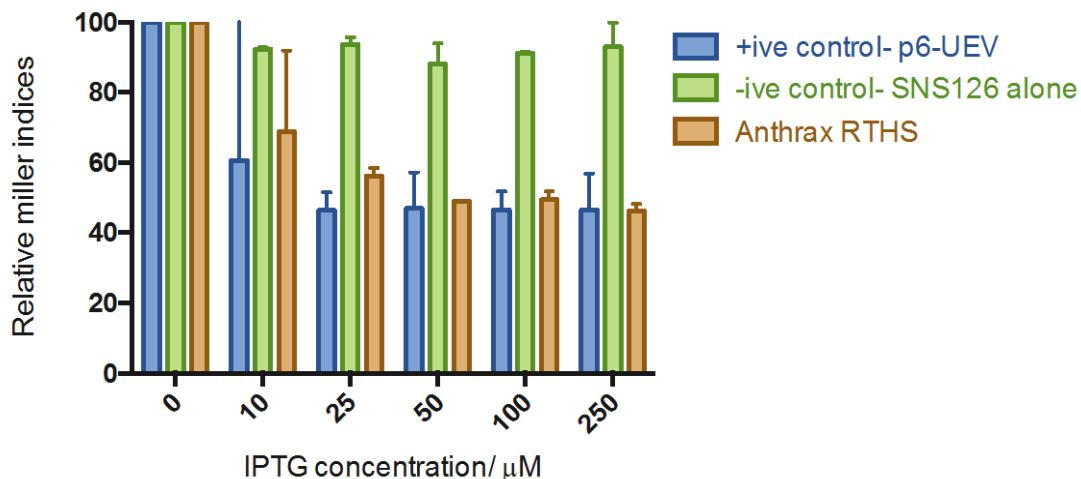
In order to assess if the decrease in the ability to grow on selective media observed in Figure 73 was due to toxicity from the 434-PA<sub>488-735</sub> or P22-CMG2<sub>38-218</sub> recombinant proteins, the Anthrax RTHS was drop-spotted onto LB agar at varying concentrations of IPTG (i.e. 0-50 µM). As IPTG concentration increased, the level of protein expression increased, and if the presence of the recombinant proteins was toxic then decreased growth would be observed even in the absence of the selection criteria (i.e. on histidine-containing media without kanamycin). No shutdown was

observed on LB agar suggesting that the effect observed in Figure 73 was due to the dimerisation of P22-CMG2<sub>38-218</sub> and 434-PA<sub>488-735</sub> inhibiting the transcription of essential genes on the reporter construct and not toxicity.

#### *4.2.2.2 Quantification of the PPI by ortho-nitrophenyl β-galactosidase assay*

As for the PspA RTHS, the dimerisation of PA<sub>488-735</sub> and CMG2<sub>38-218</sub> was linked to the expression of a reporter-construct containing *lacZ*, a gene that produces the enzyme β-galactosidase that hydrolyses β-galactosides into monosaccharides. Within the RTHS, when protein dimerisation does not occur, transcription of the reporter-construct, including the *lacZ* gene, is permitted, expressing β-galactosidase. When protein dimerisation is permitted, then transcription of the *lacZ* gene is inhibited, preventing the breakdown of the ONPG. Therefore as IPTG concentration increases, more protein is expressed, in turn, hydrolysing less ONPG. This can be visualised and the PPI quantified using an ONPG assay.

The results from the ONPG assay (Figure 74) indicated that the Anthrax RTHS followed a similar pattern to the positive control (p6-UEV RTHS, as used for drop-spotting in Section 4.2.2.1).<sup>131</sup> As IPTG levels are increased, there was a decrease in the relative miller index and at 25 μM IPTG the relative miller index plateaued as the RTHS became saturated. The negative control, however, remained constant, independent of the concentration of IPTG. As described in Chapter 3, in order to define when proteins are interacting within the system, Di Lallo *et al.* (2001) defined a limit of interaction at 50% activity repression.<sup>93</sup> The Anthrax RTHS showed an approximate 50% repression in activity at 50 μM, similar to the 54% repression observed at 25 μM for the positive control, p6-UEV RTHS. This indicated that 434-PA<sub>488-735</sub> and P22-CMG2<sub>38-218</sub> are interacting within the Anthrax RTHS.



**Figure 74: Graph of the average results from the ONPG assay.** Carried out in triplicate with the Anthrax RTHS and a positive control of p6-UEV RTHS, a previously constructed RTHS and a negative control of SNS126 with only the 434 and P22 DNA-binding domains expressed. Samples with the anthrax toxin protein PA and the receptor CMG2 present showed a decrease in ONPG breakdown as IPTG increases.

Drawing on the results from the drop-spotting and ONPG assay, the subsequent SICLOPPS screen was carried out using the more stringent conditions identified (i.e. 5 mM 3-AT, 50  $\mu\text{g}/\text{ml}$  kanamycin and IPTG 50  $\mu\text{M}$ ). These were the same conditions used to probe the PspA dimerisation in Section 3.2.2.1 but more stringent than those utilised by Tavassoli *et al.* (2008) for the SICLOPPS screening in the p6-UEV RTHS (i.e. 2.5 mM 3-AT, 25  $\mu\text{g}/\text{ml}$  kanamycin and 30  $\mu\text{M}$  IPTG).<sup>131</sup>

#### 4.2.3 Screening of the CX<sub>5</sub> library

In light of identifying suitable conditions to monitor the dimerisation of 434-PA<sub>488-735</sub> and P22-CMG2<sub>38-218</sub>, the Anthrax RTHS was used to identify cyclic peptide inhibitors of the interaction. To screen for inhibitors, the CX<sub>5</sub> library, was transformed into electro-competent Anthrax RTHS cells. The CX<sub>5</sub> library has 3.2 million unique sequences; in order to ensure that the entire library was represented during the screen the efficiency of the transformation was confirmed by plating

serial dilutions of the recovery mixture onto LB agar plates supplemented with chloramphenicol. The efficiency for this transformation was  $3.0 \times 10^7$ , ensuring that the whole library of 3.2 million members was represented ten-times. The remainder of the recovery solution was plated onto minimal media selective plates containing 5 mM 3-AT, 50 µg/ml kanamycin, 50 µM IPTG and 6.5 µM arabinose, and incubated at 37°C for 72 hours.

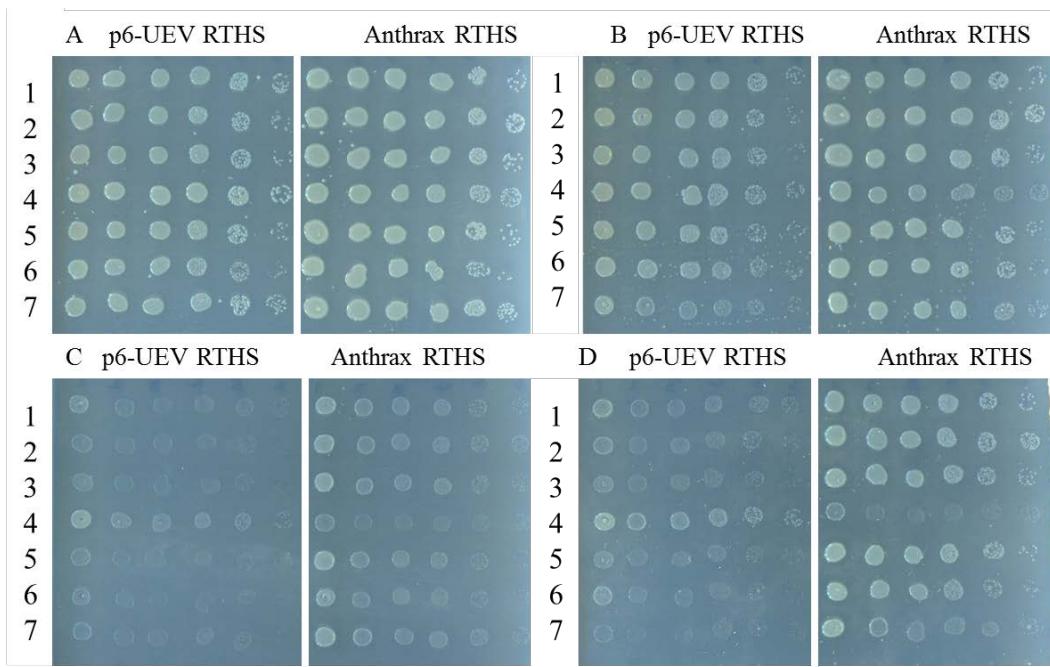
After this time, a selection of the colonies that survived this initial round of screening were drop-spotted onto selective minimal media plates. Appendix 4A-D represents a typical round of screening of 84 colonies, with 12 showing a toxic phenotype in the presence of the cyclic peptide and 15 potentially active sequences. The drop-spotting was carried out on four plates with different conditions. These were:

- 1. Minimal media only supplemented with spectinomycin and chloramphenicol (A).** On this plate growth at all dilutions was observed, as the expression of 434-PA<sub>488-735</sub> and P22-CMG2<sub>38-218</sub> proteins was repressed, meaning that the transcription of the essential genes, *his3* and *kan<sup>R</sup>* could occur.
- 2. As plate 1 with additional arabinose (6.5 µM) (B).** On this plate a mixture of growth and failed growth was observed; this was because arabinose induced the production of the cyclic peptides. When the inhibitor or its processing by-products were toxic to the system, a growth disadvantage was observed. In contrast, when there was no toxicity full growth was observed.
- 3. As plate 1 with additional IPTG (50 µM) (C).** On this plate a growth defect was observed, as expression of 434-PA<sub>488-735</sub> and P22-CMG2<sub>38-218</sub> proteins occurred. These proteins were able to dimerise and bind to their respective repressors, inhibiting the transcription of the essential genes.
- 4. As plate 1 with both IPTG (50 µM) and arabinose (6.5 µM) (D).** On this plate again a mixture of survival and growth disadvantages were observed. Both the cyclic peptides and the 434-PA<sub>488-735</sub> and P22-CMG2<sub>38-218</sub> proteins are being expressed. If the cyclic peptide was successful in disrupting the dimerisation then growth at all dilution was observed; however, if it was not

then the growth disadvantage observed on the IPTG alone plate was observed.

These results are in agreement with those observed previously in the literature. Birts *et al.* (2012) observed 100-fold decrease in growth in the presence of IPTG that was restored in the presence of arabinose when the inhibitor CP61 (*cyclo*-SGWTVVVRMY) was expressed. For the heterodimeric interaction p6-UEV, a similar 100-fold reduction in growth was observed in the presence of IPTG; however, a growth advantage was conferred by the expression of the cyclic peptide inhibitors by inclusion of arabinose.<sup>117,133</sup>

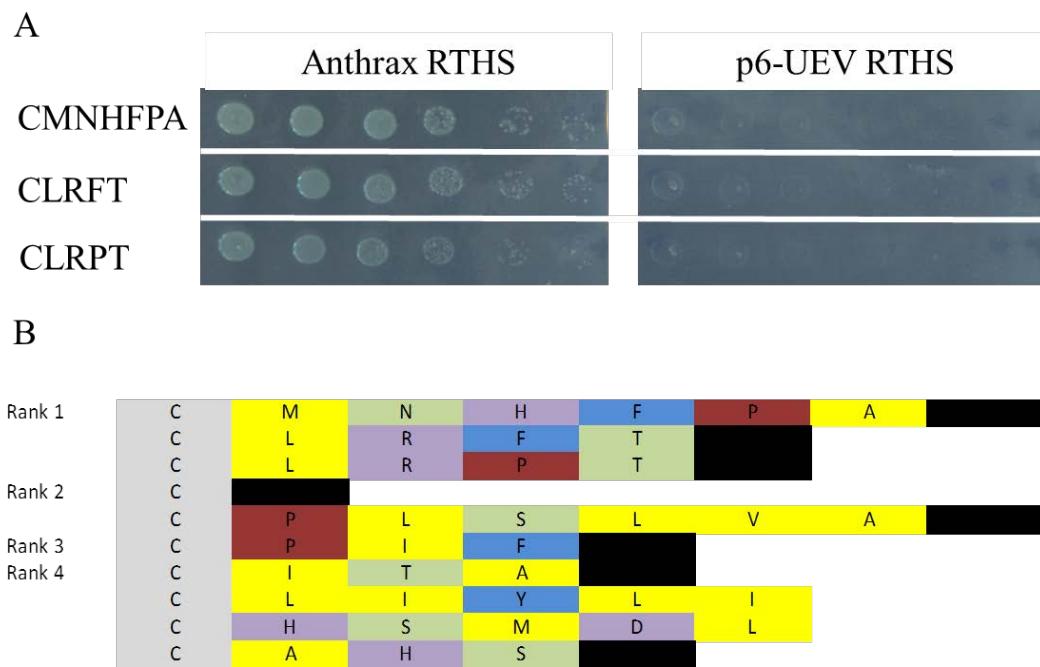
For the screen of the CX<sub>5</sub> library, 480 colonies were analysed; from this 27 potential inhibitory sequences were isolated and taken forward into the secondary screening. This involved isolating the active plasmid and retransforming it back into the Anthrax RTHS and another RTHS to check for specificity. In this case the second RTHS was the unrelated p6-UEV RTHS. Each sample that was retransformed was then drop-spotted onto the plates containing the supplements described above. The desired results were that the sequences that were transformed back into the Anthrax RTHS would retain their previously observed phenotype, however, when in the p6-UEV RTHS no restoration of growth should be observed indicating that the inhibitor is specific to the dimerisation of PA<sub>488-735</sub> and CMG2<sub>38-218</sub>. Of the 27 plasmids taken forward into the secondary screening only ten samples fulfilled the selection criteria by retaining their phenotype and being selective to the Anthrax RTHS (Figure 75).



**Figure 75: Representative secondary screen of the CX<sub>5</sub> library in the Anthrax RTHS and p6-UEV RTHS.** (A) Minimal media plate. (B) Minimal media plate supplemented with 6.5 µM arabinose. (C) Minimal media plate supplemented with 50 µM IPTG. (D) Minimal media plate supplemented with 6.5 µM arabinose and 50 µM IPTG.

#### 4.2.4 Sequencing and ranking of active SICLOPPS peptides

Once the most active sequences for the dimerisation of CMG2<sub>38-218</sub> and PA<sub>488-735</sub> had been identified they were ranked in order of activity by drop-spotting. The top three sequences are shown in Figure 76A; all the sequences that retained their phenotype and were specific were sequenced as illustrated in Figure 76B.



**Figure 76: Secondary screen of the active sequences from the CX<sub>5</sub> library.** (A) Drop-spotting of the top three sequences identified through screening on 5 mM 3-AT, 50 µg/ml kanamycin, 50 µM IPTG and 6.5 µM arabinose. (B) Sequences of active peptides identified through SICLOPPS screening. Ranking of the successful peptides according to their growth in the presence of arabinose and IPTG. The set motif from the library is in grey, yellow represents non-polar aliphatic residues, green polar uncharged residues, blue aromatic residues, red proline, black stop codons and purple represents charged residues.

It is interesting to note the high occurrence of stop codons in eight of the ten most active and specific sequences. The occurrence of stop codons here could be due to either intein toxicity or a requirement for linear peptides to disrupt the PPI. Theoretically the linear inhibitor may be binding into a groove rather than a pocket; this hypothesis will be discussed further in the discussion (Section 4.5). Additionally, two of the three most active sequences have the same residues in all bar one of the positions, CLRxT. Two other motifs were identified: a proline-hydrophobic residue motif and a histidine-serine motif. However, these emerged within the less active sequences (i.e. rank 2 and 3, and rank 4, respectively). Moreover, the sequences in the third and fourth ranking positions have a high proportion of hydrophobic residues, suggesting that these inhibitors may function by binding into a hydrophobic pocket.

The occurrence of stop codons will prevent expression of the N-terminal intein and cause the inteins not to process properly; as would a mutation in key residues like the cysteine and asparagine involved in the cyclisation<sup>100,397</sup> or the histidine; and phenylalanine in the C-terminal intein (residue 24 and 26, respectively); or the threonine and histidine in the N-terminal intein (residues 69 and 72).<sup>123</sup> The design of the SICLOPPS library at the oligonucleotide level eliminates the opal and ochre stop codon by eliminating adenine in the third position of the codon. This allows all twenty of the naturally-occurring amino acids to be produced, but excludes the stop codons UAA (ochre) and UGA (opal). UAG, the amber codon will still occur, but only in 1 in every 32 codons. This prevents the majority of the termination of the transcription and translation of the sequence.<sup>117</sup> Nevertheless, in this screen 80% of the sequences identified contained a stop codon; this is 25-times more than the anticipated 3% occurrence.

Two hypotheses about the cause of these stop codons may be proposed. One is that the occurrence of stop codons could be an artefact of selection pressure on the bacteria. Inteins have demonstrated a bias towards hydrophobic residues in the first extein position, including the aromatic residues tyrosine and tryptophan.<sup>398</sup> It could also be hypothesised that the cause of this bias may be the rate of processing. When the inteins are rapidly processing there is a high-level of intein by-products present. Consequently, one of the reasons of the bias may be that the by-products are toxic in high concentrations to the bacteria. If the sequence has a high rate of processing then the high levels of intein by-products or cyclic peptide may cause the bacteria to be under high selective pressure and show a decreased survival phenotype. This toxicity was evident by the cell death observed during the screening in the presence of only arabinose on minimal media, whereby a high number of colonies did not survive (Appendix 4). Therefore, extein sequences that have a slow rate of processing may be favoured over faster processing sequences, leading to a sequence bias. This bias would mean that the inhibitors identified during the SICLOPPS screen might not be the most active sequences. In the same way, the occurrence of stop codons prevents intein processing, and the production of the toxic intein by-products, as the N-terminal intein will not be transcribed and translated. The peptide sequence will be presented during screening attached to the C-terminal intein. This

would still allow the sequence to be probed for inhibitory activity; but, the C-terminal intein may hinder or enhance its activity.

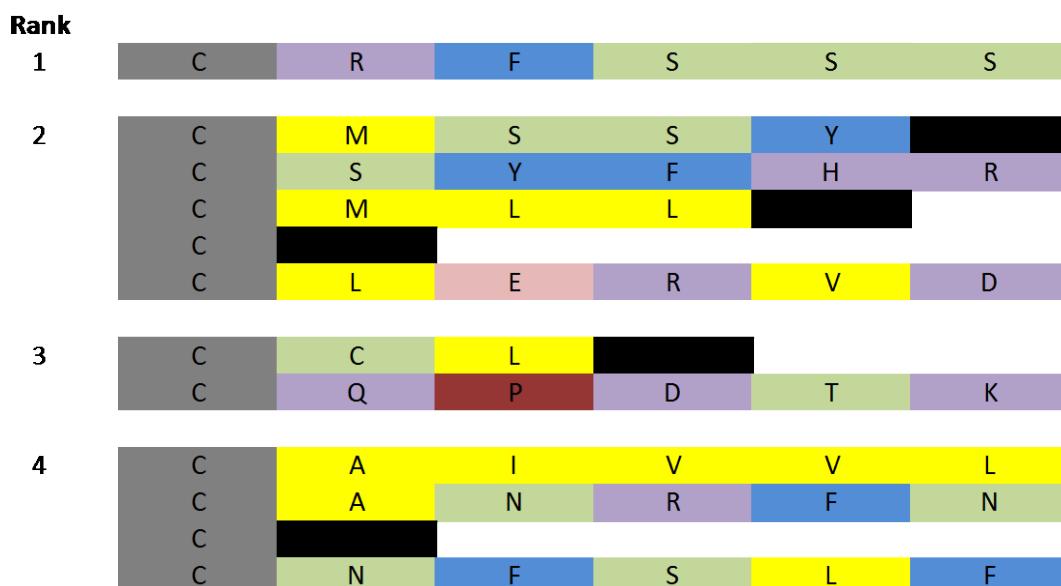
The occurrence of stop codons and mutations affecting intein processing is not well documented in the literature. There are, however, a number of explanations for their occurrence (i.e. toxicity and favoured binding pocket) and precedent in the literature to demonstrate that they may still have inhibitory function.<sup>397</sup> Nevertheless, to investigate the hypothesis that intein toxicity is mediating the types of peptides yielded, the CX<sub>5</sub> library screen was repeated. This occurred on less stringent conditions to relieve the selection pressure and encourage the survival of faster processing extein sequences.

#### **4.2.5 *Investigating intein toxicity***

Due to the high occurrence of stop codons observed in the initial screen, less stringent conditions were used to investigate whether intein toxicity was causing the bias towards linear peptides. To achieve this the Anthrax RTHS harbouring the CX<sub>5</sub> library was drop-spotted onto increased concentrations of 3-AT and kanamycin in conjunction with reduced concentrations of IPTG. If the expression of 434-PA<sub>488-735</sub> and P22-CMG2<sub>38-218</sub> was contributing to the toxicity observed in conjunction with the intein by-products then by decreasing the IPTG concentration and in turn lowering the levels of protein expression, the high levels of toxicity would not be expected to be observed. This may result in an increase in the proportion of fully processed peptides identified as active sequences.

Initially, the ideal conditions for the less stringent screen needed to be identified, to ensure that a decrease in growth was still observed in the presence of lower IPTG concentrations. The conditions were identified by drop-spotting the Anthrax RTHS on varying concentrations of 3-AT (2.5, 5.0, 7.5 and 10 mM) and kanamycin (25, 30, 35, 40 45 and 50 µg/ml) at differing IPTG levels. The results showed that a reduction in survival can still be observed with 10-25 µM IPTG but only when the concentration of 3-AT was increased to 7.5 mM and 25-35 µg/ml kanamycin.

The conditions selected for these additional screens were stringent enough to demonstrate a 10,000-fold reduction in the ability of the bacteria to grow in the presence of IPTG (i.e. 10 µM IPTG, 7.5 mM 3-AT and 30 µg/ml kanamycin). The CX<sub>5</sub> library was transformed into electro-competent Anthrax RTHS cells with an efficiency of 6.4 x 10<sup>7</sup>, ensuring that the whole library of 3.2 million members was represented. The remainder of the recovery solution was plated onto minimal media selective plates and incubated at 37°C for 72 hours. 96 colonies were analysed; this yielded 22 sequences that had a growth advantage. These potential inhibitors were subsequently re-transformed into the p6-UEV RTHS to check for specificity and again into the Anthrax RTHS to confirm retention of the phenotype. From this 12 potential inhibitors were identified and ranked (Figure 77).



**Figure 77: Sequences of active peptides identified through SICLOPPS screening on less stringent conditions.** Ranking of the successful cyclic peptides according to their growth in the presence of arabinose and IPTG. The motif from the library design is in grey, yellow represents non-polar aliphatic residues, green polar uncharged residues, blue aromatic residues, red proline, black stop codons and purple represents charged residues.

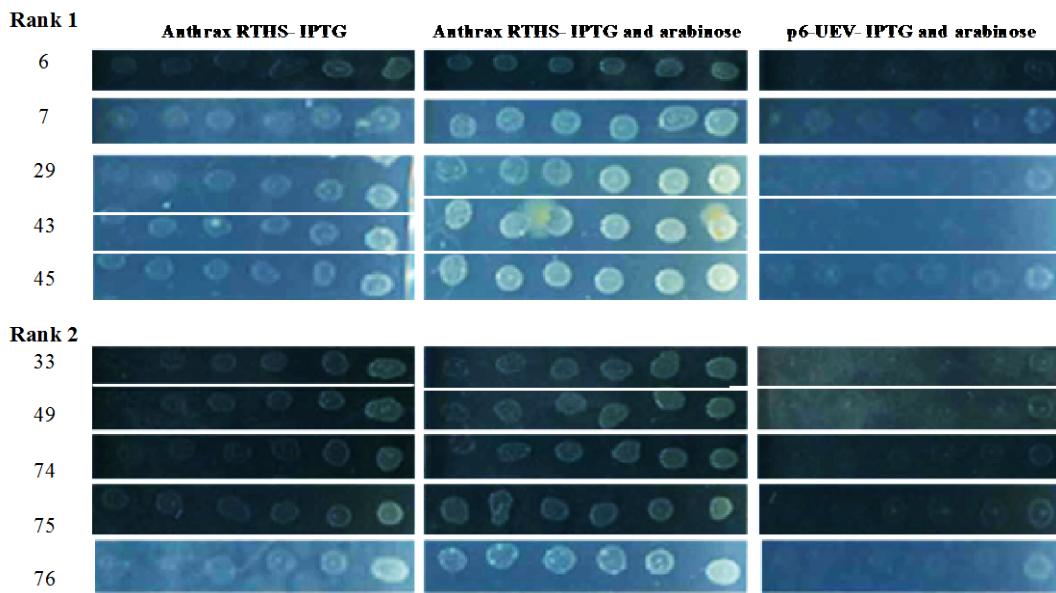
Again there was a high occurrence of stop codons (42%); this was a lower proportion than observed previously on the more stringent conditions (80%). This decrease could be attributed to lower selective pressure, permitting sequences that

have a higher processing rate but more toxic effect to be identified. There is little consensus amongst the sequences, except that in three of the top six peptides identified, there is a motif containing aromatic residues and serine. Intriguingly, the cysteine-stop codon motif was identified in this screen twice, and it was also observed in the initial screen. Despite making the selection screens less stringent stop codons and mutations were still observed, albeit slightly less frequently. Accordingly, the screening conditions cannot be the only cause of these alterations to the sequence; therefore, the requirement for a linear inhibitor was investigated.

#### **4.2.6 Screening of the SGWX<sub>5</sub> library**

As discussed previously the high occurrence of stop codons during the initial screen could have been due to either the high selection pressure or due to a requirement for a linear inhibitor to disrupt the dimerisation of PA<sub>488-735</sub> and CMG2<sub>38-218</sub>. C. Saunders (a project student in the Tavassoli group) probed the latter hypothesis using an SGWX<sub>5</sub> library. The premise behind using a larger library to screen for inhibitors was that, unlike the CX<sub>5</sub> library, the SGWX<sub>5</sub> would have a larger surface area (slightly more ‘linear surface’) and could bind into a long channel optimally mitigating the need for linear peptides.

The SGWX<sub>5</sub> library was constructed as previously described for the PspA RTHS screening (Section 3.2.2.1) and transformed into electro-competent Anthrax RTHS with an efficiency of  $4.7 \times 10^7$ , ensuring that the whole library of 3.2 million members was represented. The remainder of the recovery solution was plated onto minimal media selective plates containing 5.0 mM 3-AT, 50 µg/ml kanamycin, 50 µM IPTG and 6.5 µM arabinose, and incubated at 37°C for 72 hours. 288 colonies were analysed; this yielded 48 sequences that had a growth advantage in the presence of arabinose. These potential inhibitors were subsequently screened in the p6-UEV RTHS to check for specificity and again in the Anthrax RTHS to confirm the phenotype. From this, 25 potential inhibitors were identified and ranked: and again a high level of stop codons and additional mutations were observed. The drop-spotting for the top ten sequences is shown in Figure 78 illustrating how they have a 10,000-fold growth advantage in the presence of arabinose and IPTG in the Anthrax RTHS that is not evident in the p6-UEV RTHS where it was screened for specificity.



**Figure 78: Drop-spotting of the SGWX<sub>5</sub> library in Anthrax RTHS.** The top ten sequences for the SGWX<sub>5</sub> screen with specificity to the CMG2<sub>38-218</sub> and PA<sub>488-735</sub> PPI.

All the 25 active sequences were elucidated by DNA sequencing of the SICLOPPS plasmid (Table 41). A variety of mutations and deletions were identified from this screen: three sequences had a mutation of the serine in the SGW motif to either an asparagine or aspartic acid. This loss of the nucleophilic serine may abolish intein processing, as the nucleophilic residue is required in the intein mechanism; however, the side-chains for the amino acids contain either an amide or a carboxylic acid. Both of these residues could act as a nucleophile- albeit a very poor nucleophile- leading to slow processing. This supports the premise that the linear peptides are an artefact of toxicity leading to a bias towards slow processing inteins.<sup>398</sup> Sequence 6 has a mutation of the cysteine residue in the N-terminal intein and sequences 23 and 34 have stop codons in place of the cysteine. This residue is required for the first step of intein processing, causing an N-to-S acyl shift. Yet, like the mutations above, sequence 6 could still process via the three-step mechanism (illustrated in Figure 7), used by alanine-inteins that are deficient in the cysteine.<sup>123</sup> As observed for the CX<sub>5</sub> library, stop codons had been introduced into the degenerate extein sequence. 40% of the plasmids sequenced revealed either an SGW or SGWx sequence with a stop codon preventing transcription of the remaining degenerate extein sequence and N-

terminal intein with a further 24% having a stop codon elsewhere in the degenerate extein sequence. This would lead to the sequence being presented as an N-terminal fusion of the C-terminal intein. Only one sequence had no mutations in the inteins, extein or stop codons, allowing proper intein processing (*cyclo-SGWIYIPT*), with the same PT motif observed in the initial screen. Nevertheless, this sequence was not taken forward for further testing as it only ranked fourth.

Drawing on all the evidence, the stop codons observed in the first CX<sub>5</sub> screen (Section 4.2.4) could be due to the high pressure the bacteria are under during this screen. These pressures include the intein by-products and the recombinant P22-CMG2<sub>38-218</sub> and 434-PA<sub>488-735</sub> proteins. This conclusion was based on the slight reduction in the occurrence of stop codon in the second screen (Section 4.4.2.5 and 4.2.6) whereby the CX<sub>5</sub> library was screened on milder conditions. Nevertheless, the inhibitors identified in the first screen (CLRFT, CLRPT and CMNHFPA) were tested further to confirm their activity against the dimerisation of CMG2<sub>38-218</sub> and PA<sub>488-735</sub>.

**Table 41: Sequences of the 25 active and specific sequences from the SGWX<sub>5</sub> SICLOPPS screen in the Anthrax and p6-UEV RTHS.**

Inhibitor Reference	Mutations in C-Intein	Mutations in SGW motif	Mutations in N-Intein	Stop Codon	Sequence
6	None	Mutation W to L	Deletion of C	No	SGLVFQSVLLKFWHRNR
7	None	None	None	Yes	SGWWVP
10	None	None	None	Yes	SGWIE
11	None	None	None	Yes	SGW
13	None	None	None	Yes	SGWAHRM
15	None	None	Deletion of T	No	SGWTGIIS
17	None	None	None	Yes	SGW
20	None	None	None	Yes	SGWL
23	Frame shift, after Bg/I site	None	None	Yes	SGWCLYDK
24	Mutation N to T	None	None	Yes	SGWCSV
25	None	None	None	Yes	SGWSI
26	None	None	None	No	<i>Cyclo</i> -SGWIYIPT
27	None	Mutation S to D	None	No	DGWEFNHQ
28	None	None	None	Yes	SGWS
29	Deletion causing frame shift	None	None	No	None
32	None	None	None	Yes	SGWR
33	None	None	None	Yes	SGWN
34	None	None	None	Yes	SGWDHCYF
35	Mutation H to P	None	None	Yes	SGWK
36	None	None	None	Yes	SGW
37	Mutation N to K	Mutation S to N	None	No	NGWYNMHA
38	None	None	None	Yes	SGWS
43	None	None	None	Yes	SGW
45	Mutation H to P	Mutation S to N	None	Yes	NGWAHRT
46	Potential frame shift	None	None	No	None

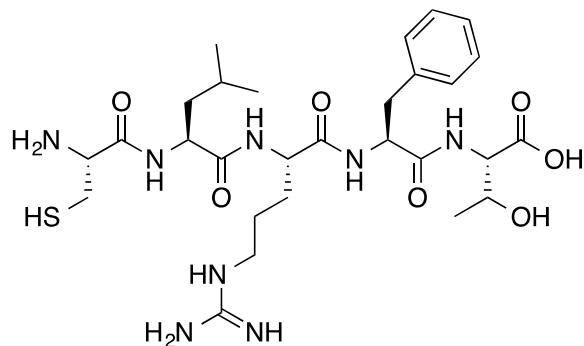
### **4.3 Results: Testing the anthrax inhibitors *in vitro***

Using a combination of a bacterial RTHS and SICLOPPS library screening, three linear peptides were identified to disrupt the heterodimerisation of the mammalian receptor CMG2 and the anthrax toxin PA. It was now necessary to focus on the synthesis of these peptides, verifying their inhibitory activity *in vitro* and attempting to identify whether the peptides are binding to the receptor or the toxin.

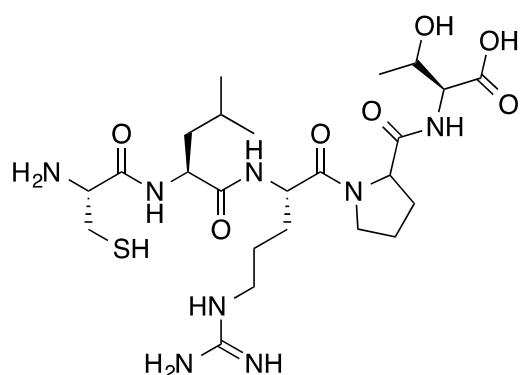
#### ***4.3.1 Synthesis of anthrax inhibitors***

The linear peptides CLRFT (9), CLRPT (10) and CMNHFPA (11) (Figure 79) were synthesised by standard Fmoc solid-phase peptide synthesis, as described previously in Section 2.6.1. All three peptides were cleaved from the Wang linker under acidic conditions in the presence of the scavengers TIS and water. They were purified by precipitation by diethyl ether and RP-HPLC. Yielding 54 mg (34%), 15 mg (21%) and 56 mg (27%), respectively, of pure white solid. Analysis of each peptide was carried out by NMR, low resolution and high resolution MS, LC-MS, IR, analytical HPLC and melting point.

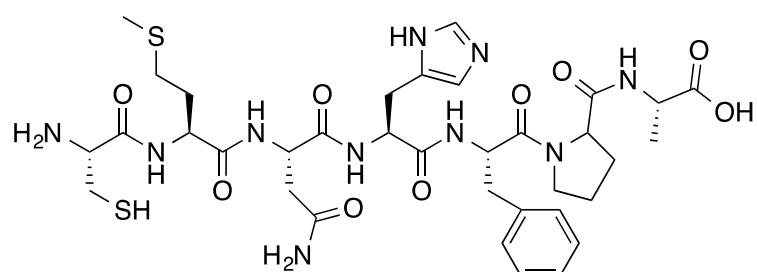
A (9)



B (10)



C (11)

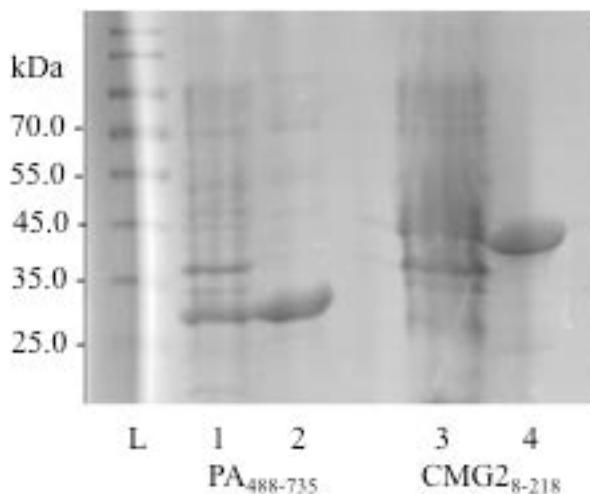


**Figure 79: Structure of CLRFT (9), CLRPT (10) and CMNHFPA (11).**

#### **4.3.2 Protein expression of GST-tagged CMG2<sub>38-218</sub> and His<sub>6</sub>-tagged PA<sub>488-735</sub>**

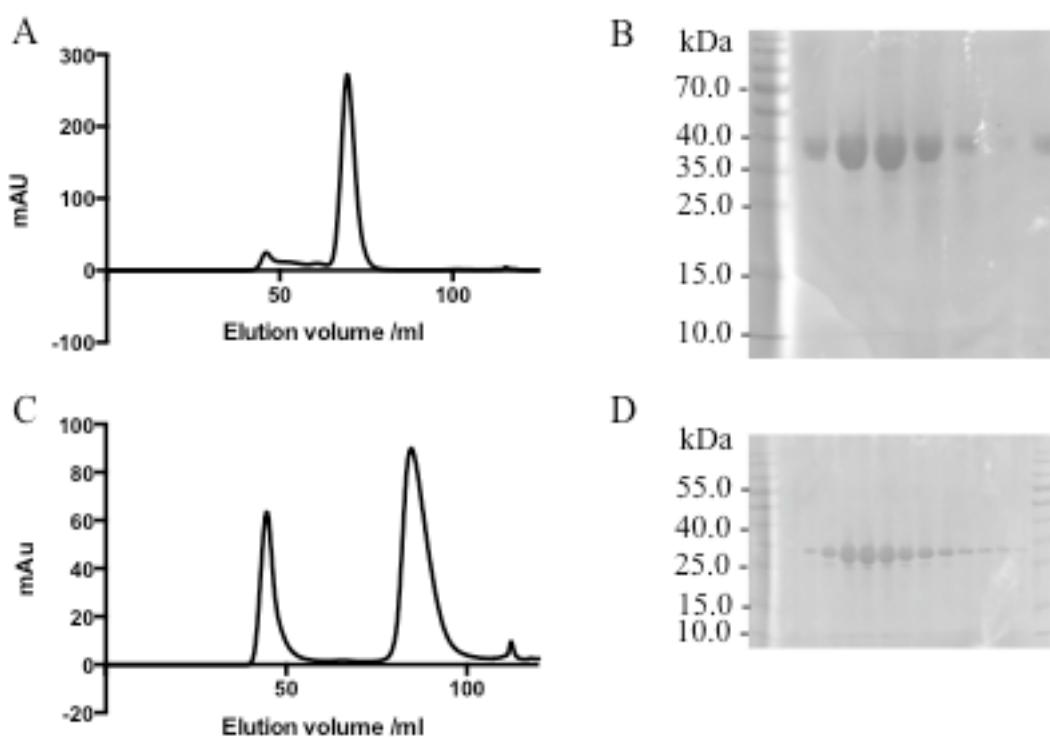
For the purpose of these *in vitro* studies N-terminal fusions with the proteins tags GST and His<sub>6</sub> were designed for CMG2<sub>38-218</sub> and PA<sub>488-735</sub>, respectively. This was done using the expression plasmids pGEX-2TK (GE Healthcare) and pET28a (Novagen), respectively.<sup>399</sup> These purification tags were selected for CMG2<sub>38-218</sub> and PA<sub>488-735</sub> as they could serve the dual purpose of facilitating the protein purification (using glutathione-coated or nickel-coated sepharose beads) as well as allowing the activity studies to be carried out by ELISA.

The plasmids were constructed by amplifying the relevant gene with primers that incorporated restriction digest sites that were compatible with the MCS downstream of the tag of each plasmid. These PCR products were then cloned into the MCS and verified by DNA sequencing. Using these newly constructed expression plasmids, GST-tagged CMG2<sub>38-218</sub> and His<sub>6</sub>-tagged PA<sub>488-735</sub> were expressed and purified. As evident in Figure 80, both proteins were expressed successfully in *E. coli* with bands corresponding to the size of each recombinant protein (i.e. GST-tagged CMG2<sub>38-218</sub> and His<sub>6</sub>-tagged PA<sub>488-735</sub>) visualised by SDS-PAGE. The anticipated sizes were 46.9 kDa and 32.0 kDa, respectively. A total of 6.0 mg of GST-tagged CMG2<sub>38-218</sub> and 7.5 mg of His<sub>6</sub>-tagged PA<sub>488-735</sub> were isolated from 500 ml cultures, as quantified by Bradford assay and absorbance at 280 nm. Typically, these proteins are expressed in a bioreactor to improve yields, within our lab this was not feasible, however, the yields described here are in line with those observed by Christensen *et al.* using shake flasks (personal communication). The buffers used for the GST-tagged CMG2<sub>38-218</sub> purification were PBS-based, with glutathione added to elute the protein from a glutathione affinity column. While the buffers for His<sub>6</sub>-tagged PA<sub>488-735</sub> purification were PBS-based and contained imidazole to allow elution from a nickel affinity column. These buffer compositions were based on those successfully used by Wigelsworth *et al.* (2004).<sup>288</sup>



**Figure 80: Affinity purification of GST-tagged CMG2<sub>38-218</sub> and His<sub>6</sub>-tagged PA<sub>488-735</sub>.** 10% SDS-PAGE gel of GST-tagged CMG2<sub>38-218</sub> where a band corresponding to 46.9 kDa was observed and His<sub>6</sub>-tagged PA<sub>488-735</sub> with a band at 32.0 kDa. *Lane 1 and 3* corresponds to the crude lysate, *Lane 2 and 4* are the purified protein. (Fisher PageRuler Prestained protein ladder 10-170 kDa).

The isolated proteins were purified further by SEC using a Superdex 200 size exclusion column (GE healthcare); GST-tagged CMG2<sub>38-218</sub> eluted at 69.4 ml, while His<sub>6</sub>-tagged PA<sub>488-735</sub> eluted at 44.6 ml and 83.7 ml, corresponding to the void volume and the monomeric protein, respectively. Figure 81A and Figure 81C depicts the elution trace for both proteins, with the corresponding SDS-PAGE gel from the dominant peak, confirming the presence of pure protein (Figure 81B and Figure 81D). Yields for GST-tagged CMG2<sub>38-218</sub> were higher than those observed for His<sub>6</sub>-tagged PA<sub>488-735</sub> after further purification by SEC. This may be due to its larger molecular weight or the removal of undesirable aggregation and misfolded protein during SEC as it eluted in the void volume.



**Figure 81: Purification of His<sub>6</sub>-tagged PA<sub>488-735</sub> and GST-tagged CMG2<sub>38-218</sub>.** (A) Chromatogram of GST-tagged CMG2<sub>38-218</sub> run on the Superdex 200 size exclusion column equilibrated in 20 mM Tris (pH 8.0). (B) 10% SDS-PAGE gel of GST-tagged CMG2<sub>38-218</sub> where a band corresponding to 46.9 kDa was observed. (C) Chromatogram of His<sub>6</sub>-tagged PA<sub>488-735</sub> run on the Superdex 200 size exclusion column equilibrated in 20 mM Tris (pH 8.0). (D) 10% SDS-PAGE gel of His<sub>6</sub>-tagged PA<sub>488-735</sub> where a band corresponding to 32.0 kDa was observed (Fisher PageRuler Prestained protein ladder 10-170 kDa).

The elution of the proteins by SEC provided further evidence of their size by comparing the elution volume to protein standards (Figure 50C). Successful expression and purification of both recombinant proteins allow their use in further studies to probe the activity of the inhibitory sequences identified by the SICLOPPS screen.

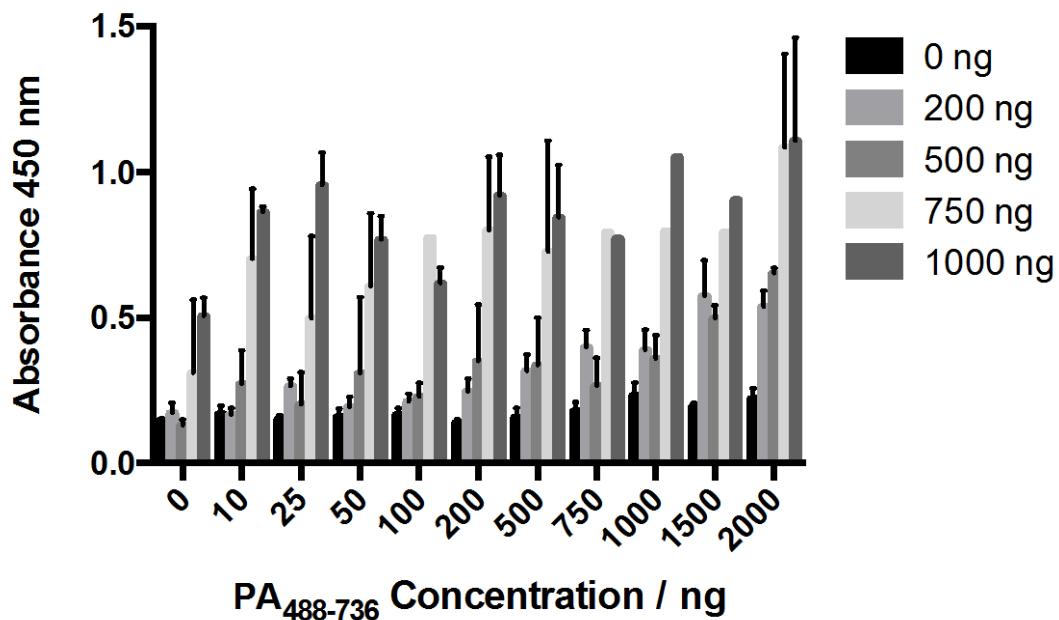


## **4.4 Results: Development of an in vitro assay to verify activity of the anthrax toxin internalisation inhibitors**

### ***4.4.1 Development of an ELISA to quantify the inhibitors activity***

In order to test the three most active inhibitors identified from the SICLOPPS screen an ELISA was established using nickel-coated 96-well plates. His<sub>6</sub>-tagged PA<sub>488-735</sub> was adsorbed onto the nickel-coated plate and then varying concentrations of inhibitor, GST-tagged CMG2<sub>38-218</sub> and MgCl<sub>2</sub> were incubated for a further 1 hour. The analysis proceeded by sequential incubations with an antibody for the GST-tag (anti-GST-primary antibody) and anti-mouse, an anti-species antibody conjugated to the enzyme horse radish peroxidase (HRP, conjugated anti-mouse-secondary antibody), then the enzyme substrate (TMB) and finally a quenching solution of 1 M sulphuric acid, allowing an absorbance reading at 450 nm. All experiments were conducted as triplicates of duplicate (excluding some optimisation experiments) with the replicate(s) giving the same result. All plots below are representative data from the repeats; however, the quantification was taken from the average of the repeats, unless otherwise stated.

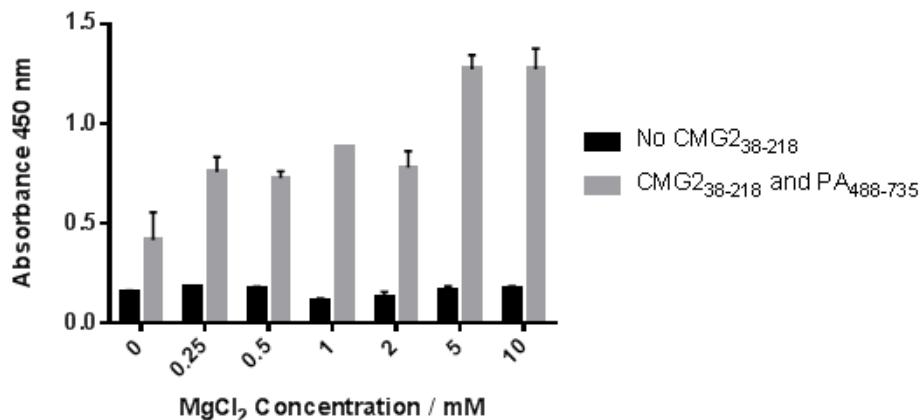
Before the inhibitors could be tested the protein, metal ion and antibody concentrations were optimised. Firstly, protein concentration was probed by binding increasing concentrations of His<sub>6</sub>-tagged PA<sub>488-735</sub> to the nickel-coated plates and then incubating with varying concentrations of GST-tagged CMG2<sub>38-218</sub> and 1 mM MgCl<sub>2</sub> (Figure 82). The result showed that at concentrations of GST-tagged CMG2<sub>38-218</sub> above 500 ng an optimal absorbance within an acceptable signal-to-noise ratio was obtained. In order to balance the levels of protein required and the signal detected future experiments were carried out using 1000 ng of both proteins.



**Figure 82: Comparison of different concentrations of His<sub>6</sub>-tagged PA<sub>488-735</sub> and GST-tagged CMG2<sub>38-218</sub>.** Error bars represent standard deviation of duplicate wells. Similar results were obtained in a repeat experiment.

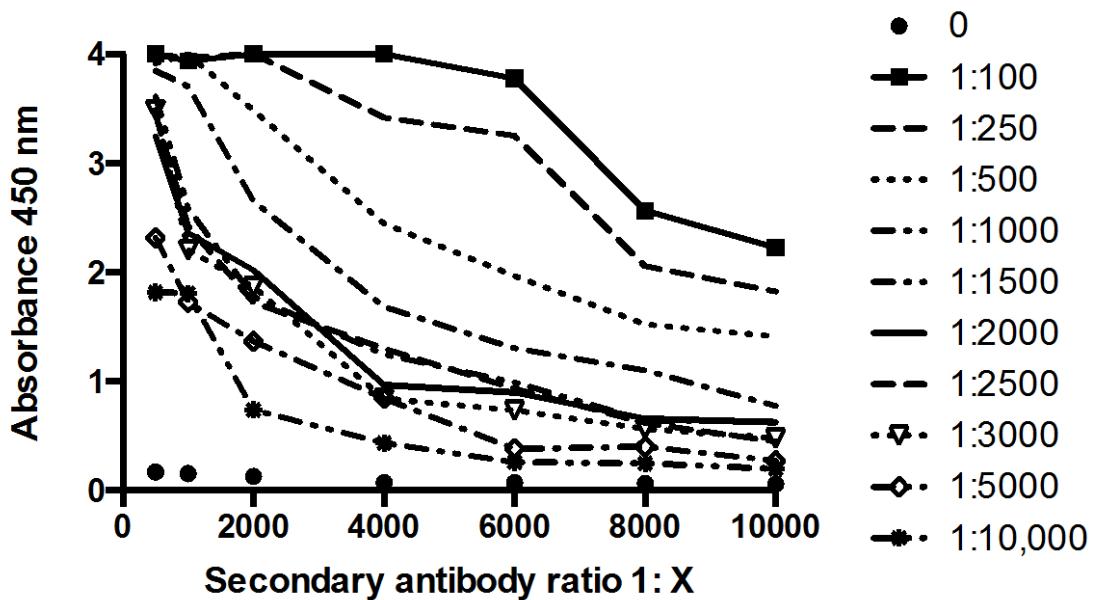
Incubation times of one hour were chosen for all experiments due to evidence from Dawson (2009) that demonstrated that after 10 minutes, 40% of the 180-minute binding value was obtained. At 1 hour this was 80% of the total protein, which was also the maximum binding observed. Therefore, longer incubation times were deemed unnecessary.<sup>290</sup>

As described in Section 4.1.5, Dawson (2009) described how the presence of a divalent cation influences binding.<sup>290</sup> The concentration of MgCl<sub>2</sub> was therefore investigated using 1000 ng His<sub>6</sub>-tagged PA<sub>488-735</sub> and GST-tagged CMG2<sub>38-218</sub>. The results showed a plateau between 0.25 mM and 2 mM (Figure 83) and then a 1.5-fold increase at higher concentrations (i.e. 5 mM and 10 mM). For subsequent experiments 1 mM MgCl<sub>2</sub> was used as this showed a four-fold increase above non-specific binding of GST-tagged CMG2<sub>38-218</sub> antibody to the nickel-coated plate and a two-fold increase from the background when no MgCl<sub>2</sub> was present. The value of 1 mM was also standardly used in the literature for *in vitro* assays, for example, ELISA and FRET.<sup>288,290,359</sup>



**Figure 83: Comparison of different concentrations of MgCl<sub>2</sub>.** Error bars represent standard deviation of duplicate wells.

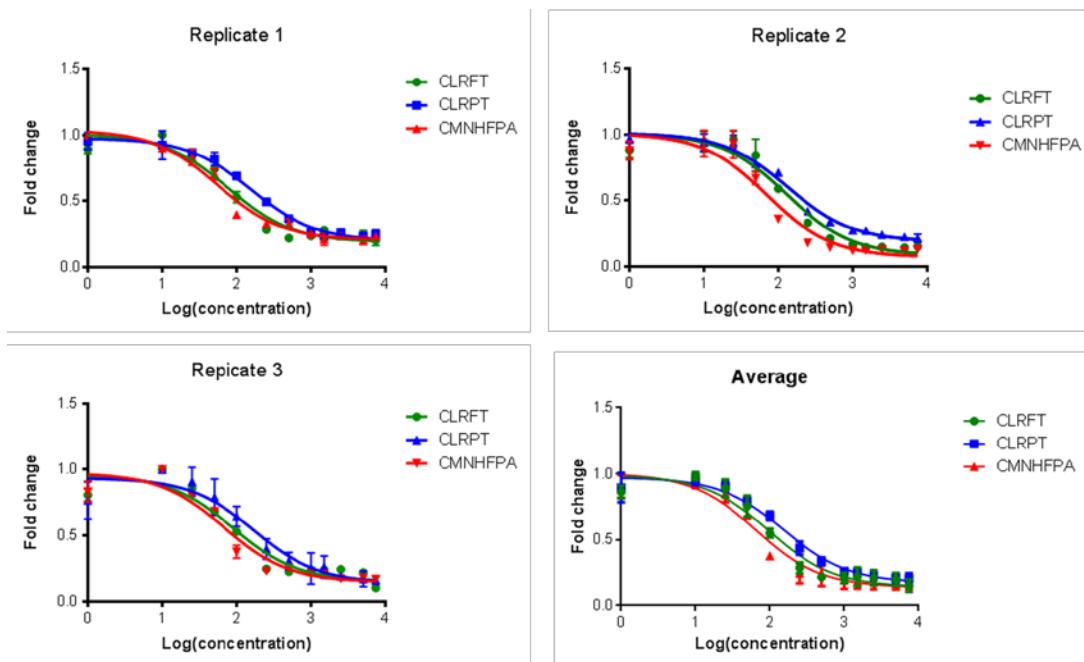
Finally, to obtain a good signal upon detection with TMB solution and absorbance reading at 450 nm, the antibody concentration was optimised (Figure 84). Previous experiments described above had been done using a 1:1000 dilution of anti-GST primary antibody, while a 1:5,000 dilution of anti-mouse-HRP secondary antibody was used. According to the data below this gives an absorbance reading of approximately 1 (consistent with previous results), this is approximately five-fold above the background. To improve the signal to background ratio the primary antibody was increased to a 1:500 dilution for future experiments giving an optimal absorbance signal of 2.0.



**Figure 84: Optimisation of primary and secondary antibody concentration.**

Each dilution is one well of a 96-well plate.

Having optimised the protein concentrations, antibody dilutions and  $MgCl_2$  concentrations, the linear inhibitors (i.e. CLRFT, CLRPT and CMNHFPA) were tested. Figure 85 shows the individual plots for each of the repeats and an average of the data. For the analysis the  $R^2$  values ranged between 0.9616-0.9858 indicating a good fit of the model.



**Figure 85: ELISA of CLRFT, CLRPT and CMNHFP against the dimerisation of GST-tagged CMG2<sub>38-218</sub> and His<sub>6</sub>-tagged PA<sub>488-735</sub>.** Nickel-coated plate incubated with His<sub>6</sub>-tagged PA<sub>488-735</sub> and sequentially treated with GST-tagged CMG2<sub>38-218</sub> and inhibitor, antibodies, enzyme substrate and quenching solution. Error bars represent standard deviation of duplicate wells. Similar results were obtained in each triplicate experiment. R<sup>2</sup> for the average of each experiment was 0.9616, 0.9858 and 0.9656, respectively.

For each inhibitor the IC<sub>50</sub> was calculated, in addition the average IC<sub>50</sub> was calculated. Table 42 shows the IC<sub>50</sub> for each inhibitor, alongside the standard error for the average. From the ELISA assay the most potent inhibitor was CMNHFP - with an IC<sub>50</sub> of  $64.92 \pm 4.70 \mu\text{M}$ , while CLRPT showed the least activity of all the three inhibitors ( $161.10 \pm 5.13 \mu\text{M}$ ). CLRFT was in the middle with an IC<sub>50</sub> of  $99.25 \pm 16.5 \mu\text{M}$ . The decrease in activity of CLRPT suggested that the phenylalanine residue was required for disrupting the interaction between CMG2<sub>38-218</sub> and PA<sub>488-</sub>

735.

**Table 42: IC<sub>50</sub> data for each repeat of the ELISA and the average and standard deviation.**

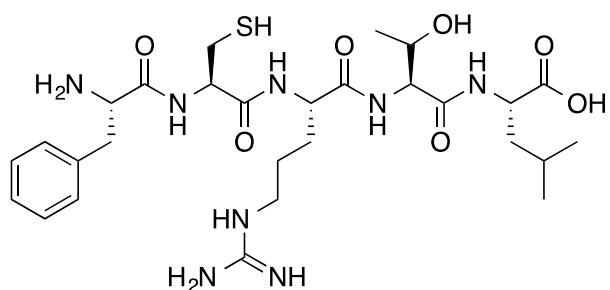
PEPTIDE	REPEAT	REPEAT	REPEAT	AVERAGE/	STANDARD
	1/ $\mu\text{M}$	2/ $\mu\text{M}$	3/ $\mu\text{M}$	$\mu\text{M}$	ERROR/ $\mu\text{M}$
CLRFT	89.13	131.50	77.12	99.25	16.5
CLRPRT	171.20	154.50	157.60	161.10	5.13
CMNHFPA	67.43	71.58	55.76	64.92	4.70

A better understanding of how the two most active inhibitors disrupt the PPI is required, before a more potent inhibitor can be developed. This can be done by looking at the sequence dependency of the inhibitors activity and at the involvement of specific residues.

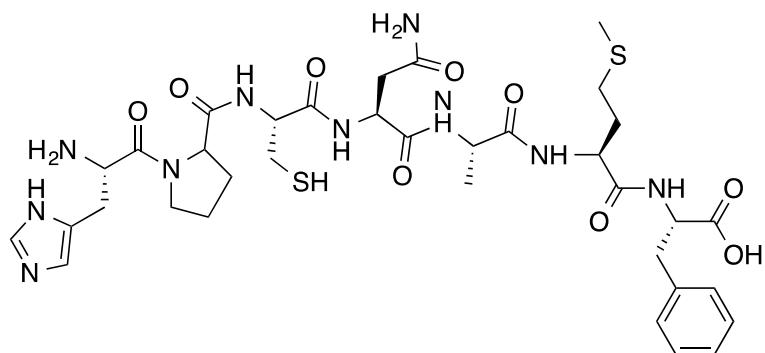
#### 4.4.2 *Investigating the importance of the amino acid sequence*

The use of an ELISA has confirmed the activity of CLRFT and CMNHFPA. Consequently the dependency of the activity of these two inhibitors (CLRFT and CMNHFPA) on the sequence was probed using scrambled analogues. This was achieved by synthesising an analogue of each inhibitor where no amino acid was in the same position; the activity of the peptides was then examined by ELISA. For CMNHFPA, the scrambled sequence HPCNAMF (13) was synthesised using Fmoc standard solid-phase synthesis techniques, along with FCRTL (12), the scrambled analogue of CLRFT (Figure 86). Both were synthesised using the procedure outlined in 2.6.1; resulting in two white solids with a yield of 45 mg (22%) for compound 13 and 56 mg (35%) for compound 12.

A

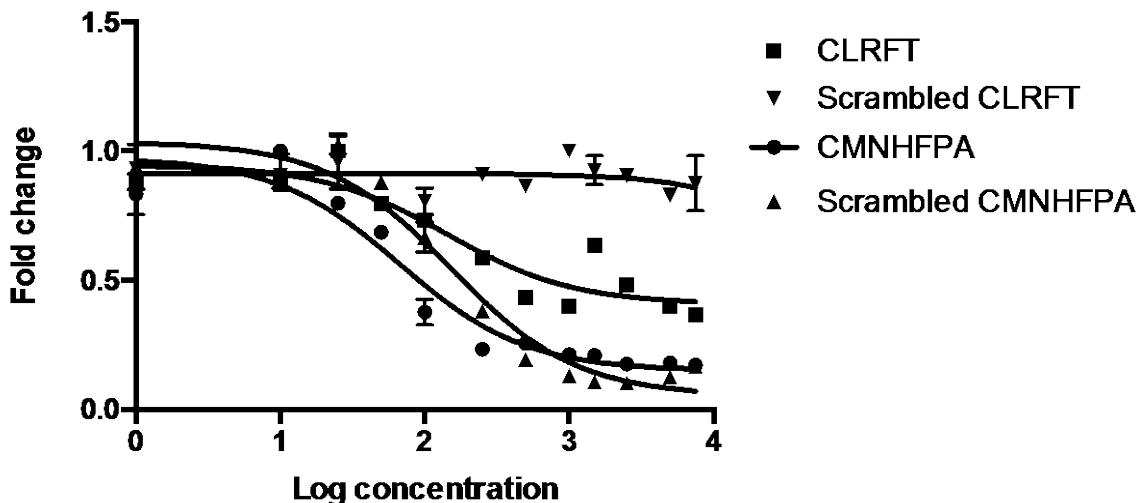


B



**Figure 86: Structure of scrambled analogues.** (A) Scrambled CLRFT (12). (B) Scrambled CMNHFP (13).

The activity of the scrambled peptides was subsequently tested in the ELISA (Figure 87). A complete loss of activity of the scrambled CLRFT analogue compared to the original peptide ( $IC_{50} = 99.25 \pm 9.52 \mu M$ ) was observed. In contrast, only an approximate two-fold loss of activity in the scrambled CMNHFP was observed. The activity of the scrambled CMNHFP may not definitively show that its activity is dependent on the sequence; it could just be that the rearrangement of residues used here could still present the active residues in the correct orientation, another scrambled analogue may not.

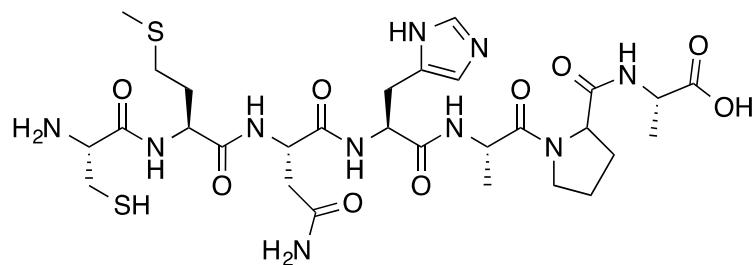


**Figure 87: Representative plot of the ELISA of CLRFT, scrambled CLRFT and CMNHFPA against the dimerisation of His<sub>6</sub>-tagged PA<sub>488-735</sub> and GST-tagged CMG2<sub>38-218</sub>.** Nickel-coated plate incubated with His<sub>6</sub>-tagged PA<sub>488-735</sub> and sequentially treated with GST-tagged CMG2<sub>38-218</sub> and inhibitor, antibodies, enzyme substrate and quenching solution. Error bars represent standard deviation of duplicate wells. Whole experiment was done in duplicate and similar results were obtained in each duplicate experiment.

The activity of CLRFT was shown to be dependent on the order of the amino acids in the sequence, whereas for CMNHFPA the results were inconclusive. No sequence dependency was observed here. The next feature of the peptides activity to examine was the amino acid residues.

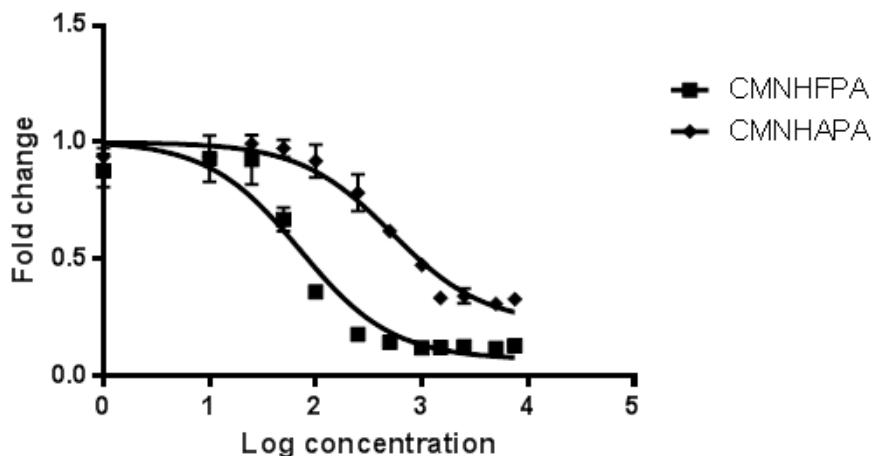
#### 4.4.3 Probing the importance of the phenylalanine residue

Both CLRFT and CMNHFPA have a phenylalanine residue near the C-terminus. When comparing the CLRFT and CLRPT inhibitors there was a 1.5-fold loss in activity with the loss of the phenylalanine ( $IC_{50} = 99.25 \pm 9.52 \mu M$  and  $161.10 \pm 2.96 \mu M$ ). The involvement of the phenylalanine residue in CMNHFPA was therefore investigated by testing an alanine analogue in the ELISA (Figure 88). Using Fmoc solid-phase peptide synthesis CMNHAPA (14) was synthesised yielding 56 mg of white solid (30%). Validation of the correct compound was done by HR-MS, analytical HPLC, NMR, melting point and LC-MS.



**Figure 88: Structure of CMNHAPA (14) the alanine analogue of CMNHFPA (11).**

The activity of the alanine analogue was tested in the ELISA and compared to the original peptide CMNHFPA (Figure 89). The ELISA showed an eight-fold loss of activity by the alanine analogue compared to the parent compound, CMNHFPA ( $IC_{50} = 488.1 \pm 48.2 \mu M$ ). This implicated the phenylalanine in the activity of CMNHFPA.



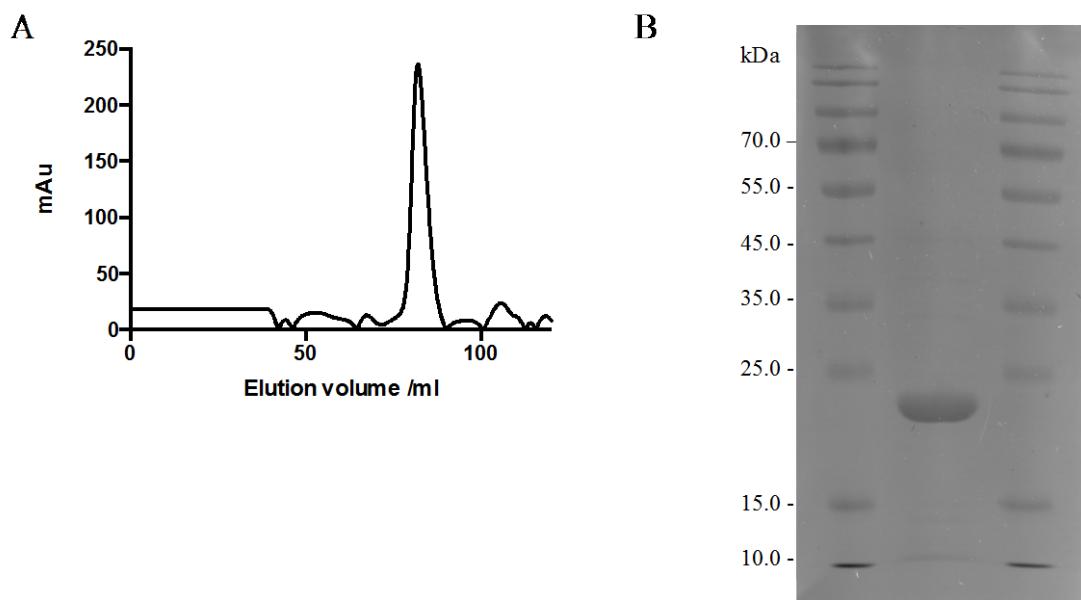
**Figure 89: Representative plot of ELISA comparing the activity of CMNHFPA and CMNHAPA, against the dimerisation of His<sub>6</sub>-tagged PA<sub>488-735</sub> and GST-tagged CMG2<sub>38-218</sub>.** Nickel-coated plate incubated with His<sub>6</sub>-tagged PA<sub>488-735</sub> and sequentially treated with GST-tagged CMG2<sub>38-218</sub> and inhibitor, antibodies, enzyme substrate and quenching solution. Error bars represent standard deviation of duplicate wells. Conducted in duplicate with similar results obtained in each duplicate experiment.

In sections 4.4.1-4.4.3 the activity of the peptides CLRFT and CMNHFPA has been investigated. Further research into the binding partner of each peptide was required to determine which peptide should be taken forward for further development.

#### ***4.4.4 Using the ELISA to ascertain which domain of PA the inhibitors disrupt the interaction with CMG2***

Due to the similarity in sequence between CLRFT and CLRPT, they may function by interacting with the same binding pocket; however, the mode of action of CMNHFPA could be by a different method. In order to investigate which domains the inhibitors interacted with, the ELISA was repeated with His<sub>6</sub>-tagged CMG2<sub>38-218</sub> bound to the nickel-coated 96-well plate and substituting His<sub>6</sub>-tagged PA<sub>488-735</sub> with only domain 4 of PA (PA<sub>596-735</sub>). Levels of PA<sub>596-735</sub> bound were detected using a PA<sub>596-735</sub> specific primary antibody.

Consequently a His<sub>6</sub>-tagged CMG2<sub>38-218</sub> recombinant protein was required to bind to the nickel-coated plates. CMG2<sub>38-218</sub> was therefore cloned into the MCS of the pET28a expression plasmid, via the *Bam*H I and *Sac*I restriction sites. This plasmid was transformed into BL21 (DE3) cells and expression of recombinant His<sub>6</sub>-tagged CMG2 was induced with 1 mM IPTG. A total of 4.9 mg of His<sub>6</sub>-tagged CMG2<sub>38-218</sub> was isolated from a 500 ml culture, as quantified by Bradford assay and absorbance at 280 nm. The buffers used for the His<sub>6</sub>-tagged CMG2<sub>38-218</sub> purification were PBS-based, with 500 mM imidazole added to elute the protein from a nickel affinity column. SEC was used to further purify the isolated protein, where His<sub>6</sub>-tagged CMG2<sub>38-218</sub> eluted at 88.0 ml, corresponding to the protein. Figure 90 shows the chromatogram, with the corresponding SDS-PAGE gel of the dominant peak (23.5 kDa), confirming the presence of pure protein.

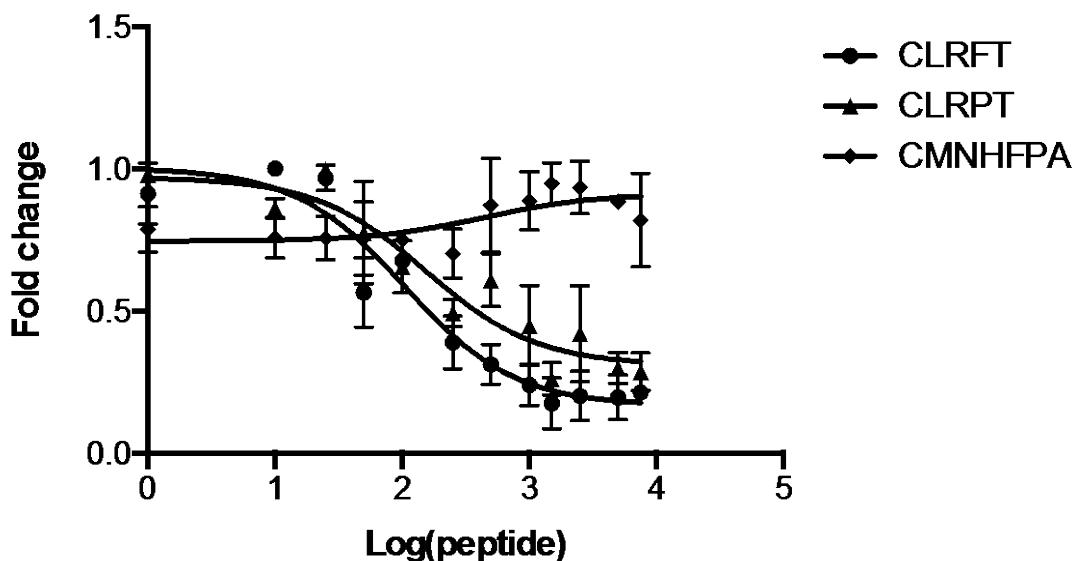


**Figure 90: Purification of His<sub>6</sub>-tagged CMG2<sub>38-218</sub>.** (A) Chromatogram of His<sub>6</sub>-tagged CMG2<sub>38-218</sub> run on the Superdex 200 size exclusion column equilibrated in 20 mM Tris (pH 8.0). (B) 10% SDS-PAGE gel of His<sub>6</sub>-tagged CMG2<sub>38-218</sub> where a band corresponding to 23.5 kDa was observed (Fisher PageRuler Prestained protein ladder 10-170 kDa).

As previously described optimisation of the antibody and the protein concentration were carried out. For this experiment 1000 ng of each protein were used, along with a 1:500 dilution of anti-PA<sub>596-735</sub> primary antibody and 1:10,000 dilution of anti-mouse-HRP secondary antibody. These concentrations were selected as they gave a six-fold increase in signal above the background (as illustrated in Appendix 5). As before 1 mM MgCl<sub>2</sub> was used as the divalent cation to enhance the interaction.

Using only the fourth domain of PA (PA<sub>596-735</sub>), CMNHFPA showed a seven-fold decrease in activity from 64.92 μM to 437.5 μM (Figure 91). This suggested the inhibitor primarily functions by disrupting the PPI of CMG2<sub>38-218</sub> and PA via domain 3 of PA. There is no evidence in the literature that domain 3 of PA directly interacts with CMG2<sub>38-218</sub>, therefore the inhibitor may be binding to domain 3 of PA and causing a conformational change in domain 4, which is in contact with CMG2<sub>38-218</sub>. It may be this conformational change that inhibits dimerisation. Meanwhile, CLRFT and CLRPT target the interaction between the VWA domain of CMG2 and the

fourth domain of PA. The IC<sub>50</sub> for both CLRFT and CLRPT remained the same as previously observed with PA<sub>488-735</sub> (100.4 µM and 154.3 µM, respectively).



**Figure 91:** Representative plot of ELISA of CLRFT, CLRPT and CMNHFP against the dimerisation of His<sub>6</sub>-tagged CMG2<sub>38-218</sub> and PA<sub>596-735</sub>. Nickel-coated plate incubated with His<sub>6</sub>-tagged CMG2<sub>38-218</sub> and sequentially treatment with PA<sub>596-735</sub> and inhibitor, antibodies, enzyme substrate and quenching solution. Error bars represent standard deviation of duplicate wells.

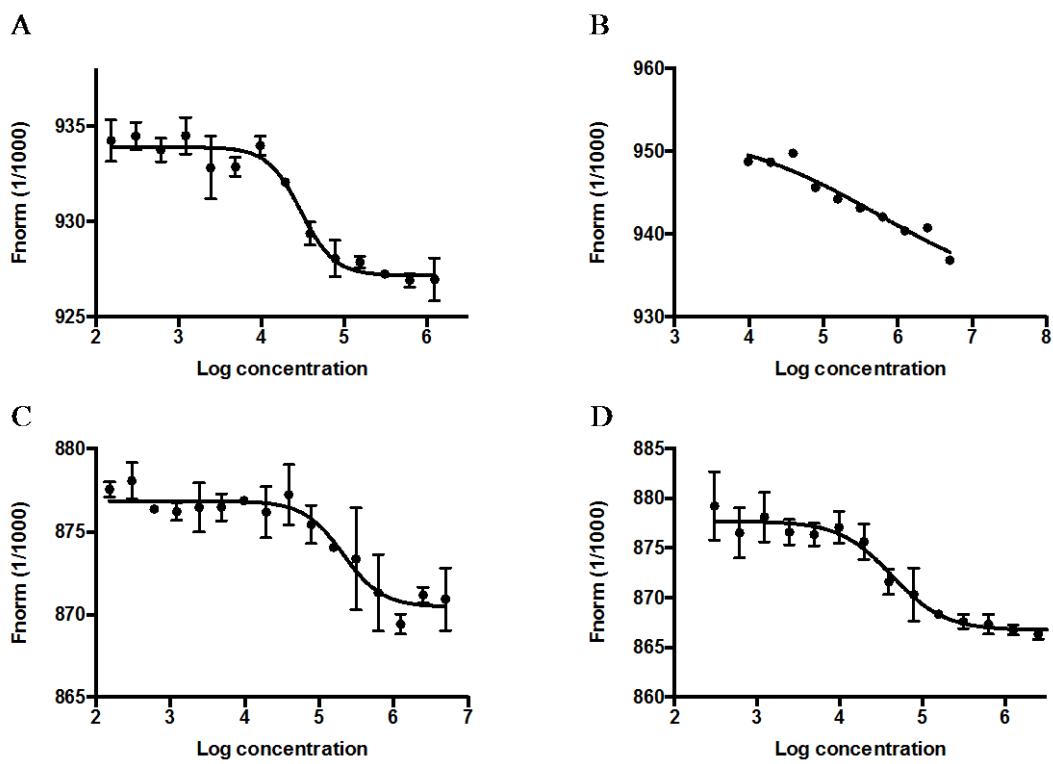
This experiment has shown how the top three inhibitors function by targeting different interactions between PA<sub>488-735</sub> and CMG2<sub>38-218</sub>. For the development of an antimicrobial of anthrax, an inhibitor whose mode of action is via the receptor and toxin interaction is desirable as it can be used as a prophylactic and post-exposure. Furthermore, an inhibitor that functions by binding to the receptor is less susceptible to bacterial resistance. From the ELISA, CMNHFP appears to function by binding to PA<sub>488-735</sub> and causes a conformational change disrupting the PPI. On the other hand CLRFT and CLRPT could bind to either PA<sub>488-735</sub> or CMG2<sub>38-218</sub>. CLRFT was more active than CLRPT, thus the binding of CLRFT was further investigated. If the inhibitor bound to the receptor it would then be optimised to produce a more potent inhibitor.

#### **4.4.5 Identification of the binding partner of CLRFT**

This section describes the use of biophysical techniques to identify the binding partner of CLRFT, in order to identify how the inhibitor functions.

##### **4.4.5.1 Identification of the binding partner of CLRFT by MST**

As described in Section 3.3.4.3, MST is a biophysical technique that can be used to establish binding partners. Both His<sub>6</sub>-tagged PA<sub>488-735</sub> and His<sub>6</sub>-tagged CMG2<sub>38-218</sub> were fluorescently labelled using an NHS-ester group that modifies the free amine group of lysines in the protein. Both of the labelled proteins were incubated with varying concentrations of CLRFT (5 mM to 152 nM) and locally heated with an IR laser and the thermophoretic movement of the proteins was monitored (Figure 92). The weakest binding was observed for His<sub>6</sub>-tagged PA<sub>488-735</sub> ( $K_d = 555.3 \pm 16.5 \mu\text{M}$ , Figure 92A); in contrast His<sub>6</sub>-tagged CMG2<sub>38-218</sub> showed stronger binding to CLRFT with a  $K_d$  more than 18-fold lower ( $K_d = 30.2 \pm 1.2 \mu\text{M}$ , Figure 92B). A seven-fold reduction in binding was observed with the scrambled analogue of CLRFT, showing that the order of the amino acids is critical not only for activity but also binding ( $K_d = 232.2 \pm 1.5 \mu\text{M}$ , Figure 92C). Similarly, reduced binding was observed with CLRPT, suggesting that the phenylalanine residue plays a role in binding ( $K_d = 43.1 \pm 1.3 \mu\text{M}$ , Figure 92D).



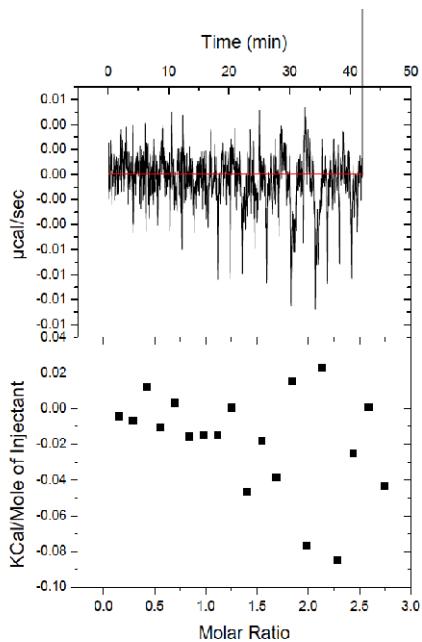
**Figure 92: MST graph of CLRFT and CLRPT against His<sub>6</sub>-tagged CMG2<sub>38-218</sub> and His<sub>6</sub>-tagged PA<sub>488-735</sub>.** Both His<sub>6</sub>-tagged CMG2<sub>38-218</sub> and His<sub>6</sub>-tagged PA<sub>488-735</sub> were fluorescently labelled (NT-647) covalently using NHS coupling. In the MST experiment the concentration of NT-647 labelled protein was kept constant (approximately 80 nM), while the concentration of the non-labelled CLRFT was varied between 5 mM to 152 nM. (A) His<sub>6</sub>-tagged CMG2<sub>38-218</sub> with CLRFT showed a K<sub>d</sub> of  $30.2 \pm 1.2 \mu\text{M}$ . (B) His<sub>6</sub>-tagged PA<sub>488-735</sub> with CLRFT showed a K<sub>d</sub> of  $555.3 \pm 16.5 \mu\text{M}$ . (C) His<sub>6</sub>-tagged CMG2<sub>38-218</sub> with scrambled CLRFT showed a K<sub>d</sub> of  $232.2 \pm 1.4 \mu\text{M}$ . (D) His<sub>6</sub>-tagged CMG2<sub>38-218</sub> with CLRPT showed a K<sub>d</sub> of  $43.1 \pm 1.3 \mu\text{M}$ . Experiments were repeated either duplicate or triplicate.

#### 4.4.5.2 Identification of the binding partner of CLRFT by ITC

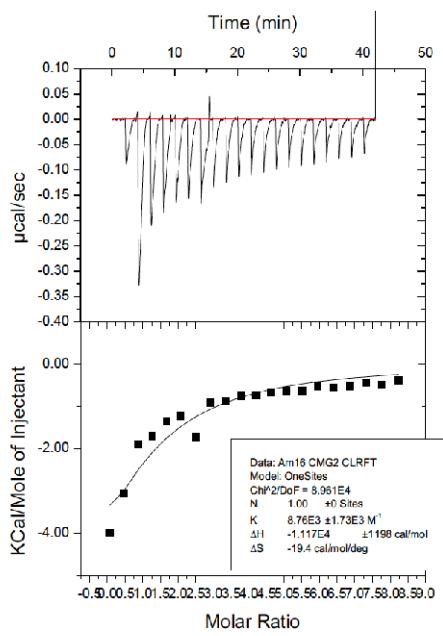
To further evaluate the binding of the inhibitor CLRFT to either His<sub>6</sub>-tagged PA<sub>488-735</sub> or His<sub>6</sub>-tagged CMG2<sub>38-218</sub> ITC was employed. Each recombinant protein was loaded into the measurement cell and CLRFT inhibitor was added in small increments under continuous stirring. The concentration of recombinant protein was 150 μM, five-orders of magnitude above the K<sub>d</sub> predicted by MST for His<sub>6</sub>-tagged

CMG2<sub>38-218</sub>, allowed incremental sub-stoichiometric amounts of inhibitor to bind the protein quantitatively up to saturation.

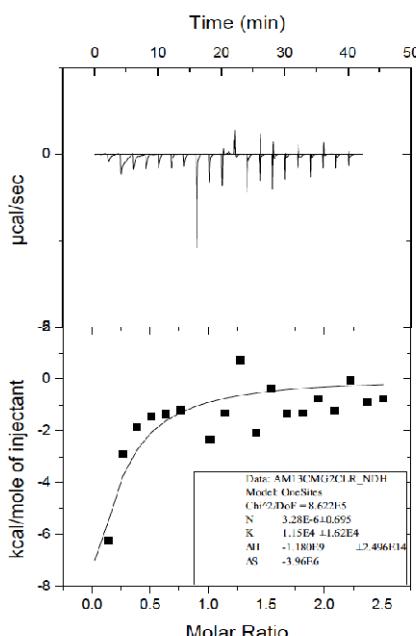
Figure 93A showed no evidence of binding between CLRFT and His<sub>6</sub>-tagged PA<sub>488-735</sub>. This was in agreement with the information obtained from MST where the K<sub>d</sub> was predicted to be 522 μM, although MST observed weak binding; this would have been out of the experimental parameters for the ITC. In order to obtain information by ITC the concentration of the protein should be approximately ten-times the K<sub>d</sub>, to have a *c* value between 10-100.<sup>282</sup> This was not feasible as a concentration of 5 mM protein would be required. Unreliable results were obtained for the binding between His<sub>6</sub>-tagged CMG2<sub>38-218</sub> and CLRFT. Figure 93B shows binding of CLRFT to His<sub>6</sub>-tagged CMG2<sub>38-218</sub>, with a K<sub>d</sub> of 114.16 μM. The trace was typical of an experiment with a low *c* value; this was not unexpected as the K<sub>d</sub> predicted by MST was 30.2 ± 1.2 μM. For an optimal experiment the concentration of protein should be 300 μM, however, recent studies have demonstrated that accurate thermodynamic information can be obtained from low *c* value experiments if the concentrations of the protein and inhibitor are accurately known.<sup>282</sup> Therefore, this experiment was repeated but the data could not be replicated, as evident in Figure 93C. The reasons for this are unclear as in Figure 93C binding is evident in the deconvoluted spectra, yet the raw data does not show a typical binding trace. The disordered nature of the raw data is unprecedented and cannot be explained; in trying to replicate Figure 93B, Figure 93C was obtained each time. Consequently, no conclusions can be drawn from this data.



A) His<sub>6</sub>-tagged PA<sub>488-735</sub> and CLRFT  
No binding observed



B) His<sub>6</sub>-tagged CMG2<sub>38-218</sub> and CLRFT  
 $K_d = 114.16 \mu\text{M}$



C) His<sub>6</sub>-tagged CMG2<sub>38-218</sub> and CLRFT  
 $K_d = 86.96 \mu\text{M}$

**Figure 93: Graph of ITC of His<sub>6</sub>-tagged PA<sub>488-735</sub> and His<sub>6</sub>-tagged CMG2<sub>38-218</sub> with CLRFT.** (A) CLRFT (2 mM) was titrated with His<sub>6</sub>-tagged PA<sub>488-735</sub> (150  $\mu\text{M}$ ) at 25°C, no binding was observed. (B and C) CLRFT (2 mM) was titrated with His<sub>6</sub>-tagged CMG2<sub>38-218</sub> (150  $\mu\text{M}$ ) at 25°C, (B) binding was observed ( $K_d = 114.16 \mu\text{M}$ ) while (C) results were ambiguous, binding was evident ( $K_d = 87.0 \mu\text{M}$ ), however, the trace was not as expected.

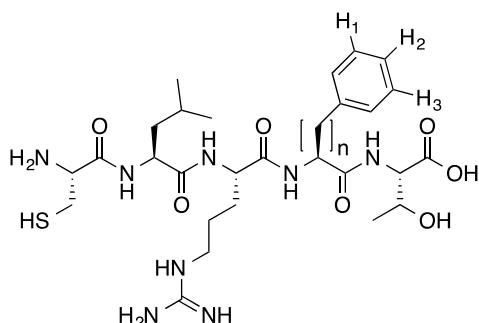
Although the ITC was inconclusive, the MST data demonstrates that CLRFT binds to CMG2<sub>38-218</sub> with a higher affinity than to PA<sub>488-735</sub>. This makes CLRFT a more desirable inhibitor than CMNHFPA, despite the lower activity, which can be optimised.

#### **4.4.6 Optimisation of the binding of CLRFT**

Our inhibitor screen was limited to the 20 naturally-occurring amino acids.<sup>117,133</sup> Therefore, to improve the activity of CLRFT, chemical spaces not accessible to our original cyclic peptide library were explored using unnatural analogues of phenylalanine. A series of related compounds containing electron-donating or electron-withdrawing *para*-substituted phenylalanine analogues were synthesised to probe the electronic and steric requirements of the phenylalanine-binding pocket. These compounds were assayed for binding to CMG2<sub>38-218</sub> using MST. Phenylalanine was selected because from the ELISA and MST data there was a 1.5-fold loss in activity and binding between CLRFT and CLRPT, implicating the phenylalanine residue in the inhibitors activity.

##### **4.4.6.1 Synthesis of the CLRFT phenylalanine analogues**

A small library of phenylalanine analogues were synthesised using standard Fmoc solid-phase peptide synthesis. A selection of modifications was chosen varying the length of the side-chain, the stereochemistry and the electronic properties. These modifications are outlined in Table 43.

**Table 43: Modifications of the phenylalanine residue in each analogue.**

NAME	COMPOUND REFERENCE	STEREOCHEMISTRY					OF PHENYLALANINE
		H <sub>1</sub>	H <sub>2</sub>	H <sub>3</sub>	N	OF	
CLRFT	9	H	H	H	1	L	
CLRF(4-NO <sub>2</sub> )T	15	H	NO <sub>2</sub>	H	1	L	
CLRYT	16	H	OH	H	1	L	
CLRY(3,5-Br <sub>2</sub> )T	17	Br	OH	Br	1	L	
CLRY(3-NO <sub>2</sub> )T	18	NO <sub>2</sub>	OH	H	1	L	
CLRF(4-Cl)T	19	H	Cl	H	1	L	
CLRF(4-F)T	20	H	F	H	1	L	
CLRF(4-CN)T	21	H	CN	H	1	L	
CLRF(4-Bz)T	22	H	Benzoyl	H	1	L	
CLR(Phg)T	23	H	H	H	0	L	
CLR(hPhe)T	24	H	H	H	2	L	
CLR <sup>d</sup> FT	25	H	H	H	1	D	

All the peptides were synthesised and analysed for the correct structure and high purity by low resolution and high resolution MS, LC-MS, NMR, analytical HPLC, IR and melting points. The binding of these analogues was subsequently analysed by MST to establish if a more potent inhibitor had been identified.

#### 4.4.6.2 Comparing the binding of the CLRFT analogues

Previously CLRFT was shown to bind preferentially to His<sub>6</sub>-tagged CMG2<sub>38-218</sub> over His<sub>6</sub>-tagged PA<sub>488-735</sub> by MST. Hence, the binding of the phenylalanine analogues of CLRFT to CMG2 was explored using MST to allow ranking of the analogues to hopefully identify a more potent inhibitor and probe the properties of the binding pocket of CLRFT.

The  $K_d$  values calculated by MST are presented in Table 44. The results indicate that the stereochemistry of the phenylalanine residue does not affect the binding of the inhibitor ( $K_d = 31.0 \pm 2.9 \mu\text{M}$ ), however, approximately a two-fold loss of binding was observed when the side-chain length was varied suggesting that the flexibility of the  $\text{CH}_2$  side-chain was required or that it was unable to fit in the pocket. A complete loss of binding was observed for CLR(4-Bz)T, this is a large bulky side-chain with an additional hydrophobic aromatic ring which could prevent it from docking in the binding pocket. There was little correlation between the electronic properties of an analogue and its ability to bind to His<sub>6</sub>-tagged CMG2<sub>38-218</sub>. The addition of the electron-donating hydroxyl group caused a three-fold reduction in the binding affinity ( $K_d = 91.9 \pm 9.5 \mu\text{M}$ ). While the addition of two bromine groups flanking the hydroxyl group restored some of the binding affinity ( $K_d = 60.0 \pm 11.1 \mu\text{M}$ ). The powerful electron-withdrawing nitro-group had little effect on the binding of CLRFT to His<sub>6</sub>-tagged CMG2<sub>38-218</sub>, yet the weaker electron-withdrawing cyanide group reduced the binding affinity by two-fold ( $K_d = 61.4 \pm 8.0 \mu\text{M}$ ). Only the chloro-analogue showed an improvement in the binding affinity, with a two-fold decrease ( $K_d = 17.7 \pm 3.23 \mu\text{M}$ ). The MST graphs for all analogues are in Appendix 6.

**Table 44:  $K_d$  of each phenylalanine analogue tested by MST.** Calculated from triplicate experiments.

NAME	COMPOUND REFERENCE	$K_d / \mu\text{M}$	ERROR IN THE $K_d / \mu\text{M}$
CLRFT	9	30.2	1.2
CLRF(4-NO <sub>2</sub> )T	15	36.2	5.5
CLRYT	16	91.9	9.5
CLRY(3,5-Br <sub>2</sub> )T	17	60.0	11.1
CLRY(3-NO <sub>2</sub> )T	18	112.0	13.4
<b>CLRF(4-Cl)T</b>	<b>19</b>	<b>17.7</b>	<b>3.23</b>
CLRF(4-F)T	20	49.9	7.1
CLRF(4-CN)T	21	61.4	8.0
CLRF(4-Bz)T	22	None	None
CLR(Phg)T	23	59.6	10.2
CLR(hPhe)T	24	74.0	10.0
CLRdFT	25	31.0	2.9



## **4.5 Summary and discussion for Chapter 4**

*B. anthracis* secretes three toxins that are essential virulence factors: once internalised they facilitate the progression of the disease, and are ultimately responsible for the death of the host. Chapter 4 focused on the identification of inhibitors for the interaction between CMG2 and PA, theoretically mitigating the progression of the disease.

Firstly, a RTHS was constructed by Dr. F. Foranov, and the PPI verified by drop-spotting assay and an ONPG assay. As noted with other PPIs probed using the RTHS, a growth retardation of *E. coli* was observed in response to increased concentrations of IPTG, and when cells were grown under more stringent conditions, relative to the less stringent conditions.<sup>126,127,131-134</sup> It is postulated that this is because increased IPTG levels induce 434-PA<sub>488-735</sub> and P22-CMG2<sub>38-218</sub> expression and dimerisation.<sup>126,127,131-134</sup> This would lead to increased binding of the recombinant proteins to the DNA-binding domains in the chromosome, thereby obstructing the binding of the RNA polymerase and preventing expression of the reporter-construct. The effects observed were confirmed to be due to the PPI and not protein toxicity by drop-spotting onto LB agar at varying IPTG concentrations where no inhibition of growth was observed.<sup>274</sup>

As for the PspA RTHS the ONPG assay was used to quantify the PPI. Similar to the PspA RTHS, as IPTG levels increased, so did the levels of protein dimerisation within the RTHS demonstrated by the decrease in the calculated relative Miller indices. The negative control remained constant regardless of IPTG concentration;<sup>134</sup> in line with previous findings whereby residual β-galactosidase synthesis of 5% was observed for the expression of the homodimeric and heterodimeric repressors, P22 and 434.<sup>90,94</sup> As described in Chapter 3, Di Lallo *et al.* (2001) defined the limit of a PPI to 50% activity repression.<sup>93</sup> The Anthrax RTHS showed a repression of β-galactosidase activity of 51%, confirming that the proteins were interacting. This repression is similar to that observed for the interaction between MinC and FtsZ (43%),<sup>93</sup> but less than the repression observed for the HIF-1α and HIF-1β heterodimerisation (80%).<sup>94,134</sup> The binding affinities for HIF heterodimerisation and FtsZ and MinC homodimerisation have been ascertained by ITC or SPR, and the

repression observed here does not correlate to the binding affinities described in the literature for CMG2 and PA.<sup>400,401</sup> For example, the binding affinity of CMG2 and PA, determined by ITC, was 170 pM,<sup>288</sup> whereas for HIF-1 $\alpha$  and the Per-Arnt-Sims domain B of HIF-1 $\beta$  the K<sub>d</sub> was 125 nM.<sup>400</sup> From this one would expect the repression observed in the ONPG assay for CMG2 and PA to be greater than for HIF heterodimerisation. In the same way, the repression for the interaction between MinC and FtsZ should be 1,000,000-times less than for the interaction between CMG2 and PA, as the binding affinity determined by SPR was 6  $\mu$ M.<sup>401</sup> Nevertheless, the differences can be attributed to the orientation of the PPI and the effect of the N-terminal binding domain hindering the interaction. As discussed in Section 3.4 the variation in IPTG concentration used during the ONPG and drop-spotting assays may also be attributed to the levels of protein expression.

Combining the results from the drop-spotting and the ONPG assays, the SICLOPPS screening was carried out using 50  $\mu$ M IPTG, 5.0 mM 3-AT and 50  $\mu$ g/ml kanamycin. These conditions are more stringent than those used in the screening of the heterodimeric p6-UEV system (i.e. 30  $\mu$ M IPTG, 25  $\mu$ g/ml kanamycin and 2.5 mM 3-AT)<sup>9</sup> and homodimeric ATIC system (i.e. 50  $\mu$ M IPTG, 25  $\mu$ g/ml kanamycin and 2.5 mM 3-AT).<sup>126,131</sup> The conditions, however, are less stringent than those used for the ribonucleotide reductase selection (i.e. 200  $\mu$ M IPTG, 50  $\mu$ g/ml kanamycin and 2.5 mM 3-AT),<sup>80</sup> as well as HIF-1 $\alpha$  and HIF-1 $\beta$  dimerisation (i.e. 100  $\mu$ M IPTG, 50  $\mu$ g/ml kanamycin and 7.5 mM 3-AT).<sup>134</sup> These differences observed between the PPIs may be due to the strength of the binding interaction. The dimerisation of the protein brings the two-repressor subunits into contact, and it may be that the orientation by which the p6-UEV proteins interact provides more optimal binding to the DNA-binding domain than that produced by CMG2<sub>38-218</sub> and PA<sub>488-</sub>

735.

Having established that the recombinant proteins dimerise within the RTHS, the CX<sub>5</sub> library was screened, and nine active sequences were identified that were specific to the disruption of CMG2 and PA. Of the nine sequences identified, seven contained a stop codon either in the degenerate region or in the N-terminal intein; this meant that the sequences were linear.<sup>114,124</sup> The occurrence of stop codons is 25-times greater

than anticipated, when considering the design of SICLOPPS.<sup>117</sup> The only two sequences that were cyclic were *cyclo-CLIHLI* and *cyclo-CHSMDL*, but these were in the fourth rank for activity. As discussed in Section 4.2.4 the occurrence of stop codons could be due to selective pressure, from intein by-product toxicity or due to a requirement for the inhibitor to be linear to fit a groove on the protein, for example the β-loops of CMG2 in the VWA domain (Figure 70), causing an allosteric change disrupting the PPI.<sup>293</sup>

The occurrence of stop codons, deletions and mutations was not discussed by Scott *et al.* (1999) in their original SICLOPPS paper.<sup>100</sup> Nevertheless, Kritzer *et al.* (2009) described three frame shifts and one stop codon during their study to identify inhibitors of α-synuclein toxicity in a yeast model using SICLOPPS technology.<sup>129</sup> The occurrence of frame shifts downstream of the library sequence has been attributed to the use of an annealing temperature less than 65°C during the initial PCR of the library insert.<sup>117</sup> The protocol used by Kritzer *et al.* (2009) was not stated, nor the location of the type of frame shifts that occurred. Therefore no conclusions can be drawn as to the cause of those mutations.<sup>129</sup> For the stop codons observed here none are frame shifts downstream of the library sequence so they cannot be attributed to a low annealing temperature during PCR, therefore further investigation into their cause was required.

The occurrence of stop codons here has led to the production of linear peptides, still attached to the C-terminal intein. There is a precedent for the successful inhibition of a PPI using a sequence that has not fully processed.<sup>397</sup> During intein processing the side-chain of the asparagine triggers cleavage of the cyclic peptide from the C-terminal intein.<sup>123</sup> A mutation of this residue to an alanine mitigates this cleavage, leaving the cyclic peptide and C-terminal intein attached in the lariat form. These lariat peptides have been used in a yeast THS to screen for inhibitors of the LexA homodimerisation, whereby the ‘lariat’ peptide had a transcription activation domain attached to the N-terminus of the C-terminal intein completing the system and allowing the screening of a combinatorial library.<sup>397</sup>

Nevertheless, a cyclic peptide that is not susceptible to metabolic degradation is desirable. Naumann *et al.* (2005) briefly discussed the advent of toxicity whereby

high level expression of the SICLOPPS library caused 30-50% of the clones to exhibit toxicity to the host,<sup>246</sup> therefore to probe whether the linear peptides were an artefact of toxicity or design, screening under less stringent conditions was carried out. Here, even under less stringent conditions, 42% of the active sequences contained a stop codon; the sequences isolated included two with stop codons immediately after the cysteine; this motif was also observed in the initial screen. The reduction in the prevalence of stop codons intimates that toxicity could be a factor in their frequency; however, the consistent occurrence of them argues that a linear structure may be required for disruption or that the C-terminal intein that is not cleaved plays a role in the disruption of the PPI.

Regardless of the speculation that the linear peptides were an artefact of toxicity, the active sequences identified in the first screen were tested for *in vitro* activity. These were CLRFT, CLRPT and CMNHFPA: the first two due to their similarity could function on a similar ‘hotspot’ of the protein while the latter may disrupt the PPI via a different mechanism. The inhibitors identified for the disruption of HIF-1 $\alpha$  and HIF-1 $\beta$  showed a similar overall consensus in the sequences identified (i.e. *cyclo*-CLLFVY and *cyclo*-CLLRMY).<sup>134</sup> These three inhibitors were synthesised using Fmoc solid-phase chemistry with high yields and tested for their ability to disrupt the PPI by ELISA. Unlike the inhibitors identified for the disruption of PspA they were water soluble, consequently no solubility tag or DMSO was required. The protocol was based on an ELISA developed by Dawson (2009).<sup>290</sup>

The three inhibitors CMNHFPA, CLRFT and CLRPT had an IC<sub>50</sub> of 64.92 ± 2.73  $\mu$ M, 99.25 ± 9.52  $\mu$ M and 161.10 ± 2.96  $\mu$ M, respectively. The small molecule inhibitor for the CMG2 and PA interaction, *cis*-platin, identified via a FRET-based assay had an IC<sub>50</sub> of 34  $\mu$ M.<sup>289</sup> Although *cis*-platin was more potent than the inhibitors described here, it was cytotoxic and highly bioactive making it unsuitable for development. Further research needs to take into account the toxicity of the inhibitors identified here, as well as potential off-target effects and bioavailability. The IC<sub>50</sub> of the linear peptides identified was in line with the linear peptide, HTSTYWLDGAP (i.e. 150  $\mu$ M) that inhibited the interaction between PA and LF.<sup>386</sup> However, further development of this peptide by conjugating the peptide to a scaffold increased the potency 7,500-fold. Similarly, a 100,000-fold increase in

activity was observed in a second screen for a truncated version of HTSTYWWLDGAP once it had been conjugated to a scaffold.<sup>391,396</sup> This significant improvement in activity illustrates how the inhibitors identified here can be optimised for improved activity, before they can compete with other inhibitors currently under investigation. As previously discussed, cyclic peptides identified using a RTHS and SICLOPPS platform have been optimised to produce a more effective inhibitor, with 120-fold improvement in activity using a truncated and capped form.<sup>122</sup>

The activity of peptides can be sequence dependent, due to the presentation of an active motif, scrambling a peptide has been used to demonstrate this effect, with the scrambled peptides showing either a complete loss or a reduction in activity.<sup>402,403</sup> To investigate the dependency of activity on the sequence, scrambled analogues of the two most active sequences were synthesised and tested by ELISA. The activity of CLRFT was dependent on the order of amino acids, while, scrambled CMNHFPA showed retention of activity. As discussed in Section 4.4.3.3 this could be due to the length of CMNHFPA allowing a greater degree of freedom to present the active motif, whereas with the short CLRFT this may not possible.<sup>402,403</sup> In addition, the ELISA was repeated with only domain 4 of PA (residues 596-735); CLRFT and CLRPT retained their activity whereas the activity of CMNHFPA was lost. PA and CMG2 interact via the 2 $\beta_2$ -2 $\beta_4$  loops of PA domain 2 and Glu122 pocket adjacent to the MIDAS in CMG2 and by a salt bridge between Arg344 in domain 4 of PA and the Glu122 pocket of CMG2.<sup>351</sup> It was hypothesised that PA Glu683 would contribute a carboxyl-group to the MIDAS metal ion coordination sphere, as observed in the ligand and  $\alpha$ -integrins interaction.<sup>295,354</sup> There is no known interaction between CMG2 and PA domain 3 suggesting that CMNHFPA is either an allosteric inhibitor or disrupted a previously unidentified interaction.

Due to the repeat occurrence of the CLRxT motif in the SICLOPPS screen, and the sequence dependency observed for CLRFT, the binding of this peptide was investigated using MST and ITC. CMG2 was identified as the binding partner by MST and tentatively by ITC. By MST the weakest binding was observed for His<sub>6</sub>-tagged PA<sub>488-735</sub> ( $K_d = 555.3 \pm 16.5 \mu\text{M}$ ), in contrast, His<sub>6</sub>-tagged CMG2<sub>38-218</sub> showed 18-fold improved binding affinity to CLRFT. This demonstrated that

CLRFT inhibits the PPI by binding to CMG2, as observed with *cis*-platin and tannic acid, identified by FRET as inhibitors of CMG2 and PA PPI.<sup>289,291</sup> Comparably to the ELISA a decreased binding affinity was observed with the scrambled analogue of CLRFT, confirming that the sequence order was critical for activity ( $K_d = 232.2 \pm 1.5 \mu\text{M}$ ). Similarly, reduced binding was observed with CLRPT, corroborating the premise that the phenylalanine residue plays a role in activity ( $K_d = 43.1 \pm 1.3 \mu\text{M}$ ). The restriction in conformational freedom introduced by the proline or the absence of an aromatic group that can bind into a hydrophobic pocket improving affinity could explain the difference in affinities. However, current opinion is that imposing conformational restraint is optimal as it provides more favourable entropy and thus improved binding affinities, for example an eight-fold loss of activity is observed when the AICAR inhibitor is linearised.<sup>126</sup>

By synthesising a library of analogues of CLRFT and obtaining their binding affinities by MST the activity of the phenylalanine was investigated. The data showed that when the phenylalanine had a chloro-group in the *para* position, a two-fold increase in binding was observed. This may be due to sterics, rather than electronics. All other modifications showed either a retention or decrease in binding affinity (Appendix 6). This method of inhibitor optimisation has been used previously to improve the activity of an inhibitor of ATIC homodimerisation.<sup>122</sup> A selection of *para*-tyrosine analogues were compared against the active motif of the original *cyclo*-CRYFNV peptide, RY. The substitution of tyrosine's hydroxyl group with the electron-withdrawing nitro-group caused a 120-fold increase in activity, producing an inhibitor with nanomolar activity.<sup>122</sup> The use of similar modifications was also employed to improve the potency of a matrix metalloproteinase inhibitor.<sup>404</sup> Using a structure-based drug design an inhibitor, based on the original malonic acid, had 500-fold improved activity. The increased potency could be attributed to three modifications: increased hydrophobic contacts in the binding pocket; the introduction of a new hydrogen bond from the amine group on the *para*-position of the aromatic group and a reduction in the size of the molecule.<sup>404</sup> Although dramatic improvements were not observed here with the addition of the chloro-group in the *para* position, other modification could be investigated to improve the potency, as discussed in Chapter 5.

As described in Section 4.1 the use of passive immunisation and antibodies has been approved for treatment, with many of them targeting PA. Although they provide good prophylactic treatments and post-exposure therapeutics, they are unable to remove existing toxins and clear the bacteria so are often used in conjunction with antibiotics. Moreover, there are limitations in their uptake into the body, and the production of an antibody therapeutic is very costly due to their size and complexity. Consequently, there are limitations in their availability and distribution. Hence diversification into the area of small molecules has been investigated to identify cheaper alternatives that are commercially viable with improved bioavailability. Although the activity of the inhibitors identified here are low and information regarding the toxicity and *in vivo* activity of the sequences is required they provide a starting point for the development of a more potent small molecule inhibitor of CMG2 and PA. Looking at the current research, most efforts look at the interaction between LF and PA; if this interaction is successfully inhibited, it may not stop the internalisation of EF. Targeting PA oligomerisation although mitigating the progression of the disease cannot be utilised as a prophylactic. However, by binding to CMG2, CLRFT is advantageous with regards to the development of resistance. Resistance can occur naturally, but it can be introduced intentionally for the use of a more virulent strain as a bioweapon. Inhibiting the interaction of a ligand with host receptors used by pathogens represents a powerful strategy to overcome this problem because extensive alterations to the pathogen will be required to allow adaptations to change to a new receptor. Nonetheless, optimisation and further investigation of CLRFT is essential. As outlined in Chapter 5 further analysis of the compound by experiments like FRET or a gel filtration assay, and inhibition of LF into CHO-K1 cells would corroborate these findings. Furthermore, the IC<sub>50</sub> and binding affinity presented here show that the compound is not as potent as inhibitors currently under investigation; this compound must, therefore, be optimised to improve its efficacy.



## **4.6 Materials and methods: Identification of inhibitors of anthrax toxin internalisation**

### ***4.6.1 Construction of the Anthrax RTHS***

Dr. F. Foranov constructed the Anthrax RTHS. Briefly, the genes encoding mouse CMG2<sub>38-218</sub> and *B. anthracis* PA<sub>488-735</sub> were integrated into the chromosome of the heterodimeric SNS126 RTHS strain, as described in Chapter 3.2.1.

### ***4.6.2 Screening of inhibitors***

Each library was plated onto minimal media plates supplemented with 50 µg/ml kanamycin, 5.0 mM 3-AT, 50 µM IPTG and 6.5 µM arabinose and incubated for 72 hours at 37°C. The libraries were generated as described for PspA.

### ***4.6.3 Anthrax protein purification***

#### ***4.6.3.1 Construction of pGEX CMG2<sub>38-218</sub> expression plasmid***

The gene-encoding mouse receptor *CMG2<sub>38-218</sub>* was amplified by PCR using the primers CMG2(*Bam*HI)-F and CMG2(*Eco*RI)-R (annealing temperature of 57°C and extension time 45 seconds) and cloned into the corresponding sites of pGEX-2TK (Novagen) by digesting with the corresponding restriction endonucleases in buffer 4 and BSA. The fragments were ligated at a 1:3 (backbone:insert) ratio at 4°C and transformed initially into chemically-competent DH5α, then for expression into chemically-competent BL21 (DE3). The resulting pGEX *CMG2<sub>38-218</sub>* plasmid encoded an N-terminally glutathione S-transferase (GST)-tagged CMG2<sub>38-218</sub>. This construct was confirmed by DNA sequencing.

#### ***4.6.3.2 Construction of pET28 CMG2<sub>38-218</sub> expression plasmid***

The gene-encoding mouse receptor *CMG2<sub>38-218</sub>* was amplified by PCR using the primers CMG2(*Bam*HI)-F and CMG2(*Sac*I)-R (annealing temperature of 56°C and extension time 45 seconds) and cloned into the corresponding sites of pET28a

(Novagen) by digesting with the corresponding restriction endonucleases in buffer 4 and BSA. The fragments were ligated at a 1:6 (backbone:insert) ratio at 4°C and transformed initially into chemically-competent DH5 $\alpha$ , then for expression into chemically-competent BL21 (DE3). The resulting pET28 CMG2<sub>38-218</sub> plasmid encoded an N-terminally His<sub>6</sub>-tagged CMG2<sub>38-218</sub>. This construct was confirmed by DNA sequencing.

#### *4.6.3.3 Construction of pET28 PA<sub>488-735</sub> expression plasmid*

The gene-encoding *B. anthracis* PA was amplified by PCR using the primers PA(*Bam*HI)-F and PA(*Sac*I)-R (annealing temperature of 56°C and extension temperature 45 seconds) and cloned into the corresponding sites of pET28a (Novagen) by digesting with the corresponding restriction endonucleases in buffer 4 and BSA. The fragments were ligated at a 1:6 (backbone:insert) ratio at 4°C and transformed initially into chemically-competent DH5 $\alpha$ , then for expression into chemically-competent BL21 (DE3). The resulting pET28 PA<sub>488-735</sub> plasmid encodes an N-terminally His<sub>6</sub>-tagged PA<sub>488-735</sub>. This construct was confirmed by DNA sequencing.

#### *4.6.3.4 Expression and purification of GST-tagged CMG2<sub>38-218</sub>*

Glutathione S-transferase (GST)-tagged CMG2<sub>38-218</sub> purification was based on the method described by Wigelsworth *et al.* (2004).<sup>288</sup> Briefly, cells harbouring pGEX CMG2<sub>38-218</sub> were grown in ECPM1 medium (Table 45) for 8 hours and protein expression was induced with 1 mM IPTG and incubated further overnight at 18°C. After incubation cells were harvested by centrifugation and stored at -80°C. Cell were resuspended in PBS and lysed by sonication. The insoluble and soluble fractions were separated by centrifugation at 8,000 rpm for 30 minutes. Then the soluble fraction was purified by glutathione affinity chromatography (as per manufacturer's protocols) using a PBS wash buffer and eluted from the column by the addition of 10 mM reduced glutathione in 50 mM Tris-HCl (pH 8.0). Proteins were analysed by Bradford assay (as per manufacturer's protocols) and 10% SDS-PAGE gels. Yields were typically between 1.1-1.8 mg/ml. The proteins were further purified by gel filtration on the Hi-Load Superdex 200 (GE Healthcare, UK)

equilibrated in 20 mM Tris-HCl (pH 8.0) buffer and stored at -80°C after the addition of 5 mM dithiothreitol.

**Table 45: Composition of ECPM1 media.**

REAGENTS	AMOUNTS
K <sub>2</sub> HPO <sub>4</sub>	4 g
KH <sub>2</sub> PO <sub>4</sub>	1 g
NH <sub>4</sub> Cl	1 g
K <sub>2</sub> SO <sub>4</sub>	2.4 g
CaCl <sub>2</sub> .H <sub>2</sub> O	0.13 g
Trace elements solution (Table 46)	10 ml
Casamino acids	20 g
Yeast extract	3 g
Glycerol	40 g
1 M MgCl <sub>2</sub>	2 ml
Sterile deionised water	Up to 1000 ml

**Table 46: Composition of trace elements solution.**

REAGENTS	AMOUNTS
EDTA	5 g
FeCl <sub>3</sub> . 6H <sub>2</sub> O	500 mg
ZnO	50 mg
CuCl <sub>2</sub> .2H <sub>2</sub> O	10 mg
Co(NO <sub>3</sub> ) <sub>2</sub> .6H <sub>2</sub> O	10 mg
(NH <sub>4</sub> ) <sub>6</sub> Mo <sub>7</sub> O <sub>24</sub> .4H <sub>2</sub> O	10 mg
Sterile deionised water	Up to 1000 ml
Use HCl to dissolve EDTA, then adjust to pH 7.0 with NaOH	

#### 4.6.3.5 Expression and purification of His<sub>6</sub>-tagged CMG2<sub>38-218</sub>

His<sub>6</sub>-tagged CMG2<sub>38-218</sub> purification procedure was based on the method described by Wigelsworth *et al.* (2004) described in Section 4.6.3.4.<sup>288</sup> Here cells harbouring pET28a CMG2<sub>38-218</sub> were used, yielding between 0.80-1.1 mg/ml.

#### *4.6.3.6 Expression and purification of His<sub>6</sub>-tagged PA<sub>488-735</sub>*

His<sub>6</sub>-tagged PA<sub>488-735</sub> purification procedure was based on the method described by Wigelsworth *et al.* (2004) described in Section 4.6.3.4.<sup>288</sup> Here cells harbouring pET28a PA<sub>488-735</sub> were used, yielding between 1.2-1.7 mg/ml.

#### *4.6.3.7 Western blot analysis of His<sub>6</sub>-tagged PA<sub>488-735</sub> and CMG2<sub>38-218</sub>*

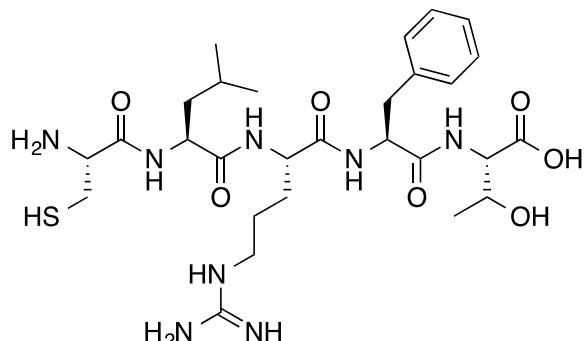
Proteins were separated by 12% SDS-PAGE gel and transferred onto nitrocellulose transfer membranes (Protran, Whatman). The blots were blocked for 1 hour in 5% milk in PBS with 0.05% Tween-20 at room temperature. They were then incubated for 16 hours at 4°C with anti-His (27-4710-01, GE Healthcare) at a 1 in 1,000 dilution in 5% milk in PBS with 0.05% Tween-20 followed by incubation with anti-mouse-HRP (NA931, GE Healthcare) at a 1 in 5,000 dilution at room temperature for 1 hour. Secondary antibody was detected using an ECL reagent (Western C, Biorad).

#### *4.6.3.8 Western blot analysis of GST-tagged CMG2<sub>38-218</sub>*

Proteins were analysed as described in Section 4.6.3.7 with anti-GST (MA4-004, Neomarkers) at a 1 in 500 dilution.

#### 4.6.4 Synthesis of anthrax inhibitors

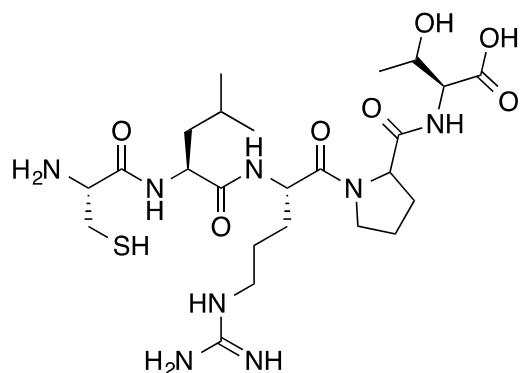
##### 4.6.4.1 Synthesis of linear CLRFT



**Figure 94: Structure of linear CLRFT (9).**

Peptide was synthesised as 2.6.1, yielding 54 mg (34%) of 9 as a white solid. M.p 103.9-124.8°C; <sup>1</sup>H NMR (600 MHz, DMSO-*d*<sub>6</sub>) δ ppm 12.62 (1 H, br. s., Thr-COOH) 8.51 (1 H, d, *J*=7.63 Hz, Leu-NH) 8.13 (1 H, d, *J*=8.24 Hz, Thr-NH) 8.12 (1 H, d, *J*=8.24 Hz, Cys-NH) 7.95 (1 H, d, *J*=7.93 Hz, Phe-NH) 7.51 (1 H, br. s., Arg-NH) 7.19 - 7.29 (4 H, m, Phe-ArH) 7.14 - 7.18 (1 H, m, Phe-ArH) 4.96 (1 H, br. s., Cys-SH or Thr-OH) 4.69 (1 H, td, *J*=8.54, 4.27 Hz, Phe-αH) 4.28 - 4.38 (1 H, m, Leu-αH) 4.20 - 4.27 (2 H, m, Thr-αH and Thr-βH) 4.15 - 4.19 (1 H, m, Arg-αH) 3.97 - 4.06 (1 H, m, Cys-αH) 3.35 (br. s., solvent- H<sub>2</sub>O) 2.95 - 3.12 (4 H, m, Arg-βH, Arg-δH or Phe-βH) 2.76 - 2.88 (2 H, m, Arg-βH or Phe-βH) 1.62 (2 H, d, *J*=5.19 Hz, Cys-βH) 1.35 - 1.51 (5 H, m, Leu-βH, Arg-γH and Leu-γH) 1.00 - 1.10 (3 H, m, Thr-γH) 0.81 - 0.91 (6 H, m, Leu-δH); Analytical HPLC (220 nm) 18.1 min; IR (neat) 3270, 1631, 1525, 1188 cm<sup>-1</sup>; MS (ESI+) *m/z* (%) 639.2 ((M + H)<sup>+</sup>, 100), 320.5 ((M + 2H)<sup>2+</sup> 41.8); HRMS (ESI+) for C<sub>28</sub>H<sub>47</sub>N<sub>8</sub>O<sub>7</sub>S (M + H)<sup>+</sup> calcd 639.3283, found 639.3270 (2.0 ppm error).

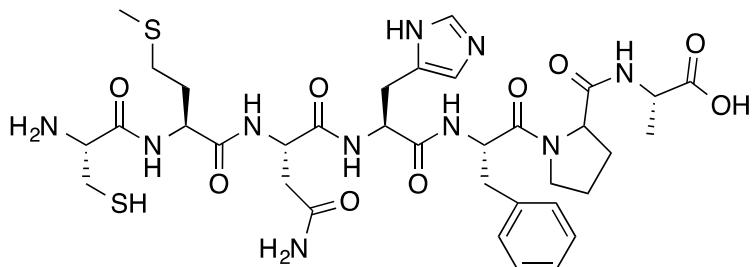
#### 4.6.4.2 Synthesis of linear CLRPT



**Figure 95: Structure of linear CLRPT (10).**

Peptide was synthesised as 2.6.1, yielding 21.0 mg (15%) of 10 as a white solid. M.p 104.6-117.9°C; <sup>1</sup>H NMR (600 MHz, DMSO-*d*<sub>6</sub>) δ ppm 12.43 - 12.71 (1 H, m, Thr-COOH) 8.52 (1 H, d, *J*=7.93 Hz, Leu-NH) 8.25 (1 H, d, *J*=7.63 Hz, Arg-NH) 8.20 (3 H br. s., Phe-NH) 7.86 (1 H, d, *J*=8.24 Hz, Pro-NH) 7.53 (1 H, t, *J*=5.49 Hz, Arg-NH<sub>side-chain</sub>) 4.91 (1 H, br. s., Thr-OH) 4.52 (1 H, dd, *J*=8.39, 3.51 Hz, Pro-αH) 4.45 - 4.50 (1 H, m, Arg-αH) 4.36 (1 H, ddd, *J*=9.84, 7.86, 5.19 Hz, Leu-αH) 4.13 - 4.18 (2 H, m, Thr-αH, Thr-βH) 4.02 - 4.07 (1 H, m, Cys-αH) 3.61 - 3.68 (1 H, m, Pro-βH) 3.02 - 3.15 (3 H, m, Arg-δH) 2.99 (1 H, d, *J*=5.19 Hz, Cys-βH) 2.88 (1 H, d, *J*=14.04 Hz, Cys-βH) 2.00 - 2.12 (1 H, m, Pro-γH) 1.82 - 1.96 (3 H, m, Pro-γH and Pro-δH) 1.62 - 1.73 (2 H, m, Leu-βH) 1.49 - 1.58 (3 H, m, Arg-γH and Leu-γH) 1.42 - 1.49 (2 H, m, Arg-βH) 1.06 - 1.09 (3 H, m, Thr-γH) 0.89 (6 H, d, *J*=6.71 Hz, Leu-δH); Analytical HPLC (220 nm) 16.1 min; IR (neat) 3278, 1655, 1182 cm<sup>-1</sup>; MS (ESI+) *m/z* (%) 589.2 ((M + H)<sup>+</sup>, 100), 315.9 ((M + 2H)<sup>2+</sup> 36.0); HRMS (ESI+) for C<sub>24</sub>H<sub>45</sub>N<sub>8</sub>O<sub>7</sub>S (M + H)<sup>+</sup> calcd 589.3125, found 589.3126 (0.2 ppm error).

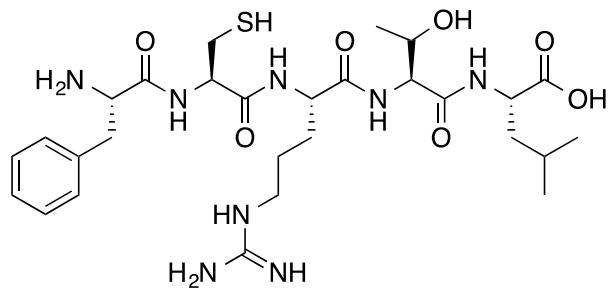
#### 4.6.4.3 Synthesis of linear CMNHFPAs



**Figure 96:** Structure of linear CMNHFPAs (11).

Peptide was synthesised as 2.6.1, yielding 56 mg (27%) of 11 as a white solid. M.p 172.4-228.9°C; <sup>1</sup>H NMR (600MHz, DMSO) δ ppm 12.57 (1H, br. s., Ala-COOH) 8.93 (2H, br. s., His-NH<sub>side-chain</sub>) 8.63 (1H, d, J = 8.5 Hz, Met-NH) 8.33 (2H, t, J = 7.9 Hz, Asn-NH and His-NH) 8.09 (1H, d, J = 7.3 Hz, Ala-NH) 8.03 (1H, d, J = 7.3 Hz, Phe-NH) 7.46 (1H, br. s., His-IndH) 7.42 (1H, m, Phe-ArH) 7.24 - 7.33 (4H, m, Phe-ArH) 7.16 - 7.25 (1H, m, His-IndH) 6.97 - 7.06 (1H, m, Cys-SH) 4.57 - 4.64 (1H, dd, His-αH) 4.46 - 4.56 (1H, m, Phe-αH) 4.39 - 4.45 (1H, m, J=8.5 Hz, Asn-αH) 4.37 (1H, d, J = 8.5 Hz, Met-αH) 4.10 - 4.25 (1H, dd, J=7.3 Hz, Ala-αH) 4.03 (1H, t, J=6, Cys-αH) 3.94 (1H, d, J = 6.1 Hz, Pro-αH) 3.53 - 3.64 (1H, m, Pro-βH) 3.14 - 3.50 (H, m, Pro-βH and Phe-βH) 3.03 (2H, d, J = 15.9 Hz, His-βH) 2.97 (2H, m, Asn-βH) 2.76 - 2.92 (3H, m, Cys-βH) 2.35 - 2.49 (4H, m, Met-γH and Pro-γH) 1.97 - 2.07 (3H, m, Met-δH) 1.81 - 1.95 (3H, m, Met-βH) 1.71 - 1.80 (2H, m, Pro-δH) 1.28 (3H, d, J = 7.3 Hz, Ala-βH); Analytical HPLC (220 nm) 17.6 min; IR (neat) 3279, 1631, 1188 cm<sup>-1</sup>; MS (ESI+) *m/z* (%) 819.2 ((M + H)<sup>+</sup>, 100.0), 410.4 ((M + 2H)<sup>2+</sup> 68.0); HRMS (ESI+) for C<sub>35</sub>H<sub>50</sub>N<sub>10</sub>O<sub>9</sub>S<sub>2</sub> (M + H)<sup>+</sup> calcd 819.3276, found 819.3828 (0.2 ppm error).

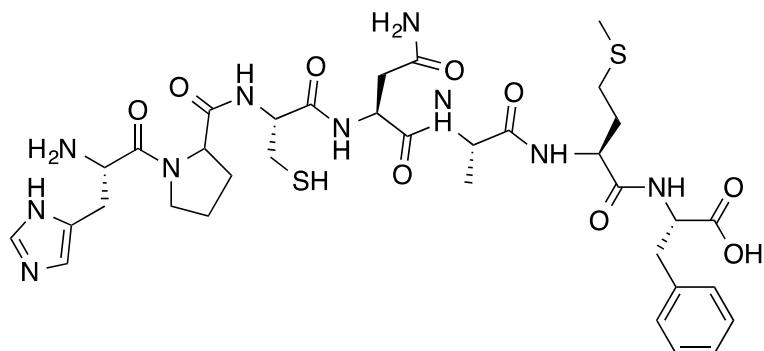
4.6.4.4 Synthesis of linear scrambled CLRFT



**Figure 97: Structure of linear scrambled CLRFT (12).**

Peptide was synthesised as 2.6.1, yielding 56 mg (35%) of 12 as a white solid. M.p 99.7-145.9°C;  $^1\text{H}$  NMR (600 MHz, DMSO- $d_6$ )  $\delta$  ppm 8.78 (1 H, d,  $J=9.77$  Hz, Cys-NH) 8.39 (1 H, d,  $J=7.32$  Hz, Leu-NH) 8.09 - 8.17 (4 H, m, Phe-NH) 7.99 (1 H, d,  $J=7.32$  Hz, Arg-NH) 7.82 (1 H, d,  $J=7.32$  Hz, Thr-NH) 7.60 (1 H, br. s., Arg-NH<sub>side-chain</sub>) 7.30 - 7.33 (2 H, m, Phe-ArH) 7.23 - 7.27 (3 H, m, Phe-ArH) 4.83 (1 H, br. s., Cys-SH) 4.50 - 4.54 (1 H, m, Cys- $\alpha$ H) 4.35 (1 H, q,  $J=7.32$  Hz, Leu- $\alpha$ H) 4.19 - 4.30 (2 H, m, Thr- $\alpha$ H, Phe- $\alpha$ H) 4.14 (1 H, br. s., Thr-OH) 3.90 - 3.97 (2 H, m, Thr- $\beta$ H, Arg- $\alpha$ H) 3.04 - 3.13 (4 H, m, Arg- $\delta$ H, Phe- $\beta$ H) 2.88 - 2.95 (1 H, m, Arg- $\gamma$ H) 2.67 - 2.82 (2 H, m, Arg- $\gamma$ H, Cys- $\beta$ H) 2.41 - 2.48 (1 H, m, Cys- $\beta$ H) 1.72 (1 H, br. s.) 1.61 - 1.68 (1 H, m, Leu- $\beta$ H) 1.49 - 1.56 (6 H, m, Leu- $\gamma$ H, Leu- $\beta$ H) 1.06 (4 H, d,  $J=7.32$  Hz, Thr- $\gamma$ H) 0.88 (5 H, d,  $J=7.32$  Hz, Leu- $\delta$ H) 0.83 (4 H, d,  $J=4.88$  Hz, Leu- $\delta$ H); Analytical HPLC (220 nm) 17.4 min; IR (neat) 3280, 1636, 1134 cm<sup>-1</sup>; MS (ESI+)  $m/z$  (%) 639.3 ((M + H)<sup>+</sup>, 61.3), 320.5 ((M + 2H)<sup>2+</sup> 100.0); HRMS (ESI+) for C<sub>28</sub>H<sub>46</sub>N<sub>8</sub>O<sub>7</sub>S (M + H)<sup>+</sup> calcd 639.3271, found 639.3283 (1.2 ppm error).

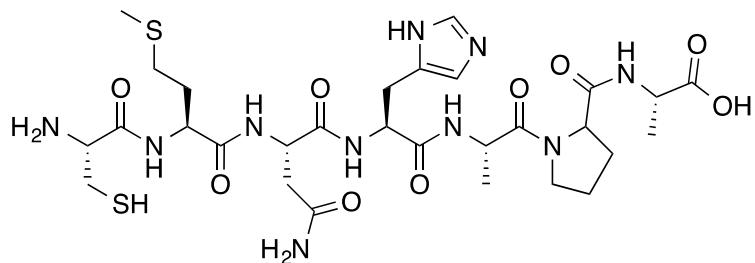
#### 4.6.4.5 Synthesis of linear scrambled CMNHFPAs



**Figure 98: Structure of linear scrambled CMNHFPAs (13).**

Peptide was synthesised as 2.6.1, yielding 45 mg (22%) of 13 as a white solid. M.p 152.4-198.4°C;  $^1\text{H}$  NMR (600 MHz, DMSO- $d_6$ )  $\delta$  ppm 8.92 (1 H, br. s., His-NH<sub>side-chain</sub>) 8.54 (1 H, d,  $J$ =4.39 Hz, Pro-NH) 8.23 (2 H, t,  $J$ =7.32 Hz, His-NH) 8.14 (1 H, d,  $J$ =7.32 Hz, Cys-NH) 8.04 (1 H, d,  $J$ =7.32 Hz, Asn-NH) 7.95 - 8.02 (2 H, m, Met-NH and Phe-NH) 7.93 (1 H, d,  $J$ =7.32 Hz, Ala-NH) 7.45 (2 H, d,  $J$ =10.25 Hz, His-ArH) 7.23 - 7.28 (3 H, m, Phe-ArH, Asn-NH<sub>2-side-chain</sub>) 7.16 - 7.22 (4 H, m, Phe-ArH) 6.98 (1 H, br. s. His-ArH) 4.51 - 4.55 (1 H, m, Cys-SH) 4.48 (1 H, d,  $J$ =7.32 Hz, His- $\alpha$ H) 4.37 - 4.45 (4 H, m, Pro- $\alpha$ H, Cys- $\alpha$ H, Asn- $\alpha$ H) 4.26 - 4.33 (1 H, m, Met- $\alpha$ H and Phe- $\alpha$ H) 4.21 (1 H, dd,  $J$ =7.32, 2.93 Hz, Ala- $\alpha$ H) 3.65 - 3.74 (2 H, m, Pro- $\delta$ H) 3.31 - 3.58 (19 H, m, Pro- $\delta$ H) 3.17 - 3.23 (2 H, m) 3.02 - 3.16 (4 H, m, Cys- $\beta$ H, Asn- $\beta$ H) 2.86 - 2.93 (1 H, m, Asn- $\beta$ H) 2.77 - 2.86 (2 H, m, Cys- $\beta$ H) 2.70 - 2.77 (1 H, m) 2.52 - 2.60 (1 H, m, His- $\beta$ H) 2.46 (2 H, dd,  $J$ =16.11, 7.32 Hz, Phe- $\beta$ H) 2.34 - 2.42 (3 H, m) 2.13 - 2.25 (3 H, m, Pro- $\gamma$ H, Met- $\beta$ H) 2.01 (3 H, s, Met- $\delta$ H) 1.96 (2 H, s) 1.83 - 1.90 (4 H, m, Pro- $\gamma$ H) 1.74 - 1.81 (1 H, m, Phe- $\beta$ H) 1.67 - 1.73 (1 H, m) 1.55 - 1.64 (1 H, m, Met- $\beta$ H) 1.13 - 1.20 (5 H, m, Ala- $\beta$ H); Analytical HPLC (220 nm) 17.8 min; IR (neat) 3261, 1622, 1132 cm<sup>-1</sup>; MS (ESI+)  $m/z$  (%) 819.4 ((M + H)<sup>+</sup>, 100.0); HRMS (ESI+) for C<sub>35</sub>H<sub>50</sub>N<sub>10</sub>O<sub>9</sub>S<sub>2</sub> (M + H)<sup>+</sup> calcd 819.3282, found 819.3276 (-0.6 ppm error).

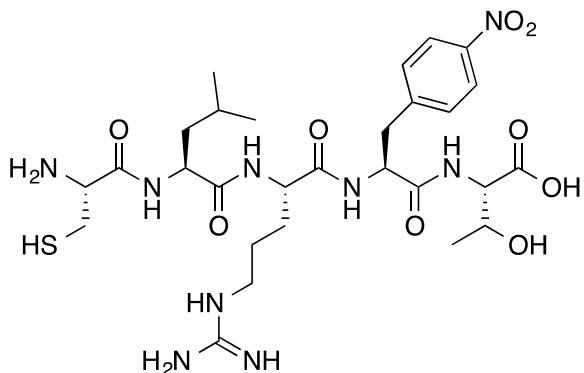
#### 4.6.4.6 Synthesis of linear CMNHAPA



**Figure 99: Structure of linear CMNHAPA (14).**

Peptide was synthesised as 2.6.1, yielding 56 mg (30%) of 14 as a white solid. M.p 169.4-205.8°C;  $^1\text{H}$  NMR (600 MHz, DMSO- $d_6$ )  $\delta$  ppm 9.45 (1 H, s, His-NH<sub>side-chain</sub>) 9.14 (1 H, d,  $J$ =7.81 Hz, Asn-NH) 8.82 (2 H, d,  $J$ =7.81 Hz, His-NH) 8.71 - 8.78 (3 H, m, Ala-NH<sub>2</sub>) 8.59 (1 H, d,  $J$ =7.81 Hz, Ala-NH<sub>1</sub>) 8.55 (1 H, d,  $J$ =7.81 Hz, Met-NH) 7.94 (1 H, br. s., Asn-NH<sub>side-chain</sub>) 7.87 (1 H, s, His- $\gamma$ H) 7.48 (1 H, br. s. Asn-NH<sub>side-chain</sub>) 4.81 - 5.04 (8 H, m, Ala<sub>1</sub>- $\alpha$ H, Met- $\alpha$ H, Asn- $\alpha$ H, His- $\alpha$ H) 4.66 (2 H, t,  $J$ =7.81 Hz, Ala<sub>2</sub>- $\alpha$ H) 4.53 (2 H, br. s., Cys- $\alpha$ H) 4.05 - 4.12 (1 H, m, Pro- $\alpha$ H) 3.99 - 4.04 (1 H, m, Pro- $\alpha$ H) 3.52 - 3.59 (1 H, m, Met- $\beta$ H) 3.42 - 3.49 (3 H, m, Cys- $\beta$ H, Met- $\beta$ H) 2.88 - 3.02 (8 H, m, Pro- $\gamma$ H, Met- $\gamma$ H, His- $\beta$ H) 2.50 - 2.52 (5 H, m, Asn- $\beta$ H) 2.32 - 2.46 (6 H, m, Pro- $\beta$ H, Asn- $\beta$ H) 2.22 - 2.30 (2 H, m, Pro- $\gamma$ H) 1.68 - 1.77 (9 H, m, Met- $\delta$ H, Ala<sub>1</sub>- $\beta$ H, Ala<sub>2</sub>- $\beta$ H); Analytical HPLC (220 nm) 15.9 min; IR (neat) 3273, 1652, 1133 cm<sup>-1</sup>; MS (ESI+)  $m/z$  (%) 743.3 ((M + H)<sup>+</sup>, 77.1), 372.5 ((M + 2H)<sup>+</sup>, 100.0)); HRMS (ESI+) for C<sub>26</sub>H<sub>46</sub>N<sub>10</sub>O<sub>9</sub>S<sub>2</sub> (M + H)<sup>+</sup> calcd 743.2953, found 743.2963 (1.0 ppm error).

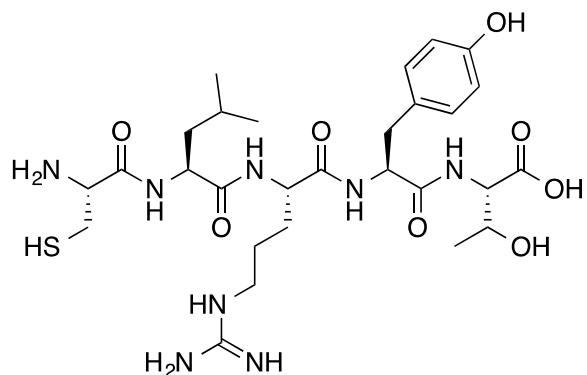
#### 4.6.4.7 Synthesis of linear CLRF(4-NO<sub>2</sub>)T (15)



**Figure 100: Structure of linear CLRF(4-NO<sub>2</sub>)T (15).**

Peptide was synthesised as 2.6.1, yielding 65 mg (40%) of 15 as a white solid. M.p 152.9-183.6°C; <sup>1</sup>H NMR (600 MHz, DMSO-*d*<sub>6</sub>) δ ppm 8.51 (1H, d, *J*=7.81, Leu-NH) 8.26 (d, *J*=7.81 Hz, 1 H, Thr-NH) 8.15 (d, *J*=7.81 Hz, 1 H, Phe-NH) 8.10 (d, *J*=7.81 Hz, 2 H, Phe-ArH) 8.03 -8.07 (m, 1 H, Arg-NH) 7.56 (d, *J*=7.81 Hz, 2 H, Phe-ArH) 5.03 (br. s., 1 H, Cys-SH) 4.75 - 4.85 (m, 1 H, Arg-αH) 4.32 (t, *J*=11.72 Hz, 1 H, Leu- αH) 4.15 - 4.26 (m, 3 H, Phe-αH, Thr-α, Thr-βH) 4.03 (br. s., 1 H, Cys-αH) 3.21 (d, *J*=13.67, 3.91 Hz, 2 H, Arg-βH) 3.02 - 3.08 (m, 2 H, Phe-βH) 2.93 (dd, *J*=13.67, 9.77 Hz, 1 H, Thr-βH) 1.60 (d, *J*=5.86 Hz, 3 H, Arg-δH, Leu-βH) 1.37 - 1.51 (m, 5 H, Leu-γH, Cys-βH, Leu-βH) 1.29 - 1.36 (m, 2 H, Arg-γH ) 1.05 (d, *J*=5.86 Hz, 3 H, Thr-δH) 0.82-0.85 (d, *J*=7.81 Hz, 6 H, Leu-δH); Analytical HPLC (220 nm) 19.8 min; IR (neat) 3275, 1637, 1135 cm<sup>-1</sup>; MS (ESI+) *m/z* (%) 684.3 ((M + H)<sup>+</sup>, 100.0); HRMS (ESI+) for C<sub>28</sub>H<sub>45</sub>N<sub>9</sub>O<sub>9</sub>S (M + H)<sup>+</sup> calcd 684.3133, found 684.3134 (0.2 ppm error).

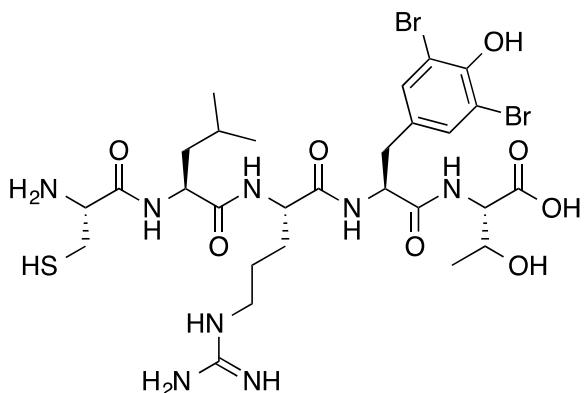
#### 4.6.4.8 Synthesis of linear CLRYT



**Figure 101: Structure of linear CLRYT (16).**

Peptide was synthesised as 2.6.1, yielding 63 mg (37%) of 16 as a white solid. M.p 103.2-124.8°C; <sup>1</sup>H NMR (600 MHz, DMSO-*d*<sub>6</sub>) δ ppm 9.16 (1 H, br. s., Tyr-OH) 8.52 (1 H, d, *J*=7.81 Hz, Leu-NH) 8.17 (1 H, d, *J*=7.81 Hz, Thr-NH) 8.04 (1 H, d, *J*=7.81 Hz, Tyr-NH) 7.86 (1 H, d, *J*=9.77 Hz, Arg-NH) 7.55 (1 H, br. s., Arg-NH<sub>side-chain</sub>) 7.01 - 7.06 (2 H, m, Tyr-ArH) 6.61 (2 H, d, *J*=7.81 Hz, Tyr-ArH) 4.94 (1 H, br. s., Cys-SH) 4.55 - 4.64 (1 H, m, Arg-αH) 4.31 - 4.38 (1 H, m, Leu-αH) 4.24 - 4.29 (1 H, m, Tyr-αH) 4.19 - 4.23 (1 H, m, Thr-OH) 4.16 (1 H, d, *J*=5.86 Hz, Thr-αH) 4.01 - 4.05 (1 H, m, Cys-αH) 3.28 - 3.50 (8 H, m, Solvent-H<sub>2</sub>O) 3.02 - 3.10 (3 H, m, Thr-βH, Tyr-βH) 2.91 - 3.01 (2 H, m, Cys-βH) 2.86 (1 H, dd, *J*=13.67, 3.91 Hz, Arg-βH) 2.70 (1 H, dd, *J*=13.67, 7.81 Hz, Arg-βH) 1.57 - 1.69 (2 H, m, Leu-βH) 1.37 - 1.52 (7 H, m, Leu-γH, Arg-γH, Thr-βH) 1.04 (3 H, d, *J*=7.81 Hz, Thr-γH) 0.83 - 0.91 (4 H, m, Leu-δH); Analytical HPLC (220 nm) 18.9 min; IR (neat) 3280, 1638, 1134 cm<sup>-1</sup>; MS (ESI+) *m/z* (%) 655.3 ((M + H)<sup>+</sup>, 68.1), 328.4 ((M + 2H)<sup>+</sup>, 100.0)); HRMS (ESI+) for C<sub>28</sub>H<sub>45</sub>N<sub>9</sub>O<sub>9</sub>S (M + H)<sup>+</sup> calcd 655.3225, found 655.3232 (1.1 ppm error).

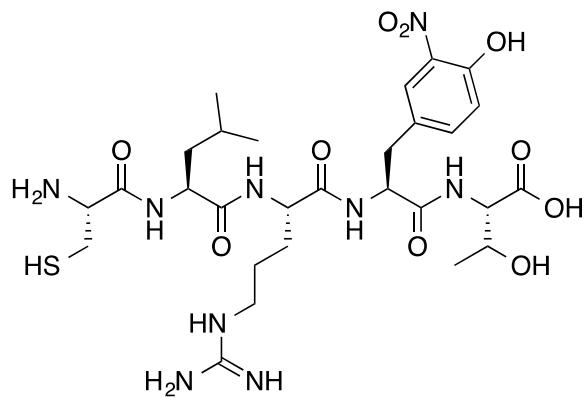
#### 4.6.4.9 Synthesis of linear CLRY(3,5-Br<sub>2</sub>)T



**Figure 102: Structure of linear CLRY(3,5-Br<sub>2</sub>)T (17).**

Peptide was synthesised as 2.6.1, yielding 27 mg (53%) of 17 as a white solid. M.p 132.5-184.9°C; <sup>1</sup>H NMR (600 MHz, DMSO-*d*<sub>6</sub>) δ ppm 12.64 (1 H, br. s., Thr-COOH) 9.66 (1 H, br. s. Tyr-OH) 8.50 (1 H, d, *J*=7.32 Hz, Leu-NH) 8.27 (1 H, d, *J*=7.32 Hz, Thr-NH) 8.19 (3 H, d, *J*=7.32 Hz, Tyr-NH) 7.84 (1 H, d, *J*=7.32 Hz, Arg-NH) 7.48 (2 H, s, Tyr-ArH) 5.02 (1 H, br. s., Cys-SH) 4.64 (1 H, br. s. Arg-αH) 4.31 - 4.38 (1 H, m, Leu-αH) 4.15 - 4.25 (3 H, m, Thr-αH, Tyr-αH) 4.02 (1 H, br. s., Cys-αH) 3.06 (3 H, d, *J*=7.32 Hz, Arg-βH, Cys-βH) 2.98 (2 H, d, *J*=12.21 Hz, Arg-βH, Arg-γH) 2.85 (1 H, d, *J*=12.21 Hz, Arg-δH) 2.66 (1 H, dd, *J*=14.65, 9.77 Hz, Arg-δH) 1.58 - 1.70 (2 H, m, Leu-βH, Thr-βH) 1.40 - 1.54 (7 H, m, Leu-βH, Leu-γH, Arg-γH) 1.04 (3 H, d, *J*=7.32 Hz, Thr-γH) 0.81 - 0.90 (6 H, m, Leu-δH); Analytical HPLC (220 nm) 19.8 min; IR (neat) 3271, 1643, 1134 cm<sup>-1</sup>; MS (ESI+) *m/z* (%) 813.1 ((M + H)<sup>+</sup>, 100.0), 407.1 ((M + 2H)<sup>+</sup>, 78.1); HRMS (ESI+) for C<sub>28</sub>H<sub>44</sub>Br<sub>2</sub>N<sub>8</sub>O<sub>8</sub>S (M + H)<sup>+</sup> calcd 811.1452, found 811.1442 (-1.2 ppm error).

**4.6.4.10 Synthesis of linear CLRY(2-NO<sub>2</sub>)T**

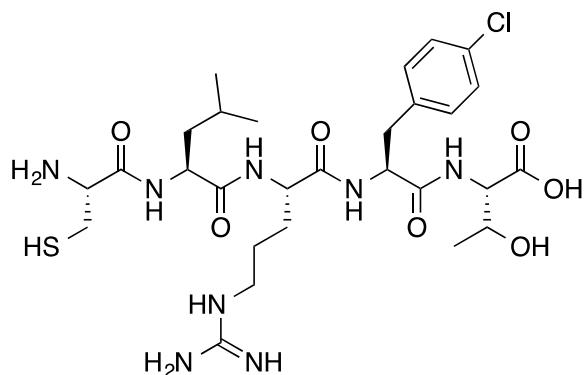


**Figure 103: Structure of linear CLRY(3-NO<sub>2</sub>)T (18).**

Peptide was synthesised as 2.6.1, yielding 66 mg (38%) of 18 as an orange solid.

M.p 161.8-184.9°C; <sup>1</sup>H NMR (600 MHz, DMSO-d<sub>6</sub>) δ ppm 10.77 (1 H, br. s., Thr-COOH) 9.23 - 9.30 (1 H, m, Tyr-OH) 8.49 (1 H, d, J=7.32 Hz, Leu-NH) 8.17 - 8.25 (4 H, m, Thr-NH) 8.15 (1 H, d, J=7.32 Hz, Tyr-NH) 7.93 (1 H, d, J=7.32 Hz, Arg-NH) 7.84 (1 H, s, Tyr-ArH) 7.52 (1 H, br. s., Arg-NH<sub>side-chain</sub>) 7.47 (1 H, d, J=9.77 Hz, Tyr-ArH) 7.00 (1 H, d, J=7.32 Hz, Tyr-ArH) 5.01 (1 H, br. s., Cys-SH) 4.65 - 4.73 (1 H, m, Arg-αH) 4.28 - 4.36 (1 H, m, Leu-αH) 4.16 - 4.26 (3 H, m, Thr-αH, Tyr-αH) 4.03 (1 H, br. s., Cys-αH) 2.94 - 3.10 (5 H, m, Arg-βH, Leu-βH, Tyr-βH) 2.82 - 2.89 (1 H, m Cys-βH) 2.75 (1 H, dd, J=14.65, 9.77 Hz, Arg-γH) 2.56 - 2.62 (1 H, m, Cys-βH) 1.61 (2 H, br. s., Leu-δH, Leu-γH) 1.34 - 1.52 (6 H, m, Arg-δH, Leu-γH, Thr-βH) 1.04 (3 H, d, J=7.32 Hz, Thr-γH) 0.82 - 0.88 (6 H, m, Leu-δH); Analytical HPLC (220 nm) 17.6 min; IR (neat) 3284, 1627, 1182 cm<sup>-1</sup>; MS (ESI+) m/z (%) 700.2 ((M + H)<sup>+</sup>, 91.0), 350.9 ((M + 2H)<sup>+</sup>, 100.0)); HRMS (ESI+) for C<sub>28</sub>H<sub>45</sub>N<sub>9</sub>O<sub>10</sub>S (M + H)<sup>+</sup> calcd 700.3072, found 700.3083 (1.5 ppm error).

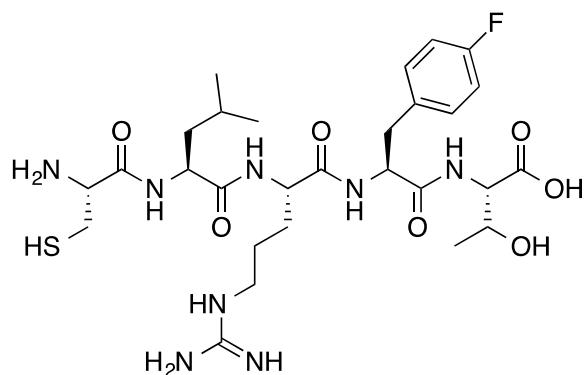
#### 4.6.4.11 Synthesis of linear CLRF(4-Cl)T (19)



**Figure 104: Structure of linear CLRF(4-Cl)T (19).**

Peptide was synthesised as 2.6.1, yielding 19 mg (11%) of 19 as a white solid. M.p 130.0-162.0°C; <sup>1</sup>H NMR (600 MHz, DMSO-*d*<sub>6</sub>) δ ppm 8.51 (1 H, d, *J*=7.93 Hz, Leu-NH) 8.15 - 8.19 (2 H, m, Arg-NH and Thr-NH) 7.97 (1 H, d, *J*=7.93 Hz, Phe-NH) 7.50 - 7.58 (1 H, m, Arg-NH<sub>side-chain</sub>) 7.23 - 7.31 (5 H, m, Phe-ArH) 4.99 (1 H, br. s., Thr-OH or Ser-OH) 4.68 (1 H, td, *J*=8.54, 4.27 Hz, Phe-αH) 4.31 - 4.36 (1 H, m, Leu-αH) 4.24 (1 H, q, *J*=7.43 Hz, Arg-αH) 4.18 - 4.22 (1 H, m) 4.17 (1 H, br. s., Thr-αH) 4.03 (1 H, t, *J*=5.04 Hz, Cys-αH) 3.33 (12 H, br. s., Solvent-H<sub>2</sub>O) 3.03 - 3.08 (4 H, m, Phe-βH, Arg-δH) 2.99 (1 H, dd, *J*=14.34, 5.80 Hz, Cys-βH) 2.85 (1 H, dd, *J*=14.34, 4.58 Hz, Cys-βH) 2.79 (1 H, dd, *J*=14.04, 9.16 Hz, Phe-βH) 1.59 - 1.66 (3 H, m, Arg-βH and Leu-βH) 1.35 - 1.51 (7 H, m, Leu-γH and Arg-γH) 1.04 (4 H, d, *J*=6.41 Hz, Thr-βH and Thr-γH) 0.89 (3 H, d, *J*=6.71 Hz, Leu-γH) 0.85 (4 H, d, *J*=6.41 Hz, Leu-δH); Analytical HPLC (220 nm) 18.6 min; IR (neat) 3271, 1629, 1133 cm<sup>-1</sup>; MS (ESI+) *m/z* (%) 673.3 ((M + H)<sup>+</sup>, 99.9), 337.2 ((M + 2H)<sup>+</sup>, 100.0)); HRMS (ESI+) for C<sub>28</sub>H<sub>45</sub>ClN<sub>8</sub>O<sub>7</sub>S (M + H)<sup>+</sup> calcd 673.2887, found 673.2893 (0.9 ppm error).

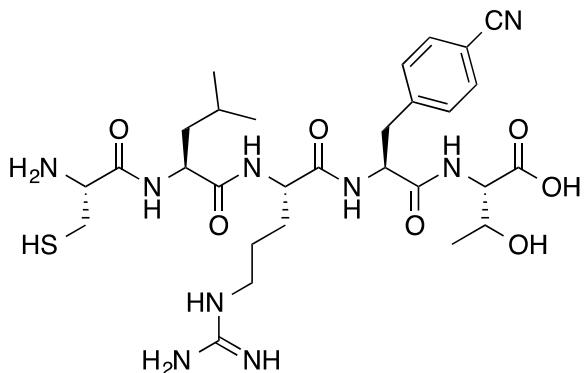
**4.6.4.12 Synthesis of linear CLRF(4-F)T (20)**



**Figure 105: Structure of linear CLRF(4-F)T (20).**

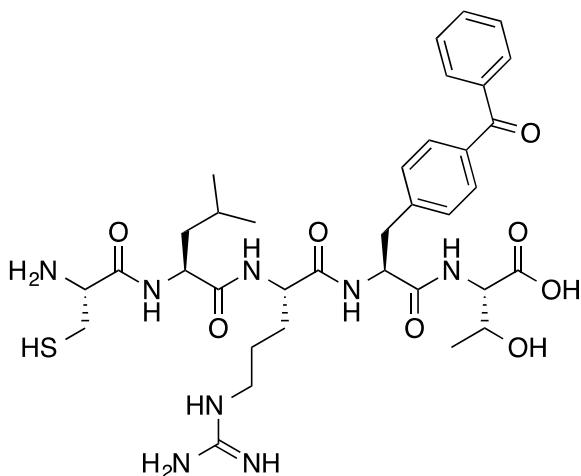
Peptide was synthesised as 2.6.1, yielding 48 mg (29%) of 20 as a white solid. M.p 163.6-189.1°C;  $^1\text{H}$  NMR (600 MHz, DMSO- $d_6$ )  $\delta$  ppm 12.63 (1 H, br. s., Thr-COOH) 8.51 (1 H, d,  $J$ =7.32 Hz, Leu-NH) 8.19 (2 H, br. s., Arg-NH) 8.15 (1 H, d,  $J$ =7.32 Hz, Thr-NH) 7.92 - 7.98 (1 H, m, Phe-NH) 7.50 (1 H, br. s., Arg-NH<sub>side-chain</sub>) 7.25 - 7.32 (2 H, m, Phe-ArH) 7.02 (2 H, t,  $J$ =7.32 Hz, Phe-ArH) 4.99 (1 H, br. s., Cys-SH) 4.66 - 4.71 (1 H, m, Phe- $\alpha$ H) 4.30 - 4.35 (1 H, m, Leu- $\alpha$ H) 4.14 - 4.27 (3 H, m, Thr- $\alpha$ H, Arg- $\alpha$ H) 4.03 (1 H, br. s., Cys- $\alpha$ H) 3.36 (6 H, br. s., Solvent-H<sub>2</sub>O) 2.96 - 3.10 (5 H, m, Phe- $\beta$ H, Arg- $\delta$ H, Thr- $\beta$ H) 2.85 (1 H, d,  $J$ =12.21 Hz, Arg- $\beta$ H) 2.78 (1 H, dd,  $J$ =14.65, 9.77 Hz, Arg- $\beta$ H) 1.62 (2 H, d,  $J$ =4.88 Hz, Leu- $\beta$ H, Thr- $\beta$ H) 1.33 - 1.51 (6 H, m, Leu- $\beta$ H, Leu- $\gamma$ H, Arg- $\gamma$ H) 1.05 (4 H, d,  $J$ =4.88 Hz, Thr- $\gamma$ H) 0.83 - 0.89 (6 H, m, Leu- $\delta$ H); Analytical HPLC (220 nm) 18.1 min; IR (neat) 3272, 1632, 1188 cm<sup>-1</sup>; MS (ESI+)  $m/z$  (%) 657.3 ((M + H)<sup>+</sup>, 51.3), 304.4 ((M + 2H)<sup>+</sup>, 100.0)); HRMS (ESI+) for C<sub>28</sub>H<sub>45</sub>FN<sub>8</sub>O<sub>7</sub>S (M + H)<sup>+</sup> calcd 657.3199, found 657.3189 (-1.6 ppm error).

#### 4.6.4.13 Synthesis of linear CLRF(4-CN)T



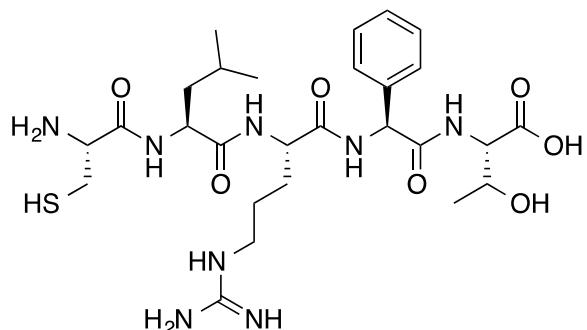
**Figure 106: Structure of linear CLRF(4-CN)T (21).**

Peptide was synthesised as 2.6.1, yielding 60 mg (36%) of 21 as a white solid. M.p 98.7118.4°C; <sup>1</sup>H NMR (600 MHz, DMSO-*d*<sub>6</sub>) δ ppm 12.65 (1 H, br. s., Thr-COOH) 8.51 (1 H, d, *J*=7.32 Hz, Leu-NH) 8.22 (1 H, d, *J*=9.77 Hz, Thr-NH) 8.19 (3 H, br. s., Arg-NH) 8.15 (1 H, d, *J*=7.32 Hz, Phe-NH) 8.02 (1 H, d, *J*=9.77 Hz, Phe-ArH) 7.68 (1 H, d, *J*=9.77 Hz, Phe-ArH) 7.48 (3 H, d, *J*=7.32 Hz, Phe-ArH, Arg-NH<sub>side-chain</sub>) 5.02 (1 H, br. s., Cys-SH) 4.73 - 4.80 (1 H, m, Phe-αH) 4.30 - 4.37 (1 H, m, Leu-αH) 4.16 - 4.26 (3 H, m, Thr-αH, Thr-βH, Arg-αH) 3.99 - 4.09 (1 H, m, Cys-αH) 3.12 - 3.19 (1 H, m, Phe-βH) 2.97 - 3.08 (3 H, m, Cys-βH, Arg-δH) 2.82 - 2.91 (2 H, m, Arg-βH, Phe-βH) 2.58 - 2.66 (1 H, m, Arg-βH) 1.55 - 1.68 (2 H, m, Leu-βH) 1.31 - 1.50 (6 H, m, Leu-γH, Arg-γH) 1.05 (3 H, d, *J*=7.32 Hz, Thr-γH) 0.83 - 0.91 (6 H, m, Leu-δH); Analytical HPLC (220 nm) 17.6 min; IR (neat) 3278, 1642, 1134 cm<sup>-1</sup>; MS (ESI+) *m/z* (%) 664.1 ((M + H)<sup>+</sup>, 76.8), 333.0 ((M + 2H)<sup>+</sup>, 100.0)); HRMS (ESI+) for C<sub>29</sub>H<sub>45</sub>N<sub>9</sub>O<sub>7</sub>S (M + 2H)<sup>+</sup> calcd 332.6660, found 332.6654 (-1.7 ppm error).

**4.6.4.14 Synthesis of linear CLRF(4-Bz)T (22)****Figure 107: Structure of linear CLRF(4-Bz)T (22).**

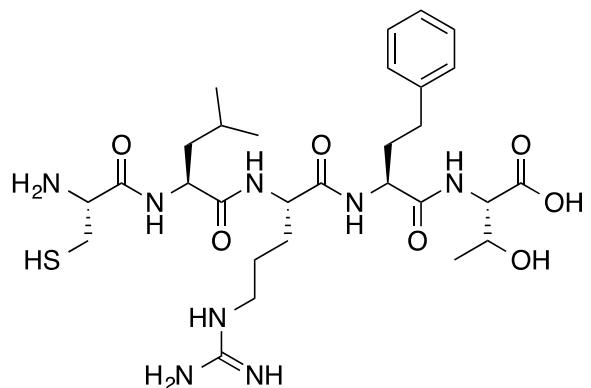
Peptide was synthesised as 2.6.1, yielding 59 mg (32%) of 22 as a white solid. M.p 108.5-127.9°C; <sup>1</sup>H NMR (600 MHz, DMSO-*d*<sub>6</sub>) δ ppm 12.64 (1 H, br. s., Thr-COOH) 8.50 (1 H, d, *J*=7.32 Hz, Leu-NH) 8.23 (2 H, d, *J*=7.32 Hz, Thr-NH, Arg-NH) 8.03 (1 H, d, *J*=9.77 Hz, Phe-NH) 7.67 - 7.72 (2 H, m, Phe-ArH) 7.63 (2 H, d, *J*=7.32 Hz, Phe-ArH) 7.56 (2 H, t, *J*=7.32 Hz, Phe-ArH) 7.47 - 7.52 (1 H, m, Arg-NH<sub>side-chain</sub>) 7.45 (2 H, d, *J*=7.32 Hz, Phe-ArH) 4.89 - 5.11 (1 H, m., Cys-SH) 4.76 - 4.81 (1 H, m, Phe-αH) 4.30 - 4.35 (1 H, m, Leu-αH) 4.21 - 4.30 (2 H, m, Thr-αH, Arg-αH) 4.18 (1 H, d, *J*=4.88 Hz, Cys-αH) 4.02 (1 H, br. s., Thr-OH) 3.18 (1 H, d, *J*=9.77 Hz, Phe-βH) 3.02 - 3.08 (3 H, m, Thr-βH, Arg-δH, Cys-βH) 2.89 - 3.01 (2 H, m, Arg-δH, Phe-βH) 2.79 - 2.88 (1 H, m, Arg-βH) 2.54 - 2.65 (1 H, m, Arg-βH) 1.55 - 1.69 (2 H, m, Leu-βH) 1.34 - 1.52 (6 H, m, Leu-γH, Arg-γH) 1.04 - 1.08 (4 H, m, Thr-γH) 0.77 - 0.81 (5 H, m, Leu-δH); Analytical HPLC (220 nm) 18.9 min; IR (neat) 3277, 1639, 1135 cm<sup>-1</sup>; MS (ESI+) *m/z* (%) 745.4 ((M + H)<sup>+</sup>, 45.5), 372.5 ((M + 2H)<sup>+</sup>, 54.3); HRMS (ESI+) for C<sub>35</sub>H<sub>50</sub>N<sub>8</sub>O<sub>8</sub>S (M + 2H)<sup>+</sup> calcd 372.1818, found 372.1809 (-2.5 ppm error).

#### 4.6.4.15 Synthesis of linear CLR(Phg)T



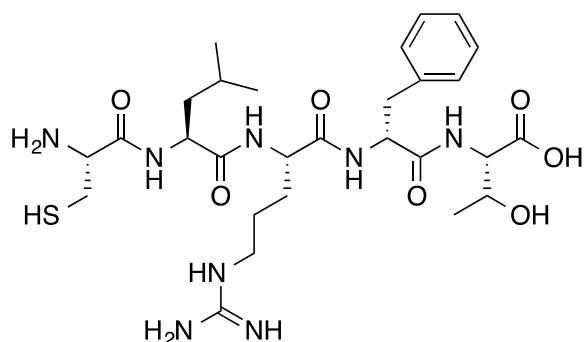
**Figure 108: Structure of linear CLR(Phg)T (23).**

Peptide was synthesised as 2.6.1, yielding 55 mg (35%) of 23 as a white solid. M.p 159.9-202.9°C; <sup>1</sup>H NMR (600 MHz, DMSO-*d*<sub>6</sub>) δ ppm 8.54 (1 H, d, *J*=7.81 Hz, Leu-NH) 8.34 (2 H, br. s., Thr-NH, Phe-NH) 8.29 (1 H, d, *J*=7.81 Hz, Arg-NH) 7.55 - 7.74 (1 H, m, Arg-NH<sub>side-chain</sub>) 7.44 (2 H, d, *J*=7.81 Hz, Phe-ArH) 7.25 - 7.34 (3 H, m, Phe-ArH) 5.66 (1 H, d, *J*=7.81 Hz, Cys-SH) 4.84 - 5.08 (1 H, m Thr-αH) 4.32 - 4.45 (2 H, m, Leu-αH, Arg-αH) 4.21 (1 H, d, *J*=5.86 Hz, Phe-αH) 4.09 - 4.17 (1 H, m, Cys-αH) 3.99 - 4.06 (1 H, m Thr-βH) 3.33 (8 H, br. s., Phe-βH, Phe-γH) 3.04 - 3.16 (2 H, m, Arg-δH) 2.96 - 3.03 (1 H, m, Cys-βH) 2.81 - 2.91 (1 H, m, Cys-βH) 1.72 (1 H, br. s., Arg-βH) 1.60 - 1.69 (1 H, m, Leu-βH) 1.42 - 1.58 (5 H, m, Leu-γH, Leu-βH, Arg-γH, Arg-βH) 1.05 (3 H, d, *J*=5.86 Hz, Thr-γH) 0.85 - 0.90 (6 H, m, Leu-δH); Analytical HPLC (220 nm) 17.4 min; IR (neat) 3270, 1630, 1137 cm<sup>-1</sup>; MS (ESI+) *m/z* (%) 625.4 ((M + H)<sup>+</sup>, 74.4), 313.5 ((M + 2H)<sup>+</sup>, 100.0); HRMS (ESI+) for C<sub>27</sub>H<sub>44</sub>N<sub>8</sub>O<sub>7</sub>S (M + H)<sup>+</sup> calcd 625.3133, found 625.3126 (-1.0 ppm error).

**4.6.4.16 Synthesis of linear CLR(hPhe)T (24)****Figure 109: Structure of linear CLR(hPhe)T (24).**

Peptide was synthesised as 2.6.1, yielding 63 mg (37%) of 24 as a white solid. M.p 109.4–117.9°C; <sup>1</sup>H NMR (600 MHz, DMSO-*d*<sub>6</sub>) δ ppm 8.54 (1 H, d, *J*=7.81 Hz, Leu-NH) 8.30 (1 H, d, *J*=7.81 Hz, Phe-NH) 8.11 (2 H, d, *J*=7.81 Hz, Arg-NH) 7.82 - 7.89 (1 H, m, Thr-NH) 7.58 (1 H, br. s., Arg-NH<sub>side-chain</sub>) 7.22 – 7.30 (2 H, m, Phe-ArH) 7.11 – 7.22 (3 H, m, Phe-ArH) 4.87 – 5.06 (1 H, m, Cys-SH, Thr-OH) 4.29 – 4.46 (3 H, m, Leu-αH, Arg-αH) 4.11 – 4.24 (2 H, m, Thr-αH) 4.00 – 4.05 (1 H, m, Cys-αH) 3.25 – 3.40 (9 H, m, Thr-βH) 3.09 (2 H, d, *J*=5.86 Hz, Arg-δH) 2.98 – 3.03 (1 H, m, Cys-βH) 2.81 – 2.90 (1 H, m, Cys-βH) 2.52 – 2.69 (3 H, m Arg-βH) 1.90 - 2.01 (1 H, m, Arg-γH) 1.77 - 1.87 (1 H, m, Arg-γH) 1.60 - 1.74 (2 H, m, Leu-βH,) 1.44 - 1.57 (6 H, m, Leu-γH, Leu-βH) 1.05 (3 H, d, *J*=5.86 Hz, Thr-γH) 0.84 - 0.88 (6 H, m, Leu-δH); Analytical HPLC (220 nm) 18.2 min; IR (neat) 3272, 1631, 1134 cm<sup>-1</sup>; MS (ESI+) *m/z* (%) 653.4 ((M + H)<sup>+</sup>, 70.8), 327.5 ((M + 2H)<sup>+</sup>, 100.0)); HRMS (ESI+) for C<sub>29</sub>H<sub>48</sub>N<sub>8</sub>O<sub>7</sub>S (M + H)<sup>+</sup> calcd 653.3430, found 653.3439 (1.5 ppm error).

#### 4.6.4.17 Synthesis of linear CLRdFT



**Figure 110:** Structure of linear CLRdFT (25).

Peptide was synthesised as 2.6.1, yielding 52 mg (33%) of 25 as a white solid.

119.0-164.9°C;  $^1\text{H}$  NMR (600 MHz, DMSO- $d_6$ )  $\delta$  ppm 8.49 (1 H, d,  $J=7.81$  Hz, Leu-NH) 8.29 (1 H, d,  $J=9.77$  Hz, Thr-NH) 8.15 (3 H, d,  $J=7.81$  Hz, Phe-NH) 8.05 - 8.09 (1 H, m, Arg-NH) 7.40 (1 H, t,  $J=5.86$  Hz, Arg-NH<sub>side-chain</sub>) 7.25 - 7.30 (2 H, m, Phe-ArH) 7.19 - 7.25 (2 H, m, Phe-ArH) 7.13 - 7.19 (1 H, m, Phe-ArH) 4.96 (1 H, br. s., Cys-SH) 4.72 (1 H, td,  $J=9.77$ , 3.91 Hz, Thr- $\alpha$ H) 4.24 - 4.34 (1 H, m, Leu- $\alpha$ H) 4.11 - 4.23 (2 H, m, Arg- $\alpha$ H, Phe- $\alpha$ H) 4.02 (1 H, br. s., Cys- $\alpha$ H) 3.35 (7 H, br. s. Phe- $\beta$ H) 3.13 (2 H, d,  $J=9.77$  Hz, Thr- $\beta$ H) 2.83 - 3.01 (3 H, m, Cys- $\beta$ H, Thr- $\beta$ H) 2.66 - 2.76 (2 H, m, Arg- $\delta$ H) 1.59 - 1.70 (1 H, m, Leu- $\beta$ H) 1.32 - 1.50 (3 H, m, Leu- $\beta$ H, Leu- $\gamma$ H, Arg- $\beta$ H) 1.22 - 1.30 (1 H, m- Arg- $\gamma$ H) 1.08 - 1.18 (1 H, m, Arg- $\gamma$ H) 1.00 (3 H, d,  $J=5.86$  Hz, Thr- $\gamma$ H) 0.82 - 0.91 (6 H, m, Leu- $\delta$ H); Analytical HPLC (220 nm) 17.7 min; IR (neat) 3278, 1643, 1133 cm<sup>-1</sup>; MS (ESI+)  $m/z$  (%) 639.3 ((M + H)<sup>+</sup>, 65.8), 320.5 ((M + 2H)<sup>+</sup>, 100.0)); HRMS (ESI+) for C<sub>28</sub>H<sub>46</sub>N<sub>8</sub>O<sub>7</sub>S (M + H)<sup>+</sup> calcd 639.3281, found 639.3283 (0.3 ppm error).

#### **4.6.5 In vitro assays for Anthrax inhibitor screening**

##### **4.6.5.1 Sandwich ELISA with GST-tagged CMG2<sub>38-218</sub> and His<sub>6</sub>-tagged PA<sub>488-735</sub>**

His<sub>6</sub>-tagged PA<sub>488-735</sub> (1,000 ng) was incubated on Ni-coated plates (Pierce, UK) for 1 hour, emptied and washed (3 x 200 µl PBS and 0.05% Tween-20). GST-tagged CMG2<sub>38-218</sub> (1,000 ng) with varying concentrations of inhibitors and 1 mM MgCl<sub>2</sub> were added and incubated for 1 hour; the plate was emptied and washed. Anti-GST (1 in 1000 dilution in blocking buffer; MA4-004, Neomarkers) was added and incubated for 1 hour and then emptied and washed (3 x 200 µl PBS and 0.05% Tween-20). Anti-mouse-HRP (1 in 6,000 dilution in blocking buffer; NA931, GE Healthcare) was added and incubated for 1 hour and then emptied and washed. 100 µl of 3,3',5,5' -tetramethylbenzide (TMB)-Ultra ELISA solution (34028, Fisher, UK) was incubated for 20 minutes and then quenched with 1 M H<sub>2</sub>SO<sub>4</sub>. Plate was analysed at 450 nm.

##### **4.6.5.2 Sandwich ELISA with His<sub>6</sub>-tagged CMG2<sub>38-218</sub> and PA<sub>596-735</sub>**

His<sub>6</sub>-tagged CMG2<sub>38-218</sub> (1,000 ng) was incubated on Ni-coated plates (Pierce, UK) for one hour, emptied and washed (3 x 200 µl PBS and 0.05% Tween-20). PA<sub>596-735</sub> (1,500 ng, gift from DSTL) with varying concentrations of inhibitors and 1 mM MgCl<sub>2</sub> were added and incubated for one hour, the plate was emptied and washed. Anti-PA<sub>596-735</sub> (1 in 500 dilution in blocking buffer; gift from DSTL) was added and incubated for 1 hour and then emptied and washed (3 x 200 µl PBS and 0.05% Tween-20). Anti-rabbit-HRP (1 in 5,000 dilution in blocking buffer; NA931, GE Healthcare) was added and incubated for 1 hour and then emptied and washed. 100 µl of TMB-Ultra ELISA solution (34028, Fisher, UK) was incubated for 20 minutes and then quenched with 1 M H<sub>2</sub>SO<sub>4</sub>. Plate was analysed at 450 nm.

##### **4.6.5.3 Isothermal titration calorimetry**

Experiments were carried out using a Micro-ITC calorimeter (Microcal) at 25°C. Multiple injections of inhibitor (either CLRFT, CLRPT or CMNHFPA) were made into the His<sub>6</sub>-tagged CMG2<sub>38-218</sub> or His<sub>6</sub>-tagged PA<sub>488-735</sub> cell under continuous stirring. Data was analysed using Microcal Origin software.

#### *4.6.5.4 Microscale thermophoresis*

MST experiments were performed on a Monolith NT.115 system (NanoTemper Technologies) using 100% LED and 40% IR-laser power. Laser on and off times were set at 30 seconds and 5 seconds, respectively. His<sub>6</sub>-tagged CMG2<sub>38-218</sub> and His<sub>6</sub>-tagged PA<sub>488-735</sub> were labelled with NT647 (NanoTemper Technologies) and used at a final concentration of 80 nM. A two-fold dilution series was prepared for the unlabelled peptide in MST-optimised buffer. Samples were filled into hydrophilic capillaries (NanoTemper Technologies) for measurement. The final concentrations of the peptides ranged from 5 mM to 152 nM.



## 5 Final conclusions and future work

The progression of disease is dependent on virulence mechanisms and the bacteria's ability to adapt to the host defence machinery. The development of antimicrobials over the last century has hindered the bacteria's ability to cause diseases. However, many antimicrobials have become ineffective in the treatment of several diseases (i.e. MRSA) due to an increase in selection pressure. Consequently, there has been a need for new antimicrobial targets that alleviate this selection pressure and, in turn, resistance. Virulence factors are regarded as such a target.

The aim of the project was to identify small molecule inhibitors of PPIs essential for virulence of pathogenic bacteria. *B. pseudomallei* and *B. anthracis* were selected due to their potential to be used as biological warfare agents. In addition, treatment processes are costly and lengthy. Currently, there is a lack of effective treatment available for *B. pseudomallei* and in the case of *B. anthracis*, the current treatments are unavailable to the general population. Holden *et al.* (2004) and Chan (2004) demonstrated that a key problem in the treatment of *B. pseudomallei* infection is its inherent antibiotic resistance even to third generation inhibitors.<sup>405,406</sup> Resistance is conferred through the presence of several resistance genes including aminoglycoside acetyltransferase, β-lactamases and multi-drug efflux systems.<sup>405</sup> Successful treatment requires a prolonged course of intravenous and oral antibiotics.<sup>407</sup> To date, there is no available vaccine against melioidosis.<sup>407</sup> On the other hand, there are therapeutics available for the treatment of anthrax including: vaccines based on PA that are expensive to mass-produce;<sup>55,408</sup> antibiotics that require long term compliance and have already shown clinically relevant resistance;<sup>304,377,378,409</sup> and raxibacumab, a monoclonal antibody that binds to PA inhibiting the interaction with the human receptor.<sup>380</sup> Raxibacumab can only be administered intravenously prior to the onset of symptoms, and used in conjunction with antibiotics.<sup>380</sup> All of the treatments for anthrax are expensive to mass-produce, susceptible to resistance and are consequently not available to the general public. Accordingly, the identification of bioavailable small molecule inhibitors for either pathogen would be advantageous.

For each pathogen a different virulence factor was targeted. The Psp response was chosen for *B. pseudomallei*; this has not been investigated as an antimicrobial target

previously. In contrast, the internalisation of anthrax toxins has previously been explored.<sup>381-383</sup> The combined use a bacterial RTHS and the SICLOPPS platform, however, have not yet been used as a screening tool. This *in vivo* screening approach has been employed to identify cyclic peptide inhibitors of the heterodimerisation of HIF-1 $\alpha$  and HIF-1 $\beta$ ,<sup>134</sup> the homodimerisation of CtBP<sup>133</sup> and ATIC,<sup>126</sup> the isolation of inhibitors of viral budding; targeting the interactions between human receptor Tsg101 and the HIV Gag protein;<sup>131</sup> and the bacterial  $\beta$ -sliding clamp in *S. auerius*.<sup>242</sup> Moreover, the SICLOPPS platform has been used in conjunction with other assays to identify inhibitors of bacterial targets.<sup>127,128</sup> This is exemplified by both the use of fluorescence-activated cell sorting with a fluorescent reporter to identify inhibitors for the bacterial ClpXP protease,<sup>127</sup> an enzyme essential for virulence in pathogenic bacteria like *Y. pestis*,<sup>410</sup> and its use in conjunction with a transposition-based genetic selection assay to screen against the protein-DNA interaction of adenine methyltransferase.<sup>128</sup>

Firstly, the screening tool was utilised to identify inhibitors of PspA oligomerisation and then the internalisation of anthrax toxins. The change in target was due to the lack of available information about the *B. pseudomallei* Psp response, the structure of PspA and assays limiting the ability to verify the activity of the cyclic peptides identified in the screen. Consequently, anthrax internalisation was targeted to demonstrate how a RTHS and the SICLOPPS platform could be utilised successfully when there is knowledge about the protein targets, their interaction and well-established assays to validate the peptides activity.<sup>292,294,295,311,322,330,340,342,349,367,411,412</sup>

These two targets have illustrated that targeting large oligomeric structures may not be feasible with the dimerisation-based RTHS and that linear peptides can be isolated from a screen, and still exhibit activity *in vitro*. Kritzer *et al.* (2009) reported how 15% of the sequences isolated against  $\alpha$ -synuclein activity contained frame shifts and 5% stop codons resulting in linear sequences; however, these were not tested further.<sup>129</sup> As described in Section 4.3.1 and 4.5, the occurrence of these linear sequences could be due to the toxicity of the system or the requirement for a linear sequence. Accordingly, their activity must be validated further. Nonetheless, if they retain their activity in future assays this would exemplify how versatile the

SICLOPPS system can be, in that it can self-regulate introducing stop codons, in order to isolate the most optimal conformation of inhibitor for a specific interaction.

Throughout the project there were limitations in the methods and targets chosen. With respect to the RTHS, as described above, it is a highly contrived system, with only dimers being investigated.<sup>106</sup> To allow accurate design of the system, knowledge of the interaction is essential, and post screen success is dependent on the breadth of information and available assays to verify the activity of the inhibitors. This was especially evident with the oligomerisation of PspA, where there is dispute about how the proteins interact, in what complex the cell membrane is restored<sup>212</sup> and how physiologically relevant the *in vitro* structures and *in vivo* localisations are.<sup>206,413</sup> In the future the anti-parallel PspA RTHS could be constructed to identify different inhibitors.

A common criticism of SICLOPPS is the functionality of the library. Although the library itself consists of 3.2 million cyclic peptides, it is restricted to the naturally-occurring amino acids. These exclude the functional groups: ketoamides, boronates, hydroxamates, Michael acceptors and N-methyl or β-amino acids. If the interaction between two proteins is well-studied by either crystal studies or NMR, for example, then a more rational approach could be taken, by expanding the genetic code to introduce non-natural amino acids by commandeering ribosomal translation machinery.<sup>414</sup>

The three active sequences identified for the disruption of CMG2 and PA PPI, have been investigated *in vitro*, with respect to their binding partner and their activity. Nevertheless, this should be verified further: for example FRET or SEC would demonstrate disruption; and ITC, SPR or a thermal-shift assay could confirm binding. Cell-based studies of the inhibitors will also be required, to confirm they can hinder toxin internalisation; this could be based on the LF-DTA and PA challenge studies of CHO-K1 cells or by monitoring the internalisation of fluorescently-labelled LF.<sup>292,325</sup> According to MST, CLRF(4-Cl)T had a two-fold improved binding affinity to CMG2. The improved activity of this analogue should be further confirmed by ELISA and other *in vitro* methods. Furthermore, FRET-based studies have shown that tannic acid and *cis*-platin, inhibitors of CMG2 and PA

PPI,<sup>289</sup> have poor bioavailability and are toxic, respectively. As the linear peptides identified here are mid-level molecules, that disobey Lipinski's rule of five for bioavailability,<sup>71</sup> it will be necessary to carry out pharmacological studies to ensure bioavailability. These could include a cell permeability assay and a serum stability assay.

The activity of the inhibitors in this study was low (i.e. IC<sub>50</sub> ranging from 50-150 μM). Therefore optimisation is required to compete with other inhibitors in the literature with nanomolar activity and improve bioavailability.<sup>391,394,396,415</sup> This could be achieved in a number of ways:

- N-methylation of the backbone to improve solubility and pharmacokinetic properties.<sup>416,417</sup> Currently, there is no information on the bioavailability and pharmacological relevance of CLRFT. As with other middle space drugs, modifications may be required in order to achieve oral absorption and cell permeability. For similar reasons, capping of the N- and C-terminus could be used to improve potency and improve solubility and pharmacokinetic properties.<sup>396</sup> However, there should be careful consideration into the choice of caps, as the properties of each capping group can greatly affect the activity of an inhibitor.<sup>132</sup>
- Alanine scanning of the peptide to identify the active residues and synthesise the active motif.<sup>122,132</sup>
- Polymerisation of the inhibitor, to improve local concentration during inhibition. Current inhibitors in the literature illustrate how polymerising the inhibitor by attaching the sequence to a β-cyclodextrin ring via a PEG-linker improves activity. These structures have been shown to improve the inhibitory effect of an LF to PA inhibitor 10,000-fold.<sup>387,391</sup>
- Enantio-substitution of L-amino acid with D-amino acid to circumvent degradation of peptides without cyclisation.<sup>418</sup> Substitution of the L-phenylalanine with D-phenylalanine did not affect the binding affinity of CLRFT to CMG2 by MST analysis, but it may improve the pharmacological properties of the molecule.

This study emphasises how successful verification of inhibitors isolated using a RTHS and SICLOPPS platform requires a good understanding of the PPI targeted, as demonstrated by targeting the unknown Psp response of *B. pseudomallei*. Here three potential peptide sequences for the inhibition of the PPI between PA and CMG2 have been identified. This PPI is essential for virulence of *B. anthracis* and illustrates how the RTHS and SICLOPPS library can be utilised to screen against PPI essential for virulence. One of the inhibitors has been shown to bind to the mammalian receptor, CMG2, opening it to be used as a prophylactic drug. Nevertheless, the activity of these inhibitors must be validated further in *in vivo* and *in vitro* assays, as well as, mouse studies, and the pharmacokinetic properties investigated. In addition, their activity and potency should be improved by altering the properties of the peptide. In conclusion, the work described here provides a foundation for the development of a more potent inhibitor of a PPI essential for virulence of *B. anthracis* that can be used in conjunction with antibiotics as a method of inhibiting anthrax disease progression.



## 6 Appendices

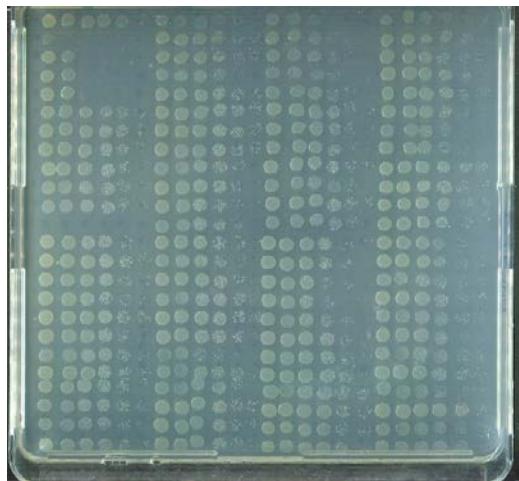
### 6.1 Appendix 1

**Table 47: PspF homologues in *B. pseudomallei* K96243.** Identified by S. Southern (DSTL) using BLAST searches with *E. coli* strain K-12 substrain MG1655 (GI: 90111246) and *Y. enterocolitica* subspecies enterocolitica 8081 (GI: 123442383).<sup>227</sup>

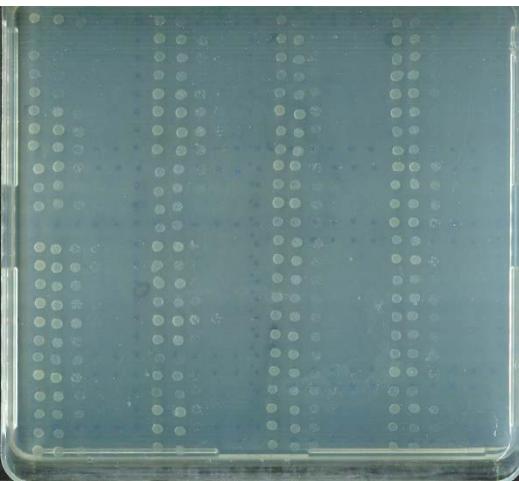
Protein	Length	Function	Identity to <i>E. coli</i>	Identity to <i>Y. enterocolitica</i>
BPSS2249	456	$\sigma^{54}$ -interacting response regulator	112/261 (43%)	115/262 (43%)
dctD (BPSL0427)	451	C4-dicarboxylate transport transcriptional regulatory protein	109/230 (47%)	110/246 (44%)
dctD (BPSS0063)	448	C4-dicarboxylate transport transcriptional response regulator	105/228 (46%)	130/330 (39%)
BPSL2475	486	$\sigma^{54}$ -dependent regulatory protein	106/254 (42%)	105/254 (41%)
glnG (BPSL2316)	511	nitrogen regulation protein NR	105/227 (46%)	109/227 (48%)
BPSL0609	462	fis family regulatory protein	116/312 (37%)	104/232 (44%)
BPSL2474	461	$\sigma^{54}$ -interacting regulatory protein	104/234 (44%)	100/230 (43%)
BPSL1562	467	putative transcriptional regulatory protein	105/236 (44%)	105/234 (44%)
BPSS2132	506	$\sigma^{54}$ -dependent transcriptional regulator	102/232 (44%)	100/232 (43%)
BPSS2250	322	$\sigma^{54}$ -interacting transcriptional regulator	122/322 (38%)	116/321 (36%)
prpR (BPSS0205)	694	$\sigma$ -interacting transcriptional regulator	112/249 (45%)	108/266 (40%)
BPSS0596	646	$\sigma^{54}$ -activated regulatory protein	97/217 (45%)	95/220 (43%)
BPSL1887	463	putative $\sigma^{54}$ -related transcriptional regulatory protein	114/325 (35%)	96/231 (41%)
BPSL2423	465	$\sigma^5$ -interacting response regulator protein	97/244 (40%)	114/342 (33%)

## **6.2 Appendix 2**

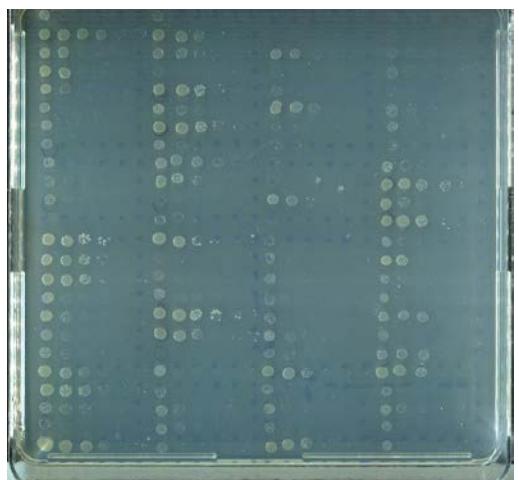
A



B

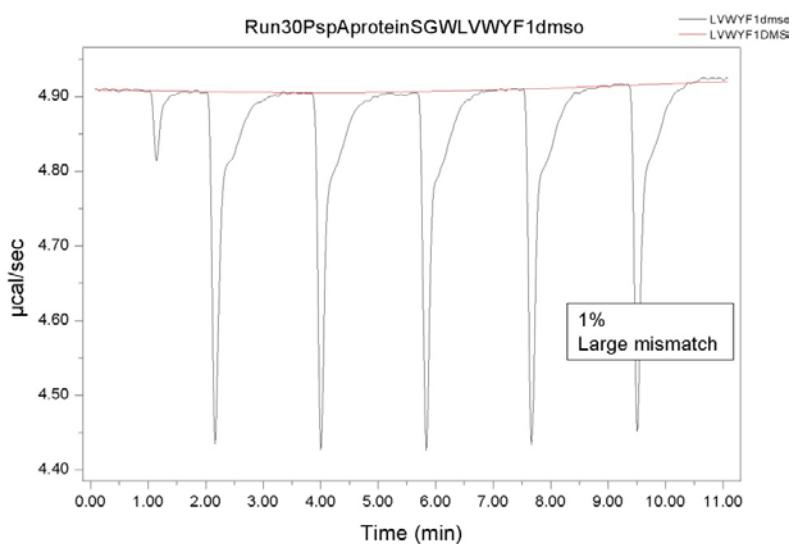
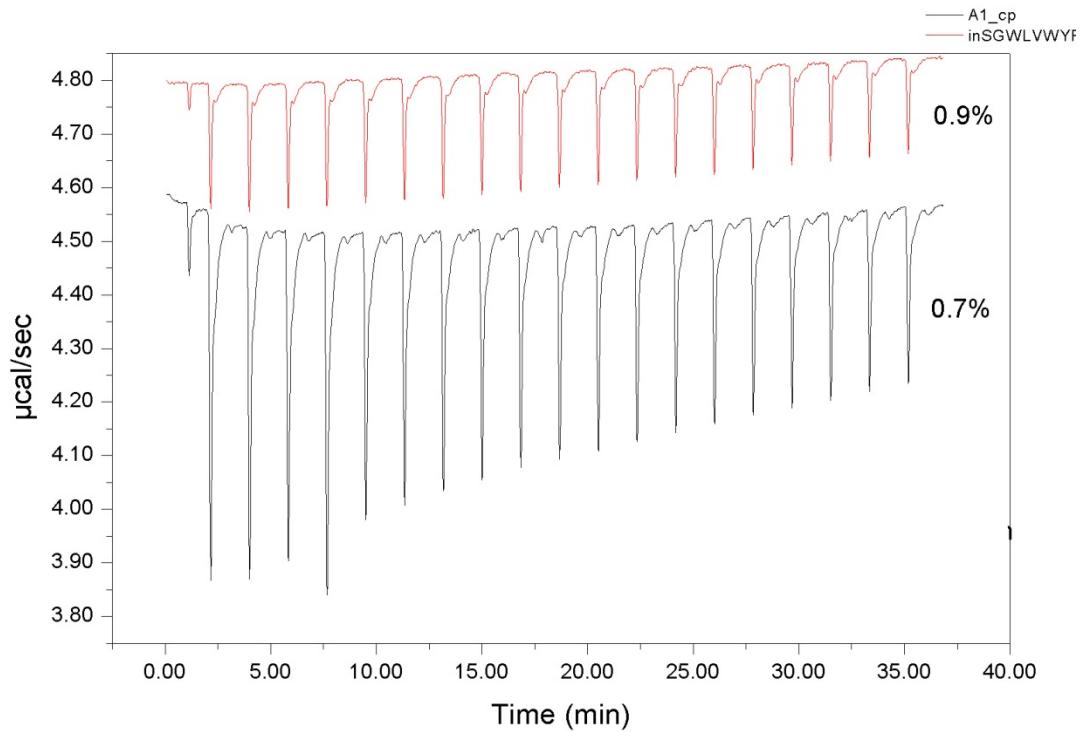


C



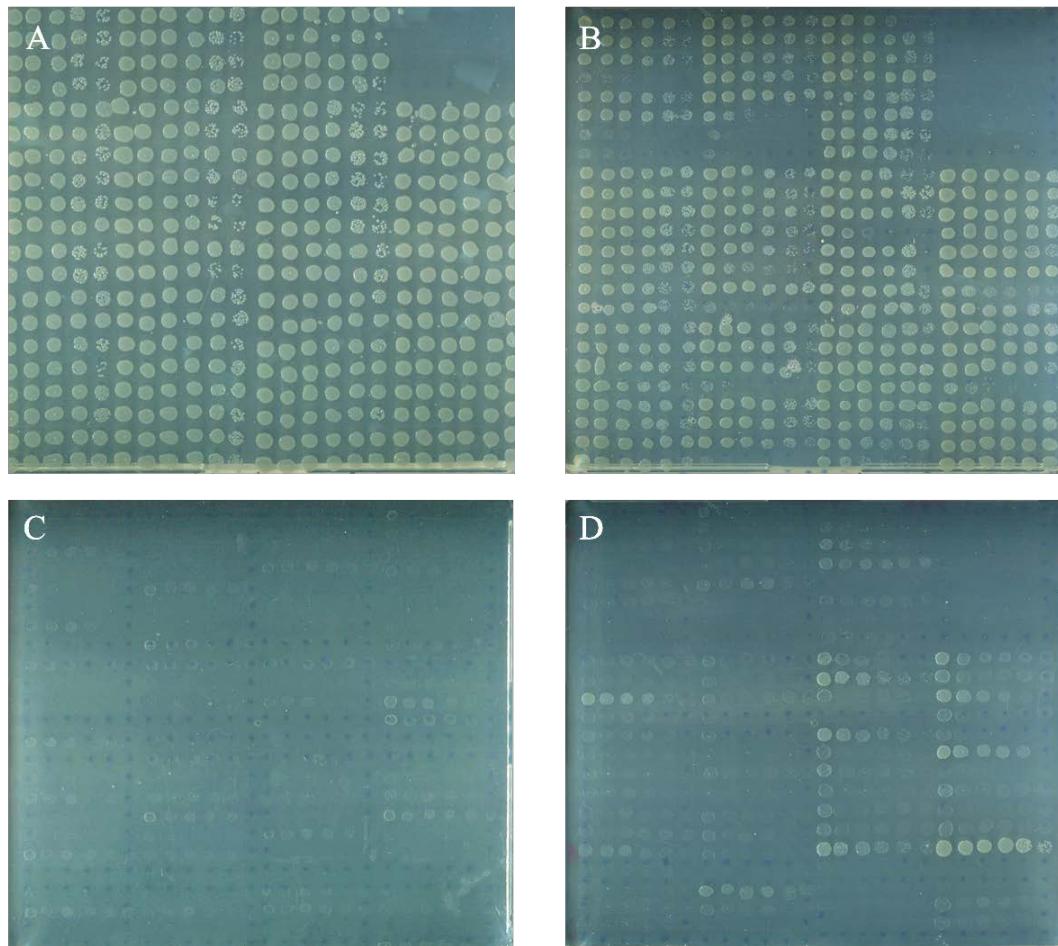
**Figure 111: Plates showing the screening of the SGWX<sub>5</sub> library.** (A) Minimal media plate where full growth was observed as there is no IPTG to induce the expression of 434-PspA recombinant protein. (B) Minimal media plate supplemented with IPTG, with an inhibition of growth observed as 434-PspA recombinant proteins are dimerising and hindering the transcription of the reporter genes. (C) Minimal media plate supplemented with IPTG and arabinose. Restoration of growth compared to IPTG plate was observed for any active cyclic peptides.

### 6.3 Appendix 3



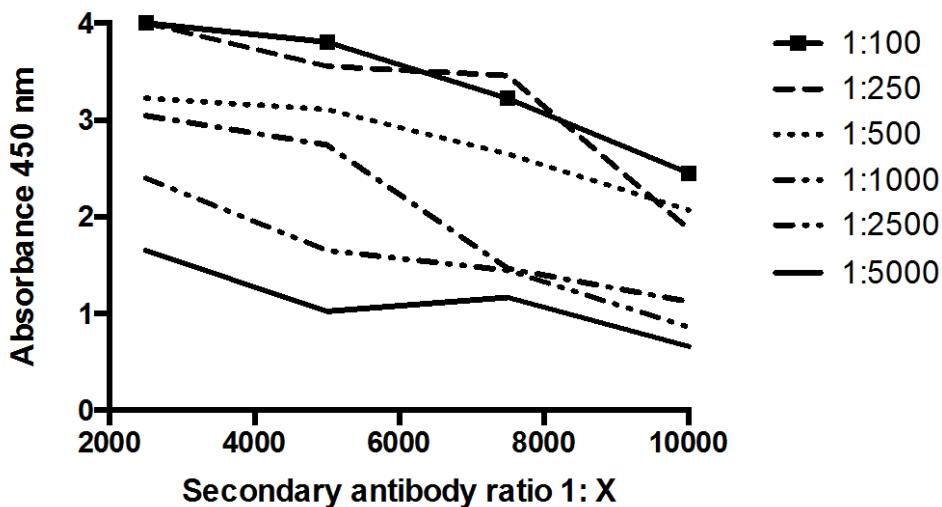
**Figure 112: Optimisation of DMSO concentration for PspA and inhibitor binding studies by ITC.** 0.7% and 1% show a large buffer mismatch; where 0.9% shows a smaller buffer mismatch suggesting that this is the optimal concentration of DMSO.

## **6.4 Appendix 4**



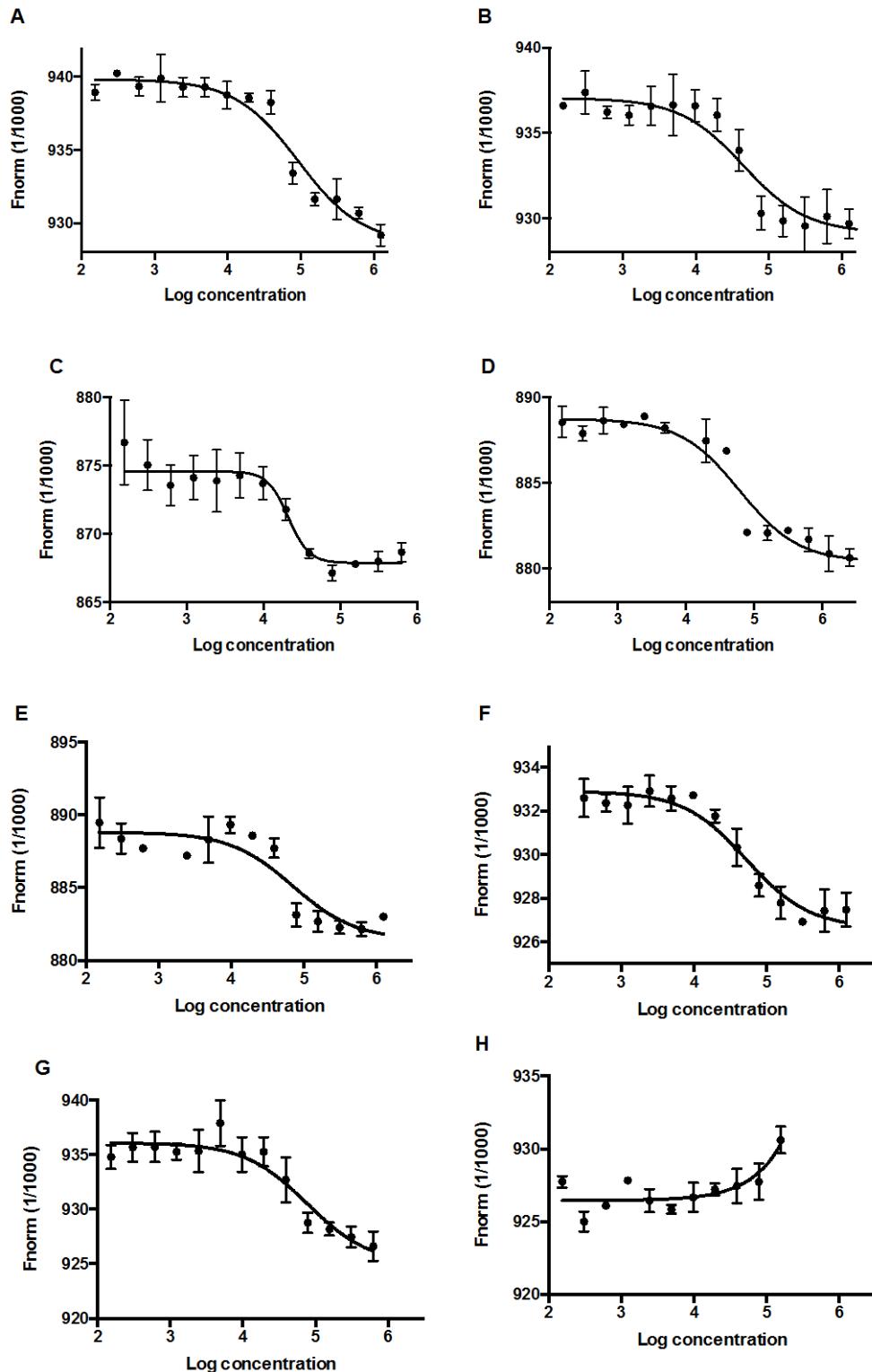
**Figure 113: Representative screen of the CX<sub>5</sub> library in the Anthrax RTHS.** (A) Minimal media plate. (B) Minimal media plate supplemented with arabinose. (C) Minimal media plate supplemented with IPTG. (D) Minimal media plate supplemented with arabinose and IPTG.

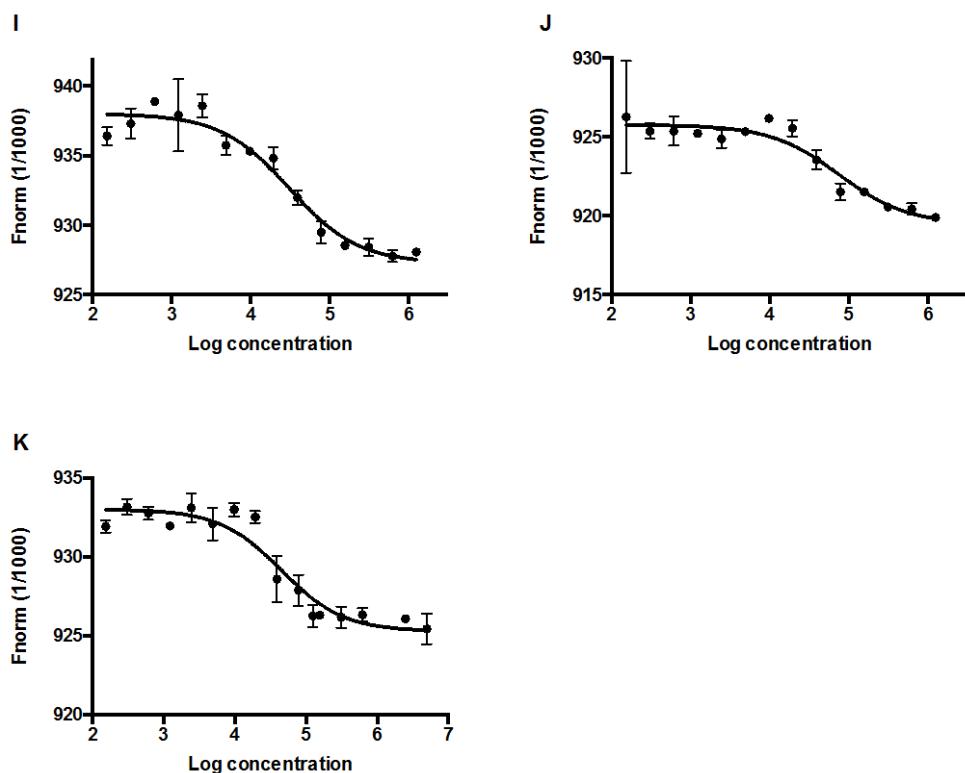
## 6.5 Appendix 5



**Figure 114: Optimisation of the concentration of the primary PA<sub>596-735</sub> antibody and secondary antibody, anti-rabbit-HRP. Each dilution is one well of a 96-well plate.**

## 6.6 Appendix 6





**Figure 115: MST of inhibitors against His<sub>6</sub>-tagged CMG2<sub>38-218</sub>.** The His<sub>6</sub>-tagged CMG2<sub>38-218</sub> was labelled covalently with NT-647 via the amine of the lysine residues. In all MST experiments the concentration of His<sub>6</sub>-tagged CMG2<sub>38-218</sub> remained constant (approximately 80 nM), while the concentration of the non-labelled inhibitor was varied between 5 mM to 152 nM. The assay was carried out in MST-optimised buffer in triplicate. (A) CLRYT, (B) CLRF(4-F)T, (C) CLRF(4-Cl)T, (D) CLRF(4-CN)T, (E) CLRY(3,5-Br<sub>2</sub>)T, (F) CLRF(4-NO<sub>2</sub>)T, (G) CLRY(3-NO<sub>2</sub>)T, (H) CLRF(4-Bz)T, (I) CLRdFT, (J) CLR(Phg)T, (K) CLR(hPhe)T.



## 7 References

- 1 Browning, C. H. & McKenzie, I. The Treatment of Syphilis by Salvarsan. *Br Med J* **2**, 654-655, (1911).
- 2 A. T. Fuller, L. F. I. C. Is p-aminobenzenesulphonamide the active agent in prontosil therapy? *Lancet* **229**, 194-198, (1937).
- 3 Otten, H. Domagk and the development of the sulphonamides. *J Antimicrob Chemother* **17**, 689-696, (1986).
- 4 Fleming, A. On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to their Use in the Isolation of *B. influenzae*. *Brit J Exp Pathol* **10**, 226-236 (1929).
- 5 Bartz, Q. R. Isolation and characterization of chloromycetin. *Journal Biol Chem* **172**, 445-450 (1948).
- 6 Duggar, B. M. Aureomycin; a product of the continuing search for new antibiotics. *Ann N Y Acad Sci* **51**, 177-181, (1948).
- 7 Lesher, G. Y., Froelich, E. J., Gruett, M. D., Bailey, J. H. & Brundage, R. P. 1,8-Naphthyridine Derivatives. A New Class of Chemotherapeutic Agents. *J Med Pharm Chem* **91**, 1063-1065, (1962).
- 8 Jones, R. N. *et al.* Piperacillin (T-1220), a new semisynthetic penicillin: in vitro antimicrobial activity comparison with carbenicillin, ticarcillin, ampicillin, cephalothin, cefamandole and cefoxitin. *J Antibiot (Tokyo)* **30**, 1107-1114, (1977).
- 9 Shah, P. P., Briedis, D. J., Robson, H. G. & Conterato, J. P. In vitro activity of piperacillin compared with that of carbenicillin, ticarcillin, ampicillin, cephalothin, and cefamandole against *Pseudomonas aeruginosa* and *Enterobacteriaceae*. *Antimicrob Agents Chemother* **15**, 346-350, (1979).
- 10 Price, P. B. The meaning of bacterostasis, bactericidal effect, and rate of disinfection. *Ann N Y Acad Sci* **53**, 76-90, (1950).
- 11 Klimek, J. W., Cavallito, C. J. & Bailey, J. H. Induced Resistance of *Staphylococcus aureus* to Various Antibiotics. *J Bacteriol* **55**, 139-145, (1948).
- 12 Jessen, O., Rosendal, K., Bulow, P., Faber, V. & Eriksen, K. R. Changing staphylococci and staphylococcal infections. A ten-year study of bacteria and cases of bacteremia. *New Engl J Med* **281**, 627-635, (1969).
- 13 Chambers, H. F. The changing epidemiology of *Staphylococcus aureus*? *Emerg Infect Dis* **7**, 178-182, (2001).
- 14 Abramson, M. A. & Sexton, D. J. Nosocomial methicillin-resistant and methicillin-susceptible *Staphylococcus aureus* primary bacteremia: at what costs? *Infect Control Hosp Epidemiol* **20**, 408-411, (1999).
- 15 Kleven, R. M. *et al.* Changes in the epidemiology of methicillin-resistant *Staphylococcus aureus* in intensive care units in US hospitals, 1992-2003. *Clin Infect Dis* **42**, 389-391, (2006).
- 16 Pray, L. Antibiotic Resistance, Mutation Rates and MRSA. *Nat Ed* **1** (2008).
- 17 Lewis, K. Platforms for antibiotic discovery. *Nat Rev Drug Discov* **12**, 371-387, (2013).
- 18 Tsiodras, S. *et al.* Linezolid resistance in a clinical isolate of *Staphylococcus aureus*. *Lancet* **358**, 207-208, (2001).

- 19 Sabol, K. *et al.* Emergence of daptomycin resistance in *Enterococcus faecium* during daptomycin therapy. *Antimicrob Agents Chemother* **49**, 1664-1665, (2005).
- 20 Hayden, M. K. *et al.* Development of Daptomycin resistance in vivo in methicillin-resistant *Staphylococcus aureus*. *J Clin Microbiol* **43**, 5285-5287, (2005).
- 21 Tupin, A., Gualtieri, M., Leonetti, J. P. & Brodolin, K. The transcription inhibitor lipiarmycin blocks DNA fitting into the RNA polymerase catalytic site. *Embo J* **29**, 2527-2537, (2010).
- 22 Rashid, M. U., Lozano, H. M., Weintraub, A. & Nord, C. E. In vitro activity of cadazolid against *Clostridium difficile* strains isolated from primary and recurrent infections in Stockholm, Sweden. *Anaerobe* **20**, 32-35, (2013).
- 23 Davies, J. & Davies, D. Origins and evolution of antibiotic resistance. *Microbiol Mol Biol Rev* **74**, 417-433, (2010).
- 24 Rao, S. N., Mookerjee, A. L., Obasanjo, O. O. & Chaisson, R. E. Errors in the treatment of tuberculosis in Baltimore. *Chest* **117**, 734-737, (2000).
- 25 Chan, Y. H. *et al.* Antibiotics nonadherence and knowledge in a community with the world's leading prevalence of antibiotics resistance: implications for public health intervention. *Am J Infect Control* **40**, 113-117, (2012).
- 26 Carey, B. & Cryan, B. Antibiotic misuse in the community--a contributor to resistance? *Ir Med J* **96**, 43-44, (2003).
- 27 Smith, D. L., Harris, A. D., Johnson, J. A., Silbergeld, E. K. & Morris, J. G., Jr. Animal antibiotic use has an early but important impact on the emergence of antibiotic resistance in human commensal bacteria. *P Natl Acad Sci USA* **99**, 6434-6439, (2002).
- 28 Witte, W. Medical consequences of antibiotic use in agriculture. *Science* **279**, 996-997, (1998).
- 29 McMurry, L., Petrucci, R. E., Jr. & Levy, S. B. Active efflux of tetracycline encoded by four genetically different tetracycline resistance determinants in *Escherichia coli*. *P Natl Acad Sci USA* **77**, 3974-3977, (1980).
- 30 Izaki, K. & Arima, K. Disappearance of Oxytetracycline Accumulation in the Cells of Multiple Drug-Resistant *Escherichia Coli*. *Nature* **200**, 384-385, (1963).
- 31 Joseph, W. B. Treatment of Staphylococcal infections with Penicillin by intermittent sterilisation. *Lancet* **244**, 497-500, (1944).
- 32 Keren, I., Kaldalu, N., Spoering, A., Wang, Y. & Lewis, K. Persister cells and tolerance to antimicrobials. *FEMS Microbiol. Lett.* **230**, 13-18, (2004).
- 33 Tuomanen, E., Cozens, R., Tosch, W., Zak, O. & Tomasz, A. The rate of killing of *Escherichia coli* by beta-lactam antibiotics is strictly proportional to the rate of bacterial growth. *J Gen Microbiol* **132**, 1297-1304, (1986).
- 34 Keren, I., Shah, D., Spoering, A., Kaldalu, N. & Lewis, K. Specialized persister cells and the mechanism of multidrug tolerance in *Escherichia coli*. *J Bacteriol* **186**, 8172-8180, (2004).
- 35 Kussell, E. & Leibler, S. Phenotypic diversity, population growth, and information in fluctuating environments. *Science* **309**, 2075-2078, (2005).
- 36 Clatworthy, A. E., Pierson, E. & Hung, D. T. Targeting virulence: a new paradigm for antimicrobial therapy. *Nat Chem Biol* **3**, 541-548, (2007).
- 37 Casadevall, A. & Pirofski, L. Host-pathogen interactions: the attributes of virulence. *J Infect dis* **184**, 337-344, (2001).

- 38 Sperandio, D. A. R. a. V. Anti-virulence strategies to combat bacteria-mediated disease. *Nature Reviews* **9**, 117-128, (2010).
- 39 Gevers, D. *et al.* The Human Microbiome Project: a community resource for the healthy human microbiome. *PLoS Biol* **10**, e1001377, (2012).
- 40 Bennedsen, M., Stuer-Lauridsen, B., Danielsen, M. & Johansen, E. Screening for antimicrobial resistance genes and virulence factors via genome sequencing. *Appl Environ Microbiol* **77**, 2785-2787, (2011).
- 41 Cheng, G., Zhang, Z., Chen, S., Bryers, J. D. & Jiang, S. Inhibition of bacterial adhesion and biofilm formation on zwitterionic surfaces. *Biomaterials* **28**, 4192-4199, (2007).
- 42 Cegelski, L. *et al.* Small-molecule inhibitors target Escherichia coli amyloid biogenesis and biofilm formation. *Nat Chem Biol* **5**, 913-919, (2009).
- 43 Pinkner, J. S. *et al.* Rationally designed small compounds inhibit pilus biogenesis in uropathogenic bacteria. *P Natl Acad Sci USA* **103**, 17897-17902, (2006).
- 44 Berg, V. *et al.* Design, synthesis and evaluation of peptidomimetics based on substituted bicyclic 2-pyridones-targeting virulence of uropathogenic E. coli. *Bioorg. Med. Chem.* **14**, 7563-7581, (2006).
- 45 Hung, D. T., Shakhnovich, E. A., Pierson, E. & Mekalanos, J. J. Small-molecule inhibitor of Vibrio cholerae virulence and intestinal colonization. *Science* **310**, 670-674, (2005).
- 46 Kauppi, A. M., Nordfelth, R., Uvell, H., Wolf-Watz, H. & Elofsson, M. Targeting bacterial virulence: inhibitors of type III secretion in Yersinia. *Chem Biol* **10**, 241-249, (2003).
- 47 Veenendaal, A. K., Sundin, C. & Blocker, A. J. Small-molecule type III secretion system inhibitors block assembly of the Shigella type III secretin. *J Bacteriol* **191**, 563-570, (2009).
- 48 Aiello, D. *et al.* Discovery and characterization of inhibitors of Pseudomonas aeruginosa type III secretion. *Antimicrob Agents Chemother* **54**, 1988-1999, (2010).
- 49 Kimura, K. *et al.* A small-molecule inhibitor of the bacterial type III secretion system protects against in vivo infection with Citrobacter rodentium. *J Antibiot (Tokyo)* **64**, 197-203, (2011).
- 50 Rasmussen, T. B. *et al.* Identity and effects of quorum-sensing inhibitors produced by Penicillium species. *Microbiology* **151**, 1325-1340, (2005).
- 51 Hentzer, M. *et al.* Attenuation of Pseudomonas aeruginosa virulence by quorum sensing inhibitors. *Embo J* **22**, 3803-3815, (2003).
- 52 Rasmussen, T. B. *et al.* Screening for quorum-sensing inhibitors (QSI) by use of a novel genetic system, the QSI selector. *J Bacteriol* **187**, 1799-1814, (2005).
- 53 McGillivray, S. M. *et al.* Pharmacological inhibition of the ClpXP protease increases bacterial susceptibility to host cathelicidin antimicrobial peptides and cell envelope-active antibiotics. *Antimicrob Agents Chemother* **56**, 1854-1861, (2012).
- 54 Svensson, A. *et al.* Design and evaluation of pilicides: potential novel antibacterial agents directed against uropathogenic Escherichia coli. *Chembiochem* **2**, 915-918, (2001).
- 55 Cirino, N. M. *et al.* Disruption of anthrax toxin binding with the use of human antibodies and competitive inhibitors. *Infect Immun* **67**, 2957-2963, (1999).

- 56 Mukhopadhyay, S. & Linstedt, A. D. Manganese blocks intracellular trafficking of Shiga toxin and protects against Shiga toxicosis. *Science* **335**, 332-335, (2012).
- 57 Wein, A. N. *et al.* Small molecule inhibitors of *Bacillus anthracis* protective antigen proteolytic activation and oligomerization. *J Med Chem* **55**, 7998-8006, (2012).
- 58 Nordfelth, R., Kauppi, A. M., Norberg, H. A., Wolf-Watz, H. & Elofsson, M. Small-molecule inhibitors specifically targeting type III secretion. *Infect Immun* **73**, 3104-3114, (2005).
- 59 Pettersson, J. *et al.* Modulation of virulence factor expression by pathogen target cell contact. *Science* **273**, 1231-1233, (1996).
- 60 Zhang, Y., Murtha, J., Roberts, M. A., Siegel, R. M. & Bliska, J. B. Type III secretion decreases bacterial and host survival following phagocytosis of *Yersinia pseudotuberculosis* by macrophages. *Infect Immun* **76**, 4299-4310, (2008).
- 61 Wolf, K. *et al.* Treatment of *Chlamydia trachomatis* with a small molecule inhibitor of the *Yersinia* type III secretion system disrupts progression of the chlamydial developmental cycle. *Mol Microbiol* **61**, 1543-1555, (2006).
- 62 Blocker, A. *et al.* Structure and composition of the *Shigella flexneri* "needle complex", a part of its type III secreton. *Mol Microbiol* **39**, 652-663, (2001).
- 63 Kauppi, A. M., Nordfelth, R., Hagglund, U., Wolf-Watz, H. & Elofsson, M. Salicylanilides are potent inhibitors of type III secretion in *Yersinia*. *Adv Exp Med Biol* **529**, 97-100, (2003).
- 64 Felise, H. B. *et al.* An inhibitor of gram-negative bacterial virulence protein secretion. *Cell Host Microbe* **4**, 325-336, (2008).
- 65 Armstrong, G. D. *et al.* A phase I study of chemically synthesized verotoxin (Shiga-like toxin) Pk-trisaccharide receptors attached to chromosorb for preventing hemolytic-uremic syndrome. *J Infect Dis* **171**, 1042-1045, (1995).
- 66 Critchley, I. A. *et al.* Spectrum of activity and mode of action of REP3123, a new antibiotic to treat *Clostridium difficile* infections. *J Antimicrob Chemother* **63**, 954-963, (2009).
- 67 Ochsner, U. A. *et al.* Inhibitory effect of REP3123 on toxin and spore formation in *Clostridium difficile*, and in vivo efficacy in a hamster gastrointestinal infection model. *J Antimicrob Chemother* **63**, 964-971, (2009).
- 68 Paschos, A. *et al.* An in vivo high-throughput screening approach targeting the type IV secretion system component VirB8 identified inhibitors of *Brucella abortus* 2308 proliferation. *Infect Immun* **79**, 1033-1043, (2011).
- 69 Gebus, C., Faudry, E., Bohn, Y. S., Elsen, S. & Attree, I. Oligomerization of PcrV and LcrV, protective antigens of *Pseudomonas aeruginosa* and *Yersinia pestis*. *J biol chem* **283**, 23940-23949, (2008).
- 70 Wells, J. A. & McClendon, C. L. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* **450**, 1001-1009, (2007).
- 71 Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliver Rev* **23**, 3-25, (1997).
- 72 Degterev, A. *et al.* Identification of small-molecule inhibitors of interaction between the BH3 domain and Bcl-xL. *Nat Cell Biol* **3**, 173-182, (2001).

- 73 Thangudu, R. R., Bryant, S. H., Panchenko, A. R. & Madej, T. Modulating protein-protein interactions with small molecules: the importance of binding hotspots. *J Mol Biol* **415**, 443-453, (2012).
- 74 Hong, Y. *et al.* The role of P53 and MDM2 polymorphisms in the risk of esophageal squamous cell carcinoma. *Cancer Res* **65**, 9582-9587, (2005).
- 75 Popowicz, G. M. *et al.* Structures of low molecular weight inhibitors bound to MDMX and MDM2 reveal new approaches for p53-MDMX/MDM2 antagonist drug discovery. *Cell Cycle* **9**, 1104-1111, (2010).
- 76 Ding, K. *et al.* Structure-based design of potent non-peptide MDM2 inhibitors. *J Am Chem Soc* **127**, 10130-10131, (2005).
- 77 Clanton, D. J. *et al.* Sulfonic acid dyes: inhibition of the human immunodeficiency virus and mechanism of action. *J Acquir Immune Defic Syndr* **5**, 771-781, (1992).
- 78 Dezube, B. J. *et al.* A fusion inhibitor (FP-21399) for the treatment of human immunodeficiency virus infection: a phase I study. *J Infect Dis* **182**, 607-610, (2000).
- 79 Fields, S. & Song, O. K. A novel genetic system to detect protein protein interactions. *Nature* **340**, 245-246, (1989).
- 80 Horswill, A. R., Savinov, S. N. & Benkovic, S. J. A systematic method for identifying small-molecule modulators of protein-protein interactions. *P Natl Acad Sci USA* **101**, 15591-15596, (2004).
- 81 Sato, T. *et al.* Interactions among members of the Bcl-2 protein family analyzed with a yeast two-hybrid system. *P Natl Acad Sci USA* **91**, 9238-9242, (1994).
- 82 Hu, J. C., O'Shea, E. K., Kim, P. S. & Sauer, R. T. Sequence requirements for coiled-coils: analysis with lambda repressor-GCN4 leucine zipper fusions. *Science* **250**, 1400-1403, (1990).
- 83 Dove, S. L., Joung, J. K. & Hochschild, A. Activation of prokaryotic transcription through arbitrary protein-protein contacts. *Nature* **386**, 627-630, (1997).
- 84 Aggarwal, A. K., Rodgers, D. W., Drottar, M., Ptashne, M. & Harrison, S. C. Recognition of a DNA Operator by the Repressor of Phage-434 - a View at High-Resolution. *Science* **242**, 899-907, (1988).
- 85 Bushman, F. D. & Ptashne, M. Activation of Transcription by the Bacteriophage-434 Repressor. *P Natl Acad Sci USA* **83**, 9353-9357 (1986).
- 86 Bushman, F. D. The bacteriophage-434 right operator roles of O(R)1, O(R)2 and O(R)3. *J Mol Biol* **230**, 28-40, (1993).
- 87 Anderson, J. E., Ptashne, M. & Harrison, S. C. Structure of the repressor-operator complex of bacteriophage 434. *Nature* **326**, 846-852, (1987).
- 88 Koudelka, G. B., Harbury, P., Harrison, S. C. & Ptashne, M. DNA twisting and the affinity of bacteriophage 434 operator for bacteriophage 434 repressor. *P Natl Acad Sci USA* **85**, 4633-4637, (1988).
- 89 Longo, F., Marchetti, M. A., Castagnoli, L., Battaglia, P. A. & Gigliani, F. A novel approach to protein-protein interaction: complex formation between the p53 tumor suppressor and the HIV Tat proteins. *Biochem Biophys Res Commun* **206**, 326-334, (1995).
- 90 Di Lallo, G., Ghelardini, P. & Paolozzi, L. Two-hybrid assay: construction of an Escherichia coli system to quantify homodimerization ability in vivo. *Microbiology* **145** ( Pt 6), 1485-1490, (1999).

- 91 Di Lallo, G., Anderluzzi, D., Ghelardini, P. & Paolozzi, L. FtsZ dimerization in vivo. *Mol Microbiol* **32**, 265-274, (1999).
- 92 Hollis, M., Valenzuela, D., Pioli, D., Wharton, R. & Ptashne, M. A Repressor Heterodimer Binds to a Chimeric Operator. *P Natl Acad Sci USA* **85**, 5834-5838, (1988).
- 93 Di Lallo, G., Castagnoli, L., Ghelardini, P. & Paolozzi, L. A two-hybrid system based on chimeric operator recognition for studying protein homo/heterodimerization in Escherichia coli. *Microbiol-Sgm* **147**, 1651-1656, (2001).
- 94 Di Lallo, G., Fagioli, M., Barionovi, D., Ghelardini, P. & Paolozzi, L. Use of a two-hybrid assay to study the assembly of a complex multicomponent protein machinery: bacterial septosome differentiation. *Microbiology* **149**, 3353-3359, (2003).
- 95 Stella, S., Spurio, R., Falconi, M., Pon, C. L. & Gualerzi, C. O. Nature and mechanism of the in vivo oligomerization of nucleoid protein H-NS. *Embo J* **24**, 2896-2905, (2005).
- 96 Joung, J. K., Ramm, E. I. & Pabo, C. O. A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions. *P Natl Acad Sci USA* **97**, 7382-7387, (2000).
- 97 Struhl, K. & Davis, R. W. Production of a functional eukaryotic enzyme in Escherichia coli: cloning and expression of the yeast structural gene for imidazole-glycerolphosphate dehydratase (his3). *P Natl Acad Sci USA* **74**, 5255-5259, (1977).
- 98 Alifano, P. *et al.* Histidine biosynthetic pathway and genes: Structure, regulation, and evolution. *Microbiol Rev* **60**, 44-69 (1996).
- 99 Buelow, P. The Onpg Test in Diagnostic Bacteriology. Methodological Investigations. *Acta Pathol Microbiol Scand* **60**, 376-386, (1964).
- 100 Scott, C. P., Abel-Santos, E., Wall, M., Wahnon, D. C. & Benkovic, S. J. Production of cyclic peptides and proteins in vivo. *P Natl Acad Sci USA* **96**, 13638-13643, (1999).
- 101 Gao, Y., Liehr, S. & Cooperman, B. S. Affinity-driven selection of tripeptide inhibitors of ribonucleotide reductase. *Bioorg Med Chem Lett* **12**, 513-515, (2002).
- 102 Furuta, E. *et al.* Targeting protein homodimerization: A novel drug discovery system. *Febs Lett* **579**, 2065-2070, (2005).
- 103 Kane, P. M. *et al.* Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase. *Science* **250**, 651-657, (1990).
- 104 Hirata, R. *et al.* Molecular structure of a gene, VMA1, encoding the catalytic subunit of H(+)-translocating adenosine triphosphatase from vacuolar membranes of *Saccharomyces cerevisiae*. *J Biol Chem* **265**, 6726-6733, (1990).
- 105 Shih, C. K., Wagner, R., Feinstein, S., Kanik-Ennulat, C. & Neff, N. A dominant trifluoperazine resistance gene from *Saccharomyces cerevisiae* has homology with FOF1 ATP synthase and confers calcium-sensitive growth. *Mol Cell Biol* **8**, 3094-3103, (1988).
- 106 Gimble, F. S. & Thorner, J. Homing of a DNA endonuclease gene by meiotic gene conversion in *Saccharomyces cerevisiae*. *Nature* **357**, 301-306, (1992).

- 107 Chong, S. & Xu, M. Q. Protein splicing of the *Saccharomyces cerevisiae* VMA intein without the endonuclease motifs. *J Biol Chem* **272**, 15587-15590, (1997).
- 108 Telenti, A. *et al.* The *Mycobacterium xenopi* GyrA protein splicing element: characterization of a minimal intein. *J Bacteriol* **179**, 6378-6382, (1997).
- 109 Wu, H., Hu, Z. & Liu, X. Q. Protein trans-splicing by a split intein encoded in a split DnaE gene of *Synechocystis* sp. PCC6803. *P Natl Acad Sci USA* **95**, 9226-9231, (1998).
- 110 Evans, T. C., Jr. *et al.* Protein trans-splicing and cyclization by a naturally split intein from the dnaE gene of *Synechocystis* species PCC6803. *J Biol Chem* **275**, 9091-9094, (2000).
- 111 Gorbalya, A. E. Non-canonical inteins. *Nucleic Acids Res.* **26**, 1741-1748, (1998).
- 112 Xu, M. Q. & Perler, F. B. The mechanism of protein splicing and its modulation by mutation. *Embo J* **15**, 5146-5153, (1996).
- 113 Xu, M. Q. *et al.* Protein splicing: an analysis of the branched intermediate and its resolution by succinimide formation. *Embo J* **13**, 5517-5522 (1994).
- 114 Xu, M. Q., Southworth, M. W., Mersha, F. B., Hornstra, L. J. & Perler, F. B. In vitro protein splicing of purified precursor and the identification of a branched intermediate. *Cell* **75**, 1371-1377, (1993).
- 115 Cooper, A. A., Chen, Y. J., Lindorfer, M. A. & Stevens, T. H. Protein splicing of the yeast TFP1 intervening protein sequence: a model for self-excision. *Embo J* **12**, 2575-2583, (1993).
- 116 Xu, M. Q. & Evans, T. C., Jr. Purification of recombinant proteins from *E. coli* by engineered inteins. *Methods Mol Biol* **205**, 43-68, (2003).
- 117 Tavassoli, A. & Benkovic, S. J. Split-intein mediated circular ligation used in the synthesis of cyclic peptide libraries in *E. coli*. *Nat Protoc* **2**, 1126-1133, (2007).
- 118 Hancock, R. E. & Chapple, D. S. Peptide antibiotics. *Antimicrob Agents Chemother* **43**, 1317-1323, (1999).
- 119 Hancock, R. E. & Sahl, H. G. Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies. *Nat Biotechnol* **24**, 1551-1557, (2006).
- 120 Gause, G. F., Brazhnikova, M. G. Gramicidin S and its use in the Treatment of Infected Wounds. *Nature* **154**, 703, (1944).
- 121 Gilon, C., Halle, D., Chorev, M., Selinger, Z. & Byk, G. Backbone Cyclization - a New Method for Conferring Conformational Constraint on Peptides. *Biopolymers* **31**, 745-750, (1991).
- 122 Spurr, I. B. *et al.* Targeting tumour proliferation with a small-molecule inhibitor of AICAR transformylase homodimerization. *Chembiochem* **13**, 1628-1634, (2012).
- 123 Perler, F. B. Protein splicing mechanisms and applications. *IUBMB Life* **57**, 469-476, (2005).
- 124 Iwai, H. & Pluckthun, A. Circular beta-lactamase: stability enhancement by cyclizing the backbone. *Febs Lett* **459**, 166-172, (1999).
- 125 Scott, C. P., Abel-Santos, E., Jones, A. D. & Benkovic, S. J. Structural requirements for the biosynthesis of backbone cyclic peptide libraries. *Chem Biol* **8**, 801-815, (2001).

- 126 Tavassoli, A. & Benkovic, S. J. Genetically selected cyclic-peptide inhibitors  
of AICAR transformylase homodimerization. *Angew Chem Int Ed Engl* **44**,  
2760-2763, (2005).
- 127 Cheng, L. *et al.* Discovery of antibacterial cyclic peptides that inhibit the  
ClpXP protease. *Protein Sci* **16**, 1535-1542, (2007).
- 128 Naumann, T. A., Tavassoli, A. & Benkovic, S. J. Genetic selection of cyclic  
peptide Dam methyltransferase inhibitors. *Chembiochem* **9**, 194-197, (2008).
- 129 Kritzer, J. A. *et al.* Rapid selection of cyclic peptides that reduce alpha-  
synuclein toxicity in yeast and animal models. *Nat Chem Biol* **5**, 655-663,  
(2009).
- 130 Kinsella, T. M. *et al.* Retrovirally delivered random cyclic Peptide libraries  
yield inhibitors of interleukin-4 signaling in human B cells. *J Biol Chem* **277**,  
37512-37518, (2002).
- 131 Tavassoli, A. *et al.* Inhibition of HIV budding by a genetically selected cyclic  
peptide targeting the Gag-TSG101 interaction. *ACS Chem Biol* **3**, 757-764,  
(2008).
- 132 Datta, S., Bucks, M. E., Koley, D., Lim, P. X. & Savinov, S. N. Functional  
profiling of p53-binding sites in Hdm2 and Hdmx using a genetic selection  
system. *Bioorg Med Chem* **18**, 6099-6108, (2010).
- 133 Birts, C. N. *et al.* A cyclic peptide inhibitor of C-terminal binding protein  
dimerization links metabolism with mitotic fidelity in breast cancer cells.  
*Chem Sci* **4**, 3046-3057, (2013).
- 134 Miranda, E. *et al.* A Cyclic Peptide Inhibitor of HIF-1 Heterodimerization  
That Inhibits Hypoxia Signaling in Cancer Cells. *J Am Chem Soc* **135**,  
10418-10425, (2013).
- 135 Lee, K. *et al.* Acriflavine inhibits HIF-1 dimerization, tumor growth, and  
vascularization. *P Natl Acad Sci USA* **106**, 17910-17915, (2009).
- 136 Rowley, G., Spector, M., Kormanec, J. & Roberts, M. Pushing the envelope:  
extracytoplasmic stress responses in bacterial pathogens. *Nat Rev Microbiol*  
**4**, 383-394, (2006).
- 137 Erickson, J. W. & Gross, C. A. Identification of the sigma E subunit of  
Escherichia coli RNA polymerase: a second alternate sigma factor involved  
in high-temperature gene expression. *Genes Dev* **3**, 1462-1471, (1989).
- 138 Wang, Q. P. & Kaguni, J. M. A novel sigma factor is involved in expression  
of the rpoH gene of Escherichia coli. *J Bacteriol* **171**, 4248-4253, (1989).
- 139 Raivio, T. L. & Silhavy, T. J. Transduction of envelope stress in Escherichia  
coli by the Cpx two-component system. *J Bacteriol* **179**, 7724-7733, (1997).
- 140 Raffa, R. G. & Raivio, T. L. A third envelope stress signal transduction  
pathway in Escherichia coli. *Mol Microbiol* **45**, 1599-1611, (2002).
- 141 Nagasawa, S., Ishige, K. & Mizuno, T. Novel members of the two-  
component signal transduction genes in Escherichia coli. *J Biol Chem* **114**,  
350-357, (1993).
- 142 Brissette, J. L., Russel, M., Weiner, L. & Model, P. Phage Shock Protein, a  
Stress Protein of Escherichia-Coli. *P Natl Acad Sci USA* **87**, 862-866, (1990).
- 143 Brissette, J. L., Weiner, L., Ripmaster, T. L. & Model, P. Characterization  
and sequence of the Escherichia coli stress-induced psp operon. *J Mol Biol*  
**220**, 35-48, (1991).
- 144 Darwin, A. J. The phage-shock-protein response. *Mol Microbiol* **57**, 621-628,  
(2005).

- 145 Merrick, M., Jones, D. H. & Thomas, C. M. Location of the rpoN gene on the physical map of Escherichia coli. *J Bacteriol* **175**, 1548-1549, (1993).
- 146 Merrick, M. J. In a class of its own--the RNA polymerase sigma factor sigma 54 (sigma N). *Mol Microbiol* **10**, 903-909, (1993).
- 147 Bury-Mone, S. *et al.* Global analysis of extracytoplasmic stress signaling in Escherichia coli. *PLoS Genet* **5**, e10000651, (2009).
- 148 Weiner, L. & Model, P. Role of an Escherichia-Coli Stress-Response Operon in Stationary-Phase Survival. *P Natl Acad Sci USA* **91**, 2191-2195, (1994).
- 149 Becker, L. A., Bang, I. S., Crouch, M. L. & Fang, F. C. Compensatory role of PspA, a member of the phage shock protein operon, in rpoE mutant *Salmonella enterica* serovar Typhimurium. *Mol Microbiol* **56**, 1004-1016, (2005).
- 150 Maxson, M. E. & Darwin, A. J. Identification of inducers of the *Yersinia enterocolitica* phage shock protein system and comparison to the regulation of the RpoE and Cpx extracytoplasmic stress responses. *J Bacteriol* **186**, 4199-4208, (2004).
- 151 Darwin, A. J. & Miller, V. L. Identification of *Yersinia enterocolitica* genes affecting survival in an animal host using signature-tagged transposon mutagenesis. *Mol Microbiol* **32**, 51-62, (1999).
- 152 DeLisa, M. P., Lee, P., Palmer, T. & Georgiou, G. Phage shock protein PspA of *Escherichia coli* relieves saturation of protein export via the Tat pathway. *J Bacteriol* **186**, 366-373, (2004).
- 153 Beloin, C. *et al.* Global impact of mature biofilm lifestyle on *Escherichia coli* K-12 gene expression. *Mol Microbiol* **51**, 659-674, (2004).
- 154 Linderoth, N. A., Simon, M. N. & Russel, M. The filamentous phage pIV multimer visualized by scanning transmission electron microscopy. *Science* **278**, 1635-1638, (1997).
- 155 Brissette, J. L. & Russel, M. Secretion and membrane integration of a filamentous phage-encoded morphogenetic protein. *J Mol Biol* **211**, 565-580, (1990).
- 156 Russel, M. & Kazmierczak, B. Analysis of the structure and subcellular location of filamentous phage pIV. *J Bacteriol* **175**, 3998-4007, (1993).
- 157 Lloyd, L. J. *et al.* Identification of a new member of the phage shock protein response in *Escherichia coli*, the phage shock protein g (PspG). *J Biol Chem* **279**, 55707-55714, (2004).
- 158 Elderkin, S., Bordes, P., Jones, S., Rappas, M. & Buck, M. Molecular determinants for PspA-mediated repression of the AAA transcriptional activator PspF. *J Bacteriol* **187**, 3238-3248, (2005).
- 159 Huvet, M. *et al.* The evolution of the phage shock protein response system: interplay between protein function, genomic organization, and system function. *Mol Biol Evol* **28**, 1141-1155, (2011).
- 160 Darwin, A. J. & Miller, V. L. The psp locus of *Yersinia enterocolitica* is required for virulence and for growth in vitro when the Ysc type III secretion system is produced. *Mol Microbiol* **39**, 429-444, (2001).
- 161 Vrancken, K., Van Mellaert, L. & Anne, J. Characterization of the *Streptomyces lividans* PspA response. *J Bacteriol* **190**, 3475-3481, (2008).
- 162 Bidle, K. A., Kirkland, P. A., Nannen, J. L. & Maupin-Furlow, J. A. Proteomic analysis of *Haloferax volcanii* reveals salinity-mediated regulation of the stress response protein PspA. *Microbiology* **154**, 1436-1443, (2008).

- 163 Kroll, D. *et al.* VIPP1, a nuclear gene of *Arabidopsis thaliana* essential for thylakoid membrane formation. *P Natl Acad Sci USA* **98**, 4238-4242, (2001).
- 164 Westphal, S., Heins, L., Soll, J. & Vothknecht, U. C. Vipp1 deletion mutant of *Synechocystis*: a connection between bacterial phage shock and thylakoid biogenesis? *P Natl Acad Sci USA* **98**, 4243-4248, (2001).
- 165 Weiner, L., Brissette, J. L. & Model, P. Stress-induced expression of the *Escherichia coli* phage shock protein operon is dependent on sigma 54 and modulated by positive and negative feedback mechanisms. *Genes Dev* **5**, 1912-1923, (1991).
- 166 Burrows, P. C., Severinov, K., Ishihama, A., Buck, M. & Wigneshweraraj, S. R. Mapping sigma 54-RNA polymerase interactions at the -24 consensus promoter element. *J Biol Chem* **278**, 29728-29743, (2003).
- 167 Barrios, H., Valderrama, B. & Morett, E. Compilation and analysis of sigma(54)-dependent promoter sequences. *Nucleic Acids Res.* **27**, 4305-4313, (1999).
- 168 Joly, N., Rappas, M., Wigneshweraraj, S. R., Zhang, X. & Buck, M. Coupling nucleotide hydrolysis to transcription activation performance in a bacterial enhancer binding protein. *Mol Microbiol* **66**, 583-595, (2007).
- 169 Rappas, M., Bose, D. & Zhang, X. Bacterial enhancer-binding proteins: unlocking sigma54-dependent gene transcription. *Curr Opin Struct Biol* **17**, 110-116, (2007).
- 170 Morett, E. & Buck, M. In vivo studies on the interaction of RNA polymerase-sigma 54 with the *Klebsiella pneumoniae* and *Rhizobium meliloti* nifH promoters. The role of NifA in the formation of an open promoter complex. *J Mol Biol* **210**, 65-77, (1989).
- 171 Patel, S. & Latterich, M. The AAA team: related ATPases with diverse functions. *Trends Cell Biol* **8**, 65-71, (1998).
- 172 Bordes, P. *et al.* The ATP hydrolyzing transcription activator phage shock protein F of *Escherichia coli*: identifying a surface that binds sigma 54. *Proc Natl Acad Sci U S A* **100**, 2278-2283, (2003).
- 173 Rappas, M., Schumacher, J., Niwa, H., Buck, M. & Zhang, X. Structural basis of the nucleotide driven conformational changes in the AAA+ domain of transcription activator PspF. *J Mol Biol* **357**, 481-492, (2006).
- 174 Joly, N. *et al.* Managing membrane stress: the phage shock protein (Psp) response, from molecular mechanisms to physiology. *FEMS Microbiol Rev* **34**, 797-827, (2010).
- 175 Hutcheson, S. W., Bretz, J., Sussan, T., Jin, S. & Pak, K. Enhancer-binding proteins HrpR and HrpS interact to regulate hrp-encoded type III protein secretion in *Pseudomonas syringae* strains. *J Bacteriol* **183**, 5589-5598, (2001).
- 176 Krzywda, S. *et al.* The crystal structure of the AAA domain of the ATP-dependent protease FtsH of *Escherichia coli* at 1.5 Å resolution. *Structure* **10**, 1073-1083, (2002).
- 177 Joly, N., Burrows, P. C., Engl, C., Jovanovic, G. & Buck, M. A lower-order oligomer form of phage shock protein A (PspA) stably associates with the hexameric AAA(+) transcription activator protein PspF for negative regulation. *J Mol Biol* **394**, 764-775, (2009).
- 178 Jovanovic, G., Rakonjac, J. & Model, P. In vivo and in vitro activities of the *Escherichia coli* sigma54 transcription activator, PspF, and its DNA-binding mutant, PspFDeltaHTH. *J Mol Biol* **285**, 469-483, (1999).

- 179 Jovanovic, G. & Model, P. PspF and IHF bind co-operatively in the psp promoter-regulatory region of *Escherichia coli*. *Mol Microbiol* **25**, 473-481, (1997).
- 180 Weiner, L., Brissette, J. L., Ramani, N. & Model, P. Analysis of the proteins and cis-acting elements regulating the stress-induced phage shock protein operon. *Nucleic Acids Res.* **23**, 2030-2036, (1995).
- 181 Kobayashi, H., Yamamoto, M. & Aono, R. Appearance of a stress-response protein, phage-shock protein A, in *Escherichia coli* exposed to hydrophobic organic solvents. *Microbiology-(UK)* **144**, 353-359, (1998).
- 182 Bergler, H., Abraham, D., Aschauer, H. & Turnowsky, F. Inhibition of lipid biosynthesis induces the expression of the PspA gene. *Microbiology-(UK)* **140**, 1937-1944, (1994).
- 183 Csonka, L. N. Physiological and genetic responses of bacteria to osmotic stress. *Microbiol Rev* **53**, 121-147, (1989).
- 184 Palade, G. Intracellular aspects of the process of protein synthesis. *Science* **189**, 867, (1975).
- 185 Coburn, B., Sekirov, I. & Finlay, B. B. Type III secretion systems and disease. *Clin Microbiol Rev* **20**, 535-549, (2007).
- 186 Galan, J. E. & Collmer, A. Type III secretion machines: bacterial devices for protein delivery into host cells. *Science* **284**, 1322-1328, (1999).
- 187 Fath, M. J. & Kolter, R. ABC transporters: bacterial exporters. *Microbiol Rev* **57**, 995-1017, (1993).
- 188 Berks, B. C. A common export pathway for proteins binding complex redox cofactors? *Mol Microbiol* **22**, 393-404, (1996).
- 189 Van Gijsegem, F., Genin, S. & Boucher, C. Conservation of secretion pathways for pathogenicity determinants of plant and animal bacteria. *Trends Microbiol* **1**, 175-180, (1993).
- 190 Seo, J., Savitzky, D. C., Ford, E. & Darwin, A. J. Global analysis of tolerance to secretin-induced stress in *Yersinia enterocolitica* suggests that the phage-shock-protein system may be a remarkably self-contained stress response. *Mol Microbiol* **65**, 714-727, (2007).
- 191 Kleerebezem, M. & Tommassen, J. Expression of the pspA gene stimulates efficient protein export in *Escherichia coli*. *Mol Microbiol* **7**, 947-956, (1993).
- 192 Kleerebezem, M., Crielaard, W. & Tommassen, J. Involvement of stress protein PspA (phage shock protein A) of *Escherichia coli* in maintenance of the protonmotive force under stress conditions. *Embo J* **15**, 162-171, (1996).
- 193 Samuelson, J. C. *et al.* Function of YidC for the insertion of M13 procoat protein in *Escherichia coli*: translocation of mutants that show differences in their membrane potential dependence and Sec requirement. *J Biol Chem* **276**, 34847-34852, (2001).
- 194 Jones, S. E., Lloyd, L. J., Tan, K. K. & Buck, M. Secretion defects that activate the phage shock response of *Escherichia coli*. *J Bacteriol* **185**, 6707-6711, (2003).
- 195 Serek, J. *et al.* *Escherichia coli* YidC is a membrane insertase for Sec-independent proteins. *Embo J* **23**, 294-301, (2004).
- 196 Hardie, K. R., Lory, S. & Pugsley, A. P. Insertion of an outer membrane protein in *Escherichia coli* requires a chaperone-like protein. *Embo J* **15**, 978-988, (1996).

- 197 Possot, O., d'Enfert, C., Reyss, I. & Pugsley, A. P. Pullulanase secretion in Escherichia coli K-12 requires a cytoplasmic protein and a putative polytopic cytoplasmic membrane protein. *Mol Microbiol* **6**, 95-105, (1992).
- 198 Russel, M., Linderoth, N. A. & Sali, A. Filamentous phage assembly: variation on a protein export theme. *Gene* **192**, 23-32, (1997).
- 199 Model, P., Jovanovic, G. & Dworkin, J. The Escherichia coli phage-shock-protein (psp) operon. *Mol Microbiol* **24**, 255-261, (1997).
- 200 Kobayashi, R., Suzuki, T. & Yoshida, M. Escherichia coli phage-shock protein A (PspA) binds to membrane phospholipids and repairs proton leakage of the damaged membranes. *Mol Microbiol* **66**, 100-109, (2007).
- 201 Kinoshita, N., Unemoto, T. & Kobayashi, H. Proton motive force is not obligatory for growth of Escherichia coli. *J Bacteriol* **160**, 1074-1077, (1984).
- 202 Wang, P., Kuhn, A. & Dalbey, R. E. Global change of gene expression and cell physiology in YidC-depleted Escherichia coli. *J Bacteriol* **192**, 2193-2209, (2010).
- 203 Jovanovic, G., Lloyd, L. J., Stumpf, M. P., Mayhew, A. J. & Buck, M. Induction and function of the phage shock protein extracytoplasmic stress response in Escherichia coli. *J Biol Chem* **281**, 21147-21161, (2006).
- 204 Jovanovic, G., Engl, C. & Buck, M. Physical, functional and conditional interactions between ArcAB and phage shock proteins upon secretin-induced stress in Escherichia coli. *Mol Microbiol* **74**, 16-28, (2009).
- 205 Engl, C. *et al.* In vivo localizations of membrane stress controllers PspA and PspG in Escherichia coli. *Mol Microbiol* **73**, 382-396, (2009).
- 206 Lenn, T. *et al.* Measuring the stoichiometry of functional PspA complexes in living bacterial cells by single molecule photobleaching. *Chem Commun (Camb)* **47**, 400-402, (2011).
- 207 Aseeva, E. *et al.* Complex formation of Vipp1 depends on its alpha-helical PspA-like domain. *J Biol Chem* **279**, 35535-35541, (2004).
- 208 Li, H. M., Kaneko, Y. & Keegstra, K. Molecular cloning of a chloroplastic protein associated with both the envelope and thylakoid membranes. *Plant Mol Biol* **25**, 619-632, (1994).
- 209 Hankamer, B. D., Elderkin, S. L., Buck, M. & Nield, J. Organization of the AAA(+) adaptor protein PspA is an oligomeric ring. *J Biol Chem* **279**, 8862-8866, (2004).
- 210 Fuhrmann, E. *et al.* The vesicle-inducing protein 1 from Synechocystis sp. PCC 6803 organizes into diverse higher-ordered ring structures. *Mol Biol Cell* **20**, 4620-4628, (2009).
- 211 Liu, C. *et al.* The chloroplast HSP70B-CDJ2-CGE1 chaperones catalyse assembly and disassembly of VIPP1 oligomers in Chlamydomonas. *Plant J* **50**, 265-277, (2007).
- 212 Standar, K. *et al.* PspA can form large scaffolds in Escherichia coli. *Febs Lett* **582**, 3585-3589, (2008).
- 213 Yamaguchi, S., Gueguen, E., Horstman, N. K. & Darwin, A. J. Membrane association of PspA depends on activation of the phage-shock-protein response in Yersinia enterocolitica. *Mol Microbiol* **78**, 429-443, (2010).
- 214 Yamaguchi, S., Reid, D. A., Rothenberg, E. & Darwin, A. J. Changes in Psp protein binding partners, localization and behaviour upon activation of the Yersinia enterocolitica phage shock protein response. *Mol Microbiol* **87**, 656-671, (2013).

- 215 Gueguen, E., Savitzky, D. C. & Darwin, A. J. Analysis of the *Yersinia enterocolitica* PspBC proteins defines functional domains, essential amino acids and new roles within the phage-shock-protein response. *Mol Microbiol* **74**, 619-633, (2009).
- 216 Maxson, M. E. & Darwin, A. J. PspB and PspC of *Yersinia enterocolitica* are dual function proteins: regulators and effectors of the phage-shock-protein response. *Mol Microbiol* **59**, 1610-1623, (2006).
- 217 Adams, H., Teertstra, W., Koster, M. & Tommassen, J. PspE (phage-shock protein E) of *Escherichia coli* is a rhodanese. *Febs Lett* **518**, 173-176, (2002).
- 218 Wallrodt, I. *et al.* The putative thiosulfate sulfurtransferases PspE and GlpE contribute to virulence of *Salmonella Typhimurium* in the mouse model of systemic disease. *Plos One* **8**, e70829, (2013).
- 219 Cheng, H., Donahue, J. L., Battle, S. E., Ray, W. K. & Larson, T. J. Biochemical and Genetic Characterization of PspE and GlpE, Two Single-domain Sulfurtransferases of *Escherichia coli*. *Open Microbiol J* **2**, 18-28, (2008).
- 220 Adams, H., Teertstra, W., Demmers, J., Boesten, R. & Tommassen, J. Interactions between phage-shock proteins in *Escherichia coli*. *J Bacteriol* **185**, 1174-1180, (2003).
- 221 Green, R. C. & Darwin, A. J. PspG, a new member of the *Yersinia enterocolitica* phage shock protein regulon. *J Bacteriol* **186**, 4910-4920, (2004).
- 222 Whitmore, A., Krishnaswami, C., . An account of the discovery of a hitherto undescribed infective disease occurring among the population of Rangoon. *Ind med Gaz* **47**, 262-267, (1912).
- 223 Currie, B. J. *et al.* Endemic melioidosis in tropical northern Australia: a 10-year prospective study and review of the literature. *Clin infect dis* **31**, 981-986, (2000).
- 224 Chaowagul, W. *et al.* Melioidosis - a Major Cause of Community-Acquired Septicemia in Northeastern Thailand. *J Infect Dis* **159**, 890-899 (1989).
- 225 Suputtamongkol, Y. *et al.* The epidemiology of melioidosis in Ubon Ratchatani, northeast Thailand. *Int J Epidemiol* **23**, 1082-1090, (1994).
- 226 White, N. J. Melioidosis. *Lancet* **361**, 1715-1722, (2003).
- 227 Southern, S. *Evaluation of Two Novel Antimicrobial Targets in Burkholderia pseudomallei* Chemistry thesis, (2014).
- 228 Novick, R. P. Plasmid incompatibility. *Microbiol Rev* **51**, 381-395, (1987).
- 229 Platt, R., Drescher, C., Park, S. K. & Phillips, G. J. Genetic system for reversible integration of DNA constructs and lacZ gene fusions into the *Escherichia coli* chromosome. *Plasmid* **43**, 12-23, (2000).
- 230 Wanner, L. Z. K. Z. a. B. L. *Methods in molecular biology: Recombinant gene expression reviews and protocols*. . 2nd edn, Vol. 267 123-133, (Human press, 2004).
- 231 Haldimann, A. & Wanner, B. L. Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. *J Bacteriol* **183**, 6384-6393, (2001).
- 232 Zhou, L. K., S.-K.; Avramova, L.; Datsenko, K. A; Wanner, B. L. *Use of conditional-replication, integration, and modular CRIM plasmids to make single-copy lacZ fusions*. 65-89, (Birkhauser Verlag, 2003).

- 233 Rossignol, M., Moulin, L. & Boccard, F. Phage HK022-based integrative vectors for the insertion of genes in the chromosome of multiply marked *Escherichia coli* strains. *FEMS Microbiol Lett* **213**, 45-49, (2002).
- 234 Nagaraja, R. & Weisberg, R. A. Specificity determinants in the attachment sites of bacteriophages HK022 and lambda. *J Bacteriol* **172**, 6540-6550, (1990).
- 235 Khlebnikov, A., Datsenko, K. A., Skaug, T., Wanner, B. L. & Keasling, J. D. Homogeneous expression of the P(BAD) promoter in *Escherichia coli* by constitutive expression of the low-affinity high-capacity AraE transporter. *Microbiology* **147**, 3241-3247, (2001).
- 236 Baba, T. *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**, 2006 0008, (2006).
- 237 Moat, A. G., Watkins-Foster, J. & Spector, M. P. in *Microbial physiology* Ch. 5, 194-236, (Wiley-Blackwell, 2002).
- 238 Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *P Natl Acad Sci USA* **97**, 6640-6645, (2000).
- 239 Miller, J. F. *Experiments in Molecular Genetics: Assays of B-galactosidase* 352-355 (CSH Laboratory Press, 1972).
- 240 Mockli, N. & Auerbach, D. Quantitative beta-galactosidase assay suitable for high-throughput applications in the yeast two-hybrid system. *Biotechniques* **36**, 872-876, (2004).
- 241 Horswill, A. R. & Benkovic, S. J. Cyclic peptides, a chemical genetics tool for biologists. *Cell Cycle* **4**, 552-555, (2005).
- 242 Kjelstrup, S., Hansen, P. M., Thomsen, L. E., Hansen, P. R. & Lobner-Olesen, A. Cyclic peptide inhibitors of the beta-sliding clamp in *Staphylococcus aureus*. *Plos One* **8**, e72273, (2013).
- 243 Pérez-Pérez, J. & Gutiérrez, J. An arabinose-inducible expression vector, pAR3, compatible with ColE1-derived plasmids. *Gene* **158**, 141-142, (1995).
- 244 Dower, W. J., Miller, J. F. & Ragsdale, C. W. High-Efficiency Transformation of *Escherichia-Coli* by High-Voltage Electroporation. *Nucleic Acids Res.* **16**, 6127-6145, (1988).
- 245 Chuang, S. E., Chen, A. L. & Chao, C. C. Growth of *Escherichia-Coli* at Low-Temperature Dramatically Increases the Transformation Frequency by Electroporation. *Nucleic Acids Res.* **23**, 1641-1641, (1995).
- 246 Naumann, T. A., Savinov, S. N. & Benkovic, S. J. Engineering an affinity tag for genetically encoded cyclic peptides. *Biotechnol Bioeng* **92**, 820-830, (2005).
- 247 Merrifield, R. Solid Phase Synthesis. I. The Synthesis of a Tetrapeptide. *J Am Chem Soc* **85**, 2149-2154, (1963).
- 248 Chang, C. D. & Meienhofer, J. Solid-phase peptide synthesis using mild base cleavage of N alpha-fluorenylmethoxycarbonyl amino acids, exemplified by a synthesis of dihydrosomatostatin. *Int J Pept Protein Res* **11**, 246-249, (1978).
- 249 Fields, G. B. & Noble, R. L. Solid phase peptide synthesis utilizing 9-fluorenylmethoxycarbonyl amino acids. *Int J Pept Protein Res* **35**, 161-214, (1990).
- 250 Jursic, B. A Simple Preparation of Amides from Acids and Amines by Heating of Their Mixture. *Synth Comm* **23**, 2761-2771, (1993).

- 251 Konig, W. & Geiger, R. [A new method for synthesis of peptides: activation of the carboxyl group with dicyclohexylcarbodiimide using 1-hydroxybenzotriazoles as additives]. *Chem Ber* **103**, 788-798, (1970).
- 252 Carpino, L. A. 1-Hydroxy-7-Azabenzotriazole - an Efficient Peptide Coupling Additive. *J Am Chem Soc* **115**, 4397-4398, (1993).
- 253 Kaiser, E., Colescott, R. L., Bossinger, C. D. & Cook, P. I. Color test for detection of free terminal amino groups in the solid-phase synthesis of peptides. *Anal Biochem* **34**, 595-598, (1970).
- 254 King, D. S., Fields, C. G. & Fields, G. B. A cleavage method which minimizes side reactions following Fmoc solid phase peptide synthesis. *Int J Pept Protein Res* **36**, 255-266, (1990).
- 255 Franzen, H., Grehn, L. & Ragnarsson, U. Synthesis, Properties, and Use of Nin-Boc-Tryptophan Derivatives. *J Chem Soc Chem Comm*, 1699-1700, (1984).
- 256 Condron, M. M., Monien, B. H. & Bitan, G. Synthesis and Purification of Highly Hydrophobic Peptides Derived from the C-Terminus of Amyloid beta-Protein. *Open Biotechnol J* **2**, 87-93, (2008).
- 257 Cuatrecasas, P., Wilchek, M. & Anfinsen, C. B. Selective enzyme purification by affinity chromatography. *P Natl Acad Sci USA* **61**, 636-643, (1968).
- 258 Porath, J., Carlsson, J., Olsson, I. & Belfrage, G. Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature* **258**, 598-599, (1975).
- 259 Elderkin, S., Jones, S., Schumacher, J., Studholme, D. & Buck, M. Mechanism of action of the Escherichia coli phage shock protein PspA in repression of the AAA family transcription factor PspF. *J Mol Biol* **320**, 23-37, (2002).
- 260 Lathe, G. H. & Ruthven, C. R. The separation of substances on the basis of their molecular weights, using columns of starch and water. *Biochem J* **60**, xxxiv, (1955).
- 261 Andrews, P. Estimation of the molecular weights of proteins by Sephadex gel-filtration. *Biochem J* **91**, 222-233, (1964).
- 262 Andrews, P. The gel-filtration behaviour of proteins related to their molecular weights over a wide range. *Biochem J* **96**, 595-606, (1965).
- 263 Baranov, V. I., Morozov, I., Ortlepp, S. A. & Spirin, A. S. Gene expression in a cell-free system on the preparative scale. *Gene* **84**, 463-466, (1989).
- 264 Studier, F. W. & Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **189**, 113-130, (1986).
- 265 Maina, C. V. *et al.* An Escherichia coli vector to express and purify foreign proteins by fusion to and separation from maltose-binding protein. *Gene* **74**, 365-373, (1988).
- 266 di Guan, C., Li, P., Riggs, P. D. & Inouye, H. Vectors that facilitate the expression and purification of foreign peptides in Escherichia coli by fusion to maltose-binding protein. *Gene* **67**, 21-30, (1988).
- 267 Doyle, M. L. Characterization of binding interactions by isothermal titration calorimetry. *Curr Opin Biotechnol* **8**, 31-35, (1997).
- 268 Wiseman, T., Williston, S., Brandts, J. F. & Lin, L. N. Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal Biochem* **179**, 131-137, (1989).

- 269 Jerabek-Willemsen, M., Wienken, C. J., Braun, D., Baaske, P. & Duhr, S. Molecular interaction studies using microscale thermophoresis. *Assay Drug Dev Technol* **9**, 342-353, (2011).
- 270 Zhao, S., Zhang, Q., Chen, Z. H. & Zhong, J. C. The factors dictating the codon usage variation among the genes in the genome of *Burkholderia pseudomallei*. *World J Microb Biot* **24**, 1585-1592, (2008).
- 271 Grote, A. (2004-2008) JAVA codon adaption tool URL: <http://www.Jcat.de> [07 July 2014]
- 272 Wand, M. E., Muller, C. M., Titball, R. W. & Michell, S. L. Macrophage and *Galleria mellonella* infection models reflect the virulence of naturally occurring isolates of *B. pseudomallei*, *B. thailandensis* and *B. oklahomensis*. *BMC Microbiol* **11**, 11, (2011).
- 273 Hoffmann, F. *et al.* Heat-inactivation of plasmid-encoded CI857 repressor induces gene expression from Ind(-) lambda prophage in recombinant *Escherichia coli*. *FEMS Microbiol Lett* **177**, 327-334, (1999).
- 274 Miranda, E., Forafonov, F. & Tavassoli, A. Deciphering interactions used by the influenza virus NS1 protein to silence the host antiviral sensor protein RIG-I using a bacterial reverse two-hybrid system. *Mol Biosyst* **7**, 1042-1045, (2011).
- 275 Wu, X. H., Upadhyaya, P., Villalona-Calero, M. A., Briesewitz, R. & Pei, D. H. Inhibition of Ras-effector interactions by cyclic peptides. *Medchemcomm* **4**, 378-382, (2013).
- 276 Bogan, A. A. & Thorn, K. S. Anatomy of hot spots in protein interfaces. *J Mol Biol* **280**, 1-9, (1998).
- 277 Kozakov, D. *et al.* Structural conservation of druggable hot spots in protein-protein interfaces. *P Natl Acad Sci USA* **108**, 13528-13533, (2011).
- 278 Tilley, J. W. *et al.* Identification of a small molecule inhibitor of the IL-2/IL-2R alpha receptor interaction which binds to IL-2. *J Am Chem Soc* **119**, 7589-7590, (1997).
- 279 Yu, F. *et al.* Approaches for identification of HIV-1 entry inhibitors targeting gp41 pocket. *Viruses* **5**, 127-149, (2013).
- 280 Moerke, N. J. *et al.* Small-molecule inhibition of the interaction between the translation initiation factors eIF4E and eIF4G. *Cell* **128**, 257-267, (2007).
- 281 Schwarz, K., Borngraber, S., Anton, M. & Kuhn, H. Probing the substrate alignment at the active site of 15-lipoxygenases by targeted substrate modification and site-directed mutagenesis. Evidence for an inverse substrate orientation. *Biochemistry-US* **37**, 15327-15335, (1998).
- 282 Turnbull, W. B. & Daranas, A. H. On the value of c: can low affinity systems be studied by isothermal titration calorimetry? *J Am Chem Soc* **125**, 14859-14866, (2003).
- 283 Corin, K. *et al.* Designer Lipid-Like Peptides: A Class of Detergents for Studying Functional Olfactory Receptors Using Commercial Cell-Free Systems. *Plos One* **6**, e25067, (2011).
- 284 Aseeva, E. *et al.* Vipp1 is required for basic thylakoid membrane formation but not for the assembly of thylakoid protein complexes. *Plant Physiol Biochem* **45**, 119-128, (2007).
- 285 Zhang, L. G., Kato, Y., Otters, S., Vothknecht, U. C. & Sakamoto, W. Essential Role of VIPP1 in Chloroplast Envelope Maintenance in *Arabidopsis*. *Plant Cell* **24**, 3695-3707, (2012).

- 286 Savva, C. G. *et al.* The holin of bacteriophage lambda forms rings with large diameter. *Mol Microbiol* **69**, 784-793, (2008).
- 287 Wolf, D. *et al.* In-depth profiling of the LiaR response of *Bacillus subtilis*. *J Bacteriol* **192**, 4680-4693, (2010).
- 288 Wigelsworth, D. J. *et al.* Binding stoichiometry and kinetics of the interaction of a human anthrax toxin receptor, CMG2, with protective antigen. *J Biol Chem* **279**, 23349-23356, (2004).
- 289 Rogers, M. S. *et al.* A FRET-based high throughput screening assay to identify inhibitors of anthrax protective antigen binding to capillary morphogenesis gene 2 protein. *Plos One* **7**, e39911, (2012).
- 290 Dawson, R. M. The CMG2 ELISA for evaluating inhibitors of the binding of anthrax toxin protective antigen to its receptor. *J Pharmacol Toxicol Methods* **59**, 50-55, (2009).
- 291 Cryan, L. M. *et al.* Identification of small molecules that inhibit the interaction of TEM8 with anthrax protective antigen using a FRET assay. *J Biomol Screen* **18**, 714-725, (2013).
- 292 Bradley, K. A., Mogridge, J., Mourez, M., Collier, R. J. & Young, J. A. Identification of the cellular receptor for anthrax toxin. *Nature* **414**, 225-229, (2001).
- 293 Lacy, D. B., Wigelsworth, D. J., Scobie, H. M., Young, J. A. & Collier, R. J. Crystal structure of the von Willebrand factor A domain of human capillary morphogenesis protein 2: an anthrax toxin receptor. *P Natl Acad Sci USA* **101**, 6367-6372, (2004).
- 294 Lacy, D. B. & Collier, R. J. Structure and function of anthrax toxin. *Curr Top Microbiol Immunol* **271**, 61-85, (2002).
- 295 Santelli, E., Bankston, L. A., Leppla, S. H. & Liddington, R. C. Crystal structure of a complex between anthrax toxin and its host cell receptor. *Nature* **430**, 905-908, (2004).
- 296 Kelly, C. P., Pothoulakis, C. & LaMont, J. T. Clostridium difficile colitis. *N Engl J Med* **330**, 257-262, (1994).
- 297 Collier, R. J. Diphtheria toxin: mode of action and structure. *Bacteriol Rev* **39**, 54-85, (1975).
- 298 Smith, H., Keppie, J. & Stanley, J. L. The chemical basis of the virulence of *Bacillus anthracis*. V. The specific toxin produced by *B. Anthracis* in vivo. *Brit J Exp Path* **36**, 460-472, (1955).
- 299 Smith, H., Tempest, D. W., Stanley, J. L., Harris-Smith, P. W. & Gallop, R. C. The chemical basis of the virulence of *Bacillus anthracis*. VII. Two components of the anthrax toxin: their relationship to known immunising aggressins. *Brit J Exp Path* **37**, 263-271, (1956).
- 300 Carter, K. C. Koch Postulates in Relation to the Work of Henle, Jacob and Klebs, Edwin. *Med Hist* **29**, 353-374, (1985).
- 301 L, P. Roux De l'attenuation des virus et de leur retoire a la virulence. . *Comptes Rendus des Seances de L'Academie des Sciences* **92**, 430-435, (1881).
- 302 Charatan, F. Anthrax blamed as two postal workers die in United States. *Br Med J* **323**, 951-951, (2001).
- 303 Keim, P. *et al.* Molecular investigation of the Aum Shinrikyo anthrax release in Kameido, Japan. *J Clin Microbiol* **39**, 4566-4567, (2001).
- 304 Inglesby, T. V. *et al.* Anthrax as a biological weapon, 2002: updated recommendations for management. *Jama* **287**, 2236-2252, (2002).

- 305 Ross, J. M. The pathogenesis of anthrax following the administration of spores by the respiratory route. *J Pathol* **73**, 485-494, (1957).
- 306 Okinaka, R. *et al.* Sequence, assembly and analysis of pXO1 and pXO2. *J Appl Microbiol* **87**, 261-262, (1999).
- 307 Robertson, D. L. & Leppla, S. H. Molecular cloning and expression in *Escherichia coli* of the lethal factor gene of *Bacillus anthracis*. *Gene* **44**, 71-78, (1986).
- 308 Robertson, D. L., Tippett, M. T. & Leppla, S. H. Nucleotide sequence of the *Bacillus anthracis* edema factor gene (*cya*): a calmodulin-dependent adenylate cyclase. *Gene* **73**, 363-371, (1988).
- 309 Vodkin, M. H. & Leppla, S. H. Cloning of the protective antigen gene of *Bacillus anthracis*. *Cell* **34**, 693-697, (1983).
- 310 Molloy, S. S., Bresnahan, P. A., Leppla, S. H., Klimpel, K. R. & Thomas, G. Human furin is a calcium-dependent serine endoprotease that recognizes the sequence Arg-X-X-Arg and efficiently cleaves anthrax toxin protective antigen. *J Biol Chem* **267**, 16396-16402, (1992).
- 311 Lacy, D. B., Wigelsworth, D. J., Melnyk, R. A., Harrison, S. C. & Collier, R. J. Structure of heptameric protective antigen bound to an anthrax toxin receptor: a role for receptor in pH-dependent pore formation. *P Natl Acad Sci USA* **101**, 13147-13151, (2004).
- 312 Cunningham, K., Lacy, D. B., Mogridge, J. & Collier, R. J. Mapping the lethal factor and edema factor binding sites on oligomeric anthrax protective antigen. *P Natl Acad Sci USA* **99**, 7049-7053, (2002).
- 313 Lacy, D. B. *et al.* A model of anthrax toxin lethal factor bound to protective antigen. *P Natl Acad Sci USA* **102**, 16409-16414, (2005).
- 314 Abrami, L., Liu, S. H., Cosson, P., Leppla, S. H. & van der Goot, F. G. Anthrax toxin triggers endocytosis of its receptor via a lipid raft-mediated clathrin-dependent process. *J Cell Biol* **160**, 321-328, (2003).
- 315 Zhang, S., Finkelstein, A. & Collier, R. J. Evidence that translocation of anthrax toxin's lethal factor is initiated by entry of its N terminus into the protective antigen channel. *P Natl Acad Sci USA* **101**, 16756-16761, (2004).
- 316 Mogridge, J., Cunningham, K., Lacy, D. B., Mourez, M. & Collier, R. J. The lethal and edema factors of anthrax toxin bind only to oligomeric forms of the protective antigen. *P Natl Acad Sci USA* **99**, 7045-7048, (2002).
- 317 Rainey, G. J. *et al.* Receptor-specific requirements for anthrax toxin delivery into cells. *P Natl Acad Sci USA* **102**, 13278-13283, (2005).
- 318 Chvyrkova, I., Zhang, X. C. & Terzyan, S. Lethal factor of anthrax toxin binds monomeric form of protective antigen. *Biochem Biophys Res Commun* **360**, 690-695, (2007).
- 319 Milne, J. C. & Collier, R. J. pH-dependent permeabilization of the plasma membrane of mammalian cells by anthrax protective antigen. *Mol Microbiol* **10**, 647-653, (1993).
- 320 Abrami, L., Bischofberger, M., Kunz, B., Groux, R. & van der Goot, F. G. Endocytosis of the anthrax toxin is mediated by clathrin, actin and unconventional adaptors. *Plos Pathog* **6**, e1000792, (2010).
- 321 Abrami, L., Lindsay, M., Parton, R. G., Leppla, S. H. & van der Goot, F. G. Membrane insertion of anthrax protective antigen and cytoplasmic delivery of lethal factor occur at different stages of the endocytic pathway. *J Cell Biol* **166**, 645-651, (2004).

- 322 Petosa, C., Collier, R. J., Klimpel, K. R., Leppla, S. H. & Liddington, R. C. Crystal structure of the anthrax toxin protective antigen. *Nature* **385**, 833-838, (1997).
- 323 Nassi, S., Collier, R. J. & Finkelstein, A. PA63 channel of anthrax toxin: an extended beta-barrel. *Biochemistry-US* **41**, 1445-1450, (2002).
- 324 Benson, E. L., Huynh, P. D., Finkelstein, A. & Collier, R. J. Identification of residues lining the anthrax protective antigen channel. *Biochemistry-US* **37**, 3941-3948, (1998).
- 325 Mourez, M. *et al.* Mapping dominant-negative mutations of anthrax protective antigen by scanning mutagenesis. *P Natl Acad Sci USA* **100**, 13803-13808, (2003).
- 326 Rajapaksha, M. *et al.* Monitoring anthrax toxin receptor dissociation from the protective antigen by NMR. *Prot Sci* **18**, 17-23, (2009).
- 327 Varughese, M., Teixeira, A. V., Liu, S. & Leppla, S. H. Identification of a receptor-binding region within domain 4 of the protective antigen component of anthrax toxin. *Infect Immun* **67**, 1860-1865, (1999).
- 328 Blaustein, R. O., Koehler, T. M., Collier, R. J. & Finkelstein, A. Anthrax toxin: channel-forming activity of protective antigen in planar phospholipid bilayers. *P Natl Acad Sci USA* **86**, 2209-2213, (1989).
- 329 Hammamieh, R., Ribot, W. J., Abshire, T. G., Jett, M. & Ezzell, J. Activity of the Bacillus anthracis 20 kDa protective antigen component. *BMC Infect Dis* **8**, 124, (2008).
- 330 Milne, J. C., Furlong, D., Hanna, P. C., Wall, J. S. & Collier, R. J. Anthrax protective antigen forms oligomers during intoxication of mammalian cells. *J Biol Chem* **269**, 20607-20612, (1994).
- 331 Krantz, B. A. *et al.* A phenylalanine clamp catalyzes protein translocation through the anthrax toxin pore. *Science* **309**, 777-781, (2005).
- 332 Kintzer, A. F. *et al.* The protective antigen component of anthrax toxin forms functional octameric complexes. *J Mol Biol* **392**, 614-629, (2009).
- 333 Feld, G. K. *et al.* Structural basis for the unfolding of anthrax lethal factor by protective antigen oligomers. *Nat Struct Mol Biol* **17**, 1383-1390, (2010).
- 334 Singh, Y., Khanna, H., Chopra, A. P. & Mehra, V. A dominant negative mutant of Bacillus anthracis protective antigen inhibits anthrax toxin action in vivo. *J Biol Chem* **276**, 22090-22094, (2001).
- 335 Sellman, B. R., Mourez, M. & Collier, R. J. Dominant-negative mutants of a toxin subunit: an approach to therapy of anthrax. *Science* **292**, 695-697, (2001).
- 336 Sellman, B. R., Nassi, S. & Collier, R. J. Point mutations in anthrax protective antigen that block translocation. *J Biol Chem* **276**, 8371-8376, (2001).
- 337 Leppla, S. H., Arora, N. & Varughese, M. Anthrax toxin fusion proteins for intracellular delivery of macromolecules. *J Appl Microbiol* **87**, 284-286, (1999).
- 338 Aulinger, B. A., Roehrl, M. H., Mekalanos, J. J., Collier, R. J. & Wang, J. Y. Combining anthrax vaccine and therapy: a dominant-negative inhibitor of anthrax toxin is also a potent and safe immunogen for vaccines. *Infect Immun* **73**, 3408-3414, (2005).
- 339 Cao, S. *et al.* Investigation of new dominant-negative inhibitors of anthrax protective antigen mutants for use in therapy and vaccination. *Infect Immun* **77**, 4679-4687, (2009).

- 340 Singh, Y., Klimpel, K. R., Quinn, C. P., Chaudhary, V. K. & Leppla, S. H. The carboxyl-terminal end of protective antigen is required for receptor binding and anthrax toxin activity. *J Biol Chem* **266**, 15493-15497, (1991).
- 341 Rosovitz, M. J. *et al.* Alanine-scanning mutations in domain 4 of anthrax toxin protective antigen reveal residues important for binding to the cellular receptor and to a neutralizing monoclonal antibody. *J Biol Chem* **278**, 30936-30944, (2003).
- 342 Scobie, H. M., Rainey, G. J., Bradley, K. A. & Young, J. A. Human capillary morphogenesis protein 2 functions as an anthrax toxin receptor. *P Natl Acad Sci USA* **100**, 5170-5174, (2003).
- 343 Reeves, C. V., Dufraine, J., Young, J. A. & Kitajewski, J. Anthrax toxin receptor 2 is expressed in murine and tumor vasculature and functions in endothelial proliferation and morphogenesis. *Oncogene* **29**, 789-801, (2010).
- 344 Verma, K., Gu, J. & Werner, E. Tumor endothelial marker 8 amplifies canonical Wnt signaling in blood vessels. *Plos One* **6**, e22334, (2011).
- 345 Dowling, O. *et al.* Mutations in capillary morphogenesis gene-2 result in the allelic disorders juvenile hyaline fibromatosis and infantile systemic hyalinosis. *Am J Hum Genet* **73**, 957-966, (2003).
- 346 Hanks, S. *et al.* Mutations in the gene encoding capillary morphogenesis protein 2 cause juvenile hyaline fibromatosis and infantile systemic hyalinosis. *Am J Hum Genet* **73**, 791-800, (2003).
- 347 Jinnin, M. *et al.* Suppressed NFAT-dependent VEGFR1 expression and constitutive VEGFR2 signaling in infantile hemangioma. *Nat Med* **14**, 1236-1246, (2008).
- 348 Scobie, H. M. *et al.* Anthrax toxin receptor 2-dependent lethal toxin killing in vivo. *Plos Pathog* **2**, e111, (2006).
- 349 Christensen, K. A., Krantz, B. A. & Collier, R. J. Assembly and disassembly kinetics of anthrax toxin complexes. *Biochemistry-US* **45**, 2380-2386, (2006).
- 350 Lee, J. O., Bankston, L. A., Arnaout, M. A. & Liddington, R. C. Two conformations of the integrin A-domain (I-domain): a pathway for activation? *Structure* **3**, 1333-1340, (1995).
- 351 Liu, S., Leung, H. J. & Leppla, S. H. Characterization of the interaction between anthrax toxin and its cellular receptors. *Cell Microbiol* **9**, 977-987, (2007).
- 352 Bradley, K. A. & Young, J. A. Anthrax toxin receptor proteins. *Biochem Pharmacol* **65**, 309-314, (2003).
- 353 Fu, S. *et al.* The structure of tumor endothelial marker 8 (TEM8) extracellular domain and implications for its receptor function for recognizing anthrax toxin. *Plos One* **5**, e11203, (2010).
- 354 Bradley, K. A. *et al.* Binding of anthrax toxin to its receptor is similar to alpha integrin-ligand interactions. *J Biol Chem* **278**, 49342-49347, (2003).
- 355 Yang, M. Y. *et al.* The cell surface structure of tumor endothelial marker 8 (TEM8) is regulated by the actin cytoskeleton. *Biochim Biophys Acta* **1813**, 39-49, (2011).
- 356 Scobie, H. M. & Young, J. A. Divalent metal ion coordination by residue T118 of anthrax toxin receptor 2 is not essential for protective antigen binding. *Plos One* **1**, e99, (2006).
- 357 Liu, S. *et al.* Capillary morphogenesis protein-2 is the major receptor mediating lethality of anthrax toxin in vivo. *P Natl Acad Sci USA* **106**, 12424-12429, (2009).

- 358 Liu, S. *et al.* Anthrax toxin targeting of myeloid cells through the CMG2 receptor is essential for establishment of *Bacillus anthracis* infections in mice. *Cell Host Microbe* **8**, 455-462, (2010)
- 359 Scobie, H. M. & Young, J. A. Interactions between anthrax toxin receptors and protective antigen. *Curr Opin Microbiol* **8**, 106-112, (2005).
- 360 Smith, H. & Keppie, J. Observations on experimental anthrax; demonstration of a specific lethal factor produced in vivo by *Bacillus anthracis*. *Nature* **173**, 869-870, (1954).
- 361 Smith, H., Keppie, J. & Stanley, J. L. Observations on the cause of death in experimental anthrax. *Lancet* **267**, 474-476, (1954).
- 362 Pezard, C., Berche, P. & Mock, M. Contribution of individual toxin components to virulence of *Bacillus anthracis*. *Infect Immun* **59**, 3472-3477, (1991).
- 363 Friedlander, A. M. Macrophages are sensitive to anthrax lethal toxin through an acid-dependent process. *J Biol Chem* **261**, 7123-7126, (1986).
- 364 Klimpel, K. R., Arora, N. & Leppla, S. H. Anthrax toxin lethal factor contains a zinc metalloprotease consensus sequence which is required for lethal toxin activity. *Mol Microbiol* **13**, 1093-1100, (1994).
- 365 Arora, N. & Leppla, S. H. Residues 1-254 of anthrax toxin lethal factor are sufficient to cause cellular uptake of fused polypeptides. *J Biol Chem* **268**, 3334-3341, (1993).
- 366 Lacy, D. B., Mourez, M., Fouassier, A. & Collier, R. J. Mapping the anthrax protective antigen binding site on the lethal and edema factors. *J Biol Chem* **277**, 3006-3010, (2002).
- 367 Pannifer, A. D. *et al.* Crystal structure of the anthrax lethal factor. *Nature* **414**, 229-233, (2001).
- 368 Duesbery, N. S. Proteolytic inactivation of MAP-kinase-kinase by anthrax lethal factor. *Science* **280**, 734-736, (1998).
- 369 Vitale, G. *et al.* Anthrax lethal factor cleaves the N-terminus of MAPKKs and induces tyrosine/threonine phosphorylation of MAPKs in cultured macrophages. *Biochem Biophys Res Commun* **248**, 706-711, (1998).
- 370 Pellizzari, R., Guidi-Rontani, C., Vitale, G., Mock, M. & Montecucco, C. Anthrax lethal factor cleaves MKK3 in macrophages and inhibits the LPS/IFNgamma-induced release of NO and TNFalpha. *Febs Lett* **462**, 199-204, (1999).
- 371 Bardwell, A. J., Abdollahi, M. & Bardwell, L. Anthrax lethal factor-cleavage products of MAPK (mitogen-activated protein kinase) kinases exhibit reduced binding to their cognate MAPKs. *Biochem J* **378**, 569-577, (2004).
- 372 Park, J. M., Greten, F. R., Li, Z. W. & Karin, M. Macrophage apoptosis by anthrax lethal factor through p38 MAP kinase inhibition. *Science* **297**, 2048-2051, (2002).
- 373 Agrawal, A. *et al.* Impairment of dendritic cells and adaptive immunity by anthrax lethal toxin. *Nature* **424**, 329-334, (2003).
- 374 Leppla, S. H. Anthrax Toxin Edema Factor - a Bacterial Adenylate-Cyclase That Increases Cyclic-Amp Concentrations in Eukaryotic Cells. *P Natl Acad Sci-Biol* **79**, 3162-3166, (1982).
- 375 Shen, Y., Zhukovskaya, N. L., Guo, Q., Florian, J. & Tang, W. J. Calcium-independent calmodulin binding and two-metal-ion catalytic mechanism of anthrax edema factor. *Embo J* **24**, 929-941, (2005).

- 376 Drum, C. L. *et al.* An extended conformation of calmodulin induces interactions between the structural domains of adenylyl cyclase from *Bacillus anthracis* to promote catalysis. *J Biol Chem* **275**, 36334-36340, (2000).
- 377 Doganay, M. & Aydin, N. Antimicrobial susceptibility of *Bacillus anthracis*. *Scand J Infect Dis* **23**, 333-335, (1991).
- 378 Athamna, A. *et al.* Selection of *Bacillus anthracis* isolates resistant to antibiotics. *J Antimicrob Chemother* **54**, 424-428, (2004).
- 379 Price, L. B. *et al.* In vitro selection and characterization of *Bacillus anthracis* mutants with high-level resistance to ciprofloxacin. *Antimicrob Agents Chemother* **47**, 2362-2365, (2003).
- 380 Migone, T. S. *et al.* Raxibacumab for the treatment of inhalational anthrax. *N Engl J Med* **361**, 135-144, (2009).
- 381 Chen, Z. *et al.* Efficient neutralization of anthrax toxin by chimpanzee monoclonal antibodies against protective antigen. *J Infect Dis* **193**, 625-633, (2006).
- 382 Vitale, L. *et al.* Prophylaxis and therapy of inhalational anthrax by a novel monoclonal antibody to protective antigen that mimics vaccine-induced immunity. *Infect Immun* **74**, 5840-5847, (2006).
- 383 Mechaly, A. *et al.* A Novel Mechanism for Antibody-based Anthrax Toxin Neutralization inhibition of prepore-to-pore conversion. *J Biol Chem* **287**, 32665-32673, (2012).
- 384 Moayeri, M., Wiggins, J. F., Lindeman, R. E. & Leppla, S. H. Cisplatin inhibition of anthrax lethal toxin. *Antimicrob Agents Chemother* **50**, 2658-2665, (2006).
- 385 Komiyama, T., Swanson, J. A. & Fuller, R. S. Protection from anthrax toxin-mediated killing of macrophages by the combined effects of furin inhibitors and chloroquine. *Antimicrob Agents Chemother* **49**, 3875-3882, (2005).
- 386 Mourez, M. *et al.* Designing a polyvalent inhibitor of anthrax toxin. *Nat Biotechnol* **19**, 958-961, (2001).
- 387 Basha, S. *et al.* Polyvalent inhibitors of anthrax toxin that target host receptors. *P Natl Acad Sci USA* **103**, 13509-13513, (2006).
- 388 Slater, L. H. *et al.* Identification of novel host-targeted compounds that protect from anthrax lethal toxin-induced cell death. *ACS Chem Bio* **8**, 812-822, (2013).
- 389 Sanchez, A. M. *et al.* Amiodarone and bepridil inhibit anthrax toxin entry into host cells. *Antimicrob Agents Chemother* **51**, 2403-2411, (2007).
- 390 Chapsky, S., Batty, S., Frost, M. & Mogridge, J. Inhibition of anthrax lethal toxin-induced cytolysis of RAW264.7 cells by celastrol. *Plos One* **3**, e1421, (2008).
- 391 Joshi, A. *et al.* Structure-based design of a heptavalent anthrax toxin inhibitor. *Biomacromolecules* **12**, 791-796, (2011).
- 392 Shoop, W. L. *et al.* Anthrax lethal factor inhibition. *P Natl Acad Sci USA* **102**, 7958-7963, (2005).
- 393 Forino, M. *et al.* Efficient synthetic inhibitors of anthrax lethal factor. *P Natl Acad Sci USA* **102**, 9499-9504, (2005).
- 394 Rai, P., Vance, D., Poon, V., Mogridge, J. & Kane, R. S. Stable and potent polyvalent anthrax toxin inhibitors: raft-inspired domain formation in liposomes that contain PEGylated lipids. *Chemistry* **14**, 7748-7751, (2008).
- 395 Gujraty, K. *et al.* Functional characterization of peptide-based anthrax toxin inhibitors. *Mol Pharm* **2**, 367-372, (2005).

- 396 Gujrati, K. V. *et al.* Synthesis of polyvalent inhibitors of controlled molecular weight: structure-activity relationship for inhibitors of anthrax toxin. *Biomacromolecules* **7**, 2082-2085, (2006).
- 397 Barreto, K. *et al.* A Genetic Screen for Isolating "Lariat" Peptide Inhibitors of Protein Function. *Chem Biol* **16**, 1148-1157, (2009).
- 398 Iwai, H., Zuger, S., Jin, J. & Tam, P. H. Highly efficient protein trans-splicing by a naturally split DnaE intein from Nostoc punctiforme. *Febs Lett* **580**, 1853-1858, (2006).
- 399 Hochuli, E. Large-scale chromatography of recombinant proteins. *J Chromatogr* **444**, 293-302, (1988).
- 400 Cardoso, R. *et al.* Identification of Cys255 in HIF-1alpha as a novel site for development of covalent inhibitors of HIF-1alpha/ARNT PasB domain protein-protein interaction. *Prot Sci* **21**, 1885-1896, (2012).
- 401 Shen, B. & Lutkenhaus, J. Examination of the interaction between FtsZ and MinCN in *E. coli* suggests how MinC disrupts Z rings. *Mol Microbiol* **75**, 1285-1298, (2010).
- 402 Seo, D. W. *et al.* An integrin-binding N-terminal peptide region of TIMP-2 retains potent angio-inhibitory and anti-tumorigenic activity in vivo. *Peptides* **32**, 1840-1848, (2011).
- 403 Valero, J. G. *et al.* Bax-derived membrane-active peptides act as potent and direct inducers of apoptosis in cancer cells. *J Cell Sci* **124**, 556-564, (2011).
- 404 Brandstetter, H. *et al.* Structure of malonic acid-based inhibitors bound to human neutrophil collagenase. A new binding mode explains apparently anomalous data. *Prot Sci* **7**, 1303-1309, (1998).
- 405 Holden, M. T. *et al.* Genomic plasticity of the causative agent of melioidosis, *Burkholderia pseudomallei*. *P Natl Acad Sci USA* **101**, 14240-14245, (2004).
- 406 Chan, Y. Y., Tan, T. M., Ong, Y. M. & Chua, K. L. BpeAB-OprB, a multidrug efflux pump in *Burkholderia pseudomallei*. *Antimicrob Agents Chemother* **48**, 1128-1135, (2004).
- 407 Wuthiekanun, V. & Peacock, S. J. Management of melioidosis. *Expert Rev Anti Infect Ther* **4**, 445-455, (2006).
- 408 Pittman, P. R. *et al.* Anthrax vaccine: immunogenicity and safety of a dose-reduction, route-change comparison study in humans. *Vaccine* **20**, 1412-1420, (2002).
- 409 Odendaal, M. W., Pieterson, P. M. & Devos, V. The Antibiotic-Sensitivity Patterns of *Bacillus-Anthracis* Isolated from the Kruger National-Park. *Onderstepoort J Vet* **58**, 17-19, (1991).
- 410 Jin, D. *et al.* Molecular typing of the pathogenic *Yersinia enterocolitica* strains with pulsed field gel electrophores isolated in China. *Zhonghua Liu Xing Bing Xue Za Zhi* **27**, 677-680, (2006).
- 411 Klimpel, K. R., Molloy, S. S., Thomas, G. & Leppla, S. H. Anthrax toxin protective antigen is activated by a cell surface protease with the sequence specificity and catalytic properties of furin. *P Natl Acad Sci USA* **89**, 10277-10281, (1992).
- 412 Mogridge, J., Cunningham, K. & Collier, R. J. Stoichiometry of anthrax toxin complexes. *Biochemistry-US* **41**, 1079-1082, (2002).
- 413 Mehta, P. *et al.* Dynamics and stoichiometry of a regulated enhancer-binding protein in live *Escherichia coli* cells. *Nat Commun* **4**, 1997, (2013).
- 414 Young, T. S. *et al.* Evolution of cyclic peptide protease inhibitors. *P Natl Acad Sci USA* **108**, 11052-11056, (2011).

- 415 Xiong, Y. *et al.* The discovery of a potent and selective lethal factor inhibitor  
for adjunct therapy of anthrax infection. *Bioorg Med Chem Lett* **16**, 964-968,  
(2006).
- 416 Sugano, H., Higaki, K. & Miyoshi, M. Synthesis and Biological-Activity of  
Peptides Related to Eledoisin .1. Hexapeptide Amides Containing Alpha-  
Hydroxy Acids. *B Chem Soc Jpn* **46**, 226-230, (1973).
- 417 Chatterjee, J., Laufer, B. & Kessler, H. Synthesis of N-methylated cyclic  
peptides. *Nat protoc* **7**, 432-444, (2012).
- 418 Shemyakin, M. M., Ovchinnikov, Y. A. & Ivanov, V. T. Topochemical  
investigations of peptide systems. *Angewandte Chemie* **8**, 492-499, (1969).