Biochemical And Biophysical Studies On SilE From The *sil* Silver Resistance Locus

Karishma Asiani BSc. (Hons)

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

July 2016

ॐ श्री गणेशाय नम

(OM Shree Ganeshaya Namah)

Ganesha, I pray to you

Dedicated to my parents and brother,

for their love, support and inspiration

"Equipped with his five senses, man explores the universe around him and calls the adventure Science."

- EDWIN POWELL HUBBLE, Harpers Magazine, **1929**

Abstract

Metal ions such as silver (Ag^{+}), mercury (Hg^{2+}), zinc (Zn^{2+}) and copper (Cu^{+}/Cu^{2+}) have a long history of antimicrobial usage and some, such as Cu^{+}/Cu^{2+}, Ag^{+} and Zn^{2+} compounds are still used as antimicrobials. Prior to the introduction of antibiotics, Ag^{+} was arguably the most important antimicrobial and with the rapid emergence of antibiotic resistance, interest in Ag^{+} and its compounds as alternative antimicrobials have recently been revived. However, resistance to Ag^{+}-based compounds has been emerging, with initial reports of carriage of silver resistance on a *Salmonella enetrica* serovar Typhimurium multi-resistance plasmid pMG101 isolated from burns patients in 1975. The proposed model for the mechanism of Ag^{+} resistance encoded by the *sil* genes from pMG101 involves export of Ag^{+} ions via an ATPase (SilP), an RND family effluxer (SilCFBA) and a periplasmic chaperone of Ag^{+} (SilE).

SilE is a periplasmic protein predicted to be intrinsically disordered until it binds Ag^{+} ions. This hypothesis was tested using structural and biophysical studies which showed that SilE is an intrinsically disordered and unstructured protein in its free *apo*-form, but folds to a compact, defined structure upon optimal binding of six Ag^{+} ions in its *holo*-form. Sequence analyses and site-directed mutagenesis established the importance of histidine and methionine containing motifs for Ag^{+}-binding, and identified a nucleation core that initiates Ag^{+}-mediated folding of SilE. The data show that SilE is a molecular metal sponge absorbing up to a maximum of eight Ag^{+} ions.

Publications Arising from Work in this Thesis

1. Asiani, K.R., Williams, H., Bird, L., Jenner, M., Searle, M.S., Hobman, J.L., Scott, D.J., and Soultanas, P., (2016). SilE is an intrinsically disordered periplasmic 'molecular sponge' involved in bacterial silver resistance. *Molecular Microbiology*. **101**, pp. 731-742.

2. Hao, X., Lüthje, F.L., Qin, Y., McDevitt, S.F., Lutay, N., Hobman, J.L., Asiani, K., Soncini, F.C., German, N., Zhang, S., Zhu, Y.G., and Rensing, C., (2015). Survival in amoeba - a major selection pressure on the presence of bacterial copper and zinc resistance determinants? Identification of a 'copper pathogenicity island'. *Applied Microbiology and Biotechnology*. **99**, pp. 5817–5824.

Many of the ideas discussed in this thesis are also discussed in these papers. Reference to these papers in the main text has been deliberately restricted to a minimum, not out of any wish to diminish the importance of these works, but rather, to prevent the need to cite these works recurrently.

Acknowledgements

These four years have been an unforgettable and incredible experience for me. I have learnt many things and this thesis is the result of the support that I have been given by those whom I wish to dedicate the next few lines to.

First and foremost, my thanks go to my three supervisors Dr. David Scott, Prof. Panos Soultanas and Dr. Jon Hobman. I thank you for placing your confidence in me four years ago and giving me the opportunity to pursue my dream. Additionally, I am grateful not only for you allowing me to use your labs, but also for your continuous encouragement, guidance and useful discussions. I express my gratitude similarly to Prof. Mark Searle and Dr. Huw Williams for taking the time to help me with the NMR and for giving my research a new sense of direction. Thanks to all of the fantastic scientists of their labs, past and present, including Dr. Geoff Briggs for acquainting me to the lab at CBS and Thomas Miller for his attempts at making the mutants.

The thesis allowed me to conduct specialised synchrotron-based biophysical techniques and for this and your expertise acknowledgements go to Dr. Nykola Jones and Dr. Soren Vronning Hoffman at the CD Beamline, ASTRID 2 as well as Dr. Katsuaki Inoue at the SAXS Beamline, Diamond. I also offer my appreciation to: the OPPF at the RCaH for your assistance in high-throughput cloning, Dr. Lisa Crossman of SequenceAnalysis.co.uk and Dr. Chrystala Constantinidou at University of Warwick for the sequencing and ICP-MS Analytical Services at Sutton

Bonington and Dr. Matthew Jenner (now at University of Warwick) for conducting the mass spectrometry experiments.

A few special people at Sutton Bonington have made working there, well, not feel like work! Support staff in Food Microbiology, Lorraine Gillett, Dawn Beatson and David Flower I warmly thank you for all of your help. My partner-in-crime Aminu Umar, I really appreciate the moral support you have given me. A special acknowledgement to Princy Varghese and Hend Gahmi. There are probably no words to express the profound respect and affection I have for you both. Your unconditional support, generosity, kindness and friendship mean a lot to me! Thank you to the amazing and wonderful Dr. Neil Doherty, Dr. Tania Perehinec, Dr. Nicola Cummings, Anja Brøegger, Dewi Yunita, Delveen Ibrahim, James Williamson, Zara Gerrard, Ahmed Saeedi, Peter O'Kane, Nacheervan Ghaffar, Nathan Meade, Rebecca Chandler and Sarah Chapman for the positive attitude and the numerous laughs.

To my very best friends, Bhavisha Chudasama, Mital Patel, Aayushi Joshi, Nikita Patel, Chrishny Kangatharan, Asha Padhiar, Manki Modhwadia, Mehreen Alam and Mohammed Bux - thank you for your unconditional support, your encouragement, kindness and continued friendship in addition to brightening my day on countless occasions. All of you are such amazing people and I know that when we get old, you will still be there as supportive and caring friends.

Last but not least, I would like to thank my parents Ramesh and Saroj Asiani, and my brother Manish Asiani for your love, patience and unwavering support at all times. You always had confidence in my ability. You have been my rock and source of energy (I literally mean that, with all of those food and drink supplies). I can't begin to describe how blessed I feel to have such an amazing family! Thank you and I love you, always!

Contents

ABSTRACT ..... **iv**

PUBLICATIONS ..... **v**

ACKNOWLEDGEMENT ..... **vi**

CONTENTS ..... **viii**

INDEX OF FIGURES ..... **xiii**

INDEX OF TABLES ..... **xvi**

INDEX OF ABBREVIATIONS ..... **xvii**

TABLE OF AMINO ACIDS ..... **xxii**

CHAPTER 1 INTRODUCTION **1**

1.1 Antimicrobials **2**

1.1.1 Bacteria 2

1.1.1.1 Bacterial morphology, structure and classification 3

1.1.1.2 Bacterial pathogenesis 6

1.1.2 Antibacterials 7

1.1.2.1 Antibacterials in action 8

1.1.3 Antimicrobial resistance 10

1.2 Metal ions as antimicrobials **13**

1.2.1 The Irving-Williams series 13

1.2.2 Lewis acids 14

1.3 Silver **16**

1.3.1 Human toxicity of silver 16

1.3.2 Antimicrobial uses of silver 17

1.3.2.1 Bactericidal mechanism of silver 19

1.4 Silver resistance **23**

1.4.1 Evolution and prevalence of silver resistance 23

1.4.2 Bacterial mechanism of silver tolerance and resistance 25

1.4.2.1 Copper homeostasis and resistance 25

1.4.2.2 The sil system 30
1.4.3 Cross-resistance to silver 33
1.4.4 Co-resistance, mobile genetic elements and co-selection 34
1.4.5 Periplasmic silver-binding SilE 36
1.4.5.1 Intrinsically disordered proteins 37
1.4.5.2 Potential metal ligand binding sites 40

1.5 Aims and objectives 43

CHAPTER 2 MATERIALS AND METHODS 46

2.1 Materials 47
2.1.1 E. coli strains 47
2.1.2 Overexpression systems, cloning and plasmid vectors 51
2.1.3 Primers 54
2.1.4 Bacterial growth media 61
2.1.4.1 Luria Broth (LB) media 61
2.1.4.2 LB agar 61
2.1.4.3 M9 minimal media 61
2.1.4.4 Power broth 62
2.1.5 Buffers and solutions 63
2.1.5.1 DNA preparation 63
2.1.5.2 Protein overexpression, extraction and purification 64
2.1.5.3 Gel electrophoresis 64
2.1.5.4 Protein experimental 65
2.1.6 Markers 66

2.2 Methodology 67
2.2.1 Bioinformatics 67
2.2.2 Gel electrophoresis 70
2.2.2.1 Agarose gel electrophoresis 70
2.2.2.2 Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE) 70
2.2.3 E. coli growths and maintenance 71
2.2.3.1 Sterilisation 71
2.2.3.2 Standard growth conditions 72
2.2.3.3 Storage of strains 72
2.2.3.4 Making chemically competent *E. coli* 73
2.2.3.5 Making electrocompetent *E. coli* 74
2.2.4 Plasmid DNA preparations 74
2.2.4.1 Plasmid DNA isolations 74
2.2.4.2 DNA sample concentration and purity 75
2.2.4.3 Transformations 76
2.2.4.4 Cloning 77
2.2.5 Protein preparations 89

2.2.5.1 Small-scale expression screens 89
2.2.5.2 SilE protein production and purification 92
2.2.6 Biophysical methods 100
2.2.6.1 Far UV-circular dichroism spectroscopy 101
2.2.6.2 Mass spectrometry 109
2.2.6.3 Nuclear magnetic resonance spectroscopy 116
2.2.6.4 Small-angle X-ray scattering 122

CHAPTER 3 THE *sil* GENE CLUSTER 125

3.1 Introduction 126

3.2 Results and discussion 128
3.2.1 Understanding the *sil* locus and its surrounding 128
3.2.1.1 Plasmid pMG101 sequencing 130
3.2.1.2 Plasmid pMG101 AMR gene analysis 132
3.2.1.3 J53 (pMG101) whole genome sequence analysis 136
3.2.1.4 The Sil proteins 145
3.2.2 Production of the Sil proteins 158
3.2.2.1 Construction of the plasmid expression vectors 158
3.2.2.2 Small scale soluble Sil protein overexpression trials 161

3.3 Summay and conclusions 165

CHAPTER 4 OVEREXPRESSION, PURIFICATION AND SOLUTION CHARACTERISATION OF THE INTRINSICALLY DISORDERED, PERIPLASMIC PROTEIN SILE 168

4.1 Introduction 169

4.2 Results and discussion 170
4.2.1 Overexpression and solubility of SilE 170
4.2.2 SilE purification 172
4.2.2.1 Extraction 172
4.2.2.2 Immobilised metal affinity chromatography 172
4.2.2.3 Gel filtration chromatography 176
4.2.2.4 HRV-3C protease treatment and HisTrap™ purification 178
4.2.3 Molecular weight of SilE by mass spectrometry 181
4.2.4 Determining disorder in SilE 183
4.2.4.1 Prediction of intrinsic disorder in SilE 184
4.2.4.2 Assessment of folding, spatial distribution and structural content of SilE 191

4.3 Summary and conclusions 213

CHAPTER 5 STUDIES ON FOLDING OF THE SILVER BINDING SILE PROTEIN 214

5.1 Introduction 215

5.2 Results and discussion 216
5.2.1 Silver solubility and the buffering system 216
5.2.1.1 Selection of other metal ions 220
5.2.2 SilE binds up to eight silver ions 222
5.2.3 SilE secondary structure predictions 225
5.2.4 SilE-metal ion bound secondary structure and folding 226
5.2.4.1 SilE-Ag^+ binding and folding analysis via far-UV CD 227
5.2.4.2 Assessment of the metal ion binding and folding specificity of SilE 228
5.2.4.3 SilE-Ag^+ far-UV CD tritration 231
5.2.5 SilE folding is independent of the environmental pH 234
5.2.6 Assessment of the thermal stability of Ag^+-bound SilE 237
5.2.7 Attempts at deducing the folded structure of the SilE-Ag^+ complex 238
5.2.7.1 SilE-Ag^+ protein crystallisation 239
5.2.7.2 The SilE-Ag^+ SAXS problem 242
5.2.7.3 Progress towards the SilE-Ag^+ NMR structure 247

5.3 Summary and conclusions 259

CHAPTER 6 CHARACTERISATION OF SILVER BINDING TO THE SILE PROTEIN 262

6.1 Introduction 263

6.2 Results and discussion 264
6.2.1 Ag^+-binding site predictions 264
6.2.1.1 *in silico* binding site prediciton 265
6.2.1.2 Sequence homology based binding site prediction 268
6.2.2 Mutagenesis of histidine and methionine residues 271
6.2.2.1 PCR based site-directed mutagenesis 272
6.2.2.2 Overexpression of the SilE histidine and methionine mutant proteins 278
6.2.3 Histidine and methionine residues are involved in Ag^+-binding 280
6.2.3.1 Ag^+-induced folding affects of the SilE mutants 280
6.2.3.2 SilE mutant Ag^+-binding analyses by the ICP-MS 289
6.2.4 Two 'core-motifs' central to SilE likely to form the nucleation site for Ag^+-induced folding 292
6.2.4.1 Production of the SilE truncation mutants 292
6.2.4.2 Truncation of SilE affects Ag^+-induced protein folding 295
6.2.4.3 ICP-MS SilE truncation Ag^+-binding analyses 297

6.2 Summary and conclusions 299

CHAPTER 7 GENERAL DISCUSSION AND FUTURE WORK 301

7.1 Discussion 302

7.2 Future work 309

REFERENCES 311

APPENDIX A 333

APPENDIX B 334

WORD COUNT - 46,699

Index of Figures

Figure 1.1: Bacterial architecture 5

Figure 1.2: Mechanisms of antibiotic resistance in a Gram-negative bacterium 12

Figure 1.3: Modern uses of antibacterial silver 19

Figure 1.4: Antimicrobial activity of silver 21

Figure 1.5: Copper resistance *pco* genes and encoded protein functions 28

Figure 1.6: Silver resistance gene cluster and encoded protein functions 31

Figure 1.7: Model for Ag^{+} binding and folding of the periplasmic Ag^{+}-binding protein SilE 42

Figure 2.1: Map of the pOPINF plasmid expression vector containing the gene for $SilE_{21\text{-}143F}$ 53

Figure 2.2: Circular polarisation of light 103

Figure 2.3: Experimental set-up of CD spectroscopy 105

Figure 2.4: Characterisatic CD spectra for secondary structure 107

Figure 2.5: Schematic of an ESI source 112

Figure 2.6: Spins in a magnetic field 118

Figure 2.7: ^{1}H-^{15}N HSQC NMR magnetisation transfer 121

Figure 2.8: Schematic of SAXS experimentation 123

Figure 3.1: The *sil-pco* island 129

Figure 3.2: Plasmid pMG101 verification PCRs 131

Figure 3.3: RGI wheel of AMR genes in pMG101 133

Figure 3.4: Mauve geneome reorder alignments with J53 and MG1655 140

Figure 3.5: Mauve geneome reorder alignments with J53 (pMG101) 144

Figure 3.6: Sil protein sequence alignments 147

Figure 3.7: Variations in the Sil proteins 157
Figure 3.8: Agarose gel of *sil* gene PCR products 159
Figure 3.9: Agarose gels of cloning verification 160
Figure 3.10: SDS-PAGE analysis of non-membrane Sil protein expression screening experiments 163
Figure 3.11: SDS-PAGE analysis of membrane Sil protein expression screening experiments 164
Figure 4.1: SDS-PAGE analysis of normalised samples from expression trial 171
Figure 4.2: Nickel affinity chromatography based protein purification 175
Figure 4.3: Protein purification by gel filtration 177
Figure 4.4: HRV-3C protease treatment and protein His-tag removal 179
Figure 4.5: Pure SilE protein 180
Figure 4.6: Mass spectrum of SilE 182
Figure 4.7: Amino acid order/disorder IDP composition profile 186
Figure 4.8: Intrinsic disorder predictions of $SilE_{21-143}$ 190
Figure 4.9: Secondary structure content of SilE 195
Figure 4.10: SilE NMR spectra 199
Figure 4.11: SAXS buffer subtraction 202
Figure 4.12: Guinier plots of SilE 205
Figure 4.13: *Pr* analysis of SilE 208
Figure 4.14: Kratky plots 210
Figure 4.15: Kratky plot of SilE 210
Figure 4.16: Bead model of SilE 212
Figure 5.1: Solubility of silver salts 220
Figure 5.2: Determination of SilE metal ion binding by ESI-MS 224
Figure 5.3: SilE amino acids 21-143 secondary structure prediction 226
Figure 5.4: *Holo*-SilE secondary structure in the presence of Ag^+ 228
Figure 5.5: Metal ion bound secondary structure content of SilE 230
Figure 5.6: SilE-Ag^+ CD-based titration 233
Figure 5.7: *Apo*-SilE ^1H NMR pH titration 236
Figure 5.8: Melting curve of SilE, in the presence of Ag^+ 238
Figure 5.9: SilE-Ag^+ crystallisation plate experiment 241

Figure 5.10: SAXS averaged scattering curves in the presence of Ag^+ 243
Figure 5.11: SAXS individual scattering patterns 244
Figure 5.12: ^{15}N-SilE overexpression and purification 250
Figure 5.13: ^{1}H-^{15}N HSQC spectra of Ag^+-SilE..... 253
Figure 5.14: Ag^+-SilE ^{1}H-^{15}N HSQC at physiological pH 255
Figure 5.15: Ag^+-SilE ^{1}H-^{15}N TROSY HSQC 256
Figure 5.16: Ag^+-SilE ^{1}H-^{15}N HSQC pH titration..... 258
Figure 6.1: COACH protein model based ligand binding-site predictions ... 267
Figure 6.2: Protein sequence analysis and evaluations..... 270
Figure 6.3: Overview of site-directed mutagenic PCR by overlap..... 275
Figure 6.4: SilE site-directed mutagenic PCR by overlap..... 277
Figure 6.5: Mutant SilE SDS-PAGE 279
Figure 6.6: SilE His mutation effects on protein secondary structure..... 281
Figure 6.7: SilE His mutants affect protein Ag^+-induced folding 283
Figure 6.8: SilE His double and Met mutation effects on protein secondary structure..... 286
Figure 6.9: SilE double His and Met mutants affect protein Ag^+-induced folding 288
Figure 6.10: ICP-MS Ag^+-binding analyses of SilE and its mutants 291
Figure 6.11: Constructing the SilE truncation mutants 294
Figure 6.12: SilE truncation effects on protein Ag^+-induced folding..... 296
Figure 6.13: ICP-MS Ag^+-binding analyses of SilE truncation mutants..... 298
Figure 7.1: Proposed role of SilE in silver resistance..... 305
Figure 7.2: Nucleation model for Ag^+-binding induced SilE folding 306

Index of Tables

Table 1.1: Antibiotics by class 9
Table 2.1: *E. coli* strains used in this study 48
Table 2.2: Details of pOPIN vectors used for cloning Sil proteins 52
Table 2.3: Primers used throughout this study 55
Table 2.4: DNA and protein gel markers 66
Table 2.5: Bioinformatics software utilised throughout study 68
Table 2.6: SDS-PAGE resolving gel components 71
Table 2.7: The Sil proteins, genes, PCR products and plasmid expression vectors 78
Table 2.8: Cloning and mutagenesis PCR experiments 82
Table 2.9: Mutant gene insert restriction digests 87
Table 2.10: Ligation reaction 88
Table 2.11: ESI-friendly buffers 111
Table 3.1: Assembly statistics for plasmid pMG101 sequencing contigs 131
Table 3.2: AMR genes carried by plasmid pMG101 134
Table 3.3: Assembly statistics for *E. coli* J53 (pMG101) 137
Table 3.4: Differences in the amino acids of predicted Sil proteins from sequencing of J53 (pMG101) compared to the original published sequence 155
Table 5.1: Silver salts 217
Table 5.2: Metal bound CD secondary structure propensity NRMSE 231
Table 6.1: Percentage of mutant SilE secondary structure calculated from far-UV CD data 284
Table 6.2: Percentage of truncation mutant SilE secondary structure calculated from far-UV CD data 297

Index of Abbreviations, Acronyms, Formulas and Symbols

A_{260}	Absorbance at 260 nm
A_{280}	Absorbance at 280 nm
AAR	Amino Acid Repeat
Ag	Silver
Ag^{+}	Silver (I) ion
$AgNO_3$	Silver Nitrate
AMR	AntiMicrobial Resistance
Amp	Ampicillin
APEC	Avian Pathogenic *Escherichia coli*
approx.	Approximately
ASTRID	Aarhus Storage Ring Denmark
bp	Base Pair(s)
$C_2H_3O_2NH_4$	Ammonium Acetate
Cam/Cat	Chloramphenicol
CAMEO	Continuous Automated Model EvalutatiOn
CARD	Comprehensive Antibiotic Resistance Database
CASP	Critical Assessment for Protein Structure Prediction
Cd	Cadmium
CD	Circular Dichrosim
CDS	CoDing Sequence
CFU	Colony-Forming Unit
CH	Charge-Hydropathy

CS	Chemical Shift
Cu	Copper
Cu^{+}	Copper (I)
Cu^{2+}	Copper (II)
$CuSO_4$	Cupric Sulfate/Copper (II) Sulfate
Dmax	Maximum inter-atomic Distance
DMSO	Dimethyl sulfoxide
DNA	DeoxyriboNucleic Acid
D_2O	Deuterium
E. coli	*Escherichia coli*
E. cloacae	*Enterobacter cloacae*
ETMD	Electron-Transfer Mediated Decay
ESI	ElectroSpray Ionisation
ETEC	Enterotoxigenic *Escherichia coli*
Fe^{2+}	Iron (II)
GC	Guanine-cytosine
H	Hydrogen
^{1}H	Proton
HDMS	High Definition Mass Spectrometry
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
HME-RND	Heavy Metal Efflux-Resistance Nodulation Cell Division
HRV-3C	Human Rhinovirus-3C
HSAB	Hard-Soft Acid Base
HSQC	Heteronuclear Single Quantum Coherence
H_2O	Water
I	Scattering Intensity
ICD	Interatomic Coulombic Decay
ICP	Inductively Coupled Plasma
ID	Intrinsic(ally) Disorder(ed)
IDP	Intrinsically Disordered Protein

IDR	Intrinsically Disordered Region
IPTG	Isopropyl β-D-1-Thiogalactopyranoside
IUP	Intrinsically Unordered Protein
IUR	Intrinsically Unordered Region
kb	Kilo Base(s)
K. pneumonia	*Klebsiella pneumoniae*
L	Litre(s)
LB	Luria Broth
LPS	Lipopolysaccharides
mdeg	Millidegree(s)
mL	Millilitre(s)
mM	Millimolar
M	Molar
MALDI	Matrix Assisted Laser Desorption/Ionization
Mg^{2+}	Magnesium (II)
$MgCl_2$	Magnesium Chloride
min(s)	Minute(s)
MRSA	Methicillin-Resistant *Staphylococcus aureus*
ms	Millisecond(s)
MOPS	3-(N-Morpholino)propanesulfonic acid
MS	Mass Spectrometry
MUSTER	MUlti-Sources ThreadER
MW	Molecular Weight
MWCO	Molecular Weight Cut Off
M^{-1} cm^{-1}	Molar Absorbance Units
µL	Microlitre
µM	Micromolar
NaCl	Sodium Chloride
Na_2HPO_4/NaH_2PO_4	Sodium Phosphate
NH_4AcO	Ammonium Acetate
Ni	Nickel

Ni^{2+}	Nickel (II)
$NiCl_2$	Nickel Chloride
nM	Nanomolar
NMR	Nucelar Magnetic Resonance
NNP	Neural Network Predictor
NRMSD/E	Normalised Root Mean Square Deviation/Error
NUP	Natively Unfolded Protein
O/N	OverNight
ORF	Open Reading Frame
PAGE	Polyacrylamide Gel Electrophoresis
Pd	Palladium
pI	Isoelectric Point
ppm	Parts Per Million
ppq	Parts Per Quadrillion
ppt	Parts Per Trillion
PrDOS	Protein DisOrder prediction System
q	Scattering Angle
RAST	Rapid Annotation Subsystem Technology
Rg	Radius of Gyration
RGI	Resistance Gene Identifier
SAXS	Small-Angle X-ray Scattering
sdH_2O	Sterile Water
SDS	Sodium Dodecyl Sulfate
sec (s)	Second(s)
S(q)	Structure Factor of Solution
SRCD	Synchrotron Radiation Circular Dichroism
Sul	Sulphonamide
Tet	Tetracycline
TOF	Time of Flight
Tris base/Trizma	Tris(hydroxymethyl)aminomethane

TROSY	Transverse Relaxation Optimised Spectroscopy
TWIM	Travelling-Wave Ion Mobility
UV	Ultraviolet
WCL	Whole Cell Lysate
WT	Wild-Type
Zn	Zinc
Zn^{2+}	Zinc (II)
$ZnCl_2$	Zinc Chloride
1D	One-Dimensional
2D	Two-Dimensional

Table of Amino Acid Abbreviations

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

CHAPTER 1

Introduction

1.1 Antimicrobials

An antimicrobial is defined as an agent that destroys or inhibits the multiplication or growth of microorganisms, particularly pathogenic microorganisms (Oxford Dictionaries, 2011). Antimicrobial agents come in the form of:

(i) Disinfectants that are used to kill microorganisms on inanimate objects or surfaces - i.e. bleach,

(ii) Detergents, which are inclusive of surfactants with cleaning properties - i.e. laundry detergents,

(iii) Antibiotics that are used in the treatment and prevention of bacterial infections and

(iv) Antiseptics, which are used as a preventative measure against infections (Mcdonnell and Russell, 1999).

The microorganisms in question are typically bacteria, although there are fungicides and virucides too.

1.1.1 Bacteria

Spanning almost four billion years of evolution, bacteria are one of the first forms of microbial life to exist on earth (DeLong and Pace, 2001). These typically single-celled prokaryotic microorganisms exist in abundance in living hosts and all over the planet. Bacteria are known to: inhabit water, soil, and also coexist symbiotically and parasitically with plants and animals. The number of bacterial cells in human flora alone is

approximately ten times greater than the number of human cells in the body (Tlaskalová-Hogenová *et al.*, 2011).

1.1.1.1 Bacterial morphology, structure and classification

Bacterial cells are microscopic in size and display diversity in shape (commonly helical/spiral, rod-shaped or spherical), yet exhibit a high level of structure and intracellular organisation (Figure 1.1). Bacteria are prokaryotes and lack membrane bound organelles that are found present in eukaryotes, such as mitochondria, Golgi bodies and endoplasmic reticulum in addition to a true nucleus (Silhavy *et al.*, 2010). Instead, chromosomal DNA (in the form of a single continuous strand) is localised at the nucleoid within the cytoplasm of a cell. Alongside the intracellular presence of ribosomes and other proteins and enzymes, extra-chromosomal genetic structures known as plasmids can also be found in the cytoplasm (Ahmed *et al.*, 2014). Plasmids are made of a circular piece of DNA that can replicate independently of the chromosome and confer some form of selective advantage on the bacterial host. Plasmids can transfer to other bacteria by conjugation and mobilisation. Other mechanisms for horizontal gene transfer between bacteria include transformation and transduction (Peters and Craig, 2001).

A cell envelope surrounds and encloses the cytoplasmic contents of a bacterial cell and acts as a protective barrier (Silhavy *et al.*, 2010). A cell envelope is generally made up of cell membrane(s), a cell wall and occasionally an exterior polysaccharide-containing capsule. The cell envelopes of bacteria fall into two major groups that can be used to classify

bacteria. These are Gram-negative and Gram-positive bacteria (Figure 1.1) – names that originate from the reaction of cells to Gram stain (Ahmed *et al.*, 2014). Whilst both Gram-negative and Gram-positive bacteria cell envelopes comprise an inner membrane and cell walls; an outer membrane (containing lipopolysaccharide) that does not exist in Gram-positive bacteria, surrounds Gram-negative bacteria cell walls. Besides lacking an outer-membrane, Gram-positive bacteria differ by possessing a thickly layered peptidoglycan cell wall (Figure 1.1) (Silhavy *et al.*, 2010).

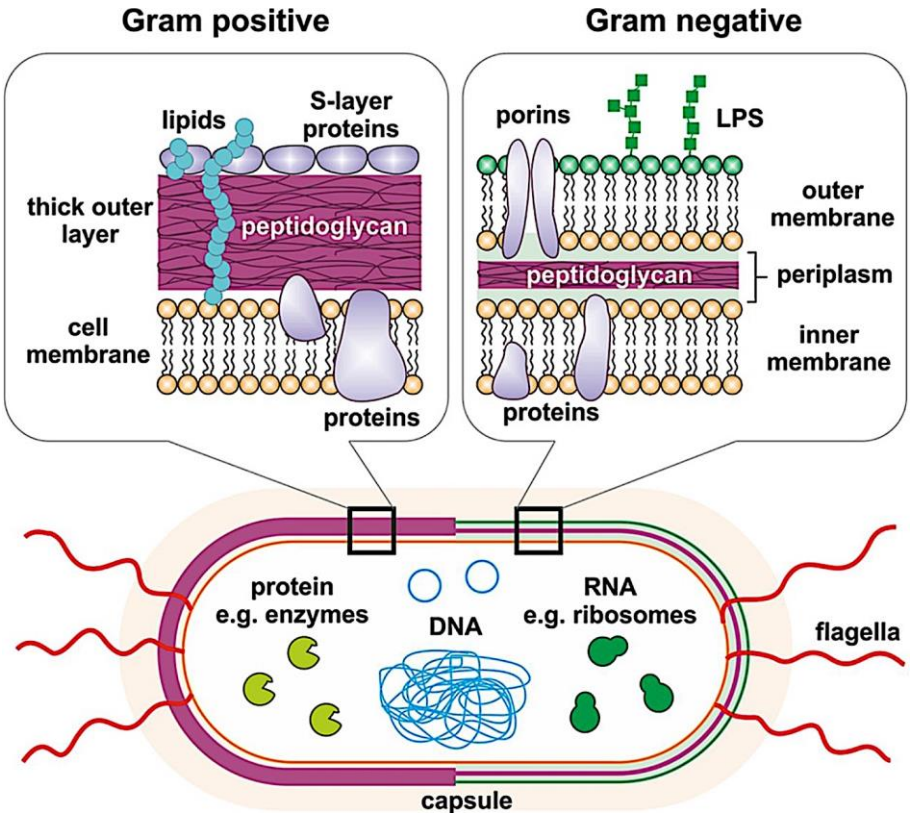


Figure 1. 1 Bacterial architecture

A diagram illustrating the structure and contents of a bacterial cell. The cell wall of Gram-positive bacteria comprises a thick layer of peptidoglycan (pink/purple), which also contains lipids and other protein components, surrounding a lipid membrane. In contrast, Gram-negative bacteria possess a much thinner peptidoglycan layer sandwiched in between two cell membranes. The outer membrane contains proteins, such as porins, as well as lipopolysaccharides (LPS), also known as endotoxin. The inner membranes of both types of bacteria contain various proteins. Both types of bacteria may have flagella. Additional intracellular contents include proteins, chromosomal and plasmid DNA and RNA. Figure used with publishers permission, from Ahmed *et al.*, (2014).

Even though most bacteria are harmless, if not beneficial, several species are known to be pathogens (i.e. a microorganism that can cause disease) (Wilson *et al.*, 2002; Tlaskalová-Hogenová *et al.*, 2011). Pathogenic Gram-negative bacteria include: *Escherichia coli* (*E. coli*),

Klebsiella pneumoniae (*K. pneumoniae*), *Acinetobacter baumanii*, *Pseudomonas aeruginosa*, *Haemophilus influenzae*, *Neisseria gonorrhoeae*, *Legionella*, *Salmonella*, *Shigella*, and *Enterobacter* species. Examples of pathogenic Gram-positive bacteria include: *Enterococcus faecium*, *Streptococcus pneumoniae*, *Staphylococcus aureus*, *Bacillus*, and *Clostridium* species. Only once a microorganism has successfully established a site of infection in a host, can disease occur. Therefore, bacterial pathogens use several mechanisms to colonise and invade human organs, despite the presence of multiple host defence systems.

1.1.1.2 Bacterial pathogenesis

When causing infection, certain structures of bacteria are known to interact with a host to evade host defence strategies and damage the host cell. Capsules can act to protect pathogenic bacteria against phagocytes, such as macrophages and neutrophils that engulf harmful foreign bodies (Tlaskalová-Hogenová *et al.*, 2011). When a host immune system lyses a bacterial cell, lipopolysaccharides (LPS - a major component of the Gram-negative outer-membrane (Figure 1.1)) are released. These are known as endotoxins and they can elicit septic shock, leading to ischemia and organ dysfunction (Wilson *et al.*, 2002). Likewise, exotoxins also serve to damage host cells and aid in host cell invasion. These toxins are mostly found in Gram-positive bacteria. Adhesins, commonly found in capsules, cell walls and/or pilli also facilitate binding of the pathogen to host surfaces (Tlaskalová-Hogenová *et al.*, 2011).

Once a bacterial pathogen has adhered to a host surface, it may then

further invade host tissues by secreting proteases and also gain access to the intracellular environment of the host cell (Wilson *et al.*, 2002). Here bacteria can cause further toxin mediated damage and prevent host nutrient uptake. One example includes the secretion of sidrephores that have a high affinity for iron and thereby take the iron away from host iron-transport proteins (Silhavy *et al.*, 2010). The degree to which these various mechanisms play a part in the pathogenesis of an infection depends on the bacterial species or strain, the site of pathogen entry, the immune status of the host and other similar factors (Wilson *et al.*, 2002). However, characterising structural features and understanding the pathophysiology of these disease-causing bacteria allowed for the development and use of therapeutic and preventative antibacterials.

1.1.2 Antibacterials

Derived from the Greek words ‘anti’ (against), ‘mikros’ (little) and ‘bios’ (life), the word ‘antimicrobial’, an antimicrobial is defined as an agent that kills or inhibits the growth of microorganisms (Oxford Dictionaries, 2011). Stemming from antimicrobials, antibacterial agents used to treat bacterial infections, are the most commonly used (Leekha *et al.*, 2011). Records contain information on the use of naturally occurring antibacterials, dating back to ancient Egyptian and Greek periods (Forrest, 1982). These include honey and plant extracts, in addition to toxic metal ions such as mercury (Hg), arsenic (As), copper (Cu) and silver (Ag) used to disinfect water and preserve food (Mcdonnell and Russell, 1999; Jung *et al.*, 2008; Alexander, 2009; Lemire *et al.*, 2013; Barillo and Marx, 2014;

Hobman and Crossman, 2015). The 19th century saw an introduction in the use of alcoholic or phenolic disinfectants, detergents and antiseptics, yet the 20th century has been noted as the most significant era for the discovery and development of a number of structurally diverse and highly effective antibacterial therapeutics. Antibacterials are commonly classified according to their mechanism of action, effect on bacteria or spectrum of activity.

1.1.2.1 Antibacterials in action

Depending on the range of bacterial species susceptible to a particular agent, the antibacterial is categorised either broad-spectrum or narrow-spectrum based on its target specificity. Broad-spectrum antibacterials are active against both Gram-negative and Gram-positive bacteria. Whereas, narrow-spectrum antibacterials are target specific and have limited activity against only a particular species i.e. Gram-negative or Gram-positive bacteria only. Due to the different mechanistic effects that various antibacterials have on bacteria, the agents can be further classified conditional to an endpoint of bacterial inactivation or death. Bactericidal agents are known to kill bacteria by targeting the bacterial cell wall or the cell membrane, or by interfering with essential bacterial enzymes – for example aminoglycosides, cephalosporins, penicillins, and quinolones. While those that target protein or DNA synthesis and merely delay or inhibit bacterial growth and replication, such as tetracyclines, sulfonamides, and macrolides, are called bacteriostatic agents (Table 1.1) (Finberg *et al.*, 2004).

Table 1. 1: Antibiotics by class

A list of some commonly used antibiotics, alongside generic examples of and their modes of action. Information gathered from Giedraitienė *et al.*, (2011) and eMedExpert.com, (2016).

Class of Antibiotic	Generic Examples	Spectrum of Activity	Mode of Action
Aminoglycosides	Streptomycin	Gram-negative aerobes and some anaerobic *Bacilli*.	Inhibit protein synthesis
Cephalosporins	Cefadroxil Ceftobiprole	First-generation cephalosporins are predominantly active against Gram-positive bacteria. Successive generations have increased activity against Gram-negative bacteria.	Class of β-lactam antibiotics - Inhibit cell wall synthesis
Macrolides	Erythromycin	Gram-positive and limited Gram-negative bacteria.	Inhibit protein synthesis
Penicillins	Ampicillin	Broad-spectrum - wide range of Gram-positive and Gram-negative bacteria.	Class of β-lactam antibiotics - Inhibit cell wall synthesis
Quinolones	Ciprofloxacin	Broad-spectrum - wide range of Gram-positive and Gram-negative bacteria.	Inhibit nucleic acid synthesis
Sulfonamides	Sulfamethoxazole	Broad-spectrum - wide range of Gram-positive and Gram-negative bacteria.	Inhibit metabolic process - folic acid pathway
Tetracyclines	Tetracycline	Broad-spectrum - wide range of Gram-positive and Gram-negative bacteria.	Inhibit protein synthesis

1.1.3 Antimicrobial resistance

Organisms develop mechanisms and actions to remove any threat posed to them. This is resistance. Resistance can be further defined as the action or ability to prevent the effects of a dangerous product or substance. Therefore any overexposure of a toxic compound, such as an antimicrobial, will allow a microorganism to become resistant to it under selective pressure, as explained below. As a consequence of the widespread and injudicious use of antimicrobials, there has been an accelerated emergence of antimicrobial resistant pathogens, resulting in an increasing health and economic problem (Giedraitienė *et al.*, 2011; Pal *et al.*, 2014). Besides the treatment of human pathogens, antimicrobials are used in a diverse range of settings. They are used: in the food and agriculture industry, to promote growth and improve feed efficiency of animals and to manage infections in aquaculture (Allen *et al.*, 2010). Through the process of natural selection, these environmental pressures have permitted the spread and increase in resistance to one or more antimicrobials (multidrug resistance). Multidrug resistance has been found present in *Pseudomonas aeruginosa*, *Acinetobacter baumannii*, *E. coli*, *K. pneumoniae*, *Enterococcus faecium*, Methicillin-resistant *Staphylococcus aureus*, *Streptococcus pneumonia*, *Mycobacterium tuberculosis*, *Salmonella enterica* serovar Typhimurium, *Shigella dysenteriae*, *Neisseria gonorrhoeae*, *Haemophilus influenzae*, *Stenotrophomonas, and Burkholderia* (Giedraitienė *et al.*, 2011; Sandegren *et al.*, 2012; Lemire *et al.*, 2013; World Health Organisation, 2014).

Bacteria may naturally present (i.e. endogenous genetic mutations) or acquire (exogenous) genes that are resistant to antimicrobials. Acquired

resistance can arise from: (i) mutations in bacterial cell genes (chromosomal mutation) leading to cross-resistance, (ii) gene transfer from one microorganism to another by plasmids (conjugation or transformation), transposons (conjugation), integrons and bacteriophages (transduction) and (iii) a combination of these mechanisms (Giedraitienė *et al.*, 2011). Following acquisition of these antimicrobial resistance genes, bacteria can use several mechanisms of resistance. Common mechanisms of resistance can be seen in Figure 1.2. These include: antimicrobial inactivation, protein target modification, altered membrane permeability, 'bypassing' metabolic pathways and efflux pumps. Efflux pumps are recognised as the main multidrug resistance mechanism in bacteria (Giedraitienė *et al.*, 2011).

Following a rise in antibiotic resistance, there is a new appreciation towards the exploitation of metal ion toxicity, being used as an effective means of controlling bacterial growth in clinical and environmental settings (Osman and Cavet, 2010).

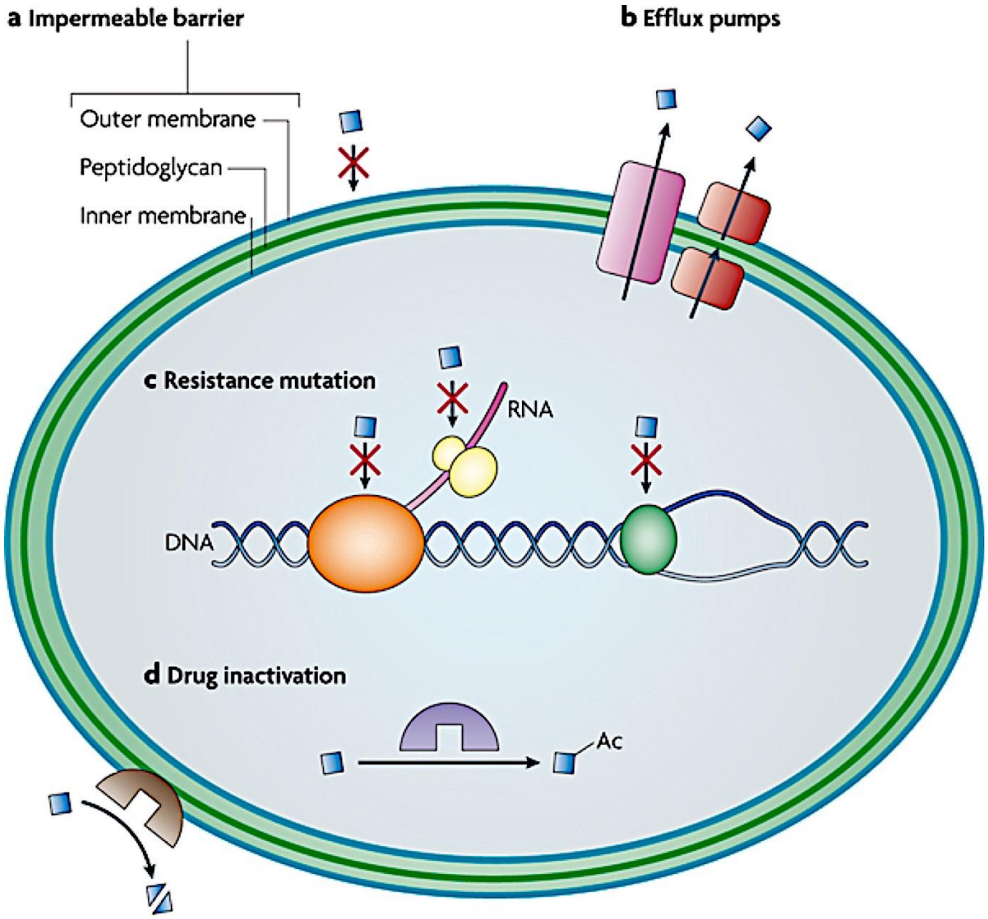


Figure 1. 2 Mechanisms of antibiotic resistance in a Gram-negative bacterium

A. Impermeable barriers. Some bacteria are intrinsically resistant to certain antibiotics (blue squares) simply because they have an impermeable membrane or lack the target of the antibiotic. **B.** Multidrug resistance efflux pumps. These pumps secrete antibiotics from the cell. Some transporters, such as those of the resistance–nodulation–cell division family (pink), can pump antibiotics directly outside the cell, whereas others, such as those of the major facilitator superfamily (red), secrete them into the periplasm. **C.** Resistance mutations. These mutations modify the target protein, for example by disabling the antibiotic-binding site but leaving the cellular functionality of the protein intact. Specific examples include mutations in the gyrase (green), which cause resistance to floroquinolones, in RNA polymerase subunit B (orange), which cause resistance to rifampicin, and in the 30S ribosomal subunit protein S12 (encoded by *rpsL*) (yellow), which cause resistance to streptomycin. **D.** Inactivation of the antibiotic. Inactivation can occur by covalent modification of the antibiotic, such as that catalysed by acetyltransferases (purple) acting on aminoglycoside antibiotics, or by degradation of the antibiotic, such as that catalysed by β-lactamases (brown) acting on β-lactam antibiotics. Ac, acetyl group. Image used with publishers permission, from Allen *et al.*, (2010).

1.2 Metal ions as antimicrobials

Living cells require some cationic metals as trace nutrients for their normal functioning. These include: copper (Cu^{2+}), zinc (Zn^{2+}), nickel (Ni^{2+}) and cobalt (Co^{2+}); known as the essential metals. While the presence of some non-essential metals that possess no known biological function, is toxic to the cell. For example, lead (Pb^{2+}), mercury (Hg^{2+}), silver (Ag^{+}) and cadmium (Cd^{2+}). The antimicrobial activities of such heavy metals have been exploited by humans in different sectors such as medicine, agriculture and other human activities (Gupta *et al.*, 2001; Lemire *et al.*, 2013).

Antimicrobial metal toxicity occurs via a number of mechanisms: by causing protein dysfunction through binding to or blocking of the functional groups of biological molecules; by displacing essential metals and reducing nutrient uptake inhibition; and through genotoxicity arising from chemical reactions harmful to the cell producing reactive oxygen species (Lemire *et al.*, 2013; Hobman and Crossman, 2015). It is apparent that it is the chemistry of the metal that determines its ability to exert antimicrobial toxic effects. This chemistry, described by the interactions of a metal ion and its biological ligand, has been classified into two categories. These are the Irving-Williams series of divalent metal complex stabilities and the classification of metals into Lewis acids.

1.2.1 The Irving-Williams series

In 1953, Harry Irving and Robert Williams observed the ligand affinity for essential divalent transition metal ions (Irving and Williams,

1953). They found that within the first-row of metal ions, the stability of complexes formed between a metal and its biological ligand generally increases across the periodic table (Appendix B): $Mn^{2+} < Fe^{2+} < Co^{2+} < Cu^{2+} > Zn^{2+}$. Evidently, Cu^{2+} has a strong affinity for biological molecules suggesting that it can displace other divalent essential metals resulting in toxicity through cellular protein dysfunction (Lemire *et al.*, 2013; Hobman and Crossman, 2015).

1.2.2 Lewis acids

Metal ions are Lewis acids that react with a Lewis base donor ligand to form a Lewis adduct. Therefore the hard-soft acid base (HSAB) theory, thus considering the strength of the metal ion as a Lewis acid, is another way of measuring metal toxicity (Hobman and Crossman, 2015). From ordering transition metals according to their preferences for specific organic ligands, the HSAB theory can predict the selectivity of metal ions for biological donor ligands (Pearson, 1963). Hard Lewis acids have a smaller ionic radius hence, weak polarisabilty (i.e. a low tendency of electron distribution to distort towards another nucleus) and prefer ionic coordination to oxygen containing ligands. By contrast, soft Lewis acids have a large ionic radius, strong polarisabilty and prefer covalent coordination to soft Lewis bases; such as the cysteine sulphydryl (R-SH) groups and nitrogen imidazole rings found in proteins. Thereby, soft Lewis acids displace hard and intermediate Lewis acids due their higher affinity for such biological molecules. Intermediate Lewis acids, known as borderline, will relatively stably coordinate to hard and soft donor ligands (Lemire *et al.*, 2013). Most

metals that are known to be toxic are soft Lewis acids. These include: Hg^{2+}, Cu^{+}, Ag^{+} and Cd^{2+}. Of these, due to the increasing multidrug resistance and the lack of new antibiotics in the pipeline, there has recently been a renewed interest in the use of Ag^{+} as an antimicrobial agent (Percival *et al.*, 2005; Chopra, 2007; Lansdown, 2010; Lemire *et al.*, 2013; Mirolo *et al.*, 2013; Durán *et al.*, 2016). Therefore, the following section discusses its uses and mechanism of microbial toxicity.

1.3 Silver

Silver (Ag) is a shiny, lustrous, precious and soft Lewis acid metal (Lansdown, 2010). With high value as a human commodity in jewellery and as an investment, and wide applications in the electronics industry; approximately 24,000 tons of silver was mined and produced in 2012 (Mijnendonckx *et al.,* 2013). Silver has also been highly valued for its broad-spectrum antimicrobial properties and has been one of the most important antimicrobial agents prior to the discovery and introduction of antibiotics.

1.3.1 Human toxicity of silver

Silver is particularly useful as an antimicrobial as it is viewed as benign, with low levels of toxicity to humans. The only negative health effects reported have been through the risks associated with absorption of silver ions (Ag^+). Alongside rare eschar formation and elevated Ag^+ levels in blood (Hobman and Crossman, 2015), it appears that the only side effect to prolonged exposure is argyria and argyriosis. Argyria and argyriosis are characterised by the deposition of Ag^+ compounds in the skin or eyes leading to a bluish grey discolouration (Mijnendonckx *et al.,* 2013). Although, this condition is irreversible it is not known to be life threatening, with individuals living for many years with the condition (Clement and Jarrett, 1994; Gupta *et al.,* 2001; Silver, 2003; Atiyeh *et al.,* 2007; Mijnendonckx *et al.,* 2013). Argyria and argyriosis are also thought to be where the term ‘blue blood’ came from, as rich families who ate mainly

with silver cutlery would sometimes develop this bluish grey hue. Interestingly, this was associated with greater immunity during epidemics of various diseases due to a lower death rate (Alexander, 2009).

1.3.2 Antimicrobial uses of silver

Historic records, dating back to the eras before the ancient Greeks, Romans and Egyptians, mention how silver preparations (in one form or another) were used to: treat ulcers, promote wound healing and to preserve food and water (Alexander, 2009; Lemire *et al.*, 2013; Barillo and Marx, 2014). Silver became the leading antimicrobial agent in the early 20th century, until its displacement in the 1940s - following the discovery of antibiotics (Chopra, 2007). Even today, silver is being used in an extensive array of applications because of its antibacterial properties and low toxicity to humans. The rapid emergence of antibiotic resistance among many bacteria has rejuvenated the interest in silver as a viable alternative antimicrobial agent (Holt and Bard, 2005; Atiyeh *et al.*, 2007; Mijnendonckx *et al.*, 2013). For example, recent research has shown that Ag^+ can be used to potentiate and expand the activity of existing antibiotics (Morones-Ramirez *et al.*, 2013).

Recent advances in technology have allowed ionisable silver compounds (with Ag^+) to be impregnated into or coat numerous materials and to produce silver nanoparticles, providing effective means of overcoming infections and bacterial biofilms (Lansdown, 2010; Habash *et al.*, 2014; Franci *et al.*, 2015; Loo *et al.*, 2015; Durán *et al.*, 2016). Silver

sulphadiazine is formulated in creams and dressings as topical silver used to treat chronic wounds (Percival *et al.*, 2005), implanted medical devices (for instance catheters, endotracheal tubes and heart valves) have silver alloyl- and hydrogel-coatings to avoid bacterial biofilm growth (Crabtree *et al.*, 2003), and other preventative measures include the use of silver nitrate for eye care (Mijnendonckx *et al.*, 2013). Silver zeolite is used in dental silver amalgams, denture resins, as well as mouthwashes to help maintain good oral hygiene (Saengmee-anupharb *et al.*, 2013). Silver-containing surfaces have been proposed for surface hygiene in hospital wards (Page *et al.*, 2009; Siedenbiedel and Tiller, 2012).

Besides its use in biomedical applications, silver is used with copper as a co-disinfectant of water systems (from public swimming pool waters to domestic water filters) and to keep food 'safer' through the incorporation of silver nanoparticles into commercially available food storage containers (Silvestre *et al.*, 2011). It has also made its way into a range of consumer and 'lifestyle' products including - clothing (particularly athletic socks and underwear), cosmetics, deodorants, detergents and antibacterial sprays (Figure 1.3) (Marambio-Jones and Hoek, 2010).

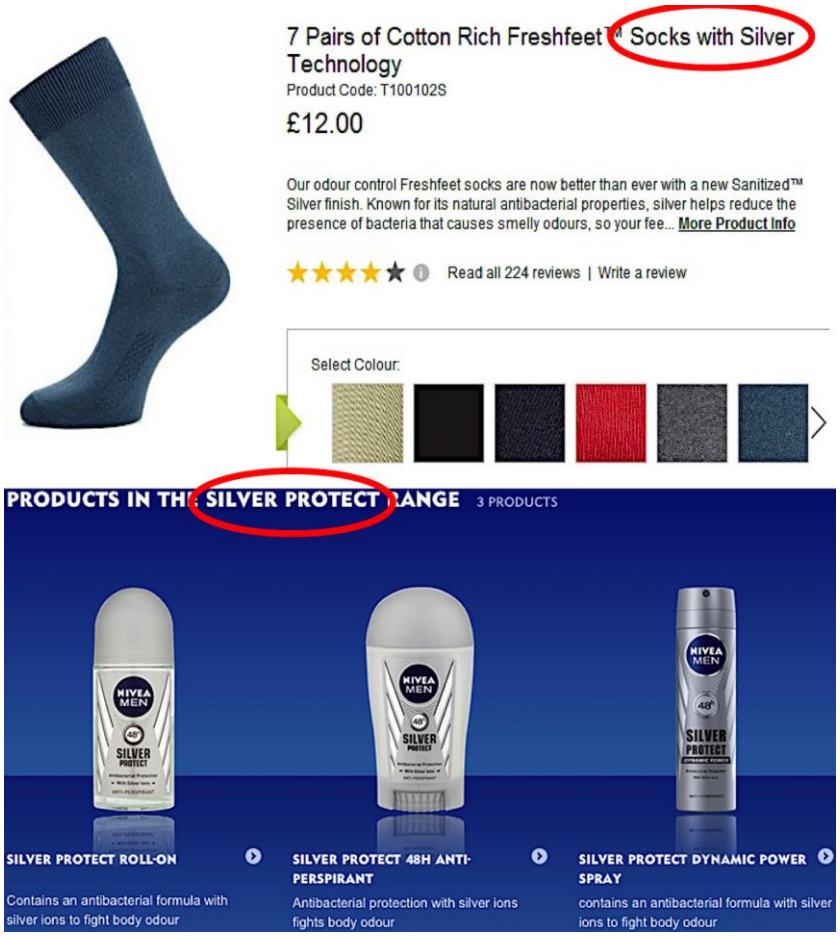


Figure 1. 3 Modern uses of antibacterial silver

Examples of the modern uses of antibacterial 'silver technology' (circled in red) in marketable goods, such as socks and deodorant, to eliminate odours caused by the presence of bacteria. Images taken from Marks and Spencer plc, 2013 and Beiersdorf, 2016.

1.3.2.1 Bactericidal mechanism of silver

Even though Ag^+ has been used as an antimicrobial since antiquity (against both Gram-negative and Gram-positive bacteria as well as fungi and viruses) and its antimicrobial activity has been studied for several decades (Mijnendonckx *et al.*, 2013); the exact antimicrobial mechanism of action of Ag^+ still remains unclear (Marambio-Jones and Hoek, 2010). Ag^+ is often described as 'oligodynamic' (Clement and Jarrett, 1994).

Oligodynamic action is the ability of small amounts of heavy metals to exert a lethal effect on microorganisms. Figure 1.4 illustrates an overview of how Ag^+ exerts its toxicity against microbes through three common proposed mechanisms: (i) the uptake of free Ag^+ subsequently interfering with ATP production and DNA replication, (ii) Ag^+ generation of Reactive Oxygen Species (ROS) and (iii) direct damage to cell membranes.

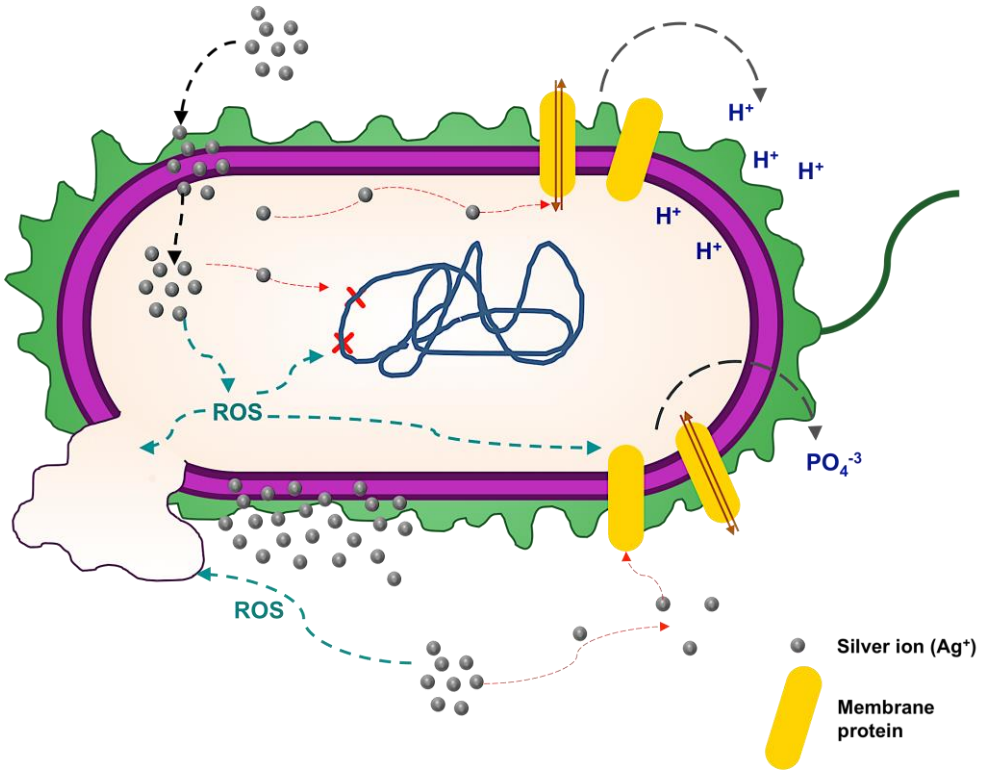


Figure 1. 4 **Antimicrobial activity of silver**

Free silver ions (Ag^+): (i) are taken up; (ii) interact with membrane proteins affect their correct function; (iii) accumulate in the cell membrane affecting membrane permeability and thus respiration (ATP production) and electron transfer; (iv) enter into the cell where they interact with DNA and proteins, disrupt DNA replication and protein function and induce Reactive Oxygen Species (ROS) production. Generated ROS may also affect DNA, cell membrane and membrane proteins. Image adapted and re-drawn with publishers permission, from Marambio-Jones and Hoek, (2010).

After 90 minutes of treatment with Ag^+, reductions of more than 5 $\log^{10}$ CFU mL^{-1} of both *E. coli* and *S. aureus* have been seen by Jung *et al.*, (2008). Consequently, studies found that Ag^+ has the ability to rapidly interact and bind to sulphhydryl (thiol) groups on the surface of microorganisms, resulting in the formation of an S-Ag bond. This bond can interfere with trans-membrane energy generation, (completely blocking respiration and electron transfer), metabolism and ion transport through

interactions with proteins associated to the cell membrane. In turn, a collapse of the proton motive force (i.e. an outflow of potassium ions) is seen to de-energise the membrane and ultimately lead to cell death (Percival *et al.*, 2005).

Disruption of the cell membrane also results in the entrance of Ag^+ in the cytoplasm where it can exert additional damage. Evidence shows that Ag^+ can also form complexes with nucleic acids, by preferential interactions with the nucleosides disrupting base pair hydrogen bonds, thereby interfering with DNA replication (Clement and Jarrett, 1994; Park *et al.*, 2009). Nevertheless, the thiol group-based interactions are thought to be causing the most profound Ag^+ effects. Through Ag^+ interaction with activity-associated key cysteine residues, enzymes and proteins are forced to undergo conformational changes, resulting in their dysfunction (Mijnendonckx *et al.*, 2013). Ag^+ also promotes ROS production, specifically superoxide and hydroxyl radicals (Park *et al.*, 2009; Gordon *et al.*, 2010). These highly toxic, reactive oxidants cause damage to proteins, DNA, RNA and lipids - inhibiting cell growth in numerous ways (Marambio-Jones and Hoek, 2010).

1.4 Silver resistance

Although Ag^+ is an extremely toxic metal to microorganisms, the widespread use of Ag^+ in medical and non-medical settings, such as environments of naturally occurring silver, has resulted in the emergence of Ag^+-resistant bacteria (Percival *et al.*, 2005; Mijnendonckx *et al.*, 2013; Hobman and Crossman, 2015). For example, a Ag^+-resistance plasmid was found in a strain of *Pseudomonas stutzeri* that has been isolated from a silver mine in Utah (Haefeli *et al.*, 1984).

1.4.1 Evolution and prevalence of silver resistance

Initial reports of bacterial resistance to Ag^+ began to emerge from the late 1960s (Jelenko, 1969; Gupta *et al.*, 1999). The first Ag^+-resistance plasmid known to be isolated was pMG101. This 182 kb large plasmid was isolated in 1973 from *Salmonella enterica* serovar Typhimurium following the death of several septicaemia patients treated with silver nitrate (McHugh et al., 1975). Subsequently, the event led to the closure of the burns unit at the Massachusetts General Hospital, Boston USA. Plasmid pMG101 (IncHI incompatibility group) reportedly carries the Ag^+-resistant determinant known as the *sil* operon (which is composed of nine open reading frames (ORFs) and the predicted roles of each protein is explained in detail in Section 1.4.2.2). The plasmid confers additional resistances to: mercuric ions (Hg^{2+}), ampicillin, chloramphenicol, tetracycline, streptomycin, and sulphonamides (McHugh *et al.*, 1975; Gupta *et al.*, 1999).

An *E. coli* K-12 strain carrying pMG101, has been shown to confer Ag^+-resistance in standard Luria-Bertani (LB) broth, without added NaCl, containing up to 600 μM of Ag^+ (as $AgNO_3$). This concentration is over six times that usually tolerated by *E. coli* K-12 strains lacking the plasmid (Gupta et al., 1999).

Since then, Ag^+-resistant bacteria have been reported periodically from a variety of sources including clinical settings (Gupta *et al.*, 2001). A few other examples of Ag^+-resistant bacterial strains that have been isolated include: *E. coli*, *Enterobacter cloacae* (*E. cloacae*), *K. pneumoniae* and *Acinetobacter baumannii* (reviewed in Percival *et al.*, 2005). Like Gupta and colleagues, Kremer and colleagues also found resistance to up to 600 μM of Ag^+ in a pathogenic strain of *E. cloacae*. Subtractive hybridisation using *sil* specific primers identified the presence of the *sil* determinant in 63 % of their *E. cloacae* isolates, thereby displaying high clinical prevalence of Ag^+-resistance (Kremer and Hoffmann, 2012). Ag^+-resistance has also been associated with *E. coli* ST131 and *K. pneumoniae*. The resistance was encoded on plasmid pUUH239.2, which also encodes resistance to various antibiotics and other heavy metals including copper and arsenic. Studies demonstrated transfer of this plasmid between *K. pneumoniae* and *E. coli*. A strain of *K. pneumoniae* containing the plasmid caused an outbreak at Uppsala University hospital in Sweden which went on to infect 248 patients (Sandegren *et al.*, 2012). Finally, Loh *et al.*, (2009) have also identified the presence of *silE* (one of the nine Ag^+-resistance genes) in isolates of another problematic multidrug resistant strain of methicillin-resistant *Staphylococcus aureus* (MRSA). Nevertheless, the prevalence of *sil* genes

was low in the isolates investigated and the presence of the *silE* gene alone did not afford full protection against Ag^+ to the organism.

Even with these sporadic reports of Ag^+-resistance in bacteria, to date, it is the pMG101 *sil* system that has been best characterised in any great detail (Gupta *et al.*, 1999).

1.4.2 Bacterial mechanism of silver tolerance and resistance

The proposed mechanism of Ag^+-resistance through the *sil* system has been described via DNA sequencing and by homologies of the Sil proteins to the bacterial *cop*, *cus* (also named *agr* by Franke *et al.*, (2001)) and *pco* copper resistance systems (Silver, 2003; Hobman and Crossman, 2015). Brief descriptions of the systems are given below.

1.4.2.1 Copper homeostasis and resistance

Copper is a redox active metal that can exist in a reduced Cu^+ cuprous form or Cu^{2+} oxidised form. Involved in donating or accepting electrons in enzymes and the electron transport chain, it is a vital metal to aerobic forms of life. However copper can also be toxic at higher cellular concentrations (as explained by the Irving-Williams series in Section 1.2.1.), particularly for prokaryotes where involvement in phagosomal killing of bacteria by engulfed macrophages has also been recognised recently (Osman *et al.*, 2013; Hao *et al.*, 2015; Hobman and Crossman, 2015; Staehlin *et al.*, 2016). In order to cope with the needs and toxicity of copper, microorganisms have developed homeostasis mechanisms that

maintain the correct compartmentalisation (i.e. not in the cytoplasm) and essentiality of copper (Osman *et al.*, 2013; Staehlin *et al.*, 2016). These systems include *cue, cus*, *pco* and *cop*.

The *cue* and *cus* systems

The *cue* and *cus* systems are chromosomally located copper homeostasis mechanisms, in Gram-negative bacteria (besides *Salmonella*, which only has the *cue* system) involved in charge modification and the efflux of copper (Hobman and Crossman, 2015). The *cue* system consists of a copper-sensing MerR family transcriptional regulator CueR, which activates the expression of: the inner membrane copper P_{1B}-type ATPase effluxer CopA and the periplasmic multi-copper oxidase CueO (predicted to oxidise Cu^{+} to Cu^{2+}). The periplasmic copper-binding CueP, is expressed in *Salmonella* and not in *E. coli* (Franke *et al.*, 2003; Zimmermann *et al.*, 2012; Osman *et al.*, 2013; Hao *et al.*, 2015; Hobman and Crossman, 2015).

Periplasmic CueO is inactive in anaerobic conditions; hence copper resistance in *E. coli* relies on the efflux complex CusCFBA of the *cus* system (Munson *et al.*, 2000; Singh *et al.*, 2011; Zimmermann *et al.*, 2012). CusCBA comprises a tripartite copper efflux pump that spans the periplasm as a proton-substrate anti-porter, expelling copper directly from the cytoplasm to the extracellular environment. CusF is a periplasmic metallochaperone believed to deliver the copper to CusB (Franke *et al.*, 2001; Franke *et al.*, 2003). The *cusCFBA* operon is transcriptionally regulated by the two-component sensor system CusRS (Munson *et al.*, 2000). CusS senses the periplasmic copper concentration, phosphorylating

CusR, which activates transcription of the cusCFBA operon. CusSR can also be induced by silver ions (Franke *et al.*, 2001).

The *pco* system

In addition to the *cue* and *cus* systems, some *E. coli* strains carry a plasmid encoded *pco* system that provides additional copper resistance. The *pco* system was originally found on plasmid pRJ1004, isolated from pigs fed on copper sulphate supplemented feed (Tetaz and Luke, 1983; Brown *et al.*, 1995). The *pco* system in Figure 1.5 is encoded by nine (or occasionally ten) ORFs, arranged as two operons and a separate gene (Zimmermann *et al.*, 2012; Hao *et al.*, 2015). The transcription of *pco* genes *pcoABCDEFG* appears to be regulated by PcoRS, a two-component regulator system homologous to CopRS. The roles of PcoF and PcoG have not been elucidated. In addition to oxidised catechol siderophores, copper may be detoxified from Cu^{+} to Cu^{2+} by the CueO homologue, the multicopper oxidase, PcoA. PcoB possibly functions as the outer membrane transporter, while PcoD sitting in the inner membrane drives the transport of Cu^{+} from the periplasm to the cytoplasm, with periplasmic PcoC chaperoning/delivering the Cu^{+} to PcoD. PcoE is an additional chaperone, co-regulated by CusRS, that may act as a first line of defence 'molecular sponge' protein by binding copper in the periplasm possibly also shuttling it to PcoA (Zimmermann *et al.*, 2012). Homologues of components of the *pco* system have been identified in the *cop* system found in plant pathogens (Hobman and Crossman, 2015).

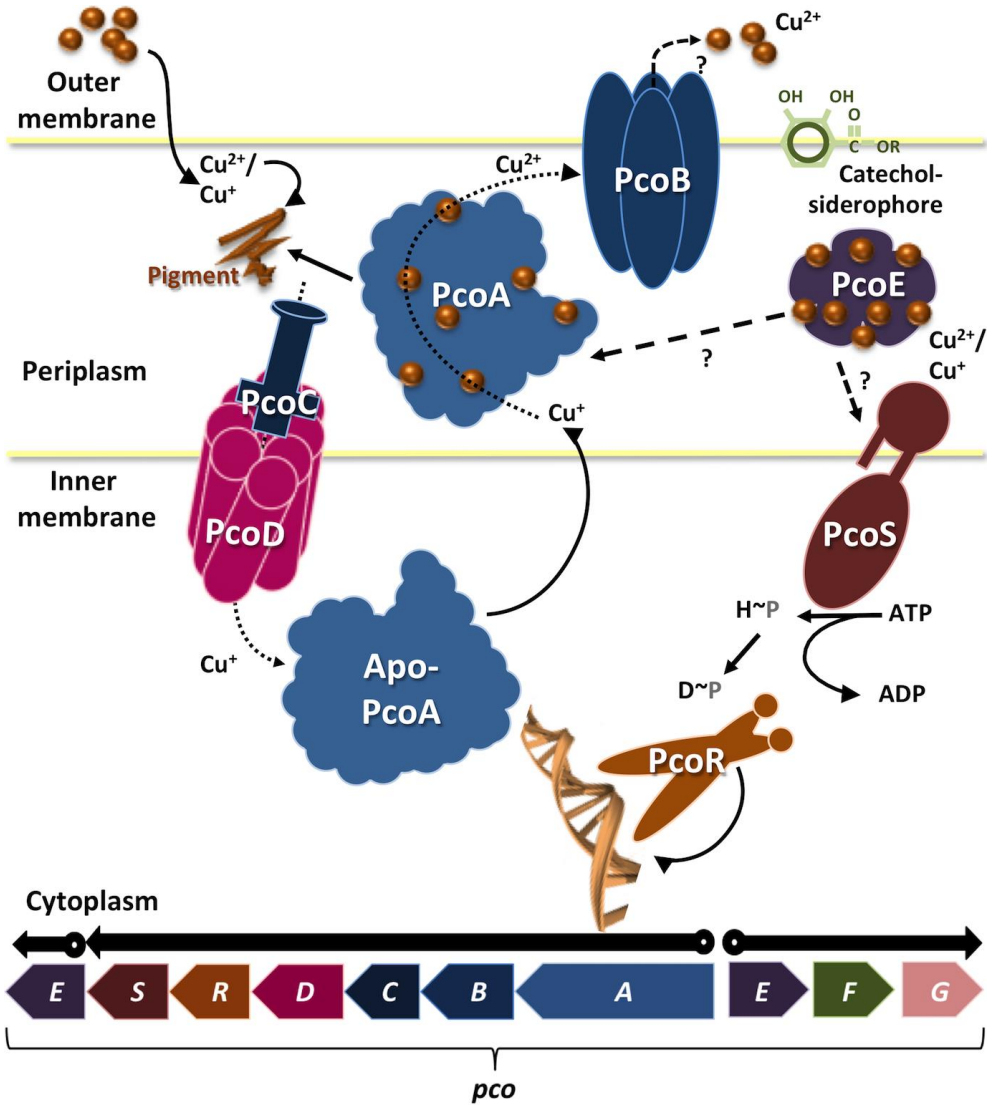


Figure 1. 5 **Copper resistance *pco* genes and encoded protein functions**

The mechanism by which the Pco system, encoded by *pcoABCDRSEGF* (bottom line), contributes to copper resistance is not fully understood. However, it is proposed to involve several periplasmic components thought to be involved in copper detoxification. Once copper has entered the cell, it is believed to be sequestered on oxidised catechol siderophores or Cu^{+} may be oxidised to the less toxic Cu^{2+} by the copper oxidase PcoA. PcoC and PcoD may transport copper across the cytoplasmic membrane and PcoE bind to copper in the periplasm, where all three are involved in the shuttling of copper to PcoA. PcoB is a probable outer membrane transporter. PcoR and PcoS appear to regulate the expression of *pcoABCD*. The exact functions of PcoG and PcoF are still unknown. This diagram has been reproduced and used with permissions from the publisher, from Hao *et al.*, (2015).

The *cop* system

Homologous copper-resistance determinants *copABCDRS* have been found in: *Pseudomonas syringae* pathovar *tomato* PT23 isolated from plants exposed to high levels of copper, *Cupriavidus metallidurans* CH34 isolated from sediments of zinc decantation basins in Belgium, *Xanthomonas axonopodis* pv. *vesicatoria* E3C5 and *Pseudomonas aeruginosa* PAO1 (Monchy *et al.*, 2006; Zimmermann *et al.*, 2012). The *copG* and *copF* genes have also been identified by Monchy *et al*., (2006) with putative descriptions of CopG containing a cysteine-rich motif, CXCC and CopF being a putative P1-type ATPase. The operon structure of the *cop* gene cluster and protein sequences are similar to the *pco* gene counterparts, suggesting a conserved mechanism of copper resistance (Zimmermann *et al.*, 2012), as explained previously.

Nevertheless, Gram-positive bacteria are known to use a different (but confusingly, similarly named) *cop* system, described by Hobman and Crossman in a recent review (Hobman and Crossman, 2015). In *Enterococcus hirae* the mechanism consists of the *copYZAB* operon: a different CopA ATPase imports copper into the cytoplasm, CopZ binds to excess copper and chaperones it to either the copper export ATPase CopB or to CopY – the copper responsive repressor for the operon itself. In *Bascillus subtilis* and *Staphylococcus*, the cytosolic copper sensor CsoR triggers the export of copper by the P_{1B}-type ATPase CopA with assistance from the copper-binding chaperone CopZ (Osman and Cavet, 2008; Hobman and Crossman, 2015). However this model has recently come into question, as ATPase pumps such as CopA are now only considered as efflux

pumps and not import pumps that export copper ions from the cytoplasm to the periplasm (Gupta and Lutsenko, 2012).

1.4.2.2 The *sil* system

The *sil* operon, characterised as the Ag^+-resistance conferring unit in plasmid pMG101, consists of a group of nine genes: *sil-PGABFCRSE* (Randall *et al.*, 2015; Hao *et al.*, 2015). These genes are divided into three transcriptional units: *silCFBAGP*, *silRS* and *silE* - each of which is expressed from a different promoter (Gupta *et al.*, 1999; Silver, 2003). All nine genes transcribe and translate into functional proteins for which the assigned putative roles are shown in Figure 1.6 (Mijnendonckx *et al.*, 2013).

SilP is a proposed cation P_{IB}-type ATPase efflux pump that sits within the inner membrane and cytoplasm and transports Ag^+ from the cell cytoplasm to the periplasm. Type IB P-type ATPases have been found to remove toxic metals from bacteria including silver, copper, zinc, iron and cadmium cations (Osman *et al.*, 2013). SilP differs from similar heavy metal P-type ATPases, primarily through a lack of cysteine residues replaced by His_5-Asp-His_2 in its N-terminal region. These histidine residues could act as potential Ag^+ ligands (Gupta *et al.*, 1999).

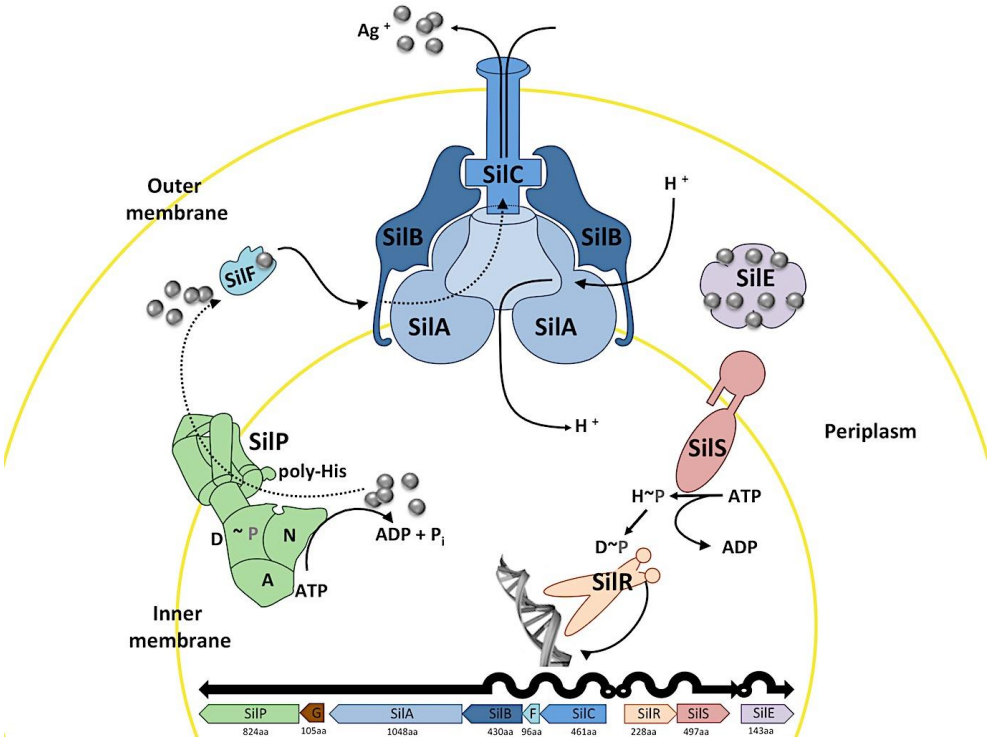


Figure 1. 6 Silver resistance gene cluster and encoded protein functions

Overview of the proposed pMG101 bacterial Ag^+-resistance *sil* system. The bottom line shows the different genes, the direction of their ORFs and the three different transcriptional start sites. Above this is an overview of the predicted function of each of the genes based on homology studies. SilP is a P-type ATPase pump that transports Ag^+ from the cytoplasm to the periplasm. SilCBA complex is a cation/proton antiporter, which transports Ag^+ from both the cytoplasm and periplasm out of the cell. SilF transports silver in the periplasm from SilP to the SilCBA complex. SilG is also a proposed periplasmic Ag^+-chaperone. SilS is a histidine kinase that phosphorylates SilR in the presence of Ag^+. Once phosphorylated SilR acts as a transcription factor for SilP, G, A, B, F, and C. SilE is a periplasmic metal binding protein controlled by its own promoter and is strongly expressed in the presence of Ag^+. It is thought to act as a first line of defence to Ag^+ whilst the other proteins are synthesised (Mijnendonckx *et al.*, 2013; Randall *et al.*, 2015). Image adapted and redrawn with publishers permission, from Silver, (2003).

Recently, Randall and colleagues showed ~45 % sequence identity between SilG and a putative copper chaperone CopG, from the *cop* system of *Cupriavidus metallidurans*. Therefore, whilst also containing the conserved metal-binding CxxC motif, SilG may serve as a periplasmic Ag^+-chaperone (Randall *et al.*, 2015). SilCBA also share similarities with *E. coli cus* copper-resistance proteins. Gupta *et al.*, (1999) suggested the role of SilCFBA upon identifying that SilC, SilB and SilA displayed 71 %, 67 % and 87 % identity to *E. coli cus* system proteins CusC, CusB and CusA, respectively. In complex, SilC, SilB and SilA are predicted to form a three-protein, entire cell membrane spanning, Ag^+/H^+ antiporter system that belongs to the heavy metal efflux-resistance nodulation cell division (HME-RND) family of efflux pumps (Figure 1.2). SilC is the outer membrane protein (also entering the periplasm) and SilA - the cation antiporter - is found intruding into the inner membrane, with the membrane fusion protein SilB spanning the periplasm and stabilising the complex. Collectively, these proteins form the efflux pump that pumps Ag^+ from the periplasm to the exterior of the cell (Mijnendonckx *et al.*, 2013). SilF is a periplasmic protein that possibly functions as a chaperone carrying Ag^+ from SilP to the SilCBA complex, in the same manner that CusF transfers Cu^+ to CusB for efflux (Osman *et al.*, 2013).

SilS and SilR (homologous to CusRS and PcoRS) are believed to transcriptionally regulate the *silCFBAGP* gene unit. Together these form a two-component sensor/regulator; in which SilS is a transmembrane histidine kinase sensor and cytoplasmic SilR is the transcriptional regulator responder (Gupta *et al.*, 1999). The two proteins are expressed via divergent

transcription to SilCFBAP, as is periplasmic protein SilE – located downstream of *silRS*. The *silE* gene is under the control of its own promoter and its expression is strongly induced in the presence of Ag^+. It has been suggested that the expression of *silE* is co-regulated by CusRS (Munson *et al.*, 2000). It codes for the 143-amino acid long periplasmic protein SilE whose precise role in Ag^+- resistance has not been experimentally confirmed (Silver, 2003; Mijnendonckx *et al.*, 2013).

To address whether all of the Sil proteins are required to mediate Ag^+-resistance, Randall *et al.*, (2015) deleted each of the *sil* genes individually. Whilst deletion of *silC, silB, silA* or *silE* led to a complete loss of Ag^+-resistance; deletion of *silF*, *silP* or *silG* showed none or limited effects on the Ag^+-resistance phenotype. The deficiency of any significant effects on the Ag^+-resistance phenotype has been considered an artifact of the Cus proteins potentially acting to complement the loss of Sil proteins, thereby establishing cross-resistance.

1.4.3 Cross-resistance to silver

Cross-resistance is a form of multi-drug resistance that can be defined as the development of a tolerance mechanism to usually toxic agents as a result of exposure to a related antimicrobial. Chromosomally located homologues of the *sil* system have also been found in *E. coli*. Although the *cus* system is mainly involved in copper resistance, it is also known to confer resistance to low levels of Ag^+. Studies by Franke *et al.*, (2001) showed that deletion of the SilA homologue CusA resulted in loss of

Ag^+-resistance. Likewise, investigations upon deletion of CusF have resulted in a decrease of the minimum inhibitory concentration of Ag^+ from more than 1 mM, in M9 minimal media, to 12 µM (Lok *et al.*, 2008). Equally suggestive of cross-resistance, the expression of Cus proteins can be induced by Ag^+, albeit to a much lesser extent than Cu^+ (Franke *et al.*, 2001). There is also evidence that Ag^+ cross-resistance is aided by decreased outer membrane permeability mediated by a deficiency in outer membrane porins (Figure 1.2) (Li *et al.*, 1997; Randall *et al.*, 2015). Li *et al.*, (1997) showed that the absence of porins also contributed to low-levels of cross-resistance to Ag^+ and structurally different antibiotics.

1.4.4 Co-resistance, mobile genetic elements and co-selection

Co-resistance refers to the presence of resistance to more than one antimicrobial. Co-resistance is likely when genes involved in different mechanisms of antimicrobial resistance are located together on 'mobile' genetic elements such as conjugative plasmids such as pMG101, transposons and integrons. Silver-resistance determinants are present in many bacteria and frequently located on mobile genetic elements, alongside numerous other heavy metal and antibiotic resistance determinants that can be horizontally transferred. Early studies carried out on the *sil* resistance genes in IncH incompatibility group plasmids found the *sil* gene locus to be highly conserved in several plasmids including: MIP233, pWR23 and MIP235 from *Salmonella enterica* serovars as well as R476b and R478 from *Serratia marcescens* (Gupta *et al.*, 2001; Kremer and Hoffmann, 2012). Kremer and Hoffmann, (2012) also identified the presence of an

IncHI-2 plasmid as the major difference between virulent and avirulent *Enterobacter cloacae*. Besides resistance to Ag^+, the plasmid in the pathogenic strains encoded resistance to several antibiotics. This association can facilitate the spread of resistances and co-selection of antimicrobial resistances.

Additionally, the *pco* operon (a cluster of seven genes – *pcoABCDRSE* on the large *E. coli* copper resistance plasmid pRJ1004 (Tetaz and Luke, 1983; Brown *et al.*, 1995; Munson *et al.*, 2000)) and *sil* operon have often been found together in a single locus of identical arrangement in plasmids and on the chromosomes of many Gram-negative bacteria (Hao *et al.*, 2015; Hobman and Crossman, 2015; Randall *et al.*, 2015; Staehlin *et al.*, 2016). Recently, Randall *et al.*, (2015) sequenced the whole genome of *E. coli* J53 (pMG101) and found that the *sil* and *pco* locus had integrated into the chromosome from pMG101. They reported the presence of Tn*7*-like elements directly upstream of the *sil* operon thus concluded that the integration may have been mediated by Tn*7*-based transposition (Randall *et al.*, 2015).

A comprehensive review on Tn*7*-based transposition has been written by Peters and Craig, (2001), in which they define a transposon as a mobile DNA segment that can translocate to non-homologus target sites within a genome, using a self-encoded recombinase. Tn*7*-like bacterial transposons have been implicated in the carriage of many genes relevant to pathogenesis, including antibiotic resistance and metal resistance systems (McKown *et al.*, 1988; Peters and Craig, 2001; Peters *et al.*, 2014). Tn*7* uses distinct but partially overlapping combinations of proteins (including

TnsABCDE with an atypical heteromeric transposase) to identify these alternative target sites. In one pathway, proteins TnsABC and TnsE proteins encourage horizontal tranmission by targetting conjugal plasmids that can move promiscuously between bacteria. In a second transposition pathway, the TnsABC and TnsD proteins direct insertion, at a high frquency, into a single specific site known as *attTn7* within the essential and highly conserved gene (*glmS*) in bacterial chromosomes (Waddell and Craig, 1989; Peters and Craig, 2001; Peters *et al.*, 2014). This is carried out without any disturbance to the *attTn7* site. Likewise, Tn*7*-like elements can form genomic islands through programmed insertion at their preferred *attTn7* attachment site and spread the associated genotype through horizontal transfer back onto mobile plasmids (Peters *et al.*, 2014).

1.4.5 Periplasmic silver-binding SilE

SilE is an indispensable key component for the exogenous Ag^+-resistance phenotype (Randall *et al.*, 2015). It has been reported to bind between five and 38 Ag^+ ions, depending on experimental conditions (Silver *et al.*, 1999; Mijnendonckx *et al.*, 2013), and is often used as a marker when confirming the presence of silver resistance genes in microbes (Mirolo *et al.*, 2013).

SilE exhibits 48% sequence identity to the periplasmic copper-binding protein PcoE (Section 1.4.2.1 and Figure 1.5) (Zimmermann *et al.*, 2012), which was originally identified by Brown *et al.*, (1995). PcoE binds up to nine Cu^+ or up to seven Ag^+ ions (Zimmermann *et al.*, 2012).

Expression of *pcoE* is rapidly controlled by the chromosomally located copper resistance *cusRS* system and can therefore act as a first line of defence while expression of the other resistance proteins proceeds (Munson *et al.*, 2000; Zimmermann *et al.*, 2012). PcoE is hypothesised to be unstructured in its *apo*-form but folds and dimerises upon Cu^{+} binding, with some α-helical content in its secondary structure (Zimmermann *et al.*, 2012). Because of its similarity to PcoE, SilE is presumed to have similar attributes as well as possess the ability to bind copper ions (Gupta *et al.*, 1999; Silver *et al.*, 1999; Silver, 2003; Zimmermann *et al.*, 2012) although currently there is no experimental data available to verify this and determine its precise function.

Unfolded and unstructured proteins that adopt a fixed three-dimensional (3D) structure after binding to a binding-partner, as PcoE and SilE are proposed to, are known to be functionally important - particularly in regulation. These types of proteins have been classified as intrinsically disordered or unstructured proteins (IDPs/IUPs) (Dyson and Wright, 2005).

1.4.5.1 Intrinsically disordered proteins

The traditional 'structure-function' paradigm dictates that a polypeptide chain requires a compact, well-defined, globular structure in order to function under physiological conditions (Wright and Dyson, 1999). However, this paradigm has been challenged by reports in the 1990s of another class of proteins and/or regions of proteins that are known to function without a well-defined three-dimensional (3D) structure under

biological conditions (Uversky, 2002a). More prevalent in eukaryotes, these proteins with intrinsically disordered regions (IDRs) are called intrinsically disordered proteins (IDPs). IDPs are also called natively unfolded (Uversky, 2002a), natively denatured (Dunker *et al.*, 2001), and intrinsically unstructured proteins (Wright and Dyson, 1999). These proteins or protein regions/domains exist as dynamic ensembles of rapidly inter-converting heterogeneous conformations rather than the relatively stable and homogenous conformations typical of folded proteins. Compared to globular proteins, IDPs have low hydrophobicity (they contain only a few Cys, Ile, Leu, Met, Phe, Trp, Tyr and Val residues) and are substantially enriched in disorder promoting polar and/or charged residues (such as Gln, Glu, Lys and Ser), as well as structure breaking residues (Pro and Gly) (Radivojac *et al.*, 2007; He *et al.*, 2009; Fontana *et al.*, 2012; Uversky, 2013a; Uversky, 2013b; Lee *et al.*, 2014). Consequently, IDPs lack the hydrophobic core that would enable them to spontaneously fold into globular proteins.

IDPs can adopt multiple distinct, yet well-defined conformations upon specific interactions with partner molecules, i.e. a coupled-folding-and-binding process when binding to other protein, DNA/RNA or metal ligands (Breydo and Uversky, 2011). These unique structural features, including accessibility and space efficiency allow for the facilitation of a large number of specific yet transient interactions; thus serving as hubs of protein-protein/molecule interaction networks (Sickmeier *et al.*, 2007; Uversky and Dunker, 2010; Receveur-Brechot and Durand, 2012; Uversky, 2013a). They are known to participate in crucial functions such as signal

transduction, transcription regulation, cell-cycle regulation as well as scavenge small ligands (Kosol, Contreras-Martos, *et al.*, 2013). Given the functional importance of IDPs, many proteins containing disordered regions have been associated with various diseases such as diabetes, amyloidosis, neurodegenerative diseases, cardiovascular disease and cancer (He *et al.*, 2009; Breydo and Uversky, 2011; Sibille and Bernadó, 2012; Uversky, 2013a; Faller *et al.*, 2014). Due to their proposed role in disease, particularly neurodegenerative, the interaction of IDPs with periodic table d-block metals has gained interest (Faller *et al.*, 2014). Some examples of neurodegenerative metal-binding IDPs include: Aβ, Tau, Prion protein and α-Synuclein (Breydo and Uversky, 2011).

Metal-binding IDPs have also been observed in prokaryotes. Protein kinase G (PknG), from *Mycobacterium tuberculosis*, is secreted into host macrophages to promote cellular survival of pathogenic *mycobacteria* by inhibiting lysosomal degradation. PknG contains a disordered N-terminal, rubredoxin-like metal-binding motif that appears to fold in its metal-bound (Zn^{2+}) state; to mediate PknG redox regulation (Wittwer and Dames, 2015). CusB of the prokaryotic CusCFBA complex from *E. coli* has also been described as a metal-binding protein consisting of a disordered N-terminal domain. Local ordering of the domain was observed upon binding to its metal-ligand, Cu^{+} (Ucisik *et al.*, 2013).

Accordingly it is possible that, upon binding Ag^{+}, SilE may fold to form a protein structure that has an increased affinity for the histidine kinase sensor SilS; consequently binding or signalling to SilS in order to activate the remainder of the silver resistance machinery. However, an alternative

theory for the role of SilE is as a Ag^+ scavenger and chaperone for the SilCFBA complex in Ag^+-resistance has been suggested by Silver, (2003) and Mirolo *et al.*, (2013), based on its potential metal-binding sites.

1.4.5.2 Potential metal-ligand binding sites

Both PcoE and SilE have a twenty amino acid leader sequence that results in their translocation to the periplasm and is followed by subsequent removal of the sequence. They also both contain a number of methionine and histidine residues, which are potential metal binding residues. Their importance is signified through sequence conservation between PcoE and SilE; >50% similarities for methionine and 100% identity for the ten histidines (Zimmermann *et al.*, 2012). The ten histidine residues are spatially conserved (Mirolo *et al.*, 2013) and following experimental and computational data, these residues have been proposed to be primary candidates for metal binding (Silver, 2003; Mirolo *et al.*, 2013).

In order to determine a potential model for SilE Ag^+-binding an unpublished study, reported of in a review, used nuclear magnetic resonance (NMR) spectroscopy and circular dichroism (CD) to analyse the binding of Ag^+ by histidines at different pH values (Silver, 2003). These experiments, together with histidine-Ag^+ complex crystal structure investigations by Mirolo *et al.*, (2013), showed that histidine binds most strongly to Ag^+ at neutral pH through its side chain imidazole ring nitrogen atoms in a linear fashion. These findings have led to a proposed mechanism (Figure 1.7) in which SilE binds five Ag^+ ions via its ten histidine residues at neutral pH. Further to this, software has been used to predict the proteins secondary

structure and which histidine residues crosslink via binding of the same Ag^+ ion (Silver, 2003). It is thought that four HX6H motifs form (each within a different helix) and bind to each other across space. The binding is not thought to occur within one of these motifs as binding between two histidines within two turns of each other would disturb the helical structure. Finally, following a change in environmental pH, protonation of these histidine residues could also partake in the release of Ag^+ into the periplasmic space with the SilCBA efflux pump ejecting the toxic monovalent metal ion, out of the cell (Silver, 2003; Mirolo *et al.*, 2013). This, however, contradicts other data showing increased binding of Ag^+ ions to SilE under acidic conditions (Silver *et al.*, 1999).

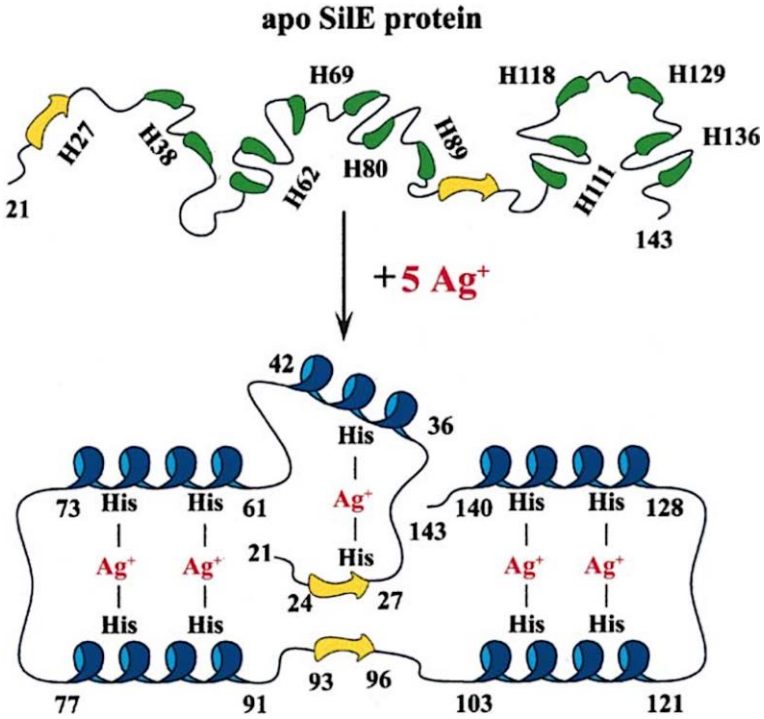


Figure 1. 7 **Model for Ag^+ binding and folding of the periplasmic Ag^+ binding protein SilE**

Top: 122-amino acid processed SilE protein after removal of 20-amino acid leader sequence (Gupta *et al.*, 1999) with positions of the 10 histidine residues noted. Bottom: secondary structure predictions of α-helical (coils) and β-sheet (arrows) regions from standard software and predicted cross-linking of five Ag(I) cations by 10 histidines. Figure used following publishers permission, from (Silver, 2003).

1.5 Aims and objectives

The introduction has discussed that bacterial Ag^+-resistance has been found within clinical and environmental settings and highlighted its prevalence. However, recent studies also suggest that full resistance to Ag^+ is still not extensive in medicine. In the aforementioned study on Ag^+-resistant MRSA, although the *silE* gene was found in two isolates, it did not actually give full resistance to Ag^+ and the bacteria were eventually completely killed by the Ag^+ impregnated bandages (Loh *et al.*, 2009). Accordingly, compared with antibiotic resistance, Ag^+-resistance is still much less widespread clinically. Yet the *sil-pco* island is becoming increasingly found in Ag^+ and Cu^+/Cu^{2+}-rich sites, such as agriculture, mines and sewage systems where it has been found in several bacterial species, inclusive of facultative anaerobes and emergent enteric pathogens too – i.e. *E. coli*, *E. cloacae*, *K. pneumoniae*, *S.* Typhimurium, *Serratia marcescens*, *Cronobacter turicensis* and *Shewanella putrefaciens* (McArthur *et al.*, 2013; Hao *et al.*, 2015; Staehlin *et al.*, 2016).

As antimicrobials are essential for the treatment of bacterial infections it is a high priority to preserve their efficacy. Therefore, it would be naive to dismiss Ag^+-resistance as something that is unlikely to become a problem, especially with the increasing use of Ag and Cu in both agriculture and commercial products. For this reason, understanding the molecular mechanism for how Ag^+-resistance works is of great importance. In order to do this, we firstly sought to interpret the genetics of the system by aiming to characterise plasmid pMG101, inclusive of the silver resistance determinant – the *sil* operon. This work is described in Chapter 3.

Secondly, few structural and biophysical studies have been carried out on the mechanism thus far. The C-terminal of SilB, from *Cupriavidus metallidurans* CH34 (39 % sequence similarity to SilB from pMG101), is the only part of the Sil system to be structurally characterised (Bersch *et al.*, 2011). Within this system, SilE has been noted as a key component. Like its copper-binding homologue SilE is proposed to be unstructured and unfolded in its *apo*-form. Models of Ag^+-binding to conserved histidine residues and Ag^+-mediated SilE folding have been suggested, but without experimental evidence. Therefore, additional objectives were to test these theories and provide further evidence on SilEs role in the mechanism of Ag^+-resistance. The additional objectives included:

1. Determining whether SilE is an IDP by native protein characterisation studies.
2. Assessing the ability of SilE to bind potential metal-ligands and its specificity towards Ag^+ through various binding and folding studies.
3. Identifying potential Ag^+-binding sites within SilE that may be involved in binding-induced folding of the protein via site-directed and truncation mutagenesis studies.

Result Chapters 4 to 6 describe the structural and biophysical approaches used to investigate these.

Note: $SilE_{21\text{-}143F}$ was the primary SilE protein that was investigated throughout these studies; therefore it will only be referred to as SilE throughout the remainder of this thesis (unless a reminder of the shortened sequence or use of construct is required).

CHAPTER 2

Materials and Methods

2.1 Materials

The majority of the chemicals and reagents (analytical grade) used were obtained from Sigma Aldrich (Poole, Dorset, UK), unless specified otherwise. Syringe filters used throughout had a pore size of 0.2 μm.

2.1.1 *E. coli* strains

To aid in the analysis of the DNA sequences and biology associated with the silver resistance *sil* operon, numerous *E. coli* strains were used as positive and negative controls. These can be found in Table 2.1. Additionally, to facilitate the overexpression of the Sil proteins and in order to clone the *sil* genes, *E. coli* OmniMAX^{TM} 2 -T1^{R} and XL1-Blue cells (Invitrogen, Paisley, UK) (Table 2.1) were used for electroporation and chemical transformation to propagate the plasmid DNA (deoxyribonucleic acid) products. For the overexpression trials of the Sil proteins, SilE, SilS, SilR, SilC, SilB, SilA and SilP *E. coli* B834 (DE3), Rosetta (DE3) pLysS (Novagen - Merck Millipore, Dermstadt, Germany), OverExpress^{TM} C41 (DE3) and OverExpressTM C43 (DE3) (Lucigen - Middleton, USA) cells were used. Finally, to overexpress the SilE proteins at a larger scale, *E. coli* Rosetta Star^{TM} (DE3) pLysS and *E. coli* BL21 Star^{TM} (DE3) pLysS cells (Invitrogen, Paisley, UK) were used (Table 2.1).

Table 2. 1: ***E. coli*** **strains used in this study**

A table listing the various *E. coli* strains that were used during the sequencing, cloning and protein overexpression projects, in addition to their defining characteristics, an original reference or source and their experimental use. Information on the different strains of cloning and overexpression cells used was located from:

http://www.invitrogen.com/site/us/en/home/Products-and-Services/Applications/Cloning/competent-cells-for-transformation/competent-cells-resources/genotypes-of-competent-cells.html,
http://www.med.unc.edu/pharm/sondeklab/Lab%20Resources/manuals/novagen_competent_cells2.pdf and
http://lucigen.com/store/docs/manuals/MA031-OverExpress-Chemicallycompetent-Cells.pdf.

*Two examples of J53 (pMG101) were used – (i) Professor Simon Silver, University of Illinois College of Medicine at Chicago, Chicago, Illinois, USA and (ii) Professor Anne Summers, University of Georgia, Athens, Georgia, USA.

Strain	Characteristics and/or Genotype	Use	Reference or Original Source
J53 (pMG101)	J53 strain containing the silver resistance plasmid pMG101. F^+ *met pro (*Amp^R Cam^R Mer^R Sil^R $Strep^R$ Sul^R Tet^R*).*	Sequenced plasmid whole genome and pMG101*, carrying positive control for silver resistance gene profiling and Sil protein overexpression plasmid construction PCRs.	(Gupta *et al.*, 1998)
J53	Derivative of K-12 strain. Used as a recipient for plasmids. F^+ *met pro*.	Non-plasmid carrying negative control for MICs and silver resistance gene profiling PCRs.	(Clowes and Rowley, 1954; Yi *et al.*, 2012)

Strain	Characteristics and/or Genotype	Use	Reference or Original Source
MG1655	A K-12 derivative strain, which approximates wild-type *E. coli*. F⁻ λ⁻ *ilvG⁻ rfb-50 rph-1*.	Negative control for silver resistance gene profiling PCRs.	(Blattner *et al.*, 1997)
***E.coli* OmniMAX™ 2 - T1^{R}**	*F´ {proAB+ lacI^{q} lacZΔM15 Tn10(Tet^{R}) Δ(ccdAB)} mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 Δ(lacZYA-argF) U169 endA1 recA1 supE44 thi-1 gyrA96 relA1 tonA panD.*	*sil* genes cloning for protein overexpression.	Invitrogen™
***E.coli* XL1-Blue**	*recA1 enA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F' proAB lacI^{q}ZΔM15 Tn10 (Tet^{R})].*	SilE mutation cloning work.	Invitrogen™
***E. coli* B834 (DE3)**	*F⁻ ompT hsdS$_{B}$(r$_{B}$⁻ m$_{B}$⁻) gal dcm met (DE3).*	Sil protein overexpression screens.	Novagen®
***E. coli* Rosetta (DE3) pLysS**	*F⁻ ompT hsdS$_{B}$(r$_{B}$⁻ m$_{B}$⁻) gal dcm (DE3) pLysSRARE^{2} (Cam^{R}).*	Sil protein overexpression screens.	Novagen®
***E. coli* OverExpress™ C41 (DE3)**	*F⁻ ompT hsdS$_{B}$(r$_{B}$⁻ m$_{B}$⁻) gal dcm (DE3).*	Sil protein overexpression screens.	Lucigen®

Strain	Characteristics and/or Genotype	Use	Reference or Original Source
***E. coli* OverExpress™ C43 (DE3)**	*F^- ompT hsdS$_B$(r_B^- m_B^-) gal dcm (DE3).*	Sil protein overexpression screens.	Lucigen®
***E. coli* Rosetta Star™ (DE3) pLysS**	*F^- ompT hsdS$_B$(r_B^-, m_B^-) gal dcm rne131 (DE3) pLysS (Cam^R).*	Large-scale protein overexpression.	Invitrogen™
***E. coli* BL21 Star™ (DE3)**	*F^- ompT hsdS$_B$(r_B^-, m_B^-) gal dcm rne131 (DE3).*	Large-scale protein overexpression.	Invitrogen™

2.1.2 Overexpression systems, cloning and plasmid vectors

E. coli is the dominant expression host in delivering soluble proteins for structural biology. Many years of study has allowed for a good understanding of the genetics and molecular biology of *E. coli* and it is therefore, relatively simple to produce heterologous protein (Bird, 2011). Nevertheless, recombinant protein expression levels and solubility in *E. coli* are unpredictable. One aspect of expression optimisation can be through the addition of a fusion protein tag that can improve the expression and/or solubility of a protein. Thus, a series of pOPIN expression vectors (OPPF-UK, Oxfordshire) (Table 2.2), based on and derived from the pTriEx plasmid (Berrow *et al.*, 2007), were used as protein expression host vectors. All contained His_6-tags (with the exception of that from pOPINI) which are proteolytically cleavable from the protein of interest, either with Human rhino virus (HRV)-3C protease, SUMO protease or Carboxypeptidase A. Figure 2.1 shows a map of the pOPINF vector, as an example of the structures and key features of the pOPIN vectors.

Table 2. 2: Details of pOPIN vectors used for cloning Sil proteins

A detailed overview of the seven vectors used for the cloning of the seven sil genes (*silE, silS, silR, silC, silB, silA* and *silP*). All of the pOPIN vectors contain His-tags, which were (with the exception of that from pOPINI) cleavable from the protein of interest - using Human rhinovirus (HRV) -3C protease, where ❶ represents the HRV-3C protease cleavage site.

Information displayed after obtaining the generous provision from OPPF-UK, 2013.

Vector	Fusion Tag	Parent Vector/ Antibiotic resistance	Forward Primer Extension	Reverse Primer Extension
pOPINE	POI…KHHHHHHSSGLEVLFQ❶GP	pTriEx2/ Ampicillin	AGGAGATATACCATG	GTGATGGTGATGTTT
pOPINE-3C-eGFP	POI…SSGLEVLFQ❶GP-GFP-HHHHHH	pTriEx2/ Ampicillin	AGGAGATATACCATG	CAGAACTTCCAGTTT
pOPINF	MAHHHHHHSSGLEVLFQ❶GP	pTriEx2/ Ampicillin	AAGTTCTGTTTCAGGGCCCG	ATGGTCTAGAAAGCTTTA
pOPINJ	MAHHHHHHSSG-*GST*-LEVLFQ❶GP	pTriEx2/ Ampicillin	AAGTTCTGTTTCAGGGCCCG	ATGGTCTAGAAAGCTTTA
pOPINI	MAHHHHHHSSG	pTriEx2/ Ampicillin	ATCATCACAGCAGCGGC	ATGGTCTAGAAAGCTTTA
pOPINM	MAHHHHHHSSG-*MBP*-LEVLFQ❶GP	pTriEx2/ Ampicillin	AAGTTCTGTTTCAGGGCCCG	ATGGTCTAGAAAGCTTTA
pOPINS3C	MAHHHHHH-SUMO-SSGLEVLQF❶GP	pTriEx2/ Ampicillin	AAGTTCTGTTTCAGGGCCCG	ATGGTCTAGAAAGCTTTA

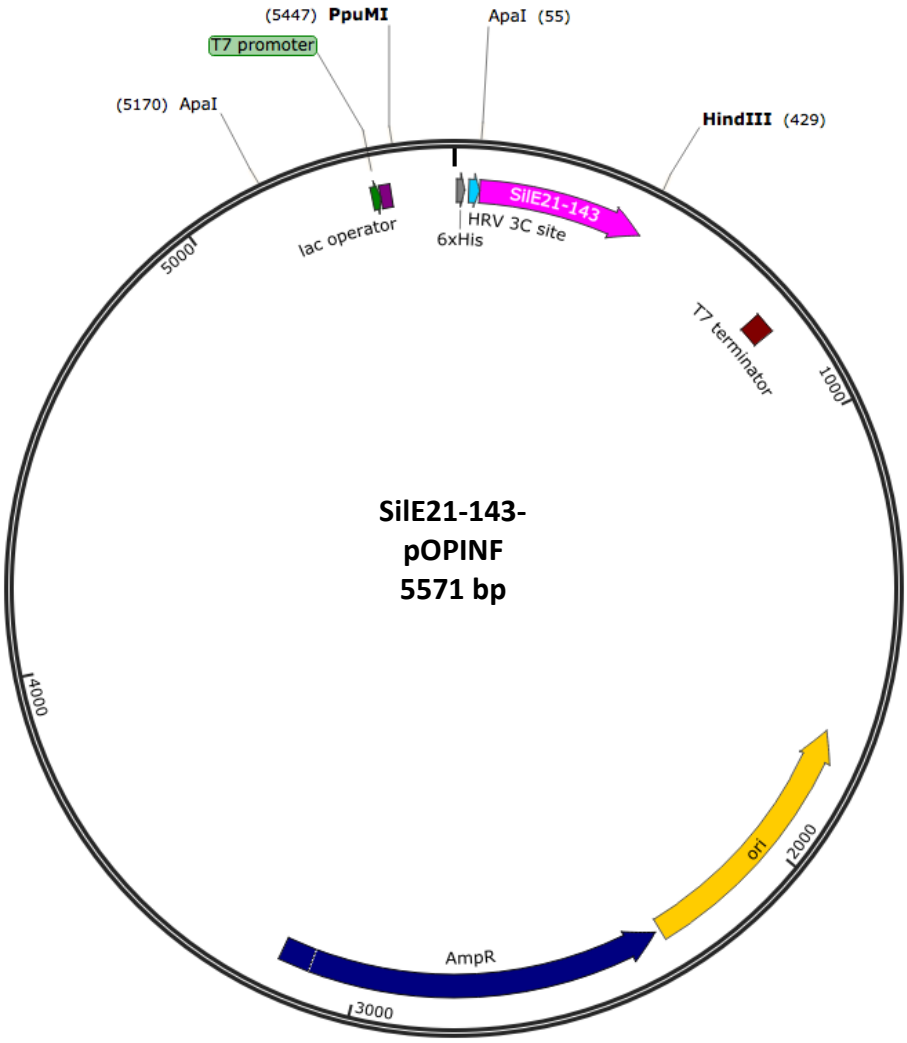


Figure 2. 3 **Map of the pOPINF plasmid expression vector containing the gene for SilE$_{21-143}$**

The *sil* genes were cloned into pOPIN plasmid expression vectors (Table 2.2), like pOPINF, at restriction sites found up- or down-stream of a His$_6$-tag sequence (HHHHHH - grey) and a linker sequence including the Human rhinovirus (HRV)-3C protease cleavage sequence (LEVLFQ/GP – sky blue). Due to the plasmids containing a *lacZ* promoter (purple), the use of the lactose analogue IPTG was permitted, to enable the induction of transcription of the Sil proteins, especially SilE$_{21-143}$ (pink). Moreover, the T7 promoter (green) allowed the initiation of transcription of the encoded genes, via the *E. coli* cells expression of T7 RNA polymerase. The ampicillin resistance gene (navy) was used as a selectable marker. The PpuMI, HindIII and ApaI (post HRV-3C site) restriction sites were used to re-insert *silE* mutants and truncations. Image has been created using SnapGene® after obtaining the information from OPPF-UK, 2013.

2.1.3 Primers

Using the known sequence of the *sil* operon in plasmid pMG101 (GenBank accession number AF067954.1), primers for the genes of seven of the nine encoded proteins - SilE, SilS, SilR, SilC, SilB, SilA and SilP (not SilF or SilG) - were designed; for their cloning. The primers were designed for In-Fusion™ cloning, using primer extensions in accordance with the fusion tag required for the resulting protein (Table 2.2). Primers were also designed and used for protein $SilE_{21\text{-}143F}$ site-directed and truncation mutagenesis. Additionally, *sil* gene and operon primers were used for the verification of the *E. coli* strains. The primers designed and used are listed in Table 2.3.

The primer oligonucleotides were supplied by Eurofins Genomics (Ebersberg, Germany) or IDT (Integrated DNA Technologies - Leuven, Belgium), in a dehydrated form. DNA was resuspended in nuclease-free water to a concentration of 100 μM; from this working stock solutions at 20 μM were also prepared and stored at -20 ^{o}C.

Table 2. 3: Primers used throughout this study

A list of the primers used for cloning, mutation work and strain verification. The sit-directed mutation introducing nucleotides are shown in **bold**, lowercase and the restriction sites in *italic*, lowercase.

Primer	Direction	Sequence (5'-3')	Use
SilE$_{2\text{-}143E}$	Forward	AGGAGATATACCATGAAAAATATCGTATTAGCATCCTTGC	Cloning
	Reverse	GTGATGGTGATGTTTGCCTGCACTGAGCATGCG	
SilE$_{21\text{-}143F}$	Forward	AAGTTCTGTTTCAGGGCCCGACTGAAACCGTGAATATCCATG	Cloning
	Reverse	ATGGTCTAGAAAGCTTTAGCCTGCACTGAGCATGCG	
SilE$_{21\text{-}143E}$	Forward	AAGTTCTGTTTCAGGGCCCGACTGAAACCGTGAATATCCATG	Cloning
	Reverse	GTGATGGTGATGTTTGCCTGCACTGAGCATGCG	
SilR$_{2\text{-}228F}$	Forward	AAGTTCTGTTTCAGGGCCCGAAAATATTAATCGTTGAAGACGACA	Cloning
	Reverse	ATGGTCTAGAAAGCTTTATGCATCCGGGATCTCCAGC	
SilR$_{2\text{-}228E}$	Forward	AAGTTCTGTTTCAGGGCCCGAAAATATTAATCGTTGAAGACGACATTAAAACAGG	Cloning
	Reverse	GTGATGGTGATGTTTTGCATCCGGGATCTCCAGC	
SilR$_{2\text{-}228M}$	Forward	AAGTTCTGTTTCAGGGCCCGCGTCAGGGGCACCTGCC	Cloning
	Reverse	ATGGTCTAGAAAGCTTTATGCATCCGGGATCTCCAGC	

$SilS_{208\text{-}497F}$	Forward	AAGTTCTGTTTCAGGGCCCGCGTCAGGGGCACCTGCCC	Cloning
	Reverse	ATGGTCTAGAAAGCTTTATTCCCAGCACTGGGTGTCC	
$SilS_{208\text{-}497J}$	Forward	AAGTTCTGTTTCAGGGCCCGCGTCAGGGGCACCTGCCC	Cloning
	Reverse	ATGGTCTAGAAAGCTTTATTCCCAGCACTGGGTGTCCG	
$SilS_{208\text{-}497S}$	Forward	AAGTTCTGTTTCAGGGCCCGCGTCAGGGGCACCTGCCC	Cloning
	Reverse	ATGGTCTAGAAAGCTTTATTCCCAGCACTGGGTGTCCG	
$SilS_{226\text{-}497F}$	Forward	AAGTTCTGTTTCAGGGCCCGTCCGAGAATCTTGATGCGCG	Cloning
	Reverse	ATGGTCTAGAAAGCTTTATTCCCAGCACTGGGTGTCC	
$SilS_{241\text{-}497F}$	Forward	AAGTTCTGTTTCAGGGCCCGGAGCTGGAGCAGCTGGTTATC	Cloning
	Reverse	ATGGTCTAGAAAGCTTTATTCCCAGCACTGGGTGTCC	
$SilS_{248\text{-}497F}$	Forward	AAGTTCTGTTTCAGGGCCCGTCGTTCAATCATATGATTGGAAAG	Cloning
	Reverse	ATGGTCTAGAAAGCTTTATTCCCAGCACTGGGTGTCC	
$SilS_{2\text{-}497E}$	Forward	AGGAGATATACCATGCATAGCAAACCGTCCAGACTAC	Cloning
	Reverse	GTGATGGTGATGTTTTTCCCAGCACTGGGTGTCC	
$SilS_{2\text{-}497G}$	Forward	AGGAGATATACCATGCATAGCAAACCGTCCAGACTAC	Cloning
	Reverse	CAGAACTTCCAGTTTTTCCCAGCACTGGGTGTCC	

SilC$_{2\text{-}461E}$	Forward	AGGAGATATACCATGTTCAAATTAAAATTACTCAGCATCAG	Cloning
	Reverse	GTGATGGTGATGTTTCTCTACCCAACCGCCACCC	
SilC$_{2\text{-}461G}$	Forward	AGGAGATATACCATGTTCAAATTAAAATTACTCAGCATCAG	Cloning
	Reverse	CAGAACTTCCAGTTTCTCTACCCAACCGCCACCC	
SilP$_{2\text{-}824E}$	Forward	AGGAGATATACCATGCTTCAGATATGCATAAGGAGAG	Cloning
	Reverse	GTGATGGTGATGTTTTTTCCCGAGCCTGACACTTTTC	
SilP$_{2\text{-}824G}$	Forward	AGGAGATATACCATGCTTCAGATATGCATAAGGAGAG	Cloning
	Reverse	CAGAACTTCCAGTTTTTTCCCGAGCCTGACACTTTTC	
SilB$_{2\text{-}430E}$	Forward	AGGAGATATACCATGGCTTCTTTAAAGATAAAATATGCTGC	Cloning
	Reverse	GTGATGGTGATGTTTGTGCCCTGAATGCATATTTACAG	
SilB$_{2\text{-}430G}$	Forward	AGGAGATATACCATGGCTTCTTTAAAGATAAAATATGCTGC	Cloning
	Reverse	CAGAACTTCCAGTTTGTGCCCTGAATGCATATTTACAG	
SilA$_{2\text{-}1048I}$	Forward	ATCATCACAGCAGCGGCATTGAATGGATTATCCGGCGC	Cloning
	Reverse	ATGGTCTAGAAAGCTTTATGACACGCTTTTTTTATGTCTGC	
SilA$_{2\text{-}1048E}$	Forward	AGGAGATATACCATGATTGAATGGATTATCCGGCGC	Cloning
	Reverse	GTGATGGTGATGTTTTGACACGCTTTTTTTATGTCTGC	

SilA$_{2\text{-}1048G}$	Forward	AGGAGATATACCATGATTGAATGGATTATCCGGCGC	Cloning
	Reverse	CAGAACTTCCAGTTTTGACACGCTTTTTTTATGTCTGC	
SilEH38A	Forward	GGCACCTGCT**gcc**CAGATGCAGT	Site-directed mutagenesis
	Reverse	ACTGCATCTG**ggc**AGCAGGTGCC	
SilEH62A	Forward	TATGGACCAG**gcc**GAACAGGCCATTATTGCTCAT	Site-directed mutagenesis
	Reverse	CATGAGCAATAATGGCCTGTTC**ggc**CTGGTCCAT	
SilEH69A	Forward	CATTATTGCT**gcc**GAAACCATGACGAACGG	Site-directed mutagenesis
	Reverse	CCGTTCGTCATGGTTTC**ggc**AGCAATAATG	
SilEH80A	Forward	GGCGGATGCG**gcc**CAGAAAATGG	Site-directed mutagenesis
	Reverse	CCATTTTCTG**ggc**CGCATCCGCC	
SilEH87A	Forward	GGTGGAAAGT**gcc**CAGAGGATGATG	Site-directed mutagenesis
	Reverse	CATCATCCTCTG**ggc**ACTTTCCACC	
SilEH111A	Forward	AATGAATGAG**gcc**GAAAGAGCTGCAGTTG	Site-directed mutagenesis
	Reverse	CAACTGCAGCTCTTTC**ggc**CTCATTCATT	

SilEH118A	Forward	TGCAGTTGCC**gcc**GAATTTATGAATAACG	Site-directed mutagenesis
	Reverse	CGTTATTCATAAATTC**ggc**GGCAACTGCA	
SilEH129A	Forward	GTCTGGCCCA**gcc**CAGGCCATGG	Site-directed mutagenesis
	Reverse	CCATGGCCTG**ggc**TGGGCCAGAC	
SilEH136A	Forward	GGCCGAAGCG**gcc**CGTCGCATGC	Site-directed mutagenesis
	Reverse	GCATGCGACG**ggc**CGCTTCG	
SilEM72L	Forward	TCATGAAACC**ctg**ACGAACGGGTC	Site-directed mutagenesis
	Reverse	GACCCGTTCGT**cag**GGTTTCATGA	
SilEM83L	Forward	GCACCAGAAA**ctg**GTGGAAAGTCATCAG	Site-directed mutagenesis
	Reverse	CTGATGACTTTCCAC**cag**TTTCTGGTGC	
SilEM90L	Forward	TCATCAGAGG**ctg**ATGGGAAGTCAGAC	Site-directed mutagenesis
	Reverse	GTCTGACTTCCCAT**cag**CCTCTGATGA	

SilEM108L	Forward	ATTAGCGGCA**ctg**AATGAGCATGAAAG	Site-directed mutagenesis
	Reverse	CTTTCATGCTCATT**cag**TGCCGCTAAT	
SilEPpuMIF	Forward	ATTCCCCGGAGTTAATCC*gggacc*TTTAATTC	Site-directed mutagenesis
SilEHindIIIR	Reverse	ATCACAAACTGGTCTAGa*aagctt*TAGCCTGC	Site-directed mutagenesis
SilE$_{46\text{-}128}$	Forward	ATTCCCCGGAGTTAATCC*GGGACC*TTTAATTC	Truncation mutagenesis
	Reverse	GCTAATG*AAAGCTT*CGGTTATTAAGGGGAAACGG	
SilE$_{21\text{-}98}$	Forward	CGATCG*gggccc*GCCTGTCGGGATCCAGGGG	Truncation mutagenesis
	Reverse	GCGCTTC*aagctt*GGCTTATTATGGGCCAG	
***silE* gene**	Forward	ATGAAAAATATCGTATTAGC	Strain verification
	Reverse	TCAGCCTGCACTGAGCATGC	
***sil* operon**	Forward	GATCTTATCGAGGAAATTAC	Strain verification
	Reverse	TGCAAAGAAAGCGGGAAATC	

2.1.4 Bacterial growth media

Media was used for bacterial cell culturing purposes prior to DNA extraction and overexpression ^{15}N-labelled and unlabelled protein. Preparations of the antibiotics used during culturing of cells can be found in Appendix A.

2.1.4.1 Luria Broth (LB) media

10g L^{-1} tryptone, 10g L^{-1} NaCl and 5g L^{-1} yeast extract were made up to the required volume by adding dH_2O and then autoclaved to sterilise. Antibiotics were added (100μg mL^{-1} ampicillin or 50μg mL^{-1} carbenicillin and 34μg mL^{-1} chloramphenicol) upon cooling, if required.

2.1.4.2 LB agar

To make LB agar, prior to autoclaving, 15g L^{-1} bacteriological agar (Oxoid Limited, Basingstoke) was added to the LB media described above.

2.1.4.3 M9 minimal media

The following was added to begin the preparation: 6 g L^{-1} Na_2PO_4, 3 g L^{-1} KH_2PO_4, 0.5 g L^{-1} NaCl, 0.3 g L^{-1} $MgSO_4$, 15 mg L^{-1} $CaCl_2$ L^{-1} g. This solution was autoclaved and allowed to cool. When the solution had cooled down, the following was added –

- 1mL L^{-1} 1000x filter sterilised d-Biotin (10mg d-Biotin, 10ml 50% EtOH)

- 1mL L^{-1} 1000x filter sterilised Thiamine (10mg Thiamine, 10mL dH_2O)
- 10 mL L^{-1} 10x stock EDTA trace elements (ingredients include: 1g EDTA, 0.32g $MnCl_2$, 0.1g $FeCl_3$, 0.01g $ZnCl_2$, 0.002g $CoCl_2$, 0.002 g $CuCl_2$, 0.002g H_3BO_3, adjusted to pH 7.0 and made up to 200mL with dH_2O).
- 10 mL L^{-1} 100x Gibco® Minimum Essential Medium Vitamin Solution (Thermo Thermo Fisher Scientific UK Ltd., Loughborough, UK)
- 4g L^{-1} filter sterilised $C_6H_{12}O_6$
- 1g L^{-1} filter sterilised $^{15}NH_4Cl$

Antibiotics (100µg mL^{-1} ampicillin or 50µg mL^{-1} carbenicillin) were then added prior to bacterial culturing.

2.1.4.4 Power Broth

Power Broth™ (recipe not provided - AthenaES™, Enzo Life Sciences (UK) Ltd. Exeter, UK). Prepared according to manufacturers instructions. The medium is buffered at pH 7.2±0.2 and rich in supplies of amino acids, vitamins, and a carbon source to improve the expression of recombinant proteins in *E. coli*.

2.1.5 Buffers and solutions

2.1.5.1 DNA preparation

Buffer NPI-10-Tween

50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, 1% v v^{-1} Tween 20, adjusted pH to 8.0 using NaOH and filtered before use. Stored at 4°C.

Buffer NPI-20-Tween (Wash Buffer)

50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, 0.05% v v^{-1} Tween 20, adjusted pH to 8.0 using NaOH and filtered before use. Stored at 4°C.

Buffer NPI-250-Tween, (Elution Buffer)

50 mM NaH_2PO_4 , 300 mM NaCl, 250 mM Imidazole, 0.05% Tween, adjusted pH to 8.0 using NaOH and filtered before use. Stored at 4°C.

DNAse I Stock Solutions (40,000 units mL^{-1})

To a 200,000 unit bottle of DNAse I (Sigma D-4527) added 5mL of UHQ Sterile Water. Aliquoted into 25µL aliquots and stored at -20°C. Added an aliquot to 25 mL of Buffer NPI-10-Tween buffer - used for lysis.

Lysozyme

Added 25 mg of freshly weighed Lysozyme to the 25 mL of Buffer NPI-10-Tween buffer used for lysis.

2X Phusion Master Mix with GC Buffer

Phusion® High-Fidelity DNA Polymerase (1 unit), 25 mM TAPS-HCl

(pH 9.3 at 25°C), 50 mM KCl, 2 mM $MgCl_2$, 1 mM beta-mercaptoethanol and 200 µM dNTPs) (Thermo Fisher Scientific UK Ltd., Loughborough, UK).

2.1.5.2 Protein overexpression, extraction and purification

IPTG

1 M stocks were prepared by dissolving 2.38 g of IPTG in 10 mL of dH_2O and filter sterilised. 1 mL aliquots were prepared and stored at -20 °C.

Lysis Buffer

50 mM Tris, 500 mM NaCl, 0.2% Tween 20, 1 mg mL^{-1} Lysozyme and pH 8.0.

Column Loading/Wash Buffer

50 mM Tris, 500 mM NaCl and pH 8.0.

Elution Buffer

50 mM Tris, 500 mM NaCl, 500 mM or 2 M Imidazole and pH 8.0.

Gel Filtration Buffer

20 mM Tris, 200 mM NaCl and pH 8.0.

2.1.5.3 Gel electrophoresis

6x DNA Sample Preparation Buffer

50 % v v^{-1} Glycerol, 50 mM EDTA, 0.25 % w v^{-1} Bromophenol Blue, 0.25 % w v^{-1} Xylene Cyanol.

4x Protein Sample Preparation Buffer

200 mM Tris, 8% w v^{-1} Sodium Dodecyl Sulphate, 0.2% w v^{-1} Bromophenol Blue, 40% v v^{-1} Glycerol, 50 mM EDTA and 400 mM DTT.

1x TAE Buffer

40 mM Tris, 20 mM acetic acid, and 1 mM EDTA and pH 8.0.

1x Tris/Glycine SDS Running Buffer

25 mM Tris, 192 mM glycine, 0.1% w v^{-1} SDS and pH 8.3.

2.1.5.4 Protein experimental

HEPES Buffer

10 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), 20 mM NaF and pH 7.5.

Ammonium Acetate

25 mM ammonium acetate ($C_2H_3O_2NH_4$) at pH 7.0.

2.1.6 Markers

Table 2. 4: DNA and protein gel markers

A list of the markers used for agarose gels and SDS-PAGE, alongside their size range and manufacturing company are shown below.

Marker	Size Range (bp or kDa)	Source
DNA Agarose Gel Electrophoresis		
HyperLadder™ 1 kb	200-10,037 bp	Bioline Reagents Ltd., London, UK
100 bp DNA Ladder	100-1,517 bp	New England Biolabs, Ipswich, UK
1 kb DNA Ladder	500-10,000 bp	New England Biolabs, Ipswich, UK
Quick-Load® 1 kb Extend Ladder	500-48,500 bp	New England Biolabs, Ipswich, UK
SDS-PAGE		
SigmaMarker™ Low Range	6.5-66 kDa	Sigma Aldrich, Poole, Dorset, UK
SigmaMarker™ High Range	36-200 kDa	Sigma Aldrich, Poole, Dorset, UK
Unstained Protein Ladder, Broad Range (10-250 kDa)	10-250 kDa	New England Biolabs, Ipswich, UK
PageRuler™ Unstained Protein Ladder	10-200 kDa	Thermo Fisher Scientific UK Ltd., Loughborough, UK

2.2 Methodology

Any primers, media, buffers or solutions used throughout the methodology can be found listed within the materials in Section 2.1.

2.2.1 Bioinformatics

To assemble and analyse sequencing information as well as assess protein sequence homology and predict structural characteristics of SilE, a number of bioinformatics techniques were used. The software used are listed in Table 2.5.

Table 2. 5 Bioinformatics software utilised throughout study

A table showing the different software that were used to generate the assembled annotated genomes and structural predictions of SilE, along with a reference and the site downloaded from.

Software	Use	Citation	Download Link or Website
SPAdes	Genome assembler.	(Bankevich *et al.*, 2012)	http://bioinf.spbau.ru/content/spades-download
RAST	Online genome annotation tool.	(Aziz *et al.*, 2008)	http://rast.nmpdr .org/
Mauve	Genome alignment tool used to align assembled genomes to reference genomes.	(Darling *et al.*, 2004)	http://darlinglab. org/mauve/down load.html
Artemis	Genome browser used to visualise the sequence and allows for sequence annotation.	(Carver *et al.*, 2012)	https://www.sanger.ac.uk/resources/software/artemis/#download
BLAST	Basic Local Alignment Search Tool that was used to search genes or proteins against the NCBI database or against specified sequences.	(Altschul *et al.*, 1990)	http://blast.ncbi.nlm.nih.gov/Blast.cgi

Software	Use	Citation	Download Link or Website
Clustal Omega	General multipurpose alignment program for nucleotides and proteins; finds the best alignment over the entire length of each sequence submitted.	(McWilliam *et al.*, 2013)	http://www.ebi.ac.uk/Tools/msa/clustalo/
Jalview	Multiple sequence alignment editing, visualisation and analysis.	(Waterhouse *et al*., 2009)	http://www.jalview.org/download
PrDOS-CNF	IDP prediction server.	(Ishida and Kinoshita, 2007)	http://prdos.hgc.jp/cgi-bin/top.cgi
GeneSilico MetaDisorder Service	IDP prediction metaserver, inclusive of GSMetaDisorderMD2 and GSMetaDisorder.	(Kozlowski and Bujnicki, 2012)	http://genesilico.pl/metadisorder/
JPred	A protein secondary structure prediction server that incorporates the Jnet algorithm to make more accurate predictions.	(Cole *et al*., 2008)	http://www.compbio.dundee.ac.uk/jpred/
COACH	A meta-server that makes protein-ligand binding site predictions.	(Yang *et al.*, 2013b)	http://zhanglab.ccmb.med.umich.edu/COACH/

2.2.2 Gel electrophoresis

2.2.2.1 Agarose gel electrophoresis

Samples were prepared with 6x DNA Sample Preparation Buffer and loaded on 0.3-1.0 % w v^{-1} TAE agarose gels. These were prepared by, dissolving agarose through melting in a microwave at full power, in 1x TAE Buffer supplemented with ethidium bromide to a final concentration of 0.5µg mL^{-1}. The gel was then run (on average) at 80 V, for 30-35 minutes, in 1x TAE Buffer. The DNA molecular weight markers used are in listed in Table 2.4.

2.2.2.2 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Running the gel

A 20 µL protein sample was obtained and to it was added 10 µL of 4x Protein Sample Preparation Buffer. Samples were heated for 10 minutes at 70 ^{o}C and then centrifuged for 10 minutes at 11,000 x g, in a micro-centrifuge. The DNA was then removed and 20 µL of each sample was loaded onto a 4–20% Mini-PROTEAN® TGX™ Precast Gel (Bio-Rad Laboratories Ltd., Hertfordshire, UK) or SDS-PAGE gel prepared at 16 or 18 % (Table 2.6). The gel was run at 200 V, 125 mA and 25 W for 35 minutes in 1x Tris/Glycine SDS Running Buffer using a Bio-Rad Protean® III System (Bio-Rad Laboratories Ltd., Hertfordshire, UK).

Table 2. 6: SDS-PAGE resolving gel components

A table listing the ingredients required for the making of either a 16 or 18 % SDS-PAGE resolving gel. 80 μL of 10 % APS and 8 μL of TEMED added for polymerisation, prior to gel setting.

Component	16 %	18 %
30 % Acrylamide	53.3 mL	60.0 mL
1.5M Tris-HCl, pH 8.8	25.0 mL	25.0 mL
10 % SDS	1.0 mL	1.0 mL
dH_2O	19.6 mL	12.9 mL

Staining and de-staining

Gels were stained using pre-prepared GelCode Blue safe protein stain (Fisher Scientific, Loughborough, UK). Gels were stained for 30-45 minutes whilst on a rocker platform. Gels were de-stained in water until protein bands were visible.

2.2.3 *E.coli* growth and maintenance

Bacterial inoculation and growth was carried out under aseptic conditions.

2.2.3.1 Sterilisation

All of the following methods were carried out under sterile conditions. Bench surfaces were regularly decontaminated using the multipurpose disinfectant Virkon® (2 % v v^{-1}) and prior to use, 70 % ethanol was used to clean the surfaces further, with all experiments involving bacterial cell culture being conducted in close proximity to a blue Bunsen

flame. Disposable equipment (Petri Dishes, Universals, Falcon tubes, syringes, needles and filters) was supplied sterile. Pipette tips, Eppendorf® tubes, conical culture flasks, Duran® bottles, growth media and all buffers were sterilised by heating to 121 ^{o}C in an autoclave (Astell, Kent, UK) at 15 psi, for 15 mins. Heat sensitive chemical reagents (i.e. antibiotics) were filter sterilised using a 0.22 μm syringe filter (Whatman® GD/X syringe filters Nylon membrane) before use.

2.2.3.2 Standard growth conditions

All *E. coli* cell growth was carried out in LB (Lysogeny broth) medium, in culture flasks, at 37 ^{o}C with shaking at 200 rpm in an orbital shaker - unless stated otherwise. To select for the cells containing the correct plasmids, 100 μg mL^{-1} ampicillin or 50 μg mL^{-1} carbenicillin was added to all of the *E. coli* liquid cell cultures and 34 μg mL^{-1} chloramphenicol (Appendix A) was additionally supplemented to the *E. coli* Rosetta (DE3) pLysS and Rosetta Star™ (DE3) pLysS cell (Table 2.1) liquid medium growths. LB agar was used as a growth medium to culture all of the *E. coli* cells at 37 ^{o}C.

2.2.3.3 Storage of strains

Glycerol stocks of the propagation and expression strains of *E. coli* were prepared to maintain and store the cloned constructs. Five mL of selective, LB medium was inoculated with a single colony of the desired cells, which was then grown overnight. The resulting culture was then

centrifuged at 3,500 x g, at room temperature for 1 minute, in a bench top micro-centrifuge. The supernatant was decanted and discarded whilst the pellet was re-suspended in a total volume of 2 mL LB (50 % v v^{-1}) mixed with sterile glycerol (50 % v v^{-1}). The glycerol stock was then distributed into 50 µl aliquots, snap frozen in liquid nitrogen and stored at -80 ^{o}C.

2.2.3.4 Making chemically competent *E. coli*

Chemically competent *E. coli* cells were prepared by initially growing a 5 mL, LB medium overnight culture, following inoculation with a single colony. Then, 100 mL of LB medium (including relevant antibiotics for modified strains) was inoculated with 1 mL of the saturated overnight culture (a 1 in 100 dilution). This culture was grown until an OD_{600} of 0.3-0.4 was reached (within a few hours), followed by a 10 minute incubation period on ice (to inhibit further growth). The following successive steps were then performed using pre-chilled containers and solutions. The cooled culture was then centrifuged, at 3,500 x g for 10 mins at 4 ^{o}C. The supernatant was removed and discarded. The pellet was re-suspended into 10 mL ice-cold 100 mM $CaCl_2$ and incubated on ice for a further 20 mins. Following the incubation period, the re-suspension was again centrifuged, at 3,500 x g for 10 mins at 4 ^{o}C. The supernatant was disposed of and the pellet re-suspended, but this time into 2 mL ice-cold 100 mM $CaCl_2$, 15 % v v^{-1} sterile glycerol. The chemically competent *E. coli* cells were then divided out into 50 µL aliquots and snap frozen in liquid nitrogen, before being stored at -80 ^{o}C.

2.2.3.5 Making electrocompetent *E. coli*

Electrocompetent *E. coli* cells were produced after inoculating 500 mL LB medium with 5 mL of a 10 mL overnight starter culture. This was cultured to an OD_{600} of 0.7, at which point the culture was incubated on ice for 30 mins (to stop cell growth). All of the subsequent steps were carried out at 4 ^{o}C, with pre-chilled containers and solutions. The cells were harvested at 3,500 x g for 15 mins at 4 ^{o}C, before the cell pellet being re-suspended and washed (by swirling) in 200 mL of sterile water. The re-suspension was then harvested at 6,000 x g for 12 mins at 4 ^{o}C. The cell pellet re-suspension, wash and centrifugation step was repeated using the same conditions. The resulting cell pellet was once more re-suspended and washed but in 100 mL sterile water and harvested using the same centrifugation parameters. The same re-suspension, wash and final centrifugation were repeated in 20 mL sterile water. The culminating cell pellet was re-suspended into 4 mL of sterile water, divided out into 50 µL aliquots, snap frozen in liquid nitrogen and stored at -80 ^{o}C.

2.2.4 Plasmid DNA preparations

2.2.4.1 Plasmid DNA isolations

For the sequence analysis of the silver resistance operon containing J53 (pMG101), DNA extraction and purification was carried out using the QIAGEN Large-Construct Kit (Qiagen, Manchester, UK). Other plasmid preparations were conducted using the QIAprep® Spin Miniprep Kit

(Qiagen, Manchester, UK). Both kits are based on the Alkaline Lysis method invented by (Birnboim and Doly, 1979). The protocols in the user manuals were strictly followed. The only deviation was a double elution step, in which the elution buffer was reloaded onto columns to provide increased plasmid yields. The procedure consists of three basic steps: firstly, the preparation and clearing of a bacterial lysate, followed by adsorption of DNA onto the QIAprep membrane and finally the washing and elution of plasmid DNA.

2.2.4.2 DNA sample concentration and purity

DNA concentration was measured using a Nanodrop™ ND1000 spectrometer (Thermo Scientific, Wilmington, USA). A measurement of ultraviolet (UV) absorbance at a wavelength of 260 nm (A_{260}) is taken for DNA samples. This is then entered into the Beer Lambert Law equation (Eq. 2.1) to calculate concentration:

Beer Lambert Law $A = \varepsilon cl$ *(Eq. 2.1)*

Where A is absorbance at A_{260}, ε = the molar absorption coefficient (L mol^{-1} cm^{-1}), c is concentration (M) and l is path length (cm). The average extinction coefficient for double stranded DNA is 0.02 $(\mu g/mL)^{-1}$ cm^{-1}. Sample purity - no contamination of other molecules (i.e. proteins and organic compounds) was assessed, by determining the ratio of absorbance at

260 and 280 nm ($A_{260/280}$). For DNA a $A_{260/280}$ ratio of 1.8-2.0 is considered as pure (Glasel, 1995).

2.2.4.3 Transformations

Heat shock

1 µL of plasmid DNA - at a concentration of 100 ng μL^{-1} - was added to an aliquot (50 µL) of thawed chemically competent *E. coli* cells, on ice. The cell-DNA mix was then incubated on ice for 30 mins. The cells were then heat-shocked for 30 seconds at 42 ^{o}C, in a water bath, before being returned to the ice for a couple of mins. During this time, 1 mL of LB medium was added to the cells. Recovery was then allowed with incubation in a shaking incubator set at 37 ^{o}C with shaking at 200 rpm, for an hour. 50, 100 or 200 µL of cells from the transformation reactions were spread onto LB agar plates containing the appropriate antibiotic selection and incubated inverted overnight at 37 ^{o}C.

Electroporation

1 µL of plasmid DNA - at a concentration of 100 ng μL^{-1} - was added to an aliquot (50 µL) of thawed electrocompetent *E. coli* cells, on ice, before being transferred to a pre-chilled, ice-cold electroporation cuvette. The cell-DNA mix was then electroporated with a pulse of 25 µF capacitance, 2.5 kV and 200 Ω resistance. Instantly, 1 mL of LB medium was added to the cells as they were allowed to recover over an hour's incubation period in a shaking incubator, set at 37 ^{o}C with shaking at 200

rpm. 50, 100 or 200 µL of cells from the transformation reactions were spread onto LB agar plates containing the appropriate antibiotic selection and incubated inverted overnight at 37 ^{o}C.

2.2.4.4 Cloning

Cloning was carried out to produce Sil protein overexpression constructs (Table 2.7). The pOPIN vectors used included an Amp resistant gene (Figure 2.1), which ensured that following transformation only colonies containing the correct plasmid grew. The pOPIN vectors also contain a *lacZ* promoter that allowed inducible expression of the SilE protein, upon the addition of the lactose analogue isopropyl β-D-1-thiogalactopyranoside (IPTG). Within the plasmids (excluding poPINI – Table 2.2), a His_6-tag and a HRV-3C protease sequence could also be found. In $SilE_{21\text{-}143F}$, the N-terminal His_6-tag and the HRV-3C protease cleavage site enabled the purification of the SilE proteins.

Table 2. 7: The Sil proteins, genes, PCR products and plasmid expression vectors

Information displayed after obtaining the generous provision from OPPF-UK, 2013.

Protein	Amino Acid Residue Numbers	PCR Product Length (bp)	Vector	Cloned Construct Size (bp)	Protein MW (kDa)	Tag Cleaved Protein MW (kDa)	Description
SilE	2-143	456	pOPINE	5577	16.7	14.6	Full cds with native periplasmic signal sequence and added C-his tag
SilE	21-143	407	pOPINF	5571	15.8	13.8	Full cds w/o periplasmic signal sequence cleavable N-his tag
SilE	21-143	399	pOPINE	5520	14.8	12.8	Full cds w/o signal periplasmic sequence and added C-his tag
SilR	2-228	719	pOPINF	5583	27.8	25.8	Full cds with cleavable N-his tag
SilR	2-228	711	pOPINE	5832	26.8	24.8	Full cds with C-his tag
SilR	2-228	719	pOPINM	7017	68.7	25.8	FL pOPINM
SilS	208-497	908	pOPINF	6072	35.7	33.7	Whole aa 208-497 cytoplasmic domain defined by TMpred; N-his tag
SilS	208-497	908	pOPINJ	6759	62.0	33.7	Whole aa 208-497 cytoplasmic domain defined by TMpred; N-his tag

Protein	Amino Acid Residue Numbers	PCR Product Length (bp)	Vector	Cloned Construct Size (bp)	Protein MW (kDa)	Tag Cleaved Protein MW (kDa)	Description
SilS	208-497	908	pOPINS3C	6360	46.9	33.6	Whole aa 208-497 cytoplasmic domain defined by TMpred; N-his tag
SilS	226-497	854	pOPINF	6018	33.7	31.6	aa 226-497 N- truncated cytoplasmic domain defined by secondary structure prediction; N-his tag
SilS	241-497	809	pOPINF	5973	32.0	30.0	aa 241-497 truncated cytoplasmic domain defined by secondary structure prediction; N-his tag
SilS	248-497	788	pOPINF	5952	31.2	29.1	aa 248-497 truncated cytoplasmic domain defined by phyre model based on PDB id 2C2A
SilS	2-497	1518	pOPINE	6639	57.2	54.7	Full cds with C-his tag
SilS	2-497	1518	pOPINE-3C-eGFP	7380	85.8	57.7	Full cds with GFP-C-his tag
SilC	2-461	1410	pOPINE	6531	52.3	50.0	Full cds with C-his tag
SilC	2-461	1410	pOPINE-3C-eGFP	7272	80.3	52.2	Full cds with GFP-C-his tag

Protein	Amino Acid Residue Numbers	PCR Product Length (bp)	Vector	Cloned Construct Size (bp)	Protein MW (kDa)	Tag Cleaved Protein MW (kDa)	Description
SilP	2-824	2499	pOPINE	7620	90.5	88.2	Full cds with C-his tag
SilP	2-824	2499	pOPINE-3C-eGFP	8361	118.5	90.4	Full cds with GFP C-his tag
SilB	2-430	1317	pOPINE	6438	49.3	47.0	Full cds with C-his tag
SilB	2-430	1317	pOPINE-3C-eGFP	7179	77.4	49.2	Full cds with GFP C-his tag
SilA	2-1048	3176	pOPINI	6317	114.7	N/A	Full cds based on PDB 3K07, with non-cleavable N-his tag
SilA	2-1048	3171	pOPINE	8292	117.1	114.7	Full cds with C-his tag
SilA	2-1048	3171	E-3C-eGFP	9033	145.1	117.0	Full cds with GFP C-his tag

PCR reactions

To obtain all of the required *sil* genes, Polymerase Chain Reactions (PCRs) were carried out using two different polymerase enzymes: KOD Xtreme™ Hot Start DNA Polymerase (Novagen - Merck Millipore, Dermstadt, Germany) and Phusion Flash High-Fidelity DNA Polymerase - as part of a PCR Master Mix - (Thermo Fisher Scientific UK Ltd., Loughborough, UK). The PCRs were set-up and run in a thermal cycler as explained in Table 2.8.

All analysis of the PCR products, following the completion of the thermal cycling, was carried out on a 1% TAE agarose gel. Once verified, the PCR products were purified using the Agencourt® AMPure® XP Magnetic Bead Purification Kit (Beckman Coulter (UK) Ltd., High Wycombe, UK), by following the manufacturer's protocol.

Table 2. 8: Cloning and mutagenesis PCR experiments

This table displays information on the different types of PCRs (cloning and mutagenesis) that were set-up using primers from Table 2.3, different polymerase enzymes and thermal cycling parameters. Cloning Ann (^{o}C) = 60, Mutagenesis Ann (^{o}C) = 63 for: H38A, H69A, H80A, H87A, H111A, H118A, M72L and M90L and 68 for: H62A, H129A, H136A, M83L and M108L. *Repeated denaturation, anneal and extension steps x 30 cycles.

PCR Method	Polymerase Enzyme	Components of Reaction Mix	Thermal Cycling Parameters	
Cloning	**KOD Xtreme™ Hot Start DNA Polymerase**	25 µL 2X KOD Hot Start Buffer, dNTPs at final concentration of 500 µM, forward and reverse primers at final concentration of 400 nM, template DNA at final concentration of 5 ng µL^{-1}, 1 µL KOD Hot Start (1 U µL^{-1}) and sterilised water taking the total reaction mix volume up to 50 µL.	94 °C 98 °C Ann °C 72°C 72 °C 4 °C	2 mins 10 secs * 30 secs * x mins* 5 mins Hold
	Phusion Flash High-Fidelity DNA Polymerase	25 µL 2X Phusion Flash Master Mix, forward and reverse primers at final concentration of 400 nM, template DNA at final concentration of 5ng µL^{-1}, 1.5 µL 100X DMSO, sterilised water taking the total reaction mix volume up to 50 µL.		
Mutagenesis PCR Round 1	**Phusion High-Fidelity DNA Polymerase**	25 µL 2X Phusion Master Mix with GC Buffer, forward and reverse primers at final concentration of 500 nM, template DNA at a final concentration of 5ng µL^{-1}, 1.5 µL 100X DMSO, sterilised water taking the total reaction mix volume up to 50 µL.	As above, but appropriate alterations were made to Ann (^{o}C)	
Mutagenesis PCR Round 2		25 µL 2X Phusion Master Mix with GC Buffer, forward and reverse primers at final concentration of 500 nM, equimolar template DNA total 18 µL, 1.5 µL 100X DMSO, sterilised water taking the total reaction mix volume up to 45 µL.		

In-fusion reactions

2 µL (200 ng) of the appropriate purified PCR insert was added to 1 µL (100 ng) of the corresponding linearised pOPIN vector (Table 2.7). 7 µL of sterile water was added and the 10 µL mix was transferred to a dry-down In-Fusion™ plate. The contents were briefly mixed by pipetting up and transferred to a thermal cycler where they were left to react for 30 mins at 42 °C. When the In-Fusion reaction was complete, the reactions were transferred to ice and diluted with the addition of 40 µL of TE buffer. The In-Fusion reactions were then transformed by heat shock into chemically competent OmniMAX^{TM} 2 -T1^{R} cells (Invitrogen, Paisley, UK) (Table 2.1). Blue-white screening was performed to detect successful recombination events, whereby white colonies were the cells containing positive recombinants.

Site-directed mutagenesis of SilE$_{21\text{-}143F}$

Overnight cultures of XL1-Blue cells containing the SilE$_{21\text{-}143}$ pOPINF plasmid were prepared via inoculation in 5 mL selective, LB broth. The cultures were spun down in a bench top micro-centrifuge at 1,000 x g for 10 mins and the supernatant was discarded. The plasmid was purified from the cell pellets using the QIAprep® Spin Miniprep Kit (Qiagen, Manchester, UK). To begin with the Q5® Site-Directed Mutagenesis kit (NEB, Hertfordshire, UK) was used in an attempt at creating nine different histidine mutants (H38A, H62A, H69A, H80A, H87A, H111A, H118A, H129A and H136A) of SilE$_{21\text{-}143F}$. Success using the kit's protocol was unachievable, for all nine of these mutants, as was trying various conditions

such as adding dimethyl sulfoxide (DMSO) and extra magnesium chloride ($MgCl_2$), in addition to lower annealing temperatures. Therefore instead, for all of the histidine and methionine SilE mutants produced, the older method of two-step PCR by overlap was executed.

The $SilE_{21\text{-}143F}$ plasmid construct served as a template for PCR-based site-directed mutagenesis. Histidine to alanine (single and double) and methionine to leucine mutations were generated in 5' and 3' DNA fragments using Phusion® High-Fidelity DNA Polymerase - as part of a PCR Master Mix - (Thermo Fisher Scientific UK Ltd., Loughborough, UK). Hence, for each mutant, two separate reactions were set up - one for the first half of the gene (Forward 5'-3' PCR), the other for the second half (Reverse 3'-5' PCR), using two different sets of primers (Table 2.3). The PCRs were set-up as described in Table 2.8. In the front half reactions the forward primer contained a restriction site for PpuMI whilst the reverse primer contained a mismatched codon in order to mutate either a histidine to an alanine, or a methionine to leucine. In the back half reactions the reverse primer contained a HindIII restriction site whilst the forward primer contained a mismatched codon in order to mutate a histidine to an alanine, or a methionine to leucine. For round one of the PCRs a slightly different set of thermal cycling parameters were performed for each mutant, as displayed in Table 2.8.

The PCR products from the first round of the two-step method were confirmed using an agarose gel (1.8 % w v^{-1}) and then extracted and cleaned-up using the GenElute™ Gel Extraction Kit (Sigma Aldrich, Poole, Dorset, UK), following the kit's protocol. In summary, the DNA fragments

of interest were extracted after solubilising the agarose gel slice via the kit's 'Gel Solubilisation Solution'. A combination of a silica-binding technology, washing and high-speed centrifugation allowed selective adsorption of the isolated DNA fragments and removal of contaminants. The bound PCR DNA products were then eluted into 40 μL Tris buffer and the concentrations measured.

The two halves of the mutated gene produced in the first round of PCR were combined to setup reactions for the second round of PCR. Table 2.8 lists the components of the reaction mix and the thermal cycling parameters, except a higher annealing temperature of 68 ^{o}C was used for mutants: H62A, H129A, H136A, M83L and M108L.

Following the cycling, 2.5 μL (500 nm) of each primer containing the restriction site (forwards and backwards) were added to the reactions. Cycling was continued as explained in Table 2.8, but for 20 cycles and at an annealing temperature of 64 ^{o}C. The GenElute™ Gel Extraction Kit (Sigma Aldrich, Poole, Dorset, UK) was used again to purify the DNA from the reactions and the PCR gene concentration measured.

Constructing $SilE_{21-143F}$ truncation mutants

The $SilE_{21-143}$-pOPINF construct served as a template for PCR-based production of the truncated polypeptides $SilE_{46-128}$ and $SilE_{21-98}$. The primers (Table 2.3) were designed to include the appropriate restriction sites - $SilE_{21-98}$ with PpuMI/HindIII sites and $SilE_{46-128}$ ApaI/HindIII sites (Figure 2.1). A procedure similar to that of producing the mutants was completed,

using only a single PCR step. The PCRs were carried out as Table 2.8 explains, yet at an annealing temperature of 64 ^{o}C was used.

The final constructs were cloned into the PpuMI/HindIII site for $SilE_{21-98}$ and into the ApaI/HindIII site for $SilE_{46-128}$, of the prokaryotic expression vector pOPINF, ensuring the presence of the N-terminal His-tag remained. Restriction digests for the $SilE_{21-98}$ truncation were setup as elucidated in Table 2.9. However, due to the presence of two ApaI restriction sites in the $SilE_{21-143}$-pOPINF construct (Figure 2.1), an individual HindII digest was followed by a partial ApaI digest (0.2 μL (10 units), 37 ^{o}C for ten mins) of the vector. The correct restriction digest products were purified using the GenElute™ Gel Extraction Kit (Sigma Aldrich, Poole, Dorset, UK) and ligated (Table 2.10).

Mutant gene insertion and ligation

The mutated genes were inserted back into the pOPINF plasmid vector (Figure 2.1) through double restriction digests of both the gene insert and the vector. The digests were set-up as explained in Table 2.9 and incubated at 37 ^{o}C for one hour.

Table 2. 9: Mutant gene insert restriction digests

This table lists the components of the double restriction digests that were used to insert the SilE mutants back into the pOPINF vector.

Component	Volume (µL)
10X Cut Smart Buffer (NEB, Hertfordshire, UK)	2.5 µL
PCR product / Vector	5 µL (~500 ng) / 2.5 µL (~ 500 ng)
HindIII	1.5 µL (25 units)
PpuMI	0.5 µL (5 units)
sdH_2O	upto 25 µL
Total Reaction Volume	**25 µL**

The DNA was purified from the digests using the GenElute™ Gel Extraction Kit (Sigma Aldrich, Poole, Dorset, UK), after being run on a 1 % w v^{-1} agarose gel and concentrations measured. Ligations of the restriction products were setup as in Table 2.10 and incubated at 16 ^{o}C overnight. Following Ligase heat inactivation (65 ^{o}C for 10 mins), the ligation products were transformed into electrocompetent XL1-Blue cells (Invitrogen, Paisley, UK) (Table 2.1).

Table 2. 10: Ligation reaction

This table lists the components of the ligation reaction used to insert the SilE mutants back into the pOPINF vector.

Component	Amount
10X T4 DNA Ligase Buffer (NEB, Hertfordshire, UK)	2 µL
Digested pOPINF vector	~ 25 ng
Digested gene insert	~ 75 ng
T4 DNA Ligase	1 µL
sdH_2O	upto 20 µL
Total Reaction Volume	**20 µL**

The same methods were applied to add a second histidine mutation, in the histidine double mutants H80A/H87A, H80A/H111A and H87A/H11A of SilE, where the identical process was repeated to introduce the additional mutation.

Overexpression construct sequencing

PCR and/or restriction mapping verified newly cloned plasmids and mutagenesis products. DNA of encouraging hits was purified using the QIAprep® Spin Miniprep Kit (Qiagen, Manchester, UK). Confirmation of complete cloning and the absence of undesired spontaneous mutations were verified by sequencing at the in-house sequencing laboratory (School of Life Sciences, University of Nottingham). T7 Promoter Forward and T7 Terminator sequencing primers were provided by the in-house service.

2.2.5 Protein preparations

2.2.5.1 Small-scale expression screens

The Sil protein expression constructs for overexpressing SilE, SilS, SilR, SilC, SilB, SilA and SilP, were transformed and expressed at a small scale in four different strains of *E. coli* cells. The cells used were: *E. coli* B834 (DE3), Rosetta (DE3) pLysS (Novagen - Merck Millipore, Dermstadt, Germany), OverExpress™ C41 (DE3) and OverExpress™ C43 (DE3) (Lucigen - Middleton, USA) (Table 2.1). All 23 of the constructs were transformed into the B834 (DE3), Rosetta (DE3) pLysS cells (Novagen - Merck Millipore, Dermstadt, Germany). Whereas, transformation of only the membrane proteins (SilS, SilC, SilB, SilA and SilP), into OverExpress™ C41 (DE3) and OverExpress™ C43 (DE3) (Lucigen - Middleton, USA) cells was conducted. This is due to known fact that membrane protein expression is effective in the latter two strains. Two different sets of expression conditions - IPTG- and auto-induced screens - were tested along with all four of these strains, to identify the optimal soluble protein expression conditions for each protein.

Expression screen with IPTG induction

Deep-well blocks were prepared by addition of 0.5 mL of Power Broth™ supplemented with 1% Glucose and the appropriate antibiotic. Individual colonies were picked into each well and the plates sealed with gas-permeable adhesive seals. The filled plates were shaken at 240 rpm at

37 °C overnight. A 24-well deep well plate was then prepared, per strain screen, by addition of 2.5 mL of Power Broth™ supplemented with 1% Glucose and supplemented with the appropriate antibiotic. Transferring 62.5 µl of overnight culture to the 24-well blocks diluted the overnight cultures. This was then shaken at 240 rpm at 37 °C for 3-5 hours (where the average OD_{600} (corresponding to an absorbance of 600 nm) reached ~0.5). At this point, shaking at 20 °C, 240 rpm for 20 mins cooled the cultures which were then induced by addition of IPTG to final concentration of 0.5 mM. Finally, the cultures were grown overnight with shaking at 200 rpm at 20 °C.

Expression screen with Overnight Express™ Instant TB Medium (Auto-Induction)

Deep-well blocks were prepared by addition of 0.5 mL of Power Broth™ supplemented with 1% Glucose and the appropriate antibiotic. Individual colonies were picked into each well and the plates sealed with gas-permeable adhesive seals. The filled plates were shaken at 240 rpm at 37 °C overnight. A 24-well deep well plate was then prepared, per strain screen, by addition of 2.5 mL of Overnight Express™ Instant TB Medium (TBONEX) (Novagen - Merck Millipore, Dermstadt, Germany) supplemented with the appropriate antibiotic. Transferring 62.5 µl of overnight culture to the 24-well blocks diluted the overnight cultures, with a 1 in 40 inoculation. This was then shaken at 240 rpm at 37 °C for 3-5 hours (where the average OD_{600} reached ~0.5). Lastly, the shaker temperature was reduced to 25 °C and protein expression and shaking continued at 200 rpm for a further 20-24 hours.

Ni^{2+}-NTA miniature expression screen

The cells were harvested by centrifugation at 5,000 xg for 15 mins, following which the media was decanted from the cell pellets. The cells were re-suspended completely, by repeated aspiration, into 210 µL of Lysis Buffer (NPI-10-Tween) supplemented with 1mg mL^{-1} Lysozyme and 400 units mL^{-1} of DNAse Type I. 30 mins were allowed for the action of the Lysozyme and DNAse before clearing the lysate by centrifuging the deep-well block at 5000 xg for 30 mins at 4°C.

The following method was carried out on the Qiagen BioRobot 8000, following the OPPF-UK, (2013) protocol. 20 µL of Ni-NTA magnetic bead suspension were dispensed into each well of a flat-bottomed 96-well micro-titre plate. The supernatant from the lysate was transferred to each well of the micro-titre plate containing the Ni-NTA magnetic beads. This was mixed for 30 mins at room temperature. The 96-well plate was placed on the 96-well magnet for one minute and the supernatant removed carefully from the beads. 200 µL of Wash Buffer (NPI-20) was added to each well, removed from the magnet and mixed on the micro-plate shaker for 5 mins. Then, place back onto the 96-well magnet for a minute, and the buffer removed. This washing step was repeated. 50 µL of Elution Buffer (NPI-250) was added to each well, mixed on the micro-plate shaker for one minute, placed on the 96-well magnet for one minute, and the eluates were transferred to a fresh micro-titre plate for analysis on SDS-PAGE.

2.2.5.2 SilE Protein production and purification

BL21 Star™ (DE3) or Rosetta Star™ (DE3) pLysS (Table 2.1) competent *E. coli* cells transformed with the pOPINF expression vector for $SilE_{21\text{-}143}$ ($SilE_{21\text{-}143F}$) were streaked onto plated selective LB agar. These inoculated LB agar plates were then incubated overnight, at 37 ^{o}C.

Protein overexpression trial of SilE

Ten mL selective, LB medium was inoculated with colonies of $SilE_{21\text{-}143F}$ transformed BL21 Star™ (DE3) or Rosetta Star™ (DE3) pLysS *E. coli* cells and grown overnight.

The OD_{600} of the grown, overnight cell culture was noted and the culture centrifuged at 3,500 xg for 15 mins (4 ^{o}C). The supernatant was removed and discarded and the pelleted sample was then re-suspended in 1 mL of LB medium, now giving a ten times greater OD_{600}. From this, a quantity of cells required to achieve an initial OD_{600} of approximately 0.1 was calculated and used to grow new cell cultures containing 50 mL selective, LB medium (Hall *et al.*, 2014). These newly inoculated cultures were then grown until an OD_{600} of between 0.6 and 0.8 (i.e. mid-log phase) was acquired. At this point, glycerol stocks of the $SilE_{21\text{-}143F}$ transformed BL21 Star™ (DE3) or Rosetta Star™ (DE3) pLysS *E. coli* cells were prepared, to maintain the cloned construct.

The expression of SilE was induced at mid-log phase (Renshaw *et al.*, 2002), by the addition of IPTG at a final concentration of 1 mM. Hourly samples were taken over the next four hours, and an overnight sample was

taken with the OD_{600} of each sample being recorded. In order to analyse these timed samples using SDS-PAGE, they were normalised to an OD_{600} of 0.8. This ensured that the same amount of cells were in each sample, allowing protein expression to be compared accurately, with minimal error.

Large-scale protein overexpression of SilE proteins

Two overnight starter cultures were initiated in 100 mL selective, LB medium, through the addition of BL21 Star™ (DE3) or Rosetta Star™ (DE3) pLysS *E. coli* cells transformed with $SilE_{21\text{-}143F}$.

The overnight cell cultures were pooled, giving a total culture volume of 200 mL and the OD_{600} was noted. The pooled cultures were then split into four equal quantities, of 50 mL before being centrifuged for 15 mins at a speed of 3,500 xg, at 4 ^{o}C. The supernatant was removed and discarded and each of the four pelleted cells was re-suspended in 500 µL of selective LB medium thereby giving an OD_{600} 100 times greater than originally noted. An appropriate quantity of cells were taken from each of the four suspensions and used to inoculate four 500 mL, selective LB cultures to a starting OD_{600} of 0.1. These were grown until an OD_{600} of around 0.6-0.8 was obtained. The cultures were then induced with IPTG at a final concentration of 1 mM and allowed to grow overnight.

The four 500 mL, IPTG induced cell cultures were pooled into two sets of 1 L. The two culture sets were centrifuged at 5,000 xg for 30 mins, at 4 ^{o}C. The supernatant was thrown away and the two pellets were scooped

and transferred into two 50 mL falcon tubes (one per pellet derived from 1 L of cells) and stored in the freezer, at -20 ^{o}C.

^{15}N labelled protein expression

A uniformly ^{15}N labelled sample of SilE was prepared from the pOPINF *E. coli* based expression vector, in *E. coli* BL21 Star^{TM} (DE3) cells. This is because; NMR requires the atomic nuclei of molecules to have magnetic spin properties in order to identify atoms that are close together in space. These properties are only possessed by certain isotopes including ^{15}N and are classified as being NMR 'active' (Howard, 1998). Therefore, for the ^{15}N labelling of SilE, the cells were grown and protein was expressed in the same manner as described for unlabelled protein; but (following the starter culture growth in LB) they were then grown in M9 minimal media instead of LB. This is due to the fact that the minimal media contained 1g L^{-1} ^{15}N-$(NH_4)_2SO_4$ as the sole nitrogen source and glucose as the only source of carbon (Marley *et al.*, 2001). Also following induction with IPTG, the growth temperature was reduced to 30 ^{o}C.

Extraction of protein SilE

The harvested pellet after overexpression was re-suspended in 20 mL Lysis Buffer, supplemented with 1 mL per 20 g of cells of DMSO solution protease inhibitor cocktail (for use in purification of His-tagged proteins) and incubated on ice for 20 mins. The cells were then sonicated using a 9.5 mm solid probe fitted to a Soniprep 150 Ultrasonic Disintegrator

(MSE UK Ltd., London, UK) with an output frequency of 23 kHz. 12 cycles of 20 seconds sonication at an amplitude of 10 microns followed by 30 seconds of recovery were performed. The lysate was then centrifuged at 16,000xg for 30 mins, at 4 ^{o}C. The supernatant was decanted, to remove any cell debris and precipitate, to a fresh falcon tube and passed through a 0.2 μm syringe filter, recovered and the soluble lysate stored at 4 ^{o}C ready for purification.

SilE protein purification

Protein purification was carried out using the techniques of nickel affinity chromatography and size-exclusion (gel filtration) chromatography. Nickel affinity chromatography exploits the ability of His to bind nickel, through natural interactions. Gel filtration relies on separating proteins according to their shape and size – larger proteins are eluted first, whereas smaller, more compact proteins are able to weave between the pores of the column. Due to the lack of light absorbing aromatic tryptophan and tyrosine residues in SilE, absorbance was followed at 254 nm.

Nickel affinity chromatography

The soluble lysate was loaded via a super-loop, through an ÄKTApurifier machine (GE Amersham Pharmacia, Buckinghamshire, UK) onto a HisTrap^{TM} HP (5 mL), nickel sepharose column (nickel cations chelated to a sepharose support). Prior to sample loading, the column was manually washed using a syringe with 30 mL of sdH_2O to remove the

ethanol used for storage of the column. The resin was charged with nickel by manually washing the column with 30 mL of 100 mM $NiCl_2$. The His_6-tag of SilE would have bound Ni^{2+} ions with a high affinity and therefore, His_6-tagged SilE would have remained attached to the Ni^{2+}-sepharose column matrix whilst all of the contamination passed directly through the column and into the flow through. With HisTrap™ HP Column Loading/Wash Buffer containing a low concentration (50 mM) of imidazole, non-specific binding to the matrix was prevented (Sigma-Aldrich, 2000). The His_6-tagged SilE was then eluted from the column by gradually adding a stepwise gradient of Elution Buffer. His_6-tagged SilE elution would have occurred as a result of the increasing imidazole concentration (up to 500 mM) out-competing the protein's His_6-tag (due to the side-chain of histidine residues also containing an imidazole ring). A flow rate of 1 mL min^{-1} was maintained throughout. Samples of the flow through and major peak fractions were taken to be analysed by SDS-PAGE.

Gel filtration chromatography

This process of size-exclusion chromatography was carried out on an ÄKTAprime plus machine (GE Amersham Pharmacia, Buckinghamshire, UK). A HighLoad™ 16/60 Superdex™ 75 pg gel-filtration column was conditioned with Gel Filtration Buffer. Subsequently, 5 mL of sample was injected onto the column via a 10mL sample loop. If necessary, the protein was concentrated to reduce volume to below 5 mL, using a VivaSpin centrifugal spin concentrator. Elution was carried out at 1

mL min^{-1} using the same buffer. The resultant fractions were pooled following SDS-PAGE analysis.

HRV-3C protease treatment and purification SilE

The His_6-tag was cleaved from the SilE protein using HRV-3C protease - a stringent sequence specific Humanrhino Virus protease that cleaves at a highly specific recognition site, with a consensus sequence of LEVLFQ/GP to remove the affinity tag (Table 2.2) and leave behind the target protein. Thus, 200 µL of HRV-3C protease stock (prepped in-house at 2 mg mL^{-1}) was added to the protein sample and mixed together. The protein and HRV-3C protease mixture was kept for 30 mins and then dialysed overnight to remove any excess imidazole, at 4°C, against Wash Buffer.

The cleaved His_6-tag and His_6-tagged HRV-3C protease were eliminated from the protein by a second round of passage through a HisTrap^{TM} HP (as described earlier). During this ÄKTApurifier run, the SilE protein was collected first, as a flow through directly passing, with contaminants - HRV-3C protease and His_6-tag - being eluted with a gradient of Elution Buffer being added. Both the flow through and contaminants were analysed by SDS-PAGE and protein was flash frozen prior to storage at -80 ^{o}C.

Protein buffer exchange and spin concentration

Due to the silver ion (Ag^+) being readily able to form compounds that precipitate (i.e. the white precipitate AgCl), thereby reducing the bioavailability of Ag^+, assorted methods for buffer exchange were used to prepare the sample for future experiments.

Spin concentration

In order to decrease protein solution volume for gel filtration and desalting chromatography, as well as concentrate the protein to higher concentrations, various Vivaspins were used. Proteins were spin concentrated (in a bench top centrifuge) using a Vivaspin 20, 2 or 500 (GE Healthcare UK Ltd., Buckinghamshire, UK) (with maximal volumes of 20 mL, 2 mL or 500 µL respectively). The spin concentrators had a 5,000 Da molecular weight cut-off (MWCO) (less than half the molecular weight of SilE), with a polyethersulfone membrane to remove small molecule and protein contaminants. The spin concentration was conducted at the manufacturers recommended centrifugal force (xg) for each concentrator size.

Desalting

Pre-packed PD-10 Desalting Columns (GE Healthcare UK Ltd., Buckinghamshire, UK) were used for convenient buffer exchange, clean-up and removal of salt (especially Cl^-) from protein samples. PD-10 Desalting Columns work in a similar principle to size exclusion chromatography, in

which the Sephadex G-25 medium allows rapid, efficient group separation of higher molecular weight molecules from lower ones. Group separation was conducted by following the 'Gravity Protocol', whereby the liquid sample passes through the column by gravity force. The columns were pre-equilibrated with new salt-free buffer and 2.5 mL of protein sample solution was loaded - the flow-through was collected and discarded. The protein was eluted with 3.5 mL of the new buffer (followed by NaCl) and presence of the protein in the eluate confirmed by SDS-PAGE.

Dialysis

Dialysis was used to dialyse in high concentrations of silver nitrate ($AgNO_3$ - up to 500 mM) for nuclear magnetic resonance (NMR) and to ensure perfect matching for small-angle X-ray scattering (SAXS). SilE was dialysed using 3,500 Da MWCO SnakeSkin^{TM} Dialysis Tubing (Thermo Fisher Scientific UK Ltd., Loughborough, UK), at 4 ^{o}C at first into 2 L of buffer containing $AgNO_3$ for four-eight hours. Followed by removal of excess ions by two buffer changes of four-hour cycles of dialysis in buffer without $AgNO_3$. Dialysis with buffer containing $AgNO_3$ was always covered with foil, in assurance of avoiding exposure to sunlight (UV light) and colour change to brown.

Protein yield quantification

$SilE_{21\text{-}143F}$ lacks UV light absorbing (ca. 280 nm) aromatic residues such as tryptophan, tyrosine therefore, its concentration was estimated via

two alternative methods. Firstly, acquiring the Brix coefficient using an Atago DD-7 Digital Differential Refractometer (Atago Co., Ltd., Tokyo, Japan) allowed calculation of the protein concentration using the following formula (Eq. 2.2) as explained elsewhere (Theisen *et al*., 2000):

(dndc sucrose/dndc protein sample) x 7.8883 x Brix coefficient

= (0.15/0.18) x 7.8883 x Brix coefficient *(Eq. 2.2)*

Secondly, by measuring the absorbance of the peptide bond at the UV wavelength of 205 nm on a Thermo Scientific NanoDrop 2000c spectrophotometer (Thermo Scientific, Wilmington, USA) (Scopes, 1974; Stoscheck, 1990). The absorbance value was entered into the Beer Lambert Law (Eq. 2.1) after determining the protein extinction coefficient at A_{205} (ε_{205}), (Anthis and Clore, 2013). Both calculations gave similar concentration values in the mg mL^{-1} range.

2.2.6 Biophysical methods

The focus of this thesis is predominantly on the periplasmic protein SilE and it's interaction with ligand Ag^{+}. In order to study protein folding, structure and it's behaviour upon ligand-binding different biophysical techniques can be used to provide complementary information about the binding event. Data obtained from these techniques report on the average properties of a protein sample. One advantage of these techniques is that

proteins can be studied directly under physiological conditions without any chemical modifications. A description of the variety of techniques used throughout the development of this investigation follow.

In order to obtain information on the SilE protein's metal binding, folding, structural features and properties, the biophysical experiments were carried out include: two forms of optical spectroscopy experiments, far-UV circular dichroism (CD) and small angle X-ray scattering (SAXS); nuclear magnetic resonance (NMR); nano-electrospray ionisation (nano-ESI) and inductively coupled plasma (ICP) mass spectrometry's (MS). Throughout these experiments, the sources of the metal ions were always soluble and excess ions were removed via extensive dialysis or using a Vivaspin device. The metal ion salts used were as follows:- Ag^{+} from $AgNO_3$, Cu^{2+} from cupric sulphate ($CuSO_4$), (Zn^{2+}) from zinc chloride ($ZnCl_2$) and Ni^{2+} from nickel chloride ($NiCl_2$).

2.2.6.1 Far-UV circular dichroism spectroscopy

Circular dichroism (CD) spectroscopy in the far-UV region is an absorption-based technique that is one of the most widely used methods for determining the secondary structure of proteins (Johnson, 1990; Sreerama and Woody, 2004; Kelly *et al.*, 2005). CD is a phenomenon, which measures the difference in absorption between counter-clockwise (left-handed) and clockwise (right-handed) circularly polarised lights in chiral chromophore containing molecules (Shapiro *et al.*, 1995). It is defined as Eq. 2.3:

$$\Delta\varepsilon(\lambda) = \varepsilon_L(\lambda) - \varepsilon_R(\lambda) \qquad \textit{(Eq. 2.3)}$$

Where ε_L and ε_R are, respectively, the extinction coefficients for the left- and right-circularly polarised components at wavelength, λ (Sreerama and Woody, 2004). Polarised light can be linear or circular (Figure 2.2). Linearly polarised light is light whose oscillations are confined to a single plane. Whereas, circularly polarised light occurs when the direction of the electric field vector rotates about its propagation direction. When polarised light passes through an asymmetric molecule, elliptically polarised light is generated – this is known as the CD signal. Therefore, a CD spectrum is a plot of $\varepsilon_L - \varepsilon_R$ against λ (Greenfield, 2007a).

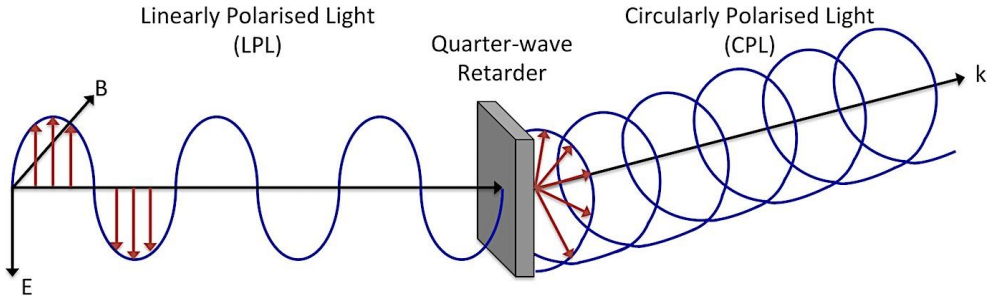


Figure 2. 2 Circular polarisation of light

Electromagnetic radiation consists of an electric (E) and magnetic (B) field, which oscillate perpendicular to one another and the propagating direction (k). At a given point, two orthogonal waves are in phase. The waves are superposed thus the resultant wave/electric field vector oscillates in one plane and is linearly polarised (LPL). Even though the amplitudes may differ, the waves must be in phase. The quarter-wave reader shifts the phase and the resultant wave is circularly polarised (CPL), where the electric field vector rotates about its propagation direction.

Figure 2.3 shows how a CD signal is measured using a spectrolarimeter. This signal can be positive or negative, depending on whether left-circularly polarised light is absorbed to a greater (CD signal positive) or a lesser (CD signal negative) extent than right-circularly polarised light. Positive or negative CD signals are presented in molecules that comprise structural asymmetry; whereas symmetrical structures exhibit a CD signal of zero. Due to the asymmetric chiral centres found in amino acids, proteins are optically active molecules that emit a CD signal. There are two types of optically active chromophores in proteins: side groups of aromatic amino acid residues, and peptide bonds. Therefore, protein tertiary structure analysis can be completed by near-UV CD measurements at the wavelength range 250-300 nm (Kelly *et al.*, 2005). At this wavelength range, side chains of aromatic amino acids (i.e. Trp and Tyr) and disulphide

bonds dominate the protein absorption spectra. Nonetheless, different secondary structures within proteins, due to the peptide bond, give rise to distinct CD signals in the far-UV (190-250 nm) region (Sreerama and Woody, 2004; Greenfield, 2007a).

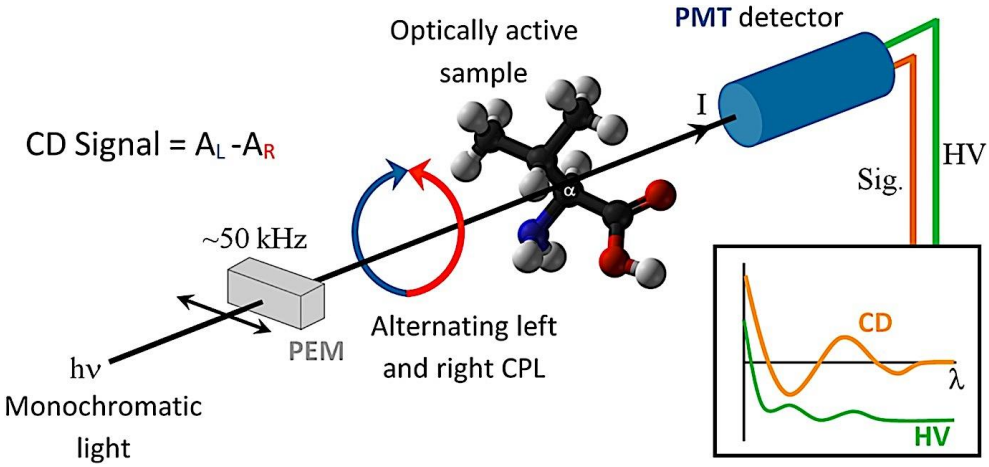


Figure 2. 3 Experimental set-up of CD spectroscopy

Monochromatised light from the beamline CD1 is passed through a Photo Elastic Modulator (PEM), which converts the linear polarised light into alternating left and right, handed polarised light. The two polarisations are differently absorbed, and the difference in absorption is detected with a Photo Multiplier Tube (PMT). Imagine obtained and used with publishers permission from:

http://www.isa.au.dk/facilities/astrid2/beamlines/AU-cd/AU-CD_3.asp

Different secondary structure elements produce characteristic far-UV CD spectra that can be easily recognised, when amide (NH) polypeptide bonds are aligned in regular arrays such as α-helices, β-sheets, β-turns or irregular unordered chains (Figure 2.4). Proteins with large α-helical content are associated with negative CD bands at 222 nm and 208 nm and a positive signal at 190 nm. The CD signal at 222 nm is roughly proportionate to the percentage of residues in an α-helical environment. Whereas CD spectra for proteins with well defined β-sheets/turns display a broad negative band around 215 nm and a positive ellipticity at 195 nm. Unstructured proteins, such as intrinsically disordered proteins (IDPs), that lack any significant organised secondary structure give the characteristic CD spectrum of an unordered polypeptide or random coil, with strong negative ellipticity near

200 nm (Dodero *et al.*, 2011). Thus, IDPs are readily recognised by CD. Typical proteins contain most or all of these conformation types, hence methods such as DichroWeb - 'on-line analysis for protein circular dichroism spectra' and BeStSEl (Beta Structure Selection) have been developed to analyse the CD data obtained from a protein to determine the fraction of residues in each conformation (Lobley *et al.*, 2002; Wallace *et al.*, 2003; Whitmore and Wallace, 2004; Sreerama and Woody, 2004; Kelly *et al.*, 2005; Miles *et al.*, 2005; Whitmore and Wallace, 2008; Micsonai *et al.*, 2015).

Since the secondary structure elements possess such distinguishing CD spectra, as well as probing the secondary structure of the protein itself, CD provides an experimentally convenient means of detecting structural changes in proteins by the binding of ligands (Rodger *et al.*, 2005). In addition, the unfolding of proteins can also be monitored using far UV-CD. Unfolding profiles can be collected as a function of temperature (Greenfield, 2007b). Finally, this technique can also be used to evaluate any structural alterations caused by a mutation. It is for these reasons that far-UV CD is a method frequently used throughout this study.

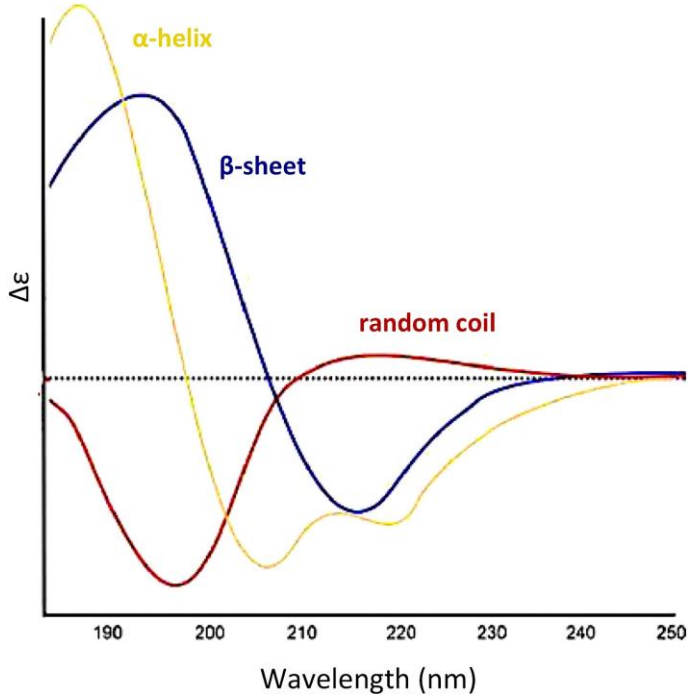


Figure 2. 4 Characteristic CD spectra for secondary structure

The 'pure' protein CD spectra for α-helical (yellow) β-sheet (blue) and random coiled (red) secondary structure - all produce distinctive lines. Image adapted and presented following publishers permission, from Dodero, Quirolo and Sequeira, 2011.

Far-UV CD experimental method

Far-UV CD was used to determine the secondary structure of *apo*- and *holo*-SilE (metal ligand free and bound, respectively) in solution. CD experiments were conducted at CD1 beam line at the ASTRID2 storage ring facility at Aarhus University, Aarhus, Denmark (Hertel and Hoffmann, 2011). Data was acquired at 25 ^{o}C from 50 µL *apo*- and metal ligand ion (Ag^{+}, Cu^{2+}, Zn^{2+}, Ni^{2+}) bound-SilE protein samples at 57 or 75 µM, in a quartz suprasil cylindrical cell (Hellma type 121.000) in a 10 mM HEPES (4-(2-hydroxyethyl)-1- piperazineethanesulphonic acid), 20 mM NaF buffer at pH 7.5, with either 2 mM or titrated quantities (1-16 equiv. Ag^{+}) of metal ions added to the *holo* samples.

Spectra were recorded from 170 to 280 nm, with the protein sample being in a 0.05 cm path length cell; scan speed of 20 nm min^{-1} and a response time of 1 s, with each spectrum representing an average of three accumulations, with an average of 15 scans per point. A scan of buffer alone was subtracted from the protein curve. The raw data were converted and plotted as delta epsilon (Δε - expressed in mdeg M^{-1} cm^{-1}) and spectra analysis was carried out by comparing the profile of the obtained curve to those illustrated and quantified in literature (Sreerama *et al.*, 1999; Greenfield, 2007a; Dodero *et al.*, 2011). Secondary structure percentages were calculated using the DichroWeb (Lobley *et al.*, 2002) interfaces analysis programme CONTINLL, which implements the locally linearised algorithm in selecting protein sets from the reference database (Provencher and Glockner, 1981; Van Stokkum *et al.*, 1990; Sreerama and Woody, 2000).

Thermal unfolding experiment

The stability of *holo*-SilE (2 mM Ag^{+}) to temperature denaturation was tested and determined by following changes in the CD spectra. The changes in the intensity of the maximal negative signal at 207 nm for *holo*-SilE (2 mM Ag^{+}) were recorded as a function of increasing temperature from 18 to 80 ^{o}C. The temperature was gradually increased at increments of 1 ^{o}C per minute and protein samples were allowed to equilibrate at each temperature, prior to recordings at intervals of 5 ^{o}C. In each case, spectra were acquired from 100 µL protein samples at 57 µM. The CD data was converted to a percentage change of the maximum CD (Δε).

2.2.6.2 Mass spectrometry

Mass spectrometry (MS) is a sensitive, analytical, chemistry-based technique used to help identify the quantity and species of chemical present, at physiological concentrations, in a sample. A mass spectrum is derived as a function of measuring the mass-to-charge ratio (m/z) and abundance of gas-phase ions. In turn, allowing for accurate determination of a samples molecular mass. While MS has been a distinguished analytical tool for some time, intact biomolecular analysis by MS only became possible in the 1980's upon the invention of two of the most widely used soft ionisation techniques – Matrix-Assisted Laser Desorption/Ionisation (MALDI) and Electrospray Ionisation (ESI) (Aebersold and Mann, 2003). With ESI and MALDI, fragile, involatile, macromolecules such as proteins are ionised and transferred into the gas-phase, from solution, without fragmentation. In recognition for their contribution and achievement in this, both Tanaka and Fenn shared the Chemistry Nobel Prize in 2002 (Tanaka, 2003; Fenn, 2003). Out of these two, only ESI was used throughout this work, alongside the earlier established Inductively Coupled Plasma (ICP) –MS technology, which is used for elemental analysis. Therefore, the following sections will briefly discuss these processes.

ESI-MS

ESI is a versatile and soft-ionisation technique that enables fragile

protein complexes to be transferred into the gas phase as intact ions, even if the interaction is weak. This means that besides deducing information on molecular mass, a range of other information including stoichiometry of non-covalent protein complexes and affinity interactions within protein complexes can be determined (Hernandez and Robinson, 2007); however very little structural information can be gained.

Common protein storage or purification buffers (Tris–HCl, HEPES, phosphate buffers, etc.) contain non-volatile electrolytes, which cause suppression of useful ESI signal and/or extensive adduct formation. Therefore, most of the protein solutions need to be buffer exchanged to ESI-friendly solvent systems (Table 2.11) prior to mass analysis (Abzalimov *et al*., 2012). The buffer of choice in this study was ammonium acetate. This is a particularly useful buffer as it can aid to prevent salt adducts. It does this by interacting with acidic and basic groups on proteins. The positive ammonium cations interact with the negative functional groups and the negative acetate anions interact with the positive groups within a protein. This interaction does not lead to adducts in the final spectrum as, once the sample reaches the heated capillary of the ESI-MS machine and the solvent starts to evaporate, the ammonium and acetate ions are volatile and will dissociate from the proteins via proton transfer reactions (Banerjee and Mazumdar, 2012).

Table 2. 11: ESI-friendly buffers

Names and chemical formulas of the volatile, ESI-friendly buffers (Abzalimov *et al*., 2012).

Name	Chemical Formula
Ammonium acetate	CH_3COONH_4
Ammonium bicarbonate	NH_4HCO_3
Ammonium trifluoroacetate	CF_3OONH_4
Ammonium hydroxide solution	NH_4OH
Methanol	CH_3OH
Formic acid	HCOOH
Ethanol	C_2H_5OH
Acetonitrile	CH_3CN
Acetic Acid	CH_3COOH

A mass spectrometer consists of three principal components – the ion source, the mass analyser and the detector (Figure 2.5). First the liquid samples, dissolved in a volatile solvent, are forced through a narrow steel needle or capillary into an ionisation source. Here, a high voltage is applied and the liquid inside the needle becomes unstable as it tries to retain more and more charge; until it can no longer sustain any more charge – the Rayleigh limit. This results in the liquid assuming a fine jet in the form of a conical shape (the Taylor cone), which consequently undergoes the Coulombic explosion and disperses as a plume of small highly charged droplets. These droplets decrease in size, whilst traversing between the needle tip and sampling cone, as solvent continuously evaporates from them - resulting in ions being produced and released. A nebulising gas, i.e. a warm flow of nitrogen (drying gas) passing across the front of the ionisation

source, also assists in the desolvation process. This process repeats itself until the radius of the droplets, free from solvent, nears the size of a single protein molecule. Multiple charged protein ions are produced from these nanodroplets and a spectrum of the multiple charge states is visualised and separated according to their mass-to-charge ratio (m/z). Therefore, the data appears in sets of charge state distributions.

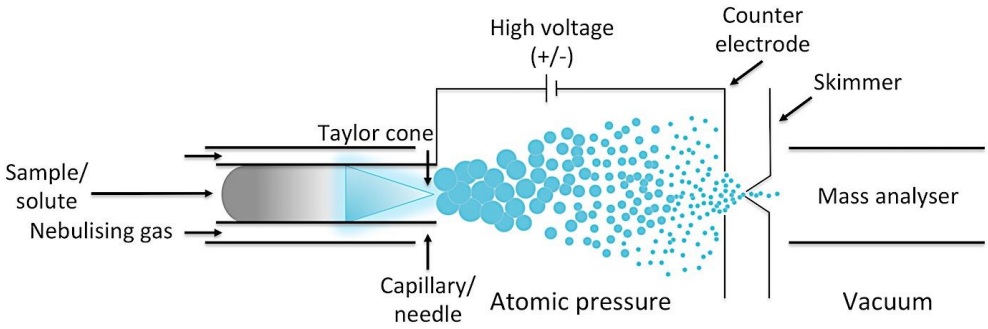


Figure 2. 5 Schematic of an ESI source

A diagrammatic depiction of the positive ion mode ESI process. Protein solution (sample/solute) is sprayed from metal capillary (blue) and a Taylor cone is formed as a result of the applied high voltage. The multiple droplet shrinkage and fission processes form multiple charged protein ions that travel through the skimmer before being detected in the mass analyser.

The m/z of that ion can then be used to calculate the charge state and subsequently define the mass of the molecule (M) using Eq. 2.4, where m is the mass of the ion (i.e. molecular mass of the protein plus adducted protons), z is the charge state and H is the mass of the proton adducts:

$$M = z\,(m/z - H) \qquad \textit{(Eq. 2.4)}$$

ESI-MS has become a useful means in the study of protein complexes with other proteins and ligands. The fast data collection and requirement of only low sample concentrations (i.e. much closer to physiological concentrations) makes the technique extremely efficient.

ESI-MS experimental method

Experiments were carried out (under the support of Dr. Matthew Jenner, University of Nottingham) and spectra were recorded on a SYNAPT High Definition Mass Spectrometry (HDMS) (Waters UK, Hertfordshire, UK) a hybrid quadrupole ion mobility time-of-flight (TOF) MS instrument, with traveling-wave ion mobility (TWIM) capability, equipped with the standard z-spray source. The instrument conditions were optimized to provide the highest relative signals for *apo*- and 2 mM metal ligand ion (Ag^{+}, Cu^{2+}, Zn^{2+}, Ni^{2+}) bound-SilE complexes at a protein concentration of 25 μM, sprayed from 25 mM ammonium acetate ($C_2H_3O_2NH_4$), pH 7.0. The nano-ESI capillary voltage was 1.5 kV; cone voltage, 20 V; extraction voltage, 5 V; transfer voltage, 5 V. Other settings were as follow: trap and transfer collision voltage, 6 and 5 V, respectively; backing pressure, 1.6 - 1.8 mbar; trap pressure, 2.1 × 10−2 mbar; TOF region pressure, 1.5 × 10−6 mbar. Instrument control as well as data processing was carried out with the Waters MassLynx 4.1 data system. All spectra were acquired in ion positive mode and the TOF analyser operated on V-mode. Minimum smoothing and background subtraction was applied to the obtained spectra prior to analysis.

ICP-MS

ICP-MS is a preferred technique for trace metal analyses, through the combination of high-temperature ICP source with a mass spectrometer. It is a highly sensitive analytical technique that is capable of determining the multi-elemental content of samples. The ICP-MS instrument can measure most of the elements in the periodic table, with detection limits as low as parts per quadrillion (ppq). It is accomplished by detecting the number of ions present at a certain mass, in mostly liquid but solid and gas state samples also. This is due to the high ion density and temperatures of plasma providing a supreme atomiser. The plasma (made up by partially ionising argon gas (Ar^+)) contains a sufficient concentration of ions and electrons to make the gas electrically conductive (Batsala *et al.*, 2012). Thus the ICP converts the atoms of the elements within the sample into detectable ions. These ions are directed into the mass spectrometer, under vacuum, via two cones in the interface region and separated on the basis of their mass-to-charge ratio (Figure 2.5). The detector then receives an ion signal proportional to the concentration (Thomas, 2001).

ICP-MS experimental method

Ag^+ elemental analysis of 50 nM protein diluted, with and without Ag^+ after binding of Ag^+ and extensive dialysis to remove excess) in 5 ml solutions of 10 mM HEPES, 20 mM NaF, pH 7.5, in 1% nitric acid (HNO_3), was undertaken by inductively coupled plasma-mass spectrometry (ICP-MS) (Thermo-Fisher Scientific iCAP-Q; Thermo Fisher Scientific, Bremen, Germany). The instrument was run employing three operational modes,

including (i) a collision-cell (Q cell) using He with kinetic energy discrimination (He-cell) to remove polyatomic interferences, (ii) standard mode (STD) in which the collision cell is evacuated and (iii) hydrogen mode (H_2-cell) in which H_2 gas is used as the cell gas. Samples were introduced from an auto-sampler (Cetac ASX-520) incorporating an ASXpress^{TM} rapid uptake module through a PEEK nebulizer (Burgener Mira Mist). An internal standard Rh (10 μg L^{-1}) in 2 % trace analysis grade (Fisher Scientific, UK) HNO_3 was introduced to the sample stream on a separate line via the ASXpress unit. External multi-element calibration standards (Claritas-PPT grade CLMS-2 from SPEX Certiprep Inc., Metuchen, NJ, USA) included Ag, Al, As, Ba, Be, Cd, Ca, Co, Cr, Cs, Cu, Fe, K, Li, Mg, Mn, Mo, Na, Ni, P, Pb, Rb, S, Se, Sr, Tl, U, V and Zn, in the range 0 – 100 μg L^{-1} (0, 20, 40, 100 μg L^{-1}). A bespoke external multi-element calibration solution (PlasmaCAL, SCP Science, France) was used to create Ca, Mg, Na and K standards in the range 0-30 mg L^{-1}. Phosphorus, boron and sulphur calibration utilized in-house standard solutions (KH_2PO4, K_2SO_4 and H_3BO_3).

In-sample switching was used to measure B and P in STD mode, Se in H_2-cell mode and all other elements in He-cell mode. Peak dwell times were 10 ms for the element with 150 scans per sample. Sample processing was undertaken using QtegraT^{M} software (Thermo Fisher Scientific, Bremen, Germany) utilising external cross-calibration between pulse-counting and analogue detector modes when required with data being acquired in μg $^{L-1}$. Protein concentrations were measured before and after the experiment. The protein and metal concentrations were determined as an

average of three independent experiments. All glassware and plasticware used for these experiments was washed with 10 % HNO_3 to remove contaminating metal. Samples were sent to Dr. Scott Young, ICP-Analytical Services (Sutton Bonington Campus, University of Nottingham) who performed the ICP-MS analysis and returned the raw data.

2.2.6.3 Nuclear magnetic resonance spectroscopy

Nuclear magnetic resonance (NMR) is a powerful field of structural biology used to determine the structure and dynamics of proteins and other biological macromolecules. Following the development of pulse Fourier transform NMR spectroscopy the growth of protein NMR, pioneered by Richard R. Ernst and Kurt Wüthrich, has been explosive (Cavanagh *et al*., 2006; Wüthrich, 2001). Today, the RCSB Protein Data Bank (PDB) contains almost 11, 500 structures solved by NMR (http://www.rcsb.org/pdb/statistics/contentGrowthChart.do?content=explMethod-nmr&seqid=100).

Alongside X-ray crystallography, NMR is currently the only other method that can be applied for the study of three-dimensional (3D) molecular structures of proteins at an atomic resolution. Being the only method that allows the determination of 3D structures of protein molecules in near physiological, solution-based conditions NMR spectroscopy can provide native-like data.

In contrast to other biophysical methods NMR elucidates atomic level chemical properties by studying individual nuclei. NMR can also

provide a wide-ranging amount of information on protein-protein and ligand interactions, kinetics and dynamics, folding and intra-molecular bonding patterns. Fundamentally, NMR spectroscopy uses the interaction between the 'little magnets' (nuclear spins) in atomic nuclei and radio frequency pulses in a magnetic field to obtain information about the nuclear environment. These atomic nuclei predominantly include: ^{1}H, ^{15}N and ^{13}C. Protein samples are thereby enriched with ^{15}N alone or ^{15}N and ^{13}C, via protein expression in bacteria grown in media with defined sources of nitrogen or carbon (Howard, 1998; Marley *et al.*, 2001). When placed in an external magnetic field (i.e. the NMR instrument) a nucleus spin of $I = ½$ can adopt two possible orientations or spin states: a lower energy level aligned in parallel with the magnetic field and the other higher energy level opposing the direction of the magnetic field (Figure 2.6). With the emission or absorption of energy, spins can then change orientation transitioning between the energy states. So differences in energy between the two spin states. This difference in energy corresponds to electromagnetic waves in the radio frequency of the spectrum, which also increases with an increasing strength of the magnetic field. The specific transition frequency of each atom within a molecule depends on its electronic environment. Electrons partially shield the nuclei from the external magnetic field, giving each nucleus a slightly different transition frequency depending on the local environment. It is these frequencies that are then observed as peaks in a NMR spectrum. The frequency of each peak is normalised against a reference and displayed on the spectra as parts per million (ppm). This normalised frequency is known as the chemical shift (δ).

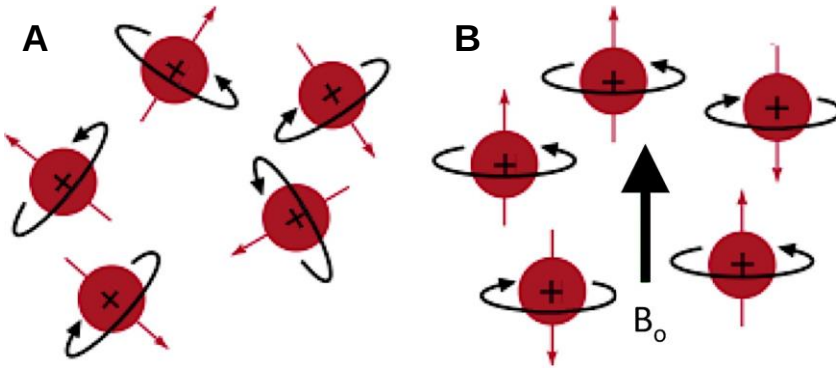


Figure 2. 6 Spins in a magnetic field

A representation of spins in the presence and absence of an external magnetic field - B_0. **A.** In the absence of an external magnetic field, the spins orient randomly. **B.** In the presence of an external magnetic field (B_0) spins align either with or against the direction of the field. More spins will align in the same direction as the external magnetic field, due to it being the lower energy state.

Traditionally, the use of NMR-based structure determination is restricted to smaller proteins. However, the former limitation has been alleviated by advances in technology involving the use of high-field spectrometers (now up to 1 GHz NMR magnets) and increased sensitivity via cryogenically cooled probes - significantly improving data quality and allowing for the study of larger proteins too (Kovacs *et al.*, 2005). The largest reported high resolution spectra of backbone NH resonances obtained to date has been derived from the 900 kDa complex of the co-chaperonin GroES bound to chaperonin GroEL (Fiaux *et al.*, 2002)!

One-dimensional NMR

A one-dimensional (1D), proton (^{1}H) NMR spectrum is the most commonly used and usually the first NMR experiment to be recorded, using non-isotopically labelled protein sample. ^{1}H is not only the most abundant isotope of hydrogen, but also ubiquitous in organic and biological molecules, interacting strongly with magnetic fields and generating good signal-to-noise (Cavanagh *et al*. 2006). Consequently, the ^{1}H experiment is used to check if the protein is suitable for further NMR-based investigations, by assessing sample purity. More importantly, a 1D spectrum can also provide information on the folding of a protein. Generally speaking, proteins with a highly ordered tertiary structure produce resonances with a wide dispersion of chemical shifts, whereas IDPs tend to produce spectra with significant overlap of resonances. More complex NMR experiments can be conducted, once the ^{1}H data has been collected, analysed and appears to be in a correct conformation.

One-dimensional NMR measurements

One-dimensional ^{1}H nuclear magnetic resonance (NMR) spectra were recorded at 25 °C on an 800 MHz Bruker Avance III NMR machine, using 500 μM protein in 10 mM sodium phosphate (Na_2HPO_4/NaH_2PO_4), 25 mM NaCl, pH 7.5 buffer with 10 % D_2O, at various pHs ranging from 5-9. The solvent water peak was attenuated using excitation sculpting or by pre-saturation.

Two-dimensional HSQC NMR

To separate peaks and facilitate analysis, proteins are most often studied using multidimensional NMR. The most commonly used experiment is the two-dimensional (2D) ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC), which correlates ^{1}H and ^{15}N chemical shifts and spreads the NH protons over another dimension, according to the chemical environment of the nitrogen atom to which they are covalently bonded (Figure 2.7). Accordingly, the chemical shift of a nucleus is quantified when a certain magnetic field is applied, which enables the measurement of the difference between the standard frequency and resonance frequency of nuclei (Cavanagh *et al*., 2006). Since the natural abundance of ^{15}N is too low to allow characterisation the protein is uniformly labelled with ^{15}N. The backbone of a protein contains an amide NH group for each amino acid thus a chemical shift or single peak is seen for each residue in a ^{1}H-^{15}N HSQC experiment (except for the N-terminal residue which has NH_3^+ protons exchanging rapidly with solvent and prolines that lack a backbone NH group, due to the missing hydrogen unable to act as a hydrogen bond donor) (Pascal, 2008). Peaks are also observed for any side chains with nitrogen-bound protons (such as tryptophan, glutamine and asparagine as well as negative peaks of arginine, lysine).

The ^{1}H-^{15}N HSQC spectrum corresponds to a unique fingerprint for each protein. Upon binding to another protein or ligand, any chemical shift perturbations are associated with a change in protein conformation due to the residue's change in environment, indicating a potential direct

interaction. Hence, HSQCs are considered to be a powerful tool to monitor ligand-binding or protein interaction-based changes (Scrofani *et al.*, 1998). In addition to studying protein dynamics, this spectrum is also the basis for 3D experiments used for backbone and side chain assignment for structure determination.

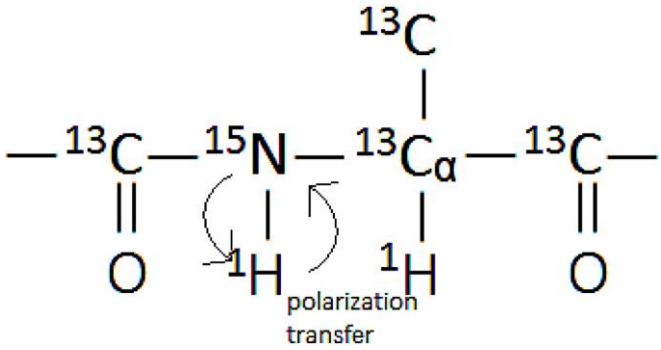


Figure 2. 7 **^{1}H-^{15}N HSQC NMR magnetisation transfer**

The ^{15}N nucleus can be detected indirectly by transferring polarisation through a ^{1}H nucleus. Magnetisation is first transferred from ^{1}H attached to ^{15}N. The chemical shift is evolved on the ^{15}N and the magnetisation is then transferred back to the 1H for detection.

Two-dimensional NMR measurements

Two-dimensional NMR based ^{15}N/^{1}H Heteronuclear Single Quantum Coherence (HSQC) spectra were acquired at 25 ^{o}C, on an 800 MHz Bruker Avance III spectrometer. Data were collected from 600 µl samples of 300-500 µM ^{15}N labelled sample of *apo*- and *holo*-Ag^{+}-bound SilE in 10 mM sodium phosphate (Na_2HPO_4/NaH_2PO_4), 25 mM NaCl, pH 7.5 and 10 mM HEPES, 20 mM NaF buffer at various pHs ranging from 9.0

to 2.0, 10 % D_2O, with and without Ag^+. NMR data was processed and analysed using Topspin package (Bruker, UK).

2.2.6.4 Small-angle X-ray scattering

For over 80 years, small-angle X-ray scattering (SAXS) has been a useful research tool in structural biology (Jacques and Trewhella, 2010). The analytical method determines low-resolution 3D structures and allows investigations in structural changes of a range of biological molecules in solution; from small peptides to large macromolecules (Svergun, 1999).

Figure 2.8 shows the principles of SAXS, in which the elastic collisions occurring between the incoming X-rays and a particle cause reflected waves to scatter in all directions. As the reflected waves interfere with one another, there will be constructive interference along certain angles, causing a scattering pattern. The scattering intensity (I), of a monodisperse sample exposed to X-rays, is recorded as a function of the scattering angle (q, Eq. 2.5).

$$q = 4\,\pi sin\theta/\lambda \qquad (Eq.\ 2.5)$$

Where, λ is the wavelength of the incident radiation, and θ is half the angle between the incident and scattered radiation (Figure 2.8).

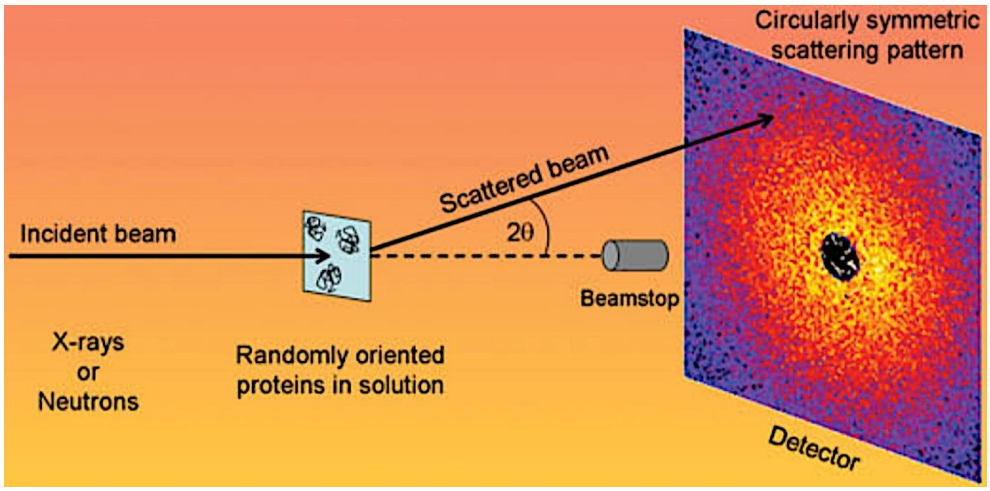


Figure 2. 8 Schematic of SAXS experimentation

The diagram outlines the proceedings of a SAXS experiment. The X-ray source (a synchrotron) emits X-rays, which are directed to the protein sample. When the sample is hit, radiation is scattered off the protein molecule at a small angle. The scattering angle relates to q (Equation 2.5). The scattered waves are recorded on a 2D detector, where a scattering pattern is built and a 1D scattering curve is extracted. The beamstop blocks most of the beamline that does not scatter. Diagram adapted with publishers permissions from Jacques and Trewhella, 2010.

As measurements are taken from the sample in a liquid state, the protein is tumbling and randomly orientated so the recorded scattering intensity is isotropic. Therefore, the scattering is due to the rotationally averaged form and structure factors (representing intra-molecular and inter-molecular scattering, respectively) of individual molecules within the sample (Svergun *et al.*, 2001).

In addition to this introduction, further details on Guinier, Kratky, Porod and Pr analysis are discussed alongside the relevant data in Chapter 4, Section 4.2.4.2.

Like NMR, a useful advantage of solution scattering is the ability to study protein structure in near physiological conditions as well as being able to analyse structural changes in response to external parameters.

SAXS beamline data collection

SAXS data were collected, at beamline BM29 (ESRF, Grenoble) and beamline B21 (Diamond Light Source, UK), from 1.25-5mg mL^{-1} *apo*- and *holo*-Ag^{+}-bound SilE in 10mM HEPES buffer, pH 7.5. Data was collected at an X-ray wavelength of 1.54Å, and the SAXS cell was equilibrated to 20 °C. Data was taken in one-minute frames, and typically 2 frames were taken for each cell. Low angle portions of successive frames were compared for evidence for aggregation of the sample due to photo-damage. Data from three concentrations of the samples (an average of 18 scans per sample) were compared and examined for aggregation and the data set was merged, so the low angle scattering came from the most dilute sample. Data was analysed, using the applications Primus (Konarev *et al.*, 2003) and ScÅtter (Biosis.net, 2016).

CHAPTER 3

The *sil* Gene Cluster

3.1 Introduction

Plasmids have been influential factors in bacterial pathogenesis and their evolution over the last three billion years (Bennett, 2008). Bacterial plasmids have served as a scaffold in aiding promiscuous gene transfer and allowed bacteria to adapt to protect themselves against toxic chemicals such as silver (Ag^+), through various mechanisms of antimicrobial resistance. Acquisition of these antimicrobial resistance genes occurs via bacterial conjugation, transformation, transposable elements (See Chapter 1, Sections 1.1.3 and 1.4.4) and integron systems, in addition to phage infection. It is therefore important to characterise these plasmids in order to understand their mechanisms of resistance, their genetic arrangements and how they may have been acquired (Bennett, 2008; Mellata *et al.*, 2009). The development of cheap, next generation sequencing (NGS, also known as high-throughput sequencing) technology has allowed this and revolutionised microbiology in terms of understanding evolution and comprehending 'systems biology' (Hall, 2007).

In 1999, Gupta and his colleagues identified the *sil* locus in plasmid pMG101 and reported this region of the plasmid as the Ag^+-resistance determinant (Gupta *et al.*, 1999). This was based on limited use of sub-clones, sequencing, and homology to known metal resistance genes. An additional study showed that expression of this *sil* locus on pMG101 was constitutively high in comparison to that resident in other IncH plasmids, which did not confer Ag^+-resistance (Gupta *et al.*, 2001). In order to understand more about the mechanism of bacterial Ag^+-resistance (at a molecular level), conservation of its determinants and possibilities of why

sil gene expression is higher in pMG101; whole genome and plasmid DNA sequencing was carried out on the megaplasmid pMG101, obtained from *E. coli* J53.

This chapter therefore describes the genes proposed to confer bacterial Ag^{+}-resistance and how they may have spread. Additionally, the chapter describes how the *sil* genes proposed to be involved in the mechanism of bacterial silver resistance were cloned into plasmid expression vectors using high-throughput cloning methods and the proteins overexpressed in *E. coli*.

3.2 Results and discussion

3.2.1 Understanding the *sil* locus and its surrounding

The 182 kb megaplasmid pMG101 (of the IncH incompatibility group) from *Salmonella eneterica* serovar Typhimurium was the original source of the putative Ag^{+}-resistance determinant; the 14.2 kb *sil* locus (McHugh *et al.*, 1975; Gupta *et al.*, 1999). It is clear that the *sil* locus is broadly distributed amongst bacteria (Hao *et al.*, 2015; Staehlin *et al.*, 2016). This probable Ag^{+}-resistance determinant has been identified on both plasmid and bacterial chromosomal DNA of pathogenic bacteria. This includes genomes and/or plasmids of: *E. coli*, *Enterobacter cloacae subsp. cloacae*, *Cronobacter sakazakii*, *Cronobacter turicensis*, *Klebsiella pneumonia* and *Serratia marcescens* (Hobman and Crossman, 2015). More recent whole genome sequence analysis, of the *sil* locus and surrounding genes in a range of Gram-negative bacteria, has shown that the *pco* (plasmid copper resistance) genes are frequently adjacent to the *sil* genes (Hao *et al.*, 2015; Hobman and Crossman, 2015; Staehlin *et al.*, 2016). A general overview of this genetic arrangement often referred to as an island can be seen in Figure 3.1.

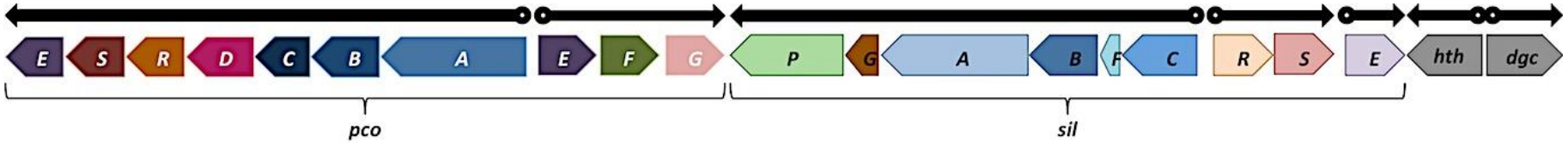


Figure 3. 1 **The *sil-pco* island**

A general overview of the *sil-pco* island arrangement, following analysis of:

Bacterial Species	Plasmid/Strain	GenBank accession number
Escherichia coli	Plasmid pAPEC-O1-R	DQ517526.1
	D6-117_07.11	CCCP010000042.1
	P12b	CP002291.1
Enterobacter hormaechei	ATCC 49162	AFHR01000042.1
Klebsiella pneumonia	Plasmid pLVPK	AY378100.1
	KPLM21	CCVM01000009.1
Salmonella enterica serovar Heidelberg	Plasmid pSH111_227	JN983042.1
Cronobacter turicensis	Plasmid pCTU3	FN543096.1
Serratia marcescens	Plasmid R478	BX664015.1

The alignment in the diagram shows the genes and their transcriptional and translational directions, with the open circles representing potential promoter regions/transcript start sites. The *pcoE* gene, found between *pcoA* and *pcoF* was not always found present in all bacterial strains. The *hth* and *dgc* genes were occasionally found in swapped positions, but this arrangement was more common. The *hth* gene is a putative helix-turn-helix restriction endonuclease and *dgc* is believed to be a diaguanylate cyclase enzyme. Details on the proposed functions of the *sil* and *pco* genes can be found in Chapter 1, Section 1.4.2.

3.2.1.1 Plasmid pMG101 sequencing

The megaplasmid pMG101 attracted attention for sequencing for a few reasons. Firstly, to analyse the presence of other AMR genes; secondly, to see if the common *sil-pco* island was present as one unit; thirdly, to assess the possibility that *sil-pco* is part of a transposable element; and finally to evaluate *sil* gene sequence conservation.

Plasmid pMG101 DNA was extracted from *E. coli* J53 (pMG101) (obtained from Professor Simon Silver, University of Illinois College of Medicine at Chicago, Chicago, Illinois, USA) using the QIAGEN® Large-Construct Kit (Qiagen, Crawley, UK) (Chapter 2, Section 2.2.4.1). This *E. coli* J53 (pMG101) strain had been stored at room temperature on sealed Dorset egg medium slope; a simple preservation method but one that does not guarantee the maintenance of a strain's viability and characteristics. Following its extraction, the identity of plasmid pMG101 was confirmed by verifying the presence of the *sil* genes, using PCR (Figure 3.2). Plasmid pMG101 was then sequenced by BGI Tech Solutions Co., Ltd. (Hong Kong) using an Illumina MiSeq machine. Data handling, assembly and annotation was performed by Dr. Lisa Crossman (SequenceAnalysis.co.uk, NRP Innovation Centre, Norwich Research Park, Norwich). Upon assembly, the total size of the sequenced plasmid pMG101 was found to be 165,548 bp, with nine scaffolds and 10 contigs (Table 3.1). This was approximately 17 kb less than the expected size of plasmid pMG101 (reported to be ~182 kb) (Gupta *et al.*, 1999).

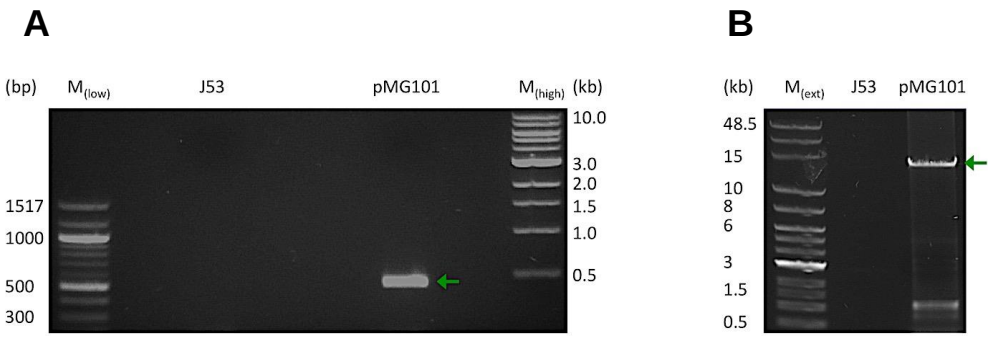


Figure 3. 2 Plasmid pMG101 verification PCRs

Agarose gels (Chapter 2, Section 2.2.2.1) showing gene amplification products of PCRs from plasmid pMG101. PCR products of **A**. the 429 bp *silE* gene amplicon electrophoresed on a at on a 1 % TAE gel and **B.** the *sil* locus amplicon of 14.2 kb on a 0.3 % TAE gel. Correct gene amplification sizes are highlighted by green arrows. The negative control (J53) showed no gene amplification and thereby no presence of the *sil* genes. The PCRs were carried out using the primers listed in Table 2.3, Chapter 2. The DNA molecular weight markers used were $M_{(low)}$ – 100 bp DNA Ladder, $M_{(high)}$ – 1kb DNA Ladder and $M_{(ext)}$ – Quick-Load® 1 kb Extend Ladder. (New England Biolabs, Ipswich, UK).

Table 3. 1: Assembly statistics for plasmid pMG101 sequencing contigs

Assembly statistics of contig for sequenced pMG101, provided by Dr. Lisa Crossman.

	Contigs
Total Number	10
Total Length (bp)	165,548
Max. Length (bp)	68,175
Min. Length	1,210
GC Content (%)	52.26

3.2.1.2 Plasmid pMG101 AMR gene analysis

As mentioned previously, plasmid pMG101 is reported to contain the *sil* locus, which is believed to confer Ag^+ resistance (Gupta *et al.*, 1999). Conjugative plasmids serve as a scaffold through which many antimicrobial resistance (AMR) genes can spread amongst bacteria (Bennett, 2008). In addition to gathering knowledge on the molecular basis of Ag^+-resistance, it was important to study the presence of other common AMR genes in pMG101. McHugh *et al.*, (1975) had identified the Ag^+-resistant strain containing plasmid pMG101, as carrying other antibiotic and metal ion resistances to: mercuric chloride, ampicillin, chloramphenicol, tetracycline, streptomycin, and sulphonamides (see Table 1.1, Chapter 1 for further information).

To gain a better understanding and to confirm the presence of the AMR genes expected to be present on the sequenced plasmid, pMG101, preliminary identification of putative AMR genes was carried out. This was done using two databases: CARD RGI (Comprehensive Antibiotic Resistance Database, (McArthur *et al.*, 2013)) and BacMet (Antibacterial Biocide & Metal Resistance Genes Database, (Pal *et al.*, 2014)). Following sequence submission, the CARD database uses the GeneMark software to identify open reading frames (ORFs). The RGI (Resistance Gene Identifier) tool in CARD then utilises a novel Antibiotic Resistance Ontology (ARO) and BLAST sequence similarity cut-offs to predict resistance genes by drug class and resistance mechanism. Figure 3.3 shows that the AMR genes present on pMG101, identified by CARD, were: *sul1*, bla_{OXA-1}, *tetD*, *aadA*, *catI* and *tetB* (Table 3.2).

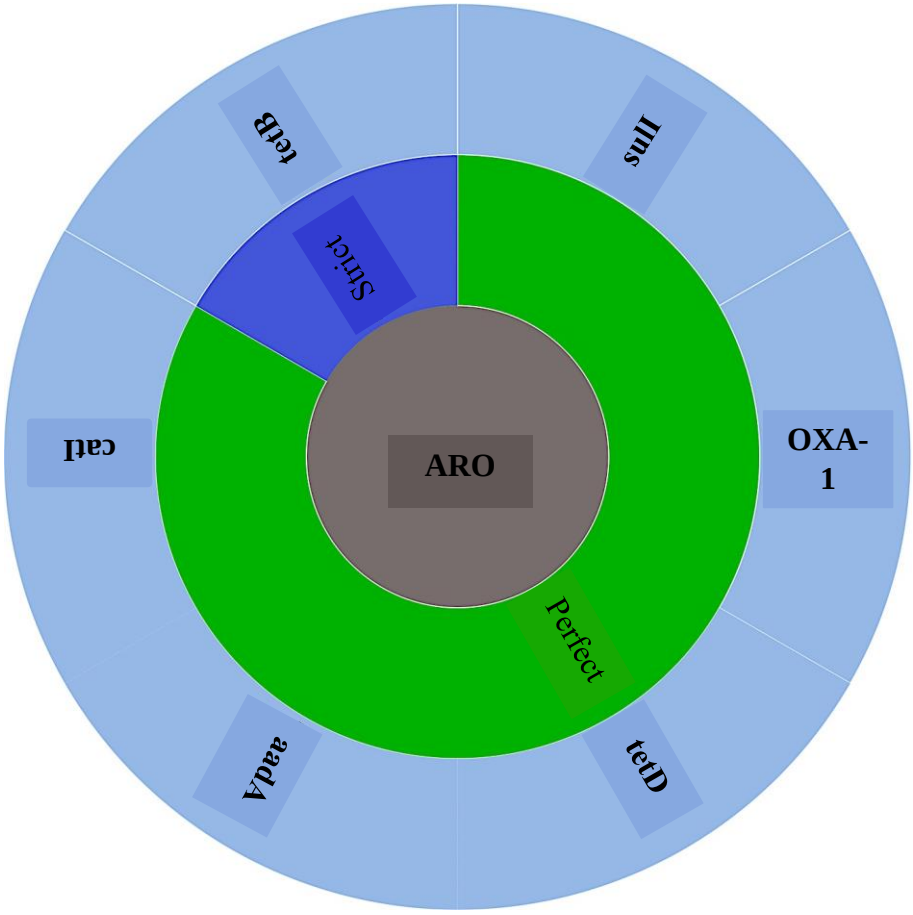


Figure 3. 3 RGI wheel of AMR genes in pMG101

RGI results produced in CARD highlighted the presence of resistance genes: sulphonamide-resistant *sul1*, β-lactam-resistant bla_{OXA-1}, tetracycline-resistant *tetB* and *tetD*, aminoglycosidase-resistant *aadA* and chloramphenicol-resistant *catI*. The 'Perfect' matches (green) were exact matches to AMR gene sequences within the database, while 'Strict' (dark blue) were unknown variants, including mutants of AMR genes within the CARD (McArthur *et al.*, 2013).

Table 3.2 provides more detail on the location, length and description of the AMR genes identified by CARD and BacMet. BacMet identifies AMR genes with the help of a bioinformatics resource of antibacterial biocide- and metal-resistance genes. The gene sequences that aligned significantly to resistance genes on the database are described in Table 3.2.

Table 3. 2: AMR genes carried by plasmid pMG101

The antibiotic- and metal-resistance genes found in pMG101 by databases CARD and BacMet. Once identified by the database, the genes were located by reviewing the sequence on Artemis (Carver *et al.*, 2012). The group of mercury resistant genes have been coloured blue.

Resistance Gene	Location	Length (bp)	Gene Description	Accession Number
AMR Genes Identified by CARD				
aadA1	5999-60778 (F)	780	Streptomycin adenylytransferase	AIC43574.1
***bla*$_{OXA-1}$**	59281-59874 (F)	594	β-lactamase OXA-1 precursor	WP_000881655.1
catAI	156785-157444 (F)	660	Type A chloramphenicol O-acetyltranferase	WP_050436713.1
sul1	61283-62122 (F)	840	Sulphonamide-resistant dihydropteroate synthase	WP_000451390.1
tetB	80066-81271 (F)	1206	Tetracycline resistance efflux pump	WP_012600019.1
tetD	82066-82482 (F)	417	Tetracycline resistance efflux pump	AAS17732.1

Resistance Gene	Location	Length (bp)	Gene Description	Accession Number
Additional AMR Genes Identified by BacMet				
merA	70873-72567 (R)	1695	Mercury (II) reductase	WP_042066333.1
merC	72619-73014 (R)	423	Mercury transport protein	WP_050858778.1
merD	70493-70855 (R)	363	Mercury resistance transcriptional regulator	WP_012644013.1
merE	70260-70496 (R)	237	Mercury ion trans-membrane transporter	EFO55297.1
merP	73077-73352 (R)	276	Mercuric transporter periplasmic component	WP_021538482.1
merR	73788-74222 (F)	435	Mercuric responsive resistance operon regulatory protein	KKA59047.1
merT	73366-73716 (R)	351	Mercuric transport protein	EFO55295.1
emrE	60942-61289 (F)	348	Multidrug resistance transporter protein	AMP54222.1

The sequenced pMG101 plasmid contains multiple resistance genes (Table 3.2), including those conferring resistance to mercury. All of the resistance genes that were found to be phenotypically present in the *S. Typhimurium* strain with plasmid pMG101 (McHugh *et al.*, 1975), were also found in the sequenced pMG101. However, none of the expected silver (*sil*) or copper (*pco*) resistance genes were found within this sequence. Manual nucleotide and amino acid sequence searches were also performed using Artemis (Carver *et al.*, 2012), to find the *sil* and *pco* genes, all without success. This suggested that neither the *sil* or *pco* genes were present in the sequence of plasmid pMG101; explaining why the sequenced plasmid size was ~17 kb shorter than the published size of ~182 kb (Gupta *et al.*, 1999). It seems likely that the *sil* gene PCR products generated from the pMG101 plasmid preparation (shown in Figure 3.2) were the result of chromosomal contamination of the plasmid preparation.

In an attempt to locate the *sil* and *pco* loci, whole genome sequencing data was collected from a strain of *E. coli* J53 (pMG101) (obtained from Professor Anne Summers, University of Georgia, Athens, Georgia, USA), that had been stored using cryopreservation – a method enabling long-term storage without bacteria undergoing change in their characteristics.

3.2.1.3 J53 (pMG101) whole genome sequence analysis

The whole genome of a new strain of *E. coli* J53 (pMG101) (received from Professor Anne Summers) was sequenced. The whole genome was extracted using the DNeasy Blood & Tissue Kit (Qiagen,

Manchester, UK) and the sequencing data was generated by Dr. Chrystala Constantinidou (University of Warwick, Coventry).

Assembly, alignment and annotation

The methods used in this section can be found in Table 2.5, Chapter 2. To allow further analysis of the genome, the assembling software SPAdes (Bankevich *et al.*, 2012) was used (with the help of Ms. Sarah Chapman, MRes, University of Nottingham) to assemble the reads into contigs. The size of the whole genome was found to be 4,983,014 bp. Assembly statistics are presented in Table 3.3.

Table 3. 3: Assembly statistics for *E. coli* J53 (pMG101)

Assembly statistics of contigs for sequenced J53 (pMG101), assembled using SPAdes (Bankevich *et al.*, 2012).

	Contigs
Total Number	157
Total Length (bp)	4,983,014
Max. Length (bp)	326,821
Min. Length	128
GC Content (%)	50.50

Aligning the contigs of a 'draft' genome to that of another assembled genome allows for an informed genome structure and more qualitative data analysis. Therefore the assembled sequence was then aligned and reordered accordingly in Mauve (Darling *et al.*, 2004; Rissman *et al.*, 2009) using *E. coli* strains J53 (Clowes and Rowley, 1954; Yi *et al.*,

2012) and MG1655 (Johnson *et al.*, 2007) as reference genomes (Figure 3.4-A). Mauve produces a file in FASTA format, consisting of contigs reordered appropriately, in alignment with the reference genomes.

Annotating genes allows for easy analysis of the genes carried by the genome. So, the Mauve generated FASTA file was then uploaded to the Rapid Annotation using Subsystem Technology (RAST) server (Aziz *et al.*, 2008) for annotation.

Locating the *sil* and *pco* genes

The annotated genome sequence was then analysed for the presence of the *sil* and *pco* genes, using Artemis (Carver *et al.*, 2012) (Table 2.5, Chapter 2). The previously determined nucleotide sequences of the *sil* (GenBank accession number AF067954.1) and *pco* genes (GenBank accession number X83541.1), served as templates for identifying and locating the *sil* and *pco* loci within the whole genome sequence (WGS). Figure 3.4-A displays how the *sil* genes were located in contig 55, within the final ~230 kb of the J53 and MG1655 aligned sequence and the *pco* genes within the last ~415 kb. Neither of the contigs in which the *sil* and *pco* genes were located (55 and 37, respectively) had aligned to the J53 and MG1655 reference genomes (Figures 3.4-A). This was indicative of *sil, pco* and any flanking genes being present in the plasmid. With both sequences present in the last 415 kb of the WGS, the data from this region were evaluated for the presence of the AMR genes found in the original plasmid pMG101 sequence (Table 3.2). Following verification of the presence of the same AMR genes (with resistances to: mercuric chloride, ampicillin,

chloramphenicol, tetracycline, streptomycin, and sulphonamides) this region of the WGS was aligned and reordered (using Mauve (Darling *et al.*, 2004; Rissman *et al.*, 2009)) using the original plasmid pMG101 sequence data as a reference (Chapter 2, Section 2.1.1). The arrangements in Figure 3.4-B show how contigs 55 and 37 (inclusive of the *sil* and *pco* loci) did not align with any of the original plasmid pMG101 sequence.

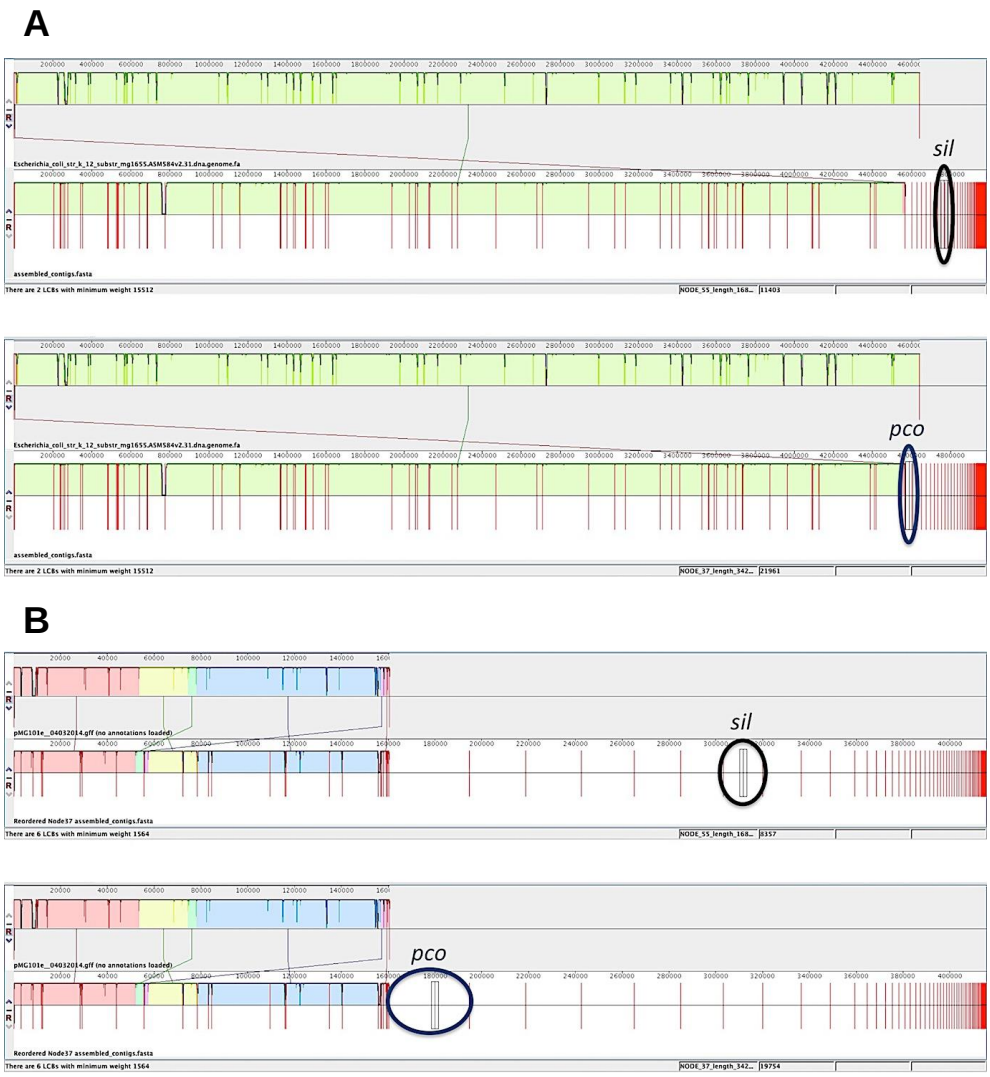


Figure 3. 4 **Mauve genome reorder alignments with J53 and MG1655**

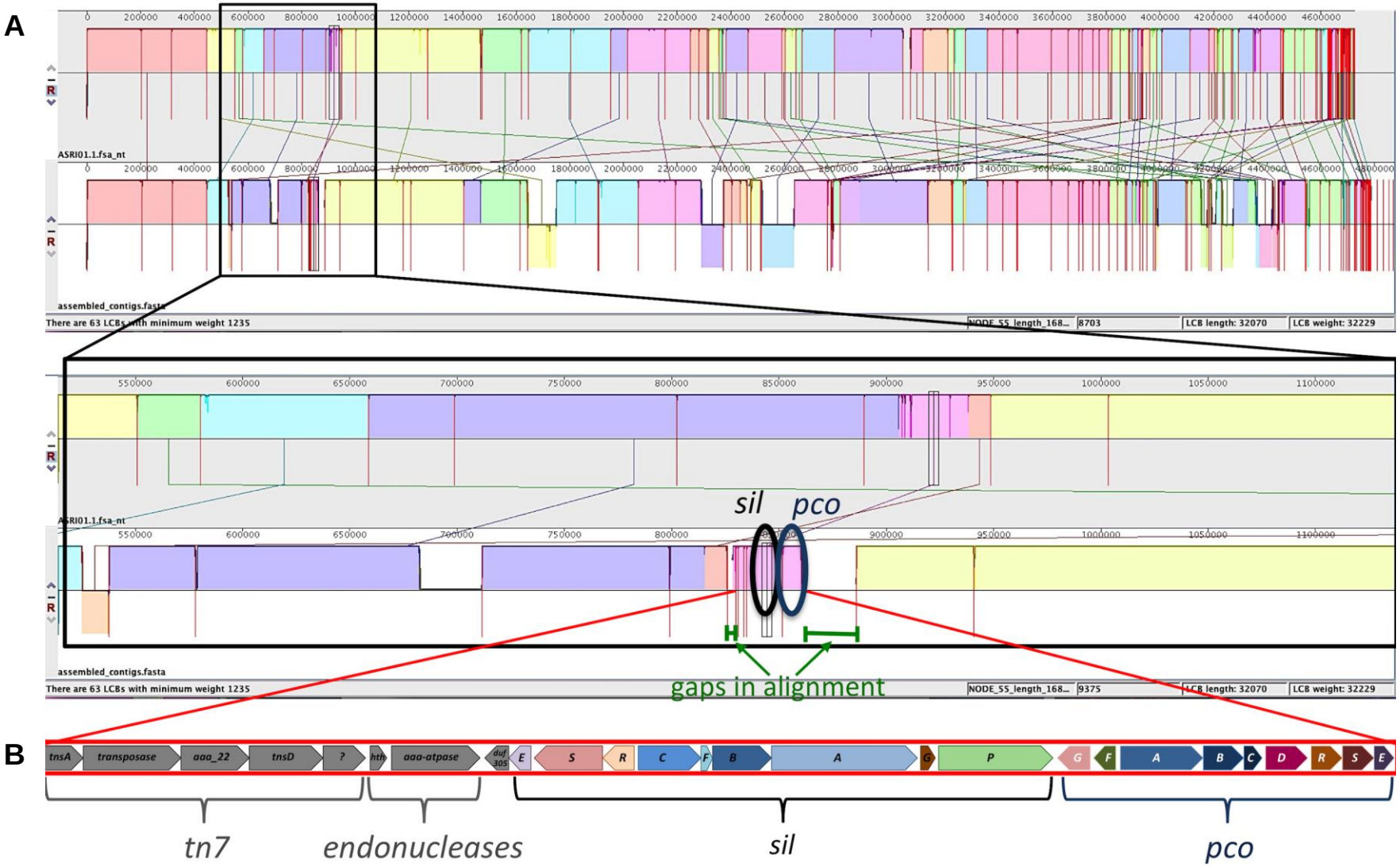
The contigs of the WGS were reordered after alignment with reference genomes **A.** *E. coli* J53 and MG1655 and **B.** originally sequenced plasmid pMG101. Arrangements and alignments were carried out using the 'move contig' tool in Mauve (Darling *et al.*, 2004; Rissman *et al.*, 2009). In the alignments, the reference genome is seen at the top and the rearrangement of our sequence below. A vertical red line separates each contig/node, in which aligned contigs are coloured and the non-aligned regions are colourless. In both cases, the *sil* (contig/node 55, circled black) and *pco* (contig/node 37, circled navy) resided outside of the aligned regions.

However, as described earlier in this chapter, the *sil* and *pco* loci are usually always found neighbouring each other, in the form of one unit (Hobman and Crossman, 2015; Hao *et al.*, 2015). Recently, Randall *et al.*, (2015), deposited the whole genome sequence of *E. coli* strain J53 (pMG101), (GenBank accession number ASRI00000000.1), in which the *sil* and *pco* genes were also found adjacent to one another. In their sequence, the presence of Tn*7*-like elements directly upstream of the *sil* locus suggested that both of the loci had integrated into the chromosome by Tn*7*-based transposition. Analysis of the genome sequence deposited by Randall and colleagues showed that both *sil* and *pco* loci appeared adjacent to each other within the same contig (GenBank accession number ASRI01000010.1); confirming that they were definitely positioned correctly within their sequence. The WGS of J53 (pMG101) was then reordered against the sequence in GenBank accession number ASRI00000000.1. The results of which can be seen in Figure 3.5-A. The newly generated arrangement of the WGS brought contigs 55 and 37, inclusive of the *sil* and *pco* genes, together (Figure 3.5-B). This formed a 21.25 kb *sil-pco* island. Comprehensive sequence analyis of the alignments showed complete matches in the region comprising the *sil* and *pco* genes. The only observed divergence was a deletion of 16 nucleotide bases in the region where contigs 55 and 37 merge. Nevertheless, the ORF present at this site was consistent with the sequence submitted by Randall *et al.*, (2015), still translating to the M23 metallopeptidase protein - PcoG (Figure 3.5B).

Having aligned the sequence of the *sil-pco* gene cluster suitably, the genome upstream and downstream of *sil-pco* was probed to provide further

information on their localisation within the WGS. Figure 3.5-B displays how, similar to Randall *et al.*, (2015), the Tn*7*-transposon-like machinery genes were found present upstream of the *sil* locus. The Tn*7*-like elements found in our sequence included a TnsD-like protein (94 % homologous to *E. coli* TnsD), which is known to be involved in direct insertion of the transposon at the *attTn7* site in the *glmS* gene (For information on Tn*7*-based transposition see Chapter 1, Section 1.4.4). This site is usually found undisturbed, directly upstream of the element upon insertion into the chromosome. Yet, no *attTn7* site was found upstream of the *sil* genes and the preceding Tn*7*-like elements. Additionally, sequences upstream of this machinery and following *pcoE* did not show any consensus with the recently published *E. coli* J53 (pMG101) sequence (Randall *et al.*, 2015) (Figure 3.5-A). Less than a quarter (~8,000 bp) of the large contig 37 (34,220 bp), comprising the final genes of the *pco* gene cluster - *pcoE*, contained sequence similarities to the submitted *E. coli* J53 (pMG101) sequence. Consequently the remaining 75 % genetic content of contig 37 was assumed to be within pMG101 and not present within the J53 genome.

Contig 37 was therefore aligned using both Mauve (Darling *et al.*, 2004) and Clustal Omega (McWilliam *et al.*, 2013) with our originally sequenced plasmid pMG101, to see if the remaining ~26,000 bp sequence showed any similarities to that of pMG101. Figure 3.5-C shows the results of the Mauve alignment wherein Clustal Omega calculated 47 % sequence identity. The similarities were not significant enough to conclude that the sequence following *pcoE* was present in plasmid pMG101.


A
ASRI01.1.fsa_nt
assembled_contigs.fasta
There are 63 LCBs with minimum weight 1235
NODE_55_length_168...
8703
LCB length: 32070
LCB weight: 32229
sil
pco
gaps in alignment
9375
B
tnsA
transposase
aaa_22
tnsD
?
hth
aaa-atpase
duf 305
E
S
R
C
F
B
A
G
P
tn7
endonucleases
sil
pco

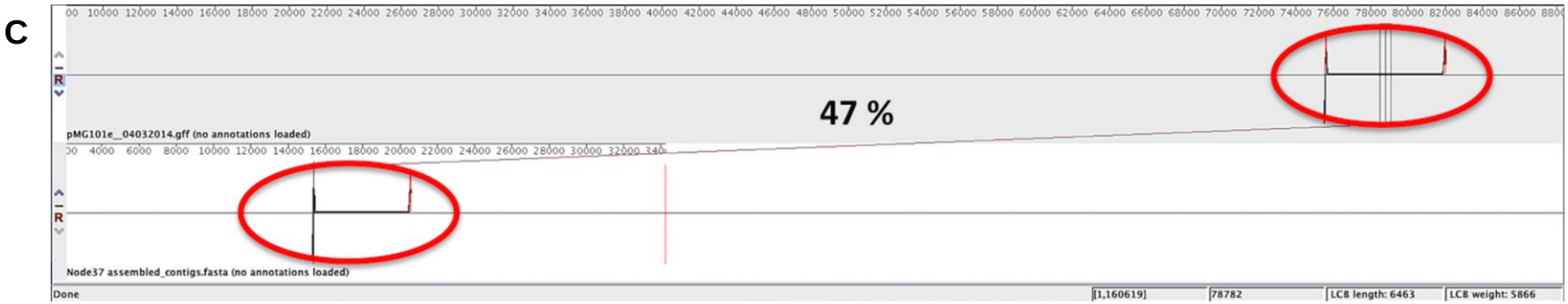


Figure 3. 5 Mauve genome reorder alignments with J53 (pMG101)

A. The contigs of the WGS were reordered after alignment (using Mauve (Darling *et al.*, 2004; Rissman *et al.*, 2009)) with the *E. coli* J53 (pMG101) sequence, GenBank accession number ASRI01000010.1 (Randall *et al.*, 2015). The enlarged view shows how contigs 55 (*sil* – black circle) and 37 (*pco* – navy circle) were reordered to neighbour each other, with *pcoG* being completed where the contigs join. However, gaps can be seen in the alignment prior to the *sil* containing contig 55 and following the *pco* genes in contig 37. These have been highlighted in green. **B.** The new alignment formed the *sil-pco* island, in which genes have been illustrated to scale. Transposon genes were found present upstream of the *sil* gene cluster, implying that the island is part of a transposable element. **C.** Alignment of contig 37 from the WGS with the originally sequenced plasmid pMG101 (using Mauve and Clustal Omega (Darling *et al.*, 2004; McWilliam *et al.*, 2013)) showed 47 % similarities between the two sequences.

Furthermore, analysis of the ORF next to *pcoE*, by using a BLAST (Basic Local Alignment Search Tool) search (Altschul *et al.*, 1990) (see Table 2.5, Chapter 2), showed the translation product had 98 % identity to a plasmid protein – 'Methyltransf_27' found in plasmid APEC O1 (GenBank accession number ABF67809.1). This was suggestive of the *sil-pco* locus being present in plasmid pMG101 of our WGS yet insufficient to conclude as such. Therefore, no conclusions could be made as to whether the *sil-pco* island was located in pMG101 or the chromosomal DNA of our WGS. Unfortunately this means that the surrounding environment of the island could not be characterised to understand more about its acquisition and organisation within the plasmid.

3.2.1.4 The Sil proteins

Discrepancies were previously found between the original nucleotide sequence of the *sil* genes (GenBank accession number AF067954.1, Gupta *et al.*, 1999) and the recently published sequence (GenBank accession number ASRI00000000.1), (Randall *et al.*, 2015). Having identified the genes encoding the ORFs for the Sil proteins, the predicted protein amino acid sequences obtained from our WGS were aligned against those found by Gupta *et al.*, (1999) and Randall *et al.*, (2015). Amino acid seqeunce alignments were carried out using Clustal Omega - the multi-purpose primary sequence alignment program for proteins (McWilliam *et al.*, 2013) (see Table 2.5, Chapter 2). Clustal Omega uses seeded guide trees and HMM (Hidden Markov Model) profile-profile techniques to generate alignments between two or more sequences. The

predicted Sil protein sequences aligned and were 100 % identical to the predicted Sil protein sequences determined by Randall *et al.*, (2015) (GenBank accession number ASRI00000000.1) (Figure 3.6). Hence, the WGS predicted Sil protein amino acid sequences also differed from the original predicted Sil proteins (GenBank accession number AF067954.1), (Gupta *et al.*, 1999). Proteins SilP, SilR and SilS had the most significant change in predicted amino acid sequence compared to the original predictions, whilst SilF had one change in one residue. The key differences have been highlighted in Figure 3.7 and Table 3.4.

To allow for a better understanding of the mechanism of silver resistance, overexpressing the Sil proteins was necessary. Whilst awaiting the results of our sequencing data, the sequences from GenBank accession number AF067954.1 (Gupta *et al.*, 1999) were used by Dr. Lousie Bird (Oxfor Protein Production Facility, Research Complex at Harwell, Oxfordshire, UK) to design primers (Table 2.3, Chapter 2) for the cloning of the *sil* genes into overepxression constructs. The sequences of the designed primers fortunately matched the nucleotide sequences of our *sil* gene sequences. The cloning of the *sil* genes and overexpression of the Sil proteins are described in the following section.

A

```
# Length: 823
# Identity:     823/823 (100.0%)
# Similarity:   823/823 (100.0%)
# Gaps:           0/823 ( 0.0%)
# Score: 4181.0
#
#
#=======================================

SilP(ASRI0000      1 MLQICIRRVTVKNDNAVEHNNQTASEQTSSPDESHALHKVRDPVCGMVIL     50
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
SilP(KA)           1 MLQICIRRVTVKNDNAVEHNNQTASEQTSSPDESHALHKVRDPVCGMVIL     50

SilP(ASRI0000     51 PDKAHSSIRYQDHQLYFCSASCESKFKAHPDHYFTEDASEHHHHHDHHEV    100
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
SilP(KA)          51 PDKAHSSIRYQDHQLYFCSASCESKFKAHPDHYFTEDASEHHHHHDHHEV    100

SilP(ASRI0000    101 SPDKIKQSHRQAEKEISEGVWTCPMHPEIRRSGPGSCPVCGMALEPLVAT    150
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
SilP(KA)         101 SPDKIKQSHRQAEKEISEGVWTCPMHPEIRRSGPGSCPVCGMALEPLVAT    150

SilP(ASRI0000    151 ASTGTSDELRDMTRRFWLGLLLAFPVLILEMGSHLFPALRNTVPPQYNTW    200
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
SilP(KA)         151 ASTGTSDELRDMTRRFWLGLLLAFPVLILEMGSHLFPALRNTVPPQYNTW    200

SilP(ASRI0000    201 LQLLLASPVVLWCGWPFFARAGMSLRNRSLNMFTLVAMGTGVAWVYSVIA    250
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
SilP(KA)         201 LQLLLASPVVLWCGWPFFARAGMSLRNRSLNMFTLVAMGTGVAWVYSVIA    250

SilP(ASRI0000    251 TVFPSWFPASFRNMDGLVAIYFEAAAVITVLVLLGQVLELRAREQTSGAI    300
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
SilP(KA)         251 TVFPSWFPASFRNMDGLVAIYFEAAAVITVLVLLGQVLELRAREQTSGAI    300

SilP(ASRI0000    301 TALLNLAPKTARRLDQDGHETDINAEDVLPGDKLRIRPGESIPVDGIVVE    350
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
SilP(KA)         301 TALLNLAPKTARRLDQDGHETDINAEDVLPGDKLRIRPGESIPVDGIVVE    350

SilP(ASRI0000    351 GKTTVDESMVTGESMPVTKTEGEPVIGGTINQTGSLIIRAEKVGDETMLS    400
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
SilP(KA)         351 GKTTVDESMVTGESMPVTKTEGEPVIGGTINQTGSLIIRAEKVGDETMLS    400

SilP(ASRI0000    401 RIVQMVADAQRSRAPIQRMADSVSGWFVPLVILIAVVAFMIWSVWGPEPR    450
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
SilP(KA)         401 RIVQMVADAQRSRAPIQRMADSVSGWFVPLVILIAVVAFMIWSVWGPEPR    450

SilP(ASRI0000    451 MAHGLIAAVSVLIIACPCALGLATPMSIMVGVGKGAQAGVLIKNAEALER    500
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
SilP(KA)         451 MAHGLIAAVSVLIIACPCALGLATPMSIMVGVGKGAQAGVLIKNAEALER    500

SilP(ASRI0000    501 LEKVDTLVVDKTGTLTEGSPTVTGIISLNPGGETSLLRVTAAVEKGSQHP    550
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
SilP(KA)         501 LEKVDTLVVDKTGTLTEGSPTVTGIISLNPGGETSLLRVTAAVEKGSQHP    550
SilP(ASRI0000    551 LGMAVVKAAQEKGIAIPAVTHFNAPSGKGVSGDVEGQRVVIGNELAMQEN    600
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
SilP(KA)         551 LGMAVVKAAQEKGIAIPAVTHFNAPSGKGVSGDVEGQRVVIGNELAMQEN    600

SilP(ASRI0000    601 SIVIDNQKAVADTLRMEGTTVIYVATDGHLAGLIAISDPVKATTPDALKA    650
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
SilP(KA)         601 SIVIDNQKAVADTLRMEGTTVIYVATDGHLAGLIAISDPVKATTPDALKA    650

SilP(ASRI0000    651 LRQAGIRIVMLTGDNQLTAEAVARKLGIDEVEAGILPDGKKAVITRLKAS    700
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
SilP(KA)         651 LRQAGIRIVMLTGDNQLTAEAVARKLGIDEVEAGILPDGKKAVITRLKAS    700

SilP(ASRI0000    701 GHVVAMAGDGVNDAPALAAADVGIAMGTGTDVAIESAGVTLLKGDLMILN    750
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
SilP(KA)         701 GHVVAMAGDGVNDAPALAAADVGIAMGTGTDVAIESAGVTLLKGDLMILN    750

SilP(ASRI0000    751 RARHLSEITMKNIRQNLFFAFIYNALGVPVAAGLLYPVYGILLSPVIAAA    800
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
SilP(KA)         751 RARHLSEITMKNIRQNLFFAFIYNALGVPVAAGLLYPVYGILLSPVIAAA    800

SilP(ASRI0000    801 AMALSSVSVIVNALRLKSVRLGK    823
                     |||||||||||||||||||||||
SilP(KA)         801 AMALSSVSVIVNALRLKSVRLGK    823
```

B

```
# Length: 146
# Identity:     146/146 (100.0%)
# Similarity:   146/146 (100.0%)
# Gaps:           0/146 ( 0.0%)
# Score: 754.0
#
#
#=======================================

SilG(ASRI0000      1 MKKVVLMALALGLSLPAMASEKVIDMYKSENCGCCSLWGKAMEKDGFEVR     50
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
SilG(KA)           1 MKKVVLMALALGLSLPAMASEKVIDMYKSENCGCCSLWGKAMEKDGFEVR     50

SilG(ASRI0000     51 THVMNDQALSALKEKHAVPAGLRSCHTAVVGNLIIEGHVPAATIHKAMQS    100
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
SilG(KA)          51 THVMNDQALSALKEKHAVPAGLRSCHTAVVGNLIIEGHVPAATIHKAMQS    100

SilG(ASRI0000    101 GSGIYGLATPGMPAGSPGMEMGARKEAYDVIAFSPEGSKKVFQRIE    146
                     ||||||||||||||||||||||||||||||||||||||||||||||||
SilG(KA)         101 GSGIYGLATPGMPAGSPGMEMGARKEAYDVIAFSPEGSKKVFQRIE    146
```

C

```
# Length: 1048
# Identity:    1048/1048 (100.0%)
# Similarity:  1048/1048 (100.0%)
# Gaps:           0/1048 ( 0.0%)
# Score: 5295.0
#
#
#=====================================

SilA(ASRI0000      1 MIEWIIRRSVANRFLVMMGALFLSIWGTWTIINTPVDALPDLSDVQVIIK     50
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
SilA(KA)           1 MIEWIIRRSVANRFLVMMGALFLSIWGTWTIINTPVDALPDLSDVQVIIK     50

SilA(ASRI0000     51 TSYPGQAPQIVENQVTYPLTTTMLSVPGAKTVRGFSQFGDSYVYVIFEDG    100
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
SilA(KA)          51 TSYPGQAPQIVENQVTYPLTTTMLSVPGAKTVRGFSQFGDSYVYVIFEDG    100

SilA(ASRI0000    101 TDLYWARSRVLEYLNQVQGKLPAGVSSEIGPDATGVGWIFEYALVDRSGK    150
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
SilA(KA)         101 TDLYWARSRVLEYLNQVQGKLPAGVSSEIGPDATGVGWIFEYALVDRSGK    150

SilA(ASRI0000    151 HDLSELRSLQDWFLKFELKTIPNVAEVASVGGVVKQYQIQVNPVKLSQYG    200
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
SilA(KA)         151 HDLSELRSLQDWFLKFELKTIPNVAEVASVGGVVKQYQIQVNPVKLSQYG    200

SilA(ASRI0000    201 ISLPEVKQALESSNQEAGGSSVEIAEAEYMVRASGYLQSIDDFNNIVLKT    250
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
SilA(KA)         201 ISLPEVKQALESSNQEAGGSSVEIAEAEYMVRASGYLQSIDDFNNIVLKT    250

SilA(ASRI0000    251 GENGVPVYLRDVARVQTGPEMRRGIAELNGQGEVAGGVVILRSGKNARDV    300
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
SilA(KA)         251 GENGVPVYLRDVARVQTGPEMRRGIAELNGQGEVAGGVVILRSGKNARDV    300

SilA(ASRI0000    301 ITAVRDKLETLKASLPEGVEIVTTYDRSQLIDRAIDNLSSKLLEEFFVVA    350
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
SilA(KA)         301 ITAVRDKLETLKASLPEGVEIVTTYDRSQLIDRAIDNLSSKLLEEFFVVA    350

SilA(ASRI0000    351 IVCALFLWHVRSALVAIISLPLGLCIAFIVMHFQGLNANIMSLGGIAIAV    400
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
SilA(KA)         351 IVCALFLWHVRSALVAIISLPLGLCIAFIVMHFQGLNANIMSLGGIAIAV    400

SilA(ASRI0000    401 GAMVDAAIVMIENAHKRLEEWDHRHPGEQIDNVTRWKVITDASVEVGPAL    450
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
SilA(KA)         401 GAMVDAAIVMIENAHKRLEEWDHRHPGEQIDNVTRWKVITDASVEVGPAL    450

SilA(ASRI0000    451 FISLLIITLSFIPIFTLEGQEGRLFGPLAFTKTYSMAGAAALAIIVIPIL    500
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
SilA(KA)         451 FISLLIITLSFIPIFTLEGQEGRLFGPLAFTKTYSMAGAAALAIIVIPIL    500

SilA(ASRI0000    501 MGFWIRGKIPAETSNPLNRVLIKAYHPLLLRVLHWPKTTLLVAALSIFTV    550
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
SilA(KA)         501 MGFWIRGKIPAETSNPLNRVLIKAYHPLLLRVLHWPKTTLLVAALSIFTV    550

SilA(ASRI0000    551 VWPLSQVGGEFLPKINEGDLLYMPSTLPGVSPAEAAALLQTTDKLIKSVP    600
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
SilA(KA)         551 VWPLSQVGGEFLPKINEGDLLYMPSTLPGVSPAEAAALLQTTDKLIKSVP    600
```

```
SilA(ASRI0000    601 EVASVFGKTGKAETATDSAPLEMVETTIQLKPEDQWRPGMTIDKIIDELD    650
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
SilA(KA)         601 EVASVFGKTGKAETATDSAPLEMVETTIQLKPEDQWRPGMTIDKIIDELD    650

SilA(ASRI0000    651 RTVRLPGLANLWVPPIRNRIDMLSTGIKSPIGIKVSGTVLSDIDATAQSI    700
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
SilA(KA)         651 RTVRLPGLANLWVPPIRNRIDMLSTGIKSPIGIKVSGTVLSDIDATAQSI    700

SilA(ASRI0000    701 EAVAKTVPGVVSVLAERLEGGRYIDIDINREKASRYGMTVGDVQLFVSSA    750
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
SilA(KA)         701 EAVAKTVPGVVSVLAERLEGGRYIDIDINREKASRYGMTVGDVQLFVSSA    750

SilA(ASRI0000    751 IGGAMVGETVEGVARYPINIRYPQDYRNSPQALKQMPILTPMKQQITLGD    800
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
SilA(KA)         751 IGGAMVGETVEGVARYPINIRYPQDYRNSPQALKQMPILTPMKQQITLGD    800

SilA(ASRI0000    801 VADINVVSGPTMLKTENARPASWIYVDARGRDMVSVVNDIKTAISEKVKL    850
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
SilA(KA)         801 VADINVVSGPTMLKTENARPASWIYVDARGRDMVSVVNDIKTAISEKVKL    850

SilA(ASRI0000    851 RPGTSVAFSGQFELLEHANKKLKLMVPMTVMIIFILLYLAFRRVDEALLI    900
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
SilA(KA)         851 RPGTSVAFSGQFELLEHANKKLKLMVPMTVMIIFILLYLAFRRVDEALLI    900

SilA(ASRI0000    901 LMSLPFALVGGIWFLYWQGFHMSVATGTGFIALAGVAAEFGVVMLMYLRH    950
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
SilA(KA)         901 LMSLPFALVGGIWFLYWQGFHMSVATGTGFIALAGVAAEFGVVMLMYLRH    950

SilA(ASRI0000    951 AIEAHPELSRKETFTPEGLDEALYHGAVLRVRPKAMTVAVIIAGLLPILW   1000
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
SilA(KA)         951 AIEAHPELSRKETFTPEGLDEALYHGAVLRVRPKAMTVAVIIAGLLPILW   1000

SilA(ASRI0000   1001 GTGAGSEVMSRIAAPMIGGMITAPLLSLFIIPAAYKLIWLRRHKKSVS   1048
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
SilA(KA)        1001 GTGAGSEVMSRIAAPMIGGMITAPLLSLFIIPAAYKLIWLRRHKKSVS   1048
```

D

```
# Length: 430
# Identity:     430/430 (100.0%)
# Similarity:   430/430 (100.0%)
# Gaps:           0/430 ( 0.0%)
# Score: 2200.0
#
#
#=======================================

SilBASRI00000      1 MASLKIKYAAIIISSLIAGGLISVTAWQYLNSSQKTVPAEQKAPEKKVLF     50
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
SilB(KA)           1 MASLKIKYAAIIISSLIAGGLISVTAWQYLNSSQKTVPAEQKAPEKKVLF     50

SilBASRI00000     51 WYDPMKPDTKFDKPGKSPFMDMDLVPKYADESGDKSSGGIRIDPTQVQNL    100
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
SilB(KA)          51 WYDPMKPDTKFDKPGKSPFMDMDLVPKYADESGDKSSGGIRIDPTQVQNL    100

SilBASRI00000    101 GLKTQKVTRGMLNYSQTIPANVSYNEYQFVIVQARSDGFVEKVYPLTIGD    150
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
SilB(KA)         101 GLKTQKVTRGMLNYSQTIPANVSYNEYQFVIVQARSDGFVEKVYPLTIGD    150

SilBASRI00000    151 HVKKGTPLIDITIPEWVEAQSEFLLLSGTGGTPTQIKGVLERLRLAGMPE    200
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
SilB(KA)         151 HVKKGTPLIDITIPEWVEAQSEFLLLSGTGGTPTQIKGVLERLRLAGMPE    200

SilBASRI00000    201 EDIQRLRSTRTIQTRFTIKAPIDGVITAFDLRTGMNISKDKVVAQIQGMD    250
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
SilB(KA)         201 EDIQRLRSTRTIQTRFTIKAPIDGVITAFDLRTGMNISKDKVVAQIQGMD    250

SilBASRI00000    251 PVWISAAVPESIAYLLKDTSQFEISVPAYPDKTFHVEKWNILPSVDQTTR    300
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
SilB(KA)         251 PVWISAAVPESIAYLLKDTSQFEISVPAYPDKTFHVEKWNILPSVDQTTR    300

SilBASRI00000    301 TLQVRLQVTNKDEFLKPGMNAYLKLNTQSQEMLLIPSQAVIDTGKEQRVI    350
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
SilB(KA)         301 TLQVRLQVTNKDEFLKPGMNAYLKLNTQSQEMLLIPSQAVIDTGKEQRVI    350

SilBASRI00000    351 TVDDEGKFVPKQIHVLHESQQQSGIGSGLNEGDTVVVSGLFLIDSEANIT    400
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
SilB(KA)         351 TVDDEGKFVPKQIHVLHESQQQSGIGSGLNEGDTVVVSGLFLIDSEANIT    400

SilBASRI00000    401 GALERMRHPEKTENSMPAMSEQPVNMHSGH    430
                     ||||||||||||||||||||||||||||||
SilB(KA)         401 GALERMRHPEKTENSMPAMSEQPVNMHSGH    430
```

E

```
# Length: 117
# Identity:     117/117 (100.0%)
# Similarity:   117/117 (100.0%)
# Gaps:           0/117 ( 0.0%)
# Score: 591.0
#
#
#=======================================

SilF(ASRI0000      1 MRNSLKAVLFGAFSVMFSAGLHAETHQHGDMNAASDASVQQVIKGTGIVK     50
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
SilF(KA)           1 MRNSLKAVLFGAFSVMFSAGLHAETHQHGDMNAASDASVQQVIKGTGIVK     50

SilF(ASRI0000     51 DIDMNSKKITISHEAIPAVGWPAMTMRFTFVNADDAINALKTGNHVDFSF    100
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
SilF(KA)          51 DIDMNSKKITISHEAIPAVGWPAMTMRFTFVNADDAINALKTGNHVDFSF    100

SilF(ASRI0000    101 IQQGNISLLKSINVTQS    117
                     |||||||||||||||||
SilF(KA)         101 IQQGNISLLKSINVTQS    117
```

F

```
# Length: 461
# Identity:     461/461 (100.0%)
# Similarity:   461/461 (100.0%)
# Gaps:           0/461 ( 0.0%)
# Score: 2295.0
#
#
#=======================================

SilC(ASRI0000      1 MFKLKLLSISTIFILAGCVSLAPEYQRPPAPVPQQFSLSKNSLTPAVNSY     50
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
SilC(KA)           1 MFKLKLLSISTIFILAGCVSLAPEYQRPPAPVPQQFSLSKNSLTPAVNSY     50

SilC(ASRI0000     51 QDTGWRNFFVDPQVSRLIGEALNNNRDLRMAALKVEEARAQFNVTDADRY    100
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
SilC(KA)          51 QDTGWRNFFVDPQVSRLIGEALNNNRDLRMAALKVEEARAQFNVTDADRY    100

SilC(ASRI0000    101 PQLNASSGITYNGGLKGDKPTTQEYDAGLELSYELDFFGKLKNMSEADRQ    150
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
SilC(KA)         101 PQLNASSGITYNGGLKGDKPTTQEYDAGLELSYELDFFGKLKNMSEADRQ    150

SilC(ASRI0000    151 NYFASEEARRAVHILLVSNVSQSYFSQQLAYEQLRIARETLKNYEQSYAF    200
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
SilC(KA)         151 NYFASEEARRAVHILLVSNVSQSYFSQQLAYEQLRIARETLKNYEQSYAF    200

SilC(ASRI0000    201 VEQQLVTGSTNVLALEQARGQIESTRAEIAKREGDLAQANNALQLVLGTY    250
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
SilC(KA)         201 VEQQLVTGSTNVLALEQARGQIESTRAEIAKREGDLAQANNALQLVLGTY    250

SilC(ASRI0000    251 RAVPSEKGIKGGEIAPVKLPPNLSSQILLQRPDIMEAEYQLKAADANIGA    300
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
SilC(KA)         251 RAVPSEKGIKGGEIAPVKLPPNLSSQILLQRPDIMEAEYQLKAADANIGA    300

SilC(ASRI0000    301 ARAAFFPSITLTSGLSSSSTELSSLFTSGSGMWNFIPKIEIPIFNAGRNK    350
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
SilC(KA)         301 ARAAFFPSITLTSGLSSSSTELSSLFTSGSGMWNFIPKIEIPIFNAGRNK    350

SilC(ASRI0000    351 ANLKLAEIRQQQSVVNYEQKIQSAFKDVSDTLALRDSLSQQLESQQRYLD    400
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
SilC(KA)         351 ANLKLAEIRQQQSVVNYEQKIQSAFKDVSDTLALRDSLSQQLESQQRYLD    400

SilC(ASRI0000    401 SLQITLQRARGLYASGAVSYIEVLDAERSLFATQQTILDLTYSRQVNEIN    450
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
SilC(KA)         401 SLQITLQRARGLYASGAVSYIEVLDAERSLFATQQTILDLTYSRQVNEIN    450

SilC(ASRI0000    451 LFTALGGGWVE    461
                     ||||||||||||
SilC(KA)         451 LFTALGGGWVE    461
```

G

```
# Length: 226
# Identity:     226/226 (100.0%)
# Similarity:   226/226 (100.0%)
# Gaps:           0/226 ( 0.0%)
# Score: 1131.0
#
#
#=======================================

SilR(ASRI0000      1 MKILIVEDDIKTGEYLSKGLTEAGFVVDHADNGLTGYHLAMTAEYDLVIL     50
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
SilR(KA)           1 MKILIVEDDIKTGEYLSKGLTEAGFVVDHADNGLTGYHLAMTAEYDLVIL     50

SilR(ASRI0000     51 DIMLPDVNGWDIIRMLRSAGKGMPVLLLTALGTIEHRVKGLELGADDYLV    100
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
SilR(KA)          51 DIMLPDVNGWDIIRMLRSAGKGMPVLLLTALGTIEHRVKGLELGADDYLV    100

SilR(ASRI0000    101 KPFAFAELLARVRTLLRRGNTMITESQLKVADLSVDLVSRKVSRAGNRIV    150
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
SilR(KA)         101 KPFAFAELLARVRTLLRRGNTMITESQLKVADLSVDLVSRKVSRAGNRIV    150

SilR(ASRI0000    151 LTSKEFSLLEFFIRHQGEVLPRSLIASQVWDMNFDSDTNAIDVAVKRLRA    200
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
SilR(KA)         151 LTSKEFSLLEFFIRHQGEVLPRSLIASQVWDMNFDSDTNAIDVAVKRLRA    200

SilR(ASRI0000    201 KIDNDYGTKLIQTVRGVGYMLEIPDA    226
                     ||||||||||||||||||||||||||
SilR(KA)         201 KIDNDYGTKLIQTVRGVGYMLEIPDA    226
```

H

```
# Length: 493
# Identity:     493/493 (100.0%)
# Similarity:   493/493 (100.0%)
# Gaps:           0/493 ( 0.0%)
# Score: 2483.0
#
#
#=======================================

SilS(ASRI0000      1 MHSKPSRRPFSLALRLTFFISLSTILAFIAFTWFMLHSVEKHFAEQDVSD     50
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
SilS(KA)           1 MHSKPSRRPFSLALRLTFFISLSTILAFIAFTWFMLHSVEKHFAEQDVSD     50

SilS(ASRI0000     51 LQQISTTLSRILQSPADPDEKKVSKIKESIASYRNVALLLLNPRGEVLYS    100
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
SilS(KA)          51 LQQISTTLSRILQSPADPDEKKVSKIKESIASYRNVALLLLNPRGEVLYS    100

SilS(ASRI0000    101 SAQGAALRPAVNSADFSEHSRARDVFLWTVEDTARAMDTGSGMKMETYRI    150
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
SilS(KA)         101 SAQGAALRPAVNSADFSEHSRARDVFLWTVEDTARAMDTGSGMKMETYRI    150

SilS(ASRI0000    151 IASSGQATFQGKQQNYVMLTGLSINFHLHYLDALKKNLIAIAVVISLLIV    200
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
SilS(KA)         151 IASSGQATFQGKQQNYVMLTGLSINFHLHYLDALKKNLIAIAVVISLLIV    200

SilS(ASRI0000    201 LIIRIAVRQGHLPLRNVSNAIKNITSENLDARLEPTRVPIELEQLVISFN    250
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
SilS(KA)         201 LIIRIAVRQGHLPLRNVSNAIKNITSENLDARLEPTRVPIELEQLVISFN    250

SilS(ASRI0000    251 HMIGKIEDVFTRQANFSADIAHEIRTPITNLVTQTEIALSQDRTQKELED    300
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
SilS(KA)         251 HMIGKIEDVFTRQANFSADIAHEIRTPITNLVTQTEIALSQDRTQKELED    300

SilS(ASRI0000    301 VLYSSLEEYNRMTKMVSDMLFLAQADNNQLIPDRVRFDLRAEVMKVFEFF    350
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
SilS(KA)         301 VLYSSLEEYNRMTKMVSDMLFLAQADNNQLIPDRVRFDLRAEVMKVFEFF    350

SilS(ASRI0000    351 EAWAEERNIMLKFNGMPCLVEGDPQMFRRAINNLLSNALRYTPEGQAITV    400
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
SilS(KA)         351 EAWAEERNIMLKFNGMPCLVEGDPQMFRRAINNLLSNALRYTPEGQAITV    400

SilS(ASRI0000    401 SIREQESFFDLVIENPGKPIPEEHLSRLFDRFYRVDPSRQRKGEGSGIGL    450
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
SilS(KA)         401 SIREQESFFDLVIENPGKPIPEEHLSRLFDRFYRVDPSRQRKGEGSGIGL    450

SilS(ASRI0000    451 AIVKSIVEAHHGRVQVESDVHSTRFILSVPRLEKMIPDTQCWE    493
                     |||||||||||||||||||||||||||||||||||||||||||
SilS(KA)         451 AIVKSIVEAHHGRVQVESDVHSTRFILSVPRLEKMIPDTQCWE    493
```

I

```
# Length: 143
# Identity:     143/143 (100.0%)
# Similarity:   143/143 (100.0%)
# Gaps:           0/143 ( 0.0%)
# Score: 731.0
#
#
#=======================================

SilE(ASRI0000      1 MKNIVLASLLGFGLISSAWATETVNIHERVNNAQAPAHQMQSAAAPVGIQ     50
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
SilE(KA)           1 MKNIVLASLLGFGLISSAWATETVNIHERVNNAQAPAHQMQSAAAPVGIQ     50

SilE(ASRI0000     51 GTAPRMAGMDQHEQAIIAHETMTNGSADAHQKMVESHQRMMGSQTVSPTG    100
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
SilE(KA)          51 GTAPRMAGMDQHEQAIIAHETMTNGSADAHQKMVESHQRMMGSQTVSPTG    100

SilE(ASRI0000    101 PSKSLAAMNEHERAAVAHEFMNNGQSGPHQAMAEAHRRMLSAG    143
                     |||||||||||||||||||||||||||||||||||||||||||
SilE(KA)         101 PSKSLAAMNEHERAAVAHEFMNNGQSGPHQAMAEAHRRMLSAG    143
```

Figure 3. 6 Sil protein sequence alignments

Sequence alignment of predicted Sil proteins from GenBank accession number (GenBank accession number ASRI00000000.1 - top (Randall *et al.*, 2015)) and our WGS (KA - bottom). The sequences were aligned using Clustal Omega (McWilliam *et al.*, 2013). All of the Sil proteins **A.** SilP **B.** SilG **C.** SilA **D.** SilB **E.** SilF **F.** SilC **G.** SilR **H.** SilS and **I.** SilE show 100 % identity (highlighted by red circle).

Table 3. 4: Differences in the amino acids of predicted Sil proteins from sequencing of J53 (pMG101) compared to the original published sequence

A list of substitution or deletion (Del) mutations found in the Sil proteins SilP, SilF, SilR and SilS; alongside any characteristic changes to the amino acid at neutral pH, in comparison to sequence data submitted by Gupta *et al.*, (1999), (GenBank accession number AF067954.1).

Protein	Mutation	Characteristic Change at Neutral pH
SilP	D23T	Loss of negative charge
	C24A	Loss of thiol side chain
	F25Del	Loss of benzyl group
	L26S	Hydroxyl, polar to nonpolar
	S27E	Hydroxyl to acidic amide, negative
	R26Q	Basic, positive to acidic amide, negative
	R32P	Loss of positive charge, basic to cyclic
	E43D	None
	S44P	Hydroxyl to cyclic
	D545E	None
SilF	S6Q	Hydroxyl to acidic
SilR	F177S	Aromatic to hydroxyl
	F178Del	Loss of aromatic
	M179Q	Loss of sulphur to acidic amide
	V182D	Gain in acidic amide, negative
	H183M	Addition of sulphur
	E184N	Loss of negative charge
	L186Del	Loss of aliphatic
	T187D	Hydroxyl to negative, acidic amide
	A188S	Gain of sulphur
	N213I	Loss of acidic amide
SilS	L8R	Gain of basic, positive charge
	Q340R	Acidic, negative to basic, positive

Protein	Mutation	Characteristic Effect at Neutral pH
SilS	S341A	Hydroxyl to aliphatic
	Q342E	Gain of negative charge None
	N343Del	None
	S344V	Loss of hydroxyl
	L345M	Gain of sulphur
	R349E	Basic, positive to acidic, negative
	V350F	Gain of aromatic
	S352Del	Loss of hydroxyl
	L355W	Gain of aromatic
	G356A	None
	P357Del	None
	K358A	Loss of basicity and positive charge
	T360E	Hydroxyl to acidic, negative charge
	P361R	Gain of basicity and positive charge
	I362N	Gain of acidity
	L363I	None
	L346M	Gain of sulphur

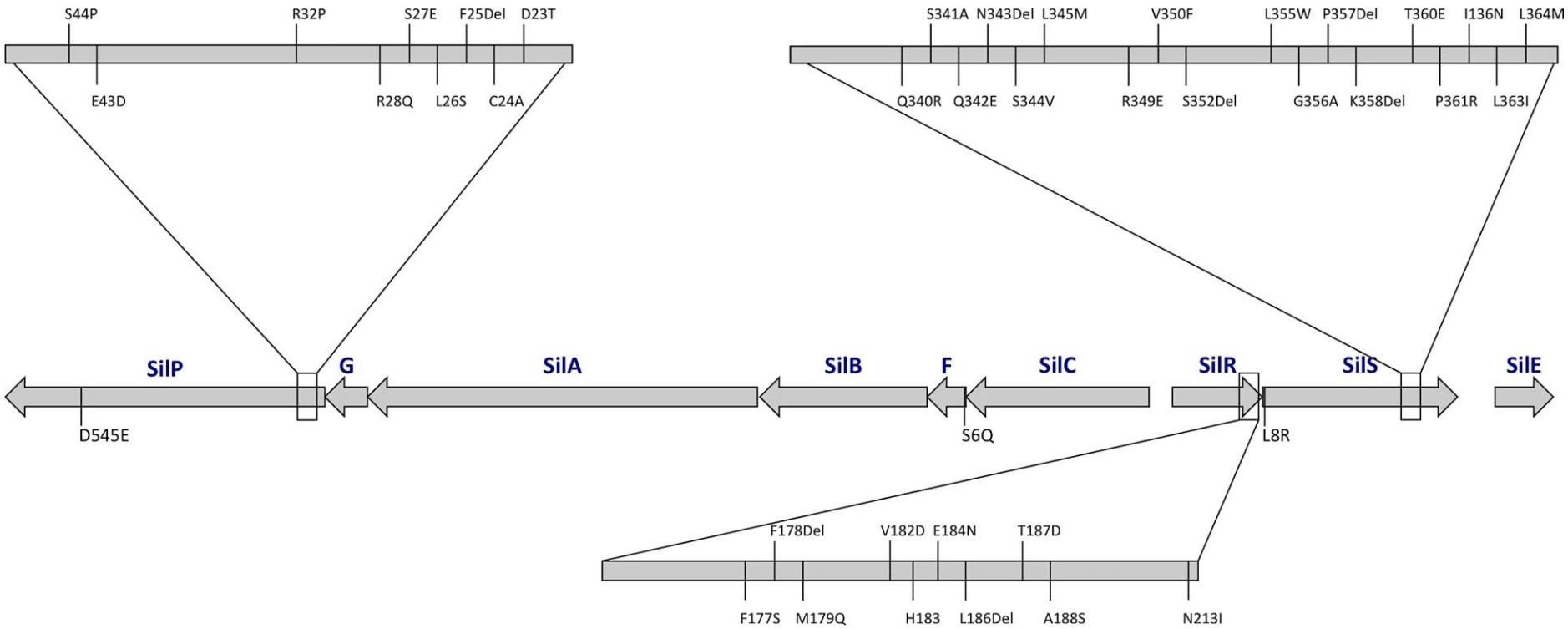


Figure 3. 7 Variations in the Sil proteins

The DNA sequencing of the *sil* genes within the WGS of *E.coli* J53 (pMG1010) showed a number of differences when compared to the previously published sequence (GenBank accession number AF067954.1, Gupta *et al.*, 1999). The grey arrows represent the predicted translation products of the *sil* genes from the WGS, to scale. Discrepancies in the predicted primary protein sequence have been noted as substitution or deletion (Del) mutations. These have been listed in Table 3.4.

3.2.2 Production of the Sil proteins

To facilitate the structural and biophysical analysis and characterisation of any of the Sil proteins, it was essential to produce substantial quantities of soluble protein, of the proteins in question. Cloning of PCR amplicons of *silP*, *A*, *B*, *C*, *E*, *R* and *S* into overexpression vectors was carried out, and the overexpression levels of the different constructs were tested and analysed (Chapter 2, Sections 2.2.4.4 and 2.2.5.1).

3.2.2.1 Construction of the plasmid expression vectors

In order to overexpress the Sil Proteins, the *sil* genes were first amplified using primers in Table 2.3, Chapter 2 (Figure 3.8). Amplicons were cloned into suitable fusion plasmid expression vectors - pOPINE, pOPINF, pOPINM, pOPINI, pOPINJ, pOPINS3C and pOPINE-3C-eGFP, (Tables 2.2, 2.7, and Figure 2.1, Chapter 2) - following a series of experiments conducted in the Oxford Protein Production Facility, described in Chapter 2, Section 2.2.4.4. Figure 3.9 shows agarose gel electrophoresis of PCR products that verified the clones, by confirming the presence of the correct *sil* genes in at least one of the screened clones. For constructs $P_{2\text{-}824E}$, $S_{2\text{-}497G}$, $A_{2\text{-}1048I}$ and $A_{2\text{-}1084G}$ second and third attempts of cloning were required (Figures 3.9-B and C). After PCR verification, the inserts were sequenced to confirm that the DNA sequences of the *sil* genes were correct.

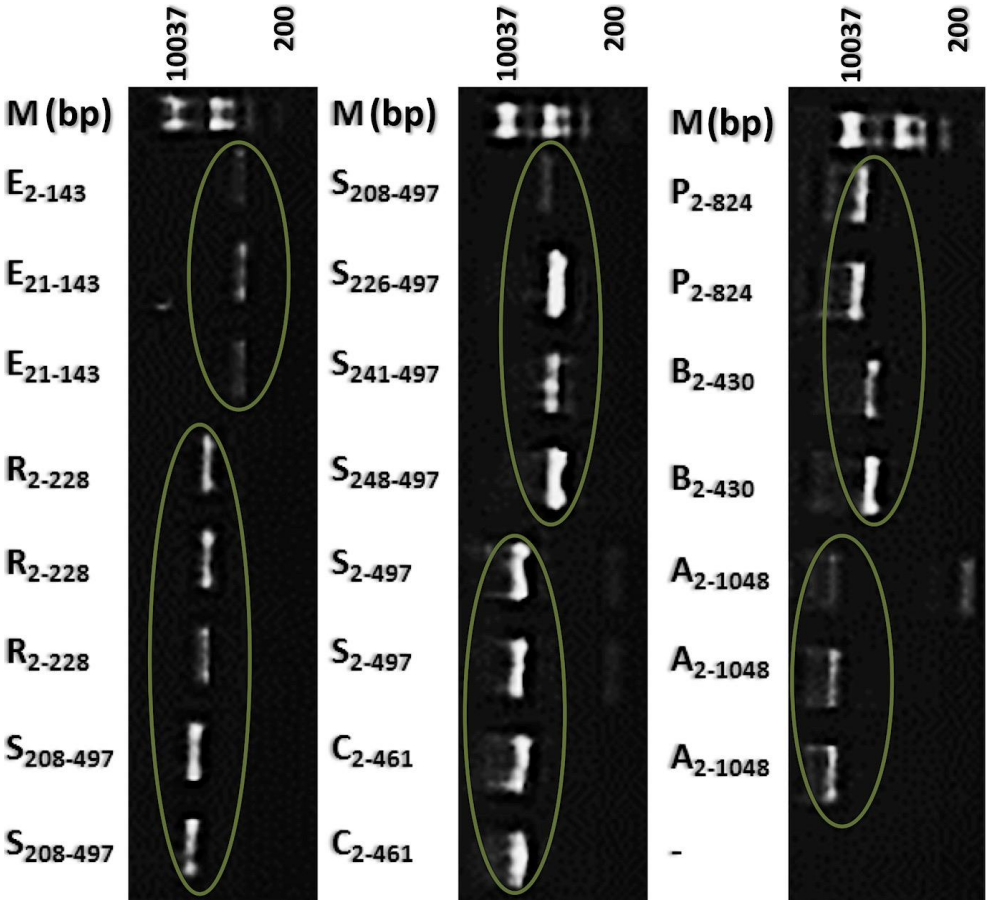


Figure 3. 8 Agarose gel of *sil* gene PCR products

A 1% TAE Agarose gel (Chapter 2, Section 2.2.2.1) illustrating how all 23 of the gene targets, for the seven Sil proteins, generated amplified products at the correct size (circled green, product lengths in Table 2.7, Chapter 2). The PCRs were carried out using the primers shown in Table 2.3, Chapter 2. The DNA molecular weight marker used was HyperLadder™ 1kb (Bioline Reagents Ltd., London, UK), ranging from bp sizes 200 to 10,037.

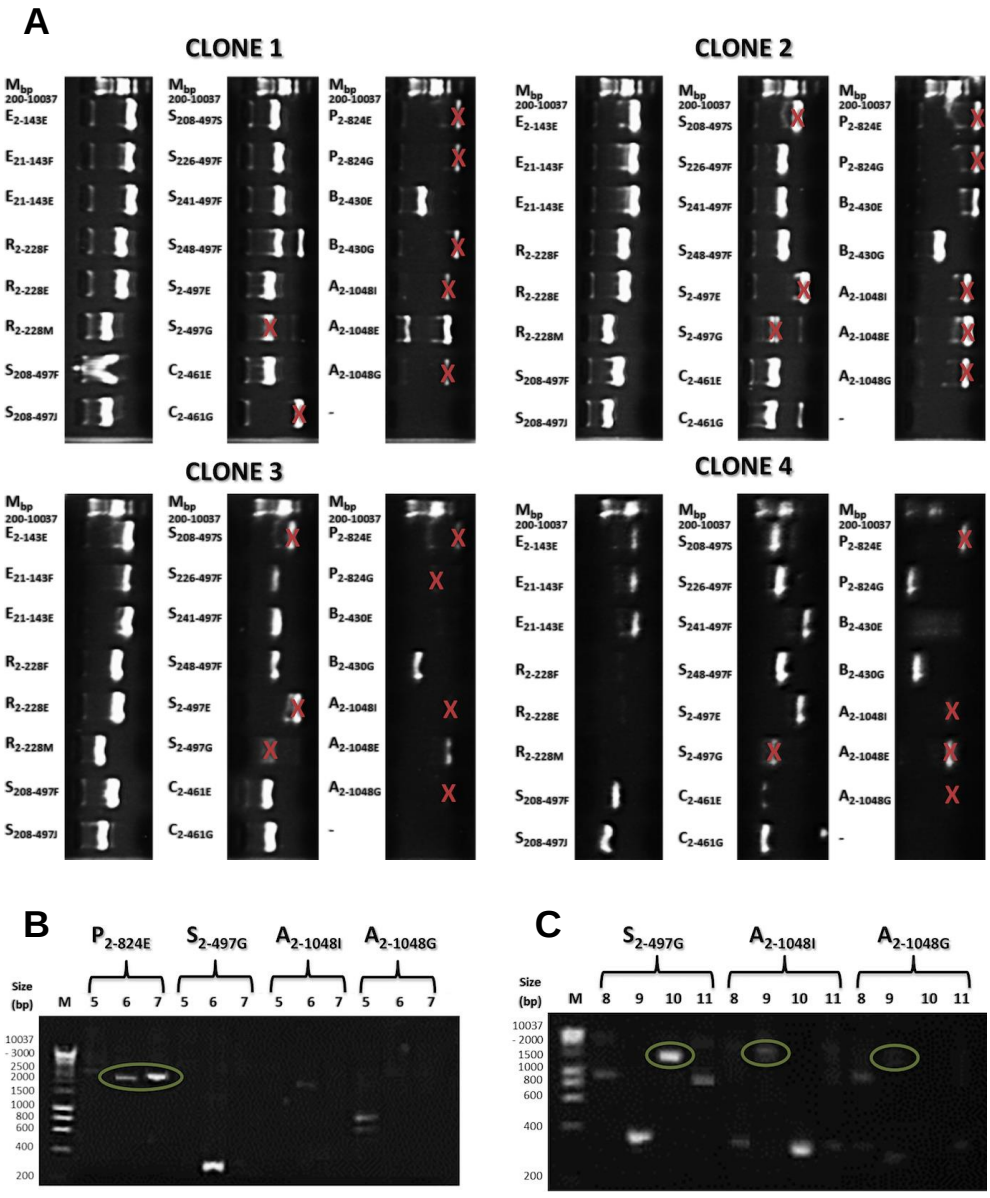


Figure 3. 9 **Agarose gels of cloning verification**

1% TAE agarose gels (Chapter 2, Section 2.2.2.1) showing that eventually all 23 of the gene targets for the seven Sil proteins (P, A, B, C, R, S, and E), were cloned into their correct fusion plasmid expression vectors, as described in Tables 2.2 and 2.7, Chapter 2. The *sil* gene PCR amplicons correspond correctly to expected gene sizes (please see table 2.7, Chapter 2) alongside the marker - HyperLadder™ 1kb (Bioline Reagents Ltd., London, UK). **A.** Any incorrectly sized PCR products have been highlighted with a red cross and repeated successful clones (5-11) in **B.** and **C.** have been circled in green.

3.2.2.2 Small-scale soluble Sil protein overexpression trials

The level of expression of the Sil proteins and their solubility was evaluated by carrying out small scale expression trials, on all 23 Sil protein overexpression constructs using high-throughput screening; as described in Section 2.2.5.1. Figures 3.10 and 3.11 show SDS-PAGE analysis of small-scale overexpression trials (Chapter 2, Section 2.2.2.2) indicating that most of the Sil proteins expressed in at least one of the *E. coli* cell strains - B834 (DE3), Rosetta (DE3) pLysS, OverExpressTM C41 (DE3) or OverExpressTM C43 (DE3) (Table 2.1, Chapter 2). Successful overexpression was seen with the non-membrane proteins: periplasmic SilE and SilB, the cytoplasmic fragment of SilS, the cytoplasm located SilR; as well as the membrane protein SilP. Overexpression of these proteins was visible with the presence of protein bands on a SDS-PAGE gel (Figures 3.10 and 3.11) at the predicted molecular weights (Table 2.7, Chapter 2), post-induction. Unfortunately, the predicted membrane proteins SilC and SilA, in addition to full-length SilS (including the membrane residing sequence) did not overexpress under any of the conditions used, in any of the overexpression strains (Figure 3.11). The full-length of SilS has subsequently been overexpressed by David Casas-Mao (personal communication).

Since no differences in amino acid sequence between the original predictions (GenBank accession number AF067954.1) and the WGS were seen in SilE (Figure 3.7), and as this protein is thought to be a key component of the *sil* system (Randall *et al.*, 2015), SilE was chosen to take forward for biophysical and structural studies. SilE is a periplasmic protein

with its first 20 amino acids being removed on movement across the membrane to the periplasm (Silver, 2003). Yields of protein overexpression were found to be higher in SilE without its periplasmic signal sequence (Figure 3.10). This is often seen during the overexpression of recombinant proteins that carry a signal sequence that, for yet poorly understood reasons, are known to be toxic to the cell (Schlegel *et al.*, 2013). As yields were higher in the protein without a periplasmic sequence and because the signal sequence is removed in its biological environment, the $SilE_{21\text{-}143F}$ polypeptide (which expressed well under both IPTG and auto-induced conditions in both the B834 (DE3) and the Rosetta (DE3) pLysS *E. coli* cell strains) was chosen to characterise.

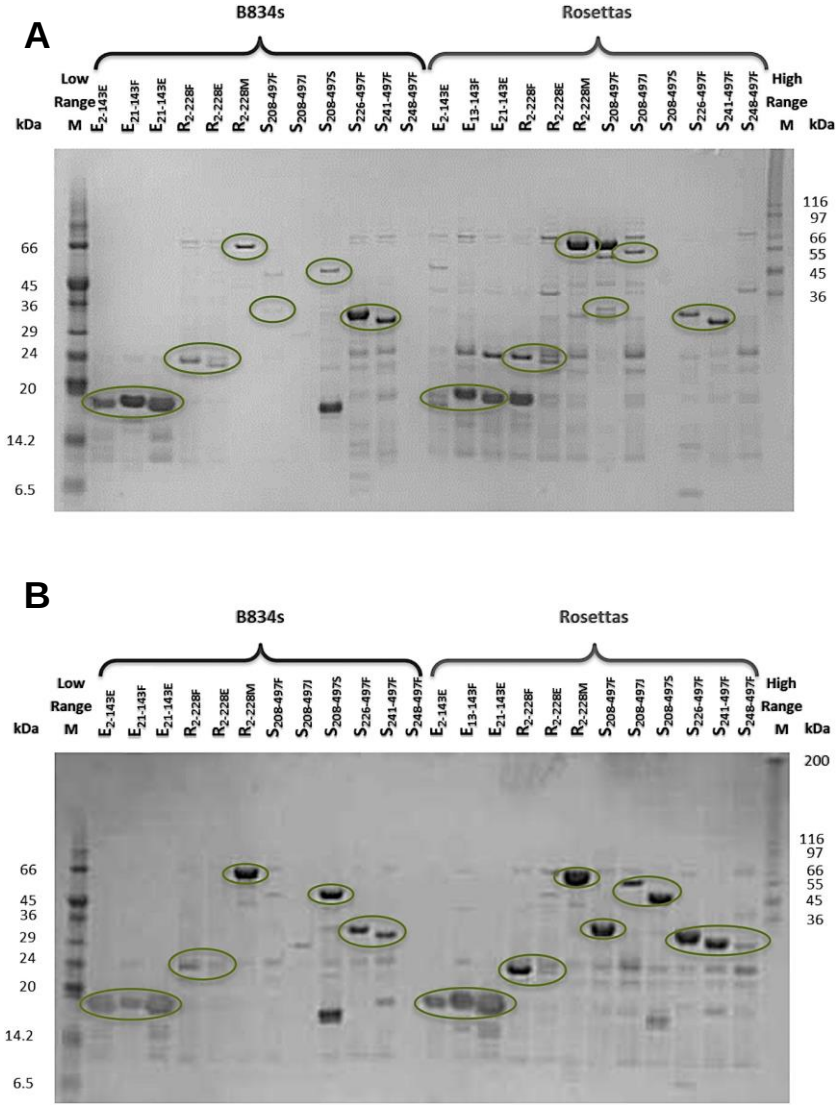


Figure 3. 10 SDS-PAGE analysis of non-membrane Sil protein expression screening experiments

The results of the expression screen of three non-membrane *E.coli* silver resistance proteins encoded by 12 variant gene constructs. The left shows the results of the screen carried out in the B834 strain of *E.coli* cells, whilst the right shows the results in Rosetta cells. **A.** Both sets of the expression screening experiments were carried out following overnight growth in Power Broth (containing the appropriate antibiotics), at 37 ^{o}C until the cultures attained an OD_{600} of ~0.5, at which point they were induced with IPTG to a final concentration of 1 mM. The growth and expression was then continued at 20 ^{o}C for a further 18 hours, with continuous shaking at 250 rpm. **B.** Both sets of the expression screening experiemnts were carried out following overnight growth in the auto-induction media TBONEX (containing the appropriate antibiotics), at 37°C until OD_{600} of ~0.5 and then at 25 ^{o}C for a further 20-24 hours, with continuous shaking at 250 rpm. On the whole, the expression screens seem to have worked for most of the proteins, circled green. (Please refer to Table 2.7, Chapter 2 for the correct molecular weight sizes). $SilE_{21\text{-}143F}$ expressed well under both conditions.

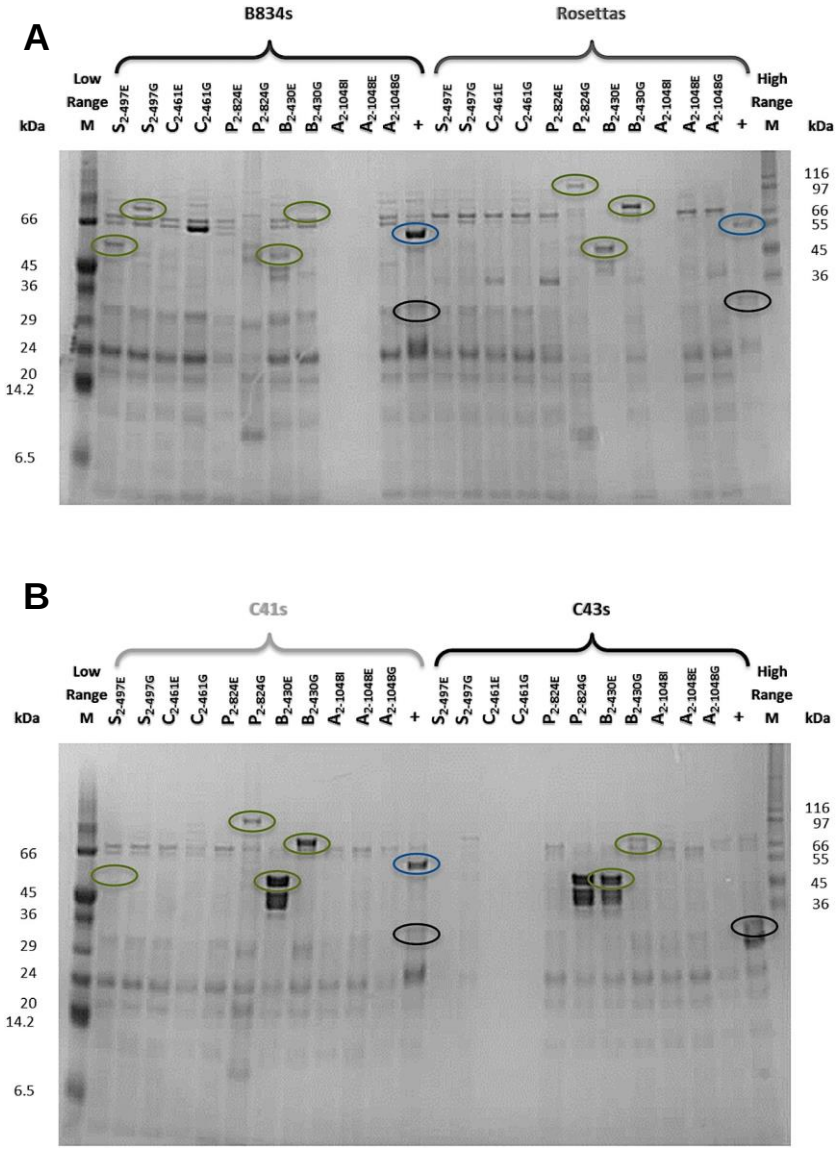


Figure 3. 11 SDS-PAGE analysis of membrane Sil protein expression screening experiments

Representative expression screening experiment results of the five membrane *E.coli* silver resistance proteins encoded by 11 variant gene constructs. All four sets of the expression screens were carried out following overnight growth in Power Broth (containing the appropriate antibiotics), at 37 ^{o}C until OD of ~0.5 at 600 nm at which point they were induced with IPTG to a final concentration of 1 mM. The growth and expression was then continued at 20 ^{o}C for a further 18 hours, with continuous shaking at 250 rpm. **A.** The left shows the products of the screen carried out in the B834 strain of *E.coli* cells, whilst the right displays the results in Rosetta cells. In these cells, the proteins that appear to have expressed correctly (circled green) are: $S_{2\text{-}497E}$, $S_{2\text{-}497G}$, $B_{2\text{-}430E}$, $B_{2\text{-}430G}$, (see Table 2.7, Chapter 2 for molecular weights). **B.** The left shows the protein bands from the screen carried out in the *E. coli* C41 strain, whilst the right displays the outcome in the C43 strain. The cells expressed $P_{2\text{-}824G}$, $B_{2\text{-}430E}$, $B_{2\text{-}430G}$ (circled green). '+' is a control that did not contain the correct eGFP (60 kDa, circled blue) but always contained a 32 kDa protein (circled black).

3.3 Summary and conclusions

The multi-resistant plasmid pMG101 contains *sil* - comprised of nine ORFs and is suggested to confer bacterial Ag^+-resistance (Silver *et al.*, 1999). Expression of the *sil* locus on pMG101 is said to be constitutively high in comparison to that of *sil* loci found in other IncH plasmids, which do not confer Ag^+-resistance (Gupta *et al.*, 2001). Additionally, the *sil* genes are usually found adjacent to the copper resistance *pco* gene cluster (Hao *et al.*, 2015; Hobman and Crossman, 2015; Staehlin *et al.*, 2016). Therefore, in order to characterise the plasmid of the proposed Ag^+-resistance determinant, to gain a better understanding of neighbouring genes and analyse *sil* gene sequence conservation, plasmid pMG101 and the whole genome of *E. coli* J53 (pMG101) including the plasmid were sequenced.

The expected size of plasmid pMG101 is ~182 kb (Gupta *et al.*, 1999; Silver, 2003; Randall *et al.*, 2015) whereas the total size of our sequenced plasmid pMG101 (from *E. coli* J53 from Professor Simon Silver, University of Illinois College of Medicine at Chicago, Chicago, Illinois, USA) was 166 kb (approx.) (Table 3.1). Whilst PCRs of the operon showed the presence of the *sil* genes before the DNA was sent for sequencing (Figure 3.2) neither the *sil* nor *pco* genes were found in the plasmid sequence. Analysis using CARD and BacMet found ampicillin, chloramphenicol, tetracycline, streptomycin, and sulphonamide resistance genes, in addition to mercury resistance genes (McHugh *et al.*, 1975) and confirmed that the correct plasmid was sequenced. Based on the findings of Randall *et al.*, (2015) this difference in size could be proposed to be the result of transposition of the *sil-pco* locus into the host chromosome, where

chromosomal contamination during plasmid isolation could explain the presence of *sil* amplification products in Figure 3.2. Indeed, upon analysing the whole genome of a new, better-preserved strain of *E. coli* J53 (pMG101) (from Professor Anne Summers, University of Georgia, Athens, Georgia, USA) this was found to be likely. This is because, like Randall *et al.*, (2015), Tn*7*-like elements were identified in close vicinity to the *sil* genes. Our WGS sequence was inclusive of a *tnsD* gene, found upstream of the *sil* genes. The TnsD protein is associated with site-specific insertion of genomic islands into the chromosome (as explained in Chapter 1, Section 1.4.4). Tn*7*-like transposons are capable of creating large mobile genomic pathogenicity, resistance or 'fitness' islands (often >30 kb) in a number of bacteria (Peters *et al.*, 2014). Hence, it can be concluded that together, the *tn7-sil-pco* resistance island associated with conferring Cu^{+}/Cu^{2+}- and/or Ag^{+}-resistance form a ~32 kb transposon that is able to transpose into the host genome. As a consequence, bacterial cells are able to easily disseminate the proposed mechanisms of Ag^{+}- and Cu^{+}/Cu^{2+}-resistance to daughter cells, without having to pay the 'fitness cost' of plasmid carriage (Sandegren *et al.*, 2012). Moreover, the preserved neighbouring of *sil* and *pco* genes (Figures 3.1 and 3.5-B) (Hobman and Crossman, 2015; Randall *et al.*, 2015; Staehlin *et al.*, 2016) not only allows for easy co-resistance and co-selection, but also questions the possibility of cross-resistance between *sil* and *pco* systems.

In addition, amino acid sequence conservation was observed amongst the majority of the Sil proteins, including SilA, B, C, E and G. Proteins SilP, SilR, SilS and SilF showed some mutations of which most

were non-conservative and conformed within certain regions of the proteins in question (Table 3.4 and Figure 3.7). Nevertheless, there seems to be no correlation in regards to the residue type that mutated. This indicates that no importance should be given to those mutations for now, especially as the mutations could just be sequence errors from manual sequencing carried out in the 90s.

Finally, all of the Sil proteins were cloned, most overexpressed (SilB, E, P, R and S) and the decision was made to take $SilE_{21\text{-}143F}$ forward for structural and biophysical characterisation, with the aim to understand the Sil system better.

CHAPTER 4

Overexpression, Purification and Solution Characterisation of the Intrinsically Disordered, Periplasmic Protein - SilE

4.1 Introduction

SilE is thought to be a putative intrinsically disordered/unstructured protein (IDP/IUP), in the same manner as its copper-binding homologue protein PcoE (Zimmermann *et al.*, 2012). It has been postulated that the protein folds and forms a structure upon binding the silver cation (Ag^{+}) (Silver, 2003), however the precise role of SilE in silver resistance has not yet been experimentally validated (Mijnendonckx *et al.*, 2013).

In order to understand how SilE contributes to the mechanism of bacterial silver resistance, it was important to characterise the structure of the native protein using biophysical and structural methods. These techniques required large quantities of pure SilE protein. Therefore this chapter describes the overexpression and purification of recombinant $SilE_{21\text{-}143F}$ (the protein without its periplasmic localisation sequence – Chapters 2 and 3, Sections 2.2.4.4 and 3.2.2). The properties and characteristics of the protein found by: mass spectrometry (MS), far-UV circular dichroism (CD), nuclear magnetic resonance (NMR) and small-angle X-ray scattering (SAXS) are also described in this chapter.

Note: As $SilE_{21\text{-}143F}$ was the primary SilE protein that was tested; it will only be referred to as SilE throughout the remainder of this thesis (unless a reminder of the shortened sequence is required).

4.2 Results and discussion

4.2.1 Overexpression and solubility of SilE

The level of SilE overexpression and the protein's solubility was evaluated by carrying out an expression trial. SilE was overexpressed using *E. coli* BL21 Star™ (DE3) or Rosetta Star™ (DE3) pLysS cells. Rosetta strains supply tRNAs for rare codons, on a chloramphenicol-resistance plasmid, to enhance the expression of proteins that contain such codons which are rarely used in *E. coli*. Whilst, the Star strains offer high mRNA stability hence higher protein yields due to a reduction in the endogenous levels of the mRNA degradation protein, RNaseE. Chemically or electrocompetent cells (Chapter 2, Sections 2.2.3.4) were transformed with the pOPINF plasmid DNA containing $SilE_{21-143}$ either using heat shock or electroporation (Chapter 2, Sections 2.2.4.3).

A small scale (50 mL) SilE overexpression time course experiment was carried out. The following conditions were used – shaking at 200 rpm, temperature at 37 ^{o}C and induction with the addition of a final concentration of 1 mM IPTG, at an OD_{600} between 0.6-0.8. Figure 4.1 shows that soluble protein of the approximately correct apparent size (calculated molecular weight of 15.8 kDa - Table 2.7, Chapter 2) was seen to accumulate over four hours and overnight (O/N = 16 hours). These results suggested that soluble protein in sufficient quantities could be produced under these standard growth conditions and negated steps to optimise growth temperatures and induction conditions.

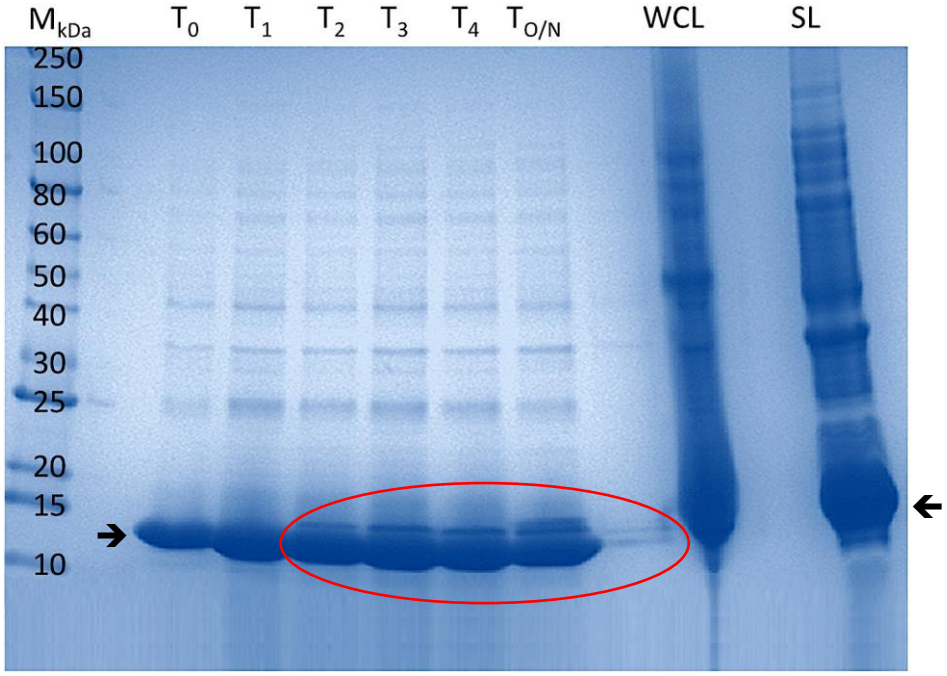


Figure 4. 1 **SDS-PAGE analysis of normalised samples from expression trial**

E. coli BL21 Star™ (DE3) or Rosetta StarT™ (DE3) pLsyS cells (Table 2.1) transformed with the $SilE_{21\text{-}143F}$ plasmid, were induced by adding 1 mM IPTG, at T_0, and left to incubate at 37 °C. A sample was taken every hour ($T_{1\text{-}4}$) and an overnight sample ($T_{O/N}$) post-induction and these were then examined for the presence of His_6-tagged SilE via gel electrophoresis (16% SDS-PAGE) alongside a high molecular weight protein marker (first lane) - Protein Ladder (10-250 kDa) (New England Biolabs, Ipswich, UK) – stained and de-stained as mentioned in Chapter 2, 2, Section 2.2.2.2. The gel illustrates how induction conditions of 1 mM IPTG and 37 °C facilitated the expression of SilE. The solubility of SilE was determined by analysing a sample of the whole cell lysate (WCL), which was acquired following cell lysis by sonication, and another sample taken from the soluble lysate (SL) after 15 minutes of centrifugation at 13,000 xg, 4 °C. The gel shows that induction with 1 mM IPTG at 37 °C yielded all soluble protein at 15.8kDa (black arrows). Note, the appearance of double bands (circled red) is explained in section 4.2.2.2.

4.2.2 SilE purification

A N-terminal His-tag was fused to the $SilE_{21-143F}$ protein, since His-tag purification is known to be a powerful method of separation (Bornhorst and Falke, 2000). The methodology used for protein extraction and purification throughout this section has been described in Chapter 2, Section 2.2.5.2 and for SDS-PAGE in Chapter 2, Section 2.2.2.2, unless mentioned otherwise.

4.2.2.1 Extraction

E. coli BL21 Star™ (DE3) or Rosetta Star™ (DE3) pLysS expressing soluble SilE protein cells were lysed as follows. Cells were harvested by centrifugation and the cell pellet was re-suspended in an appropriate amount of Wash Buffer (Chapter 2, Section 2.1.5.2). The cells were then stored at -20 ^{o}C and thawed at room temperature when required. Lysozyme was added to the Wash Buffer to a final concentration of 1 mg mL^{-1} and 1 mL of protease inhibitor cocktail in DMSO (for use in purification of His-tagged proteins) per 20 g of cells was added immediately before sonication. The sonication was performed and the soluble protein was then separated from the cell debris by centrifugation, with the supernatant being retained for nickel affinity chromatography.

4.2.2.2 Immobilised metal affinity chromatography

His_6-tagged SilE was purified by nickel affinity chromatography. Following protein preparation and soluble lysate filtration, His_6-tagged SilE

was loaded onto a HisTrap™ HP, nickel sepharose column. Protein was eluted via a step-wise elution gradient, using Elution Buffer (containing 500 mM imidazole - Chapter 2, Section 2.1.5.2). Figure 4.2 shows an 18 % SDS-PAGE analysis of the collected fractions. The PAGE gel shows significantly purer His_6-tagged SilE protein (band at 15.8kDa) eluted using an imidazole concentration gradient of approximately 200 to 400 mM. A large number of weak-column binding impurities (*E. coli* proteins) were washed out of the column at the low imidazole concentration (flow through (FT) and wash (W) 50 mM imidazole).

Interestingly, when run on a SDS-PAGE gel the SilE protein appeared anomalously as two bands on the gel (Figures 4.1 to 4.4); an initial indication that SilE is an intrinsically disordered protein (IDP). Proteins of this nature are known for their aberrant migration and appearance as multiple bands during SDS-PAGE (Tompa, 2002). One such study even found that XPA – a DNA repair protein, migrated aberrantly in denaturing SDS-PAGE as two bands. Modification of the sulphydryl groups of the cysteine in XPA resulted in the double bands becoming single (even if the size seen for the protein was still aberrant) showing that it may have been the formation of disulphide bonds in the oxidising conditions of the gel causing the double band (Iakoucheva *et al.*, 2001). This could not have been the case for SilE as it has no cysteine residues. It is thought that aberrant migration of IDPs is caused by their unusual amino acid structure leading to the binding of less SDS than a normal protein (Tompa, 2002). It might be that when disordered, SilE has two equally stable conformations, one of which allows a different level of SDS to bind when denaturation is

attempted. Mass spectrometry was therefore also used, following the final purification step, to confirm that only a single species was present in solution (Rhodes *et al.*, 2009).

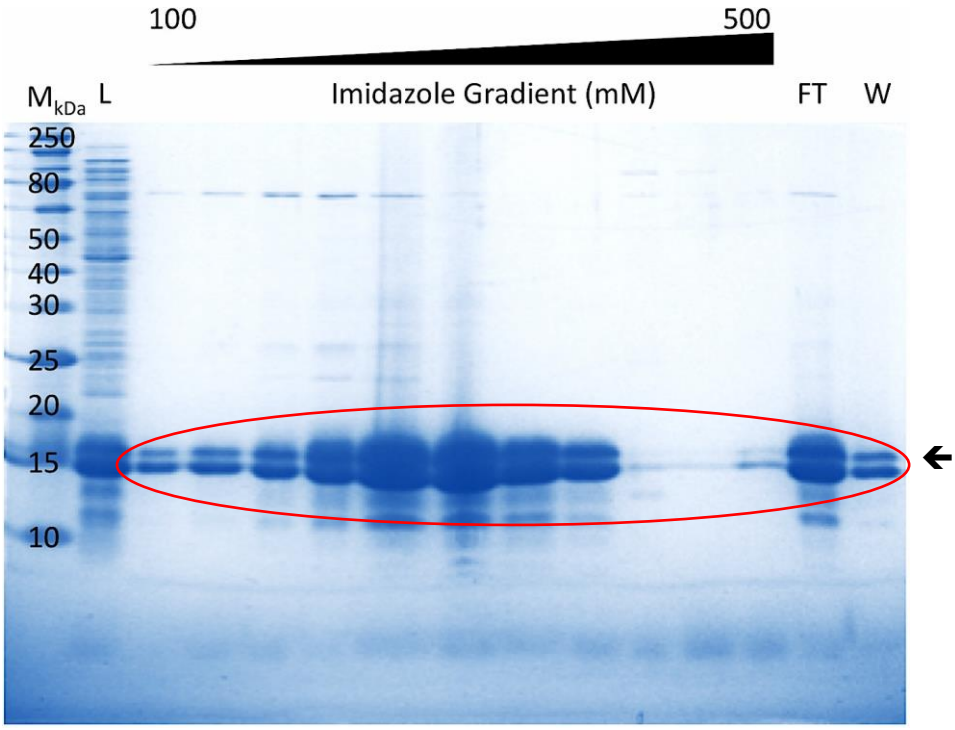


Figure 4. 2 **Nickel affinity chromatography based protein purification**

E. coli BL21 Star™ (DE3) or Rosetta Star™ (DE3) pLysS *E. coli* (Table 2.1, Chapter) cells transformed with the His_6-tagged $SilE_{21-143F}$ plasmid, were induced by adding IPTG to a final concentration of 1 mM, and left to incubate overnight at 37 ^{o}C. Expression of SilE (with a His_6-tag) can be seen in the lysed, Load sample (L). The total protein preparation, was run on a HisTrap™ HP, nickel sepharose column in a loading buffer containing 500 mM NaCl, 50 mM Tris-HCl, at pH 8.0. The column was washed in a step-wise gradient, with a second buffer containing excess imidazole Elution Buffer (500 mM NaCl, 50 mM Tris-HCl, 500 mM imidazole at pH 8.0). The flow through (FT) and washing (W with 50mM imidazole) eliminated most of the protein contaminants. His-tagged SilE (15.8 kDa – black arrow) bound to the column via histidine-nickel interactions was predominantly eluted and collected in fractions collected after addition of elution buffer containing 200-400 mM imidazole. Selected fractions eluted from the HisTrap™ HP column were then analysed, besides a high molecular weight protein marker - Protein Ladder (10-250kDa) (New England Biolabs, Ipswich, UK) - using 18 % SDS-PAGE and the gel was stained and de-stained as mentioned in Chapter 2, Section 2.2.2.2. Note, the appearance of double bands (circled red) is explained in section 5.2.2.2.

4.2.2.3 Gel filtration chromatography

The fractions eluted from the HisTrap™ HP, nickel sepharose column using a 200 to 400 mM imidazole concentration gradient were pooled together and injected onto a pre-equilibrated HighLoadTM 16/60 SuperdexTM 75 pg gel filtration column. The Superdex Gel filtration column was used to remove any non-specific binders from the fractions collected from the HisTrapTM HP, nickel sepharose column. The chromatogram in Figure 4.3-A shows two large peaks of protein eluting from the column at fractions 11-14 and 16-17. The 18 % SDS-PAGE gel presented in Figure 4.3-B shows that fractions 11-14 contained the most, pure SilE protein. These fractions were pooled together for further purification. SDS-PAGE of the other peaks did not show a visible contaminating protein or fragmentation of the protein. Still, fascinatingly you can see SilE throughout the other fractions i.e. anomalous passage through gel filtration, which is characteristic of an IDP (Dunker *et al.*, 2001; Uversky, 2002b).

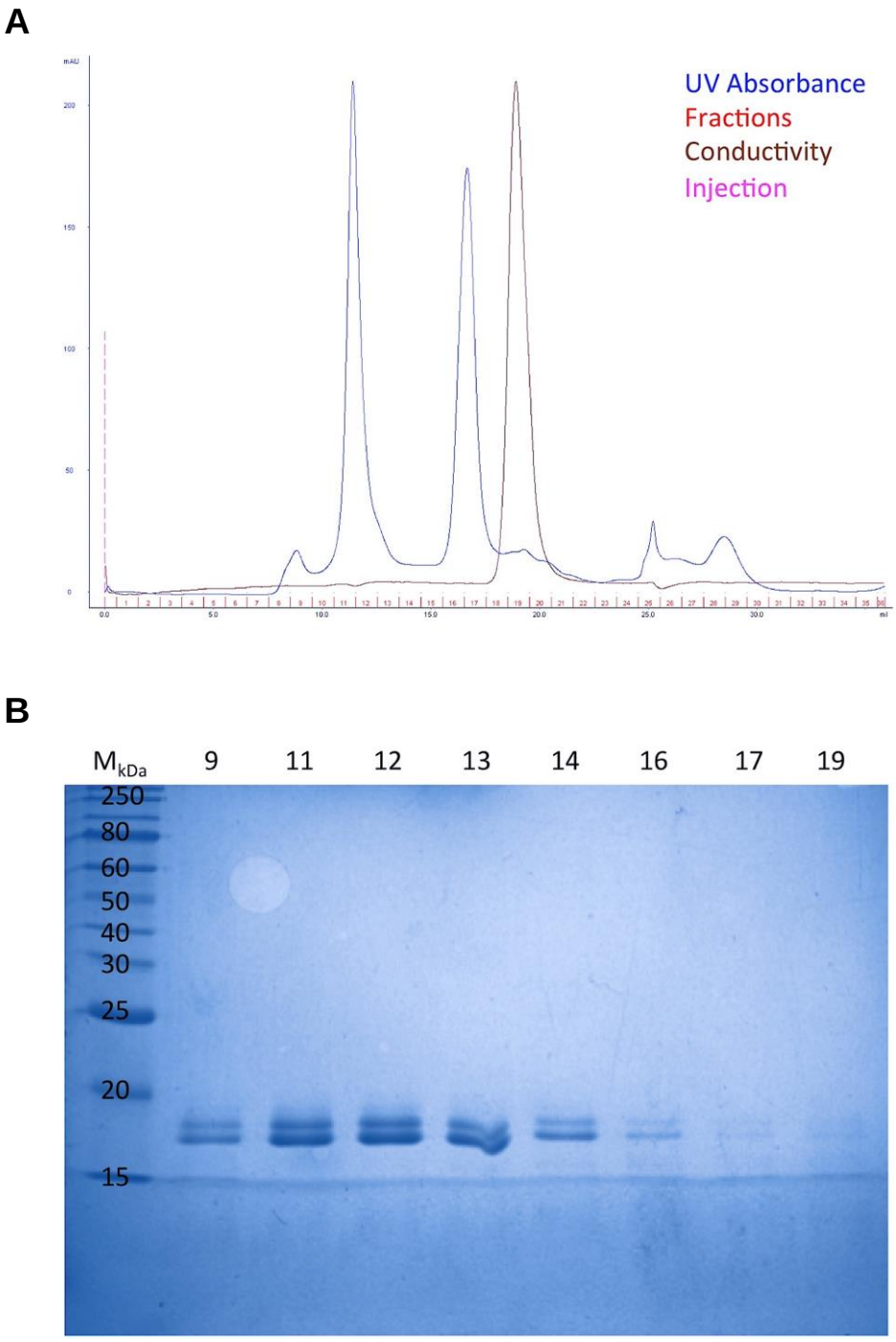


Figure 4. 3 Protein purification by gel filtration

A. Chromatogram showing the protein elution profile of gel filtration chromatography. Protein taken from the nickel affinity column (200-400 mM imidazole elution fractions) (Figure 5.2) was injected and run through a HighLoad™ 16/60 Superdex™ 75 pg gel filtration column alongside buffer (20 mM Tris, 200 mM NaCl, pH 8.0). The UV absorbance (blue) peak at fractions 11-14 contains the highest quantity of pure $SilE_{21\text{-}143F}$ protein as can be seen on the 18 % SDS-PAGE below – **B.** Protein Ladder (10-250kDa) (New England Biolabs, Ipswich, UK) and gel stained and de-stained as described in Chapter 2, Section 2.2.2.2. The protein being present throughout the other peaks is an irregularity associated with most IDPs. Double bands featured are explained in section 5.2.2.2.

4.2.2.4 HRV-3C protease treatment and HisTrap™ purification

Following HRV (Human rhinovirus) -3C protease (prepped in-house) treatment and another round of HisTrap™ HP nickel sepharose column purification of His_6-tag-cleaved SilE, an 18 % SDS-PAGE gel was generated and can be visualised in Figure 4.4. With the N-terminal His_6-tag removed, the SilE protein (13.3 kDa) passed through the column first, at low imidazole concentrations (10-100 mM). The low imidazole concentration was required due to the significant number of histidines in the protein. Any uncleaved SilE His-tagged protein and other contaminants such as the HRV-3C protease were then eluted by the addition of excess Elution Buffer (encompassing 500 mM imidazole – Chapter 2, Section 2.1.5.2). The gel in Figure 4.4 shows how the contaminants (HRV-3C Protease and His_6-tag) that appeared in the loaded sample (+3C) were removed with the Elution Buffer, in comparison to the low imidazole (10-50 mM) washes that contained the His_6-tag-cleaved SilE. The majority of the protein eluted at 50 mM imidazole.

Figure 4.5 shows the 4–20% protein gel concluding the successful purification of SilE, with and without the His_6-tag, with average protein yields of 30 mg L^{-1} culture.

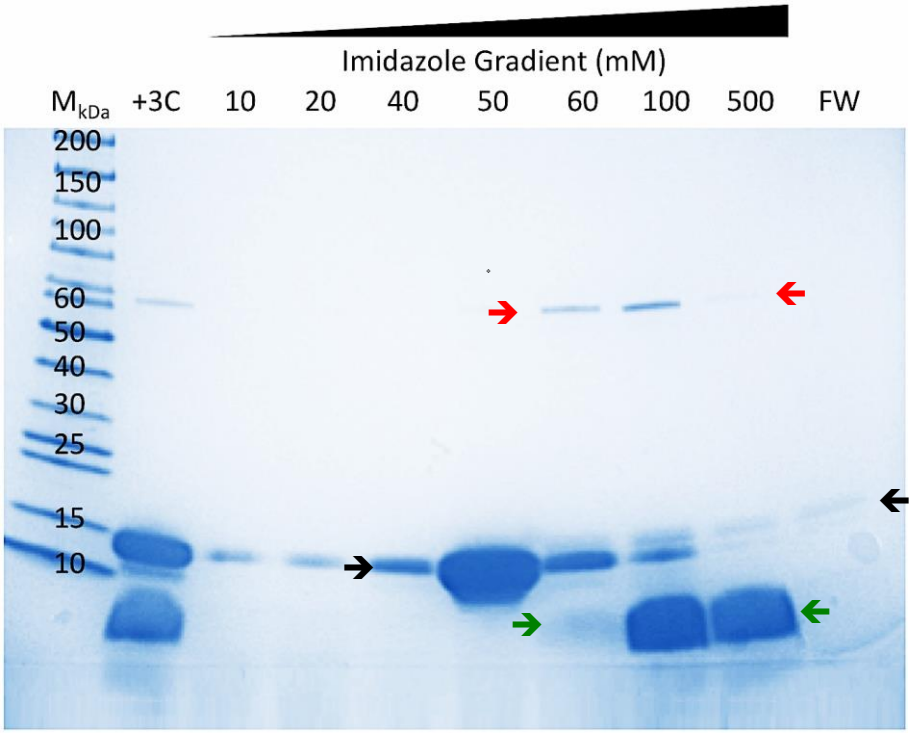


Figure 4. 4 HRV-3C protease treatment and protein His-tag removal

His_6-tagged SilE was treated with HRV-3C protease overnight (+3C), to remove the N-terminal His_6-tag. The protein-protease mix was then loaded onto a HisTrap™ HP, nickel sepharose column in a buffer containing 500mM NaCl, 50mM Tris-HCl, at pH 8.0, with varying concentrations of imidazole (up to 500 mM). Without the His_6-tag, the $SilE_{21\text{-}143F}$ protein (13.3 kDa, black arrows) eluted in the first sets of washes (largely at 50 mM imidazole) and the His_6-tagged HRV-3C protease (red arrows) and His_6-tag (green arrows) eluted at higher imidazole concentrations. The fractions were analysed besides a high molecular weight protein marker - PageRuler Unstained Protein Ladder (Thermo Fisher Scientific, Paisley, UK) - using SDS-PAGE and the gel was stained and de-stained as described in Chapter 2, Section 2.2.2.2.

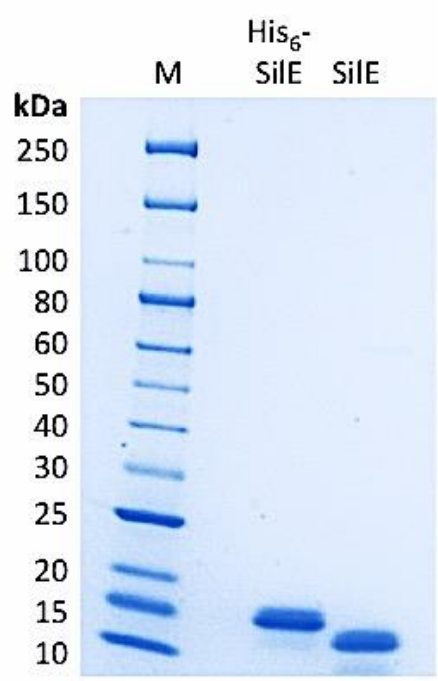


Figure 4. 5 Pure SilE protein

Representative SDS-PAGE of pure His_6-tagged and tag-cleaved SilE at 15.8 and 13.3 kDa. Proteins were analysed with a high molecular weight protein marker - Protein Ladder (10-250 kDa) (New England Biolabs, Ipswich, UK) - using a 4-20 % precast protein gel (Mini-PROTEAN® TGX™ - Bio-Rad, Hertfordshire, UK). The gel was stained and de-stained as described in Chapter 2, Section 2.2.2.2.

4.2.3 Molecular weight of SilE by mass spectrometry

Prior to using the protein for biophysical studies, confirmation was required that the correct protein ($SilE_{21\text{-}143F}$) had been produced and purified. Therefore, it was important to check the molecular weight of the protein. Nano-spray ESI-MS (with the aid of Dr. Matthew Jenner, University of Nottingham) allowed more accurate validation than visualisation by SDS-PAGE alone.

As macromolecules typically have a large number of charge states, they can be separated within the mass analyser (explained in Chapter 2, Section 2.2.6.2). Figure 4.6 shows the mass spectrum of SilE at a concentration of 90 μM; with distribution charge states from 6+ to 17+. Spectra were recorded under denaturing conditions, in 25 mM ammonium acetate ($C_2H_3O_2NH_4$), pH 7.0. The molecular mass was deconvoluted from the mass to charge ratio (m/z) observed in the spectrum, using Waters MassLynx 4.1 software (Figure 4.6B). The deconvolution uses an algorithm to calculate the molecular mass. It was noted that for each charge state there was a major peak with a molecular weight of 13, 268.46 Da (± 1.6 Da), in good agreement with the prediction of the molecular weight of $SilE_{21\text{-}143F}$ (13, 273.19 Da), calculated using monoisotropic masses (Gasteiger *et al.*, 2003).

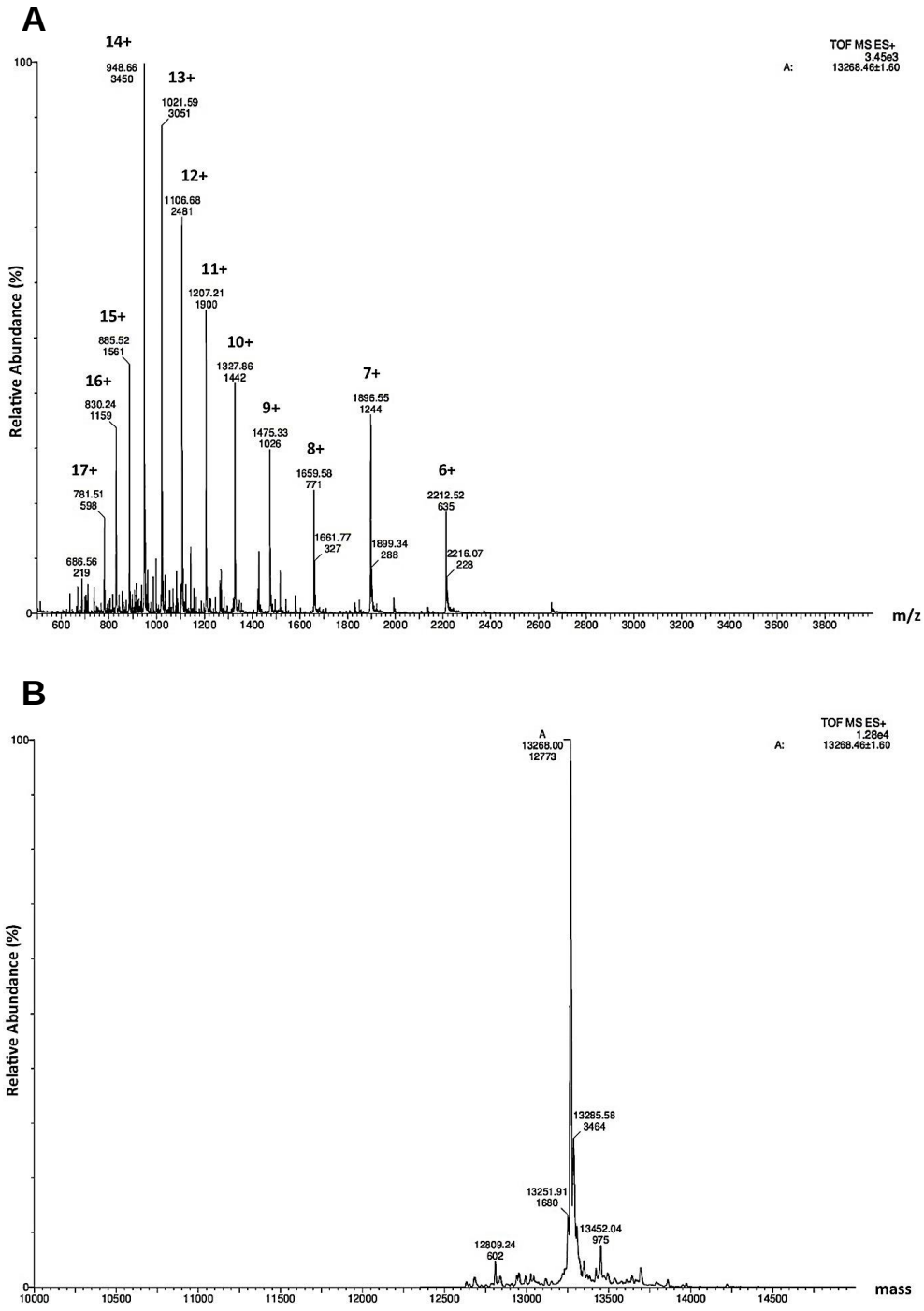


Figure 4. 6 Mass spectrum of SilE

A. Nano-spray ESI-MS spectrum of SilE at 90 μM, in 25 mM ammonium acetate, pH 7.0. Analysis of the charge states (in bold) gives a molecular weight for the largest peak, which can be seen. **B.** Transformed data showing the protein sample has a mass of 13,268.46 (± 1.6) Da after deconvolution.

4.2.4 Determining disorder in SilE

The protein structure-function paradigm states that the function of a protein is dependent on its tertiary structure, which is ultimately determined by its primary amino-acid sequence. This paradigm fundamentally conditions that a protein should present a defined and ordered tertiary structure in its native state; an essential prerequisite for protein function (Dunker *et al.*, 2001). Counter to this theory, recent work has demonstrated the existence of IDPs thereby calling for a re-assessment of the classical structure-function paradigm (Wright and Dyson, 1999; Uversky, 2002a; Uversky, 2002b). IDPs or IUPs (intrinsically unstructured proteins, also known as natively unfolded proteins - NUPs) have no single well-defined structure in their native, functional state. They exist as a rapidly fluctuating ensemble of conformations (Lee *et al.*, 2014) without a unique, well-defined state of equilibrium (Romero *et al.*, 1998; Wright and Dyson, 1999). Being very flexible, they may also adopt relatively rigid structures in the presence of natural ligands. NUPs, with their lack of ordered structure, represent attractive targets for biophysical studies of the unfolded polypeptide chain under physiological conditions *in vitro* (Uversky, 2002b). This disordered state can also be well characterised by an unusual amino acid composition – a low level of overall hydrophobic and a high level of charged residues (Romero *et al.*, 1998; Dunker *et al.*, 2001).

Silver *et al.*, (1999) described SilE as being 'predominantly unstructured' in its native form, however no experimental evidence was provided. In addition, SilE shares 48 % sequence identity with the homologous copper binding protein PcoE (Zimmermann *et al.*, 2012; Asiani

et al., 2016; Figure 6.2-A Chapter 6), which has also been proposed to be intrinsically disordered. To test whether SilE is intrinsically disordered in its native form, various studies including bioinformatics based predictions and biophysical experiments were conducted. Having obtained pure protein, optical spectrophotometry based CD, as well as NMR and SAXS experiments presented information on the structural content, spatial distribution and folding properties of SilE. (See Chapter 2, Section 2.2.6 to understand how these biophysical techniques provide this data).

4.2.4.1 Prediction of intrinsic disorder in SilE

Biochemical basis of disorder and prediction tools

IDPs can be accurately distinguished based on their primary amino acid composition. It is known that all of the necessary information for the correct folding of an ordered protein into its rigid biologically active conformation relies upon its amino-acid sequence. Nonetheless, the lack of regular structure in IDPs raised the question of specific features and properties found in this characteristic set of proteins. Comparison of the overall hydrophobicity and net charge of native (folded) proteins, and natively unfolded protein primary residue sequences, gave rise to some sequence biases coding for IUPs. Some of these common particularities found in the sequences of IDPs are: a lack of aromatic residues, many polar and charged groups (mainly negatively charged glutamic and aspartic acids) producing an overall large net charge at neutral pH and low quantities of bulky hydrophobic amino acids (Uversky, 2002b). This prevents hydrophobic interactions between amino acids and increases charge

repulsion; preventing sufficient favourable interactions to promote tertiary structure formation, resulting in an extended protein structure. Therefore, based on calculation for the predisposition of disorder for each amino acid, local regions of disorder can be determined, as shown in Figure 4.7. Hence, early intrinsic disorder predictors were based on charge-hydropathy (CH) plots, which clearly differentiated proteins of disordered content from natively folded ones; by analysing the combination of low mean hydrophobicity and relatively high, normalised net charge (Uversky *et al.*, 2000; Ferron *et al.*, 2006).

Initially, in 1979, Williams suggested using amino–acid sequence analysis for identifying proteins that form random coils rather than globular structures (He *et al.*, 2009). Yet it was not until 1997 that the first predictor of intrinsically disordered regions (IDRs) was constructed. This was only limited to training against a dataset of disordered protein regions, known as neural network predictors (NNPs) formed on the basis of PDB entries (Romero *et al.*, 1998; Radivojac *et al.*, 2007). NNPs have generally been trained on either NMR solution-based or X-ray structure data. The X-ray structure trained NNPs, built on regions lacking electron density in the crystal structure, are also commonly known as REMARK 465. Since then, bioinformatics has played an important role in identifying the important prerequisites for the absence of compact structure in proteins under physiological conditions. Now, many algorithms have been designed to predict NUPs based on the biochemical properties and biased amino acid compositions of the proteins (He *et al.*, 2009).

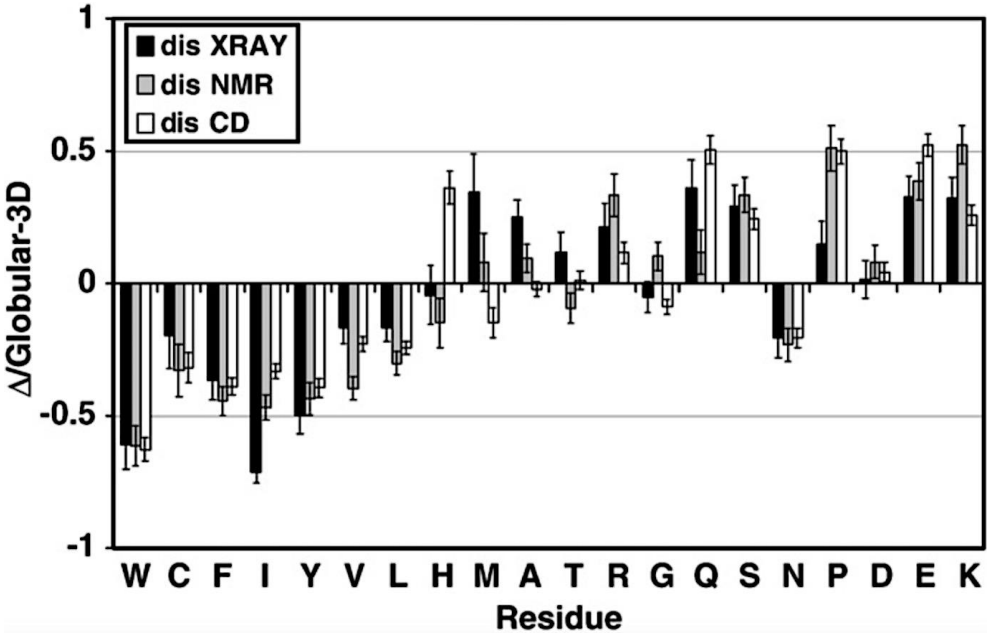


Figure 4. 7 Amino acid order/disorder IDP composition profile

Graph obtained and used from Uversky and Dunker, (2010), following permission from publisher. A comparison of amino acid compositions of ordered proteins, with each of the three databases of disordered proteins. The ordinates are (% amino acid in disordered dataset - % amino acid in ordered dataset)/(% amino acid in ordered dataset) = Δ/globular_3D. The residues are ordered according to the Vihinen's flexibility scale (Vihinen *et al.*, 1994). The names of each database indicate how the disordered regions were identified. Negative values indicate that the disordered database has less than order, positive indicates more than order.

Today, more than 60 sophisticated computational programs exist, which use various statistical and machine learning methods relying on current intrinsic disorder (ID) information or amino acid compositions and their derived properties to predict ID (Ferron *et al.*, 2006; Radivojac *et al.*, 2007; He *et al.*, 2009; DeForte and Uversky, 2016). Many of these *in silico* predictors can be found in the Disordered Protein Database (DisProt - http://www.disprot.org/predictors.php) (Sickmeier *et al.*, 2007). Current predictors now have an accuracy of 60-70 % for predicting disorder on a per

residue basis, with a best overall accuracy from the CH method at 83 % accuracy (Ferron *et al.*, 2006).

Predicting disorder in SilE

As is the case with most bioinformatics based methods, predicting ID in a protein is also more reliable when using several processes and implementations that analyse different parameters. Consequently, a single ID predictor is insufficient to produce predictions that take into account the modular organisation of a protein, (Ferron *et al.*, 2006). Therefore as suggested in the study by Li *et al.*, (2015b), an assortment of predictors relying on different concepts should be combined (i.e. meta-prediction), to provide useful consensus data (Kozlowski and Bujnicki, 2012). Thus, alongside prediction based on sequence properties and machine learning against a training set, the third general prediction strategy uses a meta-predictor that combines several independent prediction methods (Lee *et al.*, 2014).

Since 2004, due to the increased interest in the phenomenon of IDPs, disorder prediction has been included as part of the biennial Critical Assessment of Techniques for Protein Structure Prediction (CASP); experiments and assessments aimed at monitoring the state of the art ID prediction from protein sequence (Uversky and Dunker, 2010). The assessors of the recent set of data presented in CASP10 ranked PrDOS-CNF (Protein DisOrder prediction System-Conditional Neural Fields) as the best overall predictor, with an estimated accuracy of >90 % (Eickholt and Cheng, 2013; Monastyrskyy *et al.*, 2014; Jianzong Li *et al.*, 2015b). PrDOS-

CNF is a meta-prediction system composed of two predictors whose results are combined to give a final two-state prediction (order/disorder), alongside a disorder probability for each residue. The first algorithm is based on the analysis of the physico-chemical properties of the primary sequence and the second, based on the formation of a NNP from a template of proteins (i.e. homologous proteins for which structural information is available) (Ishida and Kinoshita, 2007). Figure 4.8-A shows the PrDOS prediction of $SilE_{21-143}$. The recommended false positive threshold of 4 % was used; with a cut-off of 0.55 to generate data showing us that 72 % of the protein contained IDRs. The numbers not only showed that SilE had a high proportion of disorder, but also that the disordered regions were not localised within the protein and could be found in sections throughout SilE. The regions of disorder were predicted to be associated with residue numbers: 21-30, 35-60, 71-79, 85-104 and 121-127 (Figure 4.8-A).

The CASP9 and CASP8 winning methods both come from the GeneSilico MetaDisorder Service - GSMetaDisorderMD2 and GSMetaDisorder, respectively (Noivirt-Brik *et al.*, 2009; Monastyrskyy *et al.*, 2011; Kozlowski and Bujnicki, 2012; Eickholt and Cheng, 2013). Both of these meta-servers use genetic algorithm and different weighting schemes to average models from 13 disorder-predicting servers including: DISOPRED2, DISpro, DISPROT (VSL2), GlobPlot, SPINE-D, IUPred (long and short), Pdisorder, Poodle (long and short), RONN, Spritz (long and short), DisEMBL-coils, DisEMBL-hotloops, DisEMBL remark465 (Kozlowski and Bujnicki, 2012). Predictions of ID using both GSMetaDisorderMD2 and GSMetaDisorder in Figure 4.8-B both showed a

much higher tendency of disorder throughout the entire sequence of SilE. Here, each residue whose disorder probability was over 0.5 was considered to be disordered and even though both sets of data followed the same trend, the scores of disorder were much higher from MetaDisorder compared to MetaDisorderMD2. Nevertheless, this again was indicative of the fact that SilE is an IDP, albeit 100 % disordered.

Figure 4.8-C displays an overlay of the three sets of prediction results from PrDOS, MetaDisorderMD2 and MetaDisorder. Without using a cut-off value, it is still clear to analyse how each data set followed the same movement and showed similar amino acids (21-30, 35-60, 71-79, 85-104 and 121-127) as being the high proportion of unstructured regions within the protein.

Due to their high accuracy, *in silico* disorder prediction algorithms have become an invaluable tool for providing preliminary information and directing further experimental studies. For SilE, the data from this approach in Figure 4.8 showed that there is a high propensity for disorder throughout most of the protein. This was useful initial information that led us to the biophysical experiments, which were used to test and confirm the protein's state of disorder *in vitro*.

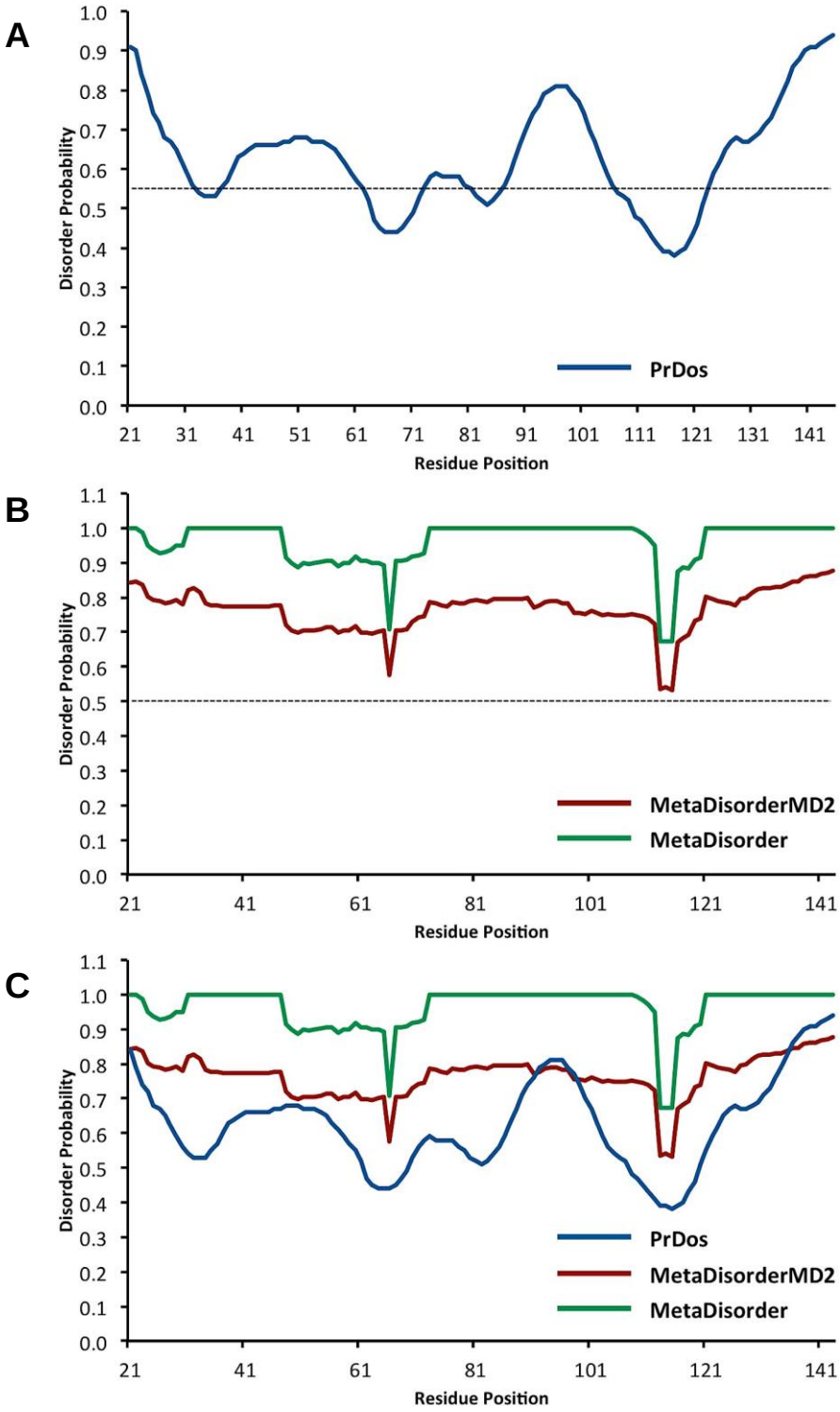


Figure 4. 8 **Intrinsic disorder predictions of $SilE_{21-143}$**

A. Prediction of $SilE_{21-143F}$ intrinsic disorder from PrDOS-CNF, cut-off 0.55. **B.** Prediction of intrinsic disorder using GS-MetaDisorder MD2 (maroon) and GS-MEtaDisorder (green), cut-off 0.5. **C.** An overlay based on the predictions from A. and B.

4.2.4.2 Assessment of folding, spatial distribution and structural content of SilE

Under physiological conditions, IDPs show none to low levels of ordered secondary structure and essentially a non-compact packed core. Several physico-chemical (biophysical and biochemical) spectroscopic methods have been found to be powerful tools for the detection of IDPs. These methods (a comprehensive list of which can be found in a review by Radivojac *et al*., 2007) can characterise protein self-organisation and discriminate between unfolded, and partially folded to folded states (Wright and Dyson, 1999; Uversky, 2002a; Radivojac *et al*., 2007). The list includes the methods used throughout this study.

The techniques used in this study include: far-UV CD, NMR and SAXS (for detailed explanations of how these experiments provide useful data, see Chapter 2, Section 2.2.6). Hence having obtained pure protein, these three methods were used in order to gain a fuller understanding of the lack of structural content, the unfolded state and spatial distribution properties of SilE.

Using CD to determine the secondary structure content of SilE

Secondary structure can be defined as the interactions that occur between the C, O, and NH groups of amino acids in a polypeptide chain which cause formation of α-helices, β-sheets, turns, loops, and other forms, and which also facilitate the folding into a three-dimensional structure.

Far-UV (or the amide region - i.e. spectral region 170-250nm) circular dichroism (CD) is a solution-based technique that is a widely used

method, to estimate the secondary structure composition of a protein (Kelly and Price, 2000). Although far-UV CD does not provide structural information at an atomic level, a measure of the overall structure can be provided. Far-UV CD is therefore an ideal technique to study protein thermodynamic properties or folding mechanisms. To identify whether SilE had a defined secondary structure, the SilE protein sample at a concentration of 57 μM (in a 10 mM HEPES, 20 mM NaF buffer at pH 7.5) was examined using far-UV CD (Chapter 2, Section 2.2.6.1). With the use of Synchrotron Radiation CD (SRCD) measurements were made down to 170 nm to allow for significantly more reliable data (Kelly *et al.*, 2005).

An ordered structure results in a spectrum with positive or negative CD signals or ellipticities. So in this manner, CD is sensitive to secondary structures by being able to sense the differences in the relative amounts of α-helices and β-sheets (Sreerama *et al.*, 1999; Martin and Schilstra, 2008), which change according to protein environment. The presence of significant α-helices, forming a protein's secondary structure, is characterised by intense negative ellipticities around 208 nm and 222 nm, with a positive signal at approximately 193 nm. Whilst a negative ellipticity at approximately 215 nm and a positive signal at about 195 nm corresponds to predominantly β-sheets, within a protein's secondary structure. A negative signal around the wavelength region of 200 nm is characteristic of a classic random coiled protein secondary structure (Sreerama *et al.*, 1999; Greenfield, 2007a; Dodero *et al.*, 2011). Accordingly, unordered polypeptides have a characteristic far-UV CD spectrum, resembling a 'random coil' polypeptide – characterised by a strong minimum centred

around 200 nm and the absence of any strong negative signals above 205 nm (i.e. those typically found in α-helix or β-sheet structure). From the spectrum found in Figure 4.9-A, as expected, obvious indications of strong negative signals around 200 nm showed that SilE was predominantly a non-structured polypeptide comprised of random coil and thereby an IDP. Also, a slight negative shoulder on the CD spectrum at 221 nm was consistent with a minor fraction of α-helical secondary structure.

Methods have been developed to analyse the CD spectrum of a protein to determine the fraction of residues within each conformational type (Chen *et al.*, 1972; Bloom and Mann, 1979; Böhm *et al.*, 1992; Sreerama and Woody, 1993; Johnson, 1999; Sreerama and Woody, 2000). Dichroweb is an interactive webserver that facilitates the calculation of each type of secondary structure content from collected CD spectroscopic data, using popular algorithms (Lobley *et al.*, 2002; Whitmore and Wallace, 2004; Whitmore and Wallace, 2008). Figure 4.9-B shows results from when data obtained from the SRCD was submitted to Dichroweb (programme CONTINLL, which implements the locally linearised algorithm in selecting protein sets from the reference database) to determine a quantitative measure of each class of secondary structure (Provencher and Glockner, 1981; Van Stokkum *et al.*, 1990; Sreerama and Woody, 2000). The protein appeared to be around 74 % unstructured (39 % unordered and 35 % turn/loop), 20 % α-helical and 6 % β-sheet.

The normalised root mean square deviation or error (NRMSD/E – Eq. 4.1) is useful as a standard goodness-of-fit parameter, to compare how the experimentally obtained data correlates with the back-calculated

spectrum; reconstructed from the predicted secondary structure (Wallace *et al.*, 2003; Whitmore and Wallace, 2008). NRMSD is defined as:

$$\Sigma[(\theta exp - \theta cal)2/(\theta exp)2]1/2 \qquad \textit{(Eq. 4.1)}$$

This is the summation over all wavelengths, in which θexp and θcal are the experimental ellipticities and back-calculated spectrum ellipticities (Miles *et al.*, 2005). Upon experimental data submission to Dichroweb (programme CONTINLL) a value of NRMSD is provided alongside a visual plot, showing the comparison of the experimental data, the fitted data and their difference. Figure 4.9-C shows this graphical plot of SilE, which obtained a NRMSD of 0.07. Generally the lower the NRMSD is, the better the experimental data represented by the fitted curve. If the NRMSD is high (i.e. >0.1), then the calculated secondary structure is more likely to be incorrect (Whitmore and Wallace, 2004). However, a low value for the NRMSD is not always a necessary condition for confirming accurate secondary structure content prediction calculations. This is because it is not always possible to allocate a more appropriate reference dataset for the calculations, especially when datasets containing more of the protein of study's component do not exist (Brahms and Brahms, 1980). As the NRMSD for the CD data obtained from SilE of 0.077 was <0.1, the values calculated for secondary structure for the secondary structure of SilE were considered correct.

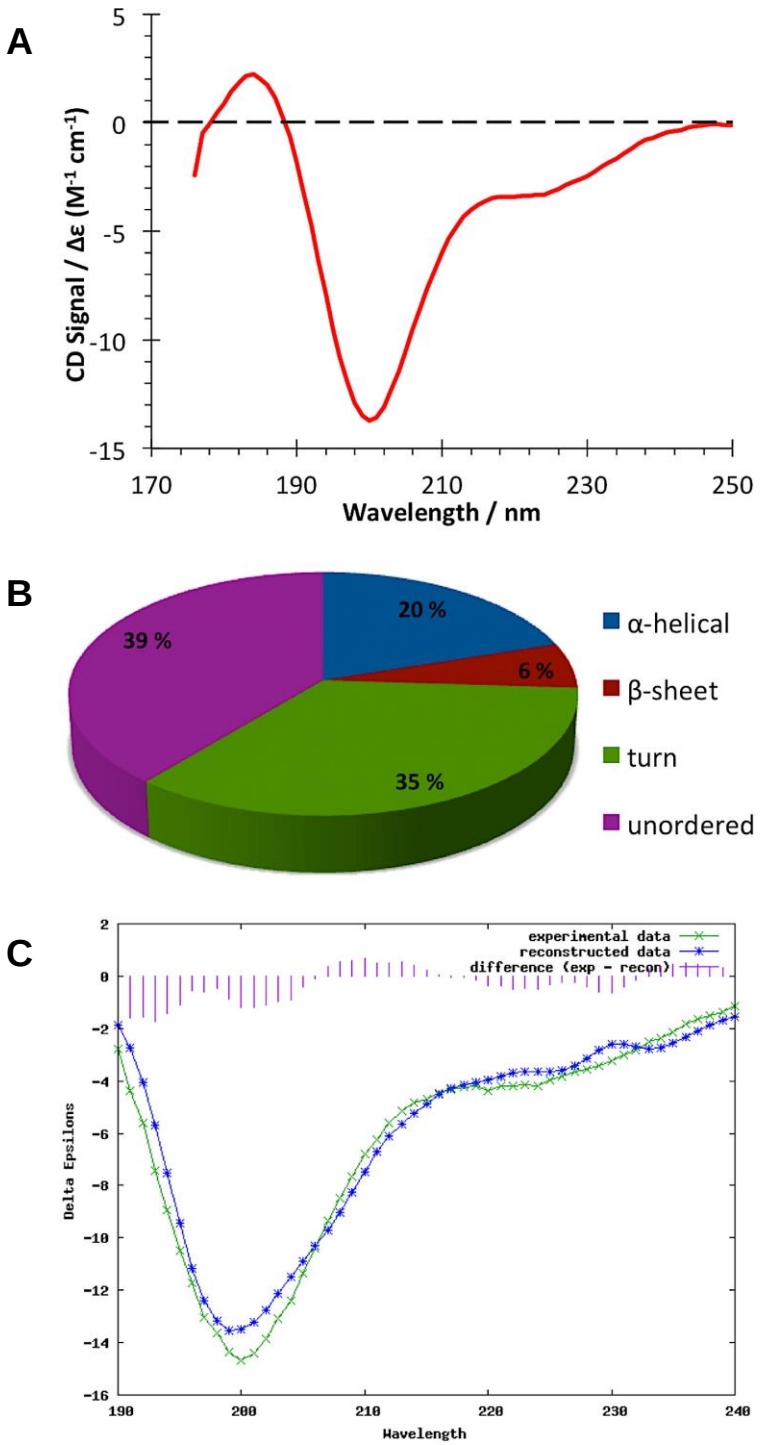


Figure 4. 9 **Secondary structure content of SilE**

A. Far-UV CD, using SRCD, of 57 μM SilE in 10 mM HEPES, 20 mM NaF buffer, at pH 7.5. The profile displays a strong negative signal at 200 nm, validating the significant presence of random coil i.e. unordered structural content. **B.** The secondary structure composition of SilE calculated using the CONTINLL algorithm within the Dichroweb server. The division of each class is as follows – 74 % unstructured (39 % unordered and 35 % turn/loop), 20 % α-helical and 6 % β-sheet. **C.** Following data submission to CONTINLL (Dichroweb), the back-calculated CD spectrum (blue), alongside the experimental data (green) and the difference of 7.7 % between the plots (pink).

Nuclear magnetic resonance applied to SilE

By showing that SilE lacked secondary structure, CD confirmed the prediction of it being an IDP. However, CD does not provide information on protein folding at the tertiary level. Therefore, a higher resolution method was required.

Intrinsically disordered systems are stabilised by interactions with solution and therefore an ideal high-resolution method would enable aqueous measurements, in order to retain native protein conformation. Nuclear Magnetic Resonance (NMR) is undoubtedly one of the most suitable tools for investigating IDPs, as it characterises dynamic behaviour and structure propensities in solution (Chapter 2, Section 2.2.6.3). The NMR spectrum of a protein sample dissolved in a solvent is well resolved, representative of a collection of signals for chemically distinct atoms within the molecule.

SilE protein NMR analysis was carried out to verify protein dynamics and obtain folding and structural information for single nuclear sites within the protein (Wüthrich, 1990). It could also judge whether further experimentation could be executed well to provide SilE's general 3D in-solution structure. Using one-dimensional (1D) proton (^{1}H) and two-dimensional (2D) ^{1}H-^{15}N Heteronuclear Single Quantum Coherence (HSQC) NMR, spectra of ^{15}N-SilE (protein overexpression and purification in Chapters 2 and 5, Sections 2.2.5.2 and 5.2.7.3 respectively), were recorded in 10 mM sodium phosphate (Na_2HPO_4/NaH_2PO_4), 25 mM NaCl, pH 7.5, 10 % D_2O, on an 800 MHz Bruker Avance NMR machine (Chapter 2, Section 2.2.6.3). Deuteration (the addition of deuterium, D_2O) was used

to enhance the data obtained, by improving the signal-to-noise ratio. This can be seen in Figure 4.10. The spectrum obtained from investigation in Figure 4.10-A is a resolution of chemical shifts in the ^{1}H direction (i.e. *x* axis) whereas; being 2D, Figure 4.10-B also determines the chemical shifts of ^{15}N in the opposite direction (*y* axis). In both sets of experiments, each residue/type of proton within a protein generally gives rise to one peak (or resonance line). In ^{15}N HSQC experiments this occurs at point of intersection of ^{15}N and ^{1}H (Howard, 1998). An exception to this is the proline residue, as it does not contain any backbone amide (NH) groups due to its nitrogen not being bound to any hydrogen when in a peptide bond (Pavlov *et al.*, 2009). Since each peak corresponds to a particular proton in the protein, $SilE_{21\text{-}143F}$ consisting of 125 amino acids should have presented roughly 1000 resonances in the ^{1}H spectrum and 125 peaks (one for each residue) in the ^{15}N HSQC spectrum. Nevertheless, neither spectrum showed that many signals (Figure 4.10), with the 2D ^{15}N-HSQC only showing ~40 resonances.

In NMR, the basic idea is that if a protein is folded the peaks in the spectrum are well dispersed and most can be individually distinguished due to their sharp and narrow appearance. The peaks provide a unique 'fingerprint' of the protein (Bodenhausen and Ruben, 1980). Instead, if the peaks are broader, there is a large cluster of overlaps and these are poorly dispersed around the middle of the spectrum; then there is an indication of significant unstructured and unfolded elements in the protein. Peaks become broad in an unfolded protein due to their chemical shifts being dependent on the protons local environment. In an unfolded protein, as the proton does not

have a distinct environment, the peak is an average of all of the different environments in which the proton is found. The averaging of these chemical shifts, based upon their local environment, is that much more difficult when a protein lacks a stable structure. Consequently, this results in a huge reduction of the chemical shifts ranges for the different nuclear spin species (^{1}H, ^{15}N) and a problem of strong overlap in the NMR spectra. Thus, disordered regions often give broad resonances, which can be difficult to define and assign (Page *et al.*, 2005).

Figure 4.10-A shows the spectrum obtained from a 1D experiment of ^{15}N-SilE at 300 μM. With little peak dispersion, i.e. no upfield shifted methyls, indicative of packing interactions and a narrow NH envelope (NH region is < 2 ppm, between 6.7 to 8.4 ppm) with no signals past 9ppm – that data was definitive that SilE was an unfolded protein. Nonetheless, spectral resolution in the ^{1}H dimension provides only limited information about through-bond and through-space connectivity. In contrast, 2D spectra present further information about connectivity and can show increased signal dispersion (Meyer and Zangger, 2014). For this purpose the 2D ^{1}H-^{15}N HSQC shown in Figure 4.10-B was carried out on 300 μM of SilE. This again displayed limited spectral resolution with few and overlapping proton resonances in a small spectral range, confirming the unstructured nature of SilE – the IDP.

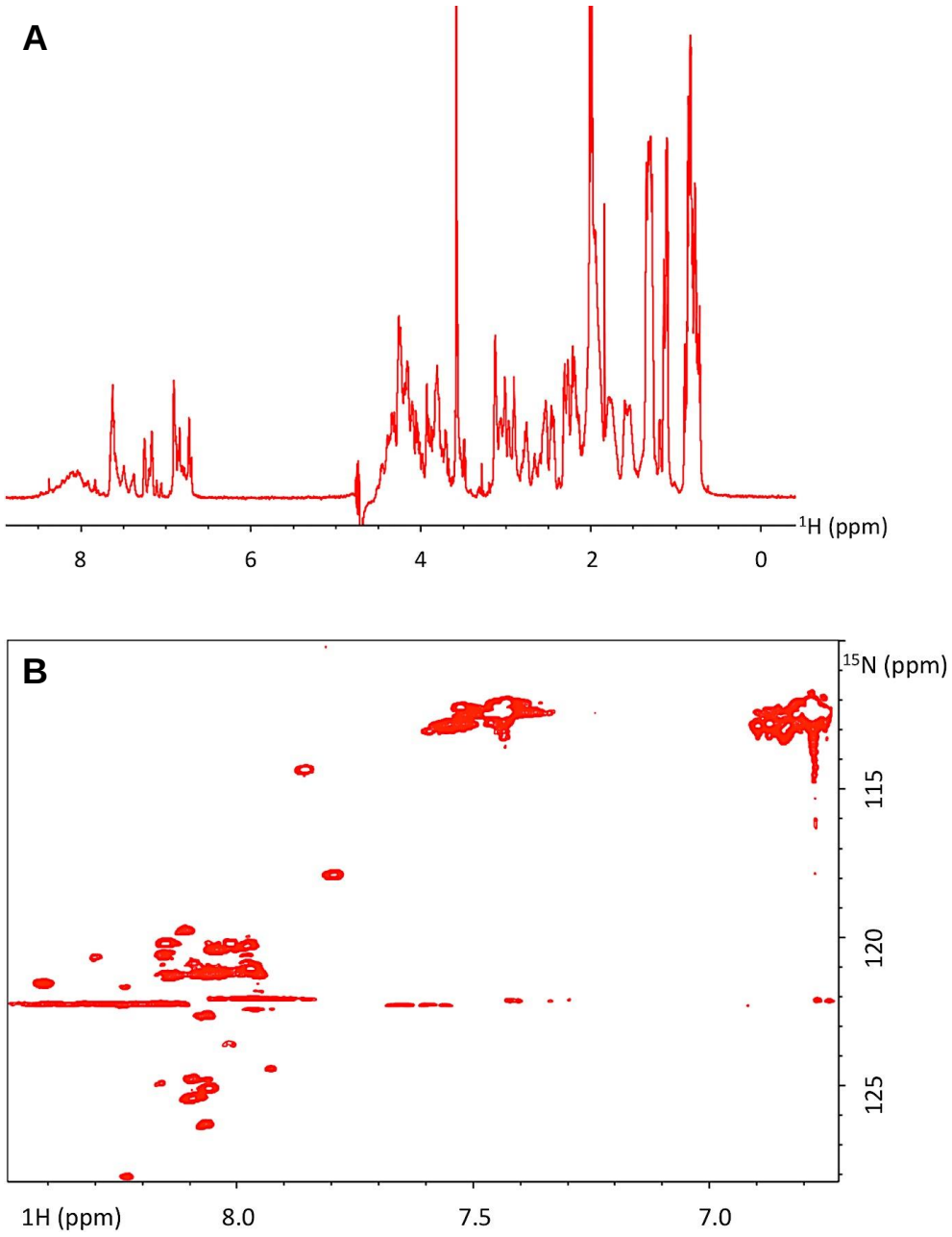


Figure 4. 10 SilE NMR spectra

NMR spectra of ^{15}N-SilE recorded in 10 mM Na_2HPO_4/NaH_2PO_4, 25 mM NaCl, pH 7.5, 10 % D_2O. **A.** ^{1}H experiment of 300 µM of ^{15}N-labelled protein. A narrow dispersion of chemical shifts suggests that the protein is unfolded. It also seems that there are no contaminants or artefacts present. **B.** Little peak distribution in the ^{1}H-^{15}N HSQC confirms the unfolded nature of SilE.

SAXS analysis of SilE

As previously noted, the inherent structural flexibility of IDPs limits the amount of structural information that can be obtained from solution NMR. Diffraction methods are among the most powerful techniques to study macromolecular structure but IDPs are difficult (if not impossible) to crystallise for high-resolution X-ray crystallography (Bernardó and Svergun, 2010). At the same time, small-angle X-ray scattering (SAXS) is one of the very few diffraction techniques that can yield valuable structural information about flexible macromolecules and IUPs. SAXS is one of the most efficient methods for structural characterisation of IDPs and allows the study of low-resolution structure of native proteins in solution (Chapter 2, Section 2.2.6.4). Due to its versatility, speed and ability to deal with disordered matter and mixtures, SAXS is now one of the major structural determination methods in the IDP field. This is because, SAXS enables in solution: (i) calculation of the molecular dimensions of a protein i.e. the *Rg* (radius of gyration), *Dmax* (largest intra-molecular distance) and *Vc* (volume-of-correlation) (ii) measurement of molecular weight (MW) and oligomeric state (iii) assessment of the flexibility of a polypeptide via qualitative analysis and (iv) deduction of a low-resolution shape of a protein using *ab initio* modelling (Receveur-Brechot and Durand, 2012; Bernadó and Svergun, 2012; Kikhney and Svergun, 2015).

Essentially, SAXS is a contrast method so the experiment depends on the square of the difference in the electron densities between the solvent ($\rho_{(s)}$) and the sample and solvent ($\rho_{(p)}$). To remove the solvent scattering, measurements are taken from the protein plus buffer and the buffer signal

alone (which also includes signals from any other background i.e. the capillary flow-cell). These curves are then subtracted at *q* to generate a curve representing the electron density of the sample ($\Delta\rho_{(p)}$), (Eq. 4.2, Figure 4.11) (Putnam *et al.*, 2007).

$$\Delta\rho_{(p)} = \rho_{(p)} - \rho_{(s)} \qquad \textit{(Eq. 4.2)}$$

Since protein particles are flexible and tumbling in solution, the SAXS profile measured for an IDP is the averaged scattering intensity of all of the conformations adopted by the protein, in different orientations, at every scattering angle *q* (*s* in earlier literature); thus presenting an exaggerated isotropic pattern. A radial integration of that pattern results in a 1D profile that is reciprocal space, in which larger particles scatter at smaller angles (low *q* range) and smaller particles scatter at larger angles. Accordingly, analysis of this 1D curve at the low *q* range yields information about the gross size and conformation of the particles in solution, whilst higher *q* data provides higher resolution information about internal structure. The curve can either be qualitatively analysed in sections or as a whole in the *q* range. Amongst these analysable parts are two significant regions – the Guinier region (analysis of the curve at the small scattering angles) and the Porod region, defined by the Porod-Debye Law (the linear, higher *q* range). Therefore, to provide information on the native proteins conformation and mass distribution around its centre of gravity, SilE SAXS data was collected (in 10 mM HEPES, 20 mM NaF, pH 7.5).

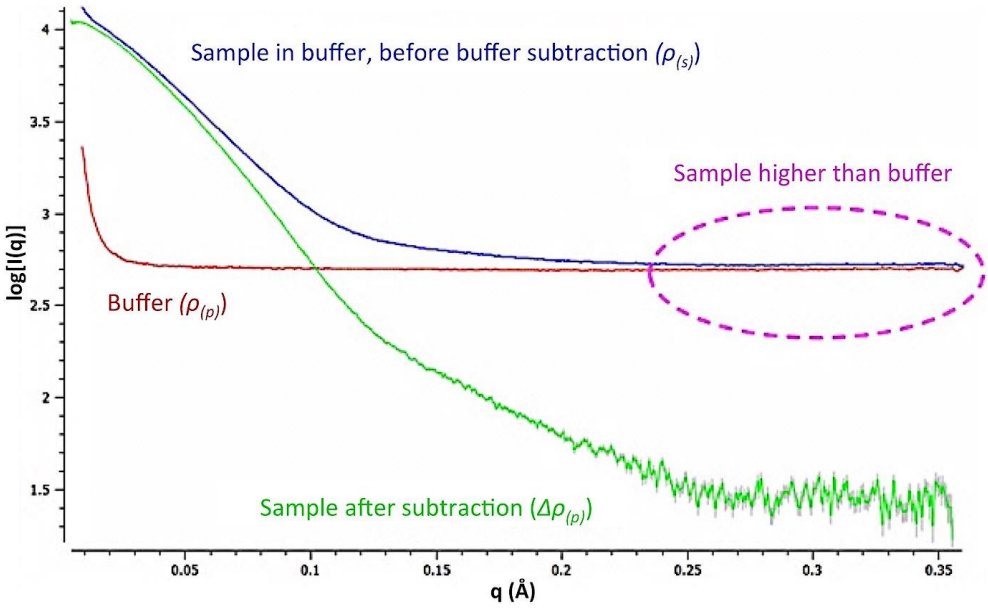


Figure 4. 11 SAXS buffer subtraction

Confirming correct buffer subtraction is essential prior to starting data analysis. A representation showing how the program PRIMUS (Konarev *et al.*, 2003) is used to ensure that the scattering intensity of sample ($\rho_{(p)}$ - blue) is higher than that of the buffer/solvent ($\rho_{(s)}$ - red) at the high *q* region. The green scattering profile is the result of the sample following buffer subtraction ($\Delta\rho_{(p)}$). Image used following publishers permission from SSRL, 2014.

Guinier analysis

Measurements of scattering pattern are typically collected from three or more concentrations of a sample, with multiple scans of each sample concentration being averaged (18 scans per sample at Beamline B21 at Diamond Light Source). However, prior to averaging per concentration, each scattering curve is overlaid and anomalous curves are excluded from further calculations. This was carried out manually on SilE using the program PRIMUS (Konarev *et al.*, 2003). Curve deviations or anomalies are usually the result of radiation damage, aggregation, interparticle interference or liquid handling robot errors (i.e. air bubbles). Where data resulting from

radiation damage and robot liquid handling errors are obvious and easily eliminated, the subtlety of curves arising from aggregated samples and interparticle interactions makes their identification a little more demanding.

The Guinier region is highly sensitive to large objects, especially aggregates. So the Guinier analysis, developed by and named after Andre Gunier in 1939, can provide useful details on sample behaviour and is useful for assessing sample aggregation, interparticle interactions and concentration-dependent scattering (Receveur-Brechot and Durand, 2012). The Guinier method uses an equation (Eq. 4.3) that extrapolates the linear low q data closest to $q_{0,}$ hence approximates scattering intensity at $I_{(0)}$ and enables determination of Rg – the mass distribution of the macromolecule around its centre of gravity.

Guinier Law $$lnI_{(q)} = lnI(0) - Rg^2/3 \,.\, q^2$$ *(Eq. 4.3)*

When assessing the Guinier plot, a non-linear dependence makes identifying an aggregated sample (upward curve) or repulsion (downward curve) obvious. Additionally, the relationship between Rg or $I(0)$ with increasing sample concentration determines the presence of any aggregation or interparticle interference, whereby a positive correlation indicates sample aggregation and a negative correlation suggests interparticle interference (Putnam *et al.*, 2007).

Figure 4.12 shows the manual and automatic Guinier analysis that were performed using the application ScÅtter (Biosis.net, 2016). In literature, for most globular molecules, the linear Guinier region should be

restricted to q_{max} <1.3/*Rg*. However, for IDPs this region is reduced to q_{max} <0.8/*Rg*. This is to ensure that the estimated parameters, such as *Rg* are within 10 % of their true value. (Putnam *et al.*, 2007; Receveur-Brechot and Durand, 2012; Bernadó and Svergun, 2012). As mentioned earlier, measurements were taken from SilE at three different concentrations at 5, 2.5 and 1.25 mg mL^{-1}, with respective *Rg* values of 3.08, 3.08 and 3.10 nm. This validated the structure factor and confirmed no interparticle interactions or significant effects of concentration on the *Rg* – 3.1 nm (rounded up to one decimal place (d.p.)). The curves were then merged via the 'merge' process using PRIMUS, to allow for better signal:noise data in the high *q* from higher concentration samples and less influence from the interparticle interactions from the low *q* range. Further assessment on the shape and size of SilE could then be continued.

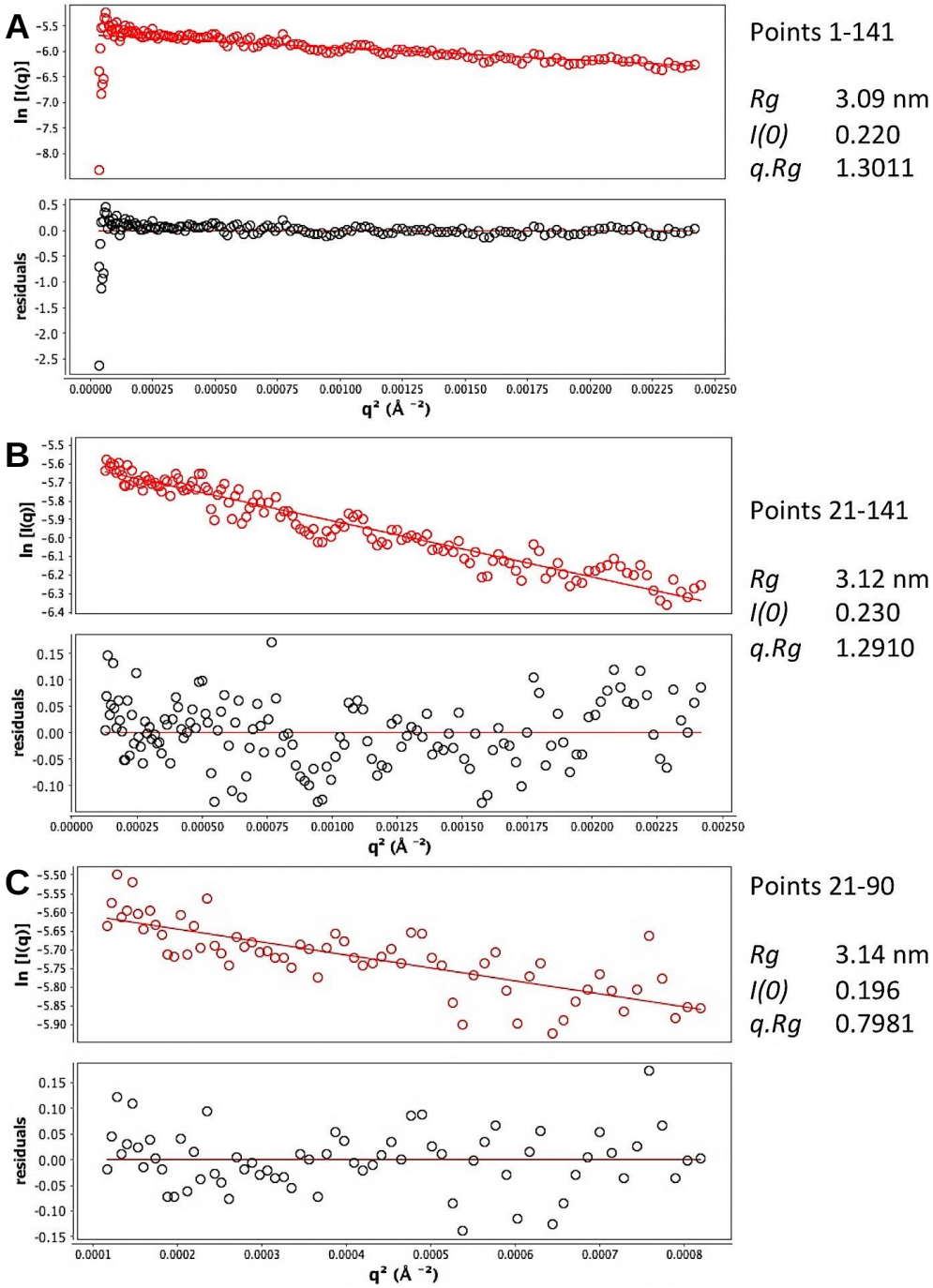


Figure 4. 12 Guinier plots of SilE

SAXS Guinier extrapolations of SilE where the Guinier region is determined by the range of points in each plot. **A.** Points 1-141 show the pre-trimmed automatically determined Guinier region. Whereas points 21-141 in **B.** and 21-90 in **C.** represent the Guinier fit by manual cropping with q_{max} <1.3/*Rg* and with q_{max} <0.8/*Rg*, respectively. Below each Guinier plot are the corresponding residuals. Any deviations of the residuals from zero indicate a non-linear Guinier fit and potential problems with the data.

Figure 4.12 represents the Guinier analysis following the combination of the data from the three concentrations. Automatic Guinier analysis on ScÅtter (Biosis.net, 2016) can provide only an initial estimation, as mild aggregation is not detected and trimmed automatically (Figure 4.12A). Therefore, manual Guinier analysis was carried out. Trimming the first 20 points and leaving points 21 to 141 allowed a slightly more accurate *Rg* of 3.12 nm to be extrapolated with *q.Rg* reducing from 1.3011 to 1.2910 (Figure 4.12B). However the *Rg* increased to 3.14 nm when the Guinier region was defined to the recommended q_{max} <0.8/*Rg* by trimming down points 21-90 (Figure 4.12C). So, in this manner the slope of the Guinier Plot inferred the size parameter *Rg* of SilE as 3.1 nm (rounded up to one d.p.). As a likely result of adopting highly extended conformations, IDPs are characterised by such large average sizes, in comparison to the number of amino acids forming this polypeptide.

Pair-wise distance distribution function

An alternative method to determine the *Rg* is using the distance distribution function - *P(r)*. By taking an indirect Fourier transform of the 1D scattering intensity, i.e. reciprocal space (*I(q)*), the real-space representation called *P(r)* can be resolved. The P(r) function is a histogram of all of the atomic distances within the molecule, which gives qualitative information about the shape and anisotropy of a protein. A Gaussian, symmetric, bell-shaped *P(r)* curve usually represents a globular, spherical protein. Whereas, the *P(r)* function of an unstructured and elongated protein tends to be asymmetric and left-skewed, as could be seen for SilE in Figure

4.13, generated using GNOM in PRIMUS (Svergun, 1992; Konarev *et al.*, 2003). In Figure 4.13 the *Pr* distribution of SilE also displayed a right-hand shoulder. This coincides with proteins that tend to have multiple domains or a dumbbell shape, in which the shoulder represents the smaller domain. The area under the scatter embodies the most distant pairs. This means that it was possible that the unfolded and unordered SilE contained more than one domain.

The *Rg* can also be calculated from the *P(r)* function. An advantage of determining your *Rg* from the *P(r)* curve is the use of the entire intensity profile, thus utilising data less sensitive to interparticle effects, rather than the Guinier region which consists of low *q* data. The more precise *Rg* obtained from the *P(r)* was 3.2 nm (to one d.p.), which correlated nicely as it was within 3 % of the *Rg* estimated by Guinier analysis.

Additionally, the *P(r)* can help to derive the *Dmax*, the maximal value of *r* for which *P(r)* does not equal zero. The *Dmax* corresponds to the maximum diameter of the protein. For SilE the GNOM computed *Dmax* was 114 Å (11.4 nm) (Svergun, 1992). This was generally quite a large value for a globular, folded protein consisting of 125 amino acids and thereby inferred an extended, elongated conformation of SilE (Bernadó and Svergun, 2010).

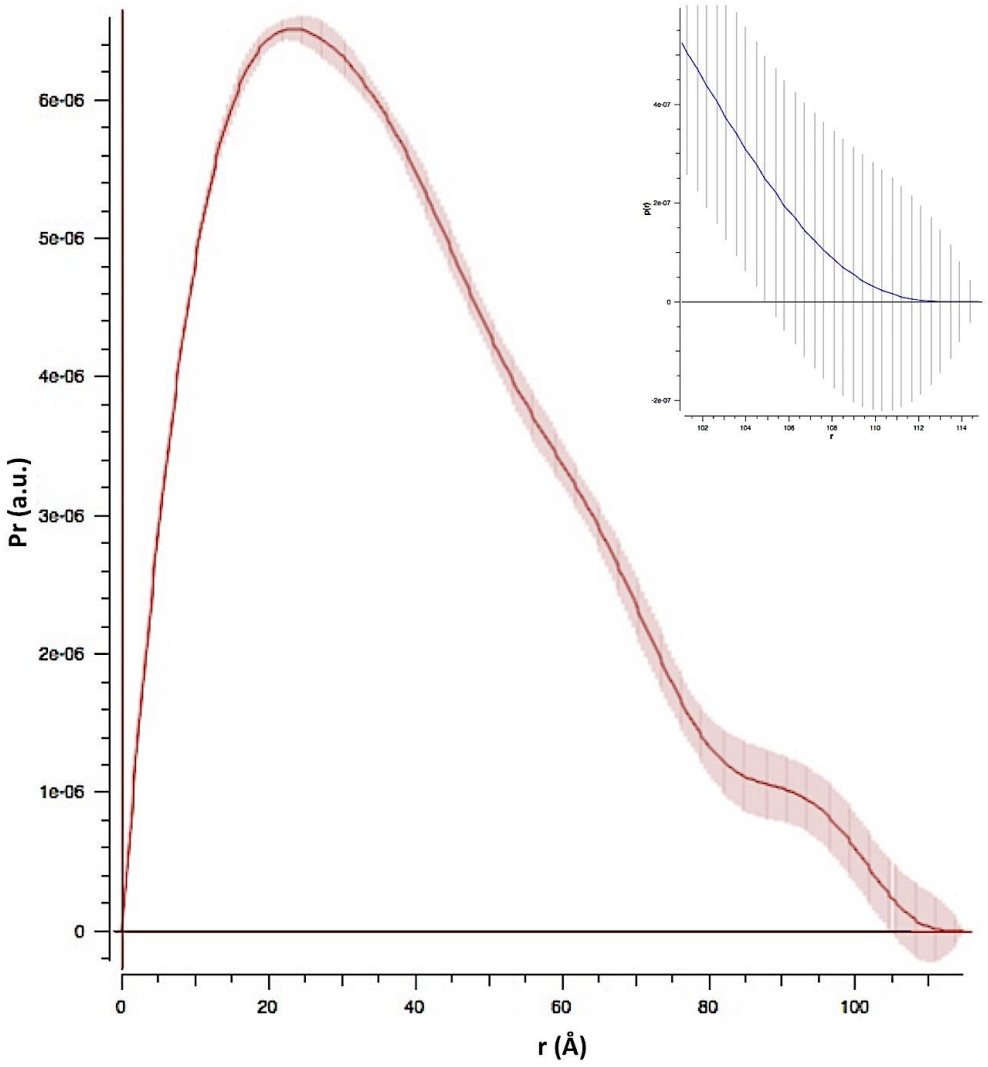


Figure 4. 13 ***Pr*** **analysis of SilE**

Pairwise distance distribution analysis (red) determines the shape distribution and *Dmax* of the protein. Note the curves gradual fall to zero (zoomed in blue curve inlayed, top right), showing a *Dmax* of 114 Å (=11.4 nm).

Kratky analysis

Folded and unfolded proteins in solution have different scattering profiles thus traditionally Kratky plots ($I(q).q^2$ versus q) are useful to identify disordered states and distinguish them from globular particles. So once data quality is confirmed, qualitative assessment using the Kratky plot (Figure 4.14) allows observation of compactness and flexibility of the protein (Kratky and Porod, 1949). Figure 4.14 shows a schematic representation of Kratky plots depending on their molecular shape and degree of flexibility. A globular molecule follows Porod's law, stating that in the high q (Porod region) $I(q)$ decreases with q^{-4}. This means well-folded proteins present a bell-shaped curve with an initial parabolic peak. More flexible and extended molecules, such as random coils (Gaussian chains) and IDPs, lack this peak and have a plateau or are divergent where I increases with q and present a positive slope. This is indicated by the Debye law (Receveur-Bréhot *et al.*, 2006; Putnam *et al.*, 2007; Rambo and Tainer, 2011). The data from the SilE Kratky plot in Figure 4.15 displayed a hyperbolic shape characteristic of an unfolded or extended protein - indicative of some degree of random coil component.

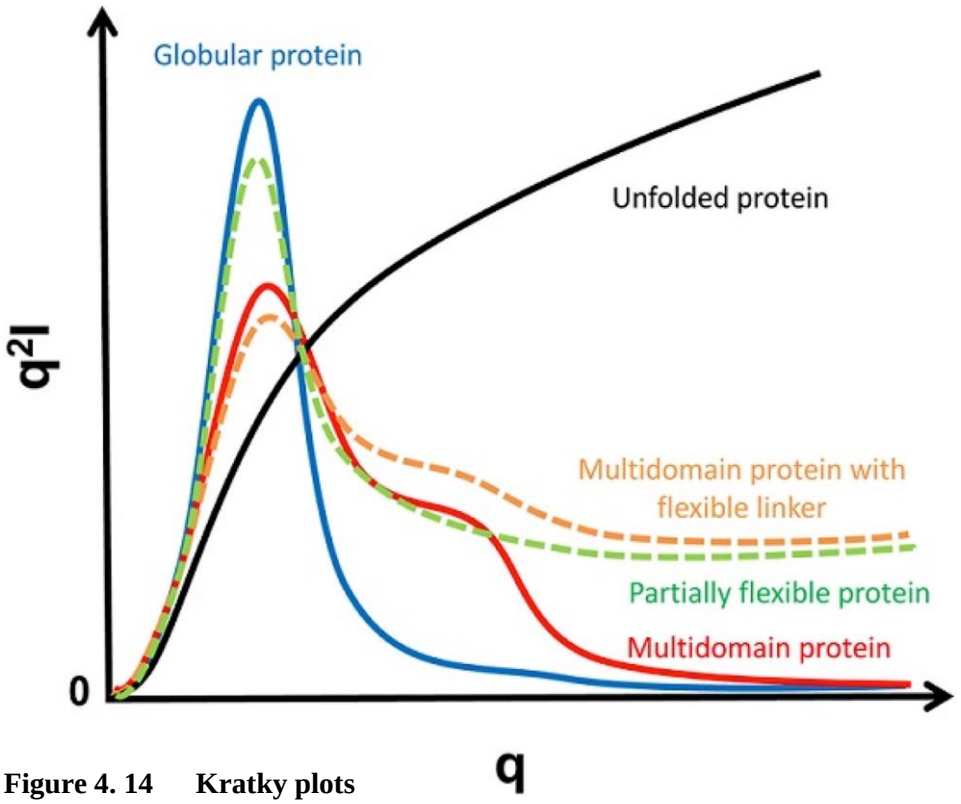


Figure 4. 14 Kratky plots

Schematic representations of the Kratky plots of: folded, globular protein (blue), partially unfolded proteins (orange and green), multidomain protein (red) and unfolded, disordered protein (black). Image used following publishers permission from SSRL, 2014).

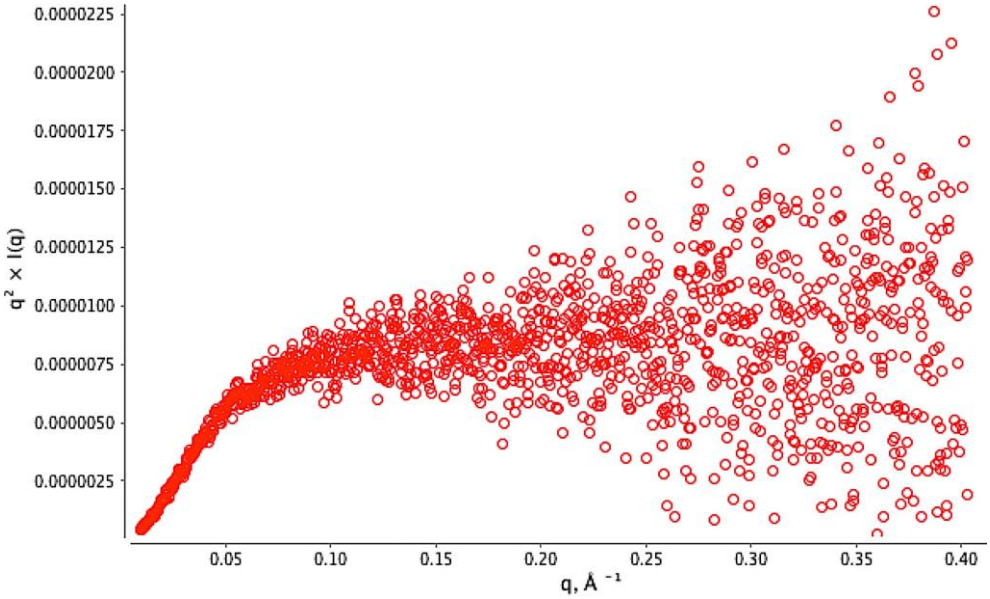


Figure 4. 15 Kratky plot of SilE

A Kratky plot of SilE, generated using ScÅtter, from data collected following SAXS experimentation. The curve definitely follows the Debye law of an extended, unfolded protein, as it plateaus in the *q* region.

Ab initio protein 3D modelling

Following the various methods of necessary data validation, SAXS allows for *ab initio* structure determination and rigid body modelling. The structural parameters observed aid in the *ab initio* determination of a molecular envelope, in which many low-resolution particle models are produced. A representative single configuration is then selected as the one that fits the experimental data (Kikhney and Svergun, 2015).

The expansion of SAXS structural biology sparked a decade ago, when new programs that resolve the envelope of a scattering object were developed. A number of programs use different algorithms and apply different restraints to calculate an overall external shape by filling a bead model of the protein (Receveur-Brechot and Durand, 2012). The most common of these, found within the All That Scattering (ATSAS) software package, include DAMMIN, DAMMIF and GASBOR (Petoukhov *et al.*, 2012). All three perform *ab initio* shape determination by simulated annealing, using a single phase dummy atom model (Svergun, 1999; Svergun *et al.*, 2001). DAMMIN and DAMMIF are known to fit the data up to a resolution of 25 Å. Yet compactness and connectivity restraints, aiming to minimise the interfacial area between the protein envelope and solvent, are explained as being inappropriate for IDPs (Receveur-Brechot and Durand, 2012). Therefore GASBOR, described as the most advanced program of its type, was used to generate a bead model of SilE. Where DAMMIN/DAMMIF represent protein structure using dummy atoms (densely packed spheres sufficiently filling the volume) GASBOR is a program used for *ab initio* reconstruction with an ensemble of dummy

residues to build a 'chain-like' model. GASBOR uses the entire SAXS curve up to a resolution of 10 Å (approx.) to generate a bead model in which each bead corresponds to a dummy residue, equal to the number of amino acids within the protein (Svergun *et al.*, 2001).

Figure 4.16 presents the *ab initio* model constructed from SilE SAXS data, using GASBOR (from ATSAS) (Svergun *et al.*, 2001; Petoukhov *et al.*, 2012). The model, consisting of the 123 protein residues, illustrated the unfolded nature of SilE. The elongated shape provided a visual insight for this disordered protein and confirmed the numerical parameters of *Rg* and *Dmax* inferred from the scattering curve.

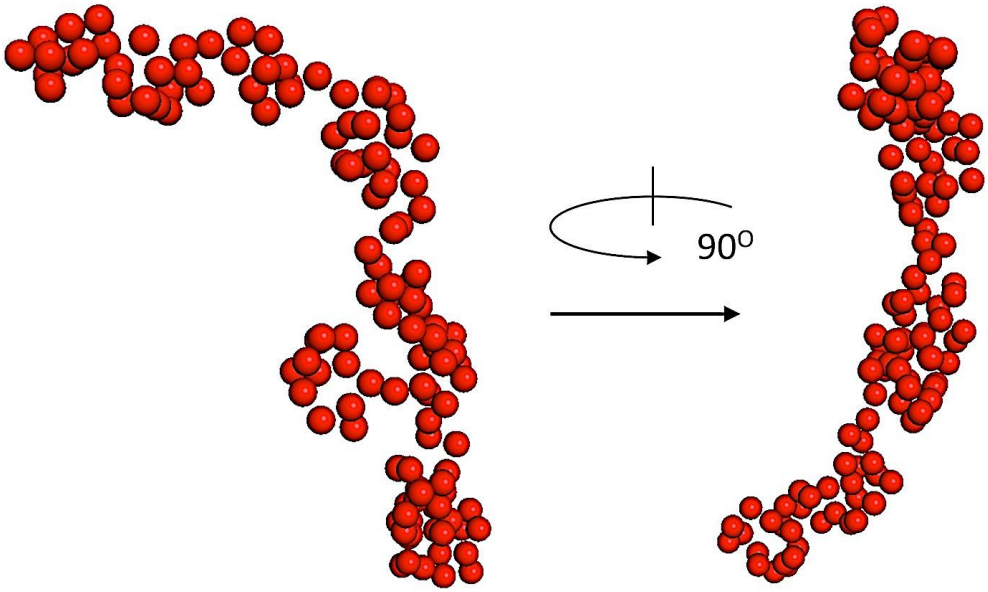


Figure 4. 16 Bead model of SilE

Output from GASBOR representing an *ab initio* model of SilE from SAXS data obtained. The model has been generated with no prior knowledge of SilE other than the polypeptide chain length. The long, extended and disordered chain of SilE is apparent in this graphic conception.

4.3 Summary and conclusions

The aim of this chapter was to successfully overexpress, purify and characterise SilE. The expression system used enabled large quantities of soluble protein to be produced – a requirement for biophysical studies. The his-tagged SilE was then successfully purified using nickel affinity and size exclusion chromatography (Figures 4.1-4.6). The purification method produced average yields of 30 mg L^{-1} of soluble protein.

Characterisation commencing with IDP predictions forecasted SilE to be an IDP, where the biophysical far-UV CD, NMR and SAXS experiments confirmed this. CD experiments showed that the protein produced secondary structure of random coil, as expected (Figure 4.9). The fraction of 74 % of SilE being unstructured coincided with the PrDOS value of predicted protein disorder (Figures 4.9 and 4.8).

Further laboratory-based analysis in the form of proton and HSQC NMR investigations suggested that the protein was not folded, as judged by poor dispersion of proton peaks (Figure 4.10). Additionally, SAXS analysis also indicated that the protein had an extended conformation (Figures 4.13, 4.15 and 4.16). Nevertheless the presence of the slight negative shoulder in the CD spectrum at 221 nm, indicative of little α-helical content (Figure 4.9), and the shoulder in the *Pr* curve (Figure 4.13) both proposed that SilE may be multi-domain protein (Bernadó *et al.*, 2007).

CHAPTER 5

Studies on Folding of the Silver Binding SilE Protein

5.1 Introduction

Chapter 4, Section 4.2 provided experimental evidence that SilE in its native form is an intrinsically disordered protein (IDP). Yet many IDPs have functional roles, which promote protein folding when interacting with d-block metal ions (Faller *et al.*, 2014; Uversky, 2015). Likewise, Silver *et al.*, (1999) stated (without presenting experimental evidence) that upon Ag^+-binding the SilE protein folded into 'a well-defined structure'. The SilE copper-binding homologue PcoE, of the *pco* copper-resistance system, was also described as an unstructured protein in its *apo*-form that folded upon binding multiple soft metal ions, Cu^+ and Ag^+ (Zimmermann *et al.*, 2012).

Therefore, this chapter discusses investigations on SilE metal binding; the proteins specificity for binding Ag^+ and the metal ligand binding induced folding properties of SilE. Following appropriate buffer selection, these were investigated using circular dichroism (CD) experiments and mass spectrometry (MS) analysis. Attempts at determining the *holo* folded structure of SilE by x-ray crystallography, small-angle x-ray scattering (SAXS) and nuclear magnetic resonance (NMR) of *holo*-SilE are also presented in this chapter.

Note: Throughout this chapter SilE refers to the protein without its periplasmic signal sequence (residues 1-20), which was produced as $SilE_{21-143F}$.

5.2 Results and discussion

5.2.1 Silver solubility and the buffering system

As soluble silver ions (Ag^+) were required in order to observe SilE Ag^+-binding and folding experiments, we first determined the best buffer system to work with. This was crucial to the study as metallic silver, and indeed many of its salts are insoluble in water - forming coloured precipitants such as the white halide silver chloride or the yellow silver phosphate (Table 5.1). Silver nitrate ($AgNO_3$) is the salt known as having the highest level of solubility, dissolving in water at up to 2,560 g L^{-1} (Seidell and Linke, 1919). Theoretically, this is because in $AgNO_3$ the NO_3^-, is significantly larger than the Ag^+, making the bond weaker, easier to break and hence easier to form interactions with water molecules. For this reason, $AgNO_3$ was used as the silver salt of choice for the *holo*-SilE studies.

An appropriate biological buffer needs to fulfil the following criteria:

(i) the pH should mimic physiological conditions,

(ii) the buffering system should have non-toxic buffering capabilities at the pH of choice (usually with a pK_a value within one pH unit of the desired pH),

(iii) contain salt to aid protein solubility and mimic biological conditions and

(iv) contain reducing agents (only necessary if the protein contains disulphide bridge forming cysteine residues) (Good *et al.*, 1966).

Table 5. 1: Silver salts

Information on the common silver salts, colour and solubility in water, at 25 ^{o}C (Patterson, 2016).

Compound	Empirical Formula	Molecular Mass (Da)	Colour	Solubility ($g\ L^{-1}$)
Silver acetate	$AgC_2H_3O_2$	166.92	White/ grey	11.11
Silver bromate	$AgBrO_3$	235.78	White	1.67
Silver bromide	AgBr	187.78	Pale yellow	0.000135
Silver carbonate	Ag_2CO_3	275.75	Pale yellow	0.033
Silver chloride	AgCl	143.32	White (grey in sunlight)	0.00193
Silver chromate	Ag_2CrO_4	331.73	Red/ brown	0.035
Silver (I) fluoride	AgF	126.87	Yellow/ brown	1000+
Silver (II) fluoride	AgF_2	145.87	White/ grey	-
Silver iodate	$AgIO_3$	282.77	White	0.44
Silver iodide	AgI	234.77	Yellow	0.0000026
Silver nitrate	$AgNO_3$	169.87	White	2570.00
Silver nitrite	$AgNO_2$	153.88	Yellow	4.20
Silver (I) oxide	Ag_2O	231.74	Grey/ black	0.0002
Silver phosphate	Ag_3PO_4	418.58	Yellow	0.00644
Silver sulfate	Ag_2SO_4	311.80	White	8.30
Silver sulfide	Ag_2S	247.80	Grey/ black	0.00014

Such criteria needed to be met in order to stabilise both the Ag^+ ion as well as the protein under experimental conditions.

The periplasmic pH of *Escherichia coli* (*E. coli*) where SilE is known to reside, has been measured to range between 7.4 and 7.6 (Stock *et al.*, 1977; Wilks and Slonczewski, 2007). Consequently, common buffers considered with a buffering capacity within these limits were – HEPES (4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid - $C_8H_{18}N_2O_4S$), MOPS (3-(*N*-Morpholino)propanesulfonic acid - $C_7H_{15}NO_4S$), sodium phosphate (Na_2HPO_4 – NaH_2PO_4), and Trizma®/Tris base (Tris(hydroxymethyl)aminomethane - $C_4H_{11}NO_3$). Addition of silver to sodium phosphate results in the formation of yellow, insoluble precipitate (Figure 5.1) - silver phosphate, via a displacement reaction:

$$3AgNO_{3(aq)} + Na_3PO_{4(aq)} \rightarrow Ag_3PO_{4(s)} + 3NaNO_{3(aq)} \quad \textit{(Eq. 5.1)}$$

Similarly, even though Trizma® cannot react directly with Ag^+ ions, the chloride ions (Cl^-) within the plentiful hydrochloric acid (HCl) required to adjust the pH of the buffer form the white crystalline solid silver chloride (AgCl), which turns grey upon exposure to sunlight (Figure 5.1):

$$AgNO_{3(aq)} + HCl_{(aq)} \rightarrow AgCl_{(s)} + HNO_{3(aq)} \quad \textit{(Eq.5.2)}$$

Because of these problems Trizma® and sodium phosphate were disregarded as buffers for any studies on SilE in the presence of Ag^+ ions. Furthermore, MOPS is known to be slightly toxic at higher concentrations and should be avoided when conducting CD experiments as it absorbs strongly at or below 200 nm. HEPES was therefore chosen as the buffering system as it stabilised the Ag^+. Due to HEPES absorbing at lower wavelengths (Kelly *et al.*, 2005), its use was restricted to a low concentration of 10 mM when continuing on to the Ag^+-SilE experiments.

Ionic strength is another important element to take into account when seeking to replicate a biological environment for proteins. Commonly used salts such as sodium chloride (NaCl) have a strong UV absorbance at low wavelengths, owing to the chloride ions (Cl^-) thereby affecting experiments involving the use of light at those wavelengths (Fox *et al.*, 1978). As explained previously, Cl^- also reduces Ag^+ to form white, insoluble AgCl. Consequently, sodium fluoride (NaF) was used as the alternative salt in the buffer, as the fluoride counterion does not absorb low wavelength light and the AgF salt has a high solubility (Table 5.1).

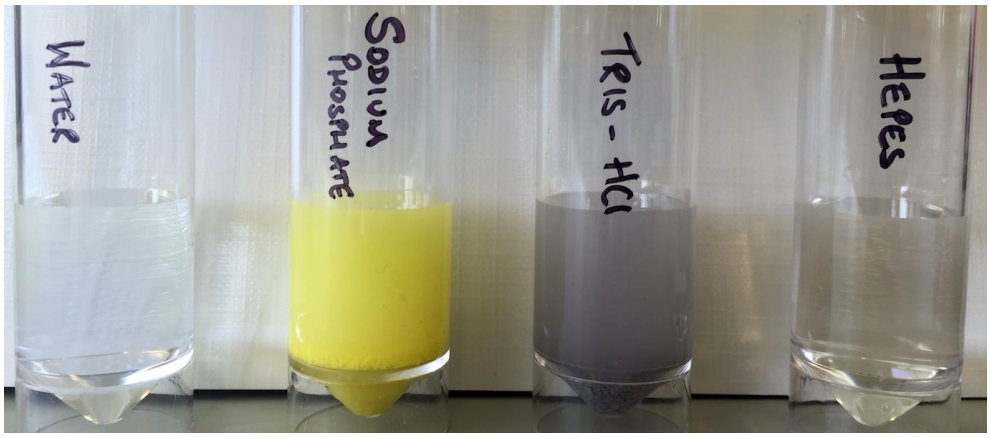


Figure 5. 1 Solubility of silver salts

100 mg of $AgNO_3$ was added to 10 mL of each liquid - water, 10 mM Na_2HPO_4 – NaH_2PO_4, 10 mM Tris-HCl and 10 mM HEPES. $AgNO_3$ is insoluble in the Na_2HPO_4 – NaH_2PO_4 and Tris-HCl buffers, forming yellow Ag_3PO_4 and grey AgCl precipitate, respectively. Yet, just like in water, it is completely soluble in HEPES (hence HEPES was used for Ag^+-binding investigations).

5.2.1.1 Selection of other metal ions

It is known from the literature that PcoE, the copper-binding (Cu^+ and Cu^{2+}) homologue of SilE, can bind other metal ions including Ag^+, Au^+ and Zn^{2+} (Zimmermann *et al.*, 2012). Therefore, it was important to determine if SilE has any specificity towards Ag^+-binding, or is just a general metal ion binding protein. To decide which other d-block metal ions were going to be used for experimentation, the following points were considered.

Firstly, copper (Cu) was the obvious choice as PcoE readily binds Ag^+ (between two and seven ions per protein molecule (Zimmermann *et al.*, 2012)) and because copper is one of the most commonly used toxic metals in medicine and agriculture (Hobman and Crossman, 2015). Additionally, Cu is found above Ag in group 11 of the transition metals on the periodic

table of elements (Appendix B); suggesting that it has similar chemical and electronic configuration (one electron in the outer shell). We used only Cu^{2+} and not Cu^{+} as SilE is located in the bacterial periplasm, an environment known naturally to be oxidising (Depuydt *et al.*, 2009; Waldron and Robinson, 2009; Lippa and Goulian, 2012). Hence, cupric sulphate (copper (II) sulphate, $CuSO_4$) was utilised as the soluble source of Cu^{2+}.

As well as Cu^{2+}, other metals were considered to test for SilE metal-binding specificity. Further analysis of the periodic table of elements led to the consideration of cadmium (Cd) and palladium (Pd), the two transition metals adjacent to Ag (Appendix B). However, both Cd and Pd are known to be extremely toxic metals: Cd has been stated to pose severe risks to human health and can lead to kidney, bone and pulmonary damages. Additionally it is a cancer risk factor due to its carcinogenic properties (Godt *et al.*, 2006). All Pd compounds are also regarded as highly toxic, acting as potential mutagens and associated with bone marrow, kidney and liver damage (Wataha and Hanks, 1996). It is for this reason that zinc (Zn) and nickel (Ni), neighbouring Cu on the periodic table (Appendix B), were used for metal binding experiments. Zn^{2+} was provided in its ionic state through zinc chloride ($ZnCl_2$) and Ni^{2+} from nickel chloride ($NiCl_2$).

5.2.2 SilE binds up to eight silver ions

Electrospray ionisation-mass spectrometry (ESI-MS) is a soft ionisation based mass spectrometry technique, which has been used extensively to observe macromolecules and macromolecular complexes in their native state (Chapter 2 Section 2.2.6.2). A fine aerosol known as an electrospray is generated by spraying the sample through a metal-coated glass needle that is kept at high voltage (c.a. 5 kV). The droplets formed are far smaller than that attained at 0 V, and are then able to be transferred into the mass spectrometer. By using several stages of elctrofocusing through low pressure gas environements it is possible to transfer the molecules intact to the gas phase without fragmenting and denaturation. To aid this, samples are generally dissolved in a volatile solvent, (i.e. aqueous ammonium acetate) (Lemaire *et al.*, 2001). The resultant ions are then separated by a mass analyser according to values of mass to charge ratio (m/z) and detected by an ultra-fast electron multiplier detector. A series of corresponding different m/z ratios are observed. For a folded protein, an apparently Gaussian envelope of charge states is observed (this was particularly clear for Ag^{+}-SilE - Figure 5.2).

ESI-MS is a very sensitive technique for the detection of non-covalent interactions between a metal ion and protein under native conditions (Gehrig *et al.*, 2000). So, the ability of SilE to interact with and bind to metal ions was assessed using nano-ESI-MS, under non-denaturing conditions in volatile 25 mM ammonium acetate buffer at pH 7.0. Solutions of 2 mM of each metal ion: Ag^{+}, Cu^{2+}, Zn^{2+} and Ni^{2+} were added to 25 μM of *apo*-SilE (molar ratio of 80:1) and de-salted using a spin concentrator.

The excess $AgNO_3$ had to be removed to prevent false positives from adducts. Electrospray mass spectra show multiple 'charge states' for a single protein. SilE-metal complexes were observed at the 7+-charge state. In each case the relative bound proportions were dependent on the protein's affinity for that particular ion type as well as the stoichiometry for the metal ion (Figure 5.2).

In Figure 5.2 the *apo* spectrum of SilE (where no metals had been added) showed a major peak at m/z 1896.92, correlating to a mass of 13,271.31 ± 1.4 Da. Multiple ionic species yielding different masses were observed as a result of Ag^+ being added to SilE. A distribution of Ag^+-bound complexes with species mostly containing 5, 6 and 7 bound Ag^+ ions (based on mass calculations for m/z peaks 1987.86, 1988.66 and 2003.96) were particularly abundant, and a maximum number of 8 were detected. Binding of other harder divalent metal ions such as Cu^{2+} and Zn^{2+} was also evident, but with different binding stoichiometries (SilE:Cu^{2+} of 1:6 and SilE:Zn^{2+} of 1:5) A lower stoichiometry was seen with Ni^{2+} (of SilE:Ni^{2+} of 1:2) (Figure 5.2).

The ESI-MS showed that the most populated ligation state of SilE was six Ag^+ (Figure 5.2) but that it was capable of binding up to eight Ag^+: a higher ligation state than for other metals.

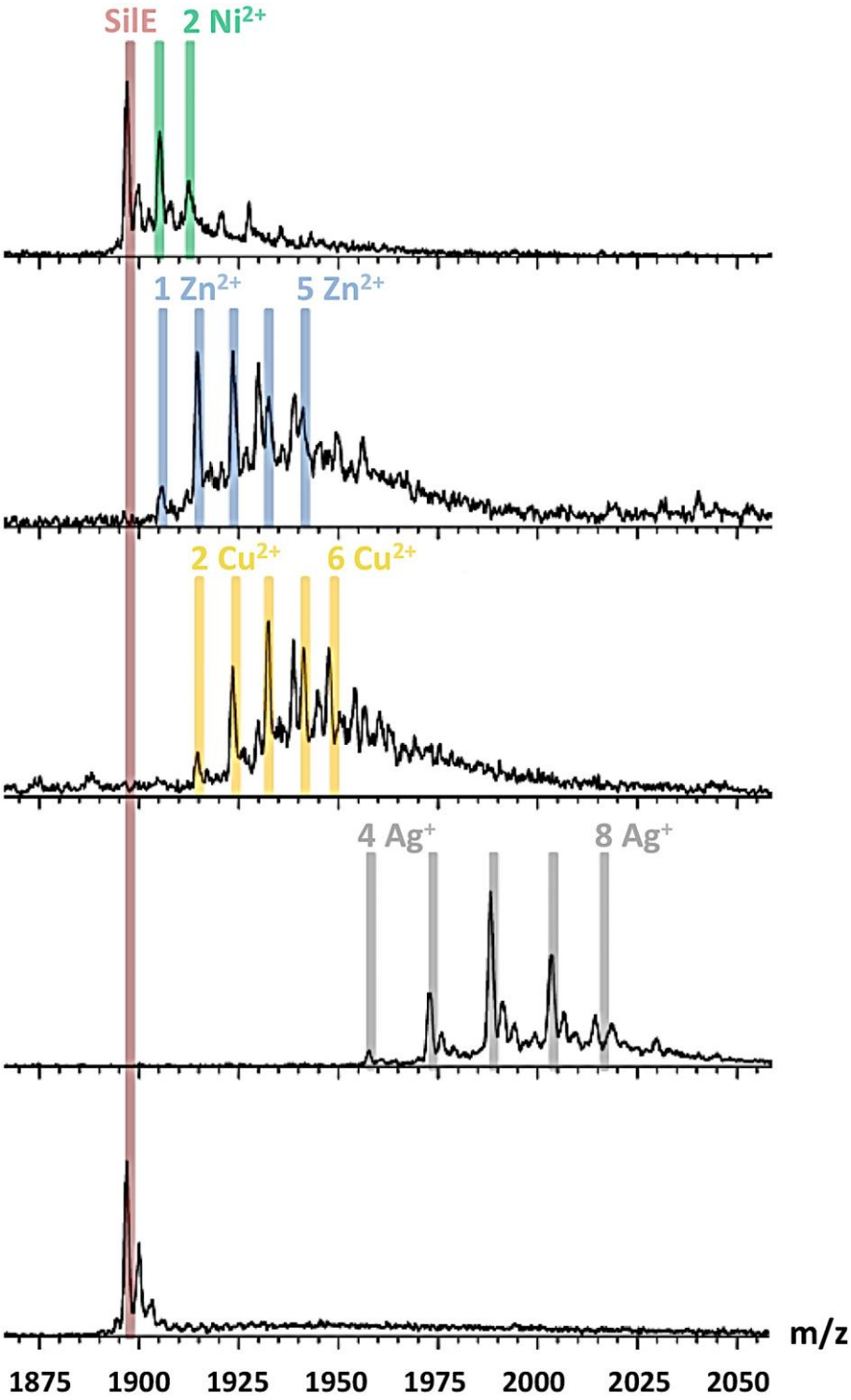


Figure 5. 2 Determination of SilE metal ion binding by ESI-MS

After incubation with 2 mM: Ni^{2+} (green), Zn^{2+} (blue), Cu^{2+} (yellow) and Ag^{+} (grey) native nano-ESI MS for SilE (25 μM in 25 mM ammonium acetate, pH 7.0) showing the 7+-charge state. The digits together with the metal elemental symbol provide the number of atoms of metal ions bound to SilE (as deduced by the incremental mass increases) in the labeled peak. *Apo*-SilE (red) has molecular mass of 13,271 Da.

5.2.3 SilE secondary structure predictions

Bioinformatics based techniques are used to predict the local secondary structure of proteins (defined in Chapter 4, Section 4.2.4.2), through knowledge of its primary amino acid sequence. Secondary structure prediction aims to locate and assign regions of the protein sequence as probable α-helices, β-sheets/strands or turns/random coils through their known propensities to form these structural elements.

The first secondary structure predictors were created over half a century ago, focusing particularly on predicting α-helices (Guzzo, 1965). A decade later saw the introduction of slightly more accurate predictors, which also included β-sheets. Amongst these were the famous methods designed by Chou and Fasman and Garnie, Osguthorbe and Robson (GOR) (Chou and Fasman, 1974; Garnier *et al.*, 1978). However, these earlier methods suffered from a paucity of structural data. Predictions were performed using only single sequences rather than utilising families of homologous sequences and comparison with just a few 3D structures were available to derive parameters from. This meant that predictions were only 50-65 % accurate (Kabsch and Sander, 1983). Nowadays, the availability of large families of homologous sequences combined with sophisticated computing techniques, such as machine learning neural networks, has led to much more accurate secondary structure identification. Modern methods can now predict secondary structure in proteins at around 80 % accuracy. JPred 4 (Incorporating Jnet) is a well-known, current prediction server that provides a consensus prediction on the secondary structure content of a protein (Drozdetskiy *et al.*, 2015).

As SilE had been experimentally proven to be an IDP (Chapter 4, Section 4.2), like homologue PcoE, it was presumed to fold upon binding to its metal ligand Ag^{+} (Zimmermann *et al.*, 2012). Figure 5.3 shows the screening of possible secondary structure within SilE performed using JPred 4 (Table 2.5, Chapter 2). Six α-helices (E^{22}-V^{30}, P^{36}-S^{42}, Q^{61}-M^{72}, A^{77}-M^{91}, E^{110}-F^{120} and G^{127}-M^{139}) were predicted across 54 % of the protein, with the remaining 46 % being predicted as random coil and no presence of β-sheet.

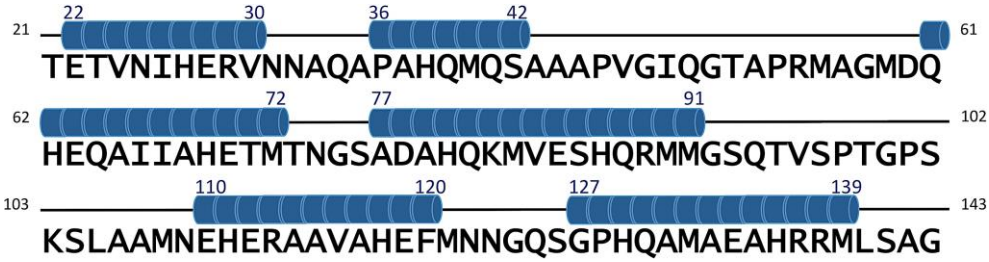


Figure 5. 3 SilE amino acids 21-143 secondary structure prediction

SilEs secondary structure fold, predicted on Jpred 4 can be seen above the sequence of SilE residues 21-143 with α-helical predominance (Drozdetskiy *et al.*, 2015). Six α-helices (residues: E^{22}-V^{30}, P^{36}-S^{42}, Q^{61}-M^{72}, A^{77}-M^{91}, E^{110}-F^{120} and G^{127}-M^{139}) are highlighted in blue, above the sequence, for the folded conformation of SilE. The predicted α-helices constitute 54 % of the secondary structure. The rest of the sequence has been predicted to consist of no secondary structure.

5.2.4 SilE-metal ion bound secondary structure and folding

In order to experimentally probe the SilE structural effects upon metal ion binding, particularly Ag^{+}, far-UV circular dichroism (CD) was conducted in 10 mM HEPES, 20 mM NaF at pH 7.5 in the presence of metal ions at 2 mM (Chapter 2, Section 2.2.6.1). The highly sensitive far-UV CD of a protein can display significant signal perturbations on a

spectrum, even upon small changes in secondary structure (Sreerama *et al.*, 1999; Martin and Schilstra, 2008). Although far-UV CD cannot provide any atomic detail, providing an average measure of change in the overall structure makes this technique very useful for studying protein thermodynamics and folding mechanisms. Furthermore, far-UV CD can monitor and provide valuable insights into secondary structure folding transitions associated with binding.

5.2.4.1 SilE-Ag^+ binding and folding analysis via far-UV CD

With a negative signal at 200 nm (Chapter 4, Section 4.2.4.2), the *apo*-SilE far-UV spectrum and Dichroweb server calculations (programme CONTINLL) (Provencher and Glockner, 1981; Van Stokkum *et al.*, 1990; Sreerama and Woody, 2000) highlighted and confirmed the ligand-unbound form of SilE as being 74 % disordered (Figures 4.9 (Chapter 4) and 5.4). Whereas the far-UV CD spectrum for *holo*-SilE (shown in Figure 5.4) showed strong negative ellipticity with minima near 207 nm and 221 nm, as well as a very convincing positive band at 190 nm, upon the addition of metal ligand Ag^+ (2 mM). This was indicative of the stabilisation of a predominantly α-helical secondary structure (Greenfield, 2007a; Dodero *et al.*, 2011). Using the measured CD data, Dichroweb (CONTINLL) analysis predicted that the Ag^+-bound protein SilE compared to the *apo*-form contained an increase in α-helix (from 20 % to 54 %) and 3 % β-sheet with a significant decrease in the absence of secondary structure from 74 % to 43 % (Figure 5.5-B).

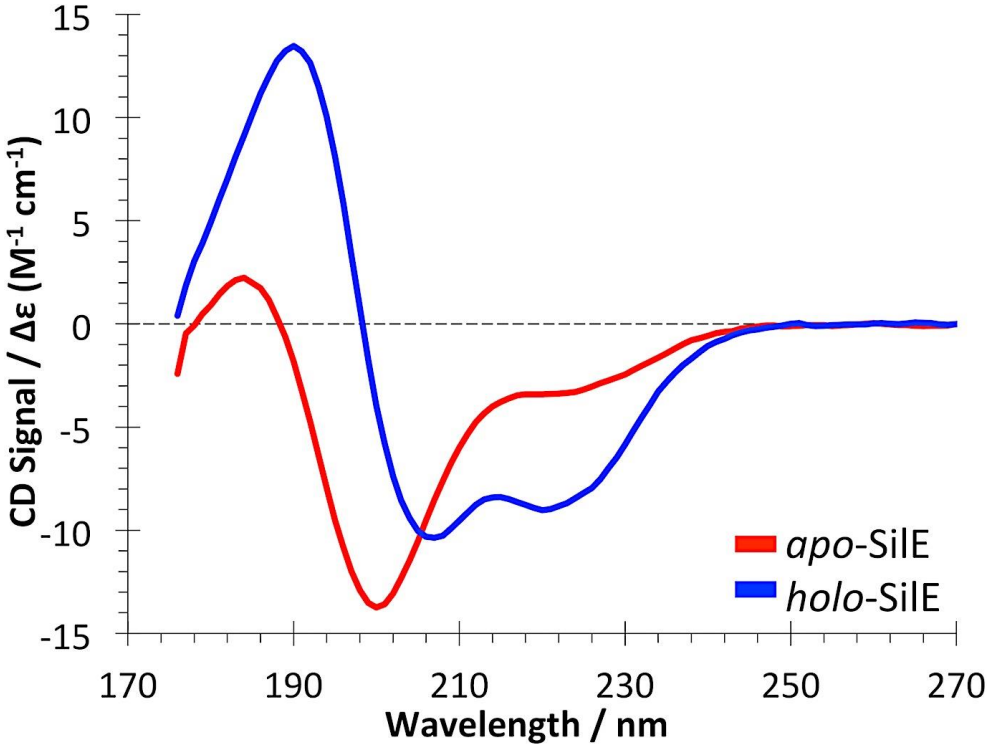


Figure 5. 4 ***Holo*-SilE secondary structure in the presence of Ag^{+}**

Folding and structure content of *apo*-SilE and *holo*-SilE (2 mM Ag^{+}), in 10 mM HEPES, 20 mM NaF, pH 7.5. Far-UV CD spectrum obtained at a concentration of 57mM for *apo*-SilE (red curve) represents very little secondary structure (negative signal around 200 nm). Upon addition of Ag^{+} (blue curve) largely negative signals are present at 207 and 221 nm typically found in proteins with significant helical structure.

5.2.4.2 Assessment of the metal ion binding and folding specificity of SilE

The secondary structure of SilE was also probed upon the addition of the divalent metal ions Cu^{2+}, Zn^{2+} and Ni^{2+} to a final concentration of 2 mM. Figure 5.5 shows that the addition of these metals did not lead to SilE adopting the same α-helical content as was seen after the addition of Ag^{+}. The far-UV CD obtained in the presence Cu^{2+} and Zn^{2+} followed the same

signal trends as the experiment upon the addition of Ag^{+} (negatives at 207 nm, 221 nm and a positive at 221 nm), yet with weaker final ellipticity intensity; thereby indicating that fewer α-helical stretches were formed than in Ag^{+}-bound SilE. The SilE-Ni^{2+} CD data on the other hand, resembled the trace of the *apo*-SilE protein (Figure 5.5-A) - representative of a favourably unfolded protein. So, where Cu^{2+} and Zn^{2+} ligation produced similar amounts of α-helical content (42 % and 40 % respectively), the addition of Ni^{2+} only increased the α-helical content by 4 % (20 % to 24 %) (Figure 5.5-C-E) (Provencher and Glockner, 1981; Van Stokkum *et al.*, 1990; Sreerama and Woody, 2000). Each fit produced low normalised root mean square error (NRMSE) values (explained in the CD section of Chapter 5, Section 5.2.4.2) in table 5.2. These reflected that the calculated percentages fit reasonably well with all of the metal-bound CD data obtained.

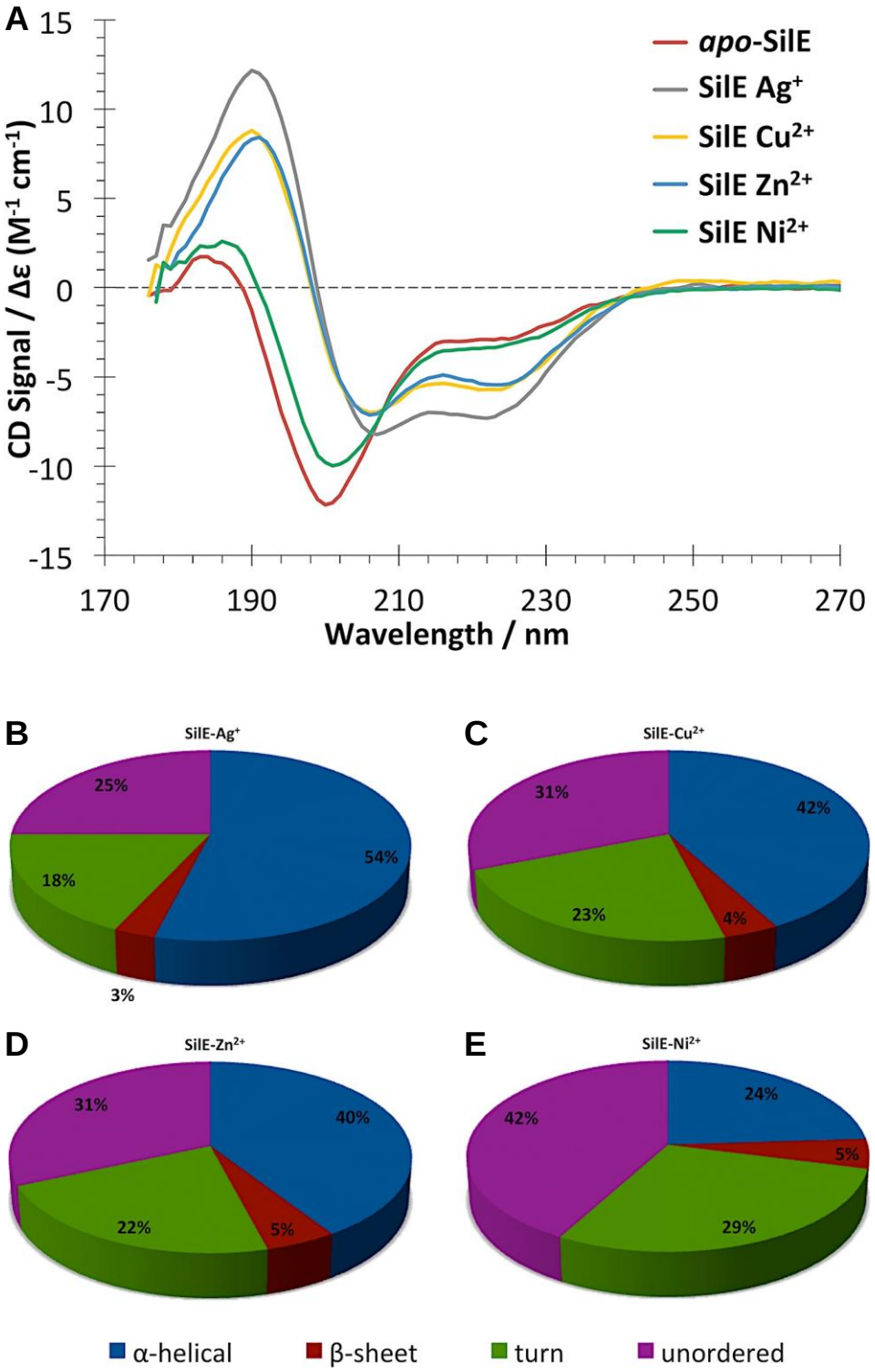


Figure 5. 5 Metal ion bound secondary structure content of SilE

A. Far-UV circular dichroism data (57 mM in 10 mM HEPES, 20 mM NaF, pH 7.5) exhibiting more alpha-helical protein content when SilE is in the presence of Ag^{+} compared to the other divalent metal ions, especially Ni^{2+}. All metal ions were soluble and added to a final concentration of 2 mM. **B-E.** Pie charts of consistent Dichroweb analysis, calculated using CONTILL, displays the percentage content of each type of secondary structure (Provencher and Glockner, 1981; van Stokkum *et al.*, 1990; Sreerama and Woody, 2000). **B.** SilE in the presence of Ag^{+} shows the highest level of helicity, whereas **C.** Cu^{2+}, **D.** Zn^{2+} and **E.** Ni^{2+} show decreasing levels, respectively.

Therefore, the CD data emphasised the specificity of SilE for Ag^{+} and better folding upon Ag^{+}-binding relative to the other metals. This suggested an important alteration in the secondary structure of SilE upon the addition and binding of Ag^{+} that crucially did not occur to the same extent with other metal ion types.

Table 5. 2: Metal bound CD secondary structure propensity NRMSE

NRMSE values for goodness-of-fit of the calculated secondary structure percentage based-model with the metal bound CD data for SilE complexes with Ag^{+}, Cu^{2+}, Zn^{2+} and Ni^{2+}.

Protein-Metal Complex	NRMSE Value
SilE-Ag^{+}	0.083
SilE-Cu^{2+}	0.103
SilE-Zn^{2+}	0.094
SilE-Ni^{2+}	0.043

5.2.4.3 SilE-Ag^{+} far-UV CD titration

Figure 5.6-A shows SilE ligand titrations, up to 16 molar equivalents of Ag^{+}, monitored using far-UV CD. It shows the spectral changes that were observed upon the successive additions of Ag^{+}, whereby a gradual shift of SilE from a mostly disordered structure over to a mostly α-helical structure was seen (Greenfield, 2007a; Dodero *et al.*, 2011). It appears that just 0.5 molar equivalents of Ag^{+} were sufficient to induce some secondary structure formation in SilE and the final shift was seen at a 1:6 SilE:Ag^{+} ratio, after which no more changes were observed. Therefore, the protein showed no further change in secondary structure content after the addition of six molar

equivalents of Ag^{+}. Hence six Ag^{+} ions appeared to be the optimum for full folding of the SilE protein. This agreed with the ESI-MS data (Figure 5.2) in which the distribution showed that SilE preferentially binds to six Ag^{+} in solution.

A specific wavelength at which the total absorption of light by a sample remains constant and the protein shows no physical change is known as the 'isosbestic' or 'isodichroic' point. The Ag^{+} CD titration showed a very clear isodichroic point around 206 nm, which was consistent with predominantly two species in solution, namely the *apo*-SilE and a predominant single 'fully loaded' Ag^{+}-bound form, rather than a heterogeneous mixture of different species with different binding stoichiometries (Kypr and Vorlícková, 1986; Holtzer and Holtzer, 1992).

Furthermore, plots of the titration CD signals at 190 and 221 nm as a function of the Ag^{+}/SilE molar ratio were produced (using GraphPadPrism for Mac OS X, Version 6.0h, GraphPad Software, La Jolla California USA, www.graphpad.com). These plots in Figure 5.6-B showed some evidence for sigmoidal curves consistent with the co-operative uptake of Ag^{+} by SilE.

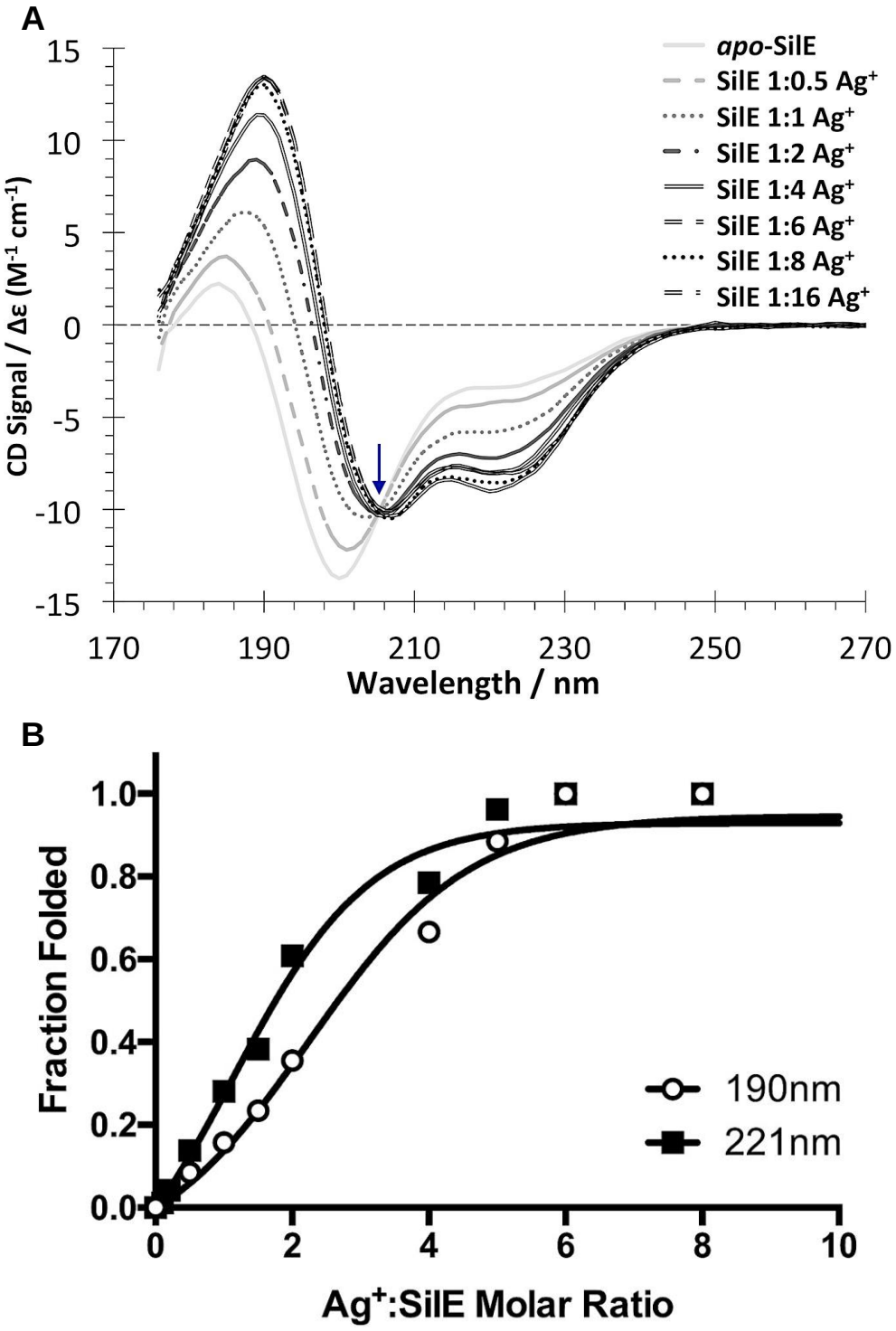


Figure 5. 6 **SilE-Ag^+ CD-based titration**

SilE titration, 75 µM in 10 mM HEPES, 20 mM NaF, pH 7.5, from 0 to 16 equivalents Ag^+ using far-UV CD. **A.** A change in protein secondary structure formation, towards helical content is evident following the addition of just 0.5 equivalents Ag^+ to SilE (dashed grey). However, there is no change in signal following Ag^+ addition beyond 6 Ag^+ equivalents to SilE (black double-line dashed). The isosbestic point at 206 nm is highlighted with a navy arrow. **B.** Plots of the CD ellipticities as a function of the Ag^+:SilE molar ratios at 190 nm (clear dots) and 221 nm (black squares) produce slight sigmoidal curves, indicative of co-operative uptake of Ag^+ ions by SilE.

5.2.5 SilE folding is independent of the environmental pH

It is well known that besides the binding of a ligand, the pH dependent change of a charge may result in protein conformational changes in the environment of the chemical modification (Kukic´ *et al.*, 2010). The protonation state of an ionisable amino acid such as aspartate, glutamate, histidine and cysteine has a large influence on the interactions of the amino acid side chains with the environment and can also affect protein structure and stability (Di Russo *et al.*, 2012). This is because altering solution pH leads to changes in the hydration of ions and hydrophobic interactions between water and hydrophobic amino acids (Kukic´ *et al.*, 2010).

Silver, (2003) and Mirolo *et al.*, (2013) proposed a model in which ten histidine residues were suggested to bind Ag^+ and coordinate the folding of SilE. Therefore, it was particularly important to assess how SilE protein folding was affected by a change in pH. One-dimensional (1D) proton (^1H) nuclear magnetic resonance (NMR) pH titrations, in 10 mM Na_2HPO_4/NaH_2PO_4 (buffer used as does not contain exchangeable protons), 25 mM NaCl at one-pH unit reductions between 9.0 and 5.0 and at pH 2.0, were performed to assess changes in the local environment of *apo*-SilE (Chapter 2, Section 2.2.6.3). Phosphoric acid (H_3PO_4) was added to the protein sample in the Na_2HPO_4/NaH_2PO_4 buffer as a means of decreasing the pH level. As explained in the NMR section of Chapter 4, Section 4.2.4.2, a folded protein tends to have well-dispersed and sharper peaks while the spectrum of an unfolded protein consists of broader, poorly dispersed peaks. *Apo*-SilE was verified as being an unfolded protein in Chapter 4 and so any change in the conformation of the protein i.e. shifts in

chemical peaks upon altering the ionic environment, were to be accounted as being pH-induced.

Figure 5.7 shows the 1D ^{1}H NMR data from pH 9.0-2.0. As is apparent, the form of *apo*-SilE stayed stable across the high to low pH ranges. With no resonance perturbations, the unfolded nature of the protein remained stable (Figure 5.7-A). A closer look at the amide (NH)/aromatic region (6-12 ppm) of the spectra in Figure 5.7-B showed the appearance of peaks, around the 8.0-8.6 ppm region, as the pH decreased past pH 7.0. Nevertheless, the presence of these peaks at the lower pH values was feasibly due to the protonation of the imidazole side chain in histidine at more acidic pH levels, and not because of a conformational change in the protein - as, the histidine imidazole ring (an aromatic heterocycle) is known to be protonated at any pH below 6.0, the pKa value of the imidazole side chain. So due to a reduced chemical exchange of the amide hydrogen atoms, this set of signals was more intense and better resolved as the pH decreased (Pineda-Sanabria *et al.*, 2013).

Consequently, the 1D ^{1}H NMR experiment in Figure 5.7 showed that SilE folding was ligand binding-induced and not solution pH dependent.

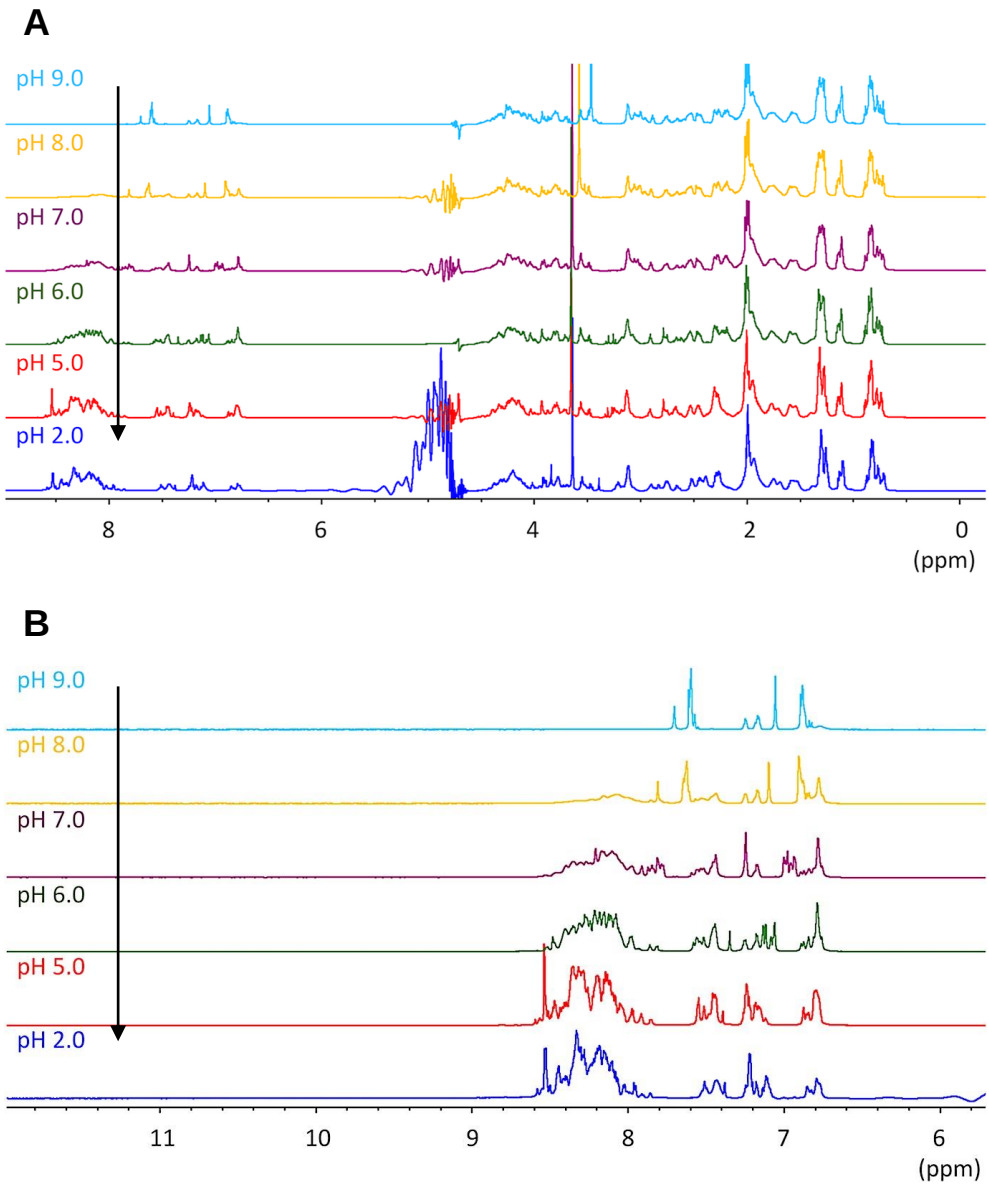


Figure 5. 7 ***Apo*-SilE ^{1}H NMR pH titration**

A. ^{1}H NMR spectrum of *apo*-SilE, in in 25 mM Na_2HPO_4/NaH_2PO_4, 25 mM NaCl buffer at various pH units. There is no apparent change in protein folding, as the pH is decreases. **B.** An expanded view of the aromatic/NH region shows signal amplification with decreasing pH levels.

5.2.6 Assessment of the thermal stability of Ag^+-bound SilE

Assessment of the temperature stability and melting point (T_m) of a protein is useful for future structure determination experiments, such as NMR, as these are likely to be relatively long (in excess of a week). T_m is defined as the temperature at which a protein denatures. In order to obtain an accurate T_m value; a fully folded SilE protein was required. Thermal unfolding studies were carried out with Ag^+ to a final concentration of 2 mM, in 10 mM HEPES, 20 mM NaF, pH 7.5. Once the initial far-UV CD scans were completed to check for presence of secondary structure, the thermal stability of the Ag^+-bound SilE complex was investigated using CD by generating a melting curve and monitoring the change in ellipticity at 207 nm (the strongest peak minima for the α-helical, *holo*-form of SilE (Figure 5.4)). The temperature was increased at 1 ^oC intervals from 18 ^oC to 80 ^oC (Chapter 2, Section 2.2.6.1). In the protein melting curve (shown in Figure 5.8), the unfolding showed a sigmoidal transition with a mid-point and hence a T_m of 52 ^oC. Moreover, the relatively sharp transition from the folded to unfolded form as the temperature increased, was characteristic of two-state proteins (Horng *et al.*, 2003), and suggested that only these two conformational states were present to any significant extent.

It appeared that SilE folding upon binding to Ag^+ occured cooperatively, as indicated by the sigmoidal thermal stability curve deduced via far-UV CD (Figure 5.8). The monophasic, cooperative unfolding of the protein confirmed that with Ag^+ present, the protein existed as a compact well-folded, stable structure up to a temperature of 42 ^oC.

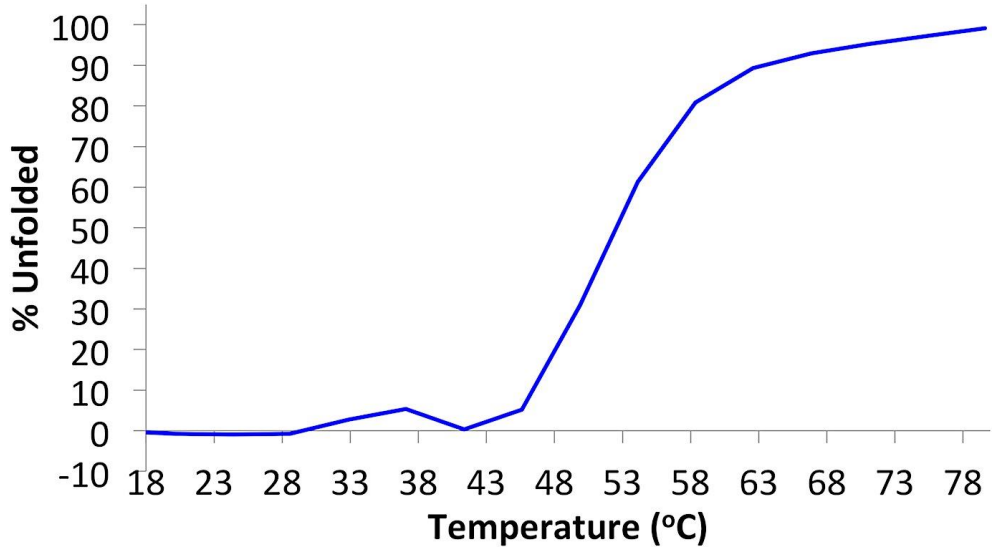


Figure 5. 8 Melting curve of SilE, in the presence of Ag^+

Temperature-induced denaturation was monitored using CD, in order to assess the thermal stability of SilE - 57 µM protein with 2 mM Ag^+ in 10 mM HEPES, 20 mM NaF buffer at pH 7.5. The CD was measured at the maximum negative signal of 207 nm and the spectrum obtained indicates that the protein is stable to about 42 ^oC. The sigmoidal temperature denaturation of *holo*-SilE illustrates the fact that the protein folds in a cooperative manner with a temperature denaturation midpoint of 52 ^oC.

5.2.7 Attempts at deducing the folded structure of the SilE-Ag^+ complex

No structures for the protein SilE or homologue PcoE have been obtained. Hence, determining a high-resolution, three-dimensional (3D) structure of SilE would prove to be beneficial in understanding the functional role of SilE at the molecular level in the mechanism of bacterial silver resistance. Therefore, various attempts in the form of protein crystallisation for X-ray crystallography, SAXS and NMR were made to deduce it.

5.2.7.1 SilE-Ag^+ protein crystallisation

X-ray crystallography uses the diffraction of X-rays to study crystals and their structure at an atomic and molecular level. Proteins will solidify into crystals under certain conditions, yet obtaining well diffracting crystals is one of the major barriers in solving the atomic-resolution structure (Geerlof *et al.*, 2006).

In order to determine the high-resolution structure of Ag^+-bound SilE, prior to x-ray crystallography, crystallisation was attempted. Crystallisation was attempted using Hampton Research (USA) Natrix and Emerald Biosystems (USA) Wizard™ I and II (screens at the Research Complex at Harwell, with the help of Dr. Anil Verma from the Oxford Protein Production Facility- (OPPF) UK). Initially SilE at a concentration of 5 mg mL^{-1} had been dialysed into 10 mM HEPES, 20 mM NaF, pH 7.5 buffer with 500 mM $AgNO_3$ followed by extensive dialysis without $AgNO_3$ to remove excess Ag^+. The PCT™ Pre-Crystallization Test (Hampton Research (USA)) was then used to determine the appropriate protein concentration for the crystallisation screening.

Whilst Natrix is a nucleic acid sparse matrix screen biased for the crystallisation of protein-nucleic acid complexes, Wizard™ I and II is a sparse matrix screen designed for the coarse screening phase of crystallising biological macromolecules, including combinations of proteins with other binding partners. Sparse matrices cover a wide range of crystallisation space by allowing the sampling of a large range of crystallants, buffers, polymers, organics, solution pH values, salt conditions and precipitants using a very small volume of protein sample (Scott *et al.*, 1995). Natrix screens cover a

pH range of 5.6 to 8.5 and Wizard^{TM} I and II include a broader range, from pH 4.5 to 10.5.

After continual monitoring of the crystallisation screening blocks stored at 21 ^{o}C (over a period of five months), it was noted that a large number of drops contained amorphous precipitant. Figure 5.9 shows one of the few that remained clear where micro-crystals were detected. However, neither of these diffracted and were logged as salt crystals. The most probable reason for this is the formation of Ag-salts, which are commonly known to precipitate (discussed in Section 5.2.1).

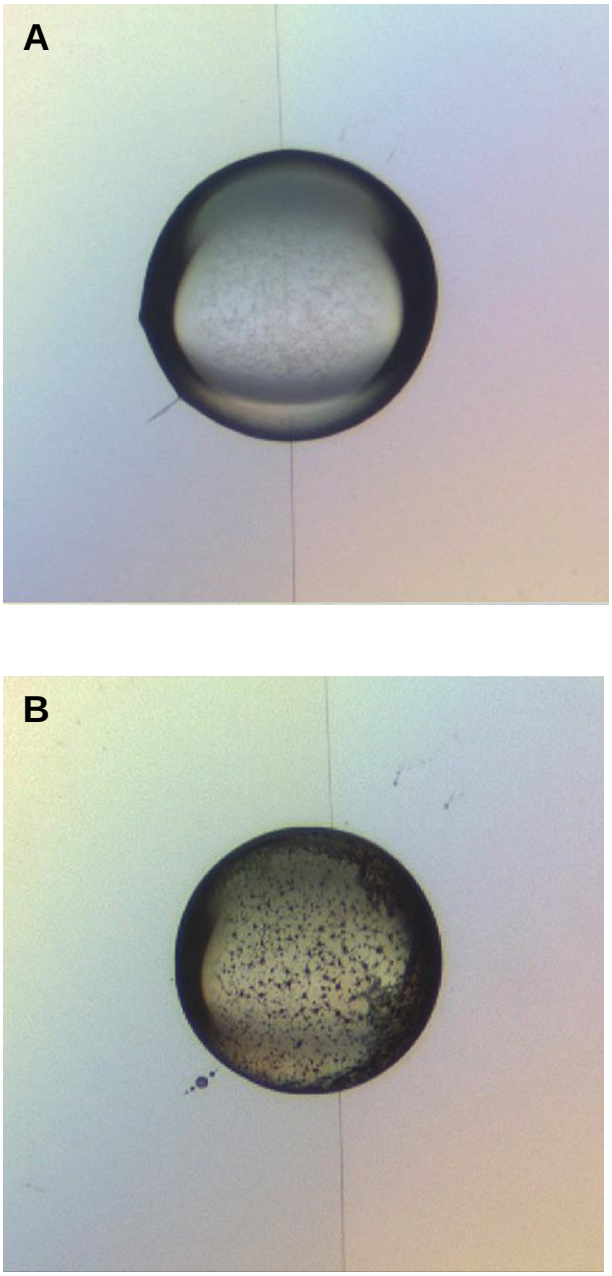


Figure 5. 9 **SilE-Ag^+ crystallisation plate experiment**

A. The first and **B.** the last image of the crystallization screen of SilE-Ag^+, originally in 10 mM HEPES, 20 mM NaF, pH 7.5.

Screen: Emerald Wizard I+II
Condition: Emerald Wizard I+II (D10)
10.0% w/v polyethylene glycol 8000
0.1 M Imidazole pH = 8.0
0.2 M Calcium acetate.

B. The visible micro-crystals are quite possibly the formation of silver acetate precipitate.

5.2.7.2 The SilE-Ag^+ SAXS problem

Following the failure of protein crystallisation, the low-resolution method of small-angle X-ray scattering (SAXS) (Chapter 2, Section 2.2.6.4) was carried out in the attempt to gather some tertiary and quaternary structure-based information on the folded, Ag^+-SilE complex. To ensure perfect buffer matching, SilE was dialysed in buffer containing 500 mM $AgNO_3$ and excess Ag^+ was removed upon extensive dialysis in buffer without any $AgNO_3$. SAXS investigations were then carried out on Ag^+-SilE, (in 10 mM HEPES, 20 mM NaF, pH 7.5,) at protein concentrations within the range of 1.25-5 mg mL^{-1}, using the outer dialysis buffer as the subtraction data.

The SAXS results in Chapter 4, Section 4.2.4.2, showed how SAXS is primarily a contrast method between the electron densities of the buffer and the protein sample (Putnam *et al.*, 2007). Therefore, it is important to obtain fitting SAXS data from both the sample in buffer solution and the buffer solution alone. This is because verification of correct buffer subtraction is fundamental before considering any further data analysis (Figure 4.11, Chapter 4). Unfortunately, the addition of Ag^+ (and the other divalent transition metal ions – Cu^{2+}, Zn^{2+} and Ni^{2+}) in the case of SAXS did not allow for this. This is because, when using ScÅtter (Biosis.net, 2016) for data evaluation, the average buffer curve subtracted from the average Ag^+-SilE curve always resulted in negative low *q* data, regardless of the sample concentration (Figure 5.10). The negative values seemed to arise from the solvent data presenting higher electron densities than the sample in

the buffer at the low *q* range, which led to over-deduction of the background.

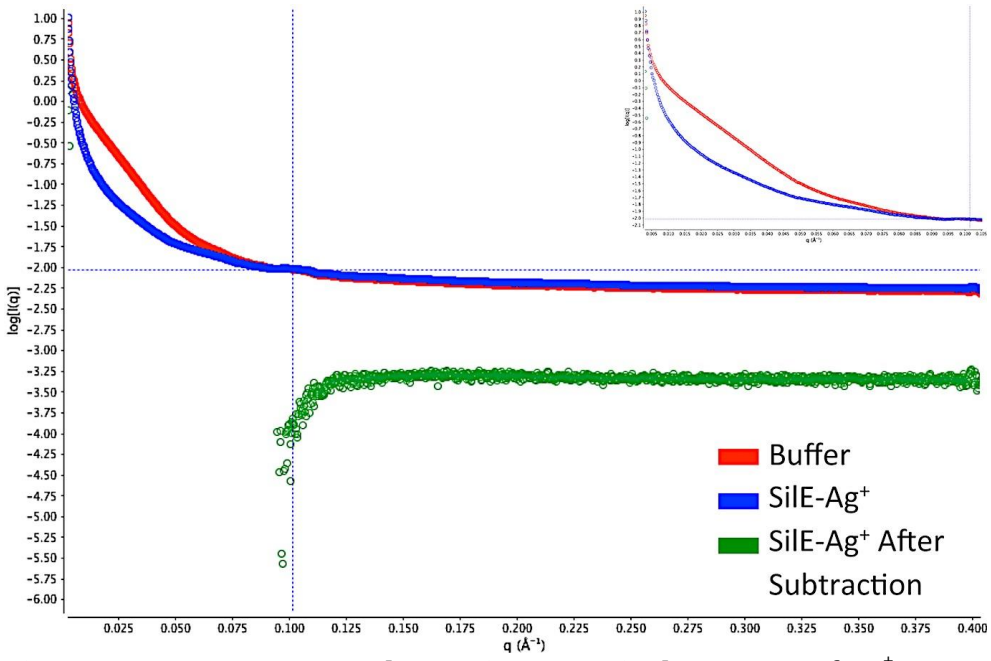


Figure 5. 10 SAXS averaged scattering curves in the presence of Ag^+

SAXS scattering patterns were obtained with 10 mM HEPES, 20 mM NaF, pH 7.5 and dialysed Ag^+ as the buffer. At low *q* values, the averaged buffer alone (red) shows higher scattering intensities than the averaged protein sample in buffer (blue), resulting in an over-subtracted curve for further SAXS analysis.

Figure 5.11 shows the individual buffer and protein sample (1.25 mg mL^{-1} SilE-Ag^+) curves generated from each of the 18 buffer scans. Upon analysis of the individual scattering patterns of the buffer alone, it became apparent that the problem was an increasing electron density of the buffer over time.

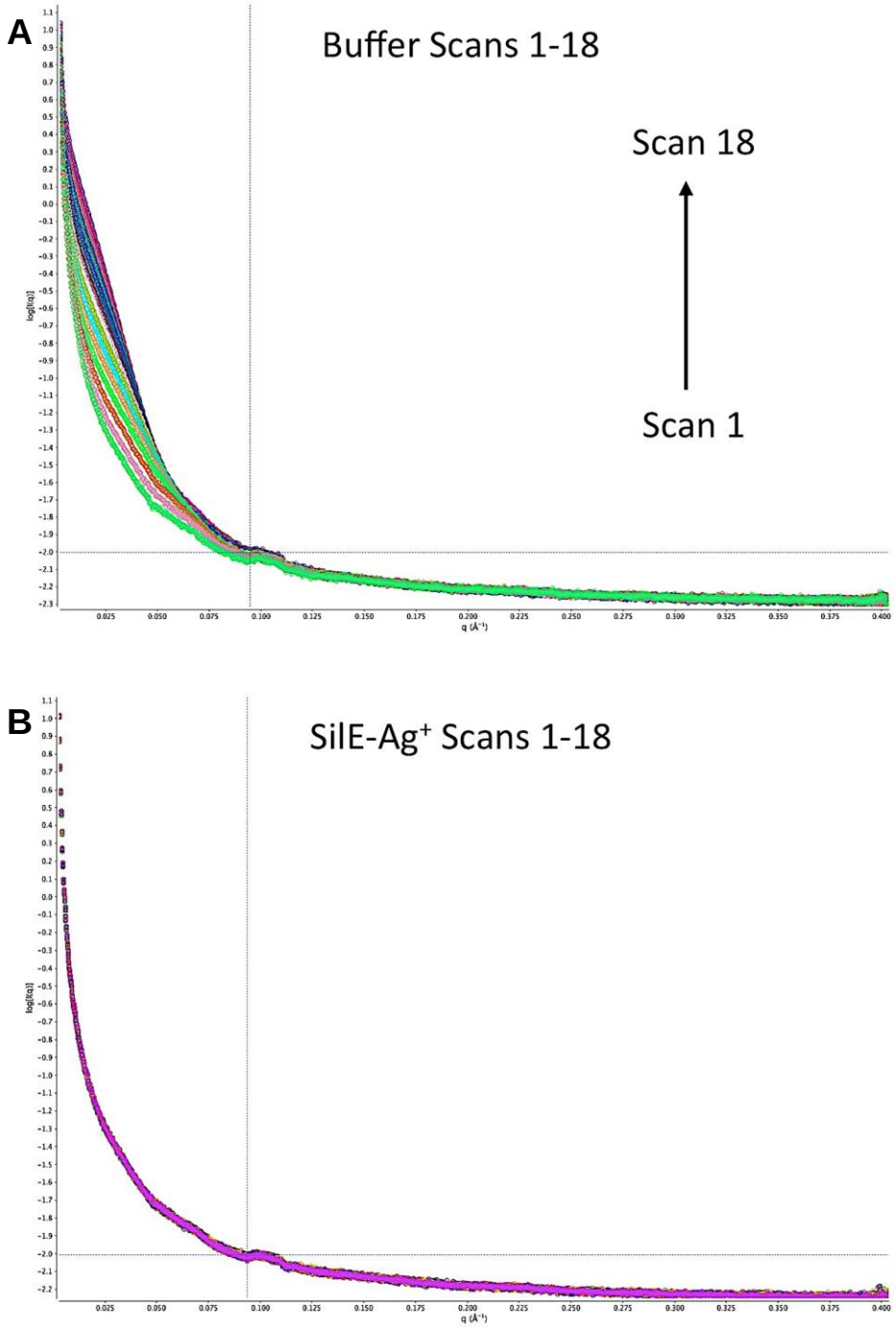


Figure 5. 11 SAXS individual scattering patterns

A. The individual SAXS scattering patterns were obtained with 10 mM HEPES, 20 mM NaF, pH 7.5 and dialysed Ag^{+} as the buffer. At low *q* values, below 0.100 the electron density of the buffer increases with each simultaneous scan (1-18). **B.** The same buffer composition containing the SilE protein sample in solution displays odd scattering at the low *q* values too. The most valid explanation for this is Ag^{+} photo-reduction and Ag aggregate formation.

Photon absorption and photo-reduction of Ag^{+}

One explanation for the high scattering at low *q* values for the buffer scans, is due to background scattering from the capillary (see Chapter 2, Section 2.2.6.4 for information on the instrument). The scattering pattern obtained from the empty capillary is far more intense than the background from the beamline. Therefore, the intensity of the capillary background directly affects the quality of data, especially with contamination on the capillary surface. Dirt, dust or sample deposits can stick to the capillary surface causing inefficient data collection. In their paper based on the use of a new sample changer, Round *et al.*, (2015) discussed the importance of capillary cleaning between all measurements to avoid cross contamination, especially because buffer and protein solute measurements should have the same background reading. They also stated how radiation-sensitive samples leave deposits on the walls of the capillary after normal cleaning.

Transition metals are known to be susceptible to radiation damage by X-ray photons (Stumpf *et al.*, 2016). Depending on the initial charge of the metal, it may become oxidised, reduced or even remain unaltered and the manifestation of its photo-reduction is often considered to be the result of this radiation damage (Mesu *et al.*, 2006; George *et al.*, 2008; George *et al.*, 2012). Stumpf, Gokhberg and Cederbaum, (2016) investigated the photo-absorption of X-ray photons in metals by studying the different effects of X-rays on magnesium cations (Appendix B) and showed different processes for the mechanisms of decay. Essentially, when a metal absorbs an X-ray photon a core electron is removed and a cascade of relaxation processes is initiated. The three main processes described as being involved

are: (i) the Auger decay, (ii) interatomic Coulombic decay (ICD) and (iii) electron-transfer-mediated decay (EMTD). Auger decay is the first step for typical metals, during which electron emission accumulates a positively charged metal species. Additionally charging the metal atoms in biological environments provides considerable amounts of energy, which encourage the energy transfer driven ICD and charge transfer driven EMTD as part of multistep cascades with the surrounding molecules. In ICD, electronically excited metal atoms relax by ionising a neighbouring molecule. Outer-valence multiple ionised species do not have sufficient excess energy for the ionisation of the neighbours and therefore, do not undergo ICD. So instead, relaxation mechanisms involving ETMD lead to the neighbour donating an electron to the metal species, which is in turn reduced with the excess energy being simultaneously used to ionise the donor or another molecule.

The elemental Ag atom has the electron configuration $1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10} 4s^2 4p^6 4d^{10} 5s^1$, with the valence electron on the outer shell presenting the 1+ charge (Appendix B). Therefore, it is more than likely that the significant quantity of unbound Ag^+ in the buffer (as opposed to the protein sample) during SAXS was photo-reduced. Photo-reduction of Ag^+ has previously been discussed whereby the photo-irradiation results in the formation of Ag nano-particles and aggregates (Remita *et al.*, 2007; Huang and Yang, 2008; Abou El-Nour *et al.*, 2010; Harada *et al.*, 2013). With the help of literature by Hada, Yonezawa and Kurakake, (1976) the photolysis of the $AgNO_3$ in aqueous solution can be explained by Eq. 5.4. As Ag^+ absorbs the X-ray photons ($h\nu$), the reaction proceeds to reduce Ag^+ via electron transfer from the water molecule.

$$Ag^{+}_{(aq)} + H_2O_{(aq)} + h\nu \rightarrow Ag_{(s)} + H^{+}_{(aq)} + OH^{-}_{(aq)} \qquad (Eq.\ 5.4)$$

Aside from water, previous work has also reported HEPES as a reducing agent in the fabrication of silver nanoparticles (Sun *et al.*, 2005; Tan *et al.*, 2007; So *et al.*, 2010; Hui Li *et al.*, 2015a). Hence, a rising population of Ag from photo-reduced Ag^{+} could essentially explain the increasing electron density that was seen at the low *q* values (below 0.100 $Å^{-1}$), from the scattering pattern of the buffer in Figure 5.11-A. In turn the formation of Ag aggregates would have stuck to the capillaries and significantly affected the scattering pattern of the buffer. Additionally, the cascades are known to cause extensive damage to surrounding molecules thus the Ag^{+}-bound SilE protein scatter would have also been disturbed.

5.2.7.3 Progress towards the SilE-Ag^{+} NMR structure

Although NMR analysis is more time consuming than other biophysical techniques, it is undoubtedly the most suitable tool for investigating the detailed properties of disordered proteins. NMR studies of IDPs have proved to be extremely useful in probing, understanding and describing their conformational preferences at the residue level. This is due to chemical shifts being very useful for monitoring disorder-to-order transitions upon binding to partners (Sibille and Bernadó, 2012). Furthermore, the wealth of observables that this technique can provide include: the measurement of detailed protein dynamics and structure, ligand

binding affinities and the residues involved in ligand binding (Wright and Dyson, 1999; Tompa, 2002; Kosol, Contreras-martos, *et al.*, 2013).

The ^{1}H-^{15}N Heteronuclear Single Quantum Coherence (HSQC) experiment is frequently one of the first NMR experiments conducted and is a key technique used for protein folding and protein-ligand binding studies (Williamson, 2013; Cavanagh). Other advantages of performing HSQC experiments include the ability to investigate a variety of buffer conditions and temperatures in order to assess the effect on a protein, and the ability to monitor protein stability over time. This allows for optimisation prior to the 3D NMR experiments and consideration on suitability of ^{13}C-carbon labelling your protein, as this is a time-consuming and expensive process. As explained in Chapter 2, Section 2.2.6.3, the HSQC is a 2D (two-dimensional) experiment, which correlates the ^{1}H and ^{15}N chemical shifts of protein NH groups and requires a ^{15}N-lablled-protein sample that has an observable nuclear spin of ½. There is a natural abundance of 0.37 % of ^{15}N, however with NMR sensitivity for ^{15}N being 1.04×10^{3} times lower than that of ^{1}H, ^{15}N-isotopic labelling provides more efficient data collection (Gust *et al.*, 1975).

Isotopically labelling SilE

^{15}N isotopically labelled SilE was overexpressed in M9 minimal media supplemented with $^{15}NH_4Cl$ as the sole source of nitrogen using *E. coli* BL21 Star^{TM} (DE3) cells (Table 2.1, Chapter 2) containing the pOPINF plasmid expressing $SilE_{21\text{-}143}$ (Chapter 2, Section 2.1.4.3). Bacterial growth and consequently, protein overexpression in minimal media, is generally

slower than in LB. Therefore, the growth temperature was reduced to 30 ^{o}C following IPTG induction to allow for a longer period of bacterial and protein growth.

Figure 5.12 shows pure ^{15}N-SilE, following identical protein purification as was used for the unlabelled protein, (described in Chapter 2, Section 2.2.5.2). A final yield of around 20 mg L^{-1} of pure ^{15}N labelled protein per litre of bacterial culture was obtained.

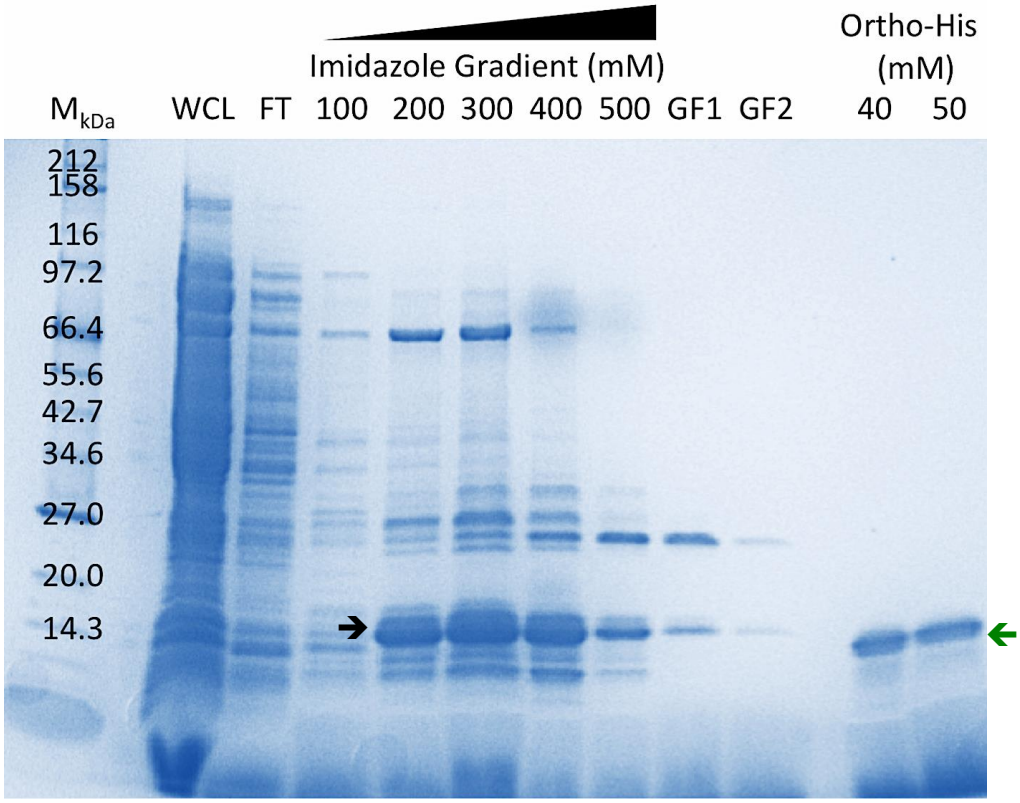


Figure 5. 12 **^{15}N-SilE overexpression and purification**

PAGE gel of protein purification alongside a high molecular weight protein marker - Unstained Protein Marker, Broad Range (2-212 kDa) (New England Biolabs, Ipswich, UK) - using 16 % SDS-PAGE. The gel was stained and de-stained as mentioned in Chapter 2.2.2.2. *E. coli* BL21 Star^{TM} (DE3) cells transformed with the His_6-tagged $SilE_{21-143F}$ plasmid, were induced in M9 minimal media by adding 1 mM IPTG, and left to incubate at 30 ^{o}C. The resultant expression of $SilE_{21-143F}$ (with a His_6-tag) can be visualised in the whole cell lysate (WCL) sample. The protein, as a component of the WCL, was run through nickel affinity chromatography and retained within the column through binding of the His_6-tag. The columns affinity for the His_6-tag was outcompeted by the addition of excess imidazole (in the elution buffer), in a step-wise gradient. The flow through (FT) and 100 mM imidazole wash eliminated most contaminants. The protein (15.8 kDa – black arrow) was predominantly eluted and collected in fractions containing 200-400 mM imidazole. Gel filtration further purified the protein (GF1 and GF2). HRV-3C protease treatment was used to the N-terminal His_6-tag. A final ortho-trap purification step eluted pure ^{15}N-SilE in 40 and 50 mM imidazole (13.2 kDa, green arrow).

Ag^+-SilE two-dimensional NMR

A ^1H-^{15}N HSQC experiment was performed using a 350 μM sample of ^{15}N-isotopically labelled Ag^+-SilE in 10 mM HEPES, 20 mM NaF (experimental method described in Chapter 2, Section 2.2.6.3) on an 800 MHz NMR instrument at the University of Nottingham. The pH of the buffer was 8.0 prior to dialysis against 500 mM Ag^+, in order to ensure that the pH was one unit above the calculated isoelectric point of the protein: 6.86 (pI), so as to avoid any protein precipitation. The buffer pH for this sample could not be measured following the addition of the Ag^+, as the metal ions form deposits on the probe and interfere with the recorded data (Riddle, 2013). Additionally, due to limited time access to the NMR machine at the time of data acquisition, the sample was almost one month old. It had been stored at varying temperatures – from 4 ^oC to room temperature (approximately 20-25 ^oC) throughout this period of time. The spectrum of the results obtained is presented in Figure 5.13. In comparison to the *apo*-SilE data (Figure 4.10-B, Chapter 4) the peaks were well dispersed. These chemical shift perturbations indicated a conformational change and protein folding through Ag^+-binding, as all of the NH protons were in distinct chemical environments. Additionally, there seemed to be no obviously unstructured regions i.e. clustered peaks as was observed for *apo*-SilE in Figure 4.10-B.

The $SilE_{21\text{-}143F}$ protein consisted of 125 residues (123 from the SilE protein itself and two remaining from the HRV-3C protease cleavage site). The number of expected peaks therefore became 117 (having removed the seven proline and initial N-terminal residues that are usually not seen in a

HSQC experiment – explained in Chapter 2, Section 2.2.6.3). Nevertheless, an initial peak count found approximately 107 peaks on the HSQC spectrum (Figure 5.13). When allowing for some overlap, this suggested that most of the residues (91 %) of SilE in the Ag^{+}-bound form were visible. To optimise the number of peaks prior to attempting triple resonance experiments and to maintain a biological environment, the experiment was repeated at physiological pH.

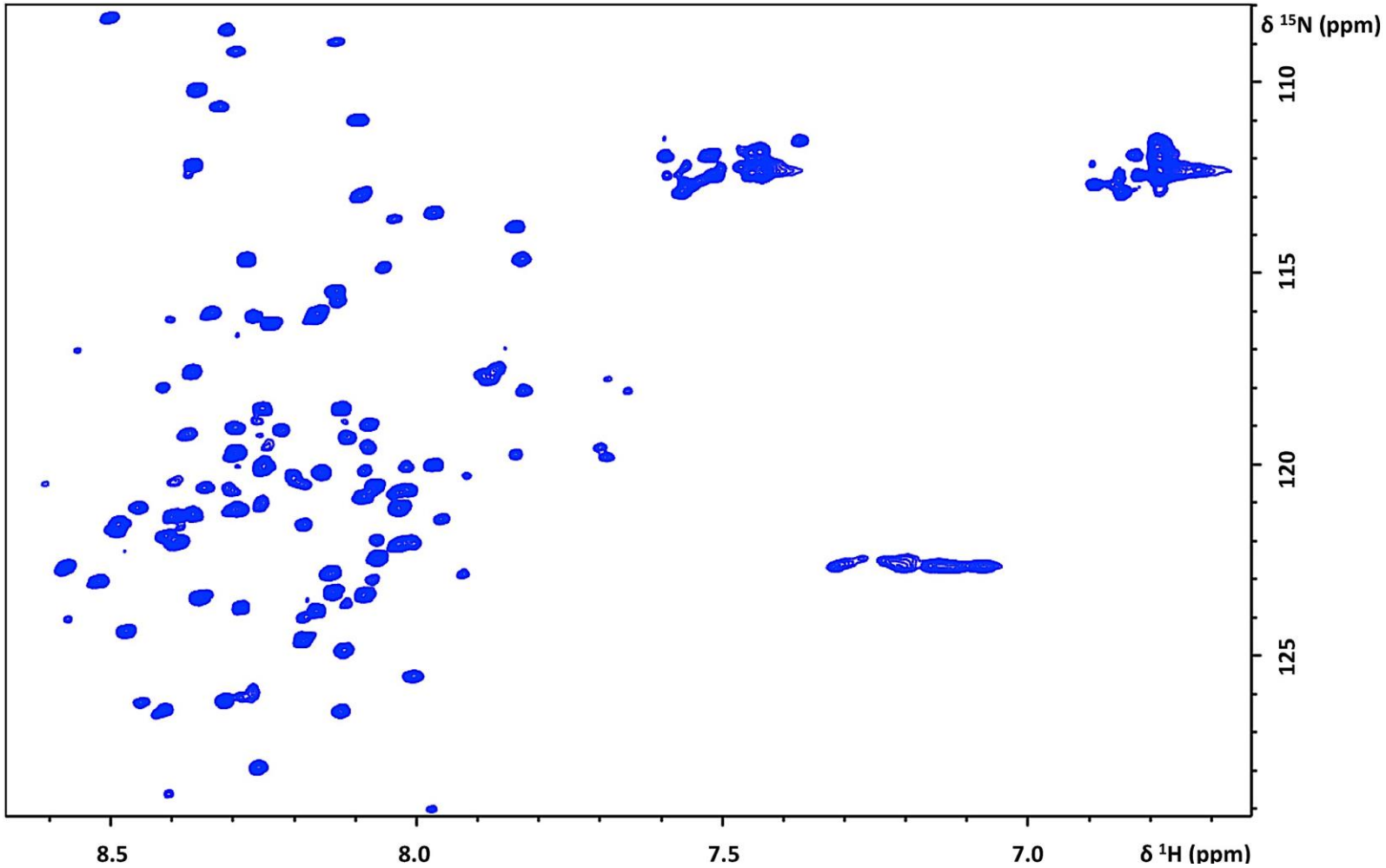


Figure 5. 13 ^{1}H-^{15}N HSQC spectra of Ag^{+}-SilE

^{1}H-^{15}N HSQC data obtained from a one month old, 350 µM ^{15}N-labelled protein sample, in 10 mM HEPES, 20 mM NaF, pH 8.0 (7.5), recorded at 25 ^{o}C. The sample was stored at both 4 ^{o}C and 20-25 ^{o}C throughout the month. The well-dispersed peaks and number of individually visible peaks (107) indicate a well folded and structured protein in the presence of Ag^{+}.

Figure 5.14 shows the repetition of the ^{1}H-^{15}N HSQC experiment of Ag^{+}-SilE at 300 μM, in which the buffer pH was measured as 7.5 before Ag^{+} dialysis. Some sample precipitation was observed and it appeared that lowering the pH had reduced the number of discernable peaks. Approximately 70 chemical shifts were seen (Figure 5.14), accounting for only 60 % of the proteins amino acids, as opposed to the 107 that were observed at pH 8.0 (overlay in Figure 5.13). Nevertheless, the protein had some structure compared to *apo*-SilE (Figure 4.10-B). For this reason the same sample was subjected to transverse relaxation-optimised spectroscopy (TROSY) measurements (Figure 5.15) in order to see if a better fingerprint of the protein, with a clearer and narrower peak resolution, could be acquired (Pervushin *et al.*, 1997).

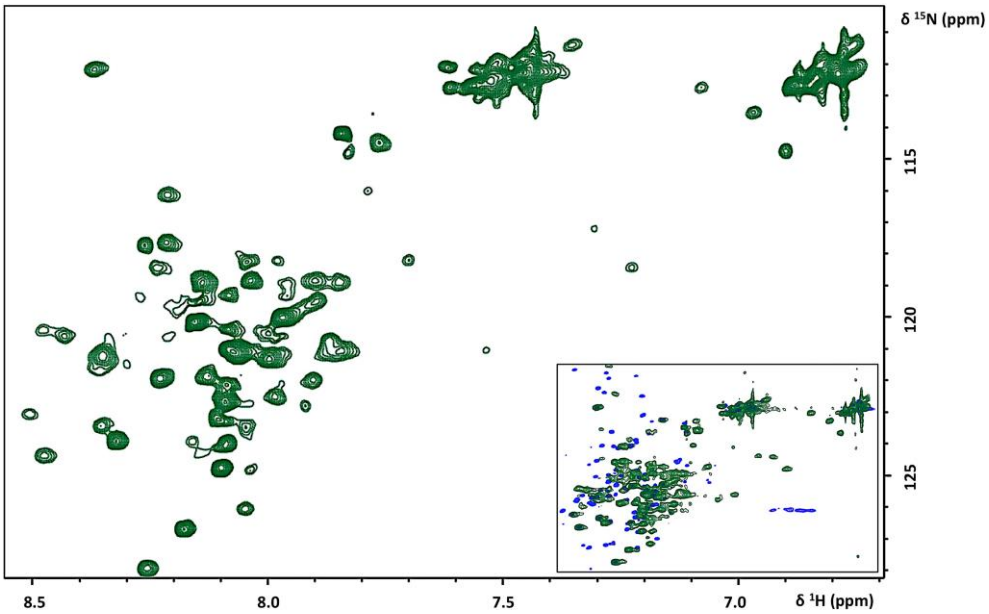


Figure 5. 14 Ag^{+}-SilE ^{1}H-^{15}N HSQC at physiological pH

^{1}H-^{15}N HSQC data obtained from 300 μM ^{15}N-labelled protein, in 10 mM HEPES, 20 mM NaF, pH 7.5 (7.0), recorded at 25 ^{o}C. The significantly lower number of distinctly obvious green peaks (70) indicates only some folding in the presence of Ag^{+} at this pH (see superposition with Figure 5.13 (blue peaks) in bottom right). Therefore the pH difference seems to be an important factor in SilE-Ag^{+} NMR data acquisition.

Figure 5.15 presents the Ag^{+}-SilE data acquired using the ^{1}H-^{15}N TROSY HSQC experiment. Where the TROSY showed some sharper signals in evaluation against the HSQC there seemed to be fewer individual chemical shifts and more signal overlap on the whole, in which the spectral resonances looked more clustered especially between 8.0 and 8.5 ppm in the ^{1}H dimension and 120 and 125 ppm in the ^{15}N dimension.

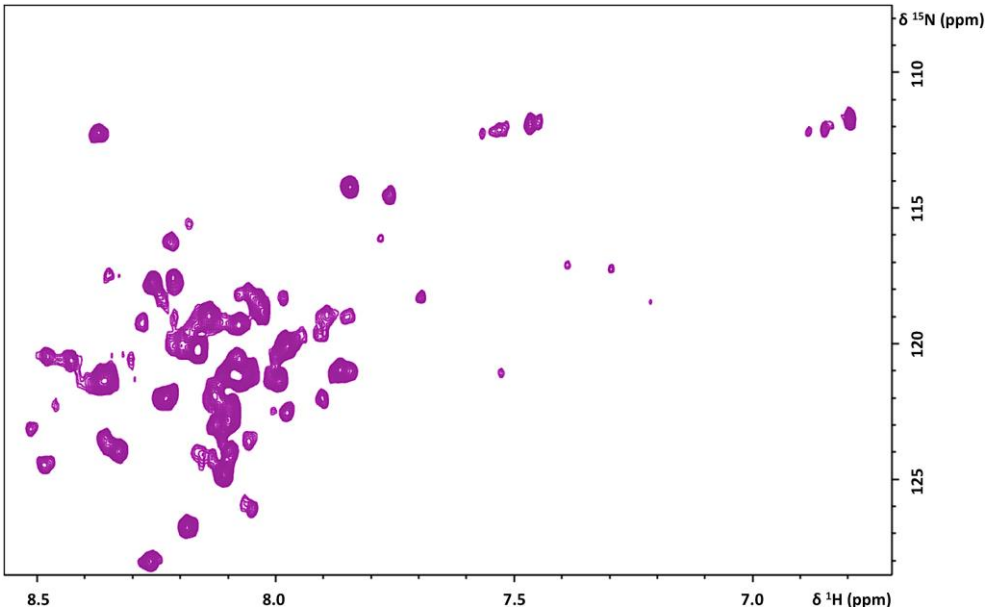


Figure 5. 15 Ag^{+}-SilE ^{1}H-^{15}N TROSY HSQC

^{1}H-^{15}N TROSY HSQC data obtained from 300 μM ^{15}N-labelled protein, in 10 mM HEPES, 20 mM NaF, pH 7.5 (7.0), recorded at 25 ^{o}C. With higher sensitivity the data shows a few sharper signals yet a substantial degree of overlap, just like the HSQC in Figure 5.14.

Due to the poorer spectral quality at the lower physiological pH, Ag^{+}-SilE pH titration-based ^{1}H-^{15}N HSQC experiments were carried out at 0.5 unit intervals, between pH 6.0 and 7.5 to assess the pH affects on protein stability and spectral quality. The adjustment of the pH value of a sample in solution is thought to be crucial for the acquisition of NMR data. Even the slightest variation of just 0.005 can have a significant affect in the chemical shift of an amino acid (especially for the commonly ionisable Asp, Glu, His, Cys, Tyr, Lys, and Arg residues) (Platzer *et al.*, 2014).

HSQC spectra measurements were taken at pH units 6.0, 6.5, 7.0 and 7.5. Figure 5.16 (A-D, respectively) exhibits the obtained data. Following Ag^{+} dialysis, the sample concentration and initial pH were measured as 350 μM and adjusted to pH 6.0. The 0.5 unit pH increments were then achieved

through the addition of small volumes of NaOH, whilst also trying to ensure the maintenance of protein concentration. However, as the pH got closer to 7.5, at each addition of NaOH, an increasing amount of SilE-Ag^+ protein precipitated.

This precipitation problem correlated with the number of resonances observed at each pH: 90 peaks at pH 6.0 (Figure 5.16-A), 78 peaks at pH 6.5 (Figure 5.16-B), 70 peaks at pH 7.0 (Figure 5.16-C) and a mere 50 at pH 7.5 (Figure 5.16-D). These corresponded to 77 %, 67 %, 60 % and 43 % of the expected number of peaks. With a decrease in the number of apparent chemical shifts as the pH was increased towards 7.5, it was clear that the declining concentrations of soluble protein played a significant part in obtaining the poor data. However, amongst the fewer peaks seen at pH 7.5 there was also more crowding in comparison to data recorded at pH 6.0. This directed towards some signal overlap and a slightly less structured SilE-Ag^+ complex at pH 7.5. Moreover, even though the chemical shifts were more dispersed at pH 6.0, the spectrum did not correlate with the data obtained from the one-month old SilE-Ag^+ in Figure 5.13.

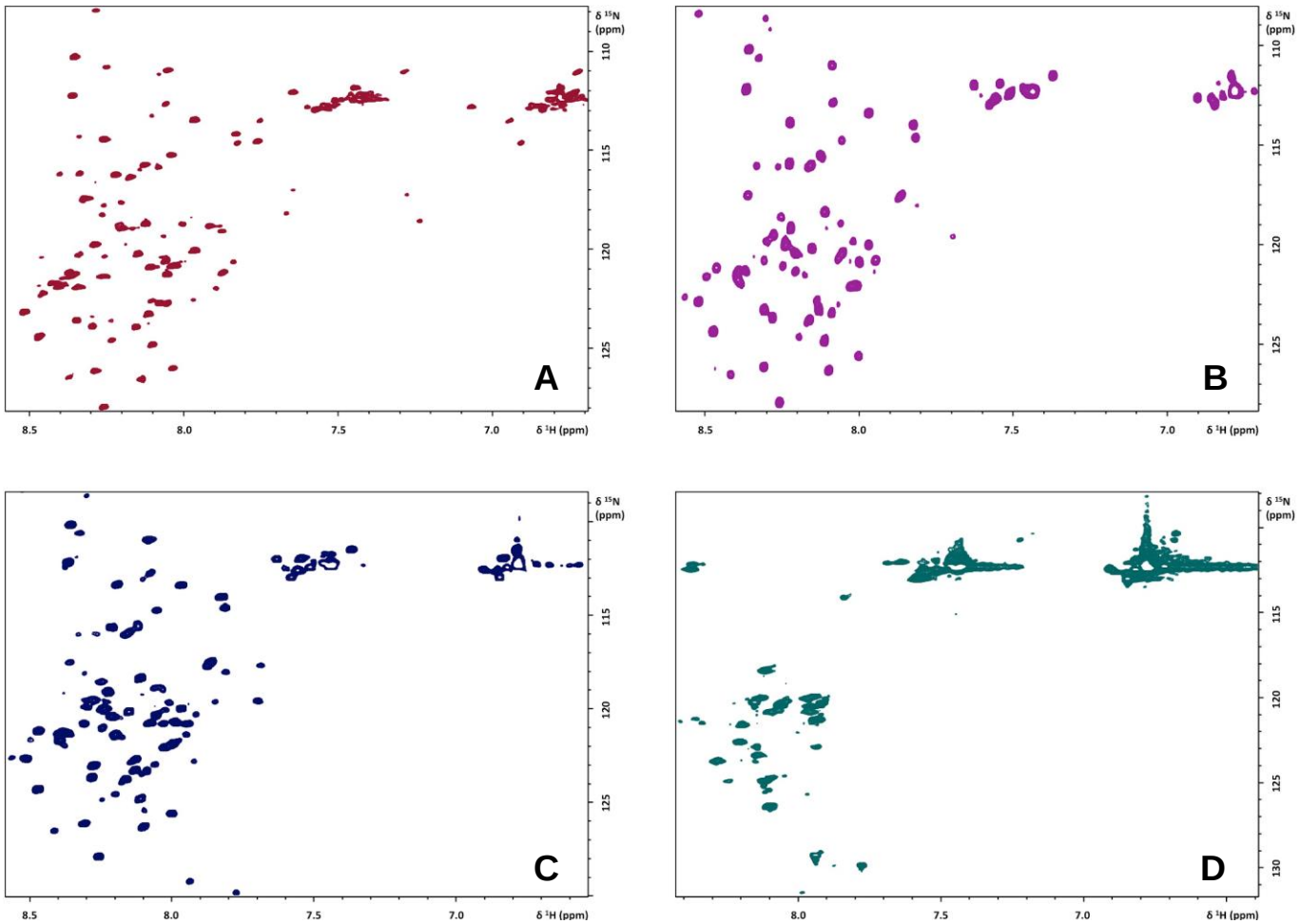


Figure 5. 16 Ag^{+}-SilE ^{1}H-^{15}N HSQC pH titration

^{1}H-^{15}N HSQC data obtained from 350 µM ^{15}N-labelled protein, in 10 mM HEPES, 20 mM NaF, recorded at 25 ^{o}C at pH **A.** 6.0, **B.** 6.5, **C.** 7.0 and **D.** 7.5. A large reduction in the number of peaks is seen as the pH was increased towards 7.5. Within this set of data, the peaks are best resolved at pH 6.0.

5.3 Summary and conclusions

In this chapter, previously proposed SilE metal ligand binding and folding studies were reviewed. The Ag^+-binding experiments were planned in terms of the choice of buffer and ionic strength, to replicate an appropriate biological environment in which both protein and ligand could remain stable and available, respectively. Consequently, 10 mM HEPES and 20 mM NaF were decided for use during metal ion binding and binding-induced folding experiments.

Silver *et al.*, (1999) previously used inductive-coupled plasma (ICP) analysis to quantify SilE:Ag^+-binding as 1:5.5 at biological pH. However, as the molecules are completely atomised, ICP analysis only provides an average number of ions bound per molecule. Therefore, ESI-MS measured the maximum number of Ag^+ that SilE can bind, which was found to be eight ions per molecule (Figure 5.2). This lies between the formerly suggested numbers of five and 10 Ag^+ ions binding to one SilE molecule (Silver *et al.*, 1999; Mirolo *et al.*, 2013). Nevertheless, the protein seemed to most strongly bind the first five to six Ag^+.

The metal binding stoichiometries were also reflected in the CD folding, where six Ag^+ appeared to be the optimum for full folding of SilE (Figure 5.6), with MS suggesting that a further two Ag^+ were capable of being bound by the folded protein. Consistent with the JPred 4 results of secondary structure prediction (Figure 5.3), CD experiments recognised that SilE cooperatively folded to a predominantly α-helical structure (54 % α-helicity), upon binding to metal ligand Ag^+ (Figures 5.4, 5.5 and Figure

5.8). Finally, the superior folding and binding of Ag^+ in comparison to the other divalent ions suggested SilEs specificity for Ag^+.

Where CD provided information on the structural content of SilE at the secondary level, SAXS and NMR were used to probe protein conformation at the tertiary level. As a result of photo-reduction, the scatter of photons from the irradiated metal ions in SAXS showed increasing electrondensity in the buffer recordings. This caused a problem in acquiring useful data. Whilst dialysing against more metal-ligand free buffer could be a method to lower the electron densities; adding a capillary 'superclean' step may prove to be more beneficial to obtain more valuable SAXS data in the future. Fast and efficient capillary cleaning, as performed at the ESRF BM29 in Grenoble, could help to avoid cross-contamination and ensure reproducibility of measurements (Round *et al.*, 2015).

There were distinct differences in the chemical shifts between the *apo*- and *holo*-SilE spectra when observed by NMR spectroscopy (Figures 4.10 and 5.13). The Ag^+-bound spectrum showed more peak dispersion that confirmed structural formation induced by metal-ligand binding. Nevertheless, using NMR for the purpose of spectral assignment and structure determination, at a molecular level, requires some optimisation. Section 5.2.7.3 explained how producing high quality spectral data at a physiological pH proved to be difficult. Defining the conformational ensemble of IDPs is a fundamental problem when attempting their structural characterisation (Konrat, 2014). They can also form dynamically disordered complexes when they interact with a target (personal conversation with Professor Peter Wright at the GRC on IDPs, 2014). Consequently, one

possibility of being unable to produce high quality spectra like the one in Figure 5.13 may have been due to insufficient amounts of Ag^+ being present and not allowing the protein to reach its 'fully' folded state. However, this seems unlikely as extensive dialysis was carried out. Secondly, the age of the sample in Figure 5.13 (approximately one month old) was significantly older than any of the other samples. Therefore, as a disordered protein originally, does SilE move and settle over time as it binds to Ag^+? Hence, the broader peaks seen otherwise could have been an artefact of a range of dynamic conformations that appear during the initial stages of Ag^+-binding? The final consideration is the decrease in sample pH upon the addition of Ag^+. Figure 5.16 showed that as the sample pH approached the pI of the protein, precipitation occurred. Therefore, in order to carry out data collection at biological pH and to avoid the precipitation problem, the sample could be prepared at a higher pH prior to Ag^+-dialysis. Consequently, more optimisation would be required to get high quality data that could lead towards assigning and defining a structure for SilE.

CHAPTER 6

Characterisation of Silver Binding to the SilE Protein

6.1 Introduction

Chapter 5 explored the Ag^+-ligand binding of wild-type SilE (residues 21-143, referred to as WT^{SilE} in this chapter) by ESI-MS (electrospray ionisation mass spectrometry), far-UV CD (far-ultraviolet circular dichroism) and NMR (nuclear magnetic resonance). The conclusion of these studies was that protein folding is coupled to binding of Ag^+.

The traditional view of amino-acid residues involved in preferential metal-binding ligands is that they usually include polar or electrically charged side chains. This would include aspartate, asparagine, cysteine, glutamate, glutamine, histidine, methionine, serine, threonine, and tyrosine (Yamashita *et al.*, 1990). However, histidine, methionine and cysteine residues are most commonly implicated in metal binding. Additionally, Silver, (2003) and Mirolo *et al.*, (2013) both proposed a model in which five Ag^+ ions were believed to be bound by ten histidine residues within SilE.

In order to understand the mechanism of WT^{SilE} Ag^+-binding and induced folding events in more detail, this chapter describes the site directed mutagenesis of selected residues and their contribution to folding and binding; studied using a combination of circular dichroism (CD) and ICP (inductively-coupled plasma) -MS.

Note: Throughout this chapter WT^{SilE} refers to the protein without its periplasmic signal sequence (residues 1-20), which was produced as $SilE_{21-143F}$.

6.2 Results and discussion

6.2.1 Ag^+-binding site predictions

Ligand molecule interactions are essential for many proteins to perform their correct biological function. In order to functionally characterise a small ligand-binding protein such as the putative Ag^+-binding SilE, it is important to identify the amino acid residues that are involved in the specific interaction. A large number of bioinformatics tools can aid in predicting ligand-binding sites. Two methods are useful for protein evaluation-based binding-site predictions. The first method of *in silico* based binding-site prediction usually requires a 3D (three-dimensional) structure (either experimentally obtained or computationally modelled using the primary sequence of the protein). Web servers use a number of methods in this way to predict binding-sites on the basis of various site-distinguishing properties. They exploit one or more of four types of properties (evolutionary, geometric, energetic, and statistical) in order to distinguish the binding site from other parts of the protein surface (Xie and Hwang, 2015). The second method uses sequence homology and alignments to assess the degree of sequence conservation between the protein of interest and homologous proteins. Accordingly, significant primary sequence features of SilE were analysed to predict the Ag^+-binding sites.

6.2.1.1 *in silico* binding site prediction

The site-distinguishing properties used for ligand-binding site predictions can be classified into the following categories: (i) template-based - utilising homologous structures with known binding sites, (ii) geometry-based – where site binding pockets are identified through computational geometric analysis, (iii) energy-based – where imaginary ligands are probed and interaction energies are computed and calculated to provide likely interaction events and (iv) propensity-based – in which a certain property shows significant preference for particular ligand-binding sites. Combinations of the methods mentioned are also sometimes used in the form of meta-servers (Xie and Hwang, 2015).

The previous set of Critical Assessment of Protein Structure prediction (CASP) experiments, evaluated the accuracy of computational methods in predicting binding residues. This was done within the 'function prediction' (prediction of binding sites, FN) category of CASP10 (Gallo Cassarino *et al.*, 2014). However, due to the recognition of inappropriate assessment, the future evaluation of ligand-binding site predictions servers was to be carried out continuously using the Continuous Automated Model EvaluatiOn (CAMEO - http://www.cameo3d.org/). CAMEO uses newly released structures from the PDB to base its predictions. The benefits of continuous evaluation are said to allow developers to constantly monitor the performance of their servers and add to the servers' development processes.

The current weekly performance of ligand binding site prediction experiments in CAMEO ranks the meta-server COACH (COnsensus ApproACH - for Protein-ligand binding site prediction) as the most accurate

ligand-binding predictor. With an average accuracy of 76 %, COACH has been ranked as the number one predictor for the past year. Consequently, COACH was used to predict the Ag^{+}-binding sites in SilE (Yang *et al.*, 2013a; Yang *et al.*, 2013b). Following the input of the entire primary sequence of SilE, COACH used MUSTER Online (MUlti-Sources ThreadER) (Wu and Zhang, 2008) to generate a 3D model (Figure 6.1-A) which was fed into the COACH pipeline for ligand-binding site prediction analysis. MUSTER is an online protein structure-threading (fold-recognition) algorithm that identifies template structures in the PDB library, which have similar folds to those of the target protein and makes a correct alignment between the target sequence and the template structure. The server then produces a model with the help of the top five template proteins and the alignments using MODELLER (Webb and Sali, 2014).

Due to insufficient homologous structures being identified, with a threading alignment z-score of 4.770 (i.e. <7.5) and a PDB sequence identity cut-off value of 61 %, MUSTER defined SilE as a 'hard target' for structure modelling. Nevertheless, the modelling was still performed and was based on the template metal binding protein Nccx (RCSB-PDB ID: 4CLV) (Ziani *et al.*, 2014). From the given structure of SilE, shown in Figure 6.1-A, COACH generated complementary ligand-binding site predictions primarily using two comparative methods – one based on binding-specific substructure comparison (TM-SITE) and another on sequence profile alignment (S-SITE). The methods successfully recognised binding residues, which combined with other methods such as COFACTOR, FINDSITE and ConCavity generated the final best ligand-

binding site predictions for SilE. Figure 6.1-B highlights the residues in SilE that COACH predicted to be Cu^{2+}, Ni^{2+} and/or Zn^{2+} ligand-binders. These were: histidine residues - H27, H38, H62, H69, H80, H87, H118 and H129, methionines – M72, M83 and M90, glutamine amino acids – Q34 and Q41 and finally glutamate (E) 119.

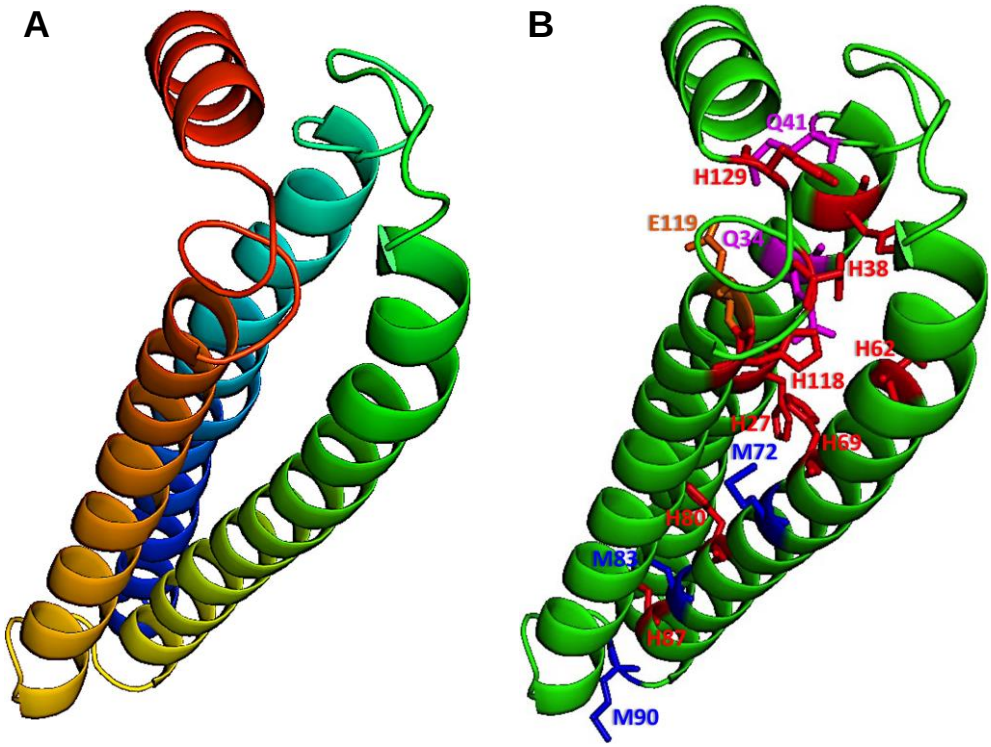


Figure 6. 1 COACH protein model based ligand binding-site predictions

A. The MUSTER (Wu and Zhang, 2008) template-based model of SilE, generated using MODELLER (Webb and Sali, 2014), z-score=4.770 (cut-off=6.1). The template used was the X-ray structure of the metal-binding NccX protein from *Cupriavidus metallidurans* (RCSB-PDB ID: 4CLV) (Ziani *et al.*, 2014). **B.** The metal ligand Cu^{2+}, Ni^{2+} and Zn^{2+} binding residues predicted by COACH are those highlighted: (histidines – red) H27, H38, H62, H69, H80, H87, H118, H129, (methionines – blue) M72, M83, M90, (glutamines – pink) Q34 and Q41 and E119 (orange). Both of the models were edited using MacPyMOL (Schrödinger, 2015).

6.2.1.2 Sequence homology based binding site prediction

A trend in the over-abundance of particular residues in a protein type (such as a metal-binding protein) shows their importance for the recognition of certain ligand types. So to begin the measurement of this tendency, sequence alignment of native SilE and its Cu-binding homologue PcoE (excluding their periplasm exporting leader sequence, residues 1-20) was carried out using the general multi-purpose primary sequence alignment program for proteins Clustal Omega (McWilliam *et al.*, 2013) (Table 2.5, Chapter 2). This is shown in Figure 6.2-A. Amino acid sequences were then coloured according to the 'Percentage Identity' between the two proteins in Jalview (Waterhouse *et al.*, 2009). The calculated sequence identity between the two homologue proteins was 48% (Zimmermann *et al.*, 2012). SilE and PcoE are rich in potential metal ligand-binding histidines (ten each) and methionines (eleven and fifteen, respectively). The primary sequence identity shows that the position of the ten histidine residues of the two proteins and eight methionine residues (of the eleven in SilE and fifteen in PcoE) are completely conserved (Figure 6.2-A), suggesting a key role in metal binding.

Furthermore, the SilE sequence was probed for amino acid repeats (AARs). AARs have functional and evolutionary roles making them abundant within protein sequences. Identifying AARs is a useful step for further investigation of a proteins biological function and evolutionary mechanism. Figure 6.2-B features two sets of non-tandem, imperfect repeats that were recognised in the sequence of SilE (Luo and Nijveen, 2014). These have been illustrated in Figure 6.2-C as the two characteristic motifs,

MxxHxxxxxHxxMxx (motif 1) and HxxMxxxHxxMxx (motif 2), each repeated twice within the sequence have conserved histidines and methionines that constitute potential metal-binding motifs (Figure 6.2-B).

Additionally, histidine and methionine residues are common features in many metal-binding, disorder to α-helical structures (Todd *et al.*, 1991; Tanaka *et al.*, 2004; Monchy *et al.*, 2006; Breydo and Uversky, 2011; Ucisik *et al.*, 2013; Faller *et al.*, 2014). The sulphur atom of methionine residues and the imidazole ring in histidnes are both well-known metal-ligand coordinators (Yamashita *et al.*, 1990; Carrera *et al.*, 2004; Long *et al.*, 2010; Rubino *et al.*, 2011; Silva *et al.*, 2014). Thus, these conserved residues were considered as feasible primary candidates for Ag^{+}-binding in SilE. Therefore, nine of the ten histidine residues (H38, H62, H69, H80, H87, H111, H118, H129 and H136) in addition to four methionine residues (M72, M83, M90 and M108) became the subjects of site-directed mutagenesis based studies.

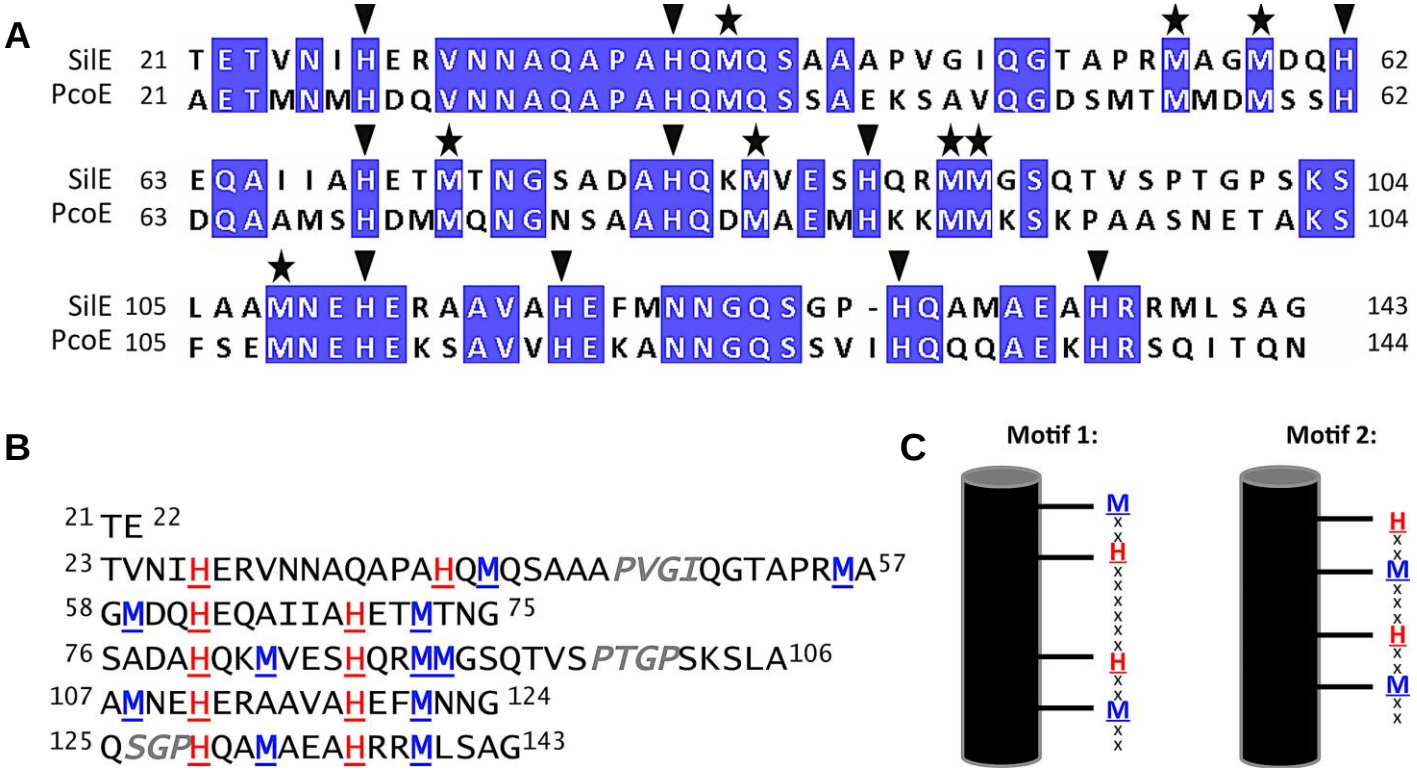


Figure 6. 2 **Protein sequence analysis and evaluations**

A. Sequence alignment of *E. coli* PcoE with SilE from *Salmonella* (lacking the leader sequence, residues 1–20). The sequences, aligned using Clustal Omega (McWilliam *et al.*, 2013) have been coloured using the 'Percentage Identity' colour-scheme in Jalview (Waterhouse *et al.*, 2009). The sequences displaying 48% identity, the 'Percentage Identity' colour-scheme has a threshold of 80% or more being conserved residues (purple) and anything below 40% as non-conserved (white), with the colour gradient between clarifying other less conserved residues. Conserved, aligned histidine and methionine residues have been highlighted above with either an inverted triangle or a star, respectively. **B.** Repeat sequence identification in the probable folded conformation of SilE. **C.** Two repeat sequence motifs (motif 1 and 2) are evident, which include the conserved histidine and methionine residues.

6.2.2 Mutagenesis of histidine and methionine residues

For the single and double histidine mutants, the commonly used method of systematic alanine scanning was implemented. Due to the small, non-bulky and non-polar nature of alanine, alanine-scanning mutagenesis is routinely used for exploring the structural and functional role of specific residues in a protein. Substitution with alanine eliminates all side-chain atoms beyond the β-carbon without altering the main-chain, protein secondary conformation (unlike glycine and proline) whilst remaining chemically inert (i.e. no direct involvement in ligand binding or catalysis) (Lefèvre *et al.*, 1997; Morrison and Weiss, 2001; Wan *et al.*, 2012). In this way, the roles of the particular mutated side-chains can be deduced (Weiss *et al.*, 2000; Van Petegem *et al.*, 2008). By substituting the hydrophilic histidine amino acid with a hydrophobic alanine residue, the main effect was the loss of the protonated imidazole ring and accordingly the removal of any potential contributions towards metal-ligand binding (Carrera *et al.*, 2004; Silva *et al.*, 2014).

For the slightly larger methionine, it was important to conserve the size of the residue, hence leucine was used as a substitute. The substitution of a methionine to a leucine residue removed the sulphur atom that could have potentially been involved in metal binding (Yamashita *et al.*, 1990; Long *et al.*, 2010).

6.2.2.1 PCR-based site-directed mutagenesis

Initial attempts at site-directed substitution mutagenesis were made (by Mr. Thomas Henry Miller, MSc, University of Nottingham and myself) to mutate histidine residues to alanine, using the Q5® Site-Directed Mutagenesis Kit. This method uses the Polymerase Chain Reaction (PCR) to produce site-specific mutagenesis. Two back-to-back primers are designed, one forward and one reverse in which one contains the desired mutation. These are then used to amplify the whole plasmid template. This method however, proved to be unsuccessful. Chapter 2, Section 2.2.4.4 describes and explains various conditions that were attempted to achieve success using this method. For example, the $SilE_{21\text{-}143F}$ plasmid template was guanine-cytosine (GC) rich in content which made it prone to deoxyribonucleic acid (DNA) secondary structure formation (such as hairpins) through hydrogen (H) –bonding. Dimethyl sulfoxide (DMSO) was added to the PCR with an aim to prevent secondary structure formation, as it interferes with the H-bonding (Mamedov *et al.*, 2008). Optimising magnesium chloride ($MgCl_2$) is important in PCR due to Taq DNA polymerase being a Mg^{2+}-dependent enzyme and the gamma phosphate of template DNA primers and dNTPs (nucleoside triphosphate containing deoxy) binding to the Mg^{2+} (Markoulatos *et al.*, 2002). So, extra $MgCl_2$ was added to improve polymerase activity and to generate more PCR product - unfortunately without any success. Finally, a range of lower annealing temperatures was tested, however this led to significant amounts of non-specific binding and incorrectly sized PCR DNA amplicons. Additionally, a change in the enzyme and reagents to Phusion® High-Fidelity DNA

Polymerase - as part of a PCR Master Mix with GC Buffer, still did not see any positive results.

On moving to the two-step PCR by overlap technique, alongside the new set of primers, the enzyme used again was Phusion® High-Fidelity DNA Polymerase with the GC Buffer. This implies that there could have been a problem with the primers initially. Nevertheless, this was highly unlikely as nine different primer sets were being used for the nine different His to Ala mutations. Where the template plasmid only had a 46 % GC content overall, one particular region contained a stretch of approximately 70 Gs and Cs which was only broken up by 3 adenosines (A). This was also near to other GC rich areas. As a result, the problem may have been insufficient quantities of DMSO that allowed melting-resistant secondary DNA structure formation, which in turn caused the polymerase to stall and hinder primer annealing (Bhagya *et al.*, 2013).

After failure of the initial method, PCR by overlap was attempted for all of the site-directed substitution mutations. This method only amplifies the gene of interest in place of the entire plasmid (Figure 6.3). The gene is amplified as two halves, with each half containing the mutation at one end and a restriction site at the other. The two halves overlap in the region of the mutation so that in a second round of PCR they can prime each other and extend to form the whole gene, which can be further amplified. The PCR product gene can then be used to reinsert the gene into the vector using the flanking restriction sites (Heckman and Pease, 2007).

Compared to usual primer design, a few further factors were taken into consideration when designing the primers for this method. First of all

forward primer ‘a’ and reverse primer ‘d’, shown in Figure 6.3, needed to contain restriction sites that are in the plasmid. As, these sites were useful for reinserting the gene into the vector following the two rounds of PCR. Primers ‘a’ and ‘d’ also consisted of different types of restriction sites so that the gene could be cloned in the correct orientation and prevented the digested vector from self-ligating. In addition, these two primers were required to have a slightly lengthier sequence prior to the restriction site at their five prime (5’) end, i.e. about sixteen base pairs. This provided regions on either side of the recognition site for restriction enzymes to bind and digest efficiently. It was also important for the primers to have similar annealing temperatures (within 5 ^{o}C of each other) for positive PCR in the successive round. For this reason, when designing reverse primer ‘b’ and forward primer ‘c’ the annealing temperatures needed to be reflected, so that the primers could be productively used with primers ‘a’ or ‘d’ during the first round of PCR (Figure 6.3). Primers ‘b’ and ‘c’ also needed to contain the desired mutation and were designed to have approximately 20-30 bases of correct sequence with at least ten correct bases each side of the mutation that allowed for efficient annealing to the template DNA. A list of the primers that were designed and used for the production of these His to Ala and Met to Leu mutants can be found in Table 2.3, Chapter 2 (Materials and Methods).

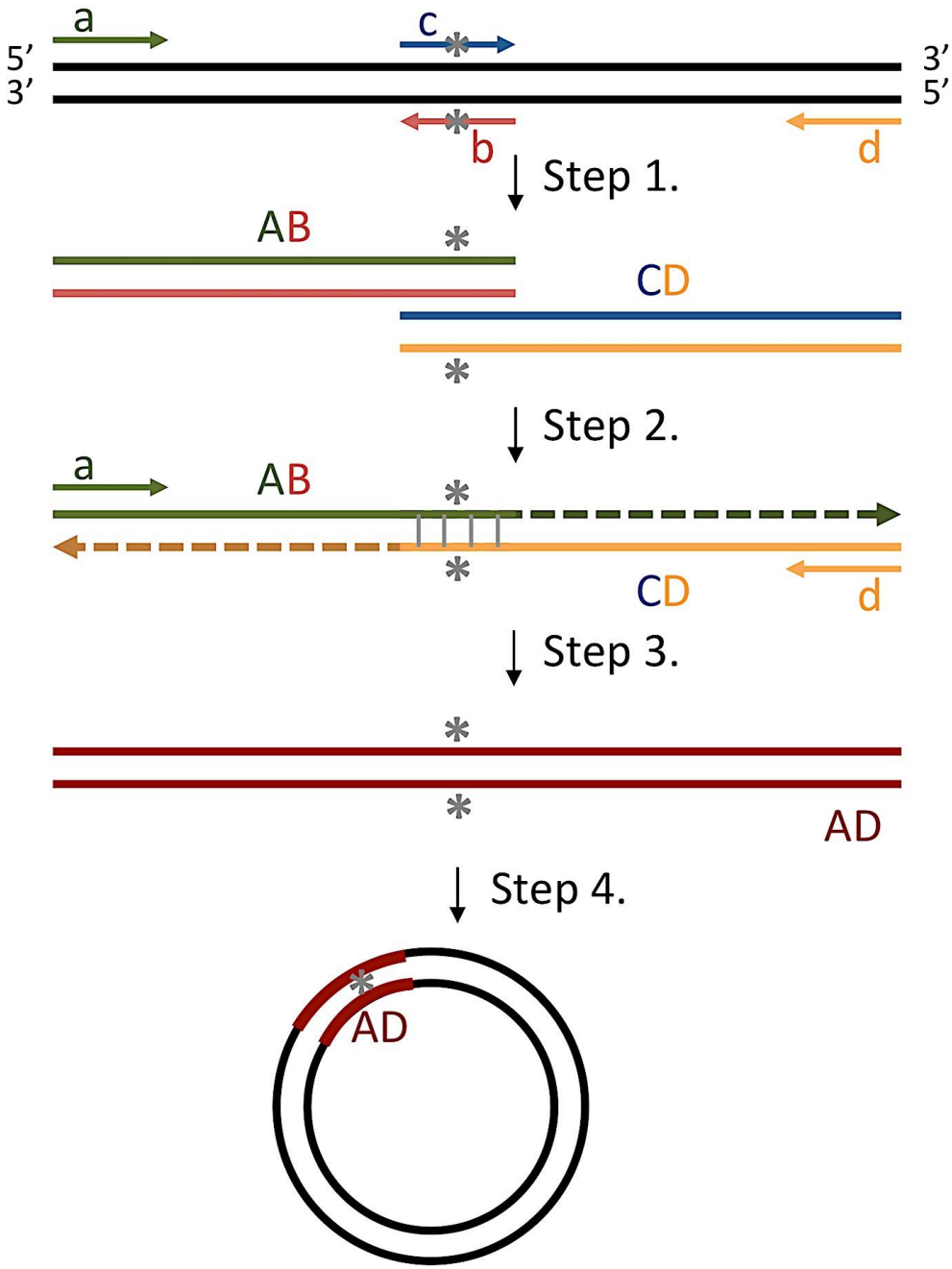


Figure 6. 3 **Overview of site-directed mutagenic PCR by overlap**

Primers a (green) and d (orange) contain restriction sites. Primers b (red) and c (blue) contain the desired mutation.

Step 1. - Two initial PCR amplifications of fragments with overlap are carried out, one with primers a and b, the other with c and d.
Step 2. - The overlapping products produced are used in a second reaction in which they prime each other, extend to form the whole gene.
Step 3. Then a and d are added in order to amplify the gene.
Step 4. Gene with desired mutation is then reinserted and ligated into the vector.

(Image has been recreated using Heckman and Pease, 2007 and Luo *et al.*, 2013 as references).

This method was extremely successful in amplifying the desired products in both rounds of PCR by overlap. Figure 6.4 shows these PCR gene products on an agarose gel. All His to Ala (H38A, H62A, H69A, H80A, H87A, H111A, H118A, H129A and H136A), Met to Leu (M72L, M83L, M90L and M108L) and double His mutants (H80A/H87A, H80A/H111A and H87A/H11A) were successfully made and confirmed to be correct by sequencing (Chapter 2, Section 2.2.4.2).

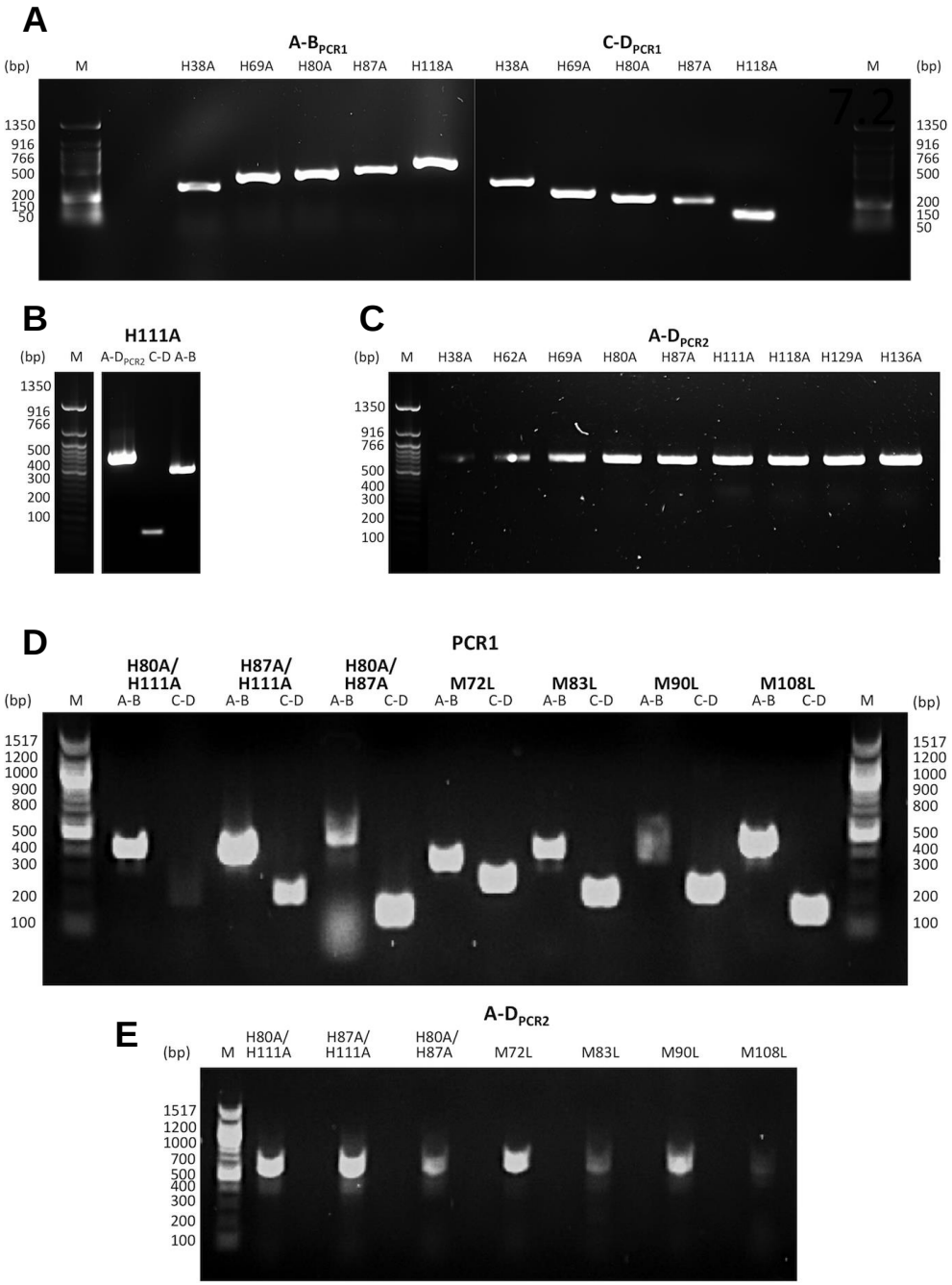


Figure 6. 4 **SilE site-directed mutagenic PCR by overlap**

A. Shows the products for both halves of the first round of PCR (forming fragments AB and CD) for five mutants – H38A, H69A, H80A, H87A and H118A. **B.** Displays the first and second rounds of PCR for the H111A mutant. **C.** Shows the final nine single His to Ala mutant gene products at 596 bp (H38A, H62A, H69A, H80A, H87A, H111A, H118A, H129A and H136A) from PCR2. **D.** Presents the gene products from both intersects of the PCR1s of the double His to Ala and Met to Leu mutants (H80A/H87A, H80A/H111A, H87A/H11A, M72L, M83L, M90L and M108L), with **E.** displaying the final mutant genes for those in D at 596 bp.

6.2.2.2 Overexpression of the SilE histidine and methionine mutant proteins

All of the mutants were expressed and purified in the same manner as described for the wild-type SilE (WT^{SilE}) protein (Chapter 3, Section 3.5). The only difference was a decrease in the concentration of imidazole required for the final elution step of the His to Ala mutants, post HRV-3C protease treatment. The expression and solubility levels of all of the protein mutants were comparable to that of WT^{SilE}, indicating that they were stable.

Figure 6.5 shows the final products of the pure mutant proteins following overexpression and purification. The protein bands appeared to be clean with no signs of degradation or contamination. ESI-MS was used to confirm that the mutants were at the expected molecular weight. They all had a similar molecular weight to WT^{SilE} (13,268-13,271 Da); single His to Ala mutants - 13,198-13,199 Da, double His to Ala mutants - 13,135-13,137 Da and Met to Leu mutants - 13,364-13,365 Da.

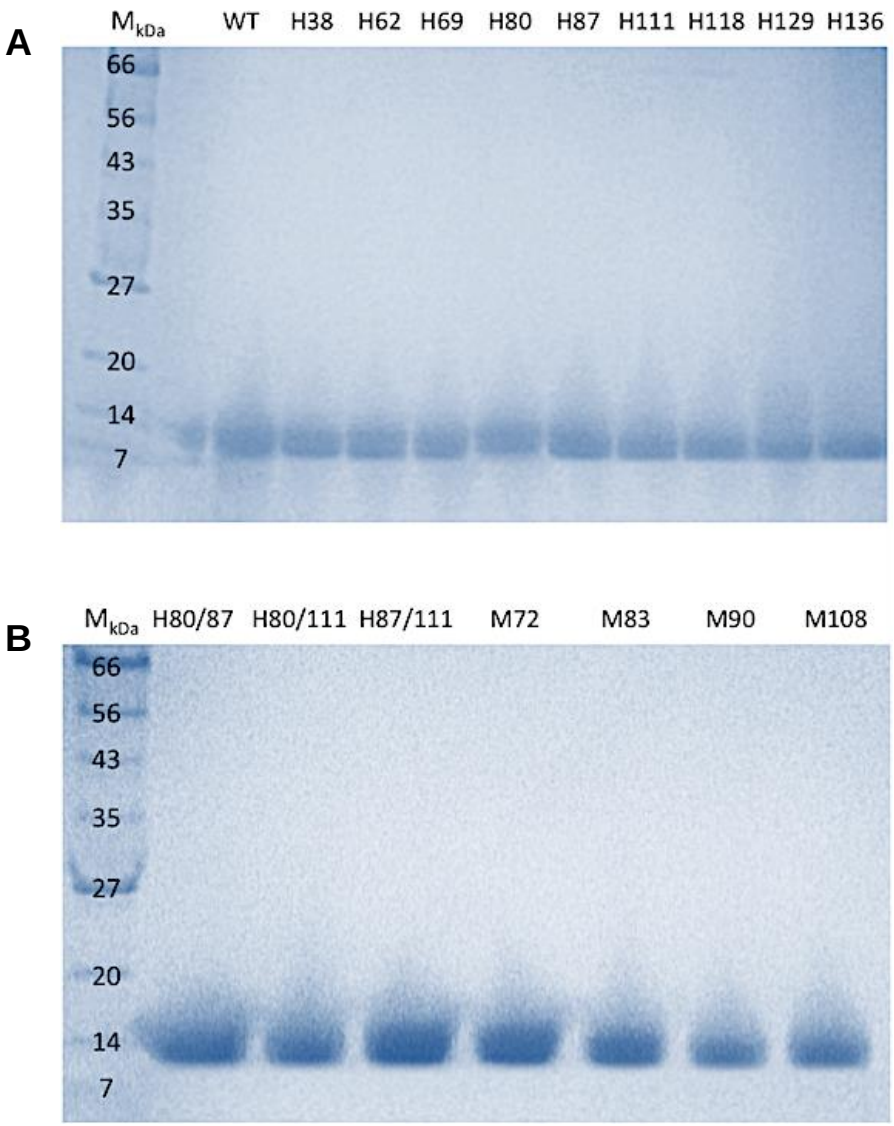


Figure 6. 5 Mutant SilE SDS-PAGE

SDS-PAGE analysis of pure SilE on a 4-20 % gradient gel (Chapter 2, Section 2.2.2.2). **A.** Single His to Ala mutants and **B.** Double His to Ala and Met to Leu mutants (expected molecular weight ~13.2-3 kDa).

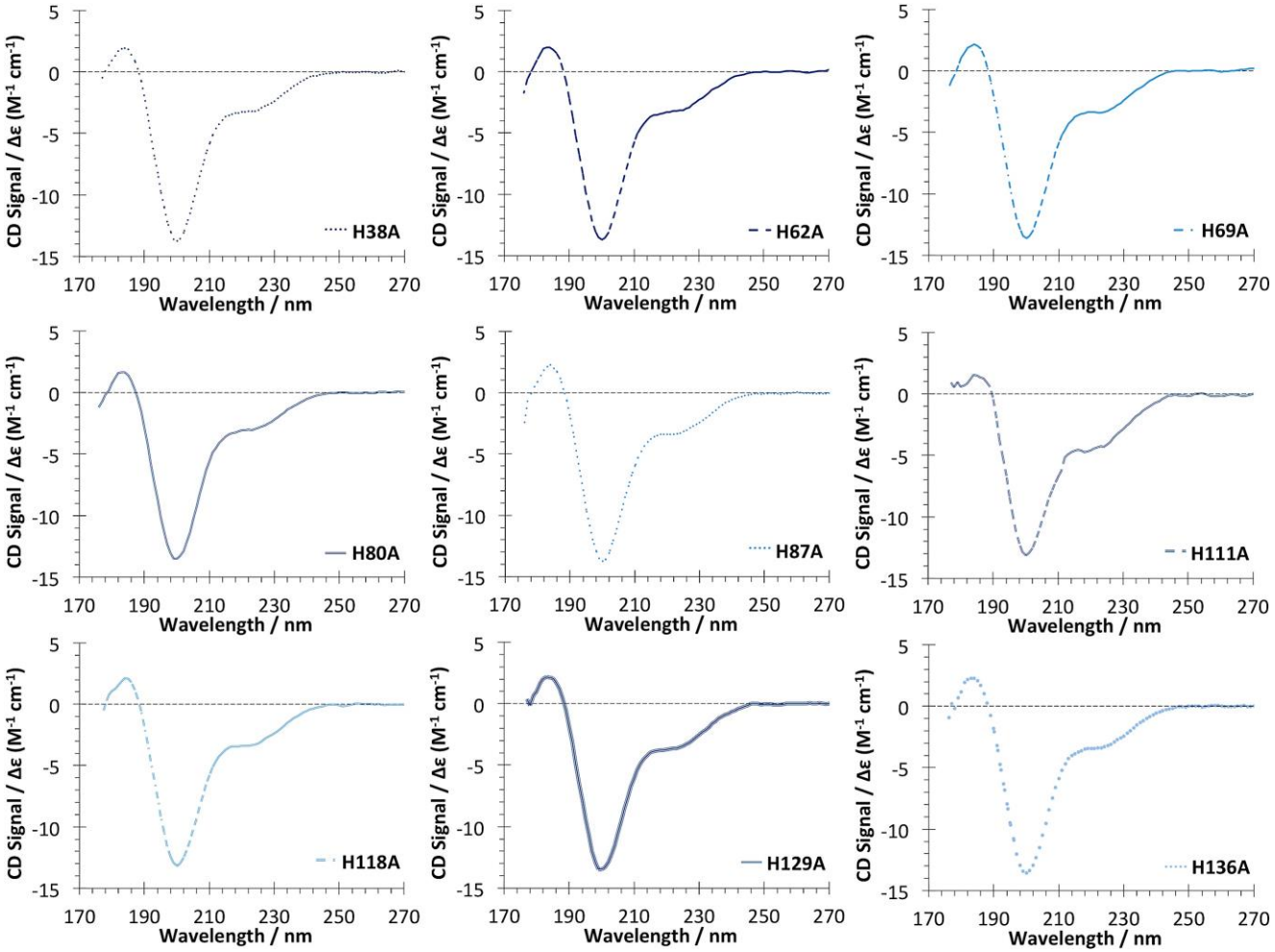
6.2.3 Histidine and methionine residues are involved in Ag^+-binding

In order to investigate the roles of conserved histidine and methionine residues in Ag^+-binding and protein folding, initially nine histidines were mutated to alanine residues (H38A, H62A, H69A, H80A, H87A, H111A, H118A, H129A and H136A) (Section 6.2.2).

6.2.3.1 Ag^+-induced folding affects of the SilE mutants

The secondary structure content of all of these His to Ala mutant SilE proteins was measured at a concentration of 75 μM (in 10 mM HEPES, 20 mM NaF, pH 7.5) using far-ultraviolet (UV) circular dichroism (CD) (Chapter 2, Section 2.2.6.1) both in their *apo*-form (Figure 6.6) and in the presence of six Ag^+ molar equivalents (Figure 6.7). Figure 6.6 shows that all of the *apo*-SilE mutants were disordered as they gave an identical CD profile to that of the *apo*-WT^{SilE} protein (Figure 6.6-B).

A


CD Signal / Δε (M-1 cm-1)
Wavelength / nm
H38A
H62A
H69A
H80A
H87A
H111A
H118A
H129A
H136A

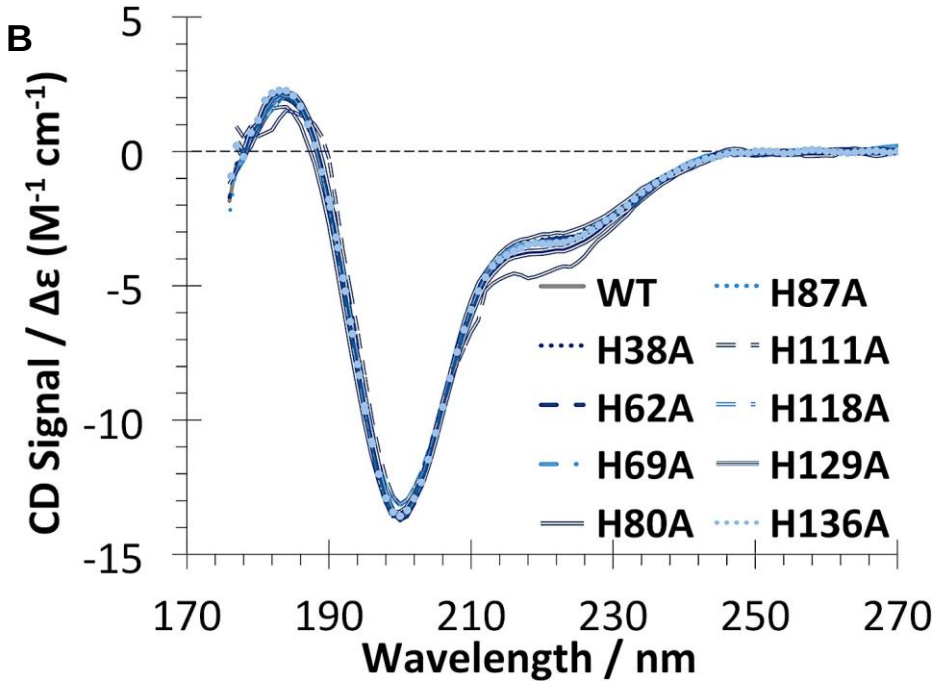


Figure 6. 6 SilE His mutation effects on protein secondary structure

A. Far-UV CD of the nine *apo*-SilE His to Ala mutants, at 75 μM in 10 mM HEPES, 20 mM NaF at pH 7.5. **B.** An overlay of all of the His to Ala mutants, with *apo*-WT^{SilE} (grey). All of the His to Ala mutants show the same CD profile as the native protein (grey), in their *apo*-form. With a strong minimum around 200 nm, they are all indicative of significant disorder.

In Figure 6.7 H38A, H62A, H69A, H118A, H129A and H136A showed little change in protein backbone conformation and gave only a slight decrease in α-helical structure (Table 6.1). In contrast, the protein mutants H80A, H87A and H111A exhibited large reductions in α-helical content in comparison to *holo*-native SilE (Figure 6.7 Table 6.1), implying that these residues are important for Ag^{+}-induced folding. H80A and H87A showed a small reduction in signal intensities. Whereas H111A exhibited a larger reduction in signal intensity: at 190 nm from 13.463 *holo*-WT^{SilE} to 4.497 M^{-1} cm^{-1}, at 207 nm -10.368 *holo*-WT^{SilE} to -6.678 M^{-1} cm^{-1} and at 221 nm from -8.981 *holo*-WT^{SilE} to 4.326 M^{-1} cm^{-1}. Moreover, peak shifts

were seen for H111A from 190 to 187 nm in the positive peak and from 207 to 204 nm and 221 to 223 nm in the negative peak (Figure 6.7). This change in the H111A CD spectra was consistent with a decrease in α-helicity, which exhibited only around 60% of the *holo*-WT^{SilE} secondary structure content (Table 6.1), calculated using the DicroWeb interface analysis programme CONTINLL (Provencher and Glockner, 1981; Van Stokkum *et al.*, 1990; Sreerama and Woody, 2000; Lobley *et al.*, 2002) (Chapter 2, Section 2.2.6.1).

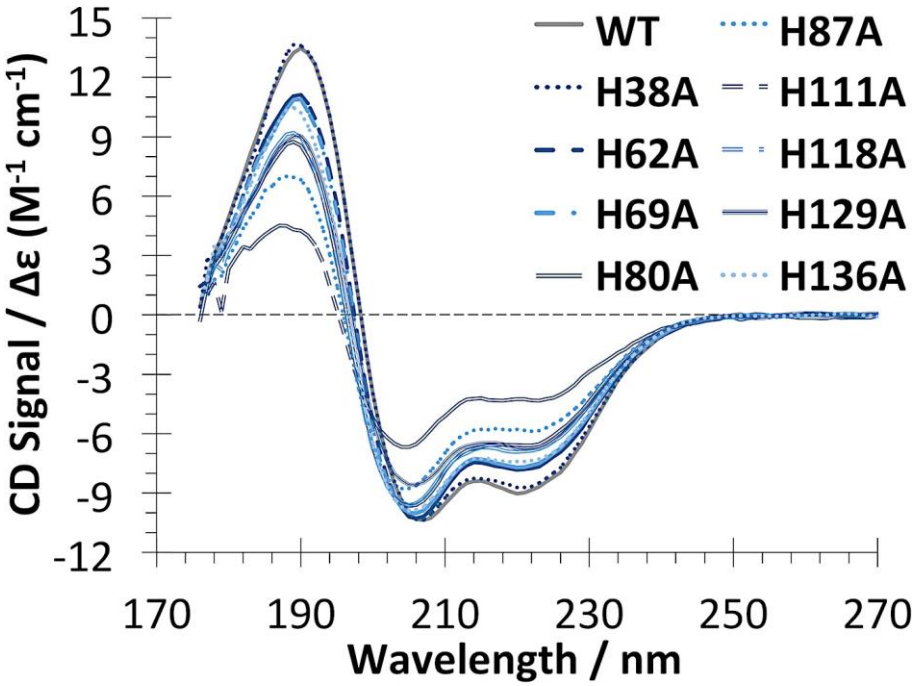


Figure 6. 7 SilE His mutants affect protein Ag^+-induced folding

Far-UV CD of *holo*-WT^{SilE} and *holo*- SilE mutants (His to Ala), with six Ag^+ molar equivalents (at a concentration of 75 μM, in 10 mM HEPES, 20 mM NaF at pH 7.5). The H111A SilE mutant (navy, dashed, double line) shows the greatest difference in comparison to WT^{SilE} (grey).

Table 6. 1: Percentage of mutant SilE secondary structure calculated from far-UV CD data

CONTINLL was used via DichroWeb to calculate the percentage of secondary structure content in WT^{SilE} and all of the mutant SilE proteins, in the presence of six Ag^+ molar equivalents. The NRMSE values (< 0.1) confirm the accuracy of the secondary structure calculations (Provencher and Glockner, 1981; Van Stokkum *et al.*, 1990; Sreerama and Woody, 2000; Lobley *et al.*, 2002) (Chapter 4, Section 4.2.4.2).

SilE Protein: 6 Ag^+	α-Helical Content (%)	β-Sheet Content (%)	Turn Content (%)	Unordered Content (%)	NRMSE Value (to 3 d.p)
WT	54	3	18	25	0.083
H38A	53	4	18	25	0.078
H62A	49	3	21	27	0.071
H69A	48	3	21	28	0.077
H80A	42	4	23	31	0.065
H87A	38	4	25	33	0.074
H111A	30	6	30	34	0.080
H118A	40	4	23	33	0.050
H129A	45	4	21	30	0.069
H136A	46	4	20	31	0.078
H80A/H87A	36	3	31	30	0.063
H80A/H111A	32	4	29	35	0.078
H87A/H111A	27	4	32	37	0.071
M72L	51	3	20	26	0.067
M83L	48	3	21	28	0.051
M90L	40	4	24	32	0.039
M108L	36	3	29	32	0.065

Based on these data, three double mutant proteins H80A/H87A, H80A/H111A and H87A/H111A and four methionine mutants (to the amino acid leucine) M72L, M83L, M90L and M108L were made (Section 6.2.2). Far-UV CD experiments were conducted to measure the secondary structure content of these mutants (75 μM, in 10 mM HEPES, 20 mM NaF, pH 7.5) in their *apo*-form (Figure 6.8) and in the presence of six Ag^{+} molar equivalents (Figure 6.9). As shown in Figure 6.8, no mutation-induced effects were present in any of the *apo*-SilE mutants.

Far-UV CD spectra of the double-His to Ala mutant proteins (Figure 6.9-A) showed a decrease in secondary structure content greater than that found with the single histidine mutant proteins, with α-helical content declining to almost half that of WT^{SilE} (Table 6.1). The most affected double mutants were the H87A/H111A and H80A/H111A both containing the H111A mutation whereas the H80A/H87A was affected the least (Figure 6.9-A and Table 6.1). For H87A/H111A the positive peak shift from 190 to 187 nm was accompanied by a decrease in signal intensity to 3.227 M^{-1} cm^{-1} and the negative shifts – 207 to 203 nm and 221 to 223 nm were accompanied with reductions to -4.843 and -2.782 M^{-1} cm^{-1}, respectively (Figure 6.9-A), with α-helical content declining to 27%.

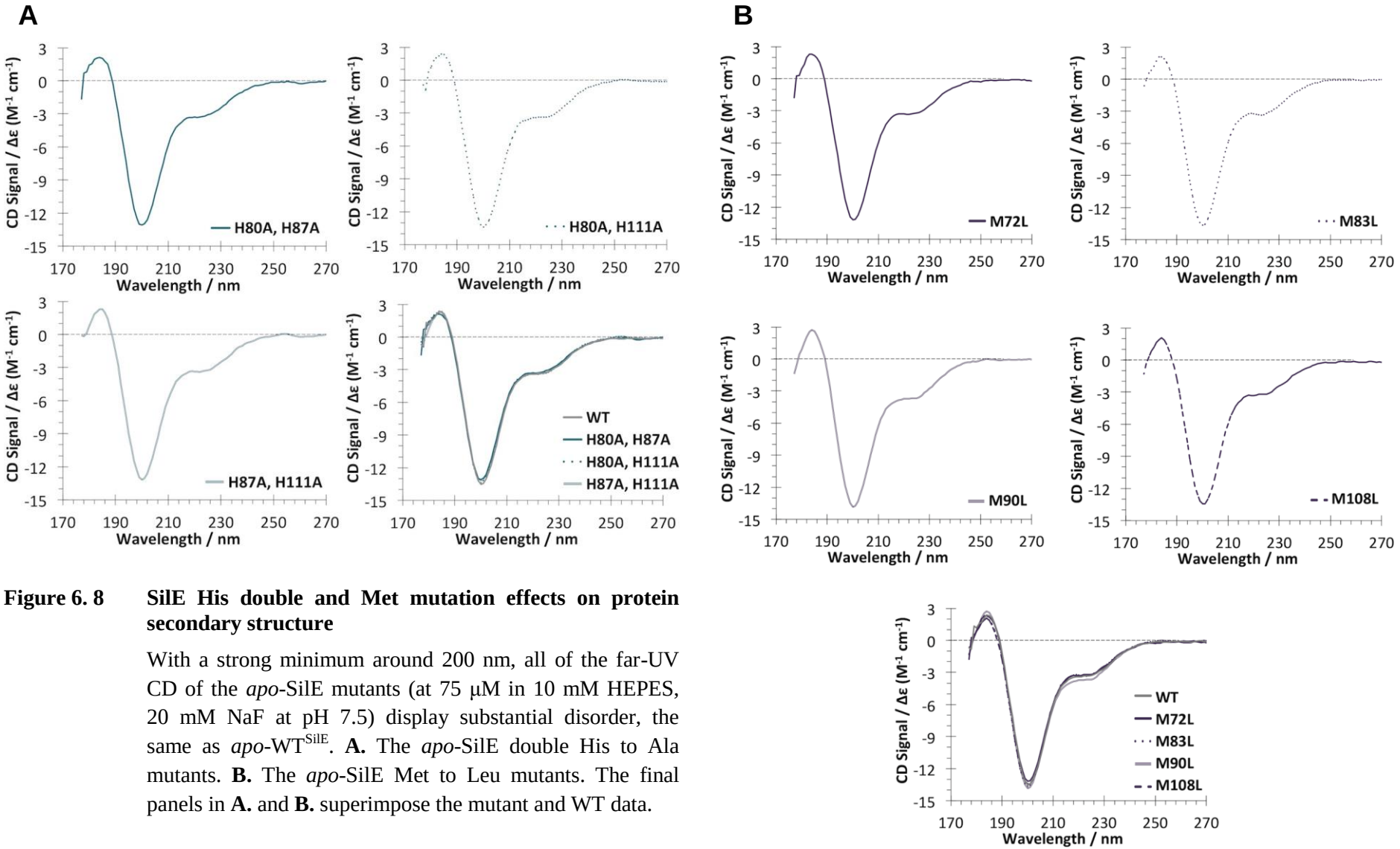


Figure 6. 8 **SilE His double and Met mutation effects on protein secondary structure**

With a strong minimum around 200 nm, all of the far-UV CD of the *apo*-SilE mutants (at 75 μM in 10 mM HEPES, 20 mM NaF at pH 7.5) display substantial disorder, the same as *apo*-WT^{SilE}. **A.** The *apo*-SilE double His to Ala mutants. **B.** The *apo*-SilE Met to Leu mutants. The final panels in **A.** and **B.** superimpose the mutant and WT data.

Similar measurements for the methionine mutants (Figure 6.9-B) found that the M108L mutant had a comparable decrease in secondary structure content to the H111A protein (Figure 6.7 and Table 6.1). In Figure 6.9-B, a positive shift to 187 nm was accompanied by a signal decrease to -6.590 M^{-1} cm^{-1} and the negative shifts to 204 nm and 223 nm were accompanied by lower signal intensities to 6.630 and 4.851 M^{-1} cm^{-1}, respectively. The other three methionine mutants (M90L, M83L and M72L) showed effects similar to those observed for the histidine mutants (compare Figure 6.7 and 6.9-B and consult Table 6.1).

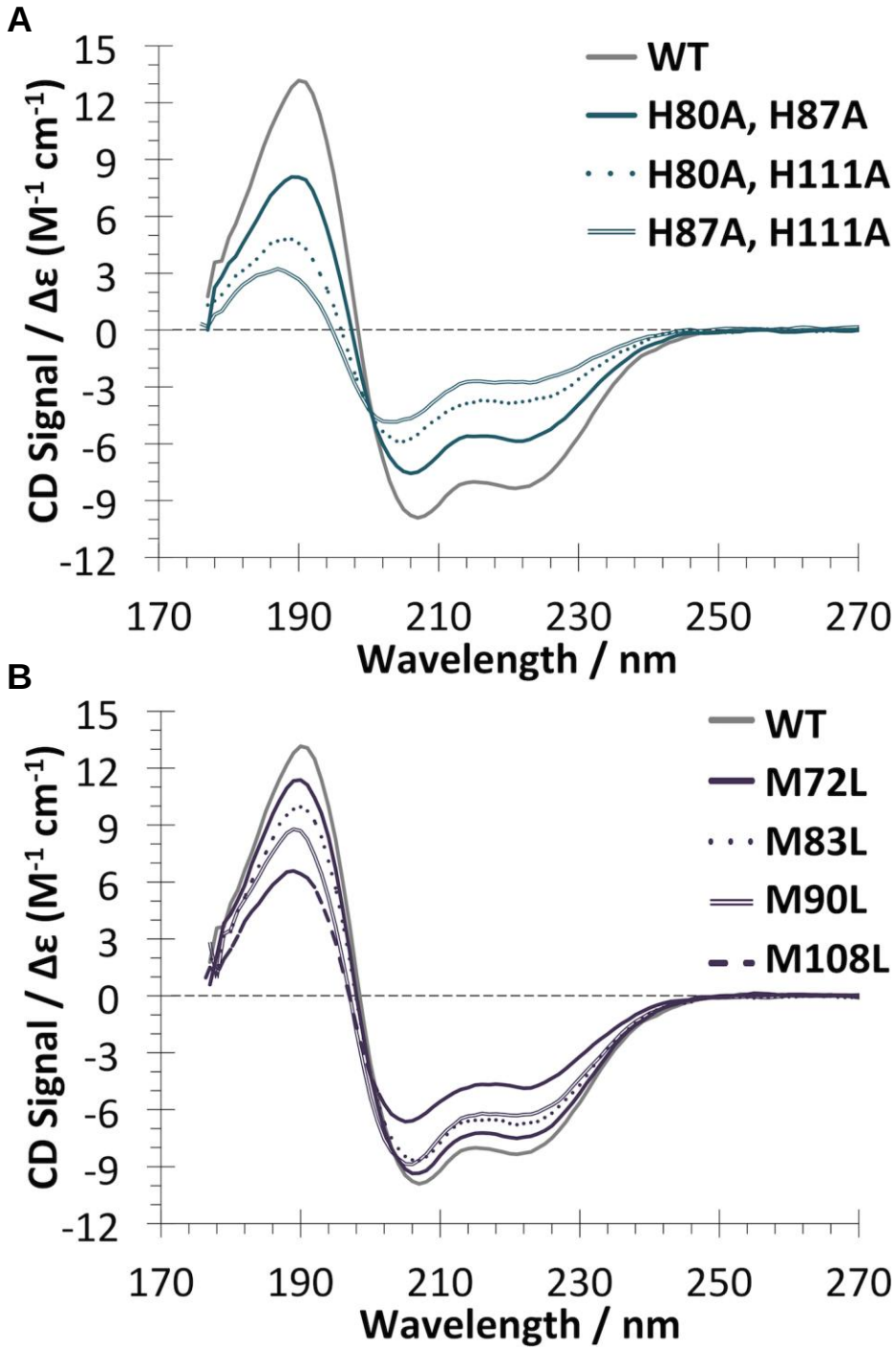


Figure 6. 9 **SilE double His and Met mutants affect protein Ag^+-induced folding**

Far-UV CD of WT *holo*-SilE and *holo*- SilE **A.** double His and **B.** Met mutants, with six Ag^+ molar equivalents (at a concentration of 75 μM, in 10 mM HEPES, 20 mM NaF at pH 7.5). The H87A/H111A mutant (**A**. - turquoise double line) shows the most significant difference in comparison to WT^{SilE} (grey). The Met mutants (**B.**) show similar profiles to the His mutants located nearby in the SilE sequence (Figure 6.7).

6.2.3.2 SilE mutant Ag^+-binding analyses by ICP-MS

In Chapter 5, Section 5.2.2, the discovery that WT^{SilE} could bind up to eight silver ions was made using electrospray ionisation mass spectrometry (Figure 5.2, Chapter 5). In an attempt to quantify SilE Ag^+-binding, Silver *et al.*, (1999) used inductively coupled plasma mass spectrometry (ICP-MS) and showed that SilE has a preference for and strong ability to bind 5.5 Ag^+ at pH 7.5. ICP-MS analysis combines the ionisation of atoms by high-temperature plasma (usually argon) with their detection and identification by mass spectrometry (Chapter 2, Section 2.2.6.2) (Thomas, 2001). The detection of ions by the spectrometer is element specific, so complex biomolecules are quantified by their known composition of elements. It is a highly sensitive technique in the range of ng L^{-1}.

ICP-MS was used to measure the number of Ag^+ ions bound to and retained by each of the SilE mutant proteins (conducted by Dr. Scott Young, ICP-Analytical Services (Sutton Bonington Campus, University of Nottingham) (Chapter 2, Section 2.6.2.2). As the molecules are completely atomised in this technique, ICP-MS only provided an average number of ions bound by SilE and not the maximum number that the protein could bind. The ICP-MS control samples, in which no Ag^+ was added to the protein, showed negligible detection of Ag^+. Figure 6.10 shows the average number of Ag^+ bound to each of the SilE proteins and demonstrates that WT^{SilE} was detected to bind an average of 6.7 Ag^+ (± 0.1) per SilE molecule. The data in Figure 6.10-A indicates that the single histidine mutants H62A, H69A, H80A, H87A, H111A, H118A and H129A all bound

on average one Ag^+ less than the *holo*-SilE. Two of the single mutants, H38A and H136A did not exhibit a clear reduction in the number of bound Ag^+ ions (Figure 6.10-A). All three double histidine mutants showed a reduction of two-bound Ag^+ ions (Figure 6.10-B) whereas all the methionine mutants showed a reduction of one bound Ag^+ (Figure 6.10-C).

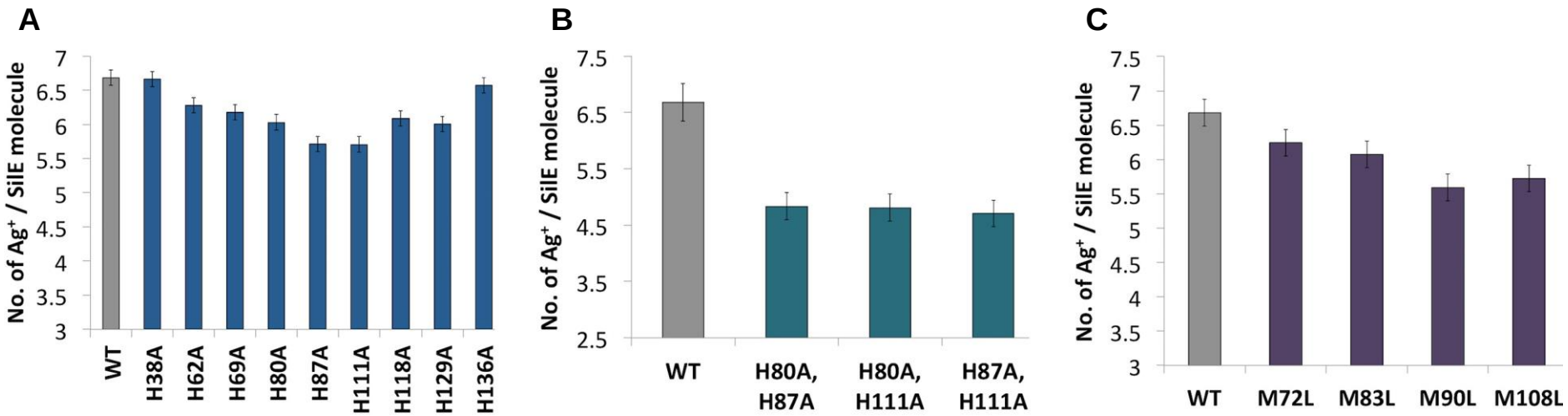


Figure 6. 10 ICP-MS Ag^+-binding analyses of SilE and its mutants

ICP-MS Ag^+-binding analyses of WT *holo*-SilE, **A.** histidine single, **B.** histidine double and **C.** methionine mutants at 50 nm, in 10 mM HEPES, 20 mM NaF, pH 7.5 in 1% HNO_3. Ag^+ levels are quantified relative to number of atoms per SilE protein molecule. Graphs show the mean of three independent replicates, from a single representative experiment, error bars represent standard error.

6.2.4 Two 'core-motifs' central to SilE likely form the nucleation site for Ag^+-induced folding

From the studies of the mutant SilE proteins (Section 6.2.3) it appeared that the H80A, H87A, H111A and M108L exhibited the largest folding defects upon Ag^+-binding. These residues are located centrally within the sequence in an apparent core comprising a motif 1 (residues A77-M90) and a motif 2 (residues E110-F120) (Figure 6.2). Given that the secondary structure content has a high sensitivity to mutation of these residues, it was proposed that they provide initial nucleation sites for Ag^+-induced folding. To test this hypothesis two truncated SilE polypeptides were engineered, one $SilE_{46-128}$ that preserved the core region (P46 to P128) but lacked the peripheral sequences at the N- and C-terminal regions, and a second $SilE_{21-98}$ (T21-P98) that lacked the C-terminal region and motif 2 from the putative nucleation core (Figure 6.11-A).

6.2.4.1 Production of the SilE truncation mutants

The truncations mutants Sil_{46-128} and $SilE_{21-98}$ were cloned into the same expression vector as WT^{SilE}, using the ApaI/HindIII and PpuMI/HindIII sites respectively (Chapter 2, Section 2.2.4.4 and Figure 2.1). The $SilE_{21-143F}$ construct was used as a template to amplify the required SilE peptide regions by PCR (Figure 6.11-B), using Phusion® High-Fidelity DNA Polymerase - as part of a PCR Master Mix with GC Buffer. These PCR products were then reinserted into the restriction-digested vector, whilst preserving the HRV-3C cleavable N-terminal His-tag for purification purposes. In order to re-insert the Sil_{46-128} gene back into the vector, an ApaI

partial digest of the vector was necessitated. Figure 6.11-C shows an agarose gel of the ApaI partial vector digest that produced the essential 5,198 bp fragment for making the correct Sil_{46-128} overexpression construct. Correct engineering of the truncation mutants was confirmed by sequencing (Chapter 2, Section 2.2.4.4).

The truncations were overexpressed and purified using the same method as the WT^{SilE} protein (Chapter 3, Section 3.5). Again, the presence of fewer histidine residues meant that the final elution step required a lower concentration of imidazole to elute the truncation mutants, post HRV-3C protease treatment. Figure 6.11-D shows the pure truncated proteins following overexpression and purification on a SDS-PAGE gel. The expression and solubility levels of all of the truncation mutants were high, suggesting that they were stable. However, protein bands on the gel did show a little sign of degradation and suggested a decrease in stability over time in comparison to WT^{SilE}. The expected molecular weights of the truncation proteins were 10.1 kDa and 10.7 kDa for $SilE_{21-98}$ and $SilE_{46-128}$, respectively.

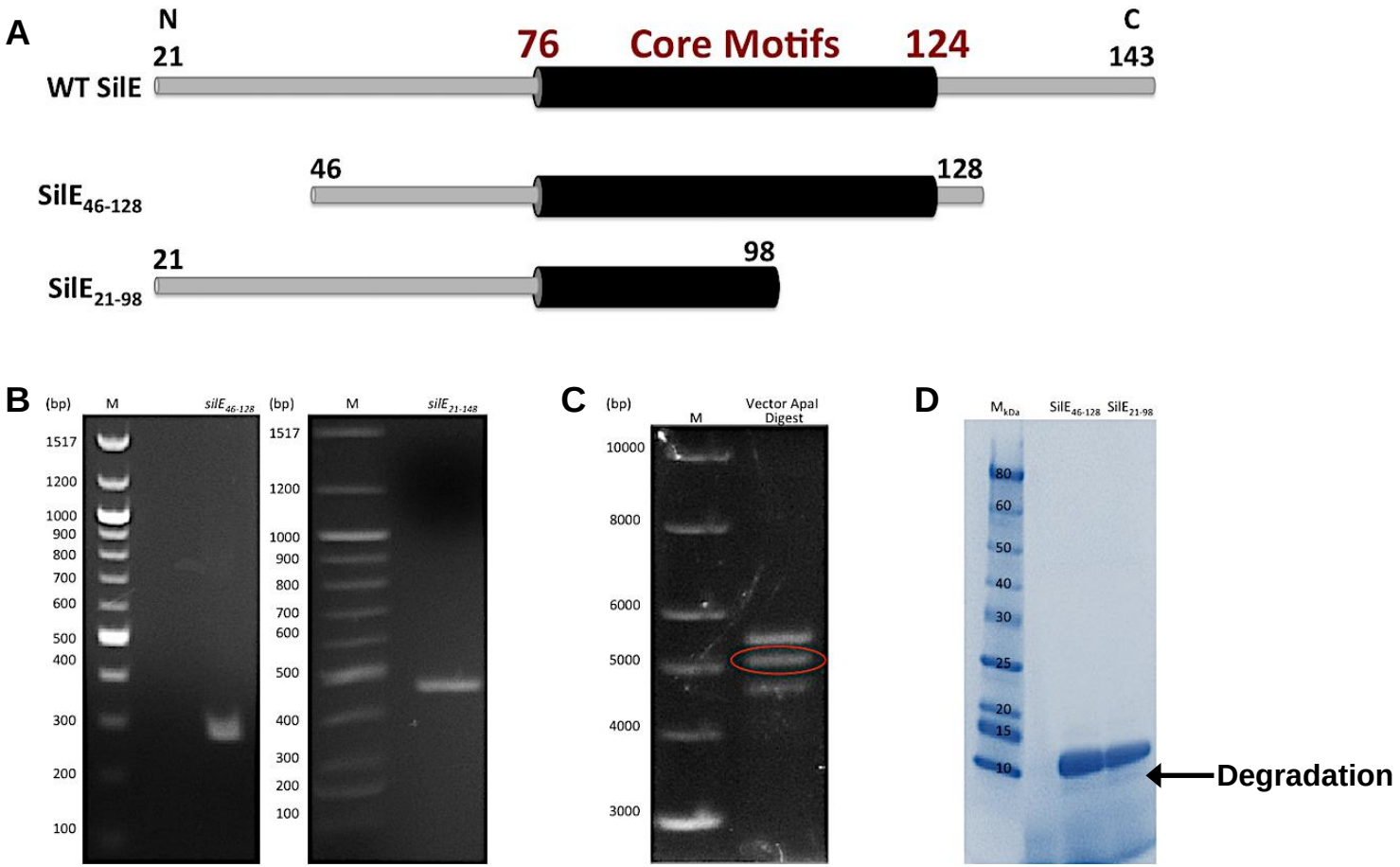


Figure 6. 11 Constructing the SilE truncation mutants

A. Schematic diagram of wild-type (WT) SilE, after removal of 20-amino acid periplasmic signal sequence, featuring proposed putative core motifs $SilE_{76\text{-}124}$ (coloured black and labeled in maroon) alongside truncated polypeptides $_{SilE46\text{-}128}$ and $SilE_{21\text{-}98}$. **B.** PCR products *silE*$_{46\text{-}128}$ (284 bp) and *silE*$_{21\text{-}98}$ (458 bp) on a 1 % agarose gel (Chapter 2, Section 2.2.2.1). **C.** ApaI partial digest of the pOPINF vector; the necessary fragment of 5,198 bp circled in red, on the 0.4 % agarose gel. **D.** 16 % SDS-PAGE of both of the truncation mutants at ~10 kDa (Chapter 2, Section 2.2.2.2). Note small amounts of degradation, designated by the black arrow.

6.2.4.2 Truncation of SilE affects Ag^+-induced protein folding

Ag^+-mediated folding of the truncation mutants $SilE_{46-128}$ and $SilE_{21-98}$ polypeptides was then studied using far-UV CD (Chapter 2, Section 2.2.6.1) in their *apo*-form (Figure 6.12-A) and in the presence of six Ag^+ molar equivalents (Figure 6.12-B), (45 μM protein, in 10 mM HEPES, 20 mM NaF at pH 7.5). In Figure 6.12-A, even though the *apo*-truncation mutants showed no secondary structural content with elipticity minima near 200 nm, they did not have an identical CD profile as the native protein. Both negative peaks presented a shift to lower wavelengths; where WT^{SilE} showed minima at 200 nm precisely, whilst $SilE_{46-128}$ displayed a negative peak at 199 nm and $SilE_{21-98}$ at 197 nm. This suggested an increase in disorder in their *apo*-form (Table 6.2).

Figure 6.12-B shows how the $SilE_{46-128}$ polypeptide exhibited some Ag^+-mediated folding similar to that of WT^{SilE}, whereas the $SilE_{21-98}$ polypeptide did not exhibit any signs of folding upon Ag^+-binding. Likewise, the secondary structure calculations carried out on Dichroweb's CONTINLL (Provencher and Glockner, 1981; Van Stokkum *et al.*, 1990; Sreerama and Woody, 2000; Lobley *et al.*, 2002), showed α-helical content (21 %) analogous to *apo*-WT^{SilE} (Chapter 4, Section 4.2.4.2).

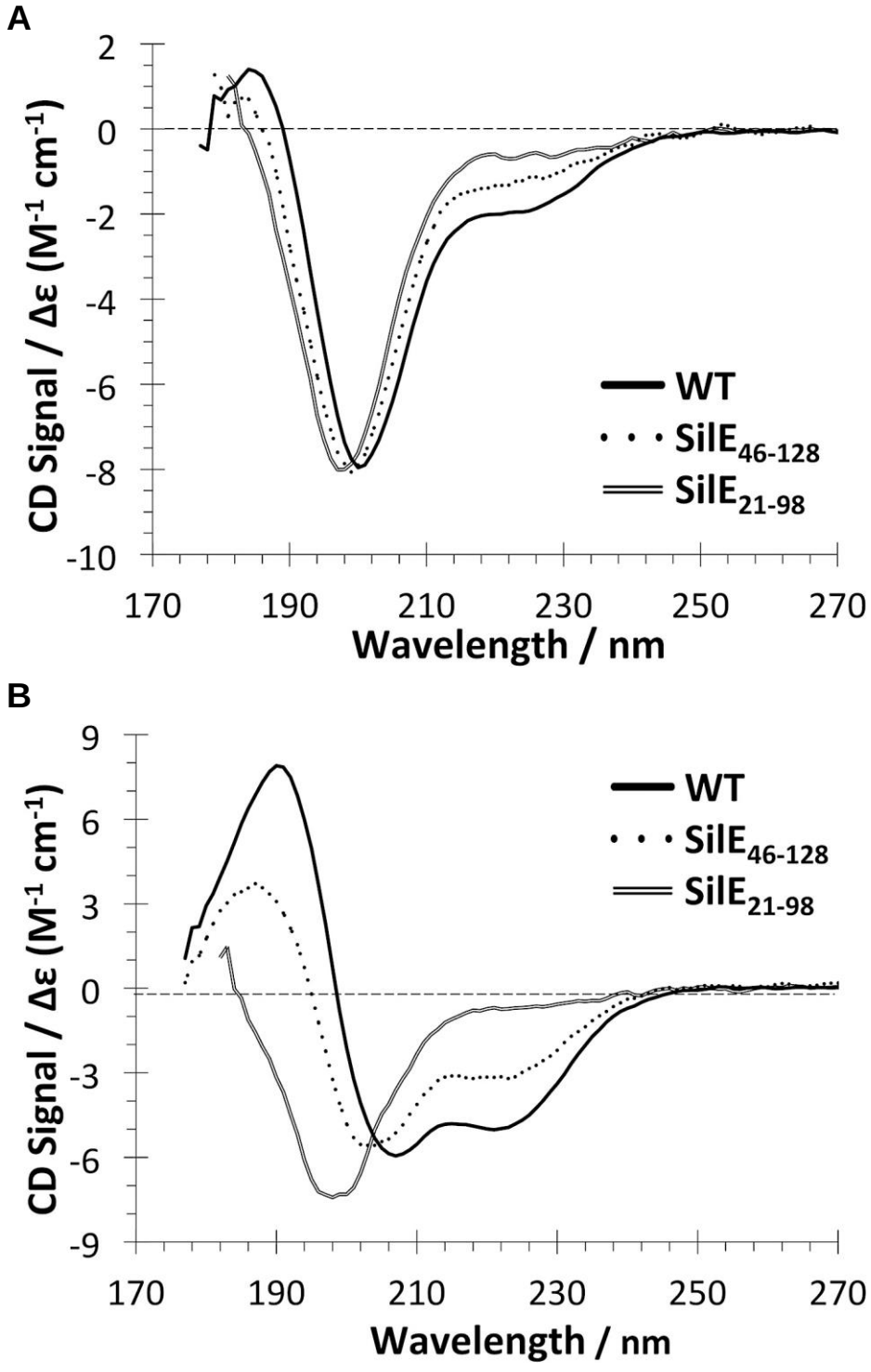


Figure 6. 12 SilE truncation effects on protein Ag^+-induced folding

Far-UV CD of WT^{SilE} and SilE truncation mutant proteins at a concentration of 45 μM, in 10 mM HEPES, 20 mM NaF at pH 7.5. **A.** *apo*-$SilE_{46-128}$ and $SilE_{21-98}$ both show a shift in negative peaks, down to lower wavelengths - 199 and 197 nm respectively. **B.** *holo*- (six Ag^+ molar equivalents) $SilE_{46-128}$ shows some formation of α-helical structure whereas $SilE_{21-98}$ displays no folding.

Table 6. 2: Percentage of truncation mutant SilE secondary structure calculated from far-UV CD data

CONTINLL was used via DichroWeb to calculate the percentage of secondary structure content in WT^{SilE} and the two truncation mutant SilE (residues 46-128 and 21-98) proteins, in the presence of six Ag^+ molar equivalents. The NRMSE values (< 0.1) confirm the accuracy of the secondary structure calculations (Provencher and Glockner, 1981; Van Stokkum *et al.*, 1990; Sreerama and Woody, 2000; Lobley *et al.*, 2002) (Chapter 4, Section 4.2.4.2).

SilE Protein: 6 Ag^+	α-Helical Content (%)	β-Sheet Content (%)	Turn Content (%)	Unordered Content (%)	NRMSE Value (to 3 d.p)
WT^{SilE}	54	3	18	25	0.083
$SilE_{46-128}$	29	4	29	38	0.067
$SilE_{21-98}$	21	6	30	44	0.180

6.2.4.3 ICP-MS SilE truncation Ag^+-binding analyses

The average number of Ag^+ ions capable of binding to each molecule of the truncation mutants was quantified by ICP-MS. The negative controls showed values of bound Ag^+ close to zero. The ICP analyses of both $SilE_{46-128}$ and $SilE_{21-98}$ truncation mutants exhibited a clear reduction in the number of bound Ag^+, when compared to WT^{SilE} (Figure 6.13). The reductions saw $SilE_{46-128}$ and $SilE_{21-98}$ bound to an average of 3.6 (±0.18) and 2.6 (±0.13) Ag^+, per SilE molecule, respectively. These are three fewer Ag^+ ions bound to $SilE_{46-128}$ and four less Ag^+ ions bound to $SilE_{21-98}$ in relation to WT^{SilE}.

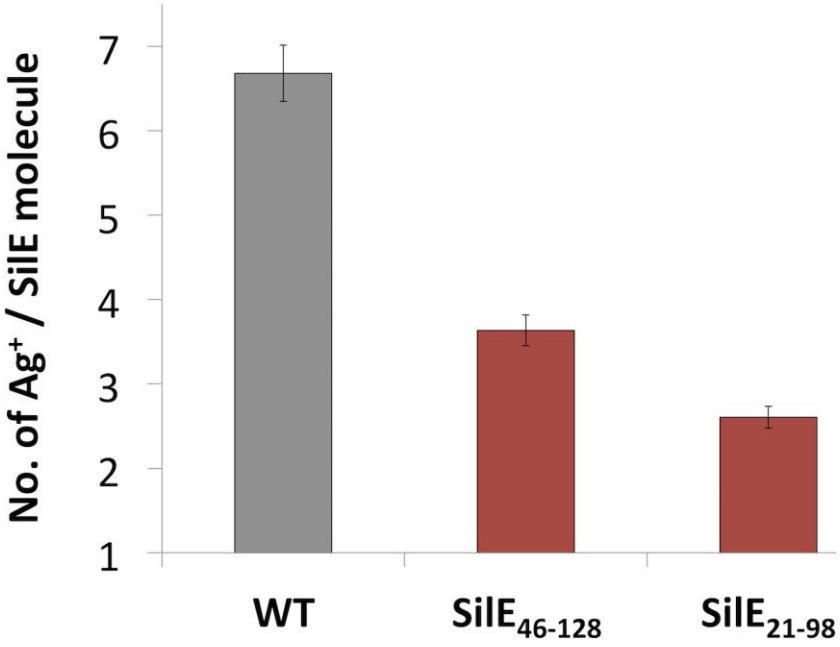


Figure 6. 13 ICP-MS Ag^+-binding analyses of SilE truncation mutants

ICP-MS Ag^+-binding analyses WT^{SilE} and SilE truncations mutants $SilE_{46-128}$ and $SilE_{21-98}$, at 50 nm, in 10 mM HEPES, 20 mM NaF, pH 7.5 in 1% HNO_3. Ag^+ levels are quantified relative to number of atoms per SilE protein molecule. Graphs show the mean of three independent replicates, from a single representative experiment, error bars represent standard error.

6.3 Summary and conclusions

Sequence analysis (Figures 6.1 and 6.2) led to the investigation of the role of a number of conserved histidine and methionine residues. Nine histdine residues were substituted with alanine residues resulting in reduced *holo*-SilE folding amongst several of the mutants (Figure 6.7). Significant folding effects were detected by far-UV CD for *holo*-mutants H87A and H111A (Figure 6.7), leading to the construction of additional mutants within this region of the protein. Similar effects on protein folding were observed for the Met to Leu mutants (Figure 6.9-B) and even more pronounced effects were seen in the double histidine mutants (Figure 6.9A). Complementary Ag^+-binding studies by ICP-MS showed reductions in the number of silver ions bound (Figure 6.10). The inability to bind Ag^+ to key residues resulted in a decrease in the amount of folded protein, as judged by the reduction in CD ellipticity in the 207-221 nm region. These data show that a number of conserved histidines and methionines are involved in Ag^+ binding in SilE, and the ability to bind Ag^+ has a direct effect on the *holo*-SilE structure.

Interestingly the mutants found within the central to C-terminally located 'core motifs' - MxxHxxxxxHxxMxx (motif 1) and HxxMxxxHxxMxx (motif 2) between residues 76-124, showed a more prominent loss in Ag^+-binding and structural folding. Following on, where both truncation mutants showed substantial reductions in the number of Ag^+ bound (Figure 6.13), the truncation mutant comprising the 'core motifs' ($SilE_{46-128}$) maintained Ag^+-mediated folding yet to a lesser extent than native SilE. Whereas, the C-terminal truncation mutant $SilE_{21-98}$ showed no

folding at all (Figure 6.12-B). Collectively, these data are consistent with a model where a nucleation core is formed by central core motifs 1 and 2, which then facilitate further folding as more Ag^+ ions are bound to the rest of the polypeptide.

CHAPTER 7

General Discussion and Future Work

7.1 Discussion

Recently, there has been a revived interest towards the use of Ag^+ compounds for clinical, agricultural and commercial purposes. However, Ag^+-resistant determinants have been found to be present in many clinical and environmental bacteria. Thus it is necessary to understand the mechanism of Ag^+-resistance in order to highlight the necessity to control the release of Ag^+ in the environment, as this can facilitate the spreading of Ag^+-resistant determinants among pathogenic microorganisms. Within this thesis, investigations were carried out in order to gain a better understanding of the mechanism of bacterial Ag^+-resistance.

At the genetic level the Ag^+-resistance *sil* genes have frequently been found located, adjacent to *pco* genes (Hao *et al.*, 2015; Hobman and Crossman, 2015; Staehlin *et al.*, 2016), on mobile elements that can be horizontally transferred (Randall *et al.*, 2015). The studies carried out in Chapter 3 of this thesis agree with these findings. Initial sequencing failed to find the *sil* and *pco* genes in plasmid pMG101 in one strain of *E. coli* J53. In another strain of *E. coli* J53 (pMG101), the *sil-pco* locus was located downstream of Tn*7*-like elements - implicated in the transposition of genetic islands into bacterial chromosomal DNA (Figure 3.5) (McKown *et al.*, 1988; Waddell and Craig, 1989; Peters and Craig, 2001; Peters *et al.*, 2014). These observations demonstrate how the Ag^+-resistance genes, alongside copper-resistance *pco* genes, are able to transpose and disseminate novel phenotypes among bacterial populations, in turn giving rise to co- and cross-resistance to antimicrobials. Further sequence analysis highlighted

conservation in the amino acid sequence of SilE (Figure 3.6), a key determinant in Ag^{+}-resistance.

Many proteins are intrinsically unstructured under physiological conditions, yet can fold when required to perform their biological functions (Dyson and Wright, 2005; Sibille and Bernadó, 2012; Kosol, Contreras-Martos, *et al.*, 2013). Previous NMR studies of the SilE homologue protein PcoE have shown that it folds to a predominantly α-helical structure upon binding ligand-metals Cu^{+} and Ag^{+} (Zimmermann *et al.*, 2012). Similarly, SilE was thought to be a putative IDP (intrinsically disordered protein) that binds to Ag^{+}, but no experimental evidence was available to determine its structure and metal binding properties. The data presented in this thesis show that *apo*-SilE is an IDP lacking significant secondary structure (Chapter 4). However, strong coupling between folding and metal-uptake was then demonstrated by showing that SilE binds Ag^{+} and folds into a moderately stable structure of high α-helical content (Chapter 5). Despite the involvement of multiple histidine residues in Ag^{+} binding, no pH induced folding of SilE was observed (Figure 5.7), verifying that SilE folding is specifically mediated by Ag^{+} binding. Each SilE molecule can bind up to a maximum of eight monovalent soft metal ions of Ag^{+} in solution (Figure 5.2) but is fully folded after binding six Ag^{+} (Figure 5.6): the last two Ag^{+} ions must therefore bind only to the fully folded protein. Cu^{2+}, Zn^{2+} and Ni^{2+} ions, exhibited lower stoichiometry, and consequently a lower degree of folding was seen (Figure 5.5), indicating that folding and binding are coupled and form part of the ion discrimination mechanism.

This use of folding and binding to enable uptake of a specific ion gives SilE the functional role of a 'molecular sponge' in the mechanism of *E. coli* Ag^+-resistance. As such its relationship to the other components in the Ag^+-resistance mechanism is likely to be one where the unfolded protein has a high affinity for silver, but a low affinity for other cellular components. In contrast the folded protein will have a high affinity for other cellular components and as such allow SilE to off-load its Ag^+ to other components in the resistance mechanism. Accordingly, SilE may act either as a metal-ion chaperone to the RND effluxer SilCBA or, as many IDPs do (Dyson and Wright, 2005; Kosol, Contreras-Martos, *et al.*, 2013) to bind or signal to the histidine kinase sensor SilS and thereby activate the remainder of the silver resistance machinery. Therefore, the notion that SilE binds Ag^+ and then initiates the entire mechanism of silver resistance in bacteria (Sil proteins P, G, A, B, F, C, R and S, as well as positively auto-regulating its own expression), by signalling or direct binding to SilS, is a highly plausible hypothesis (Figure 7.1).

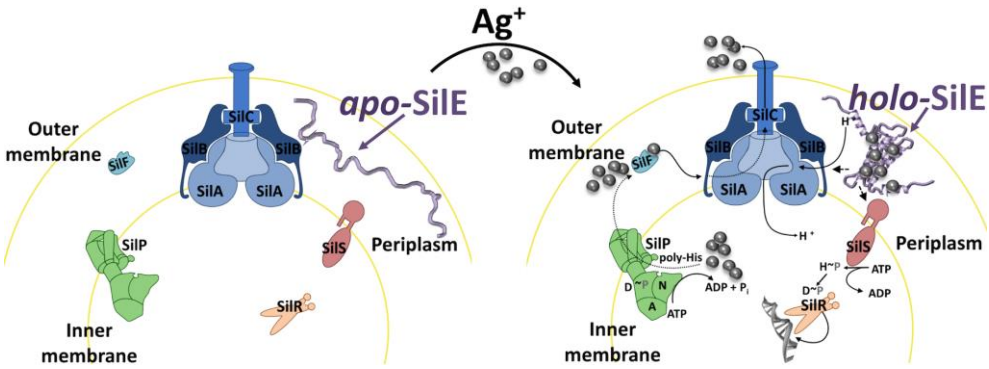


Figure 7. 1 Proposed role of SilE in silver resistance

We show that the periplasmic SilE protein, which is part of the Ag^+-resistance machinery of *E. coli*, is an intrinsically disordered protein in its *apo*-form and folds to a *holo*-form upon binding to Ag^+ ions. Upon folding, SilE may act as a metal ion chaperone to transport Ag^+ ions to the SilABC transporter and/or signal via the histidine kinase sensor SilS to activate the remainder of the Ag^+-resistance machinery.

Primary sequence alignment between SilE and its homologue PcoE revealed conserved histidine and methionine residues (Figure 6.2) (Zimmermann *et al.*, 2012) which in α-helical structures are known to be a common feature of many metal-binding proteins (Todd *et al.*, 1991; Tanaka *et al.*, 2004). This has been further confirmed to be the case in SilE by site-specific mutagenesis. Importantly, the mutagenesis studies showed that some residues - H80, H87, H111 and M108 - play a much more significant role than others in the correct Ag^+-mediated folding of SilE (Figures 6.7 and 6.9). Secondary structure predictions were consistent with our experimental data, suggesting the presence of six α-helices (Figure 5.3), however, significant helical structure was only realised in the Ag^+-bound state. In addition, two types of Ag^+-binding motifs, motif 1 MxxHxxxxxxHxxMxx and motif 2 HxxMxxxHxxMxx, were identified from the sequence in which *i* to *i+3* or *i+4* spacing between residues ensures that they appear on the

same face of a folded helix. This suggests that each single helical motif in itself is unable to bind Ag^+, but several of these together are able to (co-operatively) co-ordinate multiple Ag^+ ions leading to a compact protein fold. Two of these motifs are located at the centre of the primary SilE sequence between residues S76-G124 (Figure 6.2) and we define them as the 'core-motifs' (Figure 6.11-A and 7.2). Both H80 and H87 are located in the core-motif 2 whereas M108 and H111 are located in the core-motif 1. We suggest the possibility of a Ag^+-mediated folding mechanism whereby initial Ag^+ binding to these 'core-motifs' induces a nucleation site around which the rest of the SilE structure assembles facilitating further Ag^+ binding (Figure 7.2).

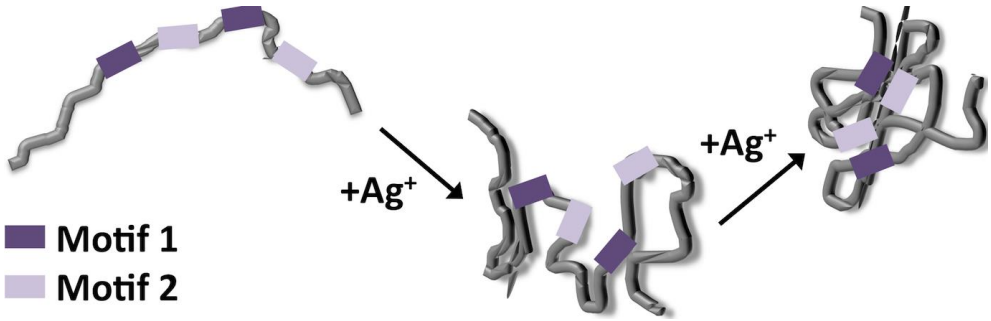


Figure 7. 2 Nucleation model for Ag^+-binding induced SilE folding

Schematic representation of the speculative nucleation model for Ag^+-mediated folding of SilE. SilE coordination of Ag^+ ions is mediated via conserved histidine and methionine residues arranged in two characteristic motifs (Motif 1: MxxHxxxxxHxxMxx and Motif 2: HxxMxxxHxxMxx) (Figure 6.2-B and -C and 6.11-A). As SilE binds Ag^+ and following initial nucleation within the putative core motifs (Figure 6.11-A) the rest of the polypeptide folds gradually to its complete *holo*-structure upon further Ag^+-binding. Maximal folding of SilE is induced upon binding to six Ag^+ ions but each *holo*-SilE molecule can bind up to eight Ag^+ ions.

This is consistent with the sigmoidal plots of CD signals at 190 and 221 nm as a function of the Ag^+/SilE molar ratio (Figure 5.6-B) which are consistent with co-operative Ag^+ binding and a folding nucleus rather than a simple sequential Ag^+ binding and folding model. We tested this model by engineering truncated SilE polypeptides. The $SilE_{46-128}$ polypeptide contained both 'core motifs' and exhibited Ag^+-mediated folding in a similar manner to the native SilE whereas the $SilE_{21-98}$ polypeptide lacking the core-motif 1 exhibited no detectable Ag^+-mediated folding; a result consistent with our nucleation model. Our double histidine mutations H80AH87, H80H111A and H87AH111A revealed loss of two Ag^+ ions and it is therefore likely that these pairs of histidines coordinate different Ag^+ ions.

Complexes of Ag^+ with histidine and imidazole can form with Ag^+:L and Ag^+:L_2 stoichiometries (where L is the ligand; histidine or imidazole) with the latter being more stable, as indicated by higher stability constants and enthalpies of formation (Czoik *et al.*, 2008). Ag^+:L_2 complexes adopt a linear geometry (Muller *et al.*, 2005; Petrovec *et al.*, 2012; Kumbhar *et al.*, 2013) and have been proposed to coordinate Ag^+ ions in SilE (Silver, 2003). Methionine residues are reported to co-ordinate Ag^+ ions in non-linear geometries with Ag^+:M_2 and Ag^+:M_3 stoichiometries in the multicopper oxidase CueO via methionine-rich helices (Singh *et al.*, 2011) while CusF co-ordinates Ag^+ ions in a manner that employs histidine, methionine and tryptophan residues (Loftin *et al.*, 2007). The human copper transporter 1 (hCtr1) protein in turn co-ordinates Cu^+ and Ag^+ ions via histidine, methionine and cysteine residues (Rubino *et al.*, 2010; Du *et al.*, 2013; Zhu

et al., 2014). It is clear that coordination of Ag^{+} ions by metal binding proteins can employ a diverse collection of amino acids and variable co-ordination geometry. The data collectively show that the previous theoretical model of SilE binding to five Ag^{+} via ten His residues (Silver, 2003) is not correct but only a high resolution structure of the *holo*-SilE will reveal the molecular details of Ag^{+} coordination in this protein.

A final point to note is that even though the periplasm where SilE resides is an oxidising environment, Cu^{+} can also be found there. However, due to the nature of these studies and certain circumstances (i.e. experiments being conducted in a number of various laboratories), reduced Cu^{+} binding and induced-folding were not tested. Therefore, as PcoE showed greater metal binding and protein folding in the presence of Cu^{+} in comparison to Cu^{2+}; SilE could possibly show similar binding and folding effects in the presence of Cu^{+} as those that were observed with Ag^{+}.

7.2 Future work

Whilst the absence of a chromosomal *attT7n* site prior to the *sil* genes and the presence of the plasmid methyltransferase protein suggested that the *sil-pco* locus was present in plasmid pMG101 (and not the chromosome of J53 of the WGS); the data was still insufficient to come to any solid conclusion on if *sil-pco* locus was within the plasmid or chromosome of our WGS data. The inconclusive nature of this sequencing data requires further analyses to decide the 'mobile' island's exact location. PCRs could be used to align contigs that could potentially bridge the gaps seen either side of the *sil-pco* locus. Further to this, the entire determinant has not been shown to confer Ag^+-resistance as yet and the role of *pco* in Ag^+-resistance is unclear. Confirmation of the presence of the *sil-pco* locus in the plasmid could allow for transposition experiments into another plasmid, to ascertain that the *sil-pco* island itself confers full resistance to Ag^+.

Secondly, even though SilE has been recognised as a key component (Randall *et al.*, 2015), its exact role in the mechanism of Ag^+-resistance is still unclear. There is no understanding of: how the protein recognises Ag^+, how this is coupled to its folding mechanism and whether it interacts with other components in the *sil* system. Hence, to understand the Ag^+-binding activity of SilE in the context of its biological role in resistance future studies can include:

(i) Characterisation of the *apo*- and *holo*-structures of SilE by NMR. Acquiring fully assigned NMR data would enable the

investigation of specific residues involved in Ag^{+}-binding mediated SilE folding and

(ii) Protein-protein interaction experiments, such as pull-down assays or the bacterial two-hybrid technique, to see whether SilE interacts with other components of the resistance mechanism. This would enable clarification on where SilE offloads the Ag^{+} and verify a functional role for the protein as a first line of defence responder, periplasmic chaperone, system regulator or 'molecular sponge'.

References

Abou El-Nour, K.M.M., Eftaiha, A., Al-Warthan, A., and Ammar, R.A.A., (2010). Synthesis and applications of silver nanoparticles. *Arabian Journal of Chemistry*. **3**, pp. 135–140.

Abzalimov, R. R., Frimpong, A. K. and Kaltashov, I. A., (2012). Detection and characterization of large-scale protein conformational transitions in solution using charge-state distribution analysis in ESI-MS, in: Uversky, V. N. and Dunker K. A., Ed. *Intrinsically Disordered Protein Analysis: Methods and Experimental Tools* (2nd ed.), Methods in Molecular Biology (896), Springer Protocols, Humana Press, Hertfordshire, UK pp. 365-374.

Aebersold, R., and Mann, M., (2003). Mass spectrometry-based proteomics. *Nature*. **422**, pp. 198–207.

Ahmed, A., Rushworth, J. V., Hirst, N.A., and Millner, P.A., (2014). Biosensors for whole-cell bacterial detection. *Clinical Microbiology Reviews*. **27**, pp. 631–646.

Alexander, J.W., (2009). History of the medical use of silver. *Surgical Infections*. **10**, pp. 289–292.

Allen, H.K., Donato, J., Wang, H.H., Cloud-Hansen, K.A., Davies, J., and Handelsman, J., (2010). Call of the wild: antibiotic resistance genes in natural environments. *Nature Reviews. Microbiology*. **8**, pp. 251–259.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J., (1990). Basic local alignment search tool. *Journal of Molecular Biology*. **215**, pp. 403–410.

Anthis, N.J., and Clore, G.M., (2013). Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm. *Protein Science*. **22**, pp. 851–858.

Asiani, K.R., Williams, H., Bird, L., Jenner, M., Searle, M.S., Hobman, J.L., Scott, D.J., and Soultanas, P., (2016). SilE is an intrinsically disordered periplasmic 'molecular sponge' involved in bacterial silver resistance. *Molecular Microbiology*. **101**, pp. 731-742.

Atiyeh, B.S., Costagliola, M., Hayek, S.N., and Dibo, S.A., (2007). Effect of silver on burn wound infection control and healing: Review of the literature. *Burns*. **33**, pp. 139–148.

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., and Zagnitko, O., (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics*. **9**, pp. 75.

Banerjee, S., and Mazumdar, S., (2012). Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. *International Journal of Analytical Chemistry*. **2012**,

pp. 1–40.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. a., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. a., and Pevzner, P. a., (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*. **19**, pp. 455–477.

Barillo, D.J., and Marx, D.E., (2014). Silver in medicine: A brief history BC 335 to present. *Burns*. **40**, pp. S3–8.

Batsala, M., Chandu, B., Sakala, B., Nama, S., and Domatoti, S., (2012). Inductively coupled plasma mass spectrometry (ICP-MS). *International Jounal of Research in Pharmacy and Chemistry*. **2**, pp. 671–680.

Beiersdorf (2016). *Nivea deodorant silver protect*. [Online]. Available at: http://www.nivea.co.uk/products/deodorants/silver-protect, [Accessed on 7 April 2016].

Bennett, P.M., (2008). Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic resistance genes in bacteria. *British Journal of Pharmacology*. **153**, pp. S347–357.

Bernadó, P., Mylonas, E., Petoukhov, M. V., Blackledge, M., and Svergun, D.I., (2007). Structural characterization of flexible proteins using small-angle X-ray scattering. *Journal of the American Chemical Society*. **129**, pp. 5656–5664.

Bernardó, P. and Svergun, D. I., (2010) Structural insight into intrinsically disordered proteins by small-angle X-ray scattering, in: Uversky V. N. and Longhi S., Eds. *Instrumental Analysis of Intrinsically Disordered Proteins: Assessing Structure And Conformation* (1st ed.), John Wiley & Sons Inc., New York pp. 451–475.

Bernadó, P., and Svergun, D.I., (2012). Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. *Molecular BioSystems*. **8**, pp. 151–167.

Berrow, N.S., Alderton, D., Sainsbury, S., Nettleship, J., Assenberg, R., Rahman, N., Stuart, D.I., and Owens, R.J., (2007). A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Research*. **35**, pp. 1–12.

Bersch, B., Derfoufi, K.M., De Angelis, F., Auquier, V., Ngonlong Ekendé, E., Mergeay, M., Ruysschaert, J.M., and Vandenbussche, G., (2011). Structural and metal binding characterization of the C-terminal metallochaperone domain of membrane fusion protein SilB from Cupriavidus metallidurans CH34. *Biochemistry*. **50**, pp. 2194–2204.

Bhagya, C.H.W.M.R.C., Wijesundera Sulochana, W.S., and Hemamali, N.P., (2013). Polymerase chain reaction optimization for amplification of Guanine-Cytosine rich templates using buccal cell DNA. *Indian Journal of Human Genetics*. **19**, pp. 78–83.

Biosis.net, (2016). *About ScÅtter*, [Online]. Available at: http://www.bioisis.net/tutorial/9, [Accessed on 16 April 2016].

Bird, L.E., (2011). High throughput construction and small scale expression screening of multi-tag vectors in *Escherichia coli*. *Methods*. **55**, pp. 29–37.

Birnboim, H.C., and Doly, J., (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research*. **7**, pp. 1513–1523.

Blattner, F.R., Plunkett, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B., and Shao, Y., (1997). The complete genome sequence of *Escherichia coli* K-12. *Science*. **277**, pp. 1453–1462.

Bloom, J.W., and Mann, K.G., (1979). Prothrombin domains: circular dichroic evidence for a lack of cooperativity. *Biochemistry*. **18**, pp. 1957–1961.

Bodenhausen, G., and Ruben, D.J., (1980). Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chemical Physics Letters*. **69**, pp. 185–189.

Böhm, G., Muhr, R., and Jaenicke, R., (1992). Quantitative analysis of protein far UV circular dichroism spectra by neural networks. *Protein Engineering*. **5**, pp. 191–195.

Bornhorst, J.A., and Falke, J.J., (2000). Purification of proteins using polyhistidine affinity tags. *Methods in Enzymology*. **326**, pp. 245–254.

Brahms, S., and Brahms, J., (1980). Determination of protein secondary structure in solution by vacuum ultraviolet circular dichroism. *Journal of Molecular Biology*. **138**, pp. 149–178.

Breydo, L., and Uversky, V., (2011). Role of metal ions in aggregation of intrinsically disordered proteins in neurodegenerative diseases. *Metallomics*. **3**, pp. 1163–1180.

Brown, N.L., Barrett, S.R., Camakaris, J., Lee, B.T., and Rouch, D.A., (1995). Molecular genetics and transport analysis of the copper-resistance determinant (pco) from *Escherichia coli* plasmid pRJ1004. *Molecular Microbiology*. **17**, pp. 1153–1166.

Carrera, F., Sa, E., Merkling, P.J., and Mun, A., (2004). Nature of metal binding sites in Cu (II) complexes with histidine and related N-coordinating ligands, as studied by EXAFS. *Inorganic Chemistry*. **43**, pp. 6674–6683.

Carver, T., Harris, S.R., Berriman, M., Parkhill, J., and McQuillan, J.A., (2012). Artemis: An integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics*. **28**, pp. 464–469.

Cavanagh, J., Fairbrother, W. J., and Palmer, I. A. G., (2006). *Protein NMR spectroscopy : Principles and practice* (2nd edn.) [e-book], Academic Press, Burlington, US, pp. 273-280. Available from ProQuest ebrary, at: http://site.ebrary.com/lib/uon/reader.action?docID=10167118&ppg=28, [Accessed 7 May 2016].

Chen, Y.-H., Yang, J.T., and Martinez, H.M., (1972). Determination of the secondary structures of proteins by circular dichroism and optical rotatory dispersion. *Biochemistry*. **11**, pp. 4120–4131.

Chopra, I., (2007). The increasing use of silver-based products as antimicrobial agents: A useful development or a cause for concern? *Journal of Antimicrobial Chemotherapy*. **59**, pp. 587–590.

Chou, P.Y., and Fasman, G.D., (1974). Prediction of protein conformation. *Biochemistry*. **13**, pp. 222–245.

Clement, J.L., and Jarrett, P.S., (1994). Antibacterial silver. *Metal-based Drugs*. **1**, pp. 467–482.

Clowes, R.C., and Rowley, D., (1954). Some observations on linkage effects in genetic recombination in *Escherichia coli* K-12. *Journal of General Microbiology*. **11**, pp. 250–260.

Crabtree, J.H., Burchette, R.J., Siddiqi, R.A., Huen, I.T., Hadnott, L.L., and Fishman, A., (2003). The efficacy of silver-ion implanted catheters in reducing peritoneal dialysis-related infections. *Peritoneal Dialysis International*. **23**, pp. 368–374.

Czoik, R., Heintz, A., John, E., and Marczak, W., (2008). Complexes of silver with histidine and imidazole investigated by the calorimetric and potentiometric methods. *Acta Physica Polonica A*. **114**, pp. 51–56.

Darling, A.C.E., Mau, B., Blattner, F.R., and Perna, N.T., (2004). Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Research*. **14**, pp. 1394–1403.

DeForte, S., and Uversky, V., (2016). Resolving the ambiguity: Making sense of intrinsic disorder when pdb structures disagree. *Protein Science*. **25**, pp. 676–688.

DeLong, E.F., and Pace, N.R., (2001). Environmental diversity of bacteria and archaea. *Systematic Biology*. **50**, pp. 470–478.

Depuydt, M., Leonard, S.E., Vertommen, D., Denoncin, K., Morsomme, P., Wahni, K., Messens, J., Carroll, K.S., and Collet, J.-F., (2009). A periplasmic reducing system protects single cysteine residues from oxidation. *Science*. **326**, pp. 1109–1111.

Dodero, V.I., Quirolo, Z.B., and Sequeira, M.A., (2011). Biomolecular studies by circular dichroism. *Frontiers in Bioscience*. **16**, pp. 61–73.

Drozdetskiy, A., Cole, C., Procter, J., and Barton, G.J., (2015). JPred4: A protein secondary structure prediction server. *Nucleic Acids Research*. **43**, pp. 389–394.

Du, X., Li, H., Wang, X., Liu, Q., Ni, J., and Sun, H., (2013). Kinetics and thermodynamics of metal binding to the N-terminus of a human copper transporter, hCTR1. *Chemical Communications*. **49**, pp. 9134–9136.

Dunker, A.K., Lawson, J.D., Brown, C.J., Williams, R.M., Romero, P., Oh, J.S., Oldfield, C.J., Campen, A.M., Ratliff, C.M., Hipps, K.W., Ausio, J., Nissen, M.S., Reeves, R., Kang, C., Kissinger, C.R., Bailey, R.W., Griswold, M.D., Chiu, W., Garner, E.C., and Obradovic, Z., (2001).

Intrinsically disordered protein. *Journal of Molecular Graphics and Modelling*. **19**, pp. 26–59.

Durán, N., Durán, M., de Jesus, M.B., Seabra, A.B., Fávaro, W.J., and Nakazato, G., (2016). Silver nanoparticles: A new view on mechanistic aspects on antimicrobial activity. *Nanomedicine: Nanotechnology, Biology and Medicine*. **12**, pp. 789–799.

Dyson, H.J., and Wright, P.E., (2005). Intrinsically unstructured proteins and their functions. *Nature Reviews. Molecular Cell Biology*. **6**, pp. 197–208.

Eickholt, J., and Cheng, J., (2013). DNdisorder: predicting protein disorder using boosting and deep networks. *BMC Bioinformatics*. **14**, pp. 88.

eMedExpert.com, (2016). *List of antibiotics*, [Online]. Available at: http://www.emedexpert.com/lists/antibiotics.shtml**,** [Accessed on 11 May 2016].

Faller, P., Hureau, C., and La Penna, G., (2014). Metal ions and intrinsically disordered proteins and peptides: from Cu/Zn amyloid-beta to general principles. *Accounts of Chemical Research*. **47**, pp. 2252–2259.

Fenn, J.B., (2003). Electrospray wings for molecular elephants (Nobel lecture). *Angewandte Chemie (International ed. in English)*. **42**, pp. 3871–94.

Ferron, F., Longhi, S., Canard, B., and Karlin, D., (2006). A practical overview of protein disorder prediction methods. *Proteins: Structure, Function and Genetics*. **65**, pp. 1–14.

Fiaux, J., Bertelsen, E.B., Horwich, A.L., and Wuthrich, K., (2002). NMR analysis of a 900K GroEL-GroES complex. *Nature*. **418**, pp. 207–211.

Finberg, R.W., Moellering, R.C., Tally, F.P., Craig, W.A., Pankey, G.A., Dellinger, E.P., West, M.A., Joshi, M., Linden, P.K., Rolston, K. V, Rotschafer, J.C., and Rybak, M.J., (2004). The importance of bactericidal drugs: Future directions in infectious disease. *Clinical Infectious Disease*. **39**, pp. 1314–1320.

Fontana, A., De Laureto, P.P., Spolaore, B., and Frare, E., (2012). Identifying disordered regions in proteins by limited proteolysis. *Methods in Molecular Biology*. **896**, pp. 297–318.

Forrest, R.D., (1982). Early history of wound treatment. *Journal of the Royal Society of Medicine*. **75**, pp. 198–205.

Fox, M.F., Barker, B.E., and Hayon, E., (1978). Far-ultraviolet solution spectroscopy of chloride ion. *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases*. **74**, pp. 1776–1785.

Franci, G., Falanga, A., Galdiero, S., Palomba, L., Rai, M., Morelli, G., and Galdiero, M., (2015). Silver nanoparticles as potential antibacterial agents. *Molecules*. **20**, pp. 8856–8874.

Franke, S., Grass, G., and Nies, D.H., (2001). The product of the ybdE gene of the *Escherichia coli* chromosome is involved in detoxification of silver ions. *Microbiology*. **147**, pp. 965–972.

Franke, S., Grass, G., Rensing, C., and Nies, D.H., (2003). Molecular analysis of the copper-transporting efflux system CusCFBA of *Escherichia coli*. *Journal of Bacteriology*. **185**, pp. 3804–3812.

Gallo Cassarino, T., Bordoli, L., and Schwede, T., (2014). Assessment of ligand binding site predictions in CASP10. *Proteins: Structure, Function and Bioinformatics*. **82**, pp. 154–163.

Garnier, J., Osguthorpe, D.J., and Robson, B., (1978). Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *Journal of Molecular Biology*. **120**, pp. 97–120.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., and Bairoch, A., (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*. **31**, pp. 3784–3788.

Geerlof, A., Brown, J., Coutard, B., Egloff, M.P., Enguita, F.J., Fogg, M.J., Gilbert, R.J.C., Groves, M.R., Haouz, A., Nettleship, J.E., Nordlund, P., Owens, R.J., Ruff, M., Sainsbury, S., Svergun, D.I., and Wilmanns, M., (2006). The impact of protein characterization in structural proteomics. *Acta Crystallographica Section D: Biological Crystallography*. **62**, pp. 1125–1136.

Gehrig, P.M., You, C., Dallinger, R., Gruber, C., Brouwer, M., Kägi, J.H., and Hunziker, P.E., (2000). Electrospray ionization mass spectrometry of zinc, cadmium, and copper metallothioneins: evidence for metal-binding cooperativity. *Protein Science: A Publication of the Protein Society*. **9**, pp. 395–402.

George, G.N., Pickering, I.J., Jake Pushie, M., Nienaber, K., Hackett, M.J., Ascone, I., Hedman, B., Hodgson, K.O., Aitken, J.B., Levina, A., Glover, C., and Lay, P.A., (2012). X-ray-induced photo-chemistry and X-ray absorption spectroscopy of biological samples. *Journal of Synchrotron Radiation*. **19**, pp. 875–886.

George, S.J., Fu, J., Guo, Y., Drury, O.B., Friedrich, S., Rauchfuss, T., Volkers, P.I., Peters, J.C., Scott, V., Brown, S.D., Thomas, C.M., and Cramer, S.P., (2008). X-ray photochemistry in iron complexes from Fe(0) to Fe(IV) - Can a bug become a feature? *Inorganica Chimica Acta*. **361**, pp. 1157–1165.

Giedraitienė, A., Vitkauskienė, A., Naginienė, R., and Pavilonis, A., (2011). Antibiotic resistance mechanisms of clinically important bacteria. *Medicina*. **47**, pp. 137–146.

Glasel, J., (1995). Validity of nucleic acid purities monitored by 260/280 absorbance ratios. *BioTechniques*. **18**, pp. 62–63.

Godt, J., Scheidig, F., Grosse-Siestrup, C., Esche, V., Brandenburg, P., Reich, A., and Groneberg, D.A., (2006). The toxicity of cadmium and resulting hazards for human health. *Journal of Occupational Medicine and*

Toxicology. **1**, pp. 1–22.

Good, N.E., Winget, G.D., Winter, W., Connolly, T.N., Izawa, S., and Singh, R.M.M., (1966). Hydrogen ion buffers for biological research. *Analytical Biochemistry*. **5**, pp. 467–477.

Gordon, O., Slenters, T.V., Brunetto, P.S., Villaruz, A.E., Sturdevant, D.E., Otto, M., Landmann, R., and Fromm, K.M., (2010). Silver coordination polymers for prevention of implant infection: Thiol interaction, impact on respiratory chain enzymes, and hydroxyl radical induction. *Antimicrobial Agents and Chemotherapy*. **54**, pp. 4208–4218.

Greenfield, N.J., (2007a). Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols*. **1**, pp. 2876–2890.

Greenfield, N.J., (2007b). Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nature Protocols*. **1**, pp. 2527–2535.

Gupta, A., and Lutsenko, S., (2012). Evolution of copper transporting ATPases in eukaryotic organisms. *Current Genomics*. **13**, pp. 124–133.

Gupta, A., Matsui, K., Lo, J.F., and Silver, S., (1999). Molecular basis for resistance to silver cations in Salmonella. *Nature Medicine*. **5**, pp. 183–188.

Gupta, A., Maynes, M., and Silver, S., (1998). Effects of halides on plasmid-mediated silver resistance in *Escherichia coli*. *Applied and Environmental Microbiology*. **64**, pp. 5042–5045.

Gupta, A., Phung, L.T., Taylor, D.E., and Silver, S., (2001). Diversity of silver resistance genes in IncH incompatibility group plasmids. *Microbiology*. **147**, pp. 3393–3402.

Gust, D., Moon, R.B., and Roberts, J.D., (1975). Applications of natural-abundance nitrogen-15 nuclear magnetic resonance to large biochemically important molecules. *Proceedings of the National Academy of Sciences of the United States of America*. **72**, pp. 4696–4700.

Guzzo, A. V, (1965). The influence of amino-acid sequence on protein structure. *Biophysical Journal*. **5**, pp. 809–22.

Habash, M.B., Park, A.J., Vis, E.C., Harris, R.J., and Khursigara, C.M., (2014). Synergy of silver nanoparticles and aztreonam against Pseudomonas aeruginosa PAO1 biofilms. *Antimicrobial Agents and Chemotherapy*. **58**, pp. 5818–5830.

Hada, H., Yonezawa, Y., and Kurakake, A., (1976). Photoreduction of silver ion in aqueous and alcoholic solutions. *The Journal of Physical Chemistry*. **80**, pp. 2728–2731.

Haefeli, C., Franklin, C., and Hardy, K., (1984). Plasmid-determined silver resistance in Pseudomonas stutzeri isolated from a silver mine. *Journal of Bacteriology*. **158**, pp. 389–392.

Hall, B.G., Acar, H., Nandipati, A., and Barlow, M., (2014). Growth rates made easy. *Molecular Biology and Evolution*. **31**, pp. 232–238.

Hall, N., (2007). Advanced sequencing technologies and their wider

impact in microbiology. *The Journal of Experimental Biology*. **209**, pp. 1518–1525.

Hao, X., Lüthje, F.L., Qin, Y., McDevitt, S.F., Lutay, N., Hobman, J.L., Asiani, K., Soncini, F.C., German, N., Zhang, S., Zhu, Y.G., and Rensing, C., (2015). Survival in amoeba - a major selection pressure on the presence of bacterial copper and zinc resistance determinants? Identification of a 'copper pathogenicity island'. *Applied Microbiology and Biotechnology*. **99**, pp. 5817–5824.

Harada, M., Yamada, M., Kimura, Y., and Saijo, K., (2013). Influence of the organization of water-in-ionic liquid microemulsions on the size of silver particles during photoreduction. *Journal of Colloid and Interface Science*. **406**, pp. 94–104.

He, B., Wang, K., Liu, Y., Xue, B., Uversky, V.N., and Dunker, A.K., (2009). Predicting intrinsic disorder in proteins: an overview. *Cell Research*. **19**, pp. 929–949.

Heckman, K.L., and Pease, L.R., (2007). Gene splicing and mutagenesis by PCR-driven overlap extension. *Nature Protocols*. **2**, pp. 924–932.

Hernandez, H., and Robinson, C. V, (2007). Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. *Nature Protocols*. **2**, pp. 715–726.

Hertel, N., and Hoffmann, S.V., (2011). ASTRID2: A new Danish low-emittance SR source. *Synchrotron Radiation News*. **24**, pp. 19–23.

Hobman, J.L., and Crossman, L.C., (2015). Bacterial antimicrobial metal ion resistance. *Journal of Medical Microbiology*. **64**, pp. 471–497.

Holt, K., and Bard, A., (2005). Interaction of silver(I) ions with the respiratory chain of *Escherichia coli*: an electrochemical and scanning electrochemical microscopy study of the antimicrobial mechanism of micromolar Ag^{+}. *Biochemistry*. **44**, pp. 13214–13223.

Holtzer, M.E., and Holtzer, A., (1992). α-Helix to random coil transitions: Determination of peptide concentration from the CD at the isodichroic point. *Biopolymers*. **32**, pp. 1675–1677.

Horng, J.C., Moroz, V., and Raleigh, D.P., (2003). Rapid cooperative two-state folding of a miniature alpha-beta protein and design of a thermostable variant. *Journal of Molecular Biology*. **326**, pp. 1261–1270.

Howard, M.J., (1998). Protein NMR spectroscopy. *Current Biology*. **8**, pp. 331–333.

Huang, H., and Yang, Y., (2008). Preparation of silver nanoparticles in inorganic clay suspensions. *Composites Science and Technology*. **68**, pp. 2948–2953.

Iakoucheva, L.M., Kimzey, A.L., Masselon, C.D., Smith, R.D., Dunker, A.K., and Ackerman, E.J., (2001). Aberrant mobility phenomena of the DNA repair protein XPA. *Protein Science: A Publication of the Protein Society*. **10**, pp. 1353–62.

Irving, H., and Williams, R.J.P., (1953). The stability of transition-metal

complexes. *Journal of the Chemical Society*. **637**, pp. 3192–3210.

Ishida, T., and Kinoshita, K., (2007). PrDOS: Prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Research*. **35**, pp. W460–464.

Jacques, D.A., and Trewhella, J., (2010). Small-angle scattering for structural biology - Expanding the frontier while avoiding the pitfalls. *Protein Science*. **19**, pp. 642–657.

Jelenko, C., (1969). Silver nitrate resistant *E. coli*: report of case. *Annals of Surgery*. **170**, pp. 296–299.

Johnson, T.J., Kariyawasam, S., Wannemuehler, Y., Mangiamele, P., Johnson, S.J., Doetkott, C., Skyberg, J.A., Lynne, A.M., Johnson, J.R., and Nolan, L.K., (2007). The genome sequence of avian pathogenic *Escherichia coli* strain O1:K1:H7 shares strong similarities with human extraintestinal pathogenic *E. coli* genomes. *Journal of Bacteriology*. **189**, pp. 3228–3236.

Johnson, W.C., (1990). Protein secondary structure and circular dichroism: a practical guide. *Proteins*. **7**, pp. 205–14.

Johnson, W.C., (1999). Analyzing protein circular dichroism spectra for accurate secondary structures. *Proteins: Structure, Function and Genetics*. **35**, pp. 307–312.

Jung, W.K., Koo, H.C., Kim, K.W., Shin, S., Kim, S.H., and Park, Y.H., (2008). Antibacterial activity and mechanism of action of the silver ion in Staphylococcus aureus and *Escherichia coli*. *Applied and Environmental Microbiology*. **74**, pp. 2171–2178.

Kabsch, W., and Sander, C., (1983). How good are predictions of protein secondary structure? *FEBS Letters*. **155**, pp. 179–182.

Kelly, S., and Price, N., (2000). The use of circular dichroism in the investigation of protein structure and function. *Current Protein & Peptide Science*. **1**, pp. 349–384.

Kelly, S.M., Jess, T.J., and Price, N.C., (2005). How to study proteins by circular dichroism. *Biochimica et Biophysica Acta - Proteins and Proteomics*. **1751**, pp. 119–139.

Kikhney, A.G., and Svergun, D.I., (2015). A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Letters*. **589**, pp. 2570–2577.

Konarev, P. V, Volkov, V. V, Sokolova, A. V, Koch, M.H.J., and Svergun, D.I., (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography*. **36**, pp. 1277–1282.

Konrat, R., (2014). NMR contributions to structural dynamics studies of intrinsically disordered proteins. *Journal of Magnetic Resonance*. **241**, pp. 74–85.

Kosol, S., Contreras-Martos, S., Cedeno, C., and Tompa, P., (2013). Structural characterization of intrinsically disordered proteins by NMR spectroscopy. *Molecules*. **18**, pp. 10802–10828.

Kovacs, H., Moskau, D., and Spraul, M., (2005). Cryogenically cooled probes - a leap in NMR technology. *Progress in Nuclear Magnetic Resonance Spectroscopy*. **46**, pp. 131–155.

Kozlowski, L.P., and Bujnicki, J.M., (2012). MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins. *BMC Bioinformatics*. **13**, pp. 111.

Kratky, O., and Porod, G., (1949). Diffuse small-angle scattering of X-rays in colloid systems. *Journal of Colloid Science*. **4**, pp. 35–70.

Kremer, A.N., and Hoffmann, H., (2012). Subtractive hybridization yields a silver resistance determinant unique to nosocomial pathogens in the Enterobacter cloacae complex. *Journal of Clinical Microbiology*. **50**, pp. 3249–3257.

Kukic´, P., Farrell, D., Søndergaard, C.R., Bjarnadottir, U., Bradley, J., Pollastri, G., and Nielsen, J.E., (2010). Improving the analysis of NMR spectra tracking pH-induced conformational changes: Removing artefacts of the electric field on the NMR chemical shift. *Proteins: Structure, Function and Bioinformatics*. **78**, pp. 971–984.

Kumbhar, S., Johannsen, S., Sigel, R.K.O., Waller, M.P., and M??ller, J., (2013). A QM/MM refinement of an experimental DNA structure with metal-mediated base pairs. *Journal of Inorganic Biochemistry*. **127**, pp. 203–210.

Kypr, J., and Vorlícková, M., (1986). Graphical analysis of circular dichroic spectra distinguishes between two-state and gradual alterations in DNA conformation. *General Physiology and Biophysics*. **5**, pp. 415–421.

Lansdown, A.B.G., (2010). A pharmacological and toxicological profile of silver as an antimicrobial agent in medical devices. *Advances in Pharmacological Sciences*. **2010**, pp. 1–16.

Lee, R. Van Der, Buljan, M., Lang, B., Weatheritt, R.J., Daughdrill, G.W., Dunker, A.K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D.T., Kim, P.M., Kriwacki, R.W., Old, C.J., Pappu, R. V, Tompa, P., Uversky, V.N., Wright, P.E., and Babu, M.M., (2014). Classification of intrinsically disordered regions and proteins. *Chemical Reviews*. **114**, pp. 6589–6631.

Leekha, S., Terrell, C.L., and Edson, R.S., (2011). General Principles of Antimicrobial Therapy. *Mayo Clinic Proceedings*. **86**, pp. 156–167.

Lefèvre, F., Rémy, M.H., and Masson, J.M., (1997). Alanine-stretch scanning mutagenesis: A simple and efficient method to probe protein structure and function. *Nucleic Acids Research*. **25**, pp. 447–448.

Lemaire, D., Marie, G., Serani, L., and Larprevote, O., (2001). Stabilization of gas-phase noncovlanet macromolecular complexes in electrospray mass spectrometry using aqueous triethylammonium bicarbonate buffer. *Analytical Chemistry*. **73**, pp. 1699–1706.

Lemire, J. a, Harrison, J.J., and Turner, R.J., (2013). Antimicrobial activity of metals: mechanisms, molecular targets and applications. *Nature Reviews. Microbiology*. **11**, pp. 371–384.

Li, H., Lu, Z., Cheng, G., Rong, K., Chen, F., and Chen, R., (2015a).

HEPES-involved hydrothermal synthesis of Fe3O4 nanoparticles and their biological application. *RSC Advances*. **5**, pp. 5059–5067.

Li, J., Feng, Y., Wang, X., Li, J., Liu, W., Rong, L., and Bao, J., (2015b). An overview of predictors for intrinsically disordered proteins over 2010-2014. *International Journal of Molecular Sciences*. **16**, pp. 23446–23462.

Li, X.Z., Nikaido, H., and Williams, K.E., (1997). Silver-resistant mutants of *Escherichia coli* display active efflux of Ag^{+} and are deficient in porins. *Journal of Bacteriology*. **179**, pp. 6127–6132.

Lippa, A.M., and Goulian, M., (2012). Perturbation of the oxidizing environment of the periplasm stimulates the PhoQ/PhoP system in *Escherichia coli*. *Journal of Bacteriology*. **194**, pp. 1457–1463.

Lobley, A., Whitmore, L., and Wallace, B.A., (2002). DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra. *Bioinformatics*. **18**, pp. 211–212.

Loftin, I.R., Franke, S., Blackburn, N.J., and McEvoy, M.M., (2007). Unusual Cu(I)/Ag(I) coordination of *Escherichia coli* CusF as revealed by atomic resolution crystallography and X-ray absorption spectroscopy. *Protein Science : A Publication of the Protein Society*. **16**, pp. 2287–2293.

Loh, J. V., Percival, S.L., Woods, E.J., Williams, N.J., and Cochrane, C.A., (2009). Silver resistance in MRSA isolated from wound and nasal sources in humans and animals. *International Wound Journal*. **6**, pp. 32–38.

Lok, C.N., Ho, C.M., Chen, R., Tam, P.K.H., Chiu, J.F., and Che, C.M., (2008). Proteomic identification of the cus system as a major determinant of constitutive *Escherichia coli* silver resistance of chromosomal origin. *Journal of Proteome Research*. **7**, pp. 2351–2356.

Long, F., Su, C., Zimmermann, M.T., Boyken, S.E., Rajashankar, K.R., Jernigan, R.L., and Yu, E.W., (2010). Crystal structures of the CusA efflux pump suggest methionine-mediatedmetal transport. *Nature*. **467**, pp. 484–488.

Loo, C.Y., Rohanizadeh, R., Young, P.M., Traini, D., Cavaliere, R., Whitchurch, C.B., and Lee, W.H., (2015). Combination of silver nanoparticles and curcumin nanoparticles for enhanced anti-biofilm activities. *Journal of Agricultural and Food Chemistry*. **64**, pp. 2513–2522.

Luo, H., and Nijveen, H., (2014). Understanding and identifying amino acid repeats. *Briefings in Bioinformatics*. **15**, pp. 582–591.

Luo, W.-G., Liu, H.-Z., Lin, W.-H., Kabir, M.H., and Su, Y., (2013). Simultaneous splicing of multiple DNA fragments in one PCR reaction. *Biological Procedures Online*. **15**, pp. 1–9.

Mamedov, T.G., Pienaar, E., Whitney, S.E., Termaat, J.R., Carvill, G., Goliath, R., Subramanian, A., and Viljoen, H.J., (2008). A fundamental study of the PCR amplification of GC-rich DNA templates. *Computational Biology and Chemistry*. **32**, pp. 452–457.

Marambio-Jones, C., and Hoek, E., (2010). A review of the antibacterial effects of silver nanomaterials and potential implications for human health and the environment. *Journal of Nanoparticle Research*. **12**, pp. 1531-1551.

Markoulatos, P., Siafakas, N., and Moncany, M., (2002). Multiplex polymerase chain reaction: a practical approach. *Journal of Clinical Laboratory Analysis*. **16**, pp. 47–51.

Marks and Spencer plc, (2013). *7 pairs of cotton rich freshfeet™ socks with silver technology*, [Online]. Available at: http://www.marksandspencer.com/Pairs-Cotton-Freshfeet-Silver-Technology/dp/B001CKAFLG, [Accessed on 1 April 2013].

Marley, J., Lu, M., and Bracken, C., (2001). A method for efficient isotopic labeling of recombinant proteins. *Journal of Biomolelcular NMR*. **20**, pp. 71–75.

Martin, S.R., and Schilstra, M.J., (2008). Circular dichroism and its application to the study of biomolecules. *Methods in Cell Biology*. **84**, pp. 263–293.

McArthur, A.G., Waglechner, N., Nizam, F., Yan, A., Azad, M.A., Baylay, A.J., Bhullar, K., Canova, M.J., De Pascale, G., Ejim, L., Kalan, L., King, A.M., Koteva, K., Morar, M., Mulvey, M.R., O'Brien, J.S., Pawlowski, A.C., Piddock, L.J. V, Spanogiannopoulos, P., Sutherland, A.D., Tang, I., Taylor, P.L., Thaker, M., Wang, W., Yan, M., Yu, T., and Wright, G.D., (2013). The comprehensive antibiotic resistance database. *Antimicrobial Agents and Chemotherapy*. **57**, pp. 3348–3357.

Mcdonnell, G., and Russell, A.D., (1999). Antiseptics and disinfectants: Activity, action, and resistance. *Clinical Microbiology Reviews*. **12**, pp. 147–179.

McHugh, G.L., Moellering, R.C., Hopkins, C.C., and Swartz, M.N., (1975). Salmonella typhimurium resistant to silver nitrate, chloramphenicol, and ampicillin. *Lancet*. **1**, pp. 235–240.

McKown, R.L., Orle, K. a., Chen, T., and Craig, N.L., (1988). Sequence requirements of *Escherichia coli att*Tn7, a specific site of transposon Tn7 insertion. *Journal of Bacteriology*. **170**, pp. 352–358.

McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y.M., Buso, N., Cowley, A.P., and Lopez, R., (2013). Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Research*. **41**, pp. W597–600.

Mellata, M., Touchman, J.W., and Curtis, R., (2009). Full sequence and comparative analysis of the plasmid pAPEC-1 of avian pathogenic *E. coli* χ7122 (O78:K80:H9). *PLoS ONE*. **4**, pp. e4232.

Mesu, J.G., Beale, A.M., De Groot, F.M.F., and Weckhuysen, B.M., (2006). Probing the influence of X-rays on aqueous copper solutions using time-resolved in situ combined video/X-ray absorption near-edge/ultraviolet-visible spectroscopy. *Journal of Physical Chemistry B*. **110**, pp. 17671–17677.

Meyer, N.H., and Zangger, K., (2014). Boosting the resolution of 1H NMR spectra by homonuclear broadband decoupling. *ChemPhysChem*. **15**, pp. 49–55.

Micsonai, A., Wien, F., Kernya, L., Lee, Y.-H., Goto, Y., Réfrégiers, M., and Kardos, J., (2015). Accurate secondary structure prediction and fold

recognition for circular dichroism spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America*. **112**, pp. E3095–3103.

Mijnendonckx, K., Leys, N., Mahillon, J., Silver, S., and Van Houdt, R., (2013). Antimicrobial silver: Uses, toxicity and potential for resistance. *BioMetals*. **26**, pp. 609–621.

Miles, A.J., Whitmore, L., and Wallace, B. A, (2005). Spectral magnitude effects on the analyses of secondary structure from circular dichroism spectroscopic data. *Protein Science : A Publication of the Protein Society*. **14**, pp. 368–374.

Mirolo, L., Schmidt, T., Eckhardt, S., Meuwly, M., and Fromm, K.M., (2013). PH-dependent coordination of AgI ions by histidine: Experiment, theory, and a model for SilE. *Chemistry - A European Journal*. **19**, pp. 1754–1761.

Monastyrskyy, B., Fidelis, K., Moult, J., Tramontano, A., and Kryshtafovych, A., (2011). Evaluation of disorder predictions in CASP9. *Proteins: Structure, Function and Bioinformatics*. **79**, pp. 107–118.

Monastyrskyy, B., Kryshtafovych, A., Moult, J., Tramontano, A., and Fidelis, K., (2014). Assessment of protein disorder region predictions in CASP10. *Proteins: Structure, Function and Bioinformatics*. **82**, pp. 127–137.

Monchy, S., Benotmane, M.A., Wattiez, R., van Aelst, S., Auquier, V., Borremans, B., Mergeay, M., Taghavi, S., van der Lelie, D., and Vallaeys, T., (2006). Transcriptomic and proteomic analyses of the pMOL30-encoded copper resistance in Cupriavidus metallidurans strain CH34. *Microbiology*. **152**, pp. 1765–1776.

Morones-Ramirez, J.R., Winkler, J.A., Spina, C.S., and Collins, J.J., (2013). Silver enhances antibiotic activity against gram-negative bacteria. *Science Translational Medicine*. **5**, pp. 190ra81.

Morrison, K.L., and Weiss, G.A., (2001). Combinatorial alanine-scanning. *Current Opinion in Chemical Biology*. **5**, pp. 302–307.

Muller, J., Bohme, D., Lax, P., Cerda, M.M., and Roitzsch, M., (2005). Metal ion coordination to azole nucleosides. *Chemistry - A European Journal*. **11**, pp. 6246–6253.

Munson, G.P., Lam, D.L., Outten, F.W., and Halloran, T.V.O., (2000). Identification of a copper-responsive two-component system on the chromosome of *Escherichia coli* K-12. *Journal of Bacteriology*. **182**, pp. 5864–5871.

Noivirt-Brik, O., Prilusky, J., and Sussman, J.L., (2009). Assessment of disorder predictions in CASP8. *Proteins: Structure, Function and Bioinformatics*. **77**, pp. 210–216.

Osman, D., and Cavet, J.S., (2008). Copper Homeostasis in Bacteria. *Advances in Applied Microbiology*. **65**, pp. 217–247.

Osman, D., and Cavet, J.S., (2010). Bacterial metal-sensing proteins exemplified by ArsR-SmtB family repressors. *Natural Product Reports*. **27**,

pp. 668–680.

Osman, D., Patterson, C.J., Bailey, K., Fisher, K., Robinson, N.J., Rigby, S.E.J., and Cavet, J.S., (2013). The copper supply pathway to a Salmonella Cu,Zn-superoxide dismutase (SodCII) involves P1B-type ATPase copper efflux and periplasmic CueP. *Molecular Microbiology*. **87**, pp. 466–477.

OPPF-UK, (2013). *Oxford Protein Production Facility UK – Division of Structural Biology*, [Online]. Available at: http://www.oppf.ox.ac.uk/OPPF/, [Accessed on 1 June 2012].

Oxford Dictionaries, (2011). *Concise Oxford English Dictionary: Main Edition*. Oxford University Press, Oxford, UK.

Page, K., Wilson, M., and Parkin, I., (2009). Antimicrobial surfaces and their potential in reducing the role of the inanimate environment in the incidence of hospital-acquired infections. *Journal of Materials Chemistry*. **19**, pp. 3819–3831.

Page, R., Peti, W., Wilson, I. a, Stevens, R.C., and Wüthrich, K., (2005). NMR screening and crystal quality of bacterially expressed prokaryotic and eukaryotic proteins in a structural genomics pipeline. *Proceedings of the National Academy of Sciences of the United States of America*. **102**, pp. 1901–1905.

Pal, C., Bengtsson-Palme, J., Rensing, C., Kristiansson, E., and Larsson, D.G.J., (2014). BacMet: Antibacterial biocide and metal resistance genes database. *Nucleic Acids Research*. **42**, pp. 737–743.

Park, H., Kim, J.J.Y., Kim, J.J.Y., Lee, J., Hahn, J.-S., Gu, M.B., Yoon, J., Yeon, J., Kim, J.J.Y., Lee, J., Hahn, J.-S., Bock, M., and Yoon, J., (2009). Silver-ion-mediated reactive oxygen species generation affecting bactericidal activity. *Water Research*. **43**, pp. 1027–1032.

Patterson, M. J., (2016) *Solubility of silver compounds in water*, [Online]. Available at:
http://www.saltlakemetals.com/Solubility_Of_Silver_Compounds.htm, [Accessed 16 April 2016].

Pavlov, M.Y., Watts, R.E., Tan, Z., Cornish, V.W., Ehrenberg, M., and Forster, A.C., (2009). Slow peptide bond formation by proline and other N-alkylamino acids in translation. *Proceedings of the National Academy of Sciences of the United States of America*. **106**, pp. 50–54.

Pearson, G., (1963). Hard and Soft Acids and Bases. *Journal of the American Chemical Society*. **85**, pp. 3533–3539.

Percival, S.L., Bowler, P.G., and Russell, D., (2005). Bacterial resistance to silver in wound care. *Journal of Hospital Infection*. **60**, pp. 1–7.

Pervushin, K., Riek, R., Wider, G., and Wüthrich, K., (1997). Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proceedings of the National Academy of Sciences of the United States of America*. **94**, pp. 12366–12371.

Van Petegem, F., Duderstadt, K.E., Clark, K.A., Wang, M., and Minor, D.L., (2008). Alanine-scanning mutagenesis defines a conserved energetic

hotspot in the CaValpha1 AID-CaVbeta interaction site that is critical for channel modulation. *Structure*. **16**, pp. 280–294.

Peters, J.E., and Craig, N.L., (2001). Tn7: smarter than we thought. *Nature Reviews. Molecular Cell Biology*. **2**, pp. 806–814.

Peters, J.E., Fricker, A.D., Kapili, B.J., and Petassi, M.T., (2014). Heteromeric transposase elements: Generators of genomic Islands across diverse bacteria. *Molecular Microbiology*. **93**, pp. 1084–1092.

Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A.G., Gajda, M., Gorba, C., Mertens, H.D.T., Konarev, P. V., and Svergun, D.I., (2012). New developments in the ATSAS program package for small-angle scattering data analysis. *Journal of Applied Crystallography*. **45**, pp. 342–350.

Petrovec, K., Ravoo, B.J., and Müller, J., (2012). Cooperative formation of silver(i)-mediated base pairs. *Chemical Communications*. **48**, pp. 11844–11846.

Pineda-Sanabria, S.E., Robertson, I.M., Li, M.X., and Sykes, B.D., (2013). Interaction between the regulatory domain of cardiac troponin C and the acidosis-resistant cardiac troponin i A162H. *Cardiovascular Research*. **97**, pp. 481–489.

Platzer, G., Okon, M., and McIntosh, L.P., (2014). PH-dependent random coil 1H, 13C, and 15N chemical shifts of the ionizable amino acids: A guide for protein pKa measurements. *Journal of Biomolecular NMR*. **60**, pp. 109–129.

Provencher, S.W., and Glockner, J., (1981). Estimation of globular protein secondary structure from circular dichroism. *Biochemistry*. **20**, pp. 33–37.

Putnam, C.D., Hammel, M., Hura, G.L., and Tainer, J. a, (2007). X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quarterly Reviews of Biophysics*. **40**, pp. 191–285.

Radivojac, P., Iakoucheva, L.M., Oldfield, C.J., Obradovic, Z., Uversky, V.N., and Dunker, A.K., (2007). Intrinsic disorder and functional proteomics. *Biophysical Journal*. **92**, pp. 1439–1456.

Rambo, R.P., and Tainer, J.A., (2011). Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers*. **95**, pp. 559–571.

Randall, C.P., Gupta, A., Jackson, N., Busse, D., and O'Neill, A.J., (2015). Silver resistance in Gram-negative bacteria: A dissection of endogenous and exogenous mechanisms. *Journal of Antimicrobial Chemotherapy*. **70**, pp. 1037–1046.

Receveur-Brechot, V., and Durand, D., (2012). How random are intrinsically disordered proteins? A small angle scattering perspective. *Current Protein & Peptide Science*. **13**, pp. 55–75.

Receveur-Bréhot, V., Bourhis, J.M., Uversky, V.N., Canard, B., and

Longhi, S., (2006). Assessing protein disorder and induced folding. *Proteins: Structure, Function and Genetics*. **62**, pp. 24–45.

Remita, S., Fontaine, P., Lacaze, E., Borensztein, Y., Sellame, H., Farha, R., Rochas, C., and Goldmann, M., (2007). X-ray radiolysis induced formation of silver nano-particles: A SAXS and UV-visible absorption spectroscopy study. *Nuclear Instruments and Methods in Physics Research, Section B: Beam Interactions with Materials and Atoms*. **263**, pp. 436–440.

Rhodes, D.G., Bossio, R.E., and Laue, T.M., (2009). Determination of size, molecular weight, and presence of subunits. *Methods in Enzymology*. **463**, pp. 691–723.

Riddle, P., (2013). pH meters and their electrodes: calibration, maintenance and use. *The Biomedical Scientist*. **April**, pp. 202–205.

Rissman, A.I., Mau, B., Biehl, B.S., Darling, A.E., Glasner, J.D., and Perna, N.T., (2009). Reordering contigs of draft genomes using the Mauve aligner. *Bioinformatics*. **25**, pp. 2071–2073.

Rodger, A., Marrington, R., Roper, D., and Windsor, S., (2005). Circular dichroism spectroscopy for the study of protein-ligand interactions. *Methods in Molecular Biology*. **305**, pp. 343–363.

Romero, P., Obradovic, Z., Kissinger, C.R., Villafranca, J.E., Garner, E., Guilliot, S., and Dunker, a K., (1998). Thousands of proteins likely to have long disordered regions. *Pacific Symposium on Biocomputing*. **3**, pp. 437–448.

Round, A., Felisaz, F., Fodinger, L., Gobbo, A., Huet, J., Villard, C., Blanchet, C.E., Pernot, P., McSweeney, S., Roessle, M., Svergun, D.I., and Cipriani, F., (2015). BioSAXS Sample Changer: A robotic sample changer for rapid and reliable high-throughput X-ray solution scattering experiments. *Acta Crystallographica Section D: Biological Crystallography*. **71**, pp. 67–75.

Rubino, J.T., Chenkin, M.P., Keller, M., Riggs-gelasco, P., and Franz, K.J., (2011). A comparison of methionine , histidine and cysteine in copper (I) -binding peptides reveals differences relevant to copper uptake by organisms in diverse environments. *Metallomics*. **3**, pp. 61–73.

Rubino, J.T., Riggs-Gelasco, P., and Franz, K.J., (2010). Methionine motifs of copper transport proteins provide general and flexible thioether-only binding sites for Cu(I) and Ag(I). *Journal of Biological Inorganic Chemistry*. **15**, pp. 1033–1049.

Di Russo, N. V., Estrin, D.A., Marti, M.A., and Roitberg, A.E., (2012). pH-Dependent conformational changes in proteins and their effect on experimental pK(a)s: The case of Nitrophorin 4. *PLoS Computational Biology*. **8**, pp. e1002761.

Saengmee-anupharb, S., Srikhirin, T., Thaweboon, B., Thaweboon, S., Amornsakchai, T., Dechkunakorn, S., and Suddhasthira, T., (2013). Antimicrobial effects of silver zeolite, silver zirconium phosphate silicate and silver zirconium phosphate against oral microorganisms. *Asian Pacific Journal of Tropical Biomedicine*. **3**, pp. 47–52.

Sandegren, L., Linkevicius, M., Lytsy, B., Melhus, Å., and Andersson, D.I., (2012). Transfer of an *Escherichia coli* ST131 multiresistance cassette has created a *Klebsiella pneumoniae*-specific plasmid associated with a major nosocomial outbreak. *Journal of Antimicrobial Chemotherapy*. **67**, pp. 74–83.

Schlegel, S., Rujas, E., Ytterberg, A.J., Zubarev, R. a, Luirink, J., and de Gier, J.-W., (2013). Optimizing heterologous protein production in the periplasm of *E. coli* by regulating gene expression levels. *Microbial Cell Factories*. **12**, pp. 24.

Schrödinger, LLC, (2015). The {PyMOL} Molecular Graphics System, Version~1.8.

Scopes, R.K., (1974). Measurement of protein by spectrophotometry at 205 nm. *Analytical Biochemistry*. **59**, pp. 277–282.

Scott, W.G., Finch, J.T., Grenfell, R., Fogg, J., Smith, T., Gait, M.J., and Klug, A., (1995). Rapid crystallization of chemically synthesized hammerhead RNAs using a double screening procedure. *Journal of Molecular Biology*. **250**, pp. 327–332.

Scrofani, S.D., Wright, P.E., and Dyson, H.J., (1998). The identification of metal-binding ligand residues in metalloproteins using nuclear magnetic resonance spectroscopy. *Protein Science : A publication of the Protein Society*. **7**, pp. 2476–2479.

Seidell, A. and Linke, W. F., (1919). *Solubilities of inorganic and organic compounds* (2nd edn.) [e-book], D. Van Nostrand Company, New York, US, pp. 617–619. Available at: http://booksnow1.scholarsportal.info/ebooks/oca1/2/2ndsolubilitieso00seiduoft/2ndsolubilitieso00seiduoft_bw.pdf, [Accessed 16 April 2016].

Shapiro, D.B., Goldbeck, R. a, Che, D., Esquerra, R.M., Paquette, S.J., and Kliger, D.S., (1995). Nanosecond optical rotatory dispersion spectroscopy: Application to photolyzed hemoglobin-CO kinetics. *Biophysical Journal*. **68**, pp. 326–334.

Sibille, N., and Bernadó, P., (2012). Structural characterization of intrinsically disordered proteins by the combined use of NMR and SAXS. *Biochemical Society Transactions*. **40**, pp. 955–962.

Sickmeier, M., Hamilton, J.A., LeGall, T., Vacic, V., Cortese, M.S., Tantos, A., Szabo, B., Tompa, P., Chen, J., Uversky, V.N., Obradovic, Z., and Dunker, A.K., (2007). DisProt: The database of disordered proteins. *Nucleic Acids Research*. **35**, pp. 786–793.

Siedenbiedel, F., and Tiller, J.C., (2012). Antimicrobial polymers in solution and on surfaces: Overview and functional principles. *Polymers*. **4**, pp. 46–71.

Sigma-Aldrich, (2000). EZview Red HIS-Select HC Nickel Affinity Gel. *Sigma-Aldrich, Inc.*, **E3528**, p. 1-6.

Silhavy, T.J., Kahne, D., and Walker, S., (2010). The bacterial cell envelope. *Cold Spring Harbor Perspectives in Biology*. **2**, pp. 1–16.

Silva, K.I., Michael, B.C., Geib, S.J., and Saxena, S., (2014). ESEEM

analysis of multi-histidine Cu(II)-coordination in model complexes, peptides, and amyloid-β. *The Journal of Physical Chemistry. B.*. **118**, pp. 8935–8944.

Silver, S., (2003). Bacterial silver resistance: Molecular biology and uses and misuses of silver compounds. *FEMS Microbiology Reviews*. **27**, pp. 341–353.

Silver, S., Gupta, A., Matsui, K., and Lo, J.-F., (1999). Resistance to Ag(I) cations in bacteria: Environments, genes and proteins. *Metal-based Drugs*. **6**, pp. 315–320.

Silvestre, C., Duraccio, D., and Cimmino, S., (2011). Food packaging based on polymer nanomaterials. *Progress in Polymer Science*. **36**, pp. 1766–1782.

Singh, S.K., Roberts, S.A., McDevitt, S.F., Weichsel, A., Wildner, G.F., Grass, G.B., Rensing, C., and Montfort, W.R., (2011). Crystal structures of multicopper oxidase CueO bound to copper(I) and silver(I): Functional role of a methionine-rich sequence. *Journal of Biological Chemistry*. **286**, pp. 37849–37857.

So, M.H., Ho, C.M., Chen, R., and Che, C.M., (2010). Hydrothermal synthesis of platinum-group-metal nanoparticles by using HEPES as a reductant and stabilizer. *Chemistry - An Asian Journal*. **5**, pp. 1322–1331.

Sreerama, N., Venyaminov, S.Y., and Woody, R.W., (1999). Estimation of the number of alpha-helical and beta-strand segments in proteins using circular dichroism spectroscopy. *Protein Science: A Publication of the Protein Society*. **8**, pp. 370–380.

Sreerama, N., and Woody, R.W., (1993). A self-consistent method for the analysis of protein secondary structure from circular dichroism. *Analytical Biochemistry*. **209**, pp. 32–44.

Sreerama, N., and Woody, R.W., (2000). Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Analytical Biochemistry*. **287**, pp. 252–260.

Sreerama, N., and Woody, R.W., (2004). On the analysis of membrane protein circular dichroism spectra. *Methods in Enzymology*. **13**, pp. 318–351.

SSRL (Stanford Synchrotron Radiation Lightsource), (2014). *Initial Analysis and Quality Assessment of Solution Scattering Data*. [Online]. Available at:
https://www-ssrl.slac.stanford.edu/~saxs/analysis/assessment.htm,
[Accessed on 1 June 2015].

Staehlin, B.M., Gibbons, J.G., Rokas, A., O'Halloran, T. V, and Slot, J.C., (2016). Evolution of a Heavy Metal Homeostasis/Resistance Island Reflects Increasing Copper Stress in Enterobacteria. *Genome Biology and Evolution*. **8**, pp. 811–826.

Stock, J.B., Rauch, B., and Roseman, S., (1977). Periplasmic space in Salmonella typhimurium and *E. coli*. *The Journal of Biological Chemistry*.

252, pp. 7850–7861.

Van Stokkum, I.H., Spoelder, H.J., Bloemendal, M., van Grondelle, R., and Groen, F.C., (1990). Estimation of protein secondary structure and error analysis from circular dichroism spectra. *Analytical Biochemistry*. **191**, pp. 110–118.

Stoscheck, C.M., (1990). Guide to protein purification. *Methods in Enzymology*. **182**, pp. 50–68.

Stumpf, V., Gokhberg, K., and Cederbaum, L.S., (2016). The role of metal ions in X-ray-induced photochemistry. *Nature Chemistry*. **8**, pp. 1–5.

Sun, R.W.-Y., Chen, R., Chung, N.P.-Y.P.-Y., Ho, C., Lin, C.-L.S., and Che, C., (2005). Silver nanoparticles fabricated in Hepes buffer exhibit cytoprotective activities toward HIV-1 infected cells. *Chemical Communications*. **1**, pp. 5059–5061.

Svergun, D.I., (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of Applied Crystallography*. **25**, pp. 495–503.

Svergun, D.I., (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophysical Journal*. **76**, pp. 2879–2886.

Svergun, D.I., Petoukhov, M. V, and Koch, M.H., (2001). Determination of domain structure of proteins from X-ray solution scattering. *Biophysical Journal*. **80**, pp. 2946–2953.

Tan, S., Erol, M., Attygalle, A., Du, H., and Sukhishvili, S., (2007). Synthesis of positively charged silver nanoparticles via photoreduction of AgNO3 in branched polyethyleneimine/HEPES solutions. *Langmuir*. **23**, pp. 9836–9843.

Tanaka, K., (2003). The origin of macromolecule ionization by laser irradiation (Nobel lecture). *Angewandte Chemie (International ed. in English)*. **42**, pp. 3860–3870.

Tanaka, T., Mizuno, T., Fukui, S., Hiroaki, H., Oku, J., Kanaori, K., Tajima, K., and Shirakawa, M., (2004). Two-metal ion , Ni(II) and Cu(II), binding alpha-helical oiled coil peptide. *Journal of the American Chemical Society*. **126**, pp. 14023–14028.

Tetaz, T.J., and Luke, R.K.J., (1983). Plasmid-controlled resistance to copper in *Escherichia coli*. *Journal of Bacteriology*. **154**, pp. 1263–1268.

Theisen, A., Johann, C., Deacon, A.P., and Harding, S.E., (2000) *Refractive increment data-book for polymer and biomolecular scientists*. (1st ed.), Nottingham University Press, pp. 5-45.

Thomas, R., (2001). A beginner's guide to ICP-MS Part I. *Spectroscopy*. **16**, pp. 38–42.

Tlaskalová-Hogenová, H., Stěpánková, R., Kozáková, H., Hudcovic, T., Vannucci, L., Tučková, L., Rossmann, P., Hrnčíř, T., Kverka, M., Zákostelská, Z., Klimešová, K., Přibylová, J., Bártová, J., Sanchez, D., Fundová, P., Borovská, D., Srůtková, D., Zídek, Z., Schwarzer, M.,

Drastich, P., and Funda, D.P., (2011). The role of gut microbiota (commensal bacteria) and the mucosal barrier in the pathogenesis of inflammatory and autoimmune diseases and cancer: contribution of germ-free and gnotobiotic animal models of human diseases. *Cellular & Molecular Immunology*. **8**, pp. 110–120.

Todd, R.J., Van Dam, M.E., Casimiro, D., Haymore, B.L., and Arnold, F.H., (1991). Cu(II)-binding properties of a cytochrome c with a synthetic metal-binding site: His-X3-His in an alpha-helix. *Proteins*. **10**, pp. 156–61.

Tompa, P., (2002). Intrinsically unstructured proteins. *Trends in Biochemical Sciences*. **27**, pp. 527–533.

Ucisik, M.N., Chakravorty, D.K., and Merz, K.M., (2013). Structure and dynamics of the N-terminal domain of the Cu(I) binding protein CusB. *Biochemistry*. **52**, pp. 6911–6923.

Uversky, V.N., (2002). Natively unfolded proteins: a point where biology waits for physics. *Protein Science: A Publication of the Protein Society*. **11**, pp. 739–756.

Uversky, V.N., (2002). What does it mean to be natively unfolded? *European Journal of Biochemistry*. **269**, pp. 2–12.

Uversky, V.N., (2013a). A decade and a half of protein intrinsic disorder: Biology still waits for physics. *Protein Science*. **22**, pp. 693–724.

Uversky, V.N., (2013b). Unusual biophysics of intrinsically disordered proteins. *Biochimica et Biophysica Acta - Proteins and Proteomics*. **1834**, pp. 932–951.

Uversky, V.N., (2015). Functional roles of transiently and intrinsically disordered regions within proteins. *The FEBS Journal*. **282**, pp. 1182–1189.

Uversky, V.N., and Dunker, A.K., (2010). Understanding protein non-folding. *Biochimica et Biophysica Acta - Proteins and Proteomics*. **1804**, pp. 1231–1264.

Uversky, V.N., Gillespie, J.R., and Fink, A.L., (2000). Why are 'natively unfolded' proteins unstructured under physiologic conditions? *Proteins*. **41**, pp. 415–427.

Vihinen, M., Torkkila, E., and Riikonen, P., (1994). Accuracy of protein flexibility predictions. *Proteins: Structure, Function and Genetics*. **19**, pp. 141–149.

Waddell, C.S., and Craig, N.L., (1989). Tn7 transposition: recognition of the attTn7 target sequence. *Proceedings of the National Academy of Sciences of the United States of America*. **86**, pp. 3958–3962.

Waldron, K.J., and Robinson, N.J., (2009). How do bacterial cells ensure that metalloproteins get the correct metal? *Nature Reviews. Microbiology*. **7**, pp. 25–35.

Wallace, B., Lees, J., and Orry, A., (2003). Analyses of circular dichroism spectra of membrane proteins. *Protein Science: A Publication of the Protein Society*. **12**, pp. 875–884.

Wan, H., Li, Y., Fan, Y., Meng, F., Chen, C., and Zhou, Q., (2012). A

site-directed mutagenesis method particularly useful for creating otherwise difficult-to-make mutants and alanine scanning. *Analytical Biochemistry*. **420**, pp. 163–170.

Wataha, J.C., and Hanks, C.T., (1996). Biological effects of palladium and risk of using palladium in dental casting alloys. *Journal of Oral Rehabilitation*. **23**, pp. 309–320.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J., (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. **25**, pp. 1189–1191.

Webb, B., and Sali, A., (2014). Comparative protein structure modeling using MODELLER. *Current Protocols in Bioinformatics*. **47**, pp. 1–32.

Weiss, G.A., Watanabe, C.K., Zhong, A., Goddard, A., and Sidhu, S.S., (2000). Rapid mapping of protein functional epitopes by combinatorial alanine scanning. *Proceedings of the National Academy of Sciences*. **97**, pp. 8950–8954.

Whitmore, L., and Wallace, B.A., (2004). DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Research*. **32**, pp. 668–673.

Whitmore, L., and Wallace, B.A., (2008). Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases. *Biopolymers*. **89**, pp. 392–400.

World Health Organisation, (2014). *Antimicrobial resistance: Global surveillance report*. World Health Organization, Geneva, Switzerland, pp. 12-27.

Wilks, J.C., and Slonczewski, J.L., (2007). pH of the cytoplasm and periplasm of *Escherichia coli*: Rapid measurement by green fluorescent protein fluorimetry. *Journal of Bacteriology*. **189**, pp. 5601–5607.

Williamson, M.P., (2013). Using chemical shift perturbation to characterise ligand binding. *Progress in Nuclear Magnetic Resonance Spectroscopy*. **73**, pp. 1–16.

Wilson, J.W., Schurr, M.J., LeBlanc, C.L., Ramamurthy, R., Buchanan, K.L., and Nickerson, C.A., (2002). Mechanisms of bacterial pathogenicity. *Postgraduate Medical Journal*. **78**, pp. 216–224.

Wittwer, M., and Dames, S.A., (2015). Expression and purification of the natively disordered and redox sensitive metal binding regions of Mycobacterium tuberculosis protein kinase G. *Protein Expression and Purification*. **111**, pp. 68–74.

Wright, P.E., and Dyson, H.J., (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *Journal of Molecular Biology*. **293**, pp. 321–331.

Wu, S., and Zhang, Y., (2008). MUSTER: Improving protein sequence profile-profile alignments by using multiple sources of structure information. *Proteins: Structure, Function and Genetics*. **72**, pp. 547–556.

Wüthrich, K., (1990). Protein structure determination in solution by

NMR spectroscopy. *Journal of Biological Chemistry*. **265**, pp. 22059–22062.

Wüthrich, K., (2001). The way to NMR structures of proteins. *Nature Structural Biology*. **8**, pp. 923–925.

Xie, Z. R. and Hwang, M. J., (2015) Methods for Predicting Protein–Ligand Binding Sites, in: Kukol, A., Ed. *Molecular Modelling of Proteins* (2nd ed.), Methods in Molecular Biology (1215), Springer Protocols, Humana Press, Hertfordshire, UK, pp. 383-398.

Yamashita, M.M., Wesson, L., Eisenmant, G., and Eisenberg, D., (1990). Where metal ions bind in proteins. *Proceedings of the National Academy of Sciences of the United States of America*. **87**, pp. 5648–5652.

Yang, J., Roy, A., and Zhang, Y., (2013a). BioLiP: A semi-manually curated database for biologically relevant ligand-protein interactions. *Nucleic Acids Research*. **41**, pp. 1096–1103.

Yang, J., Roy, A., and Zhang, Y., (2013b). Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics*. **29**, pp. 2588–2595.

Yi, H., Cho, Y.J., Yong, D., and Chun, J., (2012). Genome sequence of *Escherichia coli* J53, a reference strain for genetic studies. *Journal of Bacteriology*. **194**, pp. 3742–3743.

Zhu, Y.-C., Wang, E.-Q., Ma, G.-L., Kang, Y.-B., Zhao, L.-H., and Liu, Y.-Z., (2014). Interaction of C-terminal metal-binding domain of copper. *Acta Physico-Chimica Sinica*. **30**, pp. 1–7.

Ziani, W., Maillard, A.P., Petit-Härtlein, I., Garnier, N., Crouzy, S., Girard, E., and Covès, J., (2014). The X-ray structure of NccX from Cupriavidus metallidurans 31A illustrates potential dangers of detergent solubilization when generating and interpreting crystal structures of membrane proteins. *The Journal of Biological Chemistry*. **289**, pp. 31160–31172.

Zimmermann, M., Udagedara, S.R., Sze, C.M., Ryan, T.M., Howlett, G.J., Xiao, Z., and Wedd, A.G., (2012). PcoE - A metal sponge expressed to the periplasm of copper resistance *Escherichia coli*. Implication of its function role in copper resistance. *Journal of Inorganic Biochemistry*. **115**, pp. 186–197.

Appendix A: Antibiotics

A.1 Ampicillin (Amp) -

Stock solution of Amp at 100 mg mL^{-1} in dH_2O was prepared and filter sterilised using a 0.2 µm membrane filter in a syringe. This was stored at -20 ^oC and added, as required, to culture media to a final concentration of 100 µg mL^{-1}.

A.2 Carbenicillin (Cb) -

Stock solution of Cb at 50mg mL^{-1} in dH_2O was prepared and filter sterilised using a 0.2 µm membrane filter in a syringe and stored -20 ^oC. Cb was used, in culture media, at a final concentration of 50 µg mL^{-1} whenever needed.

A.3 Chloramphenicol (Cam) -

Stock solution of Cam at 34mg mL^{-1} in 70% EtOH was filter sterilised using a 0.2 µm membrane filter in a syringe and stored at -20 ^oC. Cam was added to culture media to a final concentration of 34 µg mL^{-1}, when necessary.

Appendix B: Periodic table of elements

Transition metals (d-block) are highlighted yellow. Image used with copyright permissions from – Helmenstine, T., (2015). *Color Periodic Table With Electron Configurations – 2015*, [Online]. Available at: http://sciencenotes.org/color-periodic-table-with-electron-configurations-2015/, [Accessed on 1 July 2016].

