

University of Glasgow

Structural and Functional Characterisation of the Protein Targets of the Anti-Virulence Compounds, the Salicylidene Acylhydrazides

A thesis submitted to the University of Glasgow for the degree of Doctor of Philosophy

Katherine S. H. Beckham BSc (Hons), MRes

Submitted February 2014

Institute of Infection, Immunity and Inflammation

College of Medical, Veterinary and Life Sciences

University of Glasgow

Acknowledgements

Firstly - a big thank you to the best supervisors a girl could ask for: Dr Andrew Roe and Dr Olwyn Byron. Thank you so much for all the help, support, encouragement and patience you have both offered me over the past four years. Andy, thanks for always being positive and enthusiastic even when three, or more, weeks later my cloning still hasn't worked... your friendly approach to supervision makes the Roe group a great team to be a part of. Olwyn, thanks for not minding my constant interruptions (even when your office door is shut) and always answering your e-mails even if I'm just sending cute pictures of pugs. You're always there to help even when it means staying awake for over forty hours to help me re-analyse data for my talk in Japan.

Next, I want to thank all the members of the Roe group, past and present. Having great people in the lab makes all the difference. I'll start the long list with Dr Mads Gabrielsen, 'my' first post doc. Mads, thanks for looking after me during my MRes rotation and the first few years of PhD before you moved onto pastures new, and then again at the end by proofreading this thesis. You really taught me a lot about proteins, crystallography and geeky science T-shirts - who knew one man could own so many? Next, I'd like to thank Dr Jayde Gawthorne. Despite declining a ticket to go and see Kylie in concert, we became good friends and you made the lab a really fun (and productive - Andy) place to be. A big thanks to my PhD buddy James Connolly. James you've been here through the PhD rollercoaster of emotions, thanks all your support and your help over the years. Finally, thanks to the recent Roe group additions Tom, Riccardo, Zoe and Claire.

Over the past four years, I have worked with numerous people outside of the lab who have helped me with experiments and offered advice that has been invaluable. This is by no means an exhaustive list but I would like to thank: Dr Dai Wang, Dr Brian Smith, Dr Gill Douce, Dr Isobel Norville, Prof Dave Gally, Dr Amin Tahoun, Dr Rudi Marquez, Dr Karen McLuskey and Prof. Allan Mowatt. Thank you to my PhD assessors Dr Richard Burchmore and Dr Dan Walker for your helpful advice.

I would like to thank the Wellcome Trust for funding my PhD. I would especially like to thank the organisers of the Wellcome Trust Programme at the University of Glasgow;

Prof. Darren Monkton and Dr Olwyn Byron. I really value having been a part of this fantastic PhD programme. I think it has been an outstanding experience and I feel that it has better prepared me for a career in science.

Of course I did have a life outside of work (... sometimes). So I'd like to say a special thanks to my long-suffering friends. To my housemates David and Charlie, it was great living with you both and having you to complain to after a rubbish day. Charlie, an extra special thank you to you for helping me through all the dramas and not making me have 'Two for Tuesday' Dominoes deals by myself. Miss Pamela, thanks for being there over the years and for letting me pretend to be your stylist on shopping trips. Thanks to my even longer suffering friends Rachel and Seb. Rach, thanks for all the proofreading you've been conscripted to do and thanks for having me back to St Andrews to hang out with you and Malfi. Seb, thanks for always being there for my irrational stresses and for always knowing how to cheer me up. Finally, thanks Kesha for all your support over the past few years, you've helped me keep it all together.

Finally, I'd like to thank my family for all their support over the years. Ma, Abby, Felicity and my cute little niece Annabelle; thanks for putting up with me and making my time at home enjoyable.

Author's Declaration

I declare that, except where explicit reference is made to the contribution of others, this dissertation is the result of my own work and has not been submitted for any other degree at the University of Glasgow or at any other institution.

Kate Beckham

February 2014

Abstract

Escherichia coli contributes to the commensal microbiota of most mammals by producing vitamins and aiding digestion. However, several strains of *E. coli* have evolved as pathogens and have become highly adapted for specific niches through the acquisition of pathogenicity islands. *E. coli* O157:H7 is a commensal bacterium of cattle but if transferred to humans, usually by contaminated food products, it can act as a pathogen. In humans *E. coli* O157:H7 causes diarrhoea, haemorrhagic colitis and in some cases haemolytic uremic syndrome which can result in death. Clinical treatment of this pathogen is difficult since the antibiotics currently available have been shown to worsen the clinical outcome of infection. Therefore, the discovery of novel anti-bacterial therapies is of high importance.

A novel approach to limit pathogenesis is to target the factors used by bacteria to cause disease - their key virulence factors. In theory, such approaches should only compromise the ability of the pathogen to cause disease, rather than its ability to survive, thereby reducing the selective pressure for the development of resistance mechanisms. The type three secretion system (LEE T3SS) is a key virulence factor for *E. coli* O157:H7 as it facilitates tight attachment to host cells and the secretion of effector proteins. The importance of this virulence factor for the disease process makes it an attractive target for anti-microbial therapies.

The salicylidene acylhydrazides (SA) are a class of compounds that inhibit the expression of the T3SS of several Gram-negative pathogens. When this study was started, the mode of action of these compounds was completely unknown. An affinity pull-down assay to identify the binding proteins of the SA compounds was conducted. Identification of the target proteins of the compounds was the first step in determining their mechanism for decreasing the expression of the LEE T3SS in *E. coli* O157. The pull-down identified nineteen putative targets, none of which had previously been linked to the regulation of the LEE T3SS.

The aim of this thesis was to systematically investigate the putative targets using a combination of approaches: phenotypic studies of deletion mutants and structural and

functional characterisation of the target proteins. From the nineteen putative targets seven were selected for further investigation namely Tpx, WbrA, FolX, FkpA, FklB, SurA and AdhE. Several of these proteins were shown to bind to the SA compounds however deletion of only one of the target proteins resulted in a decrease in LEE T3S. This target, AdhE, offers an exciting new lead in the search for novel targets for antibacterial therapies.

Table of Contents

Acknowledgements **2**

Authors Declaration **4**

Abstract **5**

Table of Contents **7**

List of Tables **12**

List of Figures **13**

Abbreviations **17**

1 Introduction **20**

1.1 *Escherichia coli* **21**

1.1.1 Pathogenic *E. coli* 21

1.1.2 Outbreaks and pathogenesis 22

1.2 *E. coli* virulence factors **23**

1.2.1 Locus of enterocyte effacement 24

1.2.2 Type three secretion system 28

1.2.3 Effector proteins 30

1.2.4 Flagella 31

1.2.5 Shiga toxin 33

1.2.6 Other virulence factors 34

1.3 Current therapies **34**

1.4 Anti-virulence compounds **36**

1.4.1 Rationale 36

1.4.2 Quorum sensing inhibitors 36

1.4.3 Adhesin inhibitors 37

1.4.4 Inhibitors of the T3SS 38

1.4.5 Salicylidene acylhydrazides 42

1.4.6 Mechanism of action of the SA compounds 45

1.5 Aims of the project **48**

1.6 Biophysical characterisation of protein targets **49**

1.7 X-ray crystallography **50**

1.7.1 X-ray crystallography theory 50

1.7.2 Data collection and processing 52

1.8 Small angle X-ray scattering **54**

1.8.1 SAXS data collection and theory 54

1.8.2 SAXS data processing 56

1.8.3 Modelling of solution structures from SAXS data 59

1.9 Analytical ultracentrifugation 61
1.9.1 Sedimentation velocity 61
1.9.2 Sedimentation equilibrium 63
2 Materials and methods 66
2.1 Chemicals, strains and growth media 67
2.1.1 Growth media 67
2.1.2 Strains 67
2.2 Maintenance and growth of bacteria 67
2.2.1 Storage of bacteria 67
2.2.2 Growth of bacteria 68
2.3 Molecular techniques 69
2.3.1 Preparation of bacterial genomic DNA 69
2.3.2 Plasmids 69
2.3.3 Oligonucleotide primers 71
2.3.4 Polymerase chain reaction 71
2.3.5 Agarose gel electrophoresis 72
2.3.6 Restriction enzyme digest 72
2.3.7 DNA gel purification 73
2.3.8 Ligation reaction 73
2.3.9 Heat shock transformation 73
2.3.10 Production of electrocompetent *E. coli* 73
2.3.11 Electroporation transformation 74
2.3.12 Extraction of plasmid DNA 74
2.3.13 Creation of deletion mutants by allelic exchange 74
2.3.14 Creation of *lux* positive *C. rodentium* strain 75
2.4 Phenotypic characterisation of strains 75
2.4.1 Secretion assay 75
2.4.2 SDS-PAGE 76
2.4.3 Western blotting 76
2.4.4 Far-western 77
2.4.5 GFP reporter fusion assay 78
2.4.6 Motility assay 78
2.5 RNA transcript sequencing 78
2.5.1 RNA extraction 78
2.5.2 RNA sequence data processing and analysis 79
2.6 Immunofluorescence microscopy 79
2.6.1 Single cell analysis 79
2.6.2 Live imaging of bacterial motility 80
2.6.3 Adhesion of *E. coli* to host cells 81
2.7 TLR5 reporter assay 82

2.8 Animal experiments **83**
2.8.1 Infection of Balb/C mice with *C. rodentium* 83
2.9 Protein overexpression and purification **84**
2.10 Analytical ultracentrifugation **85**
2.10.1 Determination of the oligomeric state of Tpx, FklB and FolX 85
2.10.2 Determination of K_d for the interaction of Tpx and ME0052 87
2.11 Small angle X-ray scattering **88**
2.11.1 Data collection and processing 88
2.12 NMR **89**
2.12.1 Acetate quantification 89
2.12.2 ^{1}H-^{15}N HSQC 89
2.13 Isothermal titration calorimetry **90**
2.14 Homology modelling **90**
2.14.1 FklB 90
2.15 X-ray crystallography **91**
2.15.1 Protein crystallisation 91
2.15.2 Data collection and processing 91
2.16 Enzymology **92**
2.16.1 Prolyl-cis isomerase assay 92
2.16.2 Inhibition by ME0052 and rapamycin 92
2.16.3 WrbA acitivty 93

3 Is thiol peroxidase a target protein of the salicylidene acylhydrazides? **94**
3.1 Introduction **95**
3.1.1 The role and enzymatic function of Tpx 95
3.1.2 The role of Tpx in virulence 97
3.1.3 The structure of Tpx 97
3.1.4 Aims of this chapter 97
3.2 Structural characterisation of *yp*Tpx **99**
3.2.1 Over-expression and purification of Tpx 99
3.2.2 Oligomeric state of Tpx 100
3.2.3 Crystallographic studies of *yp*Tpx 101
3.2.4 Crystallisation of *yp*Tpx 102
3.2.5 Structure of *yp*TpxC61S 103
3.2.6 Solution structure of *yp*Tpx 106
3.2.7 Modelling of flexible regions of *yp*Tpx using SAXS data 109
3.3 Does *yp*Tpx bind ME0052? **110**
3.3.1 Far-western analysis reveals binding of *yp*Tpx to ME0052 110
3.3.2 Determination of affinity and stoichiometry of *yp*Tpx ME0052 complex using AUC 111
3.3.3 NMR studies of ME0052-*yp*Tpx interactions 112
3.3.4 Docking of ME0052 into the binding site 114

3.4 Crystallographic analysis of a C61S mutant of Tpx from *E. coli O157:H7* **114**
3.4.1 Protein purification and crystallisation 115
3.4.2 Structure of *ec*TpxC61S 116
3.4.3 Analysis of the crystal packing of *ec*TpxC61S 118
3.5 Phenotypic analysis of the Δ*tpx* mutant of *E. coli* O157:H7 **119**
3.6 Discussion **121**
3.6.1 Structural studies of Tpx 122
3.6.2 Phenotypic studies of Δ*tpx* 124
3.6.3 Conclusion 124
4 Structural investigation of the salicylidene acylhydrazide target proteins FolX and WrbA **125**
4.1 Introduction **126**
4.1.1 FolX 126
4.1.2 WrbA 127
4.1.3 Aims of this chapter 127
4.2 Structural characterisation of FolX **128**
4.2.1 Oligomeric state of FolX 128
4.2.2 Solution structure of FolX 130
4.3 Structural characterisation of WrbA **132**
4.3.1 Expression, purification and crystallisation of WrbA 132
4.3.2 Crystal structure of *st*WrbA in complex with ME0052 133
4.3.3 Enzymatic activity of WrbA and inhibition with ME0052 139
4.4 Discussion **141**
4.4.1 FolX 141
4.4.2 WrbA 142
4.4.3 Conclusions 143
5 Are the PPIases the protein targets of the salicylidene acylhydrazides? **144**
5.1 Introduction **145**
5.2 Peptidyl-prolyl *cis-trans* isomerases **145**
5.2.1 Monitoring PPIase activity 146
5.3 The parvulins **147**
5.3.1 SurA 148
5.3.2 The FKBPs 149
5.3.3 FkpA 150
5.3.4 FklB 150
5.4 Do FkpA and SurA bind the SA compounds? **151**
5.4.1 Cloning, expression and purification of SurA and FkpA 151
5.5 Far-western analysis of the binding of the PPIases to ME0052 **152**
5.6 PPIase activity of SurA and FkpA **154**
5.7 Deletion of *surA* in *E. coli* O157 **155**
5.7.1 Δ*surA* is non-motile 156

5.7.2 The effect of *surA* deletion on the expression of the LEE T3SS and flagella 157
5.7.3 Attaching and effacing lesion formation 159
5.8 Structural and functional characterisation of FklB 160
5.8.1 Protein overexpression and purification 160
5.8.2 Homology modelling of FklB 161
5.9 Determination of the oligomeric state of FklB 162
5.10 Solution structure of FklB 166
5.10.1 Modelling FklB flexibility 169
5.11 FklB enzyme activity 171
5.11.1 Inhibition of FklB 172
5.12 Binding of rapamycin and ME0052 to FklB 174
5.12.1 ^{15}N chemical shift assay 174
5.12.2 Isothermal titration caliormetry 178
5.13 The effects of *fklB* gene deletion on the LEE T3SS 180
5.13.1 Construction of O157:H7 Δ*fklB* mutants 180
5.14 Does the deletion of *fklB* affect the expression of the LEE T3SS or flagellar system? 181
5.15 Discussion 183
5.15.1 Validation of binding targets 183
5.15.2 Characterisation of deletion mutants 184
5.15.3 Binding studies 186
5.15.4 The structure of FklB 187
5.15.5 Conclusions 187
6 Is AdhE a target of the salicylidene acylhydrazides? 189
6.1 Introduction 190
6.1.1 The function and regulation of AdhE 190
6.1.2 The structure of AdhE 192
6.1.3 AdhE and its links to virulence 193
6.1.4 Aims of this chapter 193
6.2 Characterisation of AdhE 194
6.2.1 Cloning, expression and purification of AdhE 194
6.2.2 AdhE shows binding to the SA compounds by far-western 196
6.3 Characterisation of Δ*adhE* in *E. coli* O157 197
6.3.1 Deletion of adhE 197
6.4 Type three secretion is decreased in the Δ*adhE* mutant 197
6.4.1 Effects on LEE T3SS protein expression 198
6.4.2 Effects on LEE T3SS gene expression 199
6.5 The expression of flagella is up-regulated in the Δ*adhE* mutant 202
6.5.1 Effects on flagellar protein expression 202
6.5.2 Effects on flagellar gene expression 204
6.6 Complementation of the Δ*adhE* mutant 206

6.7 The Δ*adhE* mutant is non-motile 206
6.8 Interaction with host cells 208
6.9 Metabolic differences in the *ΔadhE* mutant 210
6.9.1 The levels of acetate in the Δ*adhE* mutant are increased 210
6.10 The effects of increased acetate 211
6.10.1 The effect of acetate on *fliC* expression 211
6.10.2 The effect if acetate on LEE expression 212
6.10.3 Acetate induces protein post-translational modification 213
6.10.4 Investigation of other metabolic changes in Δ*adhE* – cAMP 214
6.11 Live imaging of bacterial swimming 216
6.12 Investigation of CheY acetylation 217
6.12.1 Cloning, overexpression and purification of CheY 217
6.12.2 CheY antibody production and testing 218
6.13 TLR5 response to the *ΔadhE* mutant 219
6.14 Is disulfiram an inhibitor of AdhE? 220
6.14.1 Effect of disulfiram on LEE T3S and flagellar gene expression 220
6.14.2 Does disulfiram affect the motility of *E. coli*? 223
6.14.3 Treatment with disulfiram reduces attachment to host cells 224
6.14.4 Disulfiram reduces *C. rodentium* colonisation in *vivo* 225
6.15 Discussion 227
6.15.1 Deletion of *adhE* mimics the phenotype associated with SA compound treatment 227
6.15.2 What is the molecular mechanism behind the observed changes in the Δ*adhE* mutant? 229
6.15.3 Inhibition of AdhE with disulfiram 234
6.15.4 Conclusion 235
7 Concluding remarks 236
References 241
Publications 265

List of Tables

Table 1-1 Nomenclature of homologous T3SS proteins found in *E. coli*, *Yersinia* and *Salmonella*. 30
Table 1-2 Chemical structure of T3SS inhibitors effective against Gram-negative pathogens 40
Table 1-3 Chemical structures of salicylidene acylhydrazide compounds that inhibit T3SSs in Gram-negative pathogens. 43
Table 1-4 Putative target proteins of the salicylidene acylhydrazides from *E. coli* O157:H7 49
Table 2-1 Media recipes 67
Table 2-2 Strains used in this study 68
Table 2-3 Plasmids used in this study 70

Table 2-4 Plasmids generated in this study....70
Table 2-5 Primers used in this study....71
Table 2-6 Standard PCR protocol....72
Table 2-7 *lux* primers....75
Table 2-8 Antibodies used in this study....77
Table 2-9 Partial specific volumes of proteins and densities and intrinsic viscosities of buffers used in analytical ultracentrifugation....86
Table 3-1 Data collection and refinement statistics for *yp*TpxC61S....103
Table 3-2 Data collection and refinement statistics for *ec*TpxC61S....117
Table 3-3 Superposition of orthorhombic *ec*TpxC61S from this study with reduced wild-type and 'forced reduced' mutated Tpx structures....118
Table 4-1 X-ray crystallography data collection and refinement statistics for *st*WrbA....134
Table 4-2 Specific activities of *st*WrbA in the presence of ME0052....140
Table 6-1 RNA sequencing data showing the expression of the LEE T3SS genes....203
Table 6-2 RNA sequencing data showing the expression of flagellar genes....205
Table 6-3 RNA sequencing data showing significant changes in *E. coli* treated with disulfiram....223

List of Figures

Figure 1-1 Organisation of the locus of enterocyte effacement (LEE) in *E. coli* O157:H7....25
Figure 1-2 Regulation of the locus of enterocyte effacement in *E. coli* O157:H7....27
Figure 1-3 The type three secretion system apparatus....29
Figure 1-4 Comparison of the structure of the LEE T3SS and the flagellar system....32
Figure 1-5 Overview of the affinity pull-down assay to determine the binding proteins of the salicylidene acylhydrazides....47
Figure 1-6 SDS-PAGE gel of the protein fractions from the affinity pull-down assay....48
Figure 1-7 Incident X-rays are elastically scattered from lattice planes according to Bragg's law....51
Figure 1-8 SAXS data collection....55
Figure 1-9 SAXS scattering curves....55
Figure 1-10 SAXS scattering curves with Guinier plot (inset)....57
Figure 1-11 p(r) distributions of different shapes....58
Figure 1-12 Kratky plots indicating the 'compactness' of a protein....59
Figure 1-13 Sedimentation velocity experiment....62
Figure 1-14 An example of sedimentation equilibrium data showing the exponential functions for the distribution of molecular masses....65
Figure 2-1 Structure of the biotinylated version of ME0052. ME0052-Bio was used in far-western analysis of the binding of the compound to target proteins in *E. coli* O157:H7....77
Figure 3-1 Catalytic mechanism of Tpx....96
Figure 3-2 Sequence alignment of Tpx from *E. coli* O157:H7 and *Y. pseudotuberculosis*....98

Figure 3-3 Purification of *yp*Tpx and *yp*TpxC61S. SDS-PAGE gel showing fractions following IMAC 100
Figure 3-4 Sedimentation velocity analysis of *yp*Tpx 101
Figure 3-5 *yp*TpxC61S crystals and diffraction 102
Figure 3-6 Overall structure of *yp*Tpx 104
Figure 3-7 Crystal structures of *yp*Tpx 105
Figure 3-8 Changes in the active site of *yp*Tpx caused by the redox state of the protein 106
Figure 3-9 Dimer interface of *yp*Tpx 107
Figure 3-10 Small angle X-ray scattering from *yp*Tpx.....108
Figure 3-11 Movement around the dimer interface of *yp*Tpx calculated from SAXS data 109
Figure 3-12 EOM modelling of the "missing" his-tag using SAXS data.....110
Figure 3-13 Far-western analysis of ME0052 binding to *yp*Tpx and *yp*TpxC61S. 111
Figure 3-14 AUC analysis of ME0052 binding to *yp*Tpx.....112
Figure 3-15 Binding region of ME0052 in *yp*Tpx 113
Figure 3-16 Docking of ME0052 into the binding site on *yp*Tpx. 114
Figure 3-17 *ec*TpxC61S crystals and diffraction. 115
Figure 3-18 The structure of orthorhombic *ec*TpxC61S. 116
Figure 3-19 Crystal packing of orthorhombic *ec*TpxC61S.....119
Figure 3-20 Transcriptomic analysis of Δ*tpx* and a *tpx*C61S complement. 120
Figure 4-1 FolX as a tetramer and octamer.....128
Figure 4-2 AUC analysis of FolX. 130
Figure 4-3 Solution structure of FolX. 131
Figure 4-4 *st*WrbA crystals and diffraction pattern.....132
Figure 4-5 Sequence alignment of WrbA from different species. 133
Figure 4-6 Structure of *st*WrbA. 135
Figure 4-7 Interactions of FMN with *st*WrbA. 136
Figure 4-8 Catalytic mechanism for the hydrolysis of ME0052.....137
Figure 4-9 Omit map showing the positions of FMN and ME0052 in the *st*WrbA structure 138
Figure 4-10 Interactions between ME0052 and FMN in the *st*WrbA active site.....139
Figure 4-11 Enzyme activities of *st*WrbA 140
Figure 5-1 *Cis-trans* isomerisation of a peptidyl-prolyl bond.....145
Figure 5-2 Monitoring peptidyl-prolyl *cis*/*trans* isomerisation using isomer-specific proteolysis. 147
Figure 5-3 Structure of SurA.....148
Figure 5-4 Purification of SurA and FkpA 153
Figure 5-5 Far-western analysis of the binding of ME0052 to the PPIase target proteins.....153
Figure 5-6 First order rate constant of SurA PPIase activity. 154
Figure 5-7 Determination of the first order rate constant of the PPIase activity of FkpA 155
Figure 5-8 Growth rate of WT and the Δ*surA* mutant.....156
Figure 5-9 Motility assay of the WT and Δ*surA* mutant. 156
Figure 5-10 Expression of gene:GFP reporter fusions in Δ*surA*. 158
Figure 5-11 Secretion profile of the WT and the Δ*surA* mutant. 159

Figure 5-12 Attaching and effacing lesion formation in the Δ*surA* mutant. 160
Figure 5-13 Purification of FklB and FklBc.. 161
Figure 5-14 FklB homology model. 162
Figure 5-15 Size distribution (c(s)) of FklB determined from SV data. 163
Figure 5-16 Sedimentation coefficient of FklB. 164
Figure 5-17 Molecular weight of FklB determined from SE data.. 165
Figure 5-18 FklB sedimentation equilibrium data fitted with a monomer-dimer model. 165
Figure 5-19 SAXS curve with Guinier analysis. 166
Figure 5-20 Kratky plot and pair-distance p(r) distribution 167
Figure 5-21 Comparison of the scattering curve calculated for the FklB homology model using CRYSOL with the experimental SAXS data. 168
Figure 5-22 *Ab inito* modelling of FklB in solution. 169
Figure 5-23 Modelling FklB flexibility using SWISS-MODEL homology models. 170
Figure 5-24 Modelling FklB flexibility using SOMO-DMD 171
Figure 5-25 Determination of the first order rate constant of the PPIase activity of FklB. 172
Figure 5-26 Inhibition of FklB with rapamycin 173
Figure 5-27 Inhibition of FklB by ME0052. 174
Figure 5-28 ^{1}H-^{15}N HSQC spectra of 100 µM FklBc in the presence and absence of DMSO 175
Figure 5-29 ^{1}H-^{15}N HSQC spectra of 100 µM FklBc with and without 200 µM ME0052. 175
Figure 5-30 ^{1}H-^{15}N HSQC spectrum of 100 µM FklB with and without 20 µM rapamycin. 176
Figure 5-31 Predicted rapamycin binding site in FklBc. 177
Figure 5-32 Calorimetric titration of FklB or FklBc to rapamycin.. 179
Figure 5-33 Calorimetric titration of ME0052 into FklB. 179
Figure 5-34 Deletion of *fklB*. 180
Figure 5-35 Expression of gene:GFP reporter fusions in Δ*fklB*. 182
Figure 5-36 Secretion profile of the WT and Δ*fklB* mutant. 183
Figure 6-1 The role of AdhE in metabolism 191
Figure 6-2 AdhE is found in several pathogens. 193
Figure 6-3 Purification of AdhE and its enzymatic domains. 194
Figure 6-4 Negative stain EM of purified AdhE. 195
Figure 6-5 Far-western analysis of the binding of AdhE to ME0052. 196
Figure 6-6 Deletion of *adhE* does not affect growth rate. 197
Figure 6-7 Analysis of the secretion profile of the Δ*adhE* mutant. 199
Figure 6-8 Expression of the LEE operons in the Δ*adhE* mutant and the WT 201
Figure 6-9 Expression of flagellar proteins in the Δ*adhE* mutant 202
Figure 6-10 Expression of a *fliC* promoter:gfp fusion is increased in the Δ*adhE* mutant 204
Figure 6-11 Complementation of Δ*adhE*. 206
Figure 6-12 Motility assay shows that the Δ*adhE* mutant is non-motile. 207
Figure 6-13 Transmission electron micrograph of WT *E. coli* O157:H7 and Δ*adhE* grown in LB medium. 208
Figure 6-14 Immunofluorescence of WT and Δ*adhE* 209

Figure 6-15 Quantification of acetate levels using ^{1}H-NMR spectroscopy.....211
Figure 6-16 The effects of extracellular acetate on *fliC* expression. 212
Figure 6-17 The effects of extracellular acetate on the expression of the LEE T3SS. 213
Figure 6-18 Acetate contributes to the post-translational modification of proteins.....214
Figure 6-19 Quantification of cAMP levels.....216
Figure 6-20 Still image from movies of bacterial swimming. 217
Figure 6-21 Purification of CheY.....218
Figure 6-22 Testing of the anti-CheY antibody.....219
Figure 6-23 The level of TLR5 stimulation in response to the Δ*adhE* mutant. 220
Figure 6-24 The effect of disulfiram on the secretion profile of *E. coli* O157 221
Figure 6-25 GFP:promoter fusions reporting on the expression of the LEE T3SS and flagellar genes in WT and DSF treated bacteria. 222
Figure 6-26 Treatment with DSF reduces bacterial motility in soft agar 224
Figure 6-27 Treatment with disulfiram reduces A/E lesion formation 225
Figure 6-28 Colonisation of BALB/c mice with *C. rodentium* cultured in 10 µM disulfiram.....226
Figure 6-29 Proposed model of the effects of *adhE* deletion in *E. coli* O157.....233

Abbreviations

°C	Degrees Celsius
Δ	Deletion
Δ*adhE*	*E. coli* O157:H7 deletion mutant for *adhE*
Δ*fklB*	*E. coli* O157:H7 deletion mutant for *fklB*
Δ*folX*	*E. coli* O157:H7 deletion mutant for *folX*
Δ*surA*	*E. coli* O157:H7 deletion mutant for *surA*
Δ*tpx*	*E. coli* O157:H7 deletion mutant for *tpx*
Δ*wrbA*	*E. coli* O157:H7 deletion mutant for *WrbA*
Å	Ångstrom
AdhE	Bifunctional acetaldehyde/alcohol dehydrogenase
Amp	Ampicillin
AUC	Analytical ultracentrifugation
BLAST	Basic local alignment search tool
bp	Base pair
BSA	Bovine serum albumin
c.f.u.	Colony forming unit
Chl	Chloramphenicol
Cra	Catabolite repressor/activator protein
dH_20	Distilled water
DMEM	Dulbelcco's Modified Eagle Medium
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EAEC	Entero-aggregative *Escherichia coli*
ECL	Enhanced chemi-luminescence
EHEC	Enterohaemorrhagic *Escherichia coli*
Ehx	Enterohaemolysin
EIEC	Enteroinvasive *Escherichia coli*
EPEC	Enteropathogenic *Escherichia coli*
Ery	Erythromycin
ETEC	Enterotoxigenic *Escherichia coli*
FCS	Foetal calf serum
FklB	FKBP-type 22KDa peptidyl-prolyl *cis-trans* isomerase
FkpA	FKBP-type peptidyl-prolyl *cis-trans* isomerase
FolX	7,8-dihydroneopterin-triphosphate-epimerase
g	Gram
gDNA	Genomic DNA
GFP	Green fluorescent protein
GI tract	Gastrointestinal tract
GrlA/R	Global regulator of LEE-activator/repressor
h	Hour

H_2O_2	Hydrogen peroxide
HGT	Horizontal gene transfer
HRP	Horseradish peroxidase
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl-β-D-thiogalactopyranoside
ITC	Isothermal titration calorimetry
Kan	Kanamycin
Kb	Kilobase pair
kDa	Kilodalton
L	Litre
LB	Luria Broth
LEE	Locus of enterocyte effacement
Ler	LEE-encoded regulator
LPS	Lipopolysaccharide
MEM	Minimum Essential Medium
mg	Milligram(s)
min	Minute(s)
ml	Millilitre (s)
mM	Milimolar
mRNA	Messenger ribonucleic acid
MW	Molecular weight
NLE	Non-LEE encoded effector(s)
nm	Nanometer
nM	Nanomolar
NMR	Nuclear magnetic resonance
$OD_{600\ nm}$	Optical density at 600 nm
ORF	Open reading frame
PAI	Pathogenicity island
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PISA	Protein Interfaces, Surfaces and Assemblies
r.m.s.d.	Root-mean-squared deviation
rpm	Revolutions per minute
RT	Room temperature (~18°C)
SA	Salicylidene acylhydrazides
SAXS	Small angle X-ray scattering
SCFA	Short chain fatty acid(s)
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
Stx	Shiga toxin
SurA	Survival protein A
T3SS	Type three secretion system
TAE	Tris-acetate-EDTA

TB	Terrific broth
Tir	Translocated intimin receptor
TLR5	Toll-like receptor five
Tpx	Thiol peroxidase
UK	United Kingdom
UPEC	Uropathogenic *Escherichia coli*
USA	United States of America
UTI	Urinary tract infection
v/v	Volume per volume
WrbA	Tryptophan repressor binding protein
WT	Wild-type
x g	Centrifugal force
µl	Microliter(s)
µm	Micrometer(s)
µM	Micromolar

1 Introduction

1.1 *Escherichia coli*

Escherichia coli (*E. coli*) is a Gram-negative, facultative anaerobe belonging to the genus Enterobacteriaceae of the Gammaproteobacteria. This bacterium can be grown readily, and its genetics are easily manipulated in the laboratory, making it a common workhorse and one of the best-studied prokaryotic model organisms.

E. coli is commonly found in the gastrointestinal tract (GI tract) of warm-blooded animals, including birds, ruminants and humans. It is considered a commensal bacterium as it colonises the GI tract of the host a few hours after birth where it forms part of the natural microflora of the host organism. In humans, the niche of this commensal is the mucosal surface of the colon. In order to outcompete other species at this site, it is thought that *E. coli* may have a metabolic advantage over other species owing to its ability to utilise gluconate as a carbon source [1]. Under normal circumstances *E. coli* lives in a mutually beneficial symbiotic relationship with its host. However commensal bacteria can cause disease, for example, when the integrity of the mucosal barrier is compromised enabling colonisation of further niches. Moreover, the acquisition of virulence factors can transform a commensal bacterium into a pathogen and several pathogenic strains of *E. coli* exist. These pathogenic strains, discussed below, can result in three main clinical presentations: diarrhoeal disease, urinary tract infections (UTIs) and meningitis/sepsis.

1.1.1 *Pathogenic E. coli*

There are several distinct pathotypes of *E. coli* that have evolved through the acquisition of different virulence factors, producing highly adapted strains able to colonise different niches, often with detrimental effects on the host. Virulence factors are transferred between strains by horizontal gene transfer (HGT) or on mobile genetic elements, thus several pathogenic strains of *E. coli* utilise homologous factors. The virulence factors used by *E. coli* will be discussed in detail in Section 1.2.

Five distinct clinical groups of diarrhoeagenic isolates have been identified: entero-aggregative *E. coli* (EAEC), enterohaemorrhagic *E. coli* (EHEC), enteropathogenic *E. coli* (EPEC), enteroinvasive *E. coli* (EIEC), and enterotoxigenic *E. coli* (ETEC). Urinary tract

infections are caused by strains belonging to the uropathogenic *E. coli* (UPEC) pathotype. Strains that are associated with meningitis and sepsis are classed as MNEC (meningitis associated *E. coli*) however pathotypes that cause extraintestinal infections are also known as ExPEC [2].

The different *E. coli* pathotypes can be classified based on their surface antigens. These include the 'O' antigen, which denotes the lipopolysaccharide (LPS) group, and the 'H' antigen, which defines the flagellar type present on the cells. Together these factors are used to determine the serotype of the *E. coli* strain. For example, the serotype of *E. coli* that is the focus of this thesis is O157:H7, which encodes the O157-type LPS and the H7 flagellin protein.

1.1.2 *Outbreaks and pathogenesis*

The first *E. coli* pathotype identified as being responsible for a disease outbreak was EPEC in 1945 [3]. Since this outbreak, the ability to isolate and characterise disease-causing agents has led to the identification of *E. coli* as a predominant causative agent of diarrheal disease. The origin of subsequent outbreaks has largely been attributed to the contamination of food and water sources. EPEC infection is associated with watery diarrhoea and fever, often affecting children under the age of one [4]. In several outbreaks mortality rates of 20-50% were reported [4–6]. However, an improvement in public sanitation and improved food-handling procedures has reduced the incidence of EPEC infections in the developed world, although they remain prevalent in the developing world.

The first reported EHEC (*E. coli* O157:H7) outbreak in 1982 was traced to the contamination of hamburger meat [7]. The symptoms associated with this outbreak were more severe than had been reported for previous *E. coli* outbreaks, predominantly linked to EPEC. In this outbreak the occurrence of haemorrhagic colitis leading to bloody diarrhoea and severe abdominal cramps was reported [7]. Later that year, EHEC infection was linked to the development of haemolytic-uremic syndrome (HUS), characterised by the combination of acute renal failure, thrombocytopenia, and microangiopathic haemolytic anaemia [8]. The development of this severe disease was associated with the production and release of a Shiga-like toxin (Stx), a cytotoxin responsible for damaging

both intestinal and renal tissue [9]. It is the presence of Stx that is a defining characteristic of the EHEC pathotype. Stx is discussed in more detail in Section 1.2.5. Since its initial identification, there have been numerous outbreaks associated with EHEC infections, with the O157:H7 serotype being the most common [10]. These outbreaks are linked with contamination of the food chain, a particular problem associated with *E. coli* O157:H7 since its natural host is cattle and other ruminants, which act as a reservoir of infection [11]. Once contamination of the food chain has occurred, the infective dose required to cause disease in healthy individuals is very low and is estimated to be approximately 50-100 colony-forming units (CFUs) [12].

EHEC O157:H7 has been recognised as a meat contaminant since 1994. However, since then six further serotypes that are associated with severe health effects have been identified as being common contaminants of meat including O26, O45, O103, O111, O121 and O145. The prevalence of these strains varies geographically although the basis of the distribution is not known [13].

The 2011 *E. coli* outbreak in Germany was significant as it led to the highest number of reported cases of HUS (908) and resulted in 34 deaths [14]. Genomic analysis showed that the pathotype responsible was an EAEC strain (O104:H4) that had acquired Stx from a lambdoid-like phage [14,15]. The acquisition of the toxin combined with the ability of the strain to form biofilms in the small intestine made treatment challenging [16]. This demonstrates how the evolution of different pathotypes through the acquisition of virulence factors leads to the development of highly adapted enteric pathogens capable of causing severe disease in humans.

1.2 *E. coli* virulence factors

The different *E. coli* pathotypes can also be classified by the virulence factors that they possess. A description of the main *E. coli* virulence factors follows, with a focus on the factors associated with EHEC.

1.2.1 *Locus of enterocyte effacement*

The locus of enterocyte effacement (LEE) is common to both EPEC and EHEC and is central to their pathogenesis. The LEE is a pathogenicity island (PAI) that has been transferred horizontally to these pathotypes from a non-*E. coli* origin. The LEE is integrated into the chromosome of a transformed strain at specific sites, which has been successfully replicated *in vitro* generating a LEE+ K12 strain [17,18]. The LEE in EHEC is slightly larger than in EPEC at 43.559 kilobase pairs (kbp) compared with 35.624 kbp [19]. The difference in size is attributed to a prophage insertion, namely prophage 933L, which encodes 13 putative open reading frames (ORFs). Excluding the 933L prophage, the LEE from EPEC and EHEC share 93.9% similarity, which encodes 41 genes arranged into five polycistronic operons shown in Figure 1-1 [19]. These 41 core genes confer the ability to form attaching and effacing (A/E) lesions, a term that describes the interface between a tightly attached bacterium and the host epithelial cell, a characteristic trait of EHEC and EPEC infection [20,21].

The formation of A/E lesions is a complex process involving the production of a multi-protein secretion apparatus known as the type three secretion system (LEE T3SS) and the secretion of effector proteins. The translocated effector proteins induce changes in host cell pathways and facilitate tight attachment and colonisation of the host epithelium. Expression of the LEE is regulated primarily by the master-regulator (Ler) which is encoded in the first operon (LEE1) [22]. Ler is a transcriptional activator that is homologous to the global regulator H-NS [23]. H-NS binds to AT rich regions (often regions that have been acquired through HGT) repressing their expression and the LEE is one of its targets [23]. Therefore Ler acts to release the repression of the LEE mediated by H-NS [24]. The expression of Ler is regulated by several other factors within the bacterium (Figure 1-2). Ler positively regulates the expression of all of the other LEE operons (LEE2-5) as well as several non-LEE encoded genes. As well as Ler, the LEE encodes two other regulatory elements: the Global Regulator of the LEE Activator (GrlA) and its opposing Repressor (GrlR) [25]. GrlA promotes Ler expression forming a positive feedback loop, and GrlR binds to GrlA to inhibit its interaction with the Ler promoter [26].

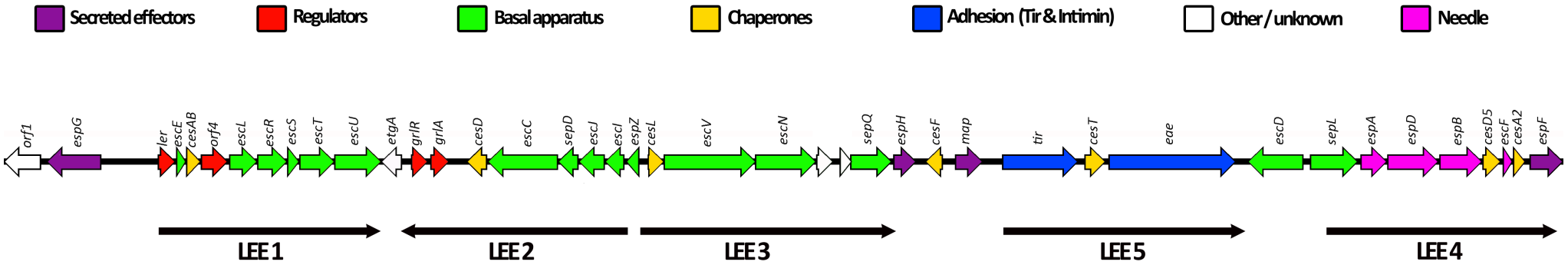


Figure 1-1 Organisation of the locus of enterocyte effacement (LEE) in *E. coli* O157:H7. Genes have been coloured according to their function.

The QseA quorum sensing regulator acts directly on the expression of the LEE1 operon [27]. The sensor kinase QseE responds to sulphate, phosphate and host-derived adrenaline leading to the phosphorylation of the QseF response regulator, which increases Ler production [27] (Figure 1-2). The sensor kinase QseC responds to host hormones adrenaline and noradrenaline as well as the EHEC produced auto-inducer (AI) molecule increasing *ler* transcription, through phosphorylation of KdpE [28].

Several other regulators of the LEE are encoded on O-islands dispersed throughout the genome. For example, EtrA and EivF are encoded in a cryptic, second LEE T3SS gene cluster in EHEC, and negatively regulate the LEE PAI [29]. The RNA chaperone Hfq has been shown to negatively affect LEE expression through interactions with *ler* mRNA, however this appears to be strain dependent [30].

Expression of the LEE is influenced by several environmental factors including nutrient levels, growth phase and growth temperature [31]. A glycolytic environment, at 0.4% glucose reduces *ler* expression, whereas 0.1% glucose mimicking gluconeogenesis conditions enhances *ler* expression [28]. In response to the metabolic status of the cells, expression of the LEE is regulated by Cra (catabolite repressor/activator protein) [28]. Furthermore, the expression of the LEE is greatly increased during growth in certain tissue culture media in particular MEM-HEPES [22,31], which is therefore used to study the expression of the LEE *in vitro*. However, the mechanisms behind this up-regulation are poorly understood.

The complex, multi-layered regulation of the LEE highlights the importance of the LEE T3SS under specific conditions and only when the environmental conditions are favourable is this virulence factor expressed. The expression of the LEE T3SS is further regulated by the expression of other virulence factors, for example the expression of flagella [26].

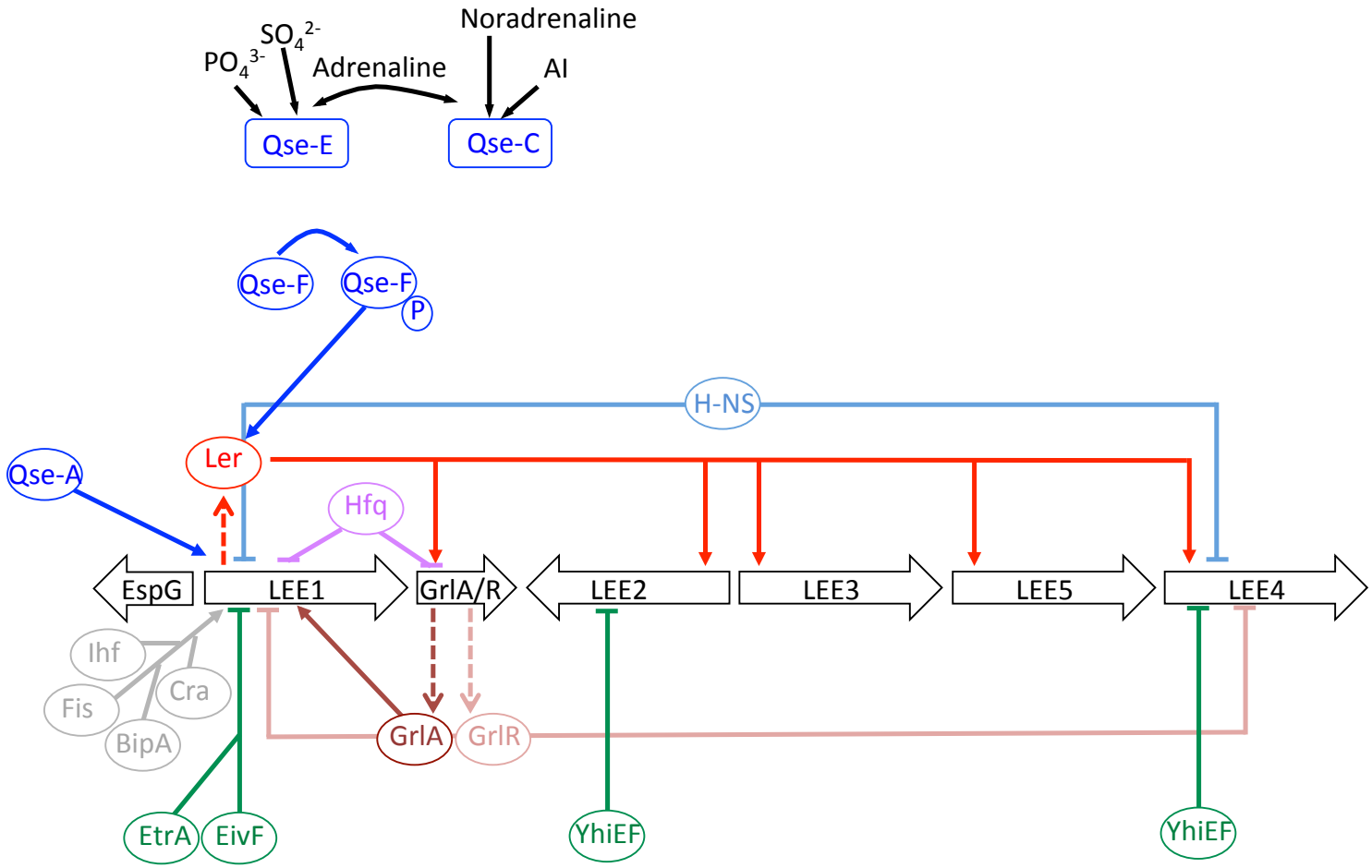


Figure 1-2 Regulation of the locus of enterocyte effacement in *E. coli* O157:H7. The LEE operons are shown as block arrows. Regulatory proteins are shown as oval shapes and for clarity are colour-coded with the arrows denoting their regulatory effects. Closed arrows represent positive regulation and blunt arrows indicate negative regulation. Dashed open arrows indicate expression of a regulatory protein from the LEE. The membrane components of the Qse system have been shown as rectangles and (P) indicates phosphorylation of Qse-F. Environmental signals are labeled with black text. See text for a detailed explanation. This figure is adapted from [32].

1.2.2 *Type three secretion system*

As mentioned above, the LEE encodes 41 genes that are required for the production of a functional LEE T3SS. The genes of the LEE can be classified into 3 main groups: the 'Esc' genes which encode the structural proteins of the LEE T3SS, the 'Ces' genes which encode chaperone proteins and the secreted effectors which are encoded by the 'Esp' genes. The LEE also encodes the adhesin intimin and its cognate receptor, Tir (translocated intimin receptor), which are primarily responsible for stabilising tight attachment to host cells.

The LEE T3SS is a multi-protein complex that spans the bacterial inner- and outer-membrane and acts as a conduit for the exported effector proteins. This leads to the production of an 'injection needle' that fuses to host cell membranes, ultimately facilitating the translocation of effector proteins into the host cell cytoplasm. The T3SS is utilised by wide range of Gram-negative animal and plant pathogens including species of *Salmonella*, *Yersinia*, *Shigella*, *Pseudomonas*, *Chlamydia*, *Citrobacter*, *Burkholderia* and *Bordetella* [33].

The LEE T3SS apparatus comprises three main structural features: the basal body, the needle filament and the translocon. The basal body of the LEE T3SS sits in the inner- and outer-membrane. EscJ is inserted into the inner-membrane, by the Sec secretory system, where it oligomerises to form a 24-mer ring structure that protrudes into the periplasmic space [34]. This structure is thought to act as a platform for the assembly of the rest of the LEE T3SS, catalysing the production of the needle complex [34]. EscJ is anchored into the membrane by N-terminal lipidation and a C-terminal transmembrane helix. Following the insertion of EscJ several other proteins are recruited to form the basal apparatus, namely EscS, EscT, EscR and EscV that sit in the cytosolic side of the inner-membrane. EscU is known as the 'switch protein' as it is thought to be partially responsible for regulating the export of protein substrates [35]. EscN is an ATPase and is thought to energise the export function of the LEE T3SS as well as forming interactions with EscL and EscQ [36]. SepL and SepD are interesting proteins as they are thought to act as the gatekeepers of the LEE T3SS by inhibiting the secretion of effector proteins until the needle complex is complete [37,38]. Both SepL and SepD are associated with the cytoplasmic side of the inner membrane [37,38].

EscC is associated with EscJ and forms an oligomeric ring structure comprising 20 subunits in the outer-membrane [39]. EscC has two domains, the N-terminal domain, thought to reside predominately in the periplasm, and the C-terminal domain, which is inserted into the outer-membrane [39].

Between the inner- and outer-membrane the initial part of the needle is formed by EspF, which interacts with both EscJ and EscC. EspF forms the base for the assembly of the needle filament, EspA. EspA self-assembles in a helical arrangement with approximately 5.6 subunits per turn [40]. In order to prevent its premature polymerisation, EspA is secreted with its chaperone protein CesA [34]. Once the needle has reached the designated length (up to 600 nm) the translocon proteins, EspD and EspB are secreted through the needle complex and inserted into the host cell membrane forming a heteropentameric complex with an EspD:EspB ratio of 4:1 [41].

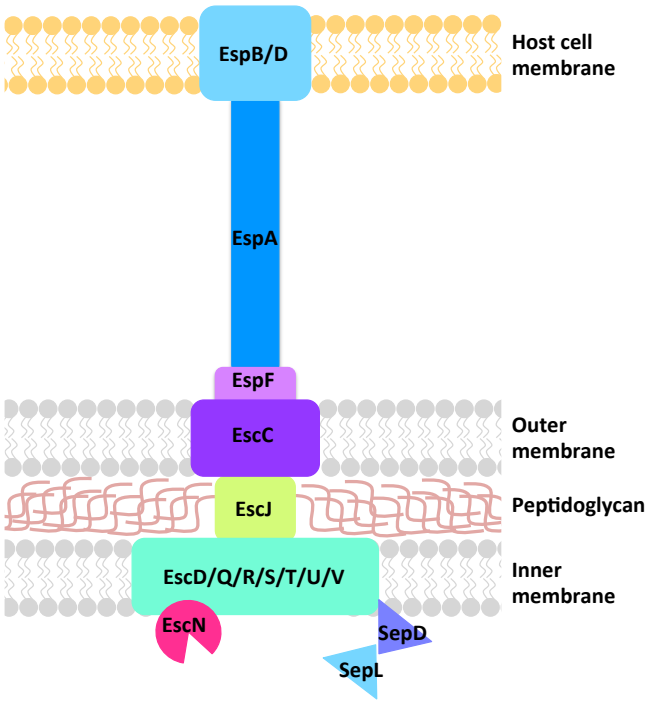


Figure 1-3 The type three secretion system apparatus. The LEE T3SS crosses the bacterial inner- and outer-membrane producing a needle filament that inserts into the host cell membrane. How this complex is assembled is described in the text above.

The overall structure of the T3SS is highly conserved between pathogens, however the nomenclature differs between species. Table 1-1 lists the homologous proteins between *E. coli, Yersinia,* and *Salmonella.* The T3SS from different pathogens will be referred to as

follows: *E. coli*, LEE T3SS; *Yersinia,* Ysc T3SS, *Salmonella,* SPI1 T3SS. Since the T3SSs from different pathogens are highly conserved findings from one pathogen can be applied to the others. For example, the mouse pathogen *Citrobacter rodentium* is often used as a model of EHEC infection since EHEC does not naturally colonise mice. In this pathogen the T3SS is also encoded on a LEE pathogenicity island, therefore the nomenclature of the of T3SS from *C. rodentium* is the same as for *E. coli* [25,42].

	E. coli	*Yersinia*	*Salmonella*
Basal body	EscC	YscC	InvG
	EscD	YscD	PrgH
	EscJ	YscJ	PrgK
Export apparatus	SepQ	YscQ	SpaO
	EscN	YecN	InvC
	EscR	YscR	SpaP
	EscS	YscS	SpaQ
	EscT	YscT	SpaR
	EscU	YscU	SpaS
	EscV	YscV	InvA
Translocon	EspD	YopB	SipB
	EspB	YopV	SipC
	EspA	LcrV	SipD
Needle	EscF	YscF	PrgI

Table 1-1 Nomenclature of homologous T3SS proteins found in *E. coli*, *Yersinia* and *Salmonella*. The function of the proteins has been indicated. The nomenclature of the T3SS in the mouse pathogen *C. rodentium* is the same as for *E. coli.*

1.2.3 *Effector proteins*

The LEE encodes seven effector proteins (Tir, Map, EspF, EspG, EspZ, EspH and EspB) as well as the adhesin, intimin, all of which are required for the formation of A/E lesions. Tir is an essential virulence determinant, owing to its role in stabilising attachment to host cells [43]. Following the formation of the LEE T3SS needle complex, Tir is one of the first effector proteins to be translocated into the host cell, where it is inserted into the membrane. Once in the membrane it binds to the bacterial outer-membrane protein, intimin, using its extracellular domains [43]. The intracellular domain of Tir possesses several phosphorylation domains facilitating its interaction with hostcell proteins such as N-WASP. N-WASP recruits the actin related protein 2/3 (Arp2/3) resulting in the

nucleation of actin filaments, a process essential to the A/E lesion formation [44]. The other LEE-encoded effectors similarly act to manipulate the host cell cytoskeleton (e.g. EspB [45] and EspF [46]) and disrupt mitochondrial function (e.g. MAP [47] and EspF [46]).

Up to 60 non-LEE encoded effectors (NLE) have been identified [25,48]. These are often encoded within pathogenicity islands or on mobile genetic elements [49]. The role of NLEs in infection has been discussed in several excellent reviews and will not be discussed further here [49].

1.2.4 *Flagella*

Flagella are assembled on the cell surface by the flagellar T3SS. The LEE T3SS and the flagellar system share several homologous proteins, especially within the basal body of these structures. The LEE T3SS and other T3SSs associated with virulence and the secretion of effector proteins have evolved from the flagellar system, which is thought to have occurred multiple times throughout bacterial evolution [50,51]. Like the LEE T3SS, the flagellar system facilitates the export of protein subunits leading to the sequential formation of a multi-protein complex. A comparison of the two structures is shown in Figure 1-4.

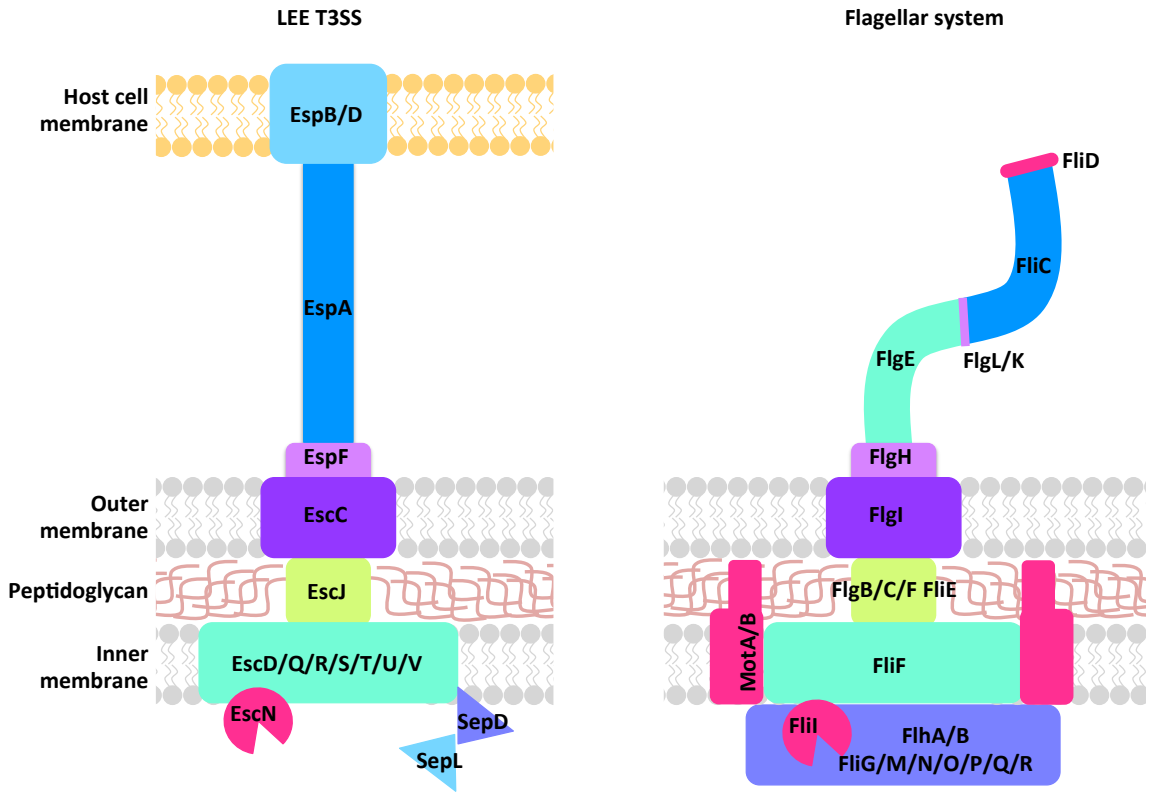


Figure 1-4 Comparison of the structure of the LEE T3SS and the flagellar system. Similar positioning and colouring highlight the functions of homologous proteins conserved between both systems.

The flagellum comprises 3 mains parts the filament that propels the bacterium, the hook and the basal body. As with LEE T3SS the flagellar system is assembled in a step-wise fashion, the specific details of which are reviewed extensively elsewhere and therefore will only be discussed briefly here [51-53]. The flagellar basal body is the first to be assembled in the membrane. It consists of six integral membrane proteins (FlhA, FlhB, FliO, FliP, FliQ, FliR) and three cytoplasmic proteins (FliH, FliI, FliJ) [53]. The basal body includes a motor-switch complex capable of utilising ATP to generate torque within the protein filament resulting in force that propells the bacteria [54]. Subsequently, the 'rod' of the flagellar complex is assembled from FliE, FlgB, FlgC, FlgF and FlgG. The hook (FlgE) joins the filament to the basal body and allows the rotation of the filament [52].The propeller of the flagellum is a helical filament that is composed of up to 20,000 subunits of a single protein (FliC) that is capped by a scaffold protein (FliD).

Bacterial swimming is a process that involves the rotation of flagella in both a clockwise and an anti-clockwise direction. The direction of flagella movement is controlled through the 'switch complex', which is composed of FliG, FliM, and FliN [52]. The directionality of the flagellar rotation is influenced through several singling pathways. For example,

phosphorylation or acetylation of the chemotactic protein CheY increases the bias of rotation towards the clockwise direction [55–57]. Movement of flagella in an anti-clockwise direction leads to the coordination of the peritrichous flagella (as is the case in *E. coli*) resulting in a 'flagellar bundle' which propels the bacterium forward thereby enabling the bacterium to direct its movement towards a chemotactic source [54]. Clockwise rotation prevents the coordination of flagellar movement, thereby causing the bacterium to 'tumble', i.e. to move without direction. This random reorientation enables the bacterium to change direction.

The flagellum is key to the infection process as it not only enables to bacterium to move towards its target tissues but it has also been shown to be important for the initial stages of host cell contact [58,59]. However, flagellin is extremely immunogenic as it is the main ligand of Toll-Like receptor five (TLR5) [60], therefore its expression must be strictly regulated.

There are three classes of promoters responsible for transcribing the flagellar operons [61]. The class I operon includes genes for the master regulators FlhD and FlhC, which dimerise and bind to the class II promoter. The class II operon includes the *fliA* gene, which is the flagellar specific sigma factor (σ^{28}), required for the expression of class III genes [61]. The expression of flagellar genes is regulated by growth stage, nutrient depletion, and quorum sensing molecules; which enable the bacteria to respond to the depletion of resources that is dependent on the population density. Furthermore, environmental signals also affect the expression of flagella. For example, short chain fatty acids (SCFAs), such as butyrate, propionate and acetate are found in concentrations ranging from 20 to 140 mM in the large intestine. SCFAs have been shown to increase the expression of flagella, indicating that they act as signalling molecules informing bacteria of their positions within the GI tract [62].

1.2.5 *Shiga toxin*

As mentioned in Section 1.1.1, Stx is the main virulence determinant of EHEC and has been reviewed extensively [63–65] therefore only a brief introduction to Stx will be given here. Stx is an AB_5 toxin, composed of the catalytic subunit A which is non-covalently bound to a pentamer of B subunits [66]. The catalytic subunit has N-glycosidase activity

that cleaves an adenosine residue from the 28S ribosomal RNA of the 60S ribosomal subunit [67], thereby inhibiting protein synthesis and leading to cell death. Subunit A enters the cell, following binding of the B subunit pentamer to the globotriaosylceramide (Gb3) receptor on the surface of some eukaryotic cells, including intestinal and kidney epithelial cells [68]. Upon receptor binding, Stx is endocytosed by the eukaryotic cell leading to apoptosis [48]. Since apoptosis is occurring specifically in the intestine and kidney the manifestation of Stx is bloody diarrhoea and renal damage, which can result in HUS [65].

1.2.6 *Other virulence factors*

EHEC O157:H7 strains carry the 93 kb pO157 plasmid that encodes several virulence factors such as KatP, Ehx, and EspP. KatP is a catalase-peroxidase which is thought to contribute to the protection of the bacterium from peroxide-mediated oxidative damage [69]. Ehx (enterohaemolysin) is a pore-forming toxin that lyses erythrocytes [70,71]. The role of Ehx in virulence is unclear, however it has been implicated in the development of HUS since it was shown to be cytotoxic to endothelial cells [72]. The serine protease EspP cleaves human coagulation factor V and complement proteins. The activity of EspP is thought to contribute to the development of HUS since Stx has been shown to induce the complement cascade [73]. Interestingly the activity of Ehx has been shown to be regulated by EspP, demonstrating how the concerted activity of virulence factors regulates the infection process [74].

1.3 Current therapies

Currently the treatment of enteric *E. coli* infection is only supportive. For example, oral rehydration therapy is often life saving in infants and children with diarrhoea. The use of antibiotics is not recommended to treat infection, especially in the case of EHEC, owing to the increase in disease severity caused by increased release of Stx from lysing cells [75]. Furthermore, treatment with antibiotics can induce the general stress response, which can up-regulate the expression of Stx [27]. There are several areas of ongoing research targeted at reducing EHEC infection and limiting the side effects caused by Stx release, a few of which are mentioned briefly below.

To reduce the damage caused by Stx, several therapies have been developed to bind to and inactivate this protein. SYNSORB Pk (SYNSORB Biotech, Alberta, Canada) is a drug that acts as an Stx receptor mimic and constitutes an inert silicone dioxide support linked to the trisaccharide moiety of the Gb3 receptor. Therefore SYNSORB Pk will compete with the Gb3 receptor for binding free Stx in the bloodstream, thus reducing tissue damage otherwise caused by Stx binding [76]. The drug was tested in humans, however it failed to significantly reduce disease severity which was attributed to the poor uptake of the drug into the bloodstream [76]. An alternative method to reduce the effects of Stx is to reduce the expression of the Gb3 receptor. A study by Silberstein *et al.* showed that an inhibitor of glucosylceramide synthase (C-9) reduced the expression of the Gb3 receptor in a human kidney cell-line [77]. In this study, rats injected with Stx died within 7 days; however treatment with C-9, both prophylactically and following infection, reduced mortality by 50% [77]. These are only two examples of anti-virulence approaches that target Stx in order to reduce the severe side effects of EHEC infection. This is an active area of research that will hopefully yield a novel therapy capable of preventing kidney damage caused by EHEC.

Several groups are investigating the use of a vaccination programme to reduce the levels of EHEC infection. A number of different approaches to the development of a vaccine are being investigated. These include using an inactivated form of Stx as an antigen, an approach that has shown some protection in a mouse model [78]. Some studies are aimed at reducing the carriage of EHEC in cattle, thereby reducing the possibility of food chain contamination (recently reviewed in [79]). Other investigations are testing the use of virulence factors as antigens, such as EspB and intimin which were shown to elicit an immune response in rabbits [80]. The possibility of generating a live attenuated vaccine for EHEC has also been investigated using an attenuated EPEC strain. EPEC strains lacking important effector proteins such as EspB and intimin were tested in human volunteers [80,81]. However both deletion mutants still caused diarrhoea in some patients indicating that a live vaccine may not be suitable for the prevention of this disease [80,81].

1.4 Anti-virulence compounds

1.4.1 *Rationale*

The lack of treatment options is not limited to EHEC infection; the increase in antibiotic resistance is a global problem for the effective treatment of both Gram-negative and Gram-positive infections. Traditional antibiotics have been selected for their ability to kill (bactericidal) or inhibit growth (bacteriostatic) *in vitro* and typically target essential cellular processes, such as peptidoglycan synthesis, transcription or translation. As the survival of the bacterium is being compromised, the selective pressure for the evolution of resistance genes is high. Therefore targeting virulence factors, that are not essential to the survival of the bacterium *per se*, may reduce the selective pressure driving the development of resistance mechanisms. Theoretically, this is one of the main advantages of an anti-virulence approach over current therapies.

Reducing the virulence of a pathogen *in vivo* would allow the host's immune system to clear the infection, which is clinically similar to vaccinating with a live attenuated strain. Specifically targeting virulence factors that are encoded only by pathogenic strains will not affect the host's natural flora. The maintenance of the natural flora, e.g. gut flora, will reduce the likelihood of opportunistic bacteria such as *Clostridium difficile* colonising and causing further disease [82]. Since the use of anti-virulence compounds to treat bacterial infections is a relatively new approach, reports on the activity of these compounds *in vivo* is limited. Examples of some of the different anti-virulence strategies being investigated will be covered briefly below followed by a detailed description of compounds that inhibit the LEE T3SS.

1.4.2 *Quorum sensing inhibitors*

The expression of virulence factors is a highly regulated process that requires input from several different signaling molecules. In order for a pathogen to effectively overcome the host defenses, the expression of virulence factors is often coordinated within a bacterial population. This requires cell-to-cell communication, a process known as quorum sensing, that uses signaling molecules that are secreted from growing bacteria. The concentration

of these signaling molecules is population size dependent [51]. Numerous diverse biological processes in both Gram-positive and Gram-negative bacteria are regulated by quorum sensing, including bioluminescence, sporulation, toxin release, virulence gene expression and biofilm formation.

In Gram-negative bacteria the signaling molecules produced are acylhomoserine lactone (AHL) molecules. AHLs are synthesised from the metabolic intermediate S-adenosylmethionine by a species specific LuxI homologue. Once AHL reaches its critical concentration the species-specific LuxR homologue (a transcriptional activator) induces the expression of a particular subset of genes. Disrupting quorum sensing would inhibit the expression of genes activated by LuxR. Inhibition of quorum sensing is widely studied and several approaches to inhibiting this system are being explored. For example, inhibiting LuxR with structural analogues of AHL has been shown to be effective in reducing virulence in several species including *Pseudomonas aeruginosa*, which encodes two quorum sensing systems. Following a screen of natural products for inhibition of quorum sensing, a halogenated furanone compound was identified [52]. Further *in vivo* studies with the compound indicated that it could reduce the virulence of this pathogen in a mouse lung infection model [52].

An alternative approach to inhibiting quorum sensing is to degrade AHL once it has been synthesised. This reaction is carried out by AHL lactonases which are found in some Gram-positive (e.g. *Bacillus*) species. Transgenic plants expressing bacterial AHL lactonases are less susceptible to *Erwinia carotovora* infection [53]. However, the use of a 'quorum quenching' approach in animals has not been reported so far.

1.4.3 *Adhesin inhibitors*

As mentioned above, UPEC is a pathotype of *E. coli* that causes UTIs. A key virulence factor that enables UPEC to effectively colonise all parts of the urinary tract is the type 1 pilus. Therefore inhibiting the production, assembly and binding of pili to host cells has been an active area of research [83]. Attachment of pili to host cells is mediated by the adhesin FimH, which specifically binds to mannosylated receptors on the host cell membrane [84]. FimH is essential to intracellular invasion of UPEC and therefore key to

the infection process. An understanding of the importance of FimH *in vivo*, the structure of FimH and an understanding of its binding activity opened the door for the development of highly specific inhibitors. The development of mannose derivative compounds and optimisation of their bioavailability resulted in compounds that reduce UPEC infection in a mouse model of chronic cystitis [85]. Administration of the compounds pre-infection reduced bacterial colonisation as well as administration during an established infection [85]. This represents an important step towards the use of anti-virulence compounds in a clinical setting. The majority of compounds discussed have been shown to be effective in reducing bacterial virulence only when the bacteria are treated with the compound before the infection. However this does not mimic a real-life scenario where a bacterial infection is established and is then treated.

1.4.4 *Inhibitors of the T3SS*

As discussed earlier (Section 1.2.2), the T3SS is a virulence factor utilised by a range of Gram-negative pathogens to attach to host cells and secrete effector proteins that manipulate host cell processes. The T3SS has been shown to be essential *in vivo* to the virulence of several pathogens, for example deletion of the Ysc T3SS renders species such as *Y. pestis* avirulent [86]. This makes inhibiting the T3SS an attractive prospect. Many groups have undertaken the search for T3SS inhibitors over the last decade and several approaches have been used to screen for potential inhibitors. These include enzyme-linked immunosorbent assays (ELISA) detecting specific secreted proteins, monitoring of host cell responses, or the inhibition of the T3SS expression that is reported by a gene:reporter fusion plasmid. Each of these approaches has yielded some potential inhibitors. The gene:reporter fusion plasmid system has been used most widely as it can easily be translated to a high throughput (HTP) screen to test large libraries of compounds. Several inhibitors of the T3SS have been identified which are discussed below. The most extensively studied inhibitors are the salicylidene acylhydrazides which will be discussed separately in Section 1.4.5. Table 1-2 shows the chemical structures of the compounds discussed below.

The first reported chemical inhibitor of the LEE T3SS was identified in 2002 by Linington *et al.* who screened chemical extracts from the marine sponge *Caminus sphaeroconia* against EPEC [87]. The screen looked for compounds that decreased the secretion of EspB

and displayed no antibacterial activity. The product caminoside was found to have these properties and had an IC_{50} of 5.1 µg/ml. Despite the promise of this compound, its cellular targets were not identified, owing to the difficulty of synthesising this natural compound [88].

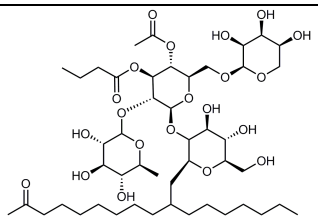
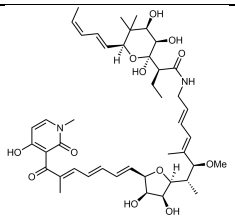
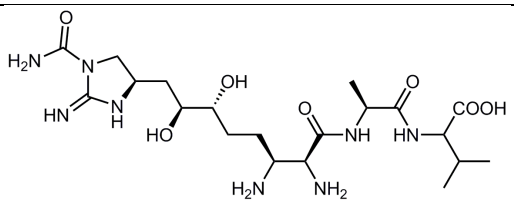
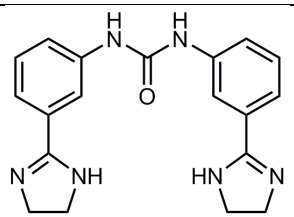
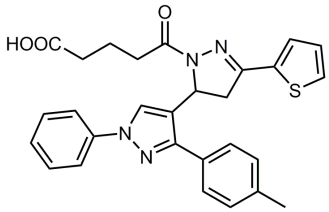
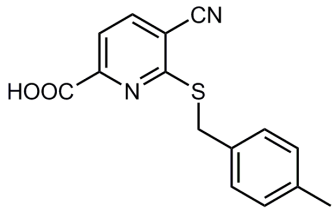
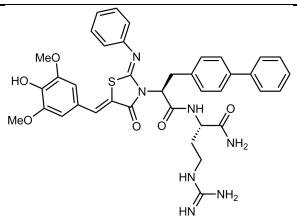
Compound	Effective against	Structure	Reference
Caminoside	EPEC (LEE T3SS)		[87]
Aurodox	EPEC (LEE T3SS) *C. rodentium* (LEE T3SS)		[89,90]
Guadinomines	EPEC (LEE T3SS)		[91]
Diproprionate	*Y. pestis* (Ysc T3SS) EPEC (LEE T3SS)		[92]
Pentanoic acid derivative	*Y. pestis* (Ysc T3SS)		[92]
Picolinic acid derivative	*Y. pestis* (Ysc T3SS) EPEC (LEE T3SS)		[92]
2-imino-5-arylidene thiazolidinones	*S. typhimurium* (SPI1 T3SS)		[93]

Table 1-2 Chemical structure of T3SS inhibitors effective against Gram-negative pathogens. The species and the T3SS against which the compounds are effective have been listed. The chemical structures were drawn using ChemDraw (PerkinElmer, UK).

Several other natural products have been shown to decrease the expression of the T3SS. Aurodox produced by *Streptomyces goldiniensis* was initially described to have anti-bacterial activity towards Gram-positive organisms and was used in livestock as a growth-promoter [89]. Aurodox was recently shown to inhibit LEE T3SS mediated haemolysis in EPEC *in vitro* with an IC_{50} of 1.8 μM [90]. It was also shown to be effective *in vivo* when tested in a mouse model of EPEC infection using the mouse pathogen *Citrobacter rodentium,* where mice treated with Aurodox survived a lethal bacterial load [90]. Aurodox was originally shown to bind to Elongation Factor-Tu, which explains its antibacterial activity [89]. However treatment of Gram-negative species with Aurodox results in a specific decrease in expression of the LEE T3SS suggesting that it may be interacting with a LEE T3SS transcriptional regulator.

Another class of compounds produced by a *Streptomyces* species (K01-0509) are the guadinomines, which were shown to inhibit LEE T3SS in EPEC *in vitro* with IC_{50} values of lower than 0.01 μg/ml [91]. Since the guadinomines appear to be highly potent with no antibacterial activity they are attractive lead compounds, however their efficacy *in vivo* has yet to be confirmed.

A recent HTP screen of over 70,000 compounds in *Y. pestis* revealed four compounds with novel structures (compared with currently identified compounds) that inhibited the Ysc T3SS [92]. This assay monitored growth following compound treatment since induction of Ysc T3SS in *Yersinia* inhibits growth. Therefore continued growth of the bacteria indicated that the Ysc T3SS had been inhibited [92]. The compounds diproprionate, a pentanoic acid derivative and a picolinic acid derivative had sub-micromolar IC_{50} values [92].

A novel class of compounds shown to have broad spectrum activity against T3S in several pathogens including *S. typhimurium, P. syringae, Francisella novicida* and *Y. enterocolitica* are the 2-imino-5-arylidene thiazolidinones [93]. A study by Felise *et al*. involved an initial HTP compound screen of over 92,000 small molecules that utilised a reporter strain to monitor the expression of SipA, a secreted effector protein produced by *Salmonella* which leads to host-actin reorganisation [93]. Compounds shown to be effective in reducing SPI1 T3S were further analysed for their inhibition of other cellular processes such as transcription, translation, sec-dependent secretion, and/or bacterial growth [93]. Only compounds that did not affect these processes were carried forward. The remaining

compounds, including the 2-imino-5-arylidene thiazolidinones, were tested for efficacy in the above pathogens and were found to have inhibitory effects on T3S and T2S but not on flagellar production [93]. From this the authors concluded that the compounds may be targeting a structural component common to both of the secretion systems, however there is currently no direct evidence for this. Also in this study a structure-activity relationship (SAR) analysis on the "hit" compounds was conducted and from this insights were gained into which chemical structures may be key to the activity of the compounds. Given the broad activity against several pathogens, the 2-imino-5-arylidene thiazolidinones are an exciting class of novel anti-virulence inhibitors.

1.4.5 *Salicylidene acylhydrazides*

The salicylidene acylhydrazides (SA) are a class of compounds that were identified from a chemical screen of 9,400 compounds carried out by Kauppi *et al.* at the University of Umeå [94]. The screen was performed on *Y. pseudotuberculosis* expressing a *yopE*:luciferase transcriptional fusion, where the *yopE* promoter was fused to a *luxAB* gene. YopE is a secreted effector protein; therefore a decrease in luciferase activity corresponded to decreased Ysc T3SS. Only compounds that showed no antibacterial activity were characterised further, leaving four lead compounds from the initial screen. These were all shown to affect the secretion of effector proteins (YopE, YopD and YopH) in a dose dependent manner with an IC_{50} of less than 50 μM. Owing to the structural similarity between the Ysc T3SS apparatus and the flagellum, the compounds were tested for inhibition of motility. Only one of the four compounds INP0010/ME0052 was shown to have an effect on motility, which may indicate that the compounds are binding to a related structural component or to a common regulator of these systems [94]. The structures of the SA compounds discussed in this Section are shown in Table 1-3.

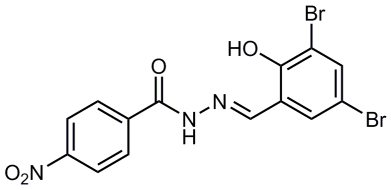
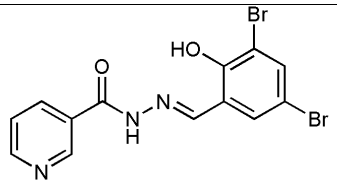
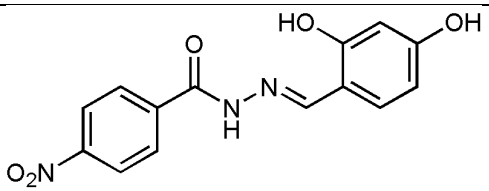
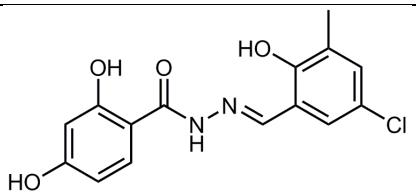
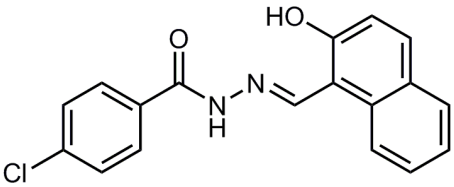
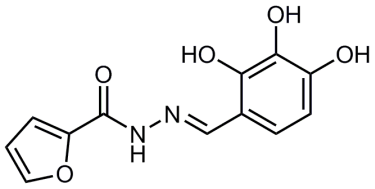
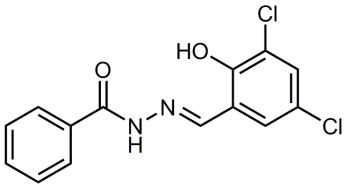
Compound	Effective against	Structure	Reference
INP0007	*Salmonella* (SPI1 T3SS) *Yersinia* (Ysc T3SS)		[94–97]
INP0010 / ME0052	*Salmonella* (SPI1 T3SS) EHEC (LEE T3SS)		[95,98–101]
INP0031 / ME0055	EHEC (LEE T3SS)		[97,99,100]
INP0341	*Chlamydia* (INC T3SS)		[102]
INP0400	*Salmonella* (SPI1 T3SS) *Shigella* (IPA T3SS)		[98,102–104]
INP0402	*Shigella* (IPA T3SS)		[104]
INP0403 / ME0053	*Salmonella* (SPI1 T3SS) EHEC (LEE T3SS)		[99,100,105]

Table 1-3 Chemical structures of salicylidene acylhydrazide compounds that inhibit T3SSs in Gram-negative pathogens. The species and the T3SS against which the compounds are effective are listed. The chemical structures were drawn using ChemDraw (PerkinElmer, UK).

A chemical screen for inhibitors of the LEE T3SS in EPEC using 20,000 compounds was carried out by Gauthier *et al.* in 2005 [106]. This screen initially looked for a decrease in EspB secretion using an enzyme-linked immunosorbent assay (ELISA). Once this criterion was met the compounds were then selected only if they specifically reduced LEE T3SS protein production, i.e. did not act by decreasing protein stability. This approach identified four compounds that were effective in inhibiting the expression of the LEE T3SS and did not directly interact with components of the LEE T3SS [106]. Interestingly these compounds are structurally similar to the salicylidene acylhydrazides identified by Kauppi *et al* [94]. In contrast to SA compounds, the compounds in this study were not shown to have effects on motility or the expression of flagellar genes [106].

Several groups later tested the SA compounds on a range of Gram-negative pathogens. The obligate intracellular pathogen *Chlamydia trachomatis* was shown to be affected by an SA compound (INP0400), identified in the original screen by Kauppi *et al*. [94,96,103]. Treatment with this compound disrupted the normal infection cycle and prevented differentiation and multiplication in mammalian cells [96,103]. At the time of the study relatively little was known about the role of the T3SS in this pathogen, thus the use of these compounds revealed insights into the regulation of the T3SS in *Chlamydia.*
The next pathogen shown to be susceptible to the SA compounds was *Salmonella enterica* serovar Typhimurium, in which the authors showed dose dependent inhibition of SPI-1, one of the two T3SSs encoded in *Salmonella* [97]. The compounds (e.g. INP0031/ME0055) had no effect on the growth of the pathogen *in vitro* and reduced SPI-1 mediated infection of HeLa cells by up to 60%. The study also showed that pre-incubation of the bacteria with the compounds before use in an *in vivo* bovine intestinal ligated loop model reduced the levels of inflammation indicating that the compounds reduced the virulence of *Salmonella in vivo* [97]. Further studies in *Salmonella* by Negrea *et al.* confirmed the ability of the SA compound (INP0400) to inhibit SPI-1 activity [98]. They also showed the compounds to be effective inhibitors of SPI-2 mediated secretion, and that treatment with the compounds reduced intracellular replication. Two of the nine compounds tested were shown to significantly reduce the motility of *Salmonella* in soft agar [98].

Veenendaal *et al.* showed that the SA compounds INP0402 and INP0400 were the most effective at reducing IPA T3SS in *Shigella flexneri,* an invasive intracellular Gram-negative

pathogen [104]. Treatment with the compound reduced its ability to invade HeLa cells and its ability to induce macrophage apoptosis, both indicative of T3SS inhibition [104].

The activity of the SA compounds against EHEC was tested by Tree *et al*. who showed the compounds to be effective at decreasing LEE T3S in a dose dependent manner [99]. This study showed INP0031 to be the most effective compound at inhibiting LEE T3S and A/E lesion formation. All of the compounds tested (INP0010, INP0103, INP0401 and INP0031) increased the production of flagella suggesting that the regulation of this factor is being affected by the compounds [99].

In summary, the SA compounds have been shown to be effective inhibitors of T3S in several species of Gram-negative pathogens. In all studies the authors ensured that no antibacterial activity was observed since anti-virulence compounds should not decrease the survival of the pathogen. Several studies showed that in addition to affecting the T3SS, the expression of motility genes was also affected, however the effects between species were not consistent. Although it is known that the compounds inhibit T3SS, the mechanism of action is yet to be elucidated. The proposed mechanisms of action of the compounds are discussed below.

1.4.6 *Mechanism of action of the SA compounds*

There are three main schools of thought about how the SA compounds function. The compounds have been proposed to act by:

i. disrupting cellular iron stores
ii. directly interacting with a component of the T3SS
iii. affecting the regulation of the T3SS

The finding that the activity of the compounds could be reversed following the addition of iron to the cell culture media was first reported by Slepenkin *et al*. [102]. This study showed that the addition of iron to HeLa cells infected with *C. trachomatis* reversed the effects of the inhibitors. This effect was not seen when other divalent metal ions were added. However, these results were somewhat inconclusive since INP compounds that did not affect the T3SS in *Chlamydia* (INP0406) chelated iron as well as INP0341, a potent inhibitor [102]. A similar study by Layton *et al*. also indicated that the effect of the SA

compounds could be partially reversed by the addition of iron [105]. Transcriptomic analysis of *Salmonella* treated with INP0403 showed a significant increase in several genes involved in iron regulation [105]. However, the addition of iron did not fully reverse the anti-SPI1 T3SS activity of INP0403. Transcriptomic analysis carried out on EHEC grown in the presence of iron found that SA compounds (INP0010 and INP0031) lead to a significant decrease in LEE T3SS [99]. Thus indicating that in this case iron is not inhibiting the action of the SA compounds. Therefore it remains unclear how iron affects the activity of the SA compounds and further work is required to clarify the effects of iron on the LEE T3SS.

Veenendaal *et al.* proposed that the compounds were acting directly on a component of the T3SS. The reports of motility also being affected by the compounds led to the conclusion that the component being targeted may be one that is homologous between the T3SS and flagellar systems [104]. The evidence for this proposed mechanism was that following SA compound treatment of *Shigella*, the needle filaments of the IPA T3SS were significantly shorter than for untreated cells indicating that the compounds were affecting needle assembly [104]. A recent study by the same group sought to determine this common component in *Salmonella*. By using strains lacking different flagellar components (mostly membrane) they hoped to identify which of these were responsible for the change in motility seen following compound treatment [107]. However this study was unable to show that the SA compounds directly affected flagellar components. The authors concluded that the SA compounds were not directly inhibiting SPI1 T3SS or flagellar components and were most likely interacting with other targets within the cell, thus indirectly affecting the expression of these virulence factors [107].

To identify the protein targets of the salicylidene acylhydrazides the Roe lab carried out an affinity pull-down assay using ME0052 (INP0010) immobilised onto an Affigel matrix, the process of which is outlined in Figure 1-5 [100]. The compounds were incubated with EHEC cell lysate and the matrix was washed extensively to remove any non-specifically bound proteins. Specifically bound proteins were eluted using high concentrations of free compound and were separated using SDS-PAGE. The eluted proteins were identified by tandem mass spectrometry.

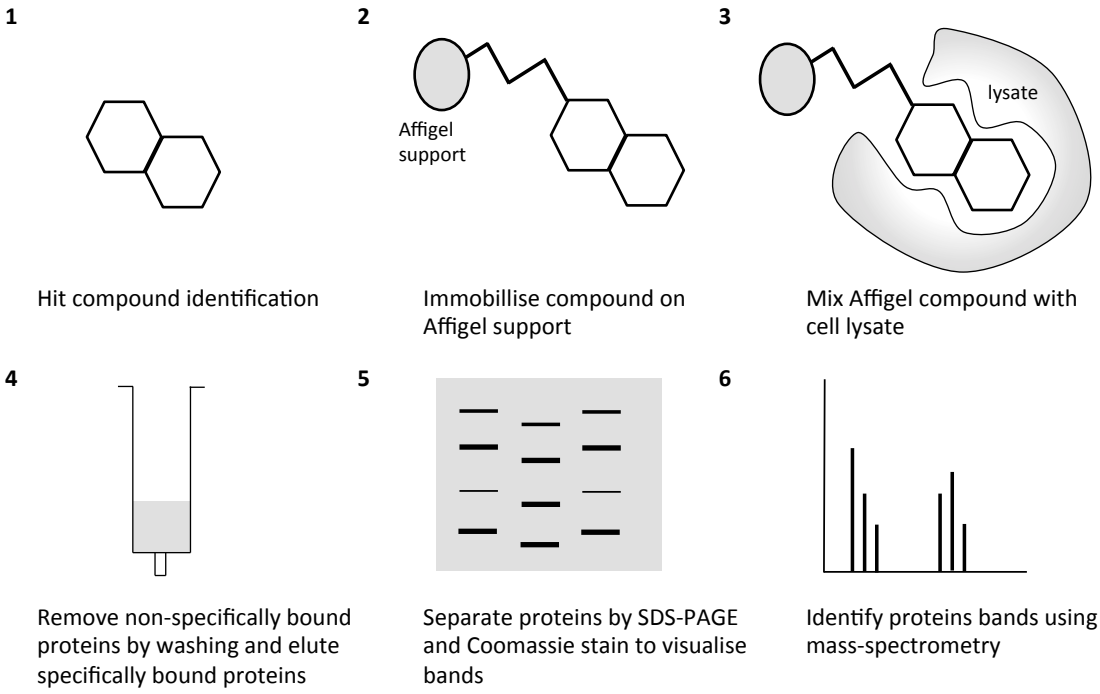


Figure 1-5 Overview of the affinity pull-down assay to determine the binding proteins of the salicylidene acylhydrazides. (1) The hit compound was identified from reporter:fusion assays looking for decreased LEE T3SS expression. **(2)** The hit compound was attached to an Affigel support. **(3)** The Affigel labelled compound was mixed with cell lysate, allowing the putative targets to bind. **(4)** Non-specifically bound proteins were removed by washing and specifically bound proteins were eluted with high concentrations of the free compound. **(5)** Eluted proteins were resolved by SDS-PAGE and protein bands visualised by Colloidal Blue Stain and excised. **(6)** Protein bands were identified by mass-spectrometry.

This assay identified nineteen putative targets of the SA compounds which are listed in Table 1-4 which were identified from a SDS-PAGE gel shown in Figure 1-6. This thesis presents an investigation of a selection of these protein targets including the cytoplasmic proteins Tpx, FolX, WrbA, and AdhE and the periplasmic proteins FklB, FkpA and SurA. Each protein target will be introduced at the beginning of the relevant Results chapter.

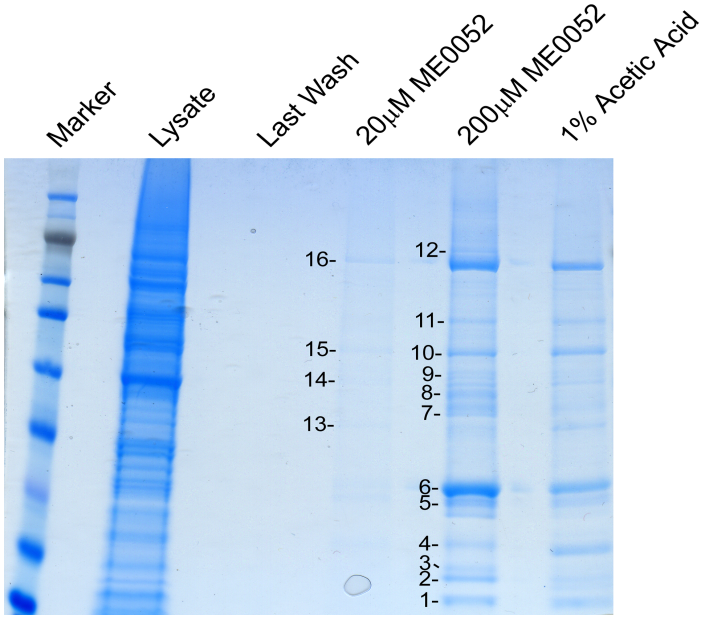


Figure 1-6 SDS-PAGE gel of the protein fractions from the affinity pull-down assay. Lysates from *E. coli* O157 were incubated with Affigel labelled with ME0052 and washed extensively (the last wash fraction is shown here). Specifically bound proteins were eluted using 20 and 200 µM unlabelled ME0052. The Affigel matrix was stripped using 1% acetic acid. Equal volumes of samples from each fraction were separated by SDS-PAGE and then visualised using colloidal blue stain. The bands were identified by mass spectrometry. The identity of each band is detailed in Table 1-4. This figure is adapted from [100].

1.5 Aims of the project

The overarching aim of this project was to determine which of the putative SA compound targets was responsible for the phenotype observed following treatment of *E. coli* O157:H7 with the SA compounds - a decrease in LEE T3SS and an increase in flagellar expression. Seven targets were selected for further investigation. For each selected target protein, the aim was to investigate the effect of gene deletion on the phenotype of *E. coli* O157. Alongside phenotypic studies, structural and functional characterisation of the target proteins was conducted. This required the use of a range of biophysical and biochemical techniques that enabled further investigation into the specific interactions occurring between the target proteins and the SA compounds. Understanding these interactions gave insights into the mechanism of action of the SA compounds and how they regulate the LEE T3SS.

Band	Genbank ID	Gene	Protein function	Cellular compartment
1	gi|15802850	*folX*	D-erythro-7,8-dihydroneopterin triphosphate 2' epimerase	cytoplasm
2	gi|15802098	Z2714	hypothetical protein	unknown
3	gi|15803192	Z3974	hypothetical protein	unknown
4	gi|15801846	*tpx*	thiol peroxidase	cytoplasm
5	gi|12515926	*yecD*	hypothetical protein	cytoplasm
6	gi|537048	*fklB*	FKBP-type 22 kDa peptidyl-prolyl *cis-trans* isomerase	inner-membrane / periplasm
6	gi|15800925	*wrbA*	tryptophan repressor binding protein	cytoplasm
7	gi|147379	*prs*	phosphoribosylpyrophosphate synthase	cytoplasm
8	gi|15804160	*tdh*	threonine dehydrogenase	cytoplasm
9	gi|42497	*proB*	gamma-glutamate kinase	cytoplasm
9	gi|15802064	*nemA*	N-ethylmaleimide reductase	cytoplasm
9	gi|396292	*gldA*	glycerol dehydrogenase	cytoplasm
10	gi|15799738	*surA*	survival protein A	periplasm
11	gi|15804455	*glnA*	glutamine synthetase	cytoplasm
12	gi|15801467	*adhE*	CoA-linked acetaldehyde dehydrogenase	cytosol
13	gi|15833452	*fkpA*	FKBP-type peptidyl-prolyl *cis-trans* isomerase	periplasm
14	gi|15801691	*yncE*	hypothetical protein	unknown
15	gi|1942721	Ef-Tu	protein chain elongation factor	cytosol
16	gi|52078252	*stcE*	metalloprotease precursor	secreted

Table 1-4 Putative target proteins of the salicylidene acylhydrazides from *E. coli* O157:H7. Target proteins identified in the study by Wang *et al.* [100]. The band column refers to the protein band presented in Figure 1-6 (band 6 and 9 appear multiple times as more than one protein was present in these bands).

1.6 Biophysical characterisation of protein targets

Several biophysical techniques were used throughout the work presented in this thesis to characterise the protein targets of the salicylidene acylhydrazides. The biophysical methods used include X-ray crystallography, small angle X-ray scattering and analytical ultracentrifugation and a brief introduction to these biophysical methods is given below.

1.7 X-ray crystallography

Protein X-ray crystallography is a widely used technique that allows the investigation of protein structure at the atomic level. This method of protein structure determination has given insights into the enzymatic mechanisms, protein-protein and protein-ligand interactions. This structural information provides the basis for structure based drug design, which can lead to the development of more specific drugs.

1.7.1 *X-ray crystallography theory*

X-rays are a form of electromagnetic radiation with wavelengths between 0.1 and 100 Å (1 Å = 0.1 nm). X-ray crystallography uses wavelengths of 0.8 - 1.6 Å which are of a similar scale to the atom bond lengths found in protein molecules and thus they can be used to resolve inter-atomic distances. When an X-ray interacts with an electron in a molecule it is scattered. X-ray crystallography relies on elastic scattering of X-rays where the wavelength of the incident X-ray beam is equivalent to the scattered X-ray.

As well as elastic scattering from electrons the other basic requirement of X-ray crystallography is the production of protein crystals. A protein crystal is a repeating pattern of molecules regularly arranged in three-dimensions. The smallest repeating unit in a crystal is known as the unit cell, defined by the length of three axes (*a, b* and c) and three angles between these axes (α,β and γ). The basic geometry of a unit cell can be classified into seven crystal systems namely: triclinic, monoclinic, orthorhombic, tetragonal, trigonal, hexagonal and cubic. Apart from the triclinic crystal system, each system has an associated minimal internal symmetry. The asymmetric unit describes the smallest unique unit within a unit cell, thus it follows that since that the triclinic unit cell has no internal symmetry, in this case, the unit cell and the asymmetric unit are the same. The asymmetric unit is the smallest amount of information required to reconstruct the entire crystal lattice.

Repeating patterns within a crystal can be used to draw 'lattice planes' (Figure 1-7). Lattice planes are equally spaced and separated by a distance (*d*). When a monochromatic X-ray beam is elastically scattered by an electron, the angle at which it is scattered is

equal to the angle of the incident beam (θ). In order to determine the distance between two lattice planes another incident X-ray needs to be scattered from a corresponding point on a different lattice plane. Assuming that the two X-rays are in phase with each other before reflection, the path-length between the planes must be an integer (*n*) of the wavelength (λ), or *n*λ, in order for them to remain in phase. Therefore when the wavelength and the reflection angle are known, the distance between the lattice planes can be calculated. This forms the basis of Bragg's law (Equation 1):

$$n\lambda = 2d \sin\theta$$ **Equation 1**

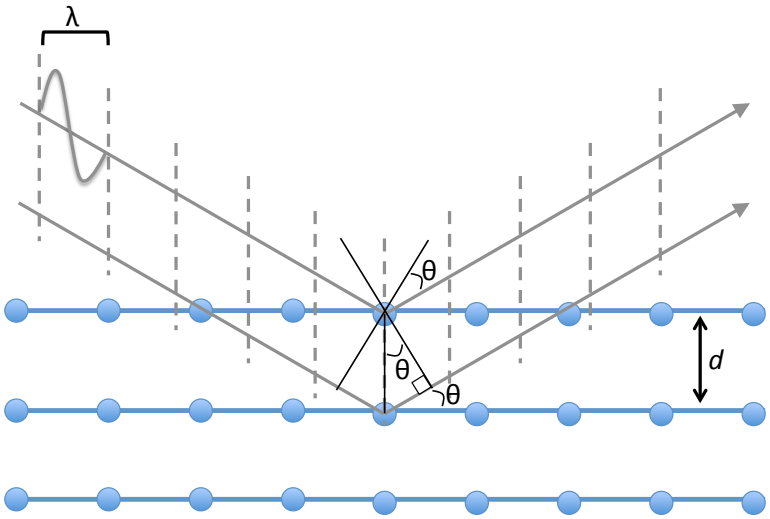


Figure 1-7 Incident X-rays are elastically scattered from lattice planes according to Bragg's law. Lattice planes are shown in blue with repeating lattice points symbolised by circles. The distance between the lattice planes is denoted by *d*. X-rays are represented as grey lines; dashed grey lines indicate that the two X-rays are in phase with each other. The angle of reflection (θ) and the distance between the planes (*d*) have been labelled.

In a crystal, X-rays are scattered from multiple planes when they are in phase, i.e. undergoing constructive interference. They can be detected experimentally as diffraction maxima. Since the scattering derives from different lattice planes, the path-length travelled by the X-rays from each plane is different. This is described by the Laue conditions that describe the Miller indices; *h, k* and *l* relate the orientation of the unit cell edges (*a, b,* and *c)* to the difference in path-length. Each *hkl* reflection, or 'spot', recorded by the detector corresponds to waves scattered from a plane in a unit cell, for each of the unit cells within the crystal.

The positions of the reflections on the detector are dictated by the unit cell and the intensity of the reflections are dictated by the composition of the molecules in the unit cell. The sum of the scattered waves within a unit cell belonging to an *hkl* family is referred to as the structure factor, $F(h,k,l)$:

$$F(h,k,l) = \sum_j f_j \exp\left[2\pi i(hx_j + ky_j + lz_j)\right] \qquad \textbf{Equation 2}$$

In Equation 2, f_j is the resolution dependent atomic scattering factor, x_j, y_j and z_j are the fractional positions of the *j*th atom in the unit cell for each of the Miller indices. The data collected allow measurement of the intensities ($I(h,k,l)$) which gives information only about the amplitudes of the diffracted waves and not their relative phase. In order to generate an electron density map, phase information must first be obtained. There are several methods for obtaining phase information and the method used in this thesis was molecular replacement, which is discussed in Section 1.7.2.

1.7.2 *Data collection and processing*

For a given crystal orientation only a subset of *hkl* families will diffract, producing spots on the detector. Therefore by changing the orientation of a crystal, different *hkl* families will be brought into the correct position for diffraction to occur. The degree of crystal symmetry dictates how many orientations of the crystal need to be collected in order to obtain a complete data set, i.e. higher symmetry requires fewer orientations.

How the diffraction data are collected is dependent on the degree of symmetry within the crystal. A crystal with high symmetry requires fewer oscillations to collect a complete data set. For each oscillation a diffraction image is collected, therefore a data set may contain several hundred images. The first step in processing the images is to index the data. Indexing identifies all the reflections in the image and determines the space group and the unit cell parameters. The reflections from all the images are then integrated. In this study both of these functions were carried out using the program Mosflm [108]. Following integration, the data are then scaled (here using SCALA [109]) to remove systematic errors caused by radiation damage or slight deviations in the wavelength

throughout the experiment. An electron density map is then calculated using phases obtained either experimentally or from an existing model (i.e. molecular replacement).

Molecular replacement (MR) uses an existing structure of a homologous protein to calculate the phases, which are used in the calculation of the electron density map. One of the disadvantages of this method is the possibility of introducing model bias, where the calculated electron density does not correspond to the structure of the molecule being determined. This can be tested for through the generation of what are termed 'omit maps', where part of the model is deleted and the electron density recalculated. If the density remains the same this is indicative that model is not biased [110].

The initial map generated from a MR solution is used to modify the model, which is manipulated using molecular graphics programs such as COOT [111]. The modified model is then refined against the data to improve the fit of the model to the data. How well the model fits the experimental data is measured by the 'R-factor' (Equation 3), which compares the observed amplitudes ($F_O(h,k,l)$) with the calculated amplitudes ($F_C(h,k,l)$):

$$R = \frac{\sum |F_O(h,k,l) - F_C(h,k,l)|}{\sum F_O(h,k,l)} \qquad \textbf{Equation 3}$$

The R-factor is compared with the R_{free} which is derived from a set of data that was not used in the refinement calculations. The R_{free} factor is used to monitor whether the data are being overfitted.

X-ray crystallography is a powerful tool for looking at protein structure, however in a crystal lattice proteins are subjected to crystal packing forces. This may disrupt the structure of the protein so that the structure obtained may not represent the conformation of the protein under physiological conditions. Furthermore, flexible regions are often difficult to observe, which means that these regions are often not modelled in crystal structures. Therefore the combination of structural data obtained from X-ray crystallography and complementary structural techniques is becoming increasingly common.

1.8 Small angle X-ray scattering

Like X-ray crystallography, small angle X-ray scattering (SAXS) relies on the coherent scattering of X-rays from proteins. However, unlike in crystallography the scattering molecules are randomly oriented in solution, thus the scattering data are derived from all possible orientations of the protein present in solution. This therefore reduces the resolution limits of this technique (10 - 20 Å) compared with X-ray crystallography where sub-Ångstrom resolution data can be obtained. SAXS data give details about the size and overall shape of the molecule and can be used to generate an *ab initio* model of the protein. This technique is applicable to a large range of molecular dimensions from micrometers to nanometers. Since the molecules are free in solution, information about the flexibility within the system can be resolved from SAXS data offering an advantage over other structural methods.

1.8.1 *SAXS data collection and theory*

SAXS data are collected on a 2D detector, as in X-ray crystallography, and since the molecules are randomly oriented in solution there is no need to reorient the sample. Therefore the data collected by the detector are isotropic and radial averaging of the data improves the signal to noise ratio (Figure 1-8). The intensity of the scattering is a function of the momentum transfer s (Equation 4):

$$s = \frac{4\pi \sin\theta}{\lambda} \qquad \textbf{Equation 4}$$

Where λ is the wavelength of the incident beam and θ is half of the scattering angle.

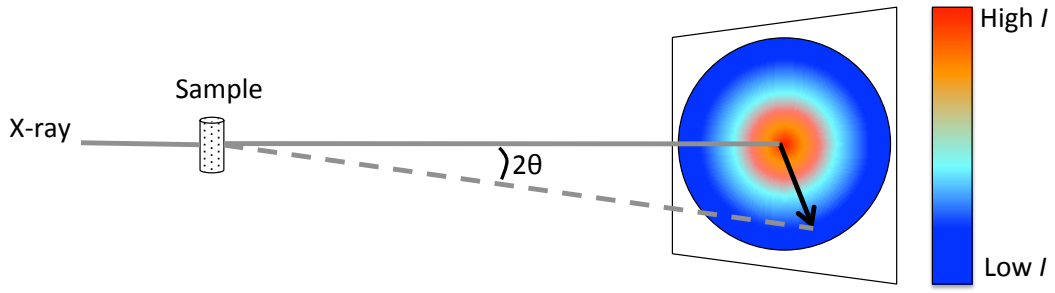


Figure 1-8 SAXS data collection. The protein sample is placed into a monochromatic X-ray beam and the scattered X-rays are measured on a 2D detector. The scattering angle is 2θ. At synchrotron X-ray sources the distance between the sample and the detector is approximately 2.7 m.

In a SAXS experiment scattering curves for the solvent and the protein in solution are collected. Both measurements are required as the scattering intensity of the protein is derived by subtracting the scattering of the solvent from the total scattering (from protein and solvent). Proteins scatter weakly with an average electron density of ~ 0.44 $e^-/Å^3$ compared with water which has an electron density of ~ 0.33 $e^-/Å^3$ [110]. Subtraction of solvent scattering data from the total scattering results in a scattering curve, an example of which is shown in Figure 1-9.

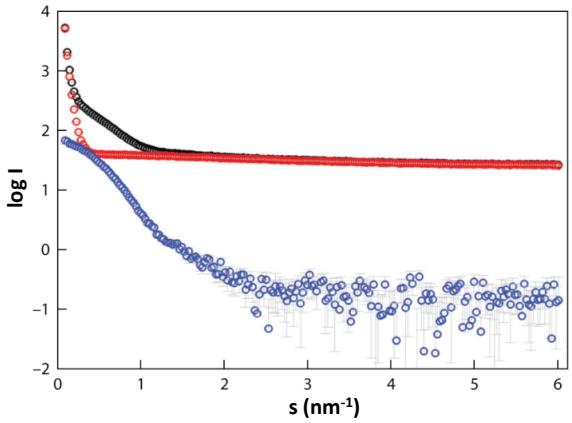


Figure 1-9 SAXS scattering curves. Example scattering curves for buffer solvent (red), protein in buffer solution (black) and the resultant scattering curve obtained by subtracting the solvent scattering curve from the protein in solution scattering curve, i.e. scattering from the protein alone (blue). This figure is adapted from [112].

Once the background scattering is accounted for, the scattering intensity $I(s)$ is proportional to the scattering of molecules in solution averaged over all orientations, which is described by the Debye equation:

$$I(s) = \sum_{i=1}^{N}\sum_{j=1}^{N} f_i(s) f_j(s) \frac{\sin(sr_{ij})}{sr_{ij}} \qquad \textbf{Equation 5}$$

Equation 5 shows that the sum of the intensities of scattering from N spherical atoms is multiplied by the scattering factors of the molecules (f) and the distance between scattering atoms (r_{ij}) averaged over all orientations. Several parameters can be extracted from SAXS scattering data including information regarding the sample quality (e.g. the presence of aggregation), the molecular mass of the protein, the radius of gyration (R_g) and the maximum dimension (D_{max}) of the protein in solution.

1.8.2 *SAXS data processing*

Initial analysis of scattering data involves analysis of the low angle part of the scattering curve, known as the Guinier region. The Guinier approximation (Equation 6) is valid only for restricted s ranges ($0 \leq sR_g \leq 1.3$ for a sphere, for a rod $0 \leq sR_g \leq 0.7$) and from this approximation the R_g and $I(0)$ of a protein can be determined. The parameter R_g describes the square root of the average squared distance of each electron from the particle centre. Plotting $log(I(s))$ against s^2 gives a linear plot in the Guinier region, the slope of which gives R_g and when extrapolated to zero angle $I(0)$ can be used to estimate the molecular mass of the protein. R_g is determined interactively using the program PRIMUS [113]. In theory, on an absolute scale $I(0)$ should equal the square of the number of electrons in the scattering particles. However, in practice this value is calculated from measurements of a well characterised protein with a known molecular weight, usually bovine serum albumin. However measurement of this parameter is highly susceptible to experimental problems such as the presence of aggregates in the sample or uncertainty in the concentration [114].

$$I\,(s) = I\,(0) \cdot \exp\left(-\frac{1}{3}R_g^2 s^2\right) \qquad \textbf{Equation 6}$$

Absence of a linear Guinier region is often an indicator of poor sample quality, however it could also be diagnostic that the protein is very elongated so that the Guinier region is at such low angles that the data are blocked by the beam stop. This region of the scattering curve is influenced by non-specific interactions occurring between particles in solution. These interactions include non-specific aggregation, which results in a sharp increase in the scattering intensity at low angles and inter-particle repulsion, which results in a decrease in the scattering intensity at low angles. These interactions result in an over- or under-estimation of R_g and $I(0)$, respectively. Examples of scattering curves along with their Guinier plots for protein samples undergoing inter-particle interference or aggregation are shown in Figure 1-10. Guinier analysis can be carried out using programs such as PRIMUS and AutoRG which are part of the ATSAS program package [113,115].

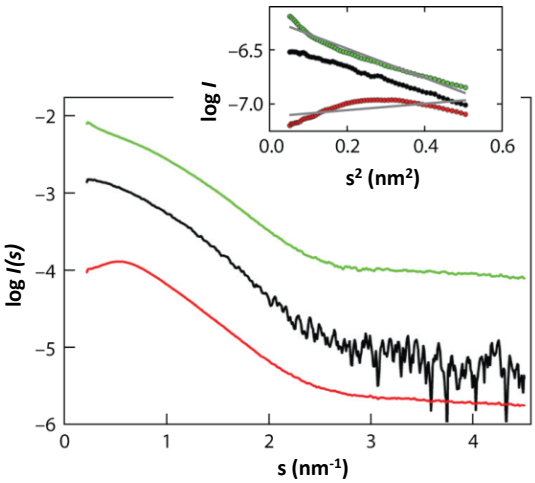


Figure 1-10 SAXS scattering curves with Guinier plot (inset). Curves are representative of an ideal protein solution (black) with its corresponding Guinier plot (inset) showing linearity; a protein solution undergoing non-specific aggregation (green) with its corresponding Guinier plot (inset) showing an increase in $\boldsymbol{I(s)}$ at low s; and a protein solution undergoing inter-particle repulsion (red) with its corresponding Guinier plot (inset) showing a decrease in $\boldsymbol{I(s)}$ at low $\boldsymbol{s}$. This figure is adapted from [112].

Although analysis of the Guinier region of the scattering curve gives a good approximation of $I(0)$ and R_g for the protein of interest, more accurate values can usually be obtained by analysing the data as a function of the pair-distance distribution (p(r)):

$$p(r) = \frac{r^2}{2\pi^r} \int_0^\infty s^2 I(s) \frac{\sin(sr)}{sr} ds$$ **Equation 7**

The p(r) function describes the distances between electrons in the protein structure. This function transforms the data from reciprocal space into real-space using an indirect Fourier transform (FT). An iterative approach to determining the D_{max} is utilised in programs such as GNOM where a range of D_{max} values are tested to determined the value that gives the best fit to the data [116]. The D_{max} and R_g values calculated from the p(r) function are generally more reliable than the values obtained from analysis of the Guinier region since all of the scattering data are used to determine these parameters. This is particularly important in cases where a limited number of data points are used in the Guinier analysis [110].

Visually the p(r) function is a useful tool for initially characterising the conformation of the protein. For example, a spherical object will give a bell-shaped p(r) curve whereas, an elongated structure gives a distribution skewed towards small s values; and a bimodal distribution is indicative of a dumbbell-shaped object (Figure 1-11). As well as being useful for the determination of the overall structure of a protein, the p(r) distribution can also be sensitive to small changes in the relative positions of residues between structures [117].

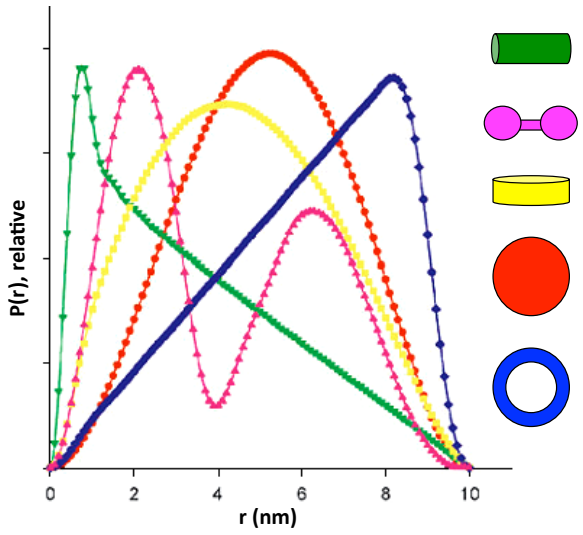


Figure 1-11 p(r) distributions of different shapes. Plot showing examples of the p(r) functions obtained from different shapes (the p(r) and the shape have been coloured correspondingly). An elongated rod (green) gives a skewed distribution, a dumbbell shape (pink) gives a bimodal distribution; a flat disk (yellow)

gives a broad distribution that is slightly skewed towards low s values; a sphere (red) gives a bell-shaped distribution; and a hollow sphere (blue) gives a distribution skewed towards larger pair-wise separations. This figure is adapted from [118].

The degree of 'compactness' or degree of folding of a protein can be determined using the Kratky plot ($s^2I(s)$ versus s) [119]. Folded globular proteins give a Kratky plot with a peak at low angles in which the scattering intensity decays as a function of $1/s^4$, whereas the Kratky plot for an unfolded protein lacks a peak at low angles and instead gives a plateau in which the scattering profile decays as a function of 1/s (Figure 1-12) [112]. A partially folded protein or a protein with flexible linker regions can give rise to a peak at low angles along with a plateau or a slightly increasing profile at higher angles (Figure 1-12).

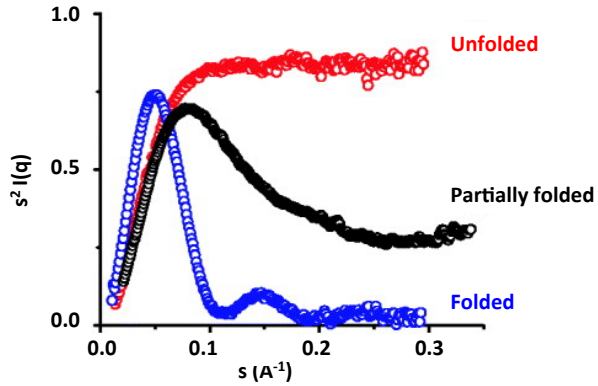


Figure 1-12 Kratky plots indicating the 'compactness' of a protein. Typical Kratky plots for an unfolded (red), partially folded (black) and a folded protein (blue) have been overlaid to show the differences in the scattering intensity at high angles. This figure was adapted from [110].

1.8.3 *Modelling of solution structures from SAXS data*

SAXS data can be used to determine a low-resolution *ab inito* shape of a protein. This is one of the advantages of using SAXS since it requires no prior knowledge of the protein structure to generate a model. However, in the case where there are some existing structural data, e.g. a crystal structure of a domain within the protein, this can be used to create a hybrid structural model.

There are several programs available for the generation of *ab initio* models from SAXS data. The program DAMMIN, which is part of the ATSAS suite of programs, is one of the most widely used SAXS modelling programs [115,120] and alongside DALAI-GA [121] was one of the first to be made widely available. DAMMIN uses the p(r) plot and D_{max} generated by GNOM as the search volume constraint. It "fills" the search volume with up to several thousand, densely packed dummy atoms. A simulated annealing algorithm is used to determine the dummy atom model that gives the best fit to the scattering data. Running DAMMIN several times for the same scattering profile generates a range of models, the variation between these models acts as an indication of the reliability of model [114]. The program DAMMIF uses a modelling approach similar to that used by DAMMIN, however the computational power and time required to run the program is decreased [122]. Furthermore, DAMMIF allows a variable search volume which avoids 'border effects' sometimes seen when the D_{max} is fixed [122]. An average DAMMIF model is generated using the DAMAVER suite of programs, which superimpose the models and filter out low occupancy beads [123]. The resulting bead model represents a low-resolution 'envelope' structure of the protein. However, the final resolution attainable from *ab initio* models varies with data quality as well as the protein size, shape and degree of flexibility [110]. For proteins where a degree of symmetry has been determined, e.g. from the oligomeric state, imposing correct symmetry in the modelling greatly enhances the resolution of final results [114].

The combination of SAXS data with other complementary structural techniques provides a powerful tool for the study of proteins in their solution state. SAXS offers insights into protein flexibility which is often unobservable with other structural methods, such as crystallography which is preferentially amenable to more rigid, ordered proteins. The flexibility of a protein can be analysed using programs such as EOM (Ensemble Optimisation Method) or SOMO-DMD [124,125]. Both programs generate a pool of possible structural conformations and then select a subset of structures that best describe the experimental scattering data. As part of this work SOMO-DMD was used to analyse the flexibility of the target protein FklB.

1.9 Analytical ultracentrifugation

Analytical ultracentrifugation (AUC) is a powerful method to characterise the size, shape and interactions of macromolecules in solution. AUC can be used to study proteins with a wide range of molecular weights from a few hundred Daltons (e.g. peptides) to several hundred-million Daltons (e.g. virus particles) [126].

The analytical ultracentrifuge (AUC) uses an optical system to monitor the movement of macromolecules (e.g proteins) in response to a gravitational force, which is generated during high-speed centrifugation. There are two main types of optical system that are widely used: absorbance and interference. Absorbance optics follow the movement of molecules at specific wavelengths (e.g. 280 nm for proteins). The interference optics system is sensitive to changes in the refractive index of the protein solution compared with a reference buffer solution. The interference optics system monitors the movement of all particles in the solution and does not rely on inherent absorbance properties of the protein. The optics systems measure the protein concentration as a function of radial position which along with the time, rotor speed and temperature are the key parameters monitored in an AUC experiment. Depletion of the protein from the solution creates a boundary region that moves throughout the course of an AUC experiment. There are two principal types of experiments that are carried out in AUC, sedimentation velocity (SV) and sedimentation equilibrium (SE), both of which are discussed below.

1.9.1 *Sedimentation velocity*

In SV experiments, high rotor speeds are used to rapidly sediment the protein towards the cell bottom. A typical AUC 'cell' centrepiece is shown in Figure 1-13A. The centrepiece has two channels into which the reference buffer and the protein solutions are loaded. During the course of a SV experiment, the sedimenting protein moves towards the bottom of the cell, leading to depletion of the protein from the meniscus. This leads to the formation of a sharp boundary between the depleted region and the region of uniform protein concentration towards the bottom of the cell. Analysis of how this boundary moves over time, shown in Figure 1-13B, allows the measurement of two main

parameters: the sedimentation coefficient (s) and the translational diffusion coefficient (D), which is explained below.

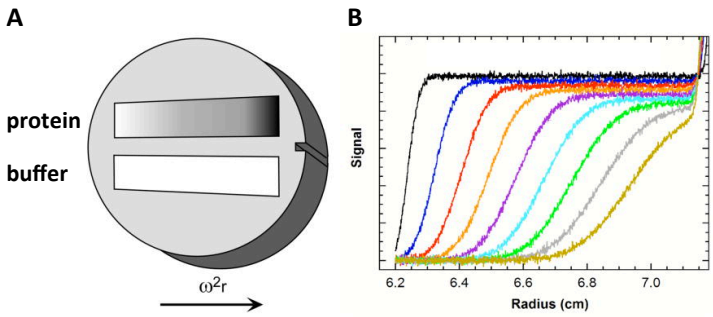


Figure 1-13 Sedimentation velocity experiment. In a typical 2-sector AUC cell **(A)** one channel contains buffer and the other contains the protein solution. The concentration of the protein at each radial position in the cell is monitored in the AUC over time during the course of an SV experiment. **(B)** The protein boundary at different time points is shown in different colours, the earliest scan taken is black and the latest scan taken is yellow. This figure is adapted from [126].

The high rotor speeds used in a SV experiment generate a large gravitational force ($M_p\omega^2 r$), where M_p is the mass of the particle, ω is the rotor speed in radians per second ($\omega = 2\pi$ x rpm/60), and r is the distance from the centre of the rotor. The mass of displaced solvent depends on the particle mass and its partial specific volume ($\bar{v}$), as well as the density of the solvent (ρ), which can be calculated using the program SEDNTERP [127]. Therefore the net, or buoyant, mass of the particle is $M_b = M(1 - \bar{v}\rho)$. Movement of the particle through the solution generates a frictional force (fu), where f is the frictional coefficient and u is the velocity of the particle. Another factor to consider during an SV experiment is the diffusion of molecules opposing the concentration gradient. The translational diffusion coefficient (D) is measured from the spreading of the boundary seen over time. The Stokes-Einstein equation shows that D and are f related by $D = RT/N_A f$, where R is the gas constant, T is the absolute temperature and N_A is Avogadro's number. So the sedimentation coefficient can then be calculated using Equation 8, where u is the radial velocity of the particle which is determined from the distance that the boundary has moved from the meniscus over time. The unit of sedimentation coefficient is seconds and is reported in Svedberg units (S, 10^{-13}s).

$$s \equiv \frac{u}{\omega^2 r} = \frac{M(1 - \bar{v}\rho)}{N_A} = \frac{MD(1 - \bar{v}\rho)}{RT} \qquad \textbf{Equation 8}$$

Since the sedimentation coefficient is influenced by several experimental variables such as the solvent used (in terms of differences in density and viscosity) as well as the temperature at which the experiment was conducted, the sedimentation coefficient is converted to a standard value, usually that which would be measured in water (w) at 20°C.

Some SV data analysis programs, such as Sedfit [128], use the Lamm equation (Equation 9) to model the evolution of the concentration distribution of a macromolecular species as a function of time (t) and radial position (r) under the influence of sedimentation and diffusion [129].

$$\frac{dc}{dt} = \frac{1}{r}\frac{d}{dr}\left[rD\frac{dc}{dr} - s\omega^2 r^2 c\right] \qquad \textbf{Equation 9}$$

There is no exact solution to the Lamm equation and Sedfit takes a numerical approach towards finding the best solution. In this thesis SV data were initially analysed with Sedfit using a sedimentation coefficient c(s) distribution analysis [128,129]. This analysis is useful as all of the species present in the solution, within a defined s range, can be observed. Assuming that the sample was pure this reveals details about higher oligomeric species that are formed. Once the number of species present has been determined by SV, further analysis of protein self-association can be conducted using SE.

1.9.2 *Sedimentation equilibrium*

SE experiments use lower rotor speeds and longer run times than SV experiments. Initially the protein in solution begins to sediment towards the cell bottom however, after a period of time, diffusion of the proteins opposes the concentration gradient that forms until these forces reach equilibrium. Once equilibrium has been reached there is no net change in the protein radial distribution. The time to achieve equilibrium is dependent on a number of experimental factors, including the mass of the particle and the volume of sample used. Therefore the experiment is monitored to ensure that equilibrium has been reached. Programs such as WinMatch (written by Jeffrey Lary, University of

Connecticut, Storrs, CT, USA) compare scans to ensure that no movement of the boundary is occurring.

At equilibrium an exponential concentration gradient is created where different molecular weight species are distributed to different radial positions within the cell, i.e. high molecular weight species will be located at the bottom of the cell whereas low molecular weight species will adopt a shallower exponential distribution. The distribution of different species is dependent on the buoyant molecular weight, the angular velocity and temperature. For a single protein, one exponential distribution, which readily allows for the determination of the apparent weight average molecular mass of this molecule, and with knowledge of the $\bar{v}$ of the protein, the molar mass is readily evaluated (Equation 10):

$$\overline{M}_{w,app}\frac{dln(c)}{dr^2}\frac{2RT}{\omega^2(1-\bar{v}\rho)} \qquad \textbf{Equation 10}$$

where c is the concentration at radius r. For a single ideal species (i.e. where particles do not non-specifically interact with each other) plotting $ln(c)$ versus r^2 generates a straight line, the gradient of which yields the mass of the protein. When this plot is not linear it may be indicative that at high concentrations there may be some non-ideality, which results in a lower $\overline{M}_{w,app}$ than expected. Non-ideality can be accounted for by introducing the molar second virial coefficient (B) to the data analysis, since $\overline{M}_{w,app} = M/(1 + BMc)$. SE data showing non-ideality were analysed using MicroCal ORIGIN (part of the Beckman XL-A/XL-I software suite), otherwise data were analysed using SEDPHAT [130,131].

Protein self-association was analysed using SEDPHAT which determines the sedimentation profile from the measured signal as a function of radial position ($a(r)$) [129].

$$a(r) = \sum_n c_{n,0}\mathcal{E}_n d\ exp\left[\frac{M_n(1-\bar{v}_n\rho)\omega^2}{2RT}(r^2 - r_0^2)\right] + \delta \qquad \textbf{Equation 11}$$

As shown in Equation 11 the summation is of all species n, $c_{n,0}$ and $\mathcal{E}_n$ are the molar concentration and extinction coefficient of species n at the reference position r_0 , d denotes the optical path length (12 or 3 mm) and δ represents the baseline offset this is represented schematically in Figure 1-14. For interference data the extinction coefficient is replaced by the specific signal increment (2.75 fringes per mg/ml) [129]. The fit to the data to different self-association models was determined from the sum of the squared residuals, the χ^2 value.

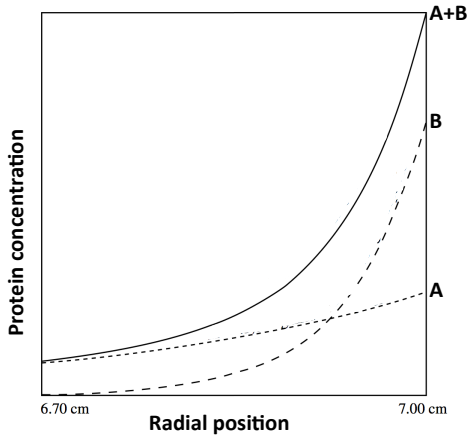


Figure 1-14 An example of sedimentation equilibrium data showing the exponential functions for the distribution of molecular masses. The data for a monomer-tetramer model are shown (A+B) with the exponential functions corresponding to the monomer (A) and the tetramer (B). This figure is adapted from [132].

The use of AUC as a precursor to further biophysical studies allows key parameters about the protein of interest to be determined. These parameters include the oligomeric state, the affinity of oligomerisation as well as the purity of the sample and are essential for other biophysical studies, including X-ray crystallography, SAXS and NMR.

2 Materials and methods

2.1 Chemicals, strains and growth media

All chemicals used in this thesis were of analytical grade and were purchased from Sigma-Aldrich, Fisher Scientific and WR-BDH unless otherwise stated.

2.1.1 *Growth media*

Luria-Bertani (LB) or Terrific broth (TB) growth media were prepared in deionised water using the recipe shown in Table 2-1 and sterilised by autoclaving. For the preparation of solid media, 15 g l^{-1} of agar was added to the media prior to autoclaving. Minimal Essential Media with HEPES modification (MEM-HEPES; Sigma M7278) and Dulbecco's Minimal Essential Media (DMEM; Sigma D5671) were purchased from Sigma.

Media	Components per litre
LB	10 g tryptone
	5 g yeast extract
	10 g NaCl
TB	12 g tryptone
	24 g yeast extract
	4 ml glycerol

Table 2-1 Media recipes

2.1.2 *Strains*

The strains used in this study are described in Table 2-2.

2.2 Maintenance and growth of bacteria

2.2.1 *Storage of bacteria*

Frozen bacterial stocks were made by adding glycerol to a final volume of 30% to an overnight LB culture which was transferred to cryo tubes. The strains were stored at -80°C in 1 ml aliquots until required.

Strain	Description and genotype	Source
Escherichia coli		
TUV 93-0	Derivative of EHEC O157:H7 strain EDL933 Shiga toxin negative (stx^-)	[133]
TUV 93-0 *ΔadhE*	Deletion of *adhE* gene generated by allelic exchange by Dr Dai Wang.	This study
TUV 93-0 *ΔsurA*	Clean deletion of *surA* gene generated by allelic exchange by Dr Dai Wang.	This study
TUV 93-0 *ΔfklB*:kan	*fklB* exchanged for a sac:kan cassette	This study
TUV 93-0 *ΔfliC*	Clean deletion of *fliC* gene generated by allelic exchange.	Prof. Dave Gally
One Shot® TOP10 competent cells	Chemically competent cells [F- *mcrA* Δ(*mrr-hsdRMS-mcrBC*) ϕ80*lac*ZΔM15 Δ*lac*X74 *recA*1*araD*139 Δ(ara-leu)7697 *galU galK* rpsL (StrR) *endA1 nupG*]	Invitrogen
StrataClone	Chemically competent - Tetr Δ(*mcrA*)183 Δ(*mcrCB-hsdSMR-mrr*)173 *endA1 supE*44 thi-1 *recA*1 *gyrA*96 *relA*1 lac Hte [F′ *proA*B lacIqZΔM15 Tn*10* (Tetr) Amy Camr]	Agilent
BL21 Star (DE3)	Protein expression strain - F^-*ompT* *hsd*S_B (r_B^-, m_B^-) *galdcmrne*131 (DE3)	Invitrogen
Citrobacter rodentium		
ATCC 51459	Wild-type strain, mouse isolate	Prof. Brett Finlay
ATCC 51459 *lux*$^+$	*lux* positive derivative generated using p16Slux [134]	This study
ICC18032	Luminescent derivative of ATCC 51459, nal^r km^r	Prof. Gad Frankel [135]

Table 2-2 Strains used in this study

2.2.2 *Growth of bacteria*

All strains were inoculated into the required liquid medium or onto agar plates with the appropriate antibiotics. When required, antibiotics were included at the following concentrations: 50 μg ml^{-1} kanamycin (Kan), 25 μg ml^{-1} chloramphenicol (Chl), 100 μg ml^{-1} ampicillin (Amp), 500 μg ml^{-1} erythromycin (Ery). For general purposes, e.g. cloning, *E. coli* cultures were propagated from a single colony and grown at 37°C at 220 rpm for aeration.

2.3 Molecular techniques

2.3.1 *Preparation of bacterial genomic DNA*

Bacterial Cells were pelleted from 1 ml of overnight LB cultures grown at 37°C, inoculated from a single colony, to an $OD_{600\ nm}$ of 3-4. Genomic DNA (gDNA) was extracted using the ChargeSwitch gDNA Mini Bacteria kit (Invitrogen, UK), as per the manufacturer's instructions. DNA was resuspended in 200 µl nuclease free water (Ambion, UK) and stored at 4°C until required.

2.3.2 *Plasmids*

The plasmids used in this study are listed in Table 2-3 and the plasmids that were generated in this study are listed in Table 2-4.

Plasmid	Vector type	Resistance	Source
StrataClone PCR UA	Cloning	Amp / Kan	Aglient
pET-28a	Expression - N-terminal His-Tag/thrombin/T7-Tag	Kan	Novagen
pET-151	Expression C-terminal His-Tag with TEV site	Amp	Invitrogen
pET-iLOV	Expression C-terminal His-tag with C3 protease cleavage site	Amp	[136]
TYB1-SurA	Expression	Amp	[137]
pRFP	Reporter - *rfp* in pACYC	Kan	Lab stock
p16Slux	Integration plasmid	Erythromycin	[134]
pSIM18	Helper plasmid	Hygromycin	[138]
LEE:GFP	Reporter - pACYC184 low copy number vector	Chl	[139]
LEE2:GFP	Reporter - pACYC184 low copy number vector	Chl	[139]
LEE3:GFP	Reporter - pACYC184 low copy number vector	Chl	[139]
tir:GFP	Reporter - pACYC184 low copy number vector	Chl	[139]
rpsM:GFP	Reporter - pACYC184 low copy number vector	Chl	[139]
fliC:GFP	Reporter - pACYC184 low copy number vector	Chl	Lab stock
pIB-307	Recombination - pMAK705-based vector for allelic exchange; temperature-sensitive replicon	Chl	[140,141]
pET-151-Tpx	Expression - *yp*Tpx-His	Amp	[100]
pET-151-FolX	Expression - FolX-His	Amp	[100]
PET-151-WrbA	Expression - WrbA-His	Amp	[100]
pET-151-*ec*TpxC61S	Expression - *ec*TpxC61S mutant	Amp	[100]

Table 2-3 Plasmids used in this study

Plasmid	Description
pET-28FklB	Over-expression of FklB
pET28-FklBc	Over-expression of C-terminal domain of FklB
p77-FkpA	Over-expression of FkpA
p77-AdhE	Over-expression of AdhE
p77-AdhE-D1	Over-expression of N-terminal domain of AdhE
p77-AdhE-D2	Over-expression of C-terminal domain of AdhE
pET28-CheY	Over-expression of CheY
pIB-FklB	Allelic exchange of *fklB* for sac-kan cassette
pAdhE	pWSK29 complement plasmid

Table 2-4 Plasmids generated in this study

2.3.3 *Oligonucleotide primers*

Oligonucleotide primers were synthesised by Invitrogen and used to amplify DNA by PCR as described in Section 2.3.4. Primers used in this study are listed in Table 2-5.

Name	Sequence 5' to 3'
pET-28FklB F	CCGTATGGATCCATGACCACCCCAACTTTTGA
pET-28FklB R	CCTTAAAGCTTATGTTAGAGGATTTCCAGCAGTT
pET28-FklB-Cterm F	CCATGGTGAATAGCACCGAATCTG
pET28-FklB-Cterm R	CCATGGGAGGATTTCCAGCAGTTC
p77-FkpA F	CGGAGCTCCGTTGAAGTTGTGGCGATTC
p77-FkpA R	CGAAGCTTGGGTAACCAGATCTTCAACAG
p77-AdhE F	CATATGATGGCTGTTACTAATGTCGC
p77-AdhE R	GGATCCTTAAGCGGATTTTTTCGC
p77-AdhE-D1 F	ACCATGGACATGGCGGTTACCAATGTTG
p77-AdhE-D1 R	ACCATGGCCGACGAAAGTAGATGCTT
p77-AdhE-D2 F	ACCATGGACATGCTGTGGCATAAGCTGCC
p77-AdhE-D2 R	ACCATGGCGCGGATTTCTTC
pET28-CheY F	GGATGACTTTTCCACCATGC
pET28-CheY R	ATAACCGCCTGCCTGCAA

Table 2-5 Primers used in this study

2.3.4 *Polymerase chain reaction*

A typical PCR used Taq polymerase (Promega, UK) master mix with 100 pM forward and reverse primers and 0.1 µM genomic DNA in a final volume of 25 µl. If the initial conditions did not amplify, a 12°C gradient of the annealing temperature was used to optimise the reaction. The standard PCR protocol used is shown in Table 2-6. For colony PCR, a single bacterial colony was diluted into 0.5 ml PBS and heated at 70°C for 10 min. Cell debris was pelleted by centrifugation and 1 µl of supernatant was added per PCR reaction.

Temperature (°C)	Step	Duration (s)
95	initial denaturing	300
95	denaturing	45
55	annealing	45
72	extension	60 (/kb)
72	final extension	600
4	hold	-

Table 2-6 Standard PCR protocol

2.3.5 *Agarose gel electrophoresis*

For a 1% w/v agarose gel, 1 g of agarose was added to 100 ml 1x Tris-acetate-EDTA (TAE). This was heated until the agarose was dissolved and allowed to cool to approximately 50°C. SYBR Safe gel stain (Life Technologies, UK) or GelRED (Cambridge Biosciences, UK) was added 1:10000 as per the manufacturer's instructions before the solution was poured into a gel tray and allowed to set. The appropriate volume of 10x loading buffer (Invitrogen, UK) was added to the DNA samples prior to loading and the gel was then run at 100 V for 45 min. The 1 kb+ size marker was purchased from Invitrogen (UK). Agarose gels were visualised using an AlphaImager transilluminator (Alpha Innotech, UK).

2.3.6 *Restriction enzyme digest*

All restriction enzyme digests were performed according to the manufacturer's instructions (NEB, UK). A typical digest reaction consisted of 1 µg of DNA, the corresponding amount of 10x restriction enzyme digestion buffer and 10 U restriction enzyme. The reaction was made up to the desired volume (typically 10 -20 µl) using nuclease free water (Ambion, UK). If more than one enzyme was used, the compatibility of the enzymes determined whether the reaction was either carried out simultaneously or sequentially. The reaction was mixed and incubated at 37°C for 2 h. The digested products were resolved by agarose gel electrophoresis to visualise and to purify the fragments of interest.

2.3.7 *DNA gel purification*

Linearised DNA products generated following PCR or restriction digest were resolved by electrophoresis. DNA bands were viewed using a transilluminator and bands of interest were excised from the agarose gel and purified using the Qiagen Gel Extraction Kit as per the manufacturer's instructions. The purified DNA was eluted into either 30 or 50 µl dH_20 depending on the concentration used in the reaction. The DNA concentration was quantified using a NanoDrop at 220 nm. For long term storage the samples were stored at -20°C.

2.3.8 *Ligation reaction*

Ligation reactions were set up according to the manufacturers' instructions (NEB, UK). A typical ligation reaction consisted of 100 ng digested plasmid, with insert added at a 1:3 molar ratio (plasmid:insert), 10x DNA ligase buffer and 5 U T4 DNA ligase (NEB, UK). The reaction volume used was either 10 or 20 µl, depending on the concentration of starting material available. The reaction was incubated at 4°C overnight and then used in transformation reactions.

2.3.9 *Heat shock transformation*

Five microliters of the ligation reaction was added to 50 µl chemically competent TOP-10 *E. coli* cells (Invitrogen, UK) and incubated on ice for 30 min to allow the cells to thaw and the DNA attach to the cell surface. The cells were heat shocked by submersion in a 42°C water bath for 45 s. The cells were immediately returned to ice for 1 min. 500 µl of prewarmed SOC medium (Invitrogen, UK) or LB was added to the transformed bacteria followed by a 2 h incubation at 37°C with agitation. The mixture was plated onto LB agar plates containing appropriate antibiotics and incubated at 37°C overnight.

2.3.10 *Production of electrocompetent E. coli*

A single bacterial colony was used to inoculate 10 ml of LB broth which was then incubated at 37°C overnight with agitation. Two millilitres of the overnight culture were added to 100 ml pre-warmed LB broth. This was incubated at 37°C with agitation until an

$OD_{600\ nm}$ = 0.6 was reached. Bacteria were chilled on ice, centrifuged at 3,000 *x g* for 10 min at 4°C and pellets resuspended in 50 ml ice-cold 10% (v/v) glycerol. This was repeated three times. After the final centrifugation step the pellet was resuspended in 50 ml ice-cold 10% (v/v) glycerol and pelleted by centrifugation. The pellet was then resuspended in 5 ml ice-cold 10% (v/v) glycerol. 40 µl of cells were used per reaction.

2.3.11 *Electroporation transformation*

Five microliters of ligation reaction was added to 40 µl of electrocompetent *E. coli* cells which were transformed by electroporation using an electroporator (Invitrogen, UK) at 2.5 kV and a capacitance of 25 µF. One millilitre of SOC medium was added to the transformation mixture and incubated at 37°C for 2 h with agitation. The mixture was plated onto LB agar plates containing appropriate antibiotics and incubated at 37°C overnight.

2.3.12 *Extraction of plasmid DNA*

A single bacterial colony was used to inoculate 5 ml LB broth containing the appropriate antibiotic and grown overnight at 37°C with agitation. The bacteria were centrifuged at 3000 *x g* for 10 min at 4°C. The plasmid was purified using the QIAprep Spin Miniprep kit as per the manufacturer's instructions.

2.3.13 *Creation of deletion mutants by allelic exchange*

The target gene *fklB* was deleted using a process called allelic exchange, which has been optimised for EHEC using the protocol by Emmerson *et al.* (2006) [140]. This relies on the creation of a temperature sensitive exchange plasmid encoding the flanking regions of the gene of interest. The 3' and 5' flanking regions were synthesised by Genescript, UK. The construct was designed with a *BamHI* site between the flanking regions. The sac-kan cassette encodes both resistance to Kan and sensitivity to sucrose, which are used as selection markers. This cassette was inserted into the plasmid between the homologous flanking regions using the *BamH1* site, creating pIB-FklB. pIB-FklB was then transformed into TUV and homologous recombination between the plasmid and the genome was

induced by growth at 42 °C (growth at this temperature also inhibits plasmid replication). Integrants were selected following growth on Kan.

2.3.14 *Creation of lux positive C. rodentium strain*

To follow *C. rodentium* infection *in vivo* the strain was marked with luminescence genes, which when expressed can be visualised using the *in vivo* imaging system (IVIS; PerkinElmer, UK). The p16S*lux* plasmid was obtained from Riedel *et al*. [134]. It encodes the *lux* operon from *Photorhabdus luminescens* and a 16S RNA gene insertion site [134]. *C. rodentium* was transformed with p16S*lux* by electroporation and plated onto LB plates containing 500 μg/ml erythromycin (LBE) and incubated at 30°C for 48 h. Colonies were checked for light emission using the IVIS and *lux* positive clones were inoculated into LBE and grown overnight at 30°C. The overnight culture was diluted 1:1000 and grown for 18 h at 42°C. Serial dilutions of the culture were plated onto LBE agar plates and incubated at 42°C overnight. Integration of the plasmid at the correct site was confirmed by PCR using the primers shown in Table 2-7. A product of 1150-bp confirmed integration.

Name	Sequence 5' to 3'
Lux-check F	ACACTGGAACTGAGACACGGTCCAGACTCC
Lux-check R	TTGTAAAACGACGGCCAGTGAGCGCGCG

Table 2-7 *lux* primers

2.4 Phenotypic characterisation of strains

2.4.1 *Secretion assay*

To induce type three secretion *E. coli* O157:H7 were cultured at 37°C at 220 rpm in MEM-HEPES to $OD_{600\ nm}$ = 0.8. 50 ml of cells were pelleted by centrifugation at 4000 x *g* for 10 min. The cell pellet was retained and lysed using Bugbuster (Millipore) as per the manufacturer's instructions and stored at -20°C for further analysis of cellular proteins by Western blot. The supernatant from 50 ml was retained and the secreted proteins were precipitated by the addition of 10% (v/v) trichloroacetic acid with an overnight incubation at 4°C. Precipitated proteins were pelleted by centrifugation at 4000 *x* g for 30 min at 4°C

and the supernatant was removed. The protein pellet was dried and resuspended in 1.5 M Tris (pH 8.8). Secreted proteins were then analysed by SDS-PAGE and/or western blotting. For secreted proteins 5 -10 µl of sample were loaded onto the gel for Coomassie staining and 2-5 µl for western blotting.

2.4.2 *SDS-PAGE*

Samples were mixed 3:1 in NuPAGE sample buffer (4x, Invitrogen), boiled at 95°C for 10 min and 20 µl loaded onto a NuPAGE 4-12% Bis-Tris pre-cast gel (Invitrogen, UK). Gels were run using either MES or MOPS buffer (Invitrogen, UK). SeeBlue+2 (Invitrogen, UK) molecular weight marker was used throughout. SDS-PAGE gels were stained using either SimplyBlue SafeStain (Invitrogen, UK) as per manufacturer's instructions or Coomassie stain (500 ml dH_20, 400 ml methanol, 100 ml acetic acid, 0.5 g Coomassie blue R250) and destained with destain solution (500 ml dH_2O, 400 ml methanol, 100 ml acetic acid) until clear.

2.4.3 *Western blotting*

Proteins were transferred from an SDS-PAGE gel using the Sure-Lock Cell Blotting Module (Invitrogen, UK) onto nitrocellulose (Amersham, UK) or PVDF (Millipore, UK) membrane at 30 V for 2 h. The membrane was was blocked for a minimum of 1 h in 3 % skimmed milk (Marvel, UK) in PBS with 0.01% Tween (PBST; (137 mM NaCl, 2.7 mM KCl, 10 mM $NaHPO_4$, 1.8 mM KH_2PO_4)) or Tris-buffered saline with 0.01% Tween (TBST; 137 mM NaCl, 50 mM Tris, pH 7.2). Primary antibody was applied for 1 h at room temperature with shaking, and unbound antibody removed with 3 x 15 min washes with PBST. The membrane was then incubated in the appropriate secondary antibody conjugated to horseradish peroxidase (HRP) (Invitrogen, UK) for 1 h at room temperature with shaking and washed as before. Table 2-8 lists the primary antibodies used in this study along with their concentrations. The secondary antibody concentration used was between 1:3000 and 1:10000 depending on the efficiency of the primary antibody. The blot was developed using enzyme linked chemiluminescence (ECL) using SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific, UK).

Primary polyclonal antibody	Dilution	Secondary antibody	Origin
EspA	1:10000	Rabbit	Prof. Dave Gally
FliC:H7	1:3000	Rabbit	Mast Assure
GroEL	1:10000	Rabbit	Enzo
Tir	1:3000	Mouse	Prof. Dave Gally
Anti-acetyl lysine	1:1000	Rabbit	Abcam
Sigma factors	1:3000	Mouse	NeoClone
Streptavidin	1:10000	Rabbit	Invitrogen

Table 2-8 Antibodies used in this study

2.4.4 *Far-western*

A Far-western is a technique used to analyse the binding of pa purified protein or a labelled ligand to a protein that is immobilised in a nitrocellulose membrane following transfer from SDS-PAGE. Here, Far-western analysis was used to determine the binding of ME0052 to the putative protein targets. Far-western analysis was performed using BL21-DE3 cell lysate from cells that were over-expressing the target proteins or using purified protein. The cell lysate/purified proteins were separated using SDS-PAGE and transferred to a nitrocellulose membrane. The membrane was incubated in 5% (v/v) Marvel at RT for approximately 4 h. The milk was removed and the membrane was incubated with 2 µM biotinylated ME0052 (Figure 2-1) in PBST overnight at 4°C. The membrane was washed in PBST and incubated in a 1 in 4000 dilution of streptavidin (anti-biotin probe conjugated to horseradish peroxidase, Invitrogen). The membrane was washed 3 times with PBST and the streptavidin antibody was detected using ECL.

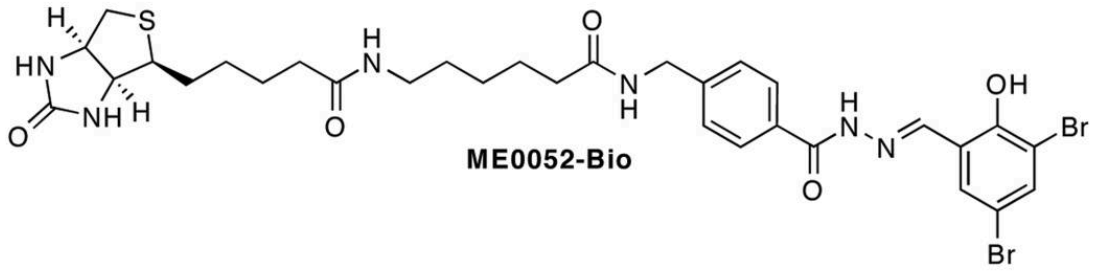


Figure 2-1 Structure of the biotinylated version of ME0052. ME0052-Bio was used in far-western analysis of the binding of the compound to target proteins in *E. coli* O157:H7.

2.4.5 *GFP reporter fusion assay*

Plasmid-borne promoter:GFP reporter fusions were used to monitor gene expression in different growth conditions. The promoter fusions used in this study are listed in Table 2-3 [139]. Plasmids were transformed into *E. coli* strains by electroporation. Transformants were cultured overnight in LB supplemented with chloramphenicol (25 µg ml^{-1}). Overnight cultures were diluted into MEM-HEPES or into fresh LB with chloramphenicol to $OD_{600\ nm}$ = 0.8. Fluorescence of GFP was measured in triplicate in 200 µl aliquots dispensed into 96-well black microtiter plates and analysed using a FLUOstar plate reader (BMG, Germany). The $OD_{600\ nm}$ of cultures was measured to monitor growth. Fluorescence was plotted against optical density using Prism (GraphPad) software, and the line of best fit was obtained. Untransformed cultures were used to correct for background fluorescence of the strain and medium. Corrected fluorescence values of the wild-type and mutant strains at $OD_{600\ nm}$ = 0.7 were compared.

2.4.6 *Motility assay*

Swarming agar was made using either LB or TB with (0.25% (w/v) agar) following the protocol described by Wolf and Berg [142]. 50 ml of agar was used per plate. The centre of the agar was inoculated with either a stab transferred from a colony or 5 µl of an overnight culture normalised to $OD_{600\ nm}$ = 0.7. Plates were incubated at 34°C overnight and the diameter of the swimming zone measured.

2.5 RNA transcript sequencing

2.5.1 *RNA extraction*

To look at the levels of gene expression on a global scale RNA sequencing of gene transcripts was carried out. This work was done with James Connolly, a PhD student in the Roe group who prepared the RNA and helped with the data analysis. RNA was extracted from *E. coli* grown in LEE T3SS inducing conditions to $OD_{600\ nm}$ = 0.8. 20 ml of bacterial culture was mixed with RNA protect (Qiagen, UK). At this step the RNeasy mini kit (Qiagen, UK) was used to extract RNA as per the manufacturer's instructions. To

ensure all DNA was removed an initial Turbo DNAse (Life Technologies, UK) step was carried out as per the manufacturer's instructions. RNA was quantified using the NanoDrop and a PCR for a housekeeping gene (*gapA*) was carried out to further ensure that the sample was free from DNA. Once a negative PCR result was obtained, the mRNA was enriched to deplete ribosomal RNA using a Microbexpress kit (Ambion). Enriched mRNA samples were sent to the University of Glasgow Polyomics facility for cDNA synthesis and RNA sequencing analysis which was carried out on an Illumina Genome Analyser IIx. Sequencing was performed using single ended reads.

2.5.2 *RNA sequence data processing and analysis*

Raw transcript data were analysed using the CLC Genomics Workbench 4.0 (Qiagen, UK). The *E. coli* EDL933 genome sequence was used as a reference for annotation (NCBI database). CLC was used to generate reads per kilobase mapped values (RPKM) for each gene as well as information regarding the depth of sequencing. Annotated sequence files were exported from CLC in ace format and converted to a custom coverage format using a python script (getCov454.py) provided by Mitchell Sullivan (University of Queensland). Coverage files were subsequently trimmed to obtain coverage files for desired regions of the genome. Corresponding regions were downloaded from the NCBI database in FASTA format to be used as the annotation reference for figure generation in EasyFig 2.1 [143]. Experiments consisted of 4 replicate WT samples and a single sample of the Δ*adhE* mutant or disulfiram-treated WT sample. The variance in WT gene expression allowed significant differentially expressed genes to be identified.

2.6 Immunofluorescence microscopy

2.6.1 *Single cell analysis*

Expression of the LEE T3SS or flagellum was analysed by immunofluorescence using an α-EspA or an α-H7 antibody, respectively. Strains transformed expressing RFP from the reporter plasmid pRFP were grown in MEM to $OD_{600\ nm}$ = 0.8 and a 100 µl aliquot was diluted 1:1 in 4% paraformaldehyde (PFA; Sigma, UK). 20 µl was dried onto a multispot microscope slide (Thermo Scientific, UK). EspA filaments were stained using 20 µl of α-

EspA antibody diluted to 1:100 in PBS with 0.01% tween and 0.1% BSA added to each spot and incubated with shaking for 1 h at room temperature in a humidity chamber. For H7 staining the same antibody dilution and staining procedure was used. The primary antibody was removed with 3 x 100 µl wash steps in PBS with 0.01% tween and 0.1% BSA. Twenty microliters of α-rabbit Alexa Fluor-488 conjugated secondary antibody (1:1250; Invitrogen, UK) was added to each spot and incubated with shaking for 1 h at room temperature in the dark in a humidity chamber. The secondary antibody was removed by washing, as above, and the spots were allowed to air dry before a coverslip was mounted with fluorescence mounting media (Dako, USA). The slides were examined on a Zeiss M1 Axioskopp microscope, using the appropriate filter sets and a x100 objective. A Z-stack of 16 images was captured at a spacing of 0.15 µm using Volocity software (PerkinElmer, UK). These images were deconvoluted to reduce the effects of light scattering thus making the image sharper.

2.6.2 *Live imaging of bacterial motility*

To image the motility of live bacteria a different approach to staining was required. Turner *et al.* (2010) describe a method using AlexaFluor esters (Invitrogen, UK), which react with primary amines of proteins forming stable dye-protein conjugates [144]. As whole cells are used, the dye binds to the surface of the bacteria as well as surface structural components such as flagella without interfering with their function. The staining protocol used here closely followed that described by Tuner *et al*. (2010). Overnight cultures of TUV, Δ*adhE*, and Δ*fliC* were inoculated into 5 ml of TB. The culture was grown to $OD_{600\ nm}$ = 0.6. The cultures were pelleted by centrifugation at 3000 *x g* for 10 min and the media removed. The pellets were resuspended in 1 ml of wash buffer (0.01 M, KH_2PO_4, pH 7, 70 mM NaCl, 0.1 mM EDTA) and transferred to a microfuge tube. Bacteria were washed 3 times at room temperature by centrifugation (2000 *x g*, 10 min) and after the final wash the bacteria were resuspended in 500 µl of wash buffer. An aliquot of Alexa Fluor-488 was resuspended in 99 µl wash buffer and 33 µl was added to 200 µl of cells and incubated in the dark at room temperature with rotation (100 rpm). Labelled cells were pelleted by centrifugation and washed 3 times, as above but in buffer containing 0.000001% Brij 35 (Sigma, St Louis MO). After the final wash the bacteria were resuspended in 500 µl of wash buffer containing 0.000001% Brij and 0.1 M glucose.

Slides were prepared using a 25-fold dilution of the labelled bacteria. Approximately 100 µl of cell suspension was sealed between the microscope slide and cover slide with a square well created using vacuum grease (Beckman; Palo Alto, CA). Slides were imaged immediately. Images were acquired on a Zeiss M1 Axioskopp, using the appropriate filter sets and a x100 objective.

2.6.3 *Adhesion of E. coli to host cells*

Embryonic bovine cells (EBL; German Collection of Microorganisms and Cell Cultures, no. ACC192) were prepared and cultured as described elsewhere [145]. Cells were cultured in DMEM (Sigma, UK) supplemented with 1 mM L-glutamine and 10% foetal calf serum (FCS; Invitrogen, UK). Cells were seeded at a density of 1 x 10^5 into a 12-well tissue culture plate with glass cover slips at the bottom of each well. Cells were incubated at 37°C, 5% CO_2 for at least 24 hours before bacterial co-culture.

Bacterial strains transformed with pRFP were cultured in MEM-HEPES to an $OD_{600\ nm}$ of 0.8 at 37°C. Prior to addition of bacteria the EBLs were washed with warm MEM-HEPES and left in 500 µl of media. Bacteria were added to the cells at a multiplicity of infection of 100:1 and centrifuged onto the cells at 1,000 *x g* for 5 min, to initiate cell contact. EBLs and bacteria were incubated for 1 h at 37°C 5% CO_2 at which point the unbound bacteria were removed by washing 3 times with warm MEM-HEPES. Cells were incubated for a further 3 h. Cells were washed in PBS three times and 250 µl of 2% PFA was added to each well and incubated at room temperature for 20 min to fix the cell. The PFA was removed and the cells washed 3 times with PBS. 250 µl of 0.5% Triton X100 was added to permeablise the host cells. After a 5 min incubation, cells were washed 3 times with PBS. 250 µl of FITC-phalloidin (Invitrogen, UK) at a 1:500 dilution was added to each well and incubated in the dark at room temperature for 1 h. Cells were washed 3 times with PBS. An additional step to stain for flagella was also carried out in some cases. Here 250 µl of 1:500 dilution of the α-H7 antibody was added to each well, and incubated with shaking for 1 h. Excess, unbound primary antibody was removed with 3 x 1 ml PBS wash steps and an α-rabbit Alexa Fluor-647 (Invitrogen, UK) conjugated secondary antibody was added at 1:1250 dilution and incubated with shaking for 1 h. Excess, unbound secondary antibody was removed with 3 x 1 ml PBS wash steps. The cover slips were removed from the well, inverted and mounted onto a slide with fluorescence mounting media (Dako, USA).

2.7 TLR5 reporter assay

Flagellin is the main agonist of Toll Like Receptor 5 (TLR5), which is responsible for inducing a pro-inflammatory response during an infection. The TLR5 response induced by different *E. coli* strains (TUV, Δ*adhE*, and Δ*fliC)* was tested using the reporter cell line HEK-Blue™-hTLR5 (Invivogen, UK). These experiments were carried out at the Roslin Institute, University of Edinburgh with help from Dr Amin Tahoun, a post-doctoral researcher with Prof. Dave Gally. The cell line was maintained following the manufacturer's instructions in the appropriate antibiotics (20 µg μl^{-1} blasticidin, 100 µg ml^{-1} Zeocin, 50 µg ml^{-1} Normocin; all from Invivogen, UK) and 10% heat inactivated FCS.

The TLR5 response generated by both the type three secreted proteins and from whole cells was tested. Here, the TUV Δ*fliC* mutant was used as a negative control as this cell line does not produce flagellin and therefore should not stimulate the TLR5 response. TUV, Δ*adhE*, and Δ*fliC* were grown to $OD_{600\ nm}$ = 0.8 and the secreted proteins were precipitated as described above. The precipitated proteins were then diluted into ten 10-fold serial dilutions in PBS. 180 µl of HEK-Blue™-hTLR5 cells was seeded into a flat bottomed 96-well plate at a density of approximately 1.4 x 10^5 cells ml^{-1} and 20 µl of each serial dilution was added per well; each dilution was tested in triplicate. The 96-well plate was then incubated overnight at 37°C. On day 2 the Quanti-Blue™ substrate (Invivogen) was prepared as per the manufacturer's instructions and 200 µl per well was transferred to a flat-bottomed 96-well plate. To these 20 µl of media from the HEK-Blue™-hTLR5 cell:bacteria culture was added. The plate was incubated at 37°C for 30 min and the absorbance at 450 nm was measured using a SPECTROstar Omega plate reader (BMG LabTech, Germany). Each absorbance value was measured in triplicate, these values were averaged and the media control (background absorbance) was subtracted. Each experiment was carried out in quadruplicate.

The TLR5 response induced by the presence of whole bacteria was also tested. Five hundred microliters of HEK-Blue™-hTLR5 cells were seeded at a density of 1.4 x 10^5 cells ml^{-1} into 24-well plates and incubated overnight at 37°C. Two hours before use, the cells were washed and left in media without antibiotics. Overnight cultures of TUV, Δ*adhE*, and Δ*fliC* were inoculated into MEM-HEPES and grown to $OD_{600\ nm}$ = 0.7. The bacteria were

then diluted five-fold into fresh MEM-HEPES media and 100 µl of the bacterial dilution was added per well. Each strain was tested in triplicate in 3 separate experiments. The bacteria were then attached to the cells by centrifugation at 300 *x g* for 3 min and incubated for 1 h at 37°C. Unbound bacteria were removed by 3 x 1 ml wash steps with fresh media and incubated for a further 3 h. As above, 20 µl of media was added to the Quanti-Blue™ and the absorbance measured.

To quantify the amount of flagellin required to stimulate the TLR5 response, 20 µl of varying concentrations of purified flagellin (produced by Dr Amin Tahoun, Roslin Institute, University of Edinburgh) was added to HEK-Blue™-hTLR5 cells in a 96-well plate as described above. The absorbance values were plotted and used as a standard curve from which the response from a defined amount of flagellin could be determined. This was used to quantify the amount of flagellin in the secreted protein preparation.

2.8 Animal experiments

2.8.1 *Infection of Balb/C mice with C. rodentium*

All animal work was carried out by Dr Gill Douce (University of Glasgow) with the help of Zoe Marjenberg, a PhD student in the Roe group. Three groups of 8 week old female Balb/C mice were infected with 200 µl of 1 x 10^9 colony forming units (c.f.u.) of *C. rodentium* ICC18032 by oral gavage 30 min after treatment with 100 µg 0.1 sodium bicarbonate by oral gavage. Group one was infected with *C. rodentium* grown in DMEM, and group 2 with *C. rodentium* grown in DMEM with 10 µM disulfiram.

Mice were anesthetised with 2-3% isofluorane and bioluminescent images acquired using an IVIS Spectrum. To measure faecal shedding, mouse faeces were collected and homogenised in PBS at a concentration of 100 mg/ml, and serially diluted onto kanamycin agar plates. The number of CFUs was used as measure of the level of *C. rodentium* colonisation.

2.9 Protein overexpression and purification

Plasmids were transformed into the *E. coli* BL21(DE3) strain and cultured in LB medium. Five millilitres overnight cultures were used to inoculate 500 ml LB for large-scale expression. For isotope labelling of proteins, bacteria were cultured in M9 media supplemented with a labelled nitrogen source ($^{15}NH_4Cl$). Cultures were grown at 37°C to $OD_{600\ nm}$ = 0.6, at which point protein expression was induced with 1 mM IPTG. At this point the temperature was reduced to 20°C and the cultures were grown overnight. Cells were harvested by centrifugation and resuspended in buffer A (20 mM Tris, 500 mM Nacl, 20 mM imidazole, pH 7.5) with protease inhibitors (Enzo Life Sciences, UK) and DNAase (Sigma, UK). Cells were lysed following passage through a French Pressure cell at 950 p.s.i. The cell lysate was cleared by centrifugation. All of the proteins produced in this study were purified by immobilised metal affinity chromatography (IMAC). Cleared lysates were loaded onto a 5 ml Ni^{2+} HisTrap Column (GE Healthcare, UK) pre-equilibrated with buffer A. The column was washed with 25 ml buffer A and the protein was eluted using buffer B (20 mM Tris, 500 mM NaCl, 300 mM imidazole, pH 7.5) in 3 ml fractions.

The purified fractions were analysed by SDS-PAGE. To remove imidazole from the protein buffer, samples were dialysed into the appropriate buffers. Further purification steps including size exclusion chromatography (SEC) and anionic exchange chromatography were carried out on some proteins. Size exclusion chromatography was carried out on an AKTA Purifier (GE Healthcare, UK) using a Superdex S75 HiLoad 26/60 column or a Superdex S200 column, depending on the size of the protein being purified. Anionic exchange was carried out using HiTrap Q Sepharose FF (GE Healthcare, UK) following dialysis of the purified proteins into a buffer containing a low salt concentration (typically 20 mM Tris pH 7.5 with 20 mM NaCl) and eluted with 20 mM Tris pH 7.5 with 500 mM NaCl. Fractions were further analysed by SDS-PAGE and the fractions containing the protein of interest were dialysed into an appropriate buffer.

2.10 Analytical ultracentrifugation

2.10.1 *Determination of the oligomeric state of Tpx, FklB and FolX*

The monodispersity of purified protein samples was determined using analytical ultracentrifugation (AUC). AUC was carried out in a Beckman Coulter (Palo Alto, CA, USA) Optima XL-I analytical ultracentrifuge. Sedimentation velocity (SV) experiments were performed at 4°C at a rotor speed of 49 krpm. 360 µl of purified protein dialysed against the buffer indicated in Table 2-9 at a range of concentrations, was loaded into double sector centrepieces. In order to impose oxidising or reducing conditions, 10 mM H_2O_2 or 5 mM DTT, respectively, were added to Tpx samples. Data were acquired with interference optics and/or absorbance optics and scans were taken every 7 min. The partial specific volume of each protein and the buffer density and viscosity values were calculated using the program SEDNTERP [146] (Table 2-9). SV data were analysed using SEDFIT [128]. The data were first analysed by c(s) and then fitted with a non-interacting discrete species model in order to obtain a value for the apparent sedimentation coefficient over the range of protein concentrations. These values were plotted against the loading concentration and the best fit through the data extrapolated to infinite dilution to obtain the sedimentation coefficient (s^0). The sedimentation coefficient under standard conditions, i.e. at 20°C in water ($s_{20,w}$) was calculated using the following equation:

$$s_{20,w} = s_{obs}\left(\frac{\eta_{T,w}}{\eta_{20,w}}\right)\left(\frac{\eta_b}{\eta_w}\right)\left(\frac{1-\bar{v}\,\rho_{20,w}}{1-\bar{v}\,\rho_{T,b}}\right)$$

where s_{obs} is the sedimentation coefficient measured in the experimental buffer at the experimental temperature, T; $\eta_{T,w}$ and $\eta_{20,w}$ are the viscosities of water at the experiment temperature and at 20°C, respectively; η_b and η_w are the viscosities of the buffer and water at a common temperature; $\rho_{20,w}$ is the density of water at 20°C and $\rho_{T,b}$ is the density of the buffer at the experimental temperature.

Protein	Temperature (°C)	Partial specific volume (g ml^{-1})	Buffer	Buffer density (g ml^{-1})	Buffer viscosity (P)	SE rotor speed (krpm)
Tpx	4	0.741	20 mM Tris pH 7.5, 50 mM NaCl	1.00264	0.0158	18, 24
	20	0.714		1.00100	0.0101	
FklB	4	0.731	20 mM Tris pH 7.5, 50 mM NaCl	1.00264	0.0158	30
	20	0.734		1.00087	0.0126	
FolX	4	0.737	20 mM Tris pH 7.5, 150 mM NaCl	1.00677	0.0156	23
	20	0.743		1.00499	0.0102	

Table 2-9 Partial specific volumes of proteins and densities and intrinsic viscosities of buffers used in analytical ultracentrifugation. Values were calculated using SEDNTERP [146].

To determine the oligomeric state of each protein, sedimentation equilibrium (SE) experiments were performed at 4°C. The rotor speeds used for each protein are indicated in Table 2-9. Samples (80 µl) were loaded under the same conditions as for the SV experiments. Scans were taken every 3 h until analysis of the scans with WinMATCH (Jeffrey Lary, University of Connecticut, Storrs, CT, USA) indicated that equilibrium had been attained. SE data for Tpx were analysed with SEDPHAT [131]. For proteins where the effects of non-ideality could be observed (i.e. FolX and FklB) the program ORIGIN (MicroCal Software Inc, USA) was used to analyse SE data. FolX equilibrium data were fitted with a tetramer-octamer model with non-ideality and Fklb data were fitted with a monomer-dimer model with non-ideality.

The dissociation constant of protein oligomerisation was calculated using the method described by Solovyova *et al*. [147]. Briefly, the 'fringe coefficient' was calculated from the SV interference and absorbance data where the concentration of a single species was determined by fitting a non-interacting discrete species model. The concentration obtained was used to convert the association constant, which is in inverse fringe units, to a dissociation constant (K_d) in molar units.

2.10.2 *Determination of K_d for the interaction of Tpx and ME0052*

Before AUC analysis, proteins were dialysed against 20 mM Tris, pH 7.4, 50 mM NaCl. SV data were collected at 49,000 rpm at 4°C for Tpx under oxidising (with 10 mM H_2O_2) and reducing (with 5 mM DTT) conditions and for the C61S mutant. 360 µl of sample (100 µM protein solution with ME0052 at concentrations ranging from 0 to 250 µM) and reference solvent (20 mM Tris, pH 7.4, 50 mM NaCl, 1% (v/v) DMSO) were loaded into 3-mm path length charcoal-filled epon double-sector centrepieces. Rayleigh interference and absorbance (at 280 and 395 nm) data were collected every 15 min for 25 h (1 replicate of absorbance data in continuous mode with a radial step size of 0.005 cm). A wavelength of 395 nm was used to monitor the sedimentation of ME0052 only, as the protein does not contribute to the signal at this wavelength.

A two-stage analysis of SV data was used to determine the amount of ME0052 bound to oxidised and reduced Tpx and the C61S mutant. The sedimentation coefficients ($s^0_{20,w}$) of oxidised and reduced Tpx and the C61S mutant were determined in this study to be 3.04, 2.62, and 2.80 S, respectively (Section 3.2.2); thus, a peak with this sedimentation coefficient in the c(s) distribution derived from SV data acquired at 395 nm could be attributed to Tpx bound to ME0052. In the control sample containing no ME0052, no peaks were resolved from the data; at higher concentrations two peaks were evident: one at very low s (corresponding to unbound ME0052) and the second close to 2.7 S (ME0052 bound to protein). The concentration of free and bound compound was determined by fitting the data with a (two species) non-interacting discrete species model in SEDFIT [128]. An extinction coefficient of 11,100 M^{-1} cm^2, determined from the linear dependence of absorbance at 395 nm with ME0052 concentration, was used to calculate the concentration of compound returned by the non-interacting discrete species model. The resultant data were then fitted with a single-site-binding non-linear regression model using the formula:

$$Y = \frac{B_{max} \times X}{(K_d + X)}$$

where *Y* is the concentration of ME0052 bound to protein (determined from the fit with the non-interacting discrete species model); *X* is the total concentration of ME0052 (i.e. bound and unbound); B_{max} is the concentration of compound required to saturate the

protein (i.e. at maximal binding); and K_d is the equilibrium dissociation constant or concentration of compound required to reach half-maximal binding. The stoichiometry of interaction between ME0052 and *yp*Tpx was determined by dividing B_{max} by the total protein concentration.

2.11 Small angle X-ray scattering

2.11.1 *Data collection and processing*

Purified proteins were taken to either the EMBLX33 beamline at DESY, Hamburg or to the BM29 beamline, ESRF Grenoble for small angle X-ray scattering data collection. The X33 beamline monochromates X-rays to a wavelength of 0.15 nm with a Pilatus 1M, 2D detector with 0.172 mm pixel size and a useful area of 67 x 420 mm^2 and a 3 ms frame rate. At the standard detector distance (2.7 m) an s range of 0.06 nm^{-1} to 6 nm^{-1} can be achieved. The BM29 beamline has an achievable s range of 0.01 nm^{-1} to 5 nm^{-1} with a Pilatus 1M detector distance of 2.87 m.

50 µl sample volumes were loaded into the automatic sample changer. For each sample and corresponding buffer 15 exposure frames were collected. Radiation damage was assessed as a change in the scattering intensity between the first and last frames. Scattering curves unaffected by aggregation were then averaged, buffer subtracted and scaled for concentration using PRIMUS [113]. Data for a range of protein concentrations were collected and analysed for interparticle interference or protein aggregation occurring at higher concentrations. p(r) analysis was carried out using GNOM [113,116]. *Ab initio* models were generated from the experimental data using DAMMIN or DAMMIF [120,122]. Twenty DAMMIF models were superimposed and averaged using DAMAVER [123]. The nominal resolution of the data was calculated from the highest angle at which usable scattering data were recorded using $2\pi/s_{max}$ where s_{max} is the reciprocal spacing of the highest-resolution data point used in the restoration process [147]. The averaged model was superimposed using SUPCOMB [148] onto the crystal structure in the case of FolX and Tpx (PDB ID: 4AEY or 2YJH, respectively) or with a homology model in the case of FklB.

SAXS data were used to model the flexibility of FklB in solution. The program SOMO-DMD was used to generate a range of models with a flexible region between residues 68 and 98 in the FklB homology model [125]. The DMD parameters were kept at the default values. The Andersen thermostat temperature set to 0.5 kcal/mol/K, this value affects the level of structural integrity of the model, i.e. higher temperatures destroy secondary structure. The simulation time used was 0.25 ns and 50 models were generated. DMD simulations were submitted to the Alamo server at UTHSCSA for processing. The models were fit to the data using a non-negative least squares analysis, which selects an ensemble of models that best fit the scattering data.

2.12 NMR

2.12.1 *Acetate quantification*

NMR is a highly quantitative method for the measurement of metabolites. Although global approaches to analysing metabolic profiles can be taken, here a targeted approached was used for the quantification of acetate levels produced by WT and Δ*adhE* mutant bacteria. All NMR studies were carried out with help and supervision of Dr Brian Smith (University of Glasgow). Bacteria were cultured in 50 ml of MEM-HEPES to an $OD_{600\ nm}$ = 0.8 and the cells were removed from the media by centrifugation at 3000 *x* g for 15 min. Growth media was filtered using a 0.2 µM filter to remove any remaining cells. One-dimensional high-resolution ^{1}H-NMR spectra were acquired on a Brucker AVANCE 600 MHz spectrometer at 298 K. Data were processed using AZARA and analysed using CCPNMR analysis [149].The acetate peak at ~1.9 ppm was identified and integrated between 1.921 and 1.888 ppm for each sample. This value was compared with a standard curve of known acetate concentrations to quantify the levels in the media. Measurements from triplicate independent experiments were averaged and differences analysed by a Student's t-test.

2.12.2 *^{1}H-^{15}N HSQC*

To look at interactions of proteins (Tpx and FklB) with different compounds ^{1}H-^{15}N heteronuclear single quantum correlation (HSQC) spectroscopy was used. Proteins were

expressed in M9 minimal media substituted with a labelled nitrogen source ($^{15}NH_4Cl$) as described in Section 2.9. A chemical shift perturbation assay was used to monitor which residues were being affected by the addition of compound. Protein samples of 100 µM dialysed against 20 mM K_2HPO_4, pH 7.4, 20 mM NaCl were used for data collection, 5% D_2O was added and used as the internal reference. ^{15}N FAST-HSQC [150] spectra were recorded at 298 K and 14.1 T using a Bruker AVANCE 600 spectrometer equipped with a 5 mm TCI cryoprobe. Compound was added from a concentrated stock solution in deuterated DMSO (dDMSO), and reference spectra recorded with an equal concentration (0.33% v/v) of dDMSO added. Data were processed using AZARA and analysed using CCPNMR analysis [149].

2.13 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) was used to determine the affinity of binding between the target protein FklB and the compounds rapamycin and ME0052. Proteins were dialysed extensively before use at 4°C into 50 mM HEPES, pH 8.0, 100 mM NaCl, 0.5 mM EDTA containing the appropriate concentration of DMSO (depending on the concentration of compound being added) to ensure that buffer mismatching was kept to a minimum in the experiment. All solutions were degassed under vacuum for 5 min with gentle stirring immediately before use. Owing to the poor solubility of the compounds, they were put into the ITC chamber at a low concentration (1 nm for rapamycin and 20 µM for ME0052 or ME0055) and titrated with the protein, which was at a high concentration in the injection syringe. Data were processed using MicroCal Origin software.

2.14 Homology modelling

2.14.1 *FklB*

There is currently no crystal structure for FklB, however FklB does share a high level of sequence identity with other PPIases therefore a homology model based on its sequence was generated using the modelling server I-TASSER [151]. This generated five models which were then analysed using the program PROCHECK [152] to identify the model that

had the best stereochemical quality. This model was then superimposed onto the crystal structure of a PPIase that has a 42% sequence identity to FklB (PDB ID:1FD9) to create an FklB dimer model.

2.15 X-ray crystallography

2.15.1 *Protein crystallisation*

Purified proteins were dialysed overnight against 20 mM Tris pH 7.5, 50 mM NaCl and kept at a concentration of approximately 8 mg ml^{-1} (for Tpx) or 4 mg ml^{-1} (for WrbA) and placed into several commercial screens using sitting drop vapour diffusion. All crystal trays were set up using a Cartesian Honeybee robot (Hamilton, UK). Crystal drops were set up 1:1 protein:buffer ratio, with a final volume of 1 µl. Where co-crystallisation was attempted, 50-100 µM of ME0052 was added to the protein before trays were set up. Crystals that formed in conditions without compound were used for compound soaks with varying concentrations of ME0052. Where cryo-protection was needed, crystals were passed through a solution of mother liquor with 10% glycerol. Crystals were frozen in liquid nitrogen or in a stream of cooled gaseous nitrogen (100K).

2.15.2 *Data collection and processing*

All diffraction data were collected at Diamond Light Source (Oxfordshire, UK). The data collection strategy was determined using the Strategy option in MOSFLM which indicated the phi range and the recommended oscillation range required to collect a complete data set [108]. Data were processed with MOSFLM [108] and scaled in SCALA [109], both parts of the CCP4 suite of programs [153]. Structures were determined by molecular replacement using PHASER [154] where the search model (PDB ID: 3HVV for Tpx, or 2R96 for WrbA) was cut back to the last identical atom using CHAINSAW [155]. The resulting models were refined using REFMAC5 [156] and visually inspected using COOT [111], where waters were added. Models were validated in COOT and by the MolProbity server [157]. Protein structures were submitted to the Protein Data Bank (PDB [158]) and are available under PDB codes 2YJH (*yp*TpxC61S) and 4A2F (*ec*TpxC61S). Images were created using PYMOL [159].

2.16 Enzymology

2.16.1 *Prolyl-cis isomerase assay*

The enzyme activity of the PPIases (FkpA, FklB and SurA) from *E. coli* O157:H7 was investigated. This work was carried out at the Defence Science Technology Laboratory, Porton Down with the help of Dr Isabel Norville. Varying concentrations of purified protein were incubated for 6 min at 6°C in 1.2 ml 35 mM HEPES buffer, pH 7.8 with 15 μl DMSO and the substrate succinyl-ala-phe-pro-phe-ρ-nitronalide (10 mg ml^{-1}, Bachem, UK). 250 μl α-chymotrypsin (5 mg ml^{-1}) was added to the cuvette and mixed. Hydrolysis of the substrate was measured at 390 and 510 nm in a Shimadzu (Japan) 1800 UV/Vis spectrophotometer at 1 s intervals for 14 min to ensure completion of the reaction.

Kinetic analysis of the enzyme activity was conducted by Dr Nic Harmer, University of Exeter following the method described in [160]. The pseudo-first-order rate constant was calculated using the equation $\ln(A_\infty - A_t) = k_{obs}t + \ln(A_\infty - A_0)$; data from 10 to 50 s (which were always after the lag phase and before substrate became limiting) were taken, and k_{obs} was calculated by linear regression. The enzymatic rate was determined by comparing the observed rate to the uncatalysed rate using the equation $k_{enz} = k_{obs} - k_{uncat}$. The specificity constant k_{cat}/K_m for the enzyme was calculated using the equation $\left(\frac{k_{cat}}{K_m}\right) = (k_{enz}/[PPIase])$ [160].

2.16.2 *Inhibition by ME0052 and rapamycin*

As an initial test for inhibition 100 nM FklB, 1 μM FkpA and SurA were incubated with 10 μM rapamycin or 100 μM ME0052 dissolved in DMSO and the above reaction was monitored. Inhibition of FklB PPIase activity was investigated further by incubating varying concentrations of rapamycin (1 – 250 nM) or ME0052 (10 – 150 μM) with 100 nM FklB. Data for inhibitor assays were fit to the equation:

$v = v_0 \frac{[E]-[I]-K+\sqrt{([E]-[I]-K)^2+4[E][K]}}{2[E]}$ [161] using least-squares nonlinear fitting. v_0 and $K_I{}^{app}$ were fit using initial estimates based on the raw data, and [E] was kept constant.

2.16.3 *WrbA activity*

WrbA is an NAD(P)H quinone oxidoreductase, an enzyme that oxidises NAD(P)H and transfers the electrons to a quinone, e.g. benzoquinone. The activity of WrbA can be monitored by following the oxidation of NADH at 340 nm using a method that was devised by Patridge *et al.* [162]. The reactions were monitored with a Shimadzu (Japan) UV-2501PC spectrophotometer with a SFA-20 Rapid Kinetics (Rapid Mix) accessory (Hi-Tech Scientific). All the measurements were obtained at 25˚C. To determine the K_m for NADH the reactions contained 400 µM 1,4-benzoquinone in 50 mM MOPS, pH 7.2 and a concentration range of NADH (5 - 250 µM) was tested. The linear region from the initial 40 s of each progress curve was used to determine the initial velocity. The K_m for NADH oxidation by WrbA was calculated using the Michaelis-Menten equation. Inhibition of WrbA by the salicylidene acylhydrazides was tested using varying concentrations of ME0052 (0.3 - 33 µM).

3 Is thiol peroxidase a target protein of the salicylidene acylhydrazides?

3.1 Introduction

Thiol peroxidase (Tpx) is one of the 19 proteins identified from *E. coli* O157:H7 as a putative target of the SA compounds. The focus of the study by Wang *et al.* (2011) was the identification of the binding proteins of the SA compounds, and included initial characterisation of the interactions between Tpx and ME0052 [100]. Further structural characterisation of Tpx conducted as part of this thesis gave rise to a more detailed understanding of the interactions between ME0052 and Tpx. This work validated Tpx as a target of the SA compounds.

3.1.1 *The role and enzymatic function of Tpx*

The reaction of molecular oxygen with free electrons in the cell generates toxic intermediates also known as reactive oxygen species (ROS), including superoxide (O_2^-), hydrogen peroxide (H_2O_2) and the hydroxyl radical (HO•). These intermediates subsequently interact with proteins and DNA in the cell, potentially causing damage or leading to mutation, respectively. ROS do not only originate from within the cell. H_2O_2 is produced as a defence against pathogens by the immune system. Since H_2O_2 can diffuse freely across the cell membrane this leaves bacteria exposed to danger from external sources of H_2O_2 [163]. For example, inside the phagosome bacteria are exposed to 5 - 10 µM H_2O_2 increasing the intracellular concentration 10-fold, a concentration high enough to kill the bacteria [164].

Protection against constant exposure to ROS in the cell requires several lines of defence. *E. coli* expresses at least 9 different enzymes that detoxify ROS. These include catalases, alkylhydroperoxide reductase (Ahp), bacterioferritin comigratory protein (BCP), glutathione peroxidase (Gpx), organic hydroperoxide resistance protein (Ohr), cytochrome C peroxidase (CCP), and peroxiredoxins [164]. This variety of enzymes with differing catalytic rates and substrate specificities emphasises the importance of regulating the levels of ROS within the cell.

Peroxiredoxins are a ubiquitous family of peroxidases that reduce hydrogen peroxide and alkyl hydroperoxides [165]. The importance of this class of enzymes is indicated by the

fact that peroxiredoxins are amongst the most abundant proteins within *E. coli* [166].Tpx is classed as an 'atypical 2-Cys peroxiredoxin', indicating that that the reaction occurs within one molecule whereas 'typical peroxiredoxins' require at least two molecules for catalysis [167]. Tpx contains 3 cysteine residues, 2 of which are involved in the catalytic mechanism (C61 and C95; residue numbering corresponds to the *E. coli* nomenclature). Catalysis involves three chemical steps: step 1 - a peroxidation reaction of C61 leading to the formation of a sulphenic acid intermediate; step 2 - the sulphenic acid residue subsequently reacts with the free thiol of C95, forming a disulphide bond between C61 and C95 and the production of H_2O; step 3 - Tpx requires interaction with thioredoxin to resolve the disulphide bond, regenerating the free thiols (Figure 3-1) [165].

Initially presumed to be localised in the periplasm [168], recent work using cross-linking and fractionation studies [169] has shown that Tpx is one of several peroxiredoxins found in the cytosol of *E. coli*. Tpx has been shown to have a high affinity for alkyl hydroperoxides indicating that it plays a key role in the detoxification of lipid hydroperoxides produced by oxidative damage [170].

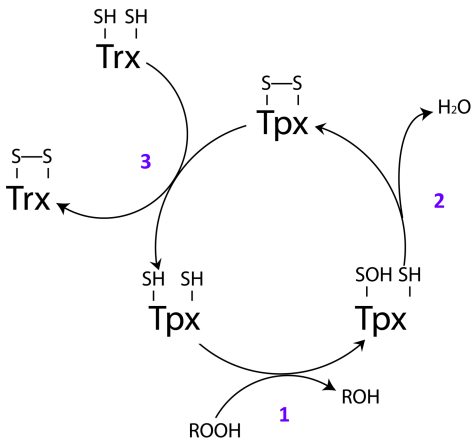


Figure 3-1 Catalytic mechanism of Tpx. (1) Tpx reduces organic hydroperoxides (ROOH) to form a sulphenic acid intermediate on C61 (Tpx SOH). (2) Water is produced when a disulphide bond is formed between the peroxidative cysteine (C61) and the resolving cysteine (C95) (Tpx S-S). (3) Thioredoxin (Trx) oxidises Tpx, regenerating it for the next reaction.

3.1.2 *The role of Tpx in virulence*

Pathogens require an effective antioxidant system within the cell in order to survive attacks from immune cells. When this system is compromised virulence is often attenuated. For example, Tpx has been shown to be important for the survival of *S. typhimurium* in macrophages, where the oxidative burst can be particularly acute [171]. However, Tpx has been reported to be only of moderate importance in protecting cells from oxidative stress [172]. For example in *P. aeruginosa* it confers some resistance to micromolar concentrations of H_2O_2 [172]. Interestingly, Cha *et al.* showed that deletion of Tpx in *E. coli* K12 severely affects growth in anaerobic conditions, suggesting that Tpx is an important enzyme under these conditions [170].

3.1.3 *The structure of Tpx*

Peroxiredoxins exhibit a conserved thioredoxin-like fold with seven β-sheets and six α-helices [167]. Tpx also has an extended N-terminal which forms a β-hairpin loop [173]. Peroxiredoxins exhibit a wide variety of oligomeric states, ranging from monomeric (YPrx) [174], to large decameric or dodecameric assemblies like TryP [175], AhpC [176] and other typical 2-Cys peroxiredoxins, including PrxIII from bovine mitochondria, which forms two concatenated dodecamers [177]. These assemblies are often dependent on redox state, dissociating into homodimers upon oxidation [167]. Previous studies of Tpx from *E. coli* K12 (*ec*Tpx) showed that the protein is a homodimer, regardless of the redox state, and despite the lack of any inter-subunit disulphide bond [165].

3.1.4 *Aims of this chapter*

This chapter aims to investigate the role of Tpx in the regulation of the LEE T3SS and to carry out a detailed structural characterisation of Tpx and its interaction with the SA compounds. The work presented in this chapter forms the basis of four publications (* denotes joint first authorship):

- Wang, D., Zetterström, C. E., Gabrielsen, M., **Beckham, K. S. H**., Tree, J. J., MacDonald, S. E., Byron, O., Mitchell, T. J., Gally, D. L., Herzyk, P., et al. (2011) Identification of bacterial target proteins for the salicylidene acylhydrazide class of

virulence-blocking compounds. Journal of Biological Chemistry, 286, 29922–29931. [100]

- Gabrielsen, M., Zetterström, C. E., Wang, D., **Beckham, K. S. H.**, Elofsson, M., Isaacs, N. W. and Roe, A. J. (2010) Expression, purification, crystallization and initial X-ray diffraction analysis of thiol peroxidase from *Yersinia pseudotuberculosis*. Acta Crystallographica Section F, 66, 1606–1609. [178]
- Gabrielsen, M., ***Beckham, K. S. H.,** Feher, V. A, Zetterström, C. E., Wang, D., Müller, S., Elofsson, M., Amaro, R. E., Byron, O. and Roe, A. J. (2012) Structural characterisation of Tpx from *Yersinia pseudotuberculosis* reveals insights into the binding of salicylidene acylhydrazide compounds. PLoS One 7, e32217. [179]
- **Beckham, K. S. H**., Byron, O., Roe, A. J. and Gabrielsen, M. (2012) The structure of an orthorhombic crystal form of a "forced reduced" thiol peroxidase reveals lattice formation aided by the presence of the affinity tag. Acta Crystallographica Section F, 68, 522–526. [180]

The majority of the structural work and investigation into the binding of the SA compounds was carried out on Tpx from *Yersinia pseudotuberculosis* (*yp*Tpx). Previous studies by Dr Dai Wang showed that despite the high level of sequence similarity with Tpx from *E. coli* O157:H7 (*ec*Tpx), this protein was highly over expressed and could be purified easily. Furthermore this protein was the first to yield structural data. Given the similarity of *yp*Tpx and *ec*Tpx (79% identity, Figure 3-2) it was presumed that data gained from studies with *yp*Tpx could be applied to *ec*Tpx.

```
ec      MSQTVHFQGNPVTVANSIPQAGSKAQTFTLVAKDLSDVTLGQFAGKRKVLNIFPSIDTGV 60
yp      MTQTVHFQGNPVSVAGKLPQIGDKAKDFTLVAKDLSDVALSSFAGKRKVLNIFPSIDTGV 60
        * ********** **   ** * **  *********** *  ******************

ec      CAASVRKFNQLATEIDNTVVLCISADLPFAQSRFCGAEGLNNVITLSTFRNAEFLQAYGV 120
yp      CAASVRKFNQLAGELENTVVLCISSDLPFAQSRFCGAEGLSNVITLSTLRGADFKQAYGV 120
        ************ *  ******** *************** ******* * * * *****

ec      AIADGPLKGLAARAVVVIDENDNVIFSQLVDEITTEPDYEAALAVLKA 168
yp      AITEGPLAGLTARAVVVLDGQDNVIYSELVNEITTEPNYDAALAALK- 167
        **  *** ** ****** *  **** * ** ****** * **** **
```

Figure 3-2 Sequence alignment of Tpx from *E. coli* O157:H7 and *Y. pseudotuberculosis.* ClustalW alignment showing protein identity, residues that are 100 % conserved are indicated by an asterisk [181].

3.2 Structural characterisation of *yp*Tpx

Tpx was identified as a target protein of the SA compounds by an affinity pull-down assay in *E. coli* O157:H7 [100]. In this study Tpx (from several Gram-negative species) was further validated as a target using far-western analysis of biotinylated ME0052 binding carried out on cell lysate from cells over-expressing Tpx from different species [100]. However, in order further validate and characterise compound binding an in-depth structural analysis of the interactions between the protein and the compound was undertaken. This required high-resolution structural data in order to determine the compound binding site. Knowing the details of these interactions will enable the future development of compounds specific to Tpx.

3.2.1 *Over-expression and purification of Tpx*

Tpx from *Y. pseudotuberculosis* and *E. coli* O157:H7 was cloned into a pET-151 expression vector by Dr Dai Wang using standard procedures. Two litres of BL21(DE3) transformed with either the *yp*Tpx-151, *yp*TpxC61S, or *ec*Tpx-151 plasmid were cultured to $OD_{600\ nm}$ = 0.6 and induced overnight at 20°C with IPTG. Tpx was purified using IMAC to about 95% purity as estimated by SDS-PAGE analysis (Figure 3-3) wherein a band corresponding to a dimer of Tpx can be observed. The identity of this band was confirmed to be Tpx by tandem mass spectrometry. For crystallisation, proteins (~8 mg ml^{-1}) were dialysed against 20 mM Tris, pH 7.5, 50 mM NaCl with or without 2 mM DTT. For AUC and SAXS the same buffer was used but with an increased NaCl concentration of 150 mM.

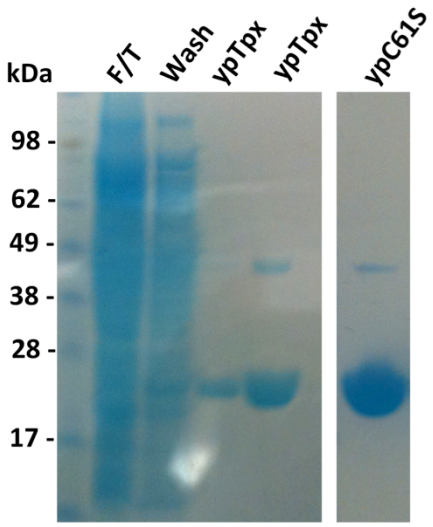


Figure 3-3 Purification of *yp*Tpx and *yp*TpxC61S. SDS-PAGE gel showing fractions following IMAC. The flow through (F/T) fraction, a fraction from the 20 mM imidazole wash step (Wash) and the final elution steps at 300 mM imidazole are indicated. For the *yp*TpxC61S protein only the elution fraction has been shown (*yp*TpxC61S). The positions of the molecular weight markers are indicated (kDa).

3.2.2 *Oligomeric state of Tpx*

Previous studies of *ec*Tpx have shown that it is dimeric [165]. However, since the oligomeric state of peroxiredoxins between species has been shown to vary greatly, it was of interest to investigate the oligomeric state of *yp*Tpx using AUC. To determine the effect of redox state on the oligomeric state of *yp*Tpx, sedimentation velocity experiments were conducted on *yp*Tpx in the presence oxidising agents (H_2O_2), reducing agents (DTT) and the 'forced reduced' catalytically inactive mutant. SV experiments revealed that all three forms of *yp*Tpx were almost completely monodisperse in solution, evidenced by a single dominant peak in the c(s) distribution (Figure 3-4). Infinite dilution sedimentation coefficients ($s^0_{20,w}$) were determined from the concentration dependence of $s_{20,w}$ (obtained from fitting the SV data with a non-interacting discrete species model in SEDFIT [128]) for the oxidised and reduced forms of *yp*Tpx ($s^0_{20,w}$= 3.04 and 2.62 S, respectively). For *yp*TpxC61S $s^0_{20,w}$ is 2.80 S, suggesting that this mutation does not induce structural instability (i.e. the value is comparable with those determined for reduced and oxidised *yp*Tpx).

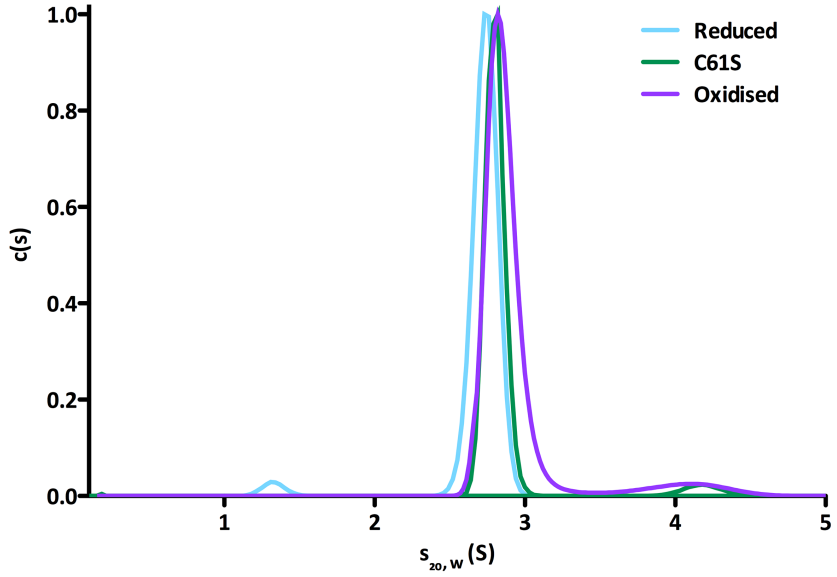


Figure 3-4 Sedimentation velocity analysis of *yp*Tpx. c(s) distributions derived via SEDFIT [128] from SV data for *yp*Tpx in oxidised (purple) and reduced (blue) states, as well as *yp*TpxC61S (green) indicate that the protein is monodisperse (dominated by a single, symmetrical peak) with essentially the same sedimentation coefficient regardless of redox state.

Sedimentation equilibrium (SE) data were fitted with the species analysis model in SEDPHAT [131]. From the concentration dependence of the resultant apparent mass of the single species, the infinite dilution mass (M^0) was determined to be 41.5 ± 3.4, 38.4 ± 2.9, 39.4 ± 0.4 kDa for oxidised, reduced and *yp*TpxC61S, respectively. The mass of the *yp*Tpx dimer, including the tag, calculated from its amino acid sequence is 42,382 Da, which is consistent with the experimentally determined masses. This indicates that *yp*Tpx is present solely as dimer in solution at the concentrations studied. It was not possible to fit the SE data with a monomer-dimer (or any other plausible) self-association model, which is further consistent with the complete dimerisation of the protein in the concentration range studied.

3.2.3 *Crystallographic studies of ypTpx*

The structure of Tpx from several different species has been solved [165]., however the structure of the protein from *Y. pseudotuberculosis* had not been solved. To determine the high-resolution structure of *yp*Tpx, recombinant *yp*Tpx in a reduced (*yp*TpxRED) and oxidised (*yp*TpxOX) state, and the mutant (*yp*TpxC61S) were crystallised. The structures of *yp*TpxOX and *yp*TpxRED were determined and refined by Dr Mads Gabrielsen, a post-

doctoral researcher in the Roe group at the time, but they are discussed here for completeness.

3.2.4 *Crystallisation of ypTpx*

The catalytically inactive mutant *yp*TpxC61S protein crystallised in 12% polyethylene glycol (PEG) 3350, 50 mM HEPES pH 7.0, 1% tryptone. The needle-like crystals appeared within 24 h with dimensions of 2 x 0.07 x 0.06 µm (Figure 3-5A). A small fragment of one of the needles was broken off and used for data collection. The protein crystallised with the space group $P6_4$. The asymmetric unit comprised one subunit with a Matthews coefficient of 2.56 $Å^3$ Da^{-1}, indicating a solvent content of 52%. The structure was determined to a resolution of 2.55 Å (Figure 3-5B). The data collection and refinement statistics for *yp*TpxC61S are presented in Table 3-1.

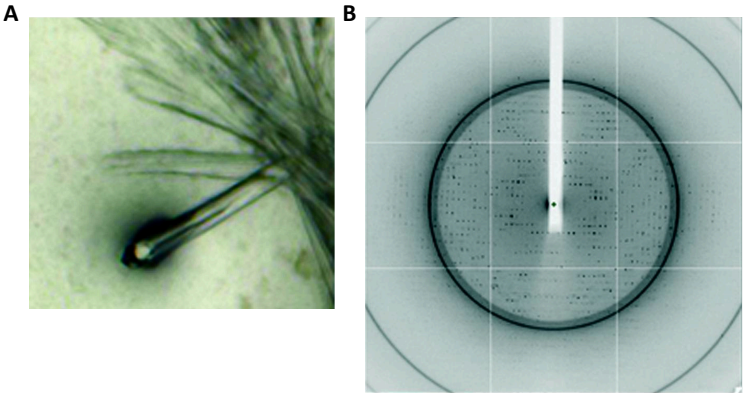


Figure 3-5 *yp*TpxC61S crystals and diffraction. (A) Needle like crystals with dimensions 2 x 0.07 x 0.06 µm in 12% PEG 3350, 50 mM HEPES pH 7.0, 1% tryptone. A small fragment of the crystal was broken off for data collection **(B)** The crystals diffracted to 2.55 Å resolution (at the detector edge) and belonged to space group $P6_4$. This figure is adapted from [180].

PDB code	2YJH
Space group	$P6_4$
Unit cell parameters (Å)	a = 65.03, c = 85.60
Resolution (Å)	56.32 - 2.55 (2.69–2.55)
Observed reflections	59770
Unique reflections	6124
Multiplicity	9.8 (9.6)
Completeness (%)	90.5 (72.6)
R_{meas} (%)	10.6 (60.6)
R_{pim} (%)	3.5 (18.9)
($I/\sigma(I)$)	11.6 (3.1)
Wilson B (Å^2)	55.8
Matthews (Å^3 Da^{-1})	2.56
Solvent content (%)	52
Water atoms	4
R_{work} (%)	31.3
R_{free} (%)	33.6
Rmsd for bond lengths (Å) / angles (°)	0.005 / 0.772
Ramachandran outliers (%)	0.3
MolProbity clash score	1.27 [100th percentile]

Table 3-1 Data collection and refinement statistics for *yp*TpxC61S. Values in parentheses are for the highest resolution shell.

3.2.5 *Structure of ypTpxC61S*

The structure of *yp*TpxC61S was determined using Phaser [154]; the existing TpxC61S structure from *E. coli* (PDB ID: 3HVV; [182]) was used as a search model for molecular replacement. The molecular-replacement solution was refined to final R_{work} and R_{free} values of 31.3% and 33.6%, respectively. There is one outlier in the Ramachandran plot and the structure is in the 100th percentile of MolProbity clash scores [157]

Overall, *yp*TpxC61S has a structure similar to that of *ec*Tpx (Figure 3-6). *yp*Tpx has a regular thioredoxin-like fold: a seven-stranded β-sheet, with β2 and β6 running anti-parallel to the rest, although with an inserted N-terminal β-hairpin (βN1-βN2) which is specific to thiol peroxidases. The central sheet is flanked by four α-helices following β3, β4, β5 and β7, and one short 3_{10} helix following β2. The numbering of the β-strands is based upon that of *ec*Tpx [173]. Nβ1-Nβ2 (Figure 3-6) forms an L-shaped hydrophobic

cleft, and it has been speculated that this cleft allows Tpx to accommodate long fatty acid hydroperoxides [183].

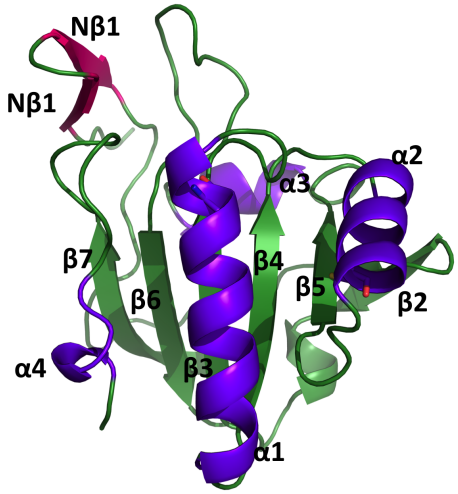


Figure 3-6 Overall structure of *yp*Tpx. Cartoon representation of *yp*TpxC61S. Strands and loops are green, helices are purple, and the catalytic residues have been shown as sticks. The N-terminal hairpin is highlighted in pink. The secondary structure elements are labelled.

The reduced form of *yp*Tpx crystallised in the presence of 2 mM DTT in 20% PEG 3350, 200 mM sodium citrate. *yp*TpxRED crystallised in the space group $P2_1$ at 2.0 Å resolution. *yp*Tpx crystallised in the oxidised state in 200 mM tripotassium citrate, 20% PEG 3350. ypTpxOX crystallised in the space group $P2_12_12_1$ and its structure was determined to 1.74 Å resolution. The change in oxidation state of the protein induces very few changes to the overall structure of *yp*Tpx (Figure 3-7A-C). The differences are localised to the region around the disulphide bond. The *yp*TpxOX structure shows an intact intramolecular disulphide bond between Cys61 and Cys95 therefore showing the oxidised form of the protein (Figure 3-7A). Whereas the *yp*TpxRED structure does not contain a disulphide bond since the presence of DTT in the crystallisation conditions created a reducing environment (Figure 3-7B). The *yp*TpxC61S structure (Figure 3-7C) represents the "forced" reduced form of the protein, as the resolving cysteine has been mutated to a serine, rendering it unable to form a disulphide bond [165].

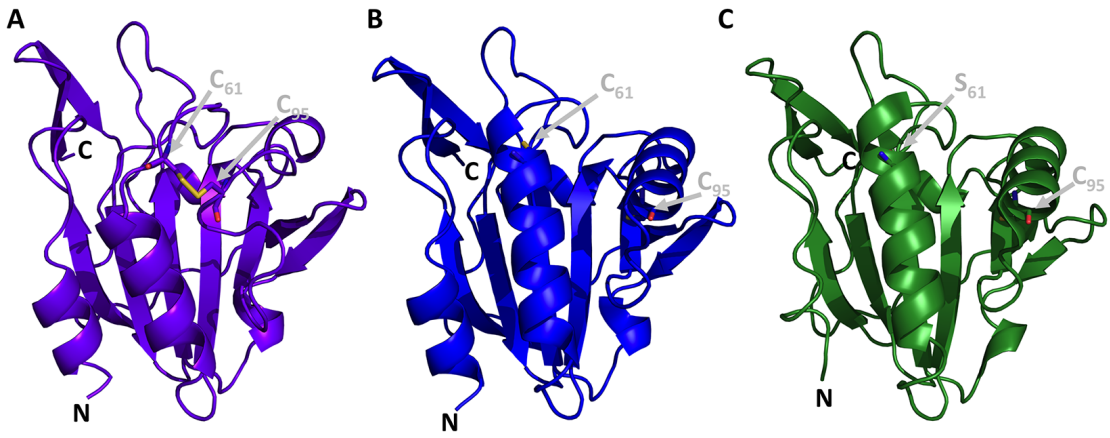


Figure 3-7 Crystal structures of *yp*Tpx. *yp*Tpx was crystallised in the oxidised **(A)**, reduced **(B)** and catalytically inactive C61S mutant **(C)** forms. The N- and C-termini have been indicated with N or C, respectively. The catalytic residues C61 and C95, or in the case of the C61S mutant - S61, are highlighted by grey arrows. The structure of the oxidised and reduced from of Tpx was solved by Dr Gabrielsen and the structure of the C61S mutant was solved in this study.

Since the C61S mutant is essentially a 'forced reduced' form of *yp*Tpx, it superimposes well with the *yp*TpxRED structure with a root-mean-squared deviation (r.m.s.d.) of 0.52 Å over 163 C_α. This fits well with the structural analysis of Hall *et al*. who used the TpxC61S mutant from *E. coli* to describe the structure of reduced Tpx [182]. Figure 3-8A shows the active site of the reduced and C61S structure and highlights the structural similarity. The changes between oxidised and reduced forms of *yp*Tpx are localised around helices α1 and α2, where there is a partial unfolding of helices α1 and α2, and a shift of 8.4 Å for C61 and 5.1 Å for C95, respectively, as presented in Figure 3-8B. The two structures superimpose well, in particular the core parts, with an r.m.s.d. of 0.7 Å over 135 C_α. Inclusion of the unfolding helices increases the r.m.s.d. to 1.02 Å.

From the AUC data (Figure 3-4) it was clear that *yp*Tpx forms a dimer even at low protein concentrations, suggesting that interactions between monomers are strong. Analysis of the *yp*Tpx structures reveals that the dimer interface comprises about 20 residues from each subunit, corresponding to 12% of total surface residues, according to the Protein Interfaces, Surfaces and Assemblies (PISA) server [184]. The interfaces are formed mostly by hydrophobic interactions, with a few hydrogen bonds, namely between R110NH1 and three main-chain carboxyl groups on the opposing subunit (G125, P126, A128). There are no salt bridges or covalent bonds between the two dimers in the reduced structure. The dimer interfaces are identical in the structures of *yp*TpxRED and *yp*TpxC61S. The dimer

interface is of similar size in the oxidised state but in addition to the hydrogen bond pattern described above, there are also salt bridges formed between D57 and R93 on opposing subunits (Figure 3-9). This is owing to the conformational change that occurs on transition from the reduced to the oxidised state.

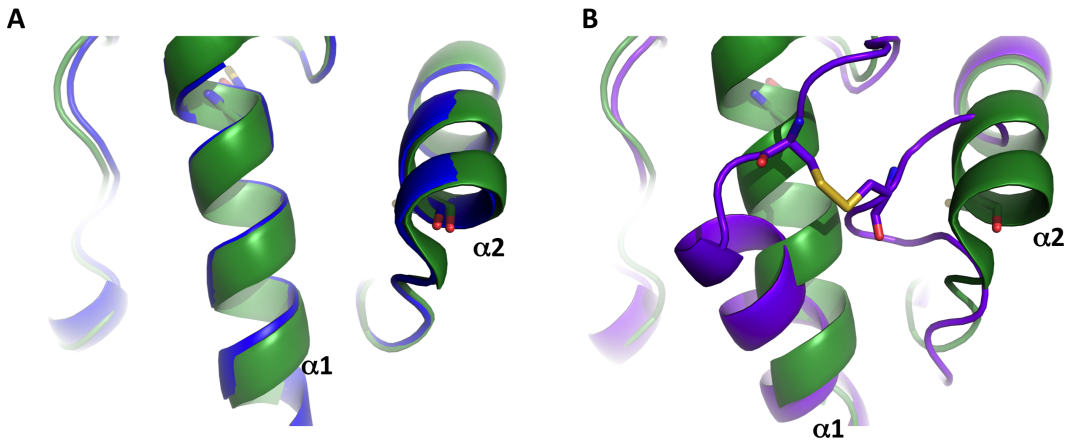


Figure 3-8 Changes in the active site of *yp*Tpx caused by the redox state of the protein. (A) Superimposition of *yp*TpxRED (blue) with the *yp*TpxC61S (green) structure showing that the α1 and α2 helices are in the same orientation. **(B)** The active *yp*TpxOX (purple) structure superimposed with that of *yp*TpxC61S (representing the forced reduced structure) highlights the structural changes in the protein between redox states. The partial unfolding of helix α2, enabling disulphide bond formation, can be seen.

3.2.6 *Solution structure of ypTpx*

The structures of *yp*Tpx and the C61S mutant in solution were investigated using SAXS [110]. Figure 3-10A shows a SAXS curve for *yp*TpxC61S, representative of the data obtained for *yp*Tpx in both oxidising and reducing conditions. It was possible to concentrate *yp*TpxC61S to a higher concentration than the other *yp*Tpx samples so higher quality SAXS data were obtained for the mutant protein. The D_{max} and R_g of *yp*Tpx and *yp*TpxC61S, obtained by indirect Fourier transform with GNOM [116] were the same (D_{max} =70.5 Å and R_g = 24.0 ± 0.2 Å) indicative that conformational changes induced by disulphide bond formation are too small to be detected by SAXS. Theoretical scattering curves of monomeric and dimeric Tpx were calculated from the atomic structures, and again confirm that Tpx is a dimer in solution (Figure 3-10A). A low-resolution (11 Å) envelope of *yp*TpxC61S (Figure 3-10B) in solution was generated using the *ab initio* modelling program DAMMIN [185]. The fit of the model to the data is shown in Figure 3-10B.

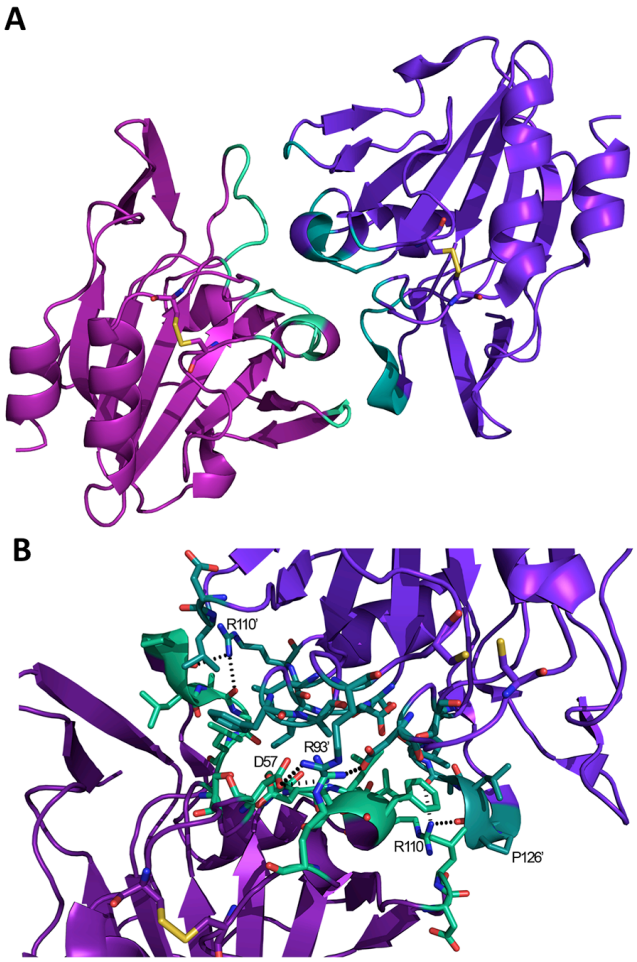


Figure 3-9 Dimer interface of *yp*Tpx. (A) The dimeric form of *yp*TpxOX has been coloured to highlight the dimer interface formed, which is shown in detail in part (B). The disulphide bond has been shown in stick models with the sulphurs coloured in yellow. **(B)** Close-up of the dimer interface with the residues involved in the dimer interface marked as sticks, and salt bridges and hydrogen bonding partners are labelled. Hydrogen bonds are indicated by dashes. This figure has been replicated from [179]**.**

The high-resolution structure superimposes well onto the low-resolution envelope (Figure 3-10B). The D_{max} of the space-fill model of the dimer crystal structure is approximately 68 Å, which agrees with the D_{max} obtained from the SAXS data (70.5 Å), indicating that the low-resolution envelope describes the *yp*Tpx dimer well. The differences in the D_{max} values obtained from the two methods are small, and may be explained by the fact that in the crystal structure there is no electron density to account for the two N-terminal residues of Tpx plus the hexa-histidine tag, therefore it has not been included in the CRYSOL calculations of the scattering curves from the high resolution structures. However, as these residues were present in the *yp*Tpx studied by SAXS, this would lead to the increased D_{max} value observed in solution by SAXS.

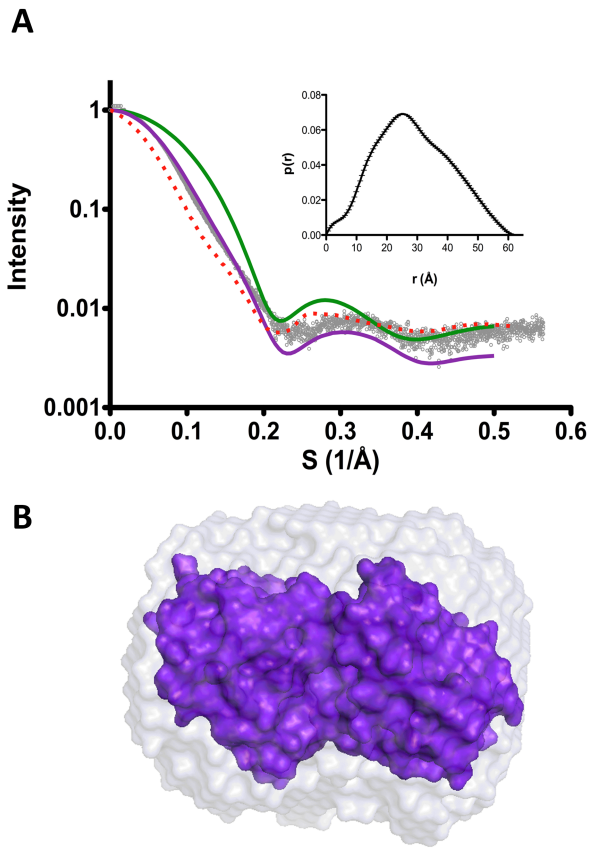


Figure 3-10 Small angle X-ray scattering from *yp*Tpx. (A) The experimental scattering curve of *yp*TpxC61S (grey), overlaid with the curves for monomeric *yp*Tpx (green) and dimeric *yp*Tpx (purple) (calculated using CRYSOL [186]) confirming the dimeric solution state of the protein. The fit of the *ab initio* model to the experimental data is shown in red. Inset is the distance distribution function (p(r) *versus* r) of *yp*TpxC61S with error bars. **(B)** Space-fill model of *yp*TpxC61S crystal structure (purple) superimposed onto the DAMMIN *ab initio* model (grey) using the program SUPCOMB [148]. This figure has been replicated from [179].

Since in solution there is presumably some movement around the dimer interface which cannot be visualised in static crystal structures, rigid body modelling of the oxidised *yp*Tpx crystal structure against the SAXS data, using the program BUNCH, was performed [187]. The search model used in BUNCH was the *yp*Tpx monomer with an imposed *P*2 symmetry (Figure 3-11). This yielded a model similar to that for the dimeric crystal structure. Comparison of the crystallographic model with the one fitted to the solution data using DYNDOM [188] yielded a rotation angle of 21.4° and a 5 Å translation. This freedom of movement corresponds well with that observed for the structures of Tpx from other species, which was determined from comparison of several crystal structures in different crystal forms [182].

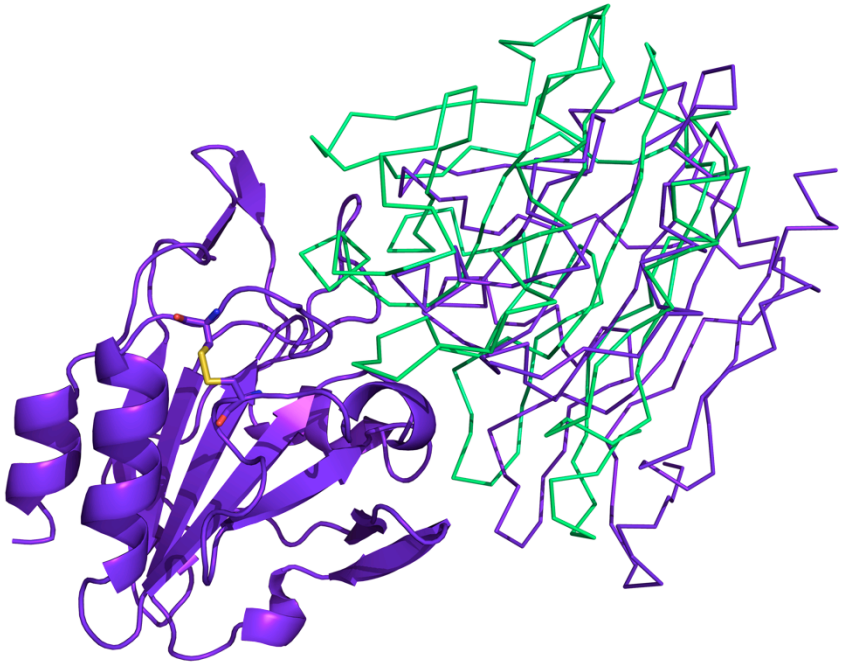


Figure 3-11 Movement around the dimer interface of *yp*Tpx calculated from SAXS data. The crystal structure of the *yp*TpxOX dimer is shown in purple. The monomer shown in cartoon form shows the position of this subunit from both the crystal structure and the BUNCH model [187]. The differently positioned second subunit (in green), modelled by BUNCH against the SAXS data, shows the degree of movement from the crystal structure seen in solution.

3.2.7 *Modelling of flexible regions of ypTpx using SAXS data*

SAXS data were collected for purified *yp*Tpx with the his-tag still present. The C_α positions of the N-terminal and his-tag residues were not observed in the crystal structure and were instead positioned in a manner consistent with the solution data using the program EOM [124] (Figure 3-12). The position of the his-tag was of interest as it was unclear if co-crystallisation of ME0052 with *yp*Tpx had been unsuccessful owing to the his-tag occluding the compound binding site. EOM generates several possible conformations of the flexible regions consistant the SAXS data. From the predicted models it was clear that the his-tag remained flexible and extended into the solvent when in solution and was therefore unlikely to be forming stable interactions with the compound binding site, preventing co-crystallisation.

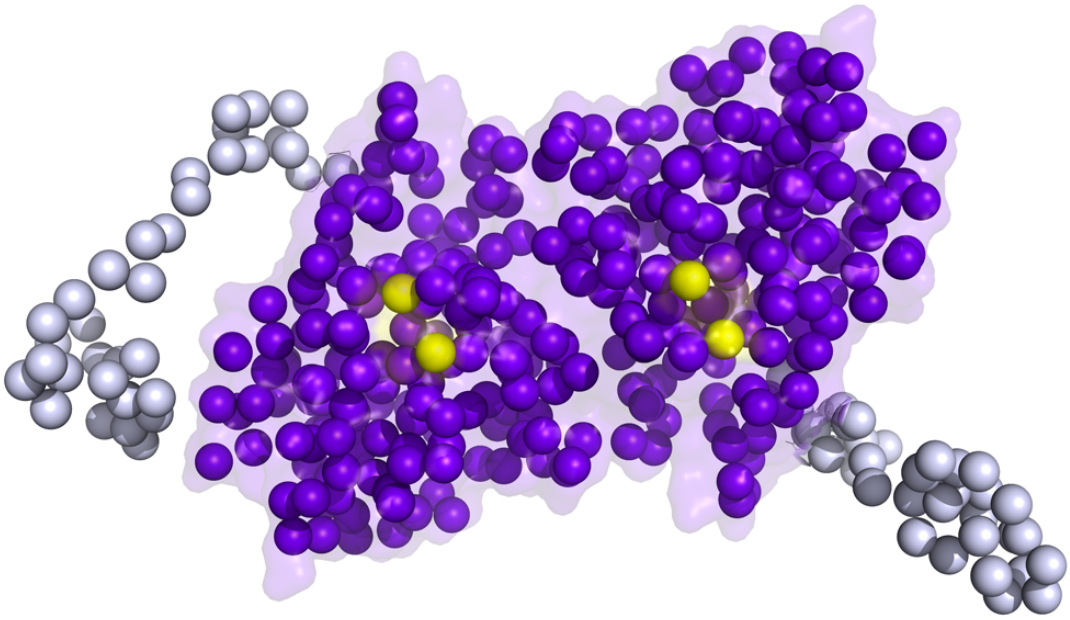


Figure 3-12 EOM modelling of the "missing" his-tag using SAXS data. The *yp*TpxOX dimer crystal structure was modelled against SAXS data with the addition of the missing his-tag to both monomers. One representative EOM model is shown, other models also showed an unstructured his-tag. The Cα crystal structure is shown using a sphere model (in purple) with the surface model overlaid: the his-tag is grey spheres and the catalytic cysteine residues are yellow.

3.3 Does *yp*Tpx bind ME0052?

3.3.1 *Far-western analysis reveals binding of ypTpx to ME0052*

Tpx was identified by an affinity pull-down with the SA compounds. However in order to determine whether *yp*Tpx is a true binding *target* of the compounds, several different binding studies were conducted. Firstly the binding was assessed using far-western analysis, where a biotinylated version of ME0052 was detected with a streptavidin-HRP conjugated secondary antibody (Figure 3-13). The presence of a band indicates that ME0052 has bound to the protein transferred onto the nitrocellulose membrane. The same samples were also analysed by SDS-PAGE and this revealed that a band corresponding to the dimeric form of *yp*Tpx can be seen on the gel. This was intriguing, as typically protein samples are sufficiently denatured during preparation such that only the monomeric form of the protein can be seen. This may suggest either that the dimer formed by *yp*Tpx is particularly resistant to denaturation or that the NU-PAGE (Invitrogen, UK) gel system used does not completely denature the protein samples. However, the appearance of the dimer band enabled investigation into the differences in binding of the

compounds in relation to the oligomeric state of the protein. A strong band corresponding to the *yp*Tpx dimer can be seen on the far-western which does not appear to be proportional to the amount of protein present on the SDS-PAGE gel. This suggested that the compound bound preferentially to the dimeric form of *yp*Tpx. The far-western analysis suggests that binding of ME0052 to *yp*TpxC61S was weaker than to the WT protein. However, this difference could also be attributed to the dimer interface of *yp*TpxC61S being weaker than the WT since it lacks a hydrogen bond formed between D57 and R'93 at the dimer interface as discussed in Section 3.2.5. The binding of ME0052 was investigated further using AUC and NMR.

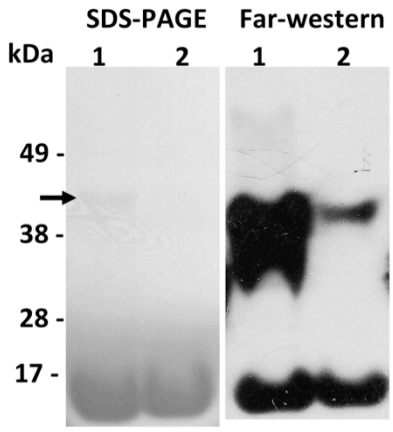


Figure 3-13 Far-western analysis of ME0052 binding to *yp*Tpx and *yp*TpxC61S. 100 µM of *yp*Tpx (1) and *yp*TpxC61S (2) analysed by SDS-PAGE and far-western analysis with biotinylated ME0052. On the SDS-PAGE gel, a faint band corresponding to the dimeric form of the protein can be seen (indicated by the arrow). The positions of the molecular weight markers have been indicated (kDa). This figure has been adapted from [100].

3.3.2 *Determination of affinity and stoichiometry of ypTpx ME0052 complex using AUC*

The SA compounds have an absorbance peak at 395 nm due to their aromatic ring structure (Figure 3-14A). This enabled binding of the compounds to *yp*Tpx in solution to be examined using AUC. This method exploits the fact that free compound does not sediment in the time course of an AUC study, at the speeds utilised, as it has a mass of only 600 Da. Therefore, following sedimentation at a wavelength of 395 nm enabled the quantification of the amount of bound compound as the peak resolved from the data obtained at this wavelength corresponded to the sedimentation of the compound when

bound to the protein. The concentration of ME0052 bound to *yp*Tpx was determined for a range of total compound concentrations. These data were used to calculate the dissociation constant (K_d) and the stoichiometry of binding. Using this approach, the K_d for binding of ME0052 to all redox states of ypTpx was determined to be in the range of 50 - 100 µM (Figure 3-14B). The correlation coefficients (R = 0.8037, 0.9133 and 0.8440 for *yp*TpxC61S, *yp*TpxRED and *yp*TpxOX, respectively) of the non-linear best fits confirmed the strength of dependence between the amount of bound drug and free drug.

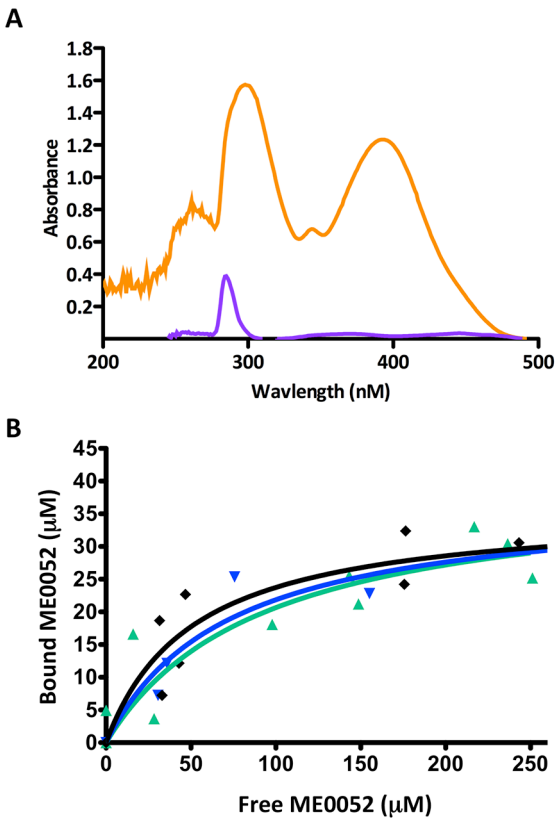


Figure 3-14 AUC analysis of ME0052 binding to *yp*Tpx. (A) The difference in absorption spectrum of ME0052 (orange) and *yp*Tpx (purple). The spectrum for *yp*Tpx starts at 250 nm. **(B)** Line graph showing the binding of ME0052 to purified *yp*Tpx during AUC experiments. The lines represent different redox states for *yp*Tpx: oxidised (black), reduced (blue) and forced reduced mutant, C61S (green). The amount of bound ME0052 was used to calculate the K_d values described in the text. This figure is adapted from [100].

3.3.3 *NMR studies of ME0052-ypTpx interactions*

A $^{1}H^{15}N$-HSQC chemical shift perturbation assay indicated that the addition of ME0052 leads to several chemical shifts within *yp*Tpx. Observed shifts of over 0.2 ppm have been mapped onto the crystal structure of *yp*TpxRED, a cut off value of 0.2 ppm was used to eliminate non-specific shifts from the data (Figure 3-15). The *yp*TpxRED structure was

used for the mapping since DTT was included during the NMR experiments to ensure that only one redox state of Tpx was being measured. The figure shows that the shifting residues localise to a region around the dimer interface. Since Tpx forms a symmetrical dimer, there are binding sites on both sides of the interface (for clarity only one side is shown in the figure). The chemical shift data indicate significant shifts in the amides of residues from both subunits, highlighting the importance of the dimer interface for the generation of the binding pocket and compound binding. The binding region is quite hydrophobic in nature and several of the shifting residues are hydrophobic (V31, A32, V60, V120 and L127). This suggests that hydrophobic interactions may be important to the binding of the compounds, which are also hydrophobic. Knowing the residues involved in compound binding allowed further analysis of how the compound may interact with these residues.

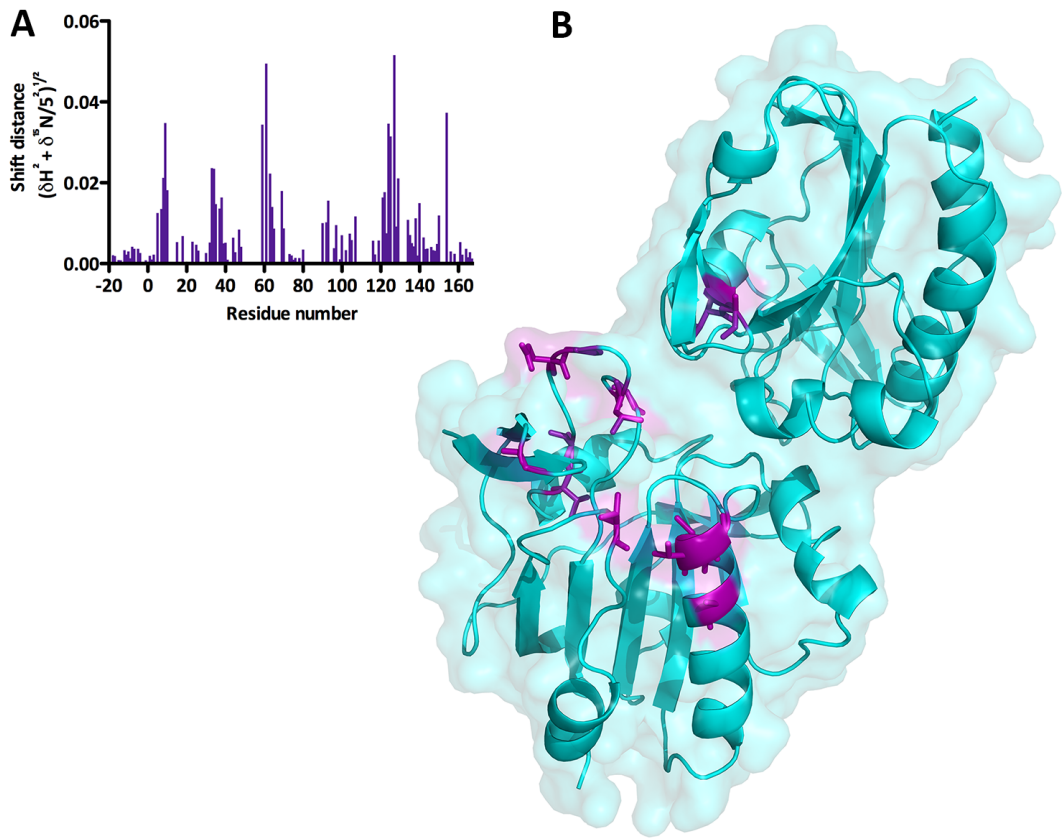


Figure 3-15 Binding region of ME0052 in *yp*Tpx. (A) The shift distances of residues in *yp*TpxRED upon addition of ME0052. **(B)** Dimer of *yp*TpxRED shown as a cartoon model overlaid with a surface representation. The shifting residues (with a distance of over 0.02 ppm) have been shown in stick representation, in purple. The shifting residues have been shown for one monomer of the dimer only.

3.3.4 *Docking of ME0052 into the binding site*

NMR analysis of the binding site defines the residues that shifted in response to compound binding, however it does not give information about the specific chemical interactions between the compound and the protein. In order to gain insight into these interactions, molecular docking was carried out in collaboration with Dr Rommie Amaro and Victoria Feher (University of California). A favorable binding pose for ME0052 in the binding site was determined using the Molecular Operating Environment docking program (Chemical Computing Group Inc., Montreal, Canada). This indicated that hydrophobic interactions were important for compound binding but also that a hydrogen bond may be formed between the ME0052 p-hydroxyl and the *yp*Tpx I153 carbonyl (Figure 3-16).

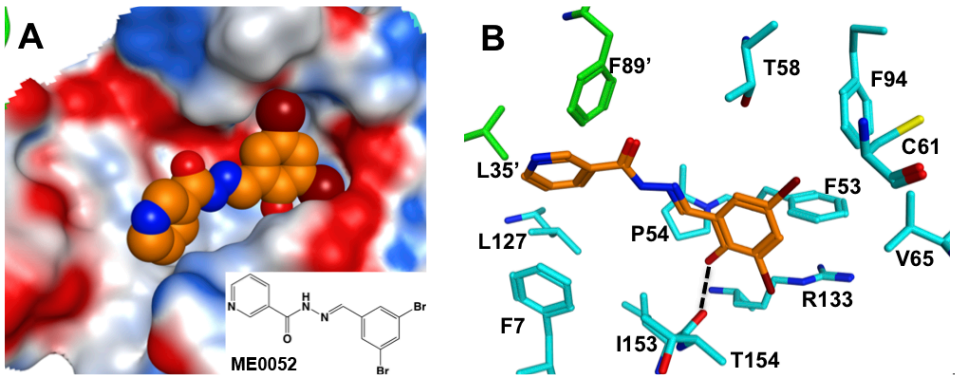


Figure 3-16 Docking of ME0052 into the binding site on *yp*Tpx. (A) Lowest energy binding mode for ME0052 (atoms represented as CPK orange for carbon, blue for nitrogen and red for bromines) and electrostatic surface for the protein binding site. ChemDraw representation of ME0052 is inset. **(B)** Detailed view of ME0052 binding mode (stick representation). *yp*TpxOX residues with atom contacts within 3.5 Å of ME0052 are shown (cyan). Residues F899 and L359 are contributed by the neighbouring monomer (green). A hydrogen bond (dashed line) is formed between ME0052 and the I153 carbonyl. The figure is adapted from [179].

3.4 Crystallographic analysis of a C61S mutant of Tpx from *E. coli O157:H7*

Several attempts to co-crystallise *yp*Tpx or *ec*Tpx with ME0052 were made; however none was successful, despite crystals growing in conditions that contained ME0052. Similarly,

attempts were made to co-crystallise the protein following his-tag cleavage, in case the tag was occluding the ME0052 binding site, however no crystals were obtained. In fact, upon further analysis of the *ec*TpxC61S crystal structures it became clear that this protein would crystallise only when the his-tag was attached, owing to crystal packing requirements.

3.4.1 *Protein purification and crystallisation*

The C61S mutant of *ec*Tpx was created previously by Dr Dai Wang using a QuikChange site-directed mutagenesis kit (Stratagene, UK). The amplified product was cloned into the TOPO pET-151 (Invitrogen) expression vector, which encodes an N-terminal affinity tag consisting of a hexahistidine motif with a TEV cleavage site and a linker region consisting of 25 residues. The resulting construct was transformed into *E. coli* BL21(DE3) cells and grown in 1 L of LB medium containing ampicillin (100 mg ml^{-1}). The protein was purified using IMAC and dialysed against 20 mM Tris pH 7.5, 50 mM NaCl.

His-tagged purified protein at approximately 8 mg ml^{-1} was screened by sitting-drop vapour diffusion using several commercial crystallisation screens with 1 µl drops consisting of a 1:1 ratio of protein and reservoir solutions. Crystals formed in 0.2 M $MgCl_2$, 0.1 M Tris pH 7, 10% PEG8000. The crystals of *ec*TpxC61S were orthorhombic, with dimensions of approximately 0.1 x 0.1 x 0.03 mm (Figure 3-17A).

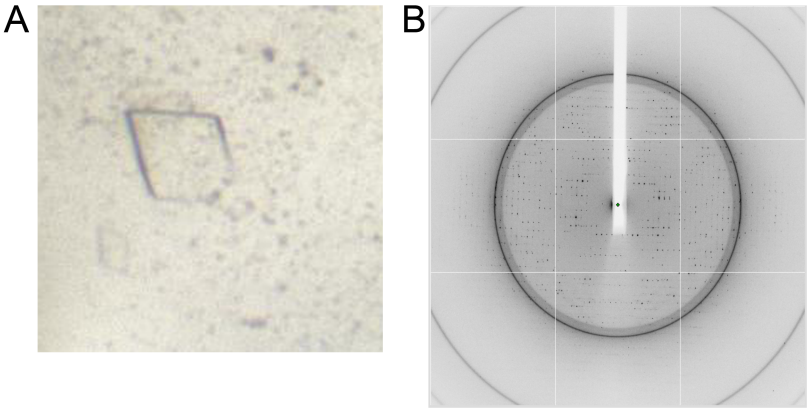


Figure 3-17 *ec*TpxC61S crystals and diffraction. (A) Orthorhombic crystals with dimensions 0.1 x 0.1 x 0.3 mm in 0.2 M $MgCl_2$, 0.1 M Tris pH 7, 10% polyethylene glycol 8000. **(B)** The crystals diffracted to 1.97 Å resolution (at the detector edge) and belonged to space group $C222_1$. This figure has been replicated from [180].

3.4.2 *Structure of ecTpxC61S*

The structure of *ec*TpxC61S was determined using Phaser [154]; the existing TpxC61S structure from *E. coli* (PDB entry 3HVV; [182]) was used as a search model for molecular replacement. Data were collected to a resolution of 1.97 Å (Figure 3-17B), and the relevant data collection and structure refinement statistics are presented in Table 3-2. The asymmetric unit comprised one subunit with a Matthews coefficient of 3.09 $Å^3$ Da^{-1}, indicating a solvent content of 60%. The data processed well, with overall R_{meas} and $R_{p.i.m.}$ values of 14.2% and 5.7%, respectively. The molecular-replacement solution was refined to final R_{work} and R_{free} values of 22.9% and 28.3%, respectively. There are no outliers in the Ramachandran plot and the structure has a MolProbity clash scores in the 94th percentile [157].

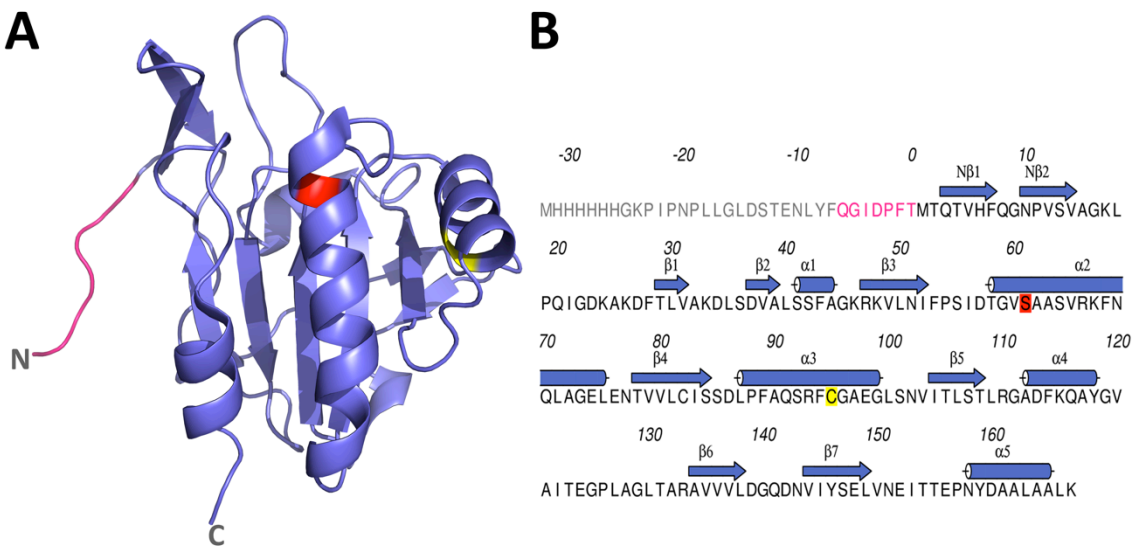


Figure 3-18 The structure of orthorhombic *ec*TpxC61S. (A) The structure of *ec*TpxC61S is shown, with S61 (red) and C95 (yellow) highlighted. The N- and C-termini of the protein have been labelled. **(B)** The sequence of the purified and crystallised protein with secondary structure elements indicated above and the catalytic residues highlighted. The affinity tag is grey, with the part of the tag that can be observed in electron density in pink. This figure is adapted from [180].

*ec*TpxC61S exhibits the same fold as described for the *yp*TpxC61S mutant (Section 3.2.5) and therefore will not be described here (Figure 3-18A). This crystal contained a monomer in the asymmetric unit with a dimer being formed by a crystallographic symmetry-related molecule. The structure of *ec*TpxC61S superimposes well onto other crystal structures of reduced Tpx (Table 3-2). This confirms the assumption made by Hall

et al. that *ec*TpxC61S represents a 'forced reduced' structure [182]. When compared with the oxidised *E. coli* structure (PDB ID: 3HVS; [182]), *ec*TpxC61S superposes with an r.m.s.d. of 0.711 Å. This reflects the local conformational change between the oxidised and the reduced structure mediated by the formation of a disulphide bond between C61 and C95, as discussed earlier.

PDB code	4A2F
Space group	$C222_1$
Unit cell parameters (Å)	a = 49.37, b = 71.75, c = 121.93
Resolution (Å)	40.97 - 1.97 (2.08 – 1.97)
Observed reflections	96980
Unique reflections	15636
Multiplicity	6.2 (6.6)
Completeness (%)	99.9 (100)
R_{meas} (%)	14.2 (67.5)
R_{pim} (%)	5.7 (26.0)
($I/\sigma(I)$)	7.5 (2.8)
Wilson B (Å^2)	22.34
Matthews (Å^3 Da^{-1})	3.09
Solvent content (%)	60.16
Water atoms	170
R_{work} (%)	22.93
R_{free} (%)	28.28
Rmsd for bond lengths (Å) / angles (°)	0.009 / 1.17
Average isotropic thermal parameters (Å^2)	
Main chain	35.34
Side chain	40.95
Water molecule	37.53
Ramachandran outliers (%)	0
MolProbity clash score	6.57 [94th percentile]

Table 3-2 Data collection and refinement statistics for *ec*TpxC61S. Values in parentheses are for the highest resolution shell.

Species	PDB code
Escherichia coli	3HVV
Yersinia pseudotuberculosis	2XPD
Yersinia pseudotuberculosis	2YJH
Mycobacterium tuberculosis	1Y25
Streptococcus pneumoniae	1PSQ
Escherichia coli	3HVS

Table 3-3 Superposition of orthorhombic *ec*TpxC61S from this study with reduced wild-type and 'forced reduced' mutated Tpx structures. The oxidised Tpx from *E. coli* (PDB ID:3HVS) has been added for completeness.

3.4.3 *Analysis of the crystal packing of ecTpxC61S*

The orthorhombic form of *ec*TpxC61S superposes well onto the existing trigonal *ec*TpxC61S structure (PDB ID 3HVV; [182]) with an r.m.s.d. of 0.207 Å, showing that despite the different space groups ($C222_1$ and $P312_1$, respectively) the structure remains essentially the same. However, when investigating the crystal packing, a noticeable difference between the two crystal forms became apparent. The unit cell of the trigonal form of *ec*TpxC61S is much more densely packed, with a solvent content of 42%. In contrast, orthorhombic *ec*TpxC61S has a solvent content of 60%. This explains the difference in the crystal contacts formed between the monomeric asymmetric unit and its environment. Whilst the trigonal crystal form has ten neighbouring *ec*TpxC61S molecules within 5.0 Å, the orthorhombic form has only six. Despite the more spacious packing of the protein in $C222_1$, the crystals diffract similarly and are robust. Analysis of the crystal lattice reveals that the linker region of the affinity-tag construct, which is accounted for in the electron density in orthorhombic *ec*TpxC61S, is involved in several crystal contacts and thus stabilises the crystal through facilitating lattice formation (Figure 3-19A). Analysis of the crystal packing in the orthorhombic crystal lattice using the PISA [184] server revealed that 17% of the total surface area of 8349 $Å^2$ was buried. The largest buried area occurs at the dimerisation interface of the protein, which accounts for about half of the buried surface area. The rest of the crystal contacts are formed between the linker region of the affinity tag of one molecule and a neighbouring molecule, as shown in Figure 3-19B, with a smaller second crystal contact with another subunit. The area of contact conferred by the linker region is stabilised by two hydrogen bonds between the

tag (I4 and F1) and a symmetry-related molecule (K'33) (Figure 3-19B). In comparison, the buried surface area between trigonal *ec*TpxC61S and symmetry-related neighbours makes up around 10% of the total surface area. However, each subunit makes direct crystal contacts with eight symmetry-related molecules. This highlights the importance of the tag in stabilising *ec*TpxC61S in the orthorhombic form.

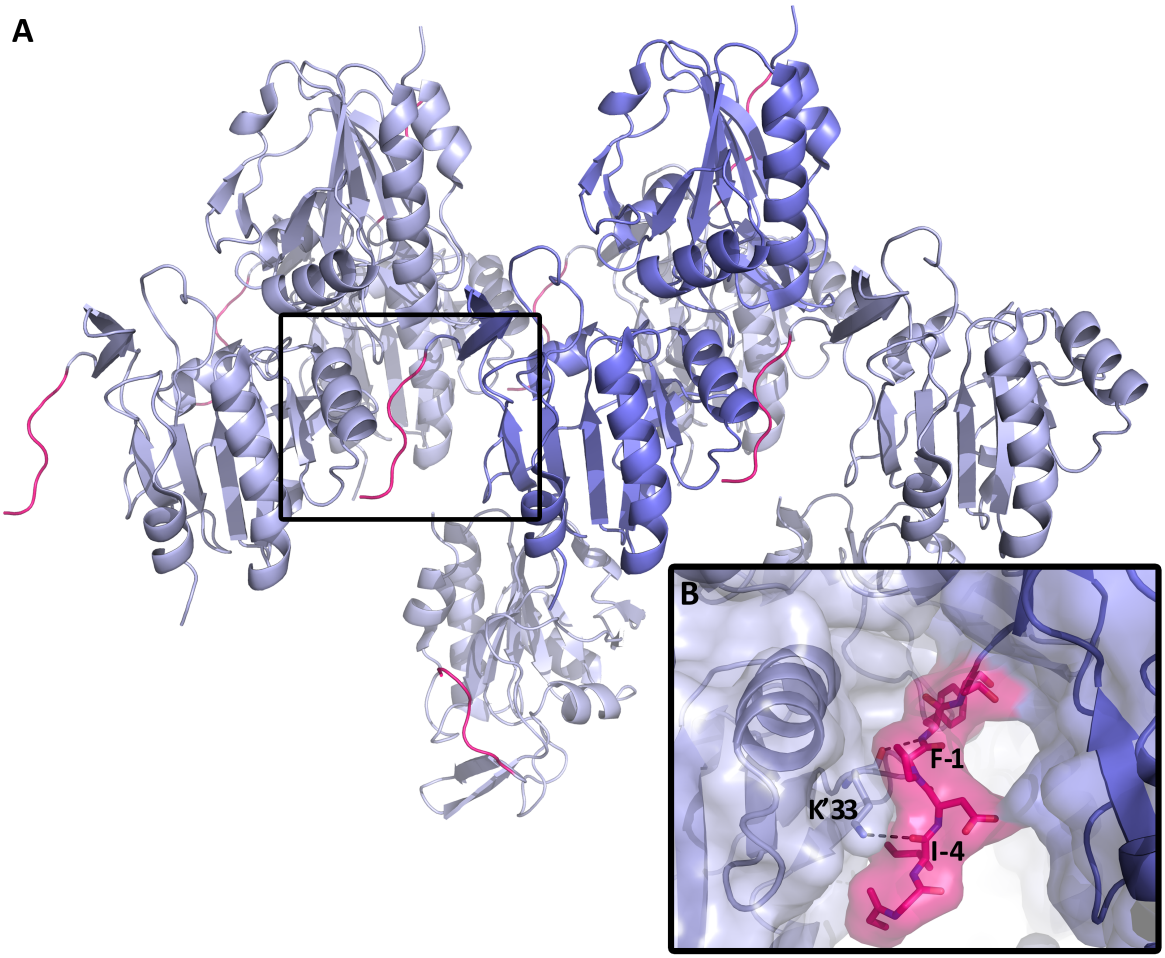


Figure 3-19 Crystal packing of orthorhombic *ec*TpxC61S. (A) The crystal lattice formed in the orthorhombic crystal form of *ec*TpxC61S highlights the importance of the linker region of the affinity tag (pink) in forming crystal contacts. The biological dimer formed by crystallographic symmetry is shown in slate, with other molecules in the crystal lattice shown in grey. **(B)** A detailed view of the crystal contacts between the linker region and a symmetry-related molecule highlighted in the black box shown in (A). The residues involved in hydrogen-bond formation have been labelled, namely K'33 N^{ε} to I4 N (3.2 Å) and K'33 O to F1 O (3.1 Å). Hydrogen bonds are shown as dashed lines. This figure has been replicated from [180].

3.5 Phenotypic analysis of the Δ*tpx* mutant of *E. coli* O157:H7

Deletion of *tpx* in *E. coli* O157 was carried out previously by Dr Dai Wang as was the phenotypic analysis of this strain [100], however some of the results will be mentioned briefly here. Gene expression of the Δ*tpx* mutant was analysed by microarray and showed

that deletion of *tpx* did have an effect on LEE T3SS and flagellar production. However deletion of this gene did not replicate the phenotype seen following addition of the SA compounds. In fact deletion of *tpx* caused a slight increase in LEE T3SS gene expression and a decrease in flagellar gene expression (Figure 3-20). Similar transcriptomic changes were seen in Δ*tpx* complemented with the *tpx*C61S gene onto the chromosome indicating that the changes observed were due to the loss of function of Tpx, since the C61S mutation inactivates the catalytic activity.

The finding that the Δ*tpx* mutant showed increased LEE T3SS gene expression was supported by analysis of the secretion profile of the Δ*tpx* mutant. This showed a slight increase in the level of secreted proteins, compared with the WT [100]. Furthermore, addition of ME0052 led to a decrease in LEE T3S, similar to the reported decrease seen in the WT upon the addition of the SA compounds [99]. Overall this suggests that Tpx may have a role in the regulation of flagellar and LEE T3SS gene expression in *E. coli*. However, deletion of *tpx* does not lead to the phenotype observed upon the addition of the SA compounds suggesting that Tpx is not a key target for the SA compounds.

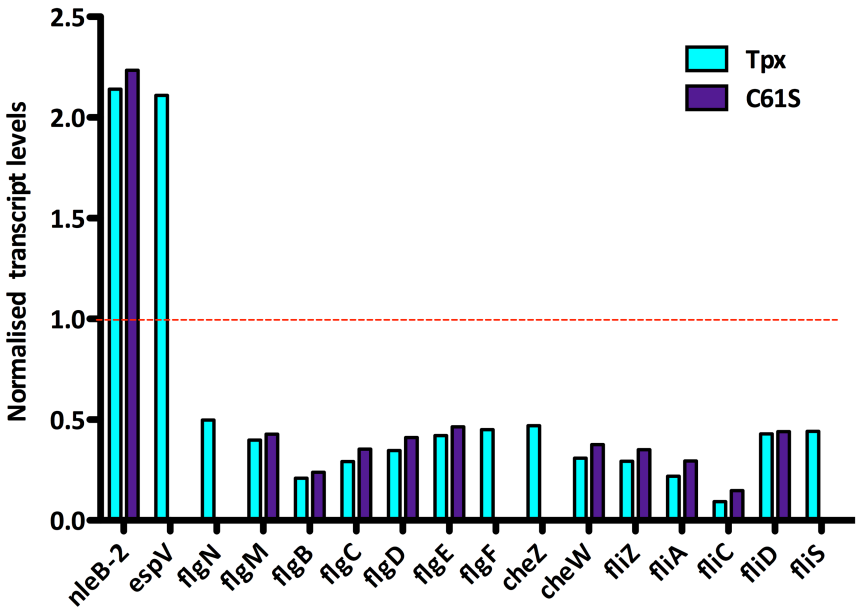


Figure 3-20 Transcriptomic analysis of Δ*tpx* and a *tpx*C61S complement. Bar chart showing significant 2-fold-changes (p <0.005) in LEE T3SS effector gene expression (*nleB*-2 and *espV*) and flagellar gene expression (*flg, che,*and *fli* genes). Data are shown for the Δ*tpx* mutant (cyan) and where significant changes were also seen for the Δ*tpx* mutant complemented with *tpx*C61S back onto the chromosome (purple). The dotted red line indicates the level of the WT transcript levels which was normalised to 1. Data were produced and analysed by Dr Dai Wang and have been taken from the supplementary data from the study by Wang *et al.* [100].

3.6 Discussion

Thiol peroxidase from *E. coli* O157:H7 was identified as a target protein of the salicylidene acylhydrazides using an affinity pull-down assay [100]. It is generally accepted that these kinds of assays can often lead to the identification of targets that are, in fact, false positives. This is not unsurprising, since proteins could either bind to the compounds non-specifically or bind to components of the assay (e.g. the Affi-Gel matrix itself) leading to the inadvertent identification of non-specific proteins. Therefore the validation of target-compound binding is essential. This was initially carried out by far-western analysis of cell lysate from cells over-expressing the target proteins, in this case Tpx. This indicated that Tpx did bind to the biotinylated form of ME0052, however the use of crude cell lysates in the initial experiment led to ambiguous results as the compounds bind to other cellular proteins. Therefore, in this study purified proteins instead of cell lysate were used and offered a less ambiguous interpretation of the binding of ME0052. This approach gave a different insight into the binding that was not seen when using cell lysates, i.e. the apparent preferential binding of the compound to the dimeric form of Tpx. Far-western analysis is not the most ideal way to examine compound-protein interactions because the protein is denatured and must then refold on the nitrocellulose membrane. Despite this inherent limitation, this observation may be accurate as the compound-binding site was found to be in the dimer interface of the protein with interactions from both monomers involved in compound binding.

Initial confirmation of Tpx as a target protein of the SA compounds warranted a more detailed two-fold investigation of the protein: firstly the phenotypic studies of the Δ*tpx* mutant analysing the production of the LEE T3SS and flagellar system and secondly structural characterisation of the interaction between *yp*Tpx and the SA compounds. Structural investigation of *yp*Tpx involved the use of several biophysical techniques including X-ray crystallography, NMR, SAXS and AUC, all of which provided complementary information about *yp*Tpx and its interactions with ME0052.

3.6.1 *Structural studies of Tpx*

In this study, *yp*Tpx was used as it purified easily with a high yield. Crystal structures of *yp*Tpx in three different redox states were obtained: an oxidised, reduced and a catalytic C61S mutant. This enabled an in-depth analysis of this protein and the conformational changes occurring in different redox conditions. Crystallographic analysis of *yp*Tpx revealed that the structure of this protein was highly similar to Tpx structures from different bacterial species. Therefore, the use of *yp*Tpx to model the *E. coli* protein was justified. Once the initial crystallisation conditions had been determined, co-crystallisation of ME0052 with *yp*Tpx was attempted, as more detailed structural information about the interactions occurring between *yp*Tpx and the compounds was an important factor in understanding the mode of action of these compounds. However a co-crystal structure could not be obtained, despite attempts at soaking *yp*Tpx crystals in different concentrations of SA compounds (both ME0052 and ME0055). The reason for this is unclear however, it may be caused by the poor solubility of the compound in aqueous environments. In addition to this, high concentrations of DMSO are known to be less than ideal for crystallisation because DMSO can disrupt the crystal lattice. Since a co-crystal structure could not be obtained, a combination of structural approaches was used to investigate the protein-compound interactions.

The use of AUC demonstrated that *yp*Tpx was a dimer in solution. This had been previously shown for *ec*Tpx, further validating the similarities of the *Yersinia* and *E. coli* proteins [182]. Characterisation of the behaviour of *yp*Tpx in solution was crucial to further biophysical characterisation. The determination of the K_d of ME0052 binding to *yp*Tpx was reliant on the use of AUC since other attempts to quantify this parameter were unsuccessful owing to either the low solubility of the compounds at higher concentrations or the effects of DMSO (detrimental to investigation by ITC).

The development of a novel AUC-based method for this study exploited the fact that the compounds absorb light at several different wavelengths. Therefore the sedimentation of protein bound to compound could be monitored at a wavelength allowing quantification of the compound only. This was used to determine the amount of bound versus free compound, which was then used to determine the binding constant. Determination of the binding affinity of ME0052 to *yp*Tpx in three different redox states revealed very similar

K_d values indicating that the redox state of the protein did not affect compound binding. This was an important finding that aided further structural characterisation using NMR. Since initial NMR ^{1}H-^{15}N HSQC spectra were hard to resolve for the oxidised form of Tpx, DTT was added to improve the quality of the spectra making residue assignment easier.

^{1}H-^{15}N HSQC is a powerful method for investigating the effects on a protein of compound binding. In a ^{1}H-^{15}N HSQC, each peak corresponds to the amine nitrogen from each amino acid present in the protein sequence (apart from proline). The NMR studies showed that ME0052 does indeed bind to Tpx, since the chemical shift perturbations were limited to a small number of residues that clustered to a discrete region in the *yp*Tpx structure. The degree of chemical shift perturbation was dependent on the amount of compound added (i.e. it was concentration dependent), however the poor solubility of the compounds restricted the analysis of these data. Ideally a titration with different concentrations of compound would have been carried out. The concentration of compound could then be plotted against the chemical shift distance and a K_d could have been determined. This would have been a useful result to validate the AUC derived K_d values.

In order to determine the identity of the shifting residues, the residues were assigned by Dr Brian Smith following triple isotope labelling of Tpx (^{15}N, ^{13}C, in D_2O). The combination of these data and the crystallographic data allowed the binding site to be identified. Mapping the shifting residues onto the crystal structure showed that the binding region consisted of a cluster of residues close to the active site and the dimer interface of Tpx. The NMR data did not reveal the specific interactions occurring between ME0052 and Tpx, only the residues involved in the interactions. Therefore, docking approaches were used to predict the chemical interactions taking place. Docking studies indicated that compound binding was primarily mediated by hydrophobic interactions. Since the dimer interface of *yp*Tpx is primarily stabilised by hydrophobic interactions, this may suggest that the compounds are likely to bind to non-specifically to hydrophobic regions on proteins. This affinity for hydrophobic regions may be a reason for the identification of several putative targets. The work in this thesis therefore demonstrated that the compounds do bind to Tpx. However, it is critical to also determine if the target is important for the desired phenotype: an "on target" hit or whether it is an "off target" hit, a non-desirable interaction. To evaluate this, the phenotype of the knockout mutant

needs to correspond to the phenotype seen following treatment of whole cells with the compounds, i.e. a decrease in LEE T3S and an increase in flagellar expression.

3.6.2 *Phenotypic studies of Δtpx*

In-depth structural studies of *yp*Tpx and its binding interactions with ME0052 revealed that *yp*Tpx is in fact a target of the SA compounds. However, phenotypic studies of a *tpx* deletion mutant of *E. coli* O157 did not show a marked effect on LEE T3S. In fact the pattern of expression was reversed compared with the gene expression patterns observed following treatment with the SA compounds, i.e. there was an increase in LEE T3SS gene expression but a decrease in flagellar gene expression. This suggests that although there is an interaction between ME0052 and *yp*Tpx, this interaction is not responsible for the SA induced phenotype. Since both flagellar and LEE T3SS genes are affected, it would appear that Tpx does play a role in regulating the expression of these virulence factors. The mechanism underlying its role is not clear, but since Tpx is one of the proteins responsible for counteracting oxidative stress, it is possible that an increase in oxidative stress may be responsible. The expression of the alternative sigma factor RpoS in *E. coli*, for example, is increased by several different environmental stresses, including oxidative stress [189]. This sigma factor has been shown to positively regulate the expression of *LEE3* and *tir* in *E. coli* O157 [190]. However the transcriptomics data did not show a significant increase in the expression of these genes in the Δ*tpx* mutant.

3.6.3 *Conclusion*

Tpx was one of the 19 putative binding targets of the SA compounds. This study shows that Tpx binds ME0052, validating that the affinity pull-down approach did identify at least one target that interacts with the SA compounds. That said, this study also shows that targeting Tpx does not lead to a decrease in LEE T3S and an increase in flagellar expression, the phenotype observed following SA treatment. Therefore, it would appear that Tpx is not the key target of the SA compounds. Knowing this, other targets were further investigated.

4 Structural investigation of the salicylidene acylhydrazide target proteins FolX and WrbA

4.1 Introduction

Two other proteins identified in *E. coli* O157 as putative SA targets are 7,8-dihydroneopterin-triphosphate-epimerase (FolX) and the tryptophan repressor binding protein (WrbA). The main focus of this chapter is the structural characterisation of these two proteins with an emphasis on demonstrating that they are indeed genuine SA binding proteins. In the publication characterising the putative SA targets, Dr Dai Wang carried out far-western analysis showing binding of ME0052 to both FolX and WrbA from several Gram-negative pathogens including *E. coli, Salmonella*, *Shigella*, *Psuedomonas* and *Yersinia* [100]. Furthermore, transcriptomic analyses on the deletion mutants Δ*folX* and Δ*wrbA* were carried out. This work will be mentioned here for completeness in the discussion, however the main focus of this chapter is the structural characterisation of these proteins that was undertaken during this PhD.

4.1.1 *FolX*

Relatively little is known about the enzyme FolX and thus far it has been identified only within the *Gammaproteobacteria*. FolX is connected to folate biosynthesis as it catalyses the conversion of 7,8-dihydroneopterin triphosphate to dihydromonapterin-triphosphate [191], which redirects 7,8-dihydroneopterin triphosphate away from the synthesis of tetrahydrofolate. However the importance of this step is unknown since *folX* deletion mutants show no obvious defects [191]. Recently, it has been shown that deletion of *folX* confers some resistance to drugs that target folate biosynthesis, such as sulphamonomethoxine and trimethoprim, potentially by altering the flux of metabolites through this pathway [192].

The crystal structure of FolX from *E. coli* has been solved to 2.9 Å [193]. The *ec*FolX structure suggests that this protein forms an octamer however the oligomeric state of FolX was investigated further here. To characterise the binding interactions between FolX and the SA compounds co-crystallisation with FolX from several different species was attempted. However, only the holo-form of FolX from *P. aeruginosa* crystallised, the structure of which was determined by Dr Mads Gabrielsen. Since FolX from *P. aeruginosa* shares only 62% identity with the *E. coli* homologue, it was unknown whether FolX would

adopt the same oligomeric state as the *E. coli* protein. This was investigated using AUC and SAXS.

4.1.2 *WrbA*

WrbA was originally identified as a binding partner of the tryptophan repressor protein hence the name tryptophan (W) repressor binding protein (WrbA) [194]. However since its discovery, its role in the regulation of the tryptophan regulon has been disputed and it has been implicated in several other cellular processes, such as redox regulation [195]. WrbA is a 21 kDa protein that forms a dimer which participates in a dimer-tetramer equilibrium [196]. Sequence analysis revealed that WrbA belongs to the flavodoxin-like family of proteins. WrbA binds to a flavin mononucleotide (FMN) cofactor, with one FMN per monomer [196]. WrbA from *E. coli* and *Archeoglobus fulgidus* display FMN-dependent NAD(P)H:quinone oxidoreductase activity [162].

The expression of *wrbA* is partially controlled by the stress response sigma factor (RpoS) and several studies report the up-regulation of WrbA in response to H_2O_2, acids and increased salt concentrations [194,196]. This led to the hypothesis that WrbA may be important during the course of an infection.

4.1.3 *Aims of this chapter*

The aims of this chapter were to carry out structural characterisation of the target proteins FolX and WrbA using a combination of biophysical approaches including X-ray crystallography, small angle X-ray scattering and analytical ultracentrifugation. The work carried out on FolX forms the basis of the following publication (* denotes joint first authorship):

- *Gabrielsen, M., ***Beckham, K. S. H**., Cogdell, R. J., Byron, O. and Roe, A. J. (2012) FolX from *Pseudomonas aeruginosa* is octameric in both crystal and solution. FEBS Lett., **586**, 1160–5 [197]

4.2 Structural characterisation of FolX

4.2.1 *Oligomeric state of FolX*

FolX crystallised in space group *I*432, with a single chain in the asymmetric unit. The PISA server [184] suggests that the oligomeric state, based on the crystallographic symmetry, is tetrameric (Figure 4-1A). However, previous studies have suggested that FolX may be an octamer [191,193]. Analysis of the interfaces formed in the octamer using PISA [184] showed that the interface between the dimer of tetramers (Figure 4-1B) involves 7% of the solvent accessible area (SAA). This is considerably smaller than the SAA involved in stabilising the FolX tetramer, which accounts for 18% of the SAA. Analysis of the interface between the tetrameric rings, using PISA, indicates that there are 16 hydrogen bonds and 8 salt bridges connecting the two tetramers, which suggests that this interaction is important despite the relatively small interface area.

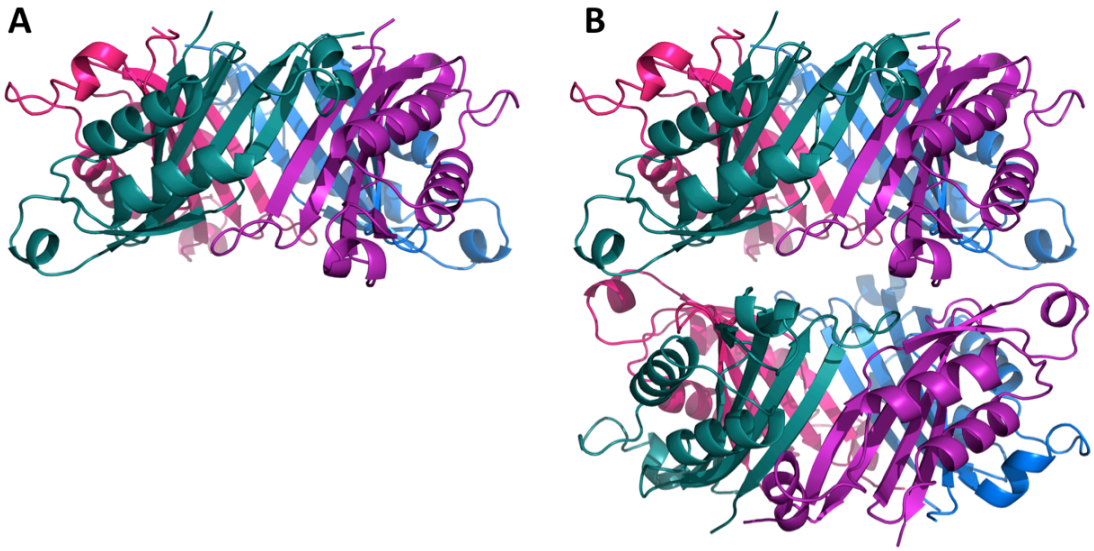


Figure 4-1 FolX as a tetramer and octamer. (A) In the crystal structure (PDB ID: 4AEY [197]) FolX forms a tetramer. However, this tetramer can dimerise forming an octamer **(B)**, as seen in the *E. coli* structure (PDB ID: 1B9L [192]). Monomer subunits have been highlighted in different colours. This figure has been replicated from [197].

In order to try to obtain conclusive evidence for the quaternary state of FolX, AUC was conducted. Sedimentation velocity experiments revealed that FolX was present as a single

species in solution, evidenced by a single peak in the c(s) distribution (Figure 4-2A). The infinite dilution sedimentation coefficient ($s^0_{20,w}$) of FolX, derived from the concentration dependence of $s_{20,w}$, determined by fitting the data with a non-interacting discrete species model, is 6.09 ± 0.03 S. This corresponds with the value of $s^0_{20,w}$ computed using the program SOMO [125] for the octamer crystal structure (5.97 S) and not with that computed for the tetramer (3.62 S). Sedimentation equilibrium data fitted with a single species model indicated the presence of a species with a mass of 141.500 ± 6.660 kDa at infinite dilution (M^0). This value is slightly lower than the calculated octamer mass of 143.864 kDa. In order to improve the fit of the model parameters, the effects of non-ideality and the presence of a tetramer-octamer equilibrium were introduced into the data analysis. Inclusion of non-ideality improved the χ^2 of the global fit from 0.01890 to 0.00316. Extending the model to include a tetramer-octamer equilibrium further improved the fit to a χ^2 of 0.00297 and gave a K_d of 0.887 µM. The fit to the data along with the resultant residuals is shown in Figure 4-2B and C.

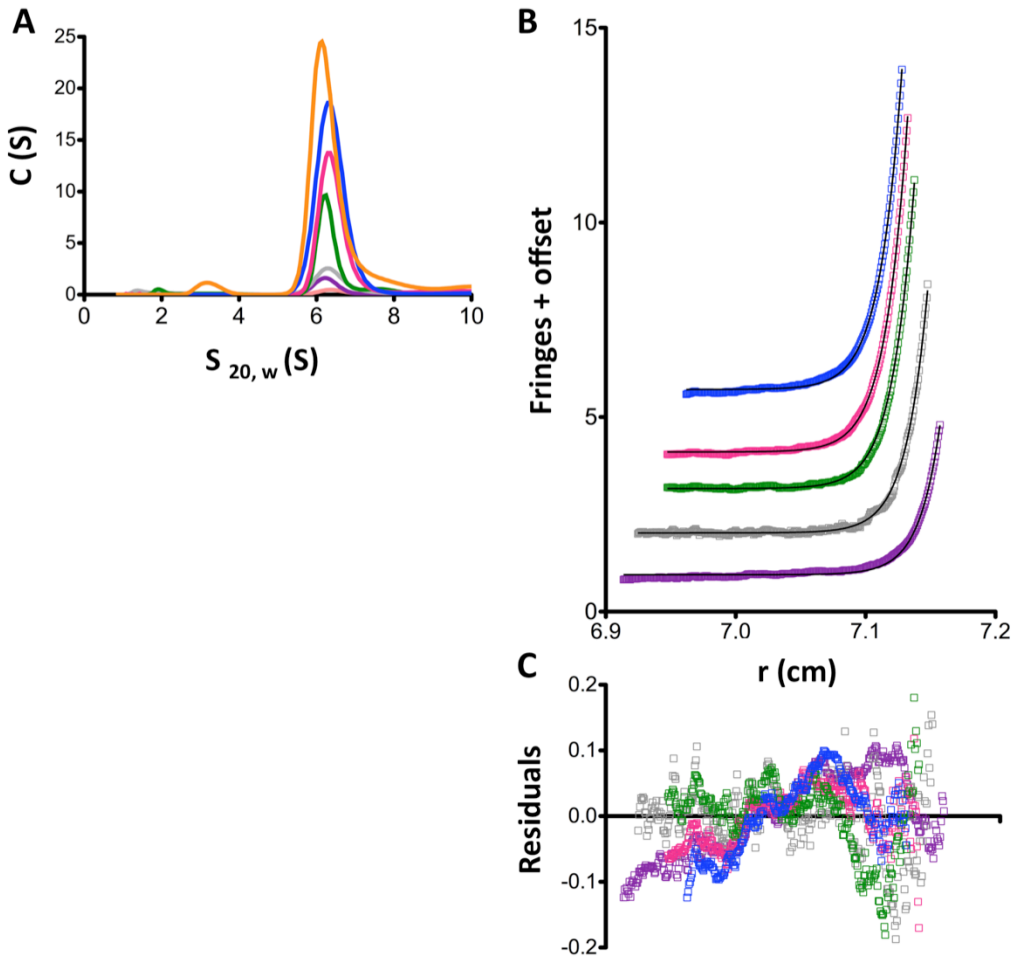


Figure 4-2 AUC analysis of FolX. (A) c(s) distributions derived via SEDFIT from SV interference data for varying concentrations of FolX are dominated by a peak at $s_{20,w} \approx 6$ S. Data for FolX at different concentrations are shown in different colours: 0.2 mg ml^{-1}, light pink; 0.5 mg ml^{-1}, purple; 1 mg ml^{-1}, grey; 2.5 mg ml^{-1}, green; 5 mg ml^{-1}, pink, 7.5 mg ml^{-1}, blue; 10 mg ml^{-1}, orange. **(B & C)** The global fit to SE absorbance data using a non-ideal tetramer-octamer self-association model. The experimental data and fits are shown in (B); the residuals for each fit are plotted in (C). Different protein concentrations are represented by the colour scheme used in (A).The 10 mg ml^{-1} data are not shown here since the absorbance of the sample was too high. Similarly, the signal from lowest concentration (0.2 mg ml^{-1}) was too weak to analyse and these SE data are also omitted. This figure has been replicated from [197].

4.2.2 *Solution structure of FolX*

To further investigate the oligomeric state of FolX, the solution structure was determined using SAXS. Indirect Fourier transformation of the scattering data using GNOM [116] indicated a D_{max} of 177 Å and an R_g of 44.1 ± 0.6 Å. An *ab initio* model of FolX was generated using DAMMIF [122], imposing *P*4 symmetry, based on the crystal structure, and the fit of the final averaged model to the experimental data is shown in Figure 4-3A.

Using sedimentation equilibrium the K_d of FolX oligomerisation was determined to be 0.89 µM, therefore at the protein concentration used in the SAXS study 99.5% (by mass) of the protein would have been in the octameric state. The crystal structure of the octamer was superimposed onto the 10.5 Å resolution envelope of FolX in solution, (Figure 4-3B). The octamer crystal structure fits well into the envelope confirming that this is the true oligomeric state in solution.

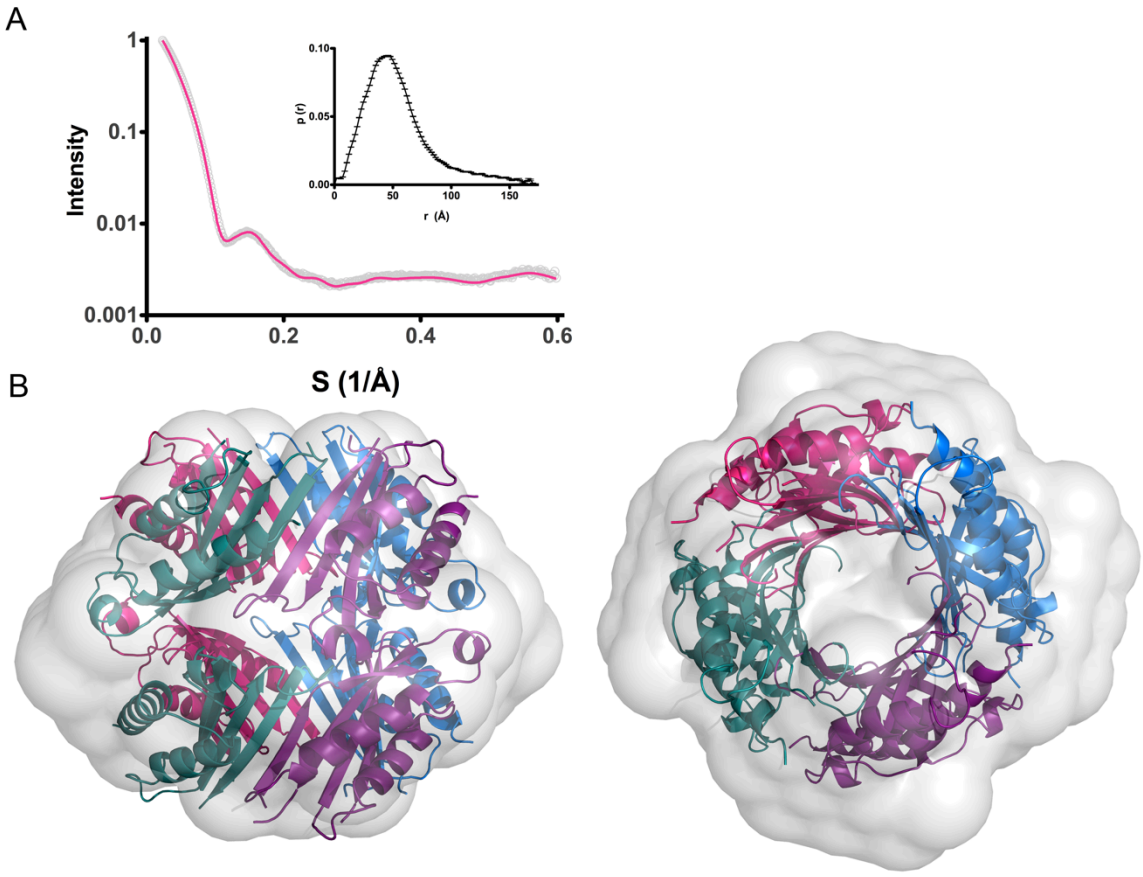


Figure 4-3 Solution structure of FolX. (A) The experimental scattering data (grey) with the fit of the DAMMIF model to the data shown in pink. Inset is the pairwise distribution p(r) function of the data. **(B)** The octamer crystal structure superimposed onto the DAMMIF model shown as a side and top view.

4.3 Structural characterisation of WrbA

4.3.1 *Expression, purification and crystallisation of WrbA*

WrbA from *E. coli, S. typhimurium* (*st*) and *Y. pseudotuberculosis* were previously cloned into pET-151 by Dr Dai Wang [100]. The proteins were overexpressed and purified by IMAC and placed into commercial crystallisation screens as described in Sections 2.9 and 2.15, respectively. Crystals formed for *st*WrbA in several conditions however the crystal described below formed in 8% (v/v) tacsimate, pH 8, 20% (w/v) PEG 3350 with 100 µM ME0052. The crystals were cubic with dimensions of 0.01 x 0.01 x 0.01 mm and were yellow (Figure 4-4A). The crystals diffracted to 2.9 Å resolution (Figure 4-4B).

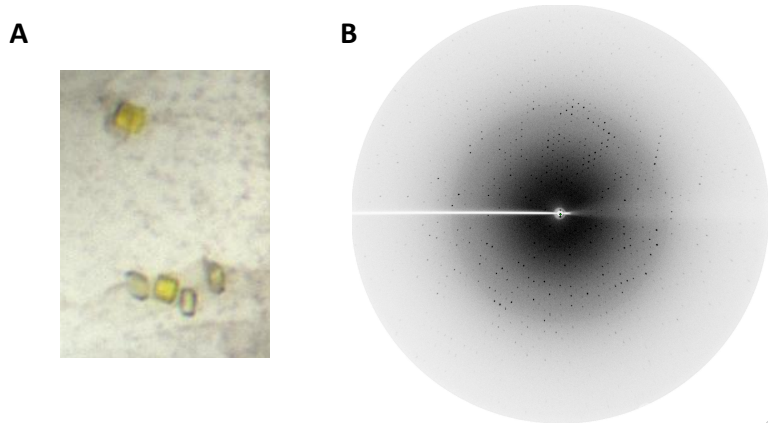


Figure 4-4 *st*WrbA crystals and diffraction pattern. (A) Yellow cubic crystals with dimensions 0.01 x 0.01 x 0.01 mm in 8% (v/v) tacsimate, pH 8, 20% (w/v) polyethylene glycol 3350 with 100 µM ME0052. **(B)** The crystals diffracted to 2.9 Å resolution (at the detector edge) and belonged to space group $P2_12_12_1$.

Crystallisation screens were set up both with and without 200 µM ME0052. Although the level of sequence similarity of WrbA is high between species (Figure 4-5) it is possible that the slight differences in their primary sequences would make one protein more amenable to crystallisation than the others. This was found to be the case as no crystals were obtained for *ec*WrbA despite being 98% identical to *st*WrbA and only poorly diffracting crystals were obtained for *yp*WrbA.

```
ec  MAKVLVLYYSMYGHIETMARAVAEGASKVDGAEVVVKRVPETMPPQLFEKAGGKT-QTAP 59
st  MAKILVLYYSMYGHIETMAHAVAEGAKKVDGAEVIIKRVPETMPPEIFAKAGGKT-QNAP 59
yp  MAKILVLYYSMYGHIETLAGAIAEGARKVSGVDVTIKRVPETMPAEAFAKAGGKTNQQAP 60
    *** ************* * * **** ** *  *  ********   * ****** * **

ec  VATPQELADYDAIIFGTPTRFGNMSGQMRTFLDQTGGLWASGALYGKLASVFSSTGTGGG 119
st  VATPQELADYDAIIFGTPTRFGNMSGQMRTFLDQTGGLWASGSLYGKLGSVFSSTGTGGG 119
yp  VATPHELADYDGIIFGTPTRFGNMSGQMRTFLDQTGGLWASGALYGKVASVFASTGTGGG 120
    **** ****** ****************************** ****  *** *******

ec  QEQTITSTWTTLAHHGMVIVPIGYAAQELFDVSQVRGGTPYGATTIAGGDGSRQPSQEEL 179
st  QEQTITSTWTTLAHHGMVIVPIGYAAQELFDVSQVRGGTPYGATTIAGGDGSRQPSQEEL 179
yp  QEHTITSTWTTLAHHGFIIVPIGYGAKELFDVSQTRGGTPYGATTIAGGDGSRQPSAEEL 180
    ** *************  ****** * ******* ********************** ***

ec  SIARYQGEYVAGLAVKLNG 198
st  SIARYQGEYVAGLAVKLNG 198
yp  AIARFQGEHVAKITAKLKG 199
     *** *** **    ** *
```

Figure 4-5 Sequence alignment of WrbA from different species. ClustalW [181] sequence alignment of the protein sequence of WrbA from *E. coli* (*ec*), *S. typhimurium* (*st*), and *Y. pseudotuberculosis* (*yp*). The level of sequence conservation between the three species is indicated, * indicates 100% sequence identity.

WrbA from *S. typhimurium* crystallised in the space group $P2_12_12_1$ with unit cell parameters of a = 92.16, b = 92.81, c = 93.97 Å. The asymmetric unit contained a tetramer with Matthews coefficient of 2.24 $Å^3$ Da^{-1}, indicating a solvent content of 44.9%. The data processed well, with overall R_{meas} and $R_{p.i.m.}$ values of 15.3% and 4.2%, respectively. The molecular-replacement solution was refined to final R_{work} and R_{free} values of 17.73% and 24.54%, respectively. There are no outliers in the Ramachandran plot and the structure is in the 99th percentile of MolProbity clash scores [157]. All relevant structure refinement statistics are listed in Table 4-1.

4.3.2 *Crystal structure of stWrbA in complex with ME0052*

The structure of *st*WrbA was determined by molecular replacement using Phaser [154] using the existing WrbA structure from *E. coli* (PDB ID: 2R96 [198]) as the search model. Overall the structure of *st*WrbA is similar to that of *ec*WbrA and superimposes well with an r.m.s.d. of 0.235 Å. *st*WbrA crystallised as a tetramer in the asymmetric unit (Figure 4-6A) and the four chains superimpose well with an r.m.s.d. of less than 0.23 Å. Each monomer contained an FMN cofactor and three of the 4 monomers showed evidence for the presence of ME0052 in the electron density, discussed below.

Space group	$P2_12_12_1$
Unit cell parameters (Å)	a = 92.16, b = 92.81, c = 93.97
Resolution (Å)	66.04 - 2.29 (2.30- 2.42)
Observed reflections	48244
Unique reflections	36656
Multiplicity	11.8 (12.8)
Completeness (%)	98.6 (98.4)
R_{meas} (%)	15.3 (3.2)
R_{pim} (%)	4.2 (0.9)
($I/\sigma(I)$)	57.4 (3.6)
Wilson B (Å^2)	32.32
Matthews (Å^3 Da^{-1})	2.24
Solvent content (%)	44.9
Water atoms	147
R_{work} (%)	17.43
R_{free} (%)	24.54
Rmsd for bond lengths (Å) / angles (°)	0.015 / 1.84
Ramachandran outliers (%)	0
Molprobity clash score	6.09 [99th percentile]

Table 4-1 X-ray crystallography data collection and refinement statistics for *st*WrbA. Values in parentheses correspond to the highest resolution shell.

Previous studies have reported that WrbA is exists in a dimer-tetramer equilibrium [196]. In the structure chains A and C, and B and D form homodimers. *st*WrbA exhibits the typical WrbA fold consisting of 5 parallel β-strands (β1-5) which form a twisted β-sheet (Figure 4-6B). Surrounding the twisted β-sheet are seven α-helices, which are insertions to the sequence and follow each β-sheet. The interface between dimers is formed primarily through interactions between α4 from each partner. This interface accounts for 7.8% of the SAA and involves 26 residues and 12 hydrogen bonds. The dimer interface is formed between α5 and the loop region between α5 and β5, reciprocally, from chain A to B and from C to D, and accounts for 10% of the total SAA of a monomer involving 28 residues forming 8 hydrogen bonds.

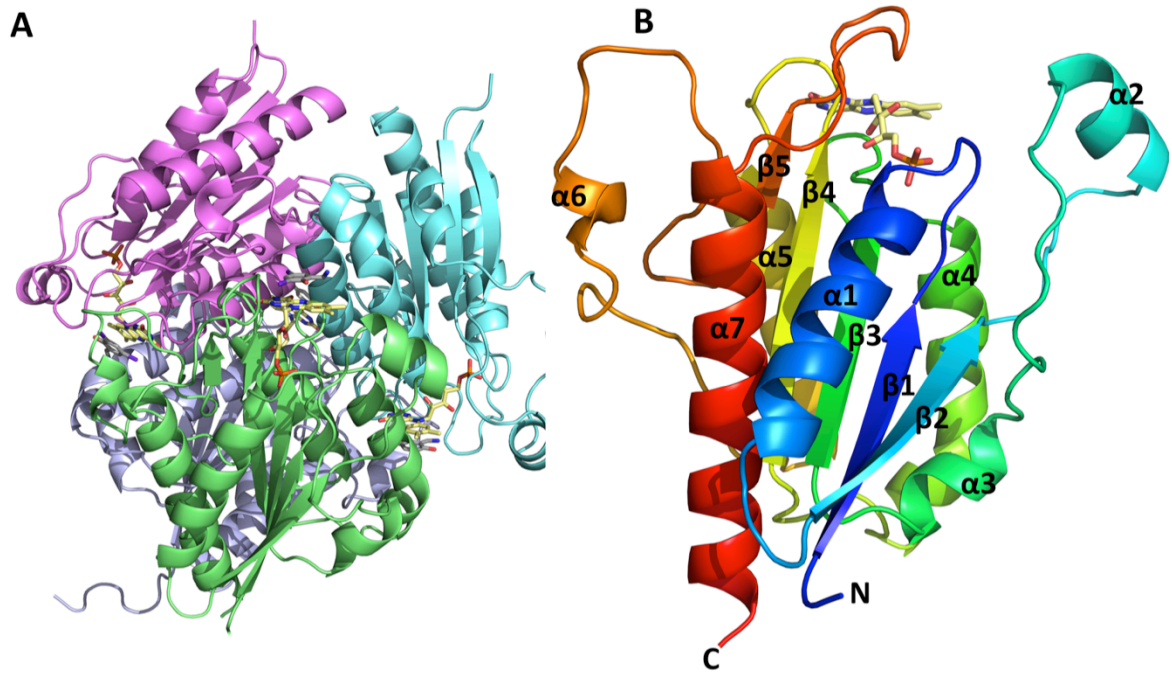


Figure 4-6 Structure of *st*WrbA. (A) *st*WrbA forms a tetramer in the asymmetric unit of the crystal. The tetramer is composed of a dimer of dimers (pink with light purple and blue with green) and each monomer has its own FMN cofactor (yellow). The ME0052 fragment that co-crystallised in the structure is shown in grey stick representation. **(B)** Monomer of *st*WrbA showing the overall fold of the protein. The N-terminus is blue and the C-terminus red. α-helices and β-sheets have been labelled according to their order in the protein sequence. The FMN cofactor is yellow with nitrogen and oxygen atoms in blue or red, respectively.

Refinement revealed density corresponding to FMN being present in the structure, i.e. the holo-enzyme had been crystallised. The FMN cofactor binds to the periphery of the monomer, interacting with residues from the loop regions at the C-terminal end of the β-strands. The interactions formed between the FMN and the protein are shown in Figure 4-7. As seen in the *E. coli* protein, the N5 atom of the isoalloxazine ring forms a long hydrogen bond (3.13 Å) to the backbone nitrogen of Phe 80 [199]. Below the FMN, Arg 79 stabilises the cofactor through hydrophobic stacking interactions. Several other hydrogen bonds are formed between protein backbone nitrogen atoms and oxygen atoms in the FMN, especially around the FMN phosphate moiety. Tetramerisation of *st*WrbA appears to increase interactions with the FMN since His 133 from chain C forms a hydrogen bond with the O4 of the isoalloxazine ring in chain B (this interaction is also seen in the other monomers).

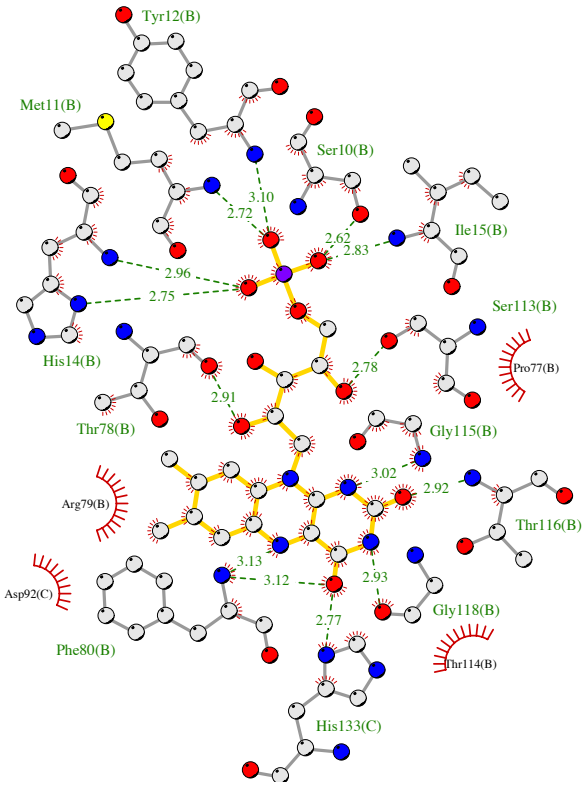


Figure 4-7 Interactions of FMN with *st*WrbA. Ligplot+ [200] representation of the interactions between FMN (yellow) and *st*WbrA (grey) [201]. The interactions have been shown for only one monomer, however this is representative of the other FMN cofactors in the structure. Atoms have been coloured as follows: oxygen, red; nitrogen, blue; carbon, grey; sulphur, yellow; phosphorous, purple. Hydrogen bond distances are shown in green and red lines indicatedhydrophobic interactions.

The FMN cofactor clearly interacts with *st*WrbA, however, inserting FMN was not sufficient to explain all the electron density features in the active site. Further rounds of refinement revealed clear density corresponding to a small molecule. Since the crystal was grown in the presence of ME0052 this was also inserted into the density. However, the electron density could not accommodate the whole compound structure. Only the halogenated portion of the molecule inserted credibly into the density. Owing to the nature of the acid sensitive bond, it is likely that ME0052 is cleaved in half. The suggested catalytic mechanism for the hydrolysis of the ME0052 imine bond is shown in Figure 4-8.

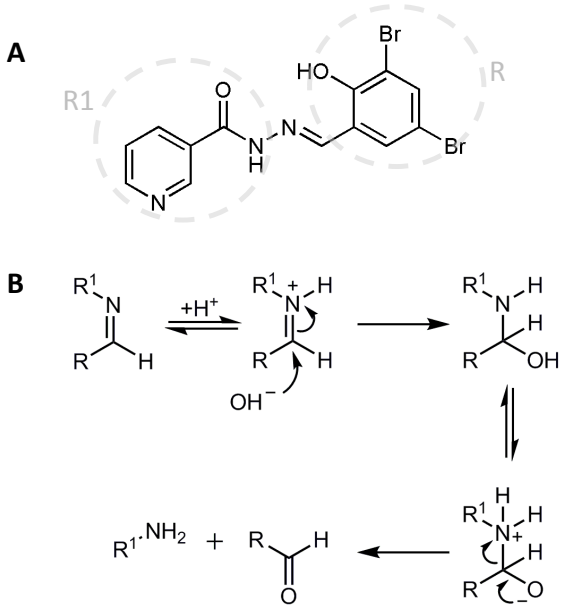


Figure 4-8 Catalytic mechanism for the hydrolysis of ME0052. (A) The complete structure of ME0052 is shown with the corresponding 'R' groups shown in (B). (**B)** There is an equilibrium between the imine bond and the iminium ion. The iminium ion is attacked by a hydroxy ion, which comes from water, this attack in turn forms a quaternary intermediate. This intermediate is in equilibrium with a second intermediate, where the oxygen's hydrogen is transferred to the nitrogen. At this point the lone pair of electrons in the oxygen displaces the nitrogen, forming the corresponding aldehyde and amine. Chemical structures were drawn using ChemDraw (PerkinElmer, UK).

The electron density for the *st*WbrA active site is shown in Figure 4-9, where an omit map indicates the presence of both the FMN and the compound in the density. The occupancy of ME0052 is less than 1, as the density for each monomer is not of equal quality. This indicates that not all subunits in the asymmetric unit are bound to ME0052 which may be reflective of a low affinity of binding or that the concentration used in the crystallisation screen was insufficient to allow full occupancy. Bromines are more electron dense than other atoms in the structure and should therefore generate more scattering. Evidence for the presence of these heavy atoms can be seen in the $2F_o$-F_c map at high signal to noise levels (5σ), further validating the presence of the compound in the structure.

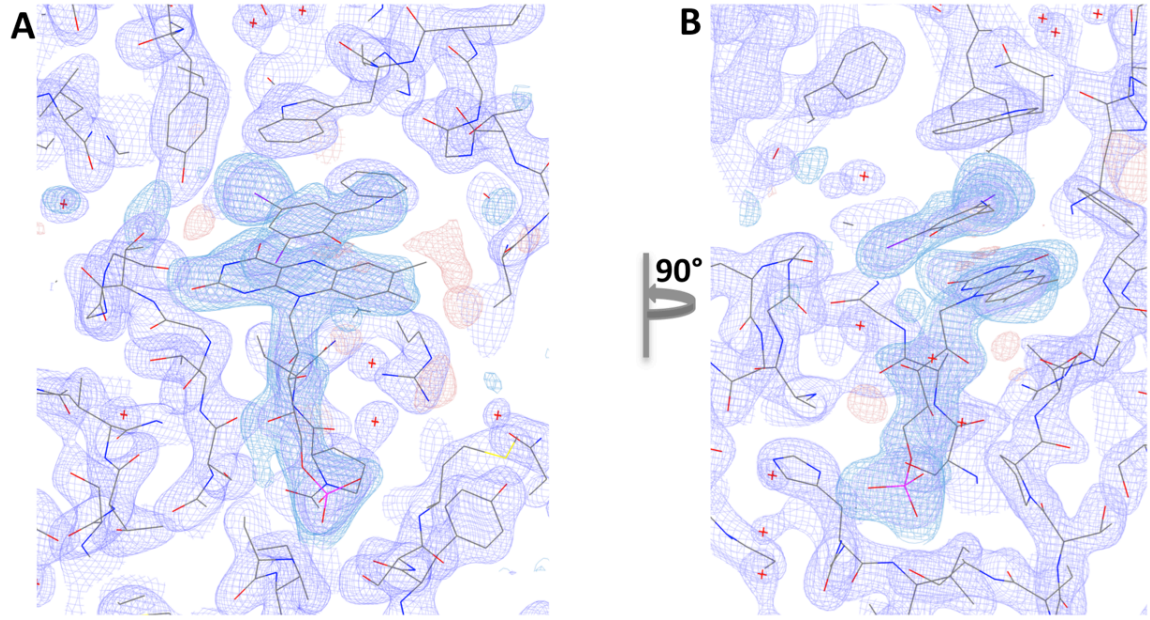


Figure 4-9 Omit map showing the positions of FMN and ME0052 in the *st*WrbA structure. An omit map was generated by removing the FMN and ME0052 from the completed structure and performing 20 rounds of restrained refinement in REFMAC. The generated $2F_o$-F_c (dark blue) and mF_o-DF_c (light blue shows positive density and red shows negative density) maps are shown with the completed structure highlighting the position of the FMN and ME0052 fragment. The density is shown from two viewpoints **(A)** and **(B)**.

Analysis of the ME0052 binding site reveals that the compound fragment, which is essentially an aromatic ring, sits between the isoalloxozine ring of the FMN cofactor and the indole side chain of Trp 98 from a neighbouring subunit (Figure 4-10A). The distances between the ring systems are favourable for the occurrence of π-stacking interactions (less than 5 Å). Analysis of the crystal structure of *ec*WrbA with benzoquinone in the active site (PDB ID: 3B6K [199]), showed that the haloagenated portion of ME0052 overlapped with the benzoquinone in this structure. Considering the similar chemical properties of both compounds it follows that ME0052 could be mimicking benzoquinone in its interaction with the FMN in *st*WrbA. ME0052 is held in place through hydrophobic interactions with FMN, Phe 80 and Gly 95 from within the monomer and with Trp 98 from a neighbouring subunit.

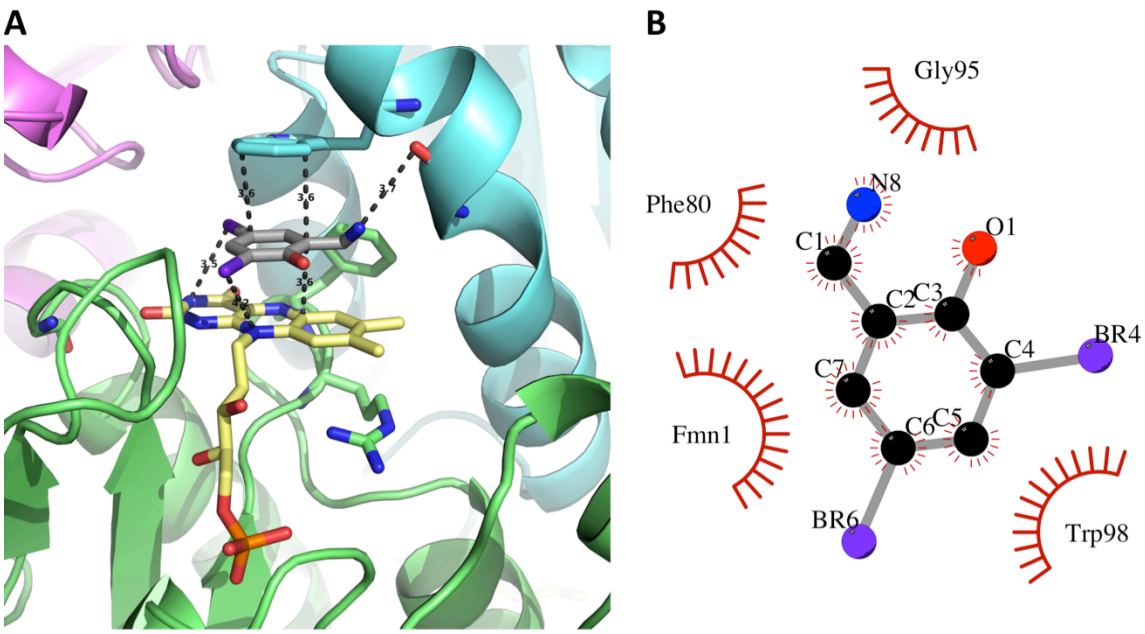


Figure 4-10 Interactions between ME0052 and FMN in the *st*WrbA active site. (A) Close-up view of the active site of one WrbA monomers; each chain is highlighted in a different colour. The FMN is shown in yellow stick representation with the fragment of ME0052 in grey. From the neighbouring subunit Trp 98 has been highlighted in stick representation. Atoms are coloured as follows, nitrogen = blue, oxygen = red, phosphate = orange, bromine = purple. The distances between the FMN, ME0052 and Trp 98 are indicated. **(B)** Ligplot+ [201] representation of the ME0052 interactions within the active site. Red lines indicate hydrophobic interactions.

4.3.3 *Enzymatic activity of WrbA and inhibition with ME0052*

To ensure that purified *st*WrbA was active, initial enzymatic characterisation of the protein was conducted following the protocol determined by Patridge *et al.* used to characterise the enzymatic activities of WrbA from *E. coli* and *A. fulgidus* [162]. The oxidation of NADH was monitored at 405 nm and benzoquinone was used as the electron acceptor. Varying concentrations of NADH were tested until the enzymatic rate did not increase, i.e. the reaction was saturated. The K_m for NADH was determined as 45 µM which is in line with the values obtained by Patridge *et al.* for WrbA from *E. coli* (14 µM) [162] (Figure 4-11).

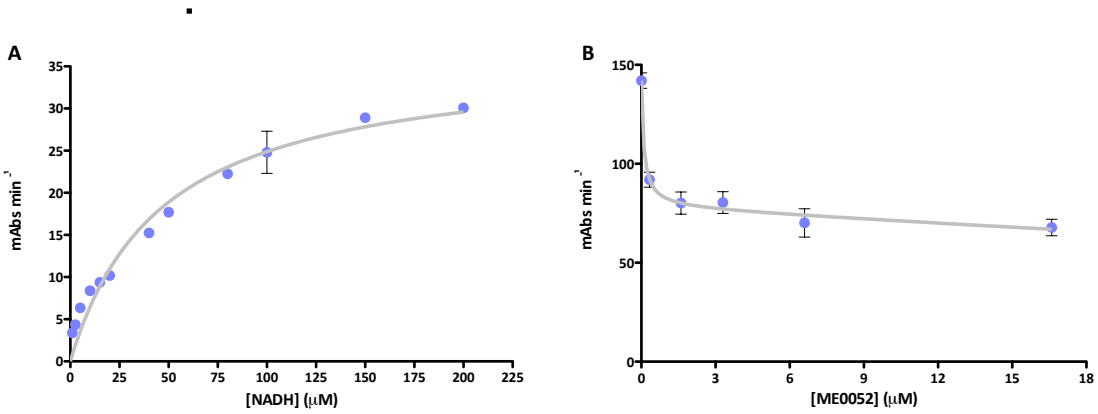


Figure 4-11 Enzyme activities of *st*WrbA. (A) Enzymatic activity of WrbA in the presence of varying concentrations of NADH. The data were fitted with a Michaelis-Menten model to determine the K_m (45 µM). **(B)** Inhibtion of *st*WrbA by ME0052. The enzyme reaction was carried out in the presence of varying concentrations of ME0052. Since complete inhibition could not be achieved, the K_i could not be calculated and the data were fitted with a one-site binding model to illustrate the decrease in enzyme activity.

Far-western analysis indicated that *st*WrbA binds to biotinylated ME0052 [100]. To confirm this, the effect of ME0052 on WrbA enzyme activity was tested. In this experiment the concentrations of both electron donor (NADH) and acceptor (benzoquinone) were in excess and the reaction was conducted in the presence of varying concentrations of ME0052. The addition of ME0052 decreased the specific activity of the enzyme by approximately 50% (Table 4-2). Complete inhibition of the enzyme could not be achieved owing to the poor solubility of the compound at high concentrations.

ME0052 (µM)	Specific activity ± sd (U mg^{-1})
0	14.2 ± 0.0
0.3	7.4 ± 0.0
1.6	6.4 ± 0.5
3.3	6.4 ± 0.4
6.6	5.6 ± 0.6
16.6	5.5 ± 0.3

Table 4-2 Specific activities of *st*WrbA in the presence of ME0052. The specific activity of the reaction was determined from the consumption of NADH (in µmol mg^{-1} = U) per mg of protein.

4.4 Discussion

FolX and WrbA were both identified in the affinity pull-down assay as putative targets of the salicylidene acylhydrazides. There are no reported links between either of these proteins and virulence gene expression and the roles of these proteins are poorly understood. However, both proteins were shown to bind to biotinylated ME0052 by far-western analysis, therefore further investigation of these proteins was conducted.

4.4.1 *FolX*

Far-western analysis of the binding of FolX to ME0052 indicated that there was some variability between species in the degree of binding [100]. FolX from *S. typhimurium*, *P. aeruginosa*, and *E. coli* were overexpressed and the cell lysate was used in the far-western. This showed that *pa*FolX appeared to bind more strongly to the biotinylated SA compound. Attempts were made to co-crystallise FolX with ME0052, however these attempts were unsuccessful. Crystals were obtained from the holo-form of the protein and since there is currently only one other structure of FolX (from *E. coli*) in the PDB it was interesting to further investigate the structure of *pa*FolX.

The crystal structures of *ec*FolX (PDB ID: 1B9L [182]) and *pa*FolX (PDB ID: 4AEY [186]) both show that FolX forms an octamer, which is formed by a dimer of tetramers. Analysis of the crystal structures using PISA indicate that the interface formed between the dimer of tetramers accounts for a relatively small area, just 7% of the SAA. This raised the question as to whether FolX is an octamer or whether this was an artefact of crystallisation. Therefore the oligomeric state of FolX was investigated using AUC and SAXS enabling the characterisation of FolX in solution. SV experiments indicated that FolX is present primarily as a single species in solution and SE experiments indicated that the main species was an octamer. The octamer is in equilibrium with a tetratmer but the K_d is in the micromolar range indicating that the interaction between tetramers is strong. This corresponds with the strength of the dimer interface predicated by PISA, which indicates that it is stabilised by 16 hydrogen bonds and 8 salt bridges. SAXS experiments carried out at concentrations where the predominant species was an octamer revealed that the solution structure of FolX is similar to the crystal structure.

From this investigation no further insight into whether FolX is a true target of the SA compounds was gained. Furthermore, transcriptomic analysis of the Δ*folX* knockout mutant did not show a decrease in LEE T3SS gene expression with an increase in flagellar gene expression [100]. From this it was concluded that FolX was not the main target of the SA compounds and that its enzymatic activity was not responsible for the desired phenotype.

4.4.2 *WrbA*

Far-western analysis of the binding of biotinylated ME0052 to WrbA from *Y. pseudotuberculosis*, *Shigella flexneri*, *S. typhimurium*, *P. aeruginosa*, and *E. coli* showed no difference in the binding between species. Therefore, proteins from all species were put into crystallisation screens as slight differences in protein sequence may result in one variant being more amenable to crystallisation. This proved to be the case since good quality, diffracting crystals were obtained only with *st*WrbA. Crystallisation was also undertaken in the presence of 200 µM ME0052, wherein good quality crystals were also obtained.

Electron density corresponding to the halogenated fragment of ME0052 could be seen in the crystal structure. The compound was positioned between the isoalloxazine rings of the FMN cofactor and the indole ring of a tryptophan side chain from a neighbouring subunit. Since only half of ME0052 can be seen in the density, it appears that the compound has been cleaved, perhaps through acid hydrolysis of the imine bond. Further evidence is required to fully validate that the density corresponds to ME0052. This could be acquired from co-crystallisation of the compound with a range of concentrations of ME0052, which should therefore alter the occupancy of the ligand in the binding site. In the current structure it would appear that the compound is not present at full occupancy, so increasing the concentration of ME0052 may improve this. In the future if any co-crystals were taken to a synchrotron for data collection, an X-ray Absorption Near Edge Structure (XANES) scan could be performed. This involves irradiating the crystal with a wavelength that causes bromine atoms to fluoresce, which is then detected. Detecting fluorescence would be a strong indication that bromine is present in the crystal.

The inhibition of WrbA enzyme activity further supports the hypothesis that WrbA does bind to ME0052. The specific activity of WrbA was decreased by approximately 50% by ME0052. ME0052 is poorly soluble at high concentrations and absorbs at the wavelength used to monitor the reaction. Therefore the reaction could not be tested at concentrations high enough for the reaction to be completely inhibited. Further investigations into how the SA compounds could be modified to increase their affinity of binding to WrbA are being conducted by collaborators at the University of California (Amaro group).

Overall, the biochemical data support the hypothesis that WrbA binds to the salicylidene acylhydrazides. However, the deletion of *wrbA* did not replicate the phenotype seen following treatment of whole cells with the SA compounds. Phenotypic studies of the mutant carried out by Dr Dai Wang showed that, as with Tpx, an opposite pattern of gene expression was seen, i.e. an increase in LEE T3S and a decrease in flagellar gene expression. The transcriptomics data showed significant changes in the expression of genes involved in T2S. This may give some insight into the role of WrbA and requires further investigation.

4.4.3 *Conclusion*

Both FolX and WrbA were identified as putative targets of the SA compounds. However, deletion of either FolX or WrbA does not decrease the expression of the LEE T3SS, therefore it was determined that neither of these proteins are the key target of the SA compounds.

5 Are the PPIases the protein targets of the salicylidene acylhydrazides?

5.1 Introduction

The SA compound affinity pull-down assay identified three proteins belonging to the same family, known as the peptidyl-prolyl *cis-trans* isomerases (PPIases) [100]. Specifically, the putative target proteins identified were SurA, FkpA and FklB. Whilst SurA and FkpA have been studied in some detail and have been linked to virulence in several reports [202–205], relatively little is known about the role of FklB.

5.2 Peptidyl-prolyl *cis-trans* isomerases

The peptide bond between amino acids can be connected in one of two conformations during biosynthesis at the ribosome, either *trans* or *cis*. Protein folding experiments revealed that the formation of the *trans*-form of the peptide bond is thermodynamically favourable for most amino acids and is therefore the most common [206]. However, the conversion of the petidyl-prolyl bond from the *cis* to *trans* conformation is less thermodynamically favourable, therefore this bond can exist in the *cis* or *trans* conformation [206]. The presence of the *cis* peptide bond can have a significant effect on the overall structure of proteins, therefore in most cases the *cis* bond needs to be converted to a *trans* bond to produce a functional protein. Naturally this conversion occurs very slowly and therefore imposes a rate-limiting step on protein folding [207,208]. To overcome this thermodynamic barrier, cells express a variety of enzymes that catalyse this reaction. The PPIases are a ubiquitous and highly conserved superfamily of proteins that catalyse the slow, rate-limiting *cis* to *trans* isomerisation of peptidyl-prolyl bonds shown in Figure 5-1 [206]. Many proteins require PPIases for efficient folding thus implicating these proteins in several essential cellular processes.

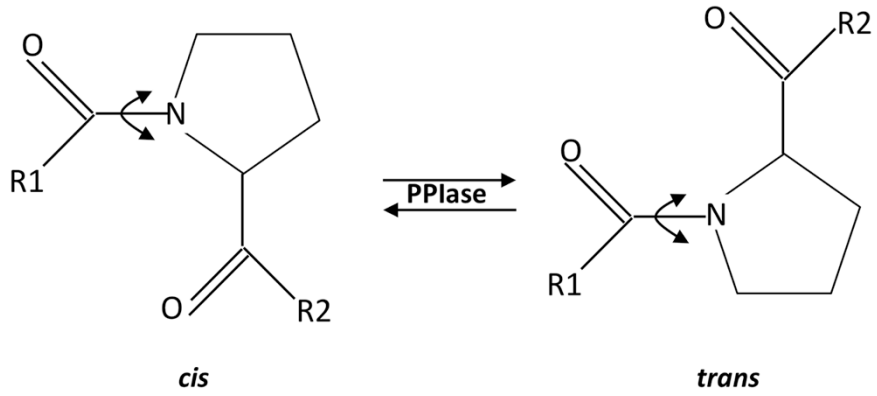


Figure 5-1 *Cis-trans* isomerisation of a peptidyl-prolyl bond. Figure modified from [209].

To date six PPIases have been described in *E. coli*: SurA, PpiD, PpiA, Par10, FkpA and FklB [210,211]. There are three main classes of PPIase, which are unrelated in amino acid sequence and classified on the basis of their binding partners: (i) FK506-binding proteins (FKBPs) bind to the compounds FK506 and rapamycin; (ii) cyclophilins bind to cyclosporine A; and (iii) parvulins bind to juglone [209]. Since FkpA and FklB belong to the FKBPs and SurA belongs to the parvulin family of PPIases, only these two classes of PPIase will be discussed further here.

5.2.1 *Monitoring PPIase activity*

PPIase activity can be measured using several different methods including spectrophotometric assays and NMR to determine the rate of *cis* to *trans* isomerisation. Here only the protease-coupled spectrophotometric assay will be discussed (Figure 5-2). This assay relies on the fact that most proteases specifically cleave protein sequences where all residues are in the *trans* conformation. However, in approximately 10% of proline-containing protein sequences the proline is in the *cis* conformation and therefore is not cleaved by the protease (reaction 2 in Figure 5-2). Consequently, before it can cleave, a PPIase must first catalyse the isomerisation of the *cis* bond (reaction 1 in Figure 5-2). To measure the conversion of a *cis* to a *trans* bond, a tetrapeptide substrate, usually with the sequence Succinyl-Ala-Xaa-Pro-Phe-para-nitroanilide (Xaa is any amino acid) is used. Cleavage of the nitroanilide bond releases the chromophore para-nitroanilide, the absorbance of which can be monitored at 390 nm (reaction 3 in Figure 5-2). The protease α-chymotrypsin is generally used in this assay. Since 90% of the substrate has proline in the *trans* confirmation the initial spectrophotometric signal is from cleavage of the trans substrate independent of the PPIase activity. This rapid phase is then followed by the slow kinetic phase of the reaction where the rate of chromophore release is a direct consequence of PPIase activity [209].

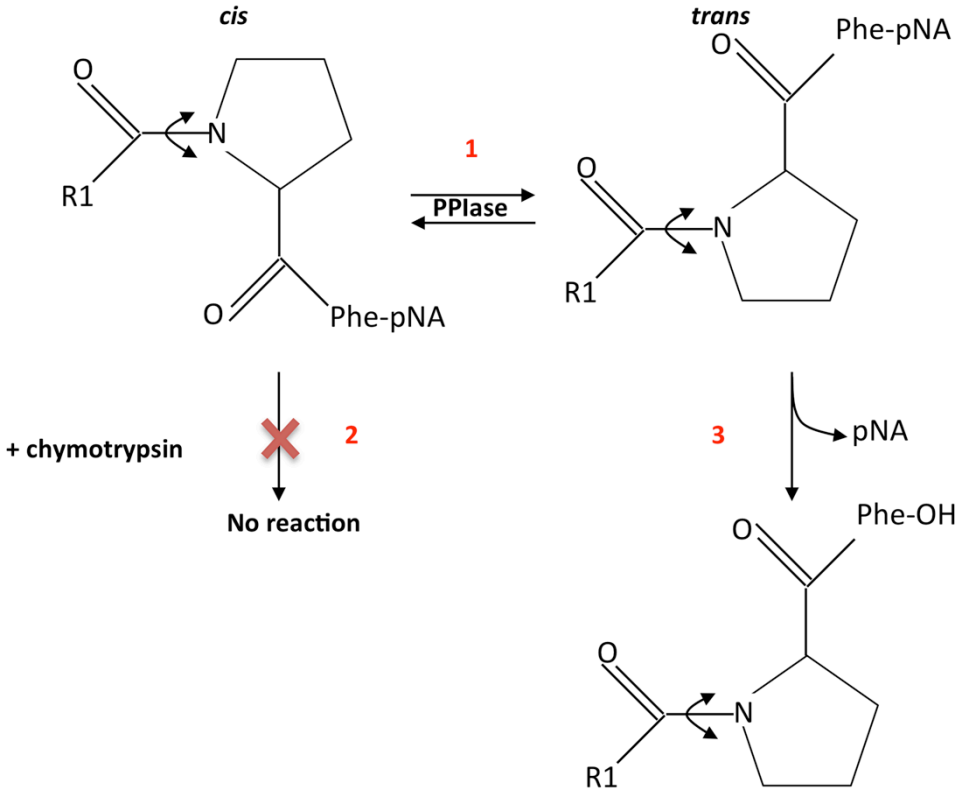


Figure 5-2 Monitoring peptidyl-prolyl *cis*/*trans* isomerisation using isomer-specific proteolysis. (1) PPIase activity isomerises the peptidyl-prolyl bond from the *cis* to the *trans* conformation, any substrate that is in the *cis* conformation cannot be cleaved by α-chymotrypsin and para-nitroanilide (pNA) is not released (2). (3) Only when the substrate contains the proline in the *trans* conformation, following isomerisation, can α-chymotrypsin cleave pNA from the peptide. The cleavage of the chromophore can be monitored at 390 nm. Figure modified from [209].

5.3 The parvulins

Parvulins are found ubiquitously throughout all kingdoms of life. They can vary in size from 10 kDa up to 68 kDa [212]. There are three proteins in *E. coli* that belong to the parvulin family, these include Par10, PPiD and SurA [211]. Par10 is the smallest known parvulin with only 92 residues (10 kDa) and represents the smallest unit required for the *cis*/*trans* isomerisation of the peptidyl-prolyl bond [212]. The structure of Par10 shows the archetypal parvulin fold of four α-helices and four β-sheets. This domain is often duplicated in larger parvulins, such as SurA [212].

5.3.1 *SurA*

SurA, or Survival protein A, was first identified as being essential for survival during the stationary growth phase in *E. coli* [213]. SurA was confirmed to have PPIase activity [214] and sequence analysis revealed that it belonged to the parvulin class of PPIases [211]. SurA has been shown to have a specific and essential role in the folding of outer-membrane proteins (OMPs) [205,215]. Although SurA is a PPIase, its enzymatic function is thought to be secondary to its role as a protein chaperone. SurA consists of an N-terminal region, two parvulin domains, only one of which has PPIase activity (domain 2), followed by a C-terminal region (Figure 5-3A and B) [214]. The second parvulin domain has no PPIase activity but instead has the role of binding folding intermediates with exposed hydrophobic residues [216]. The chaperone activity has been shown to be mediated primarily by the N- and C-terminal regions of SurA [217].

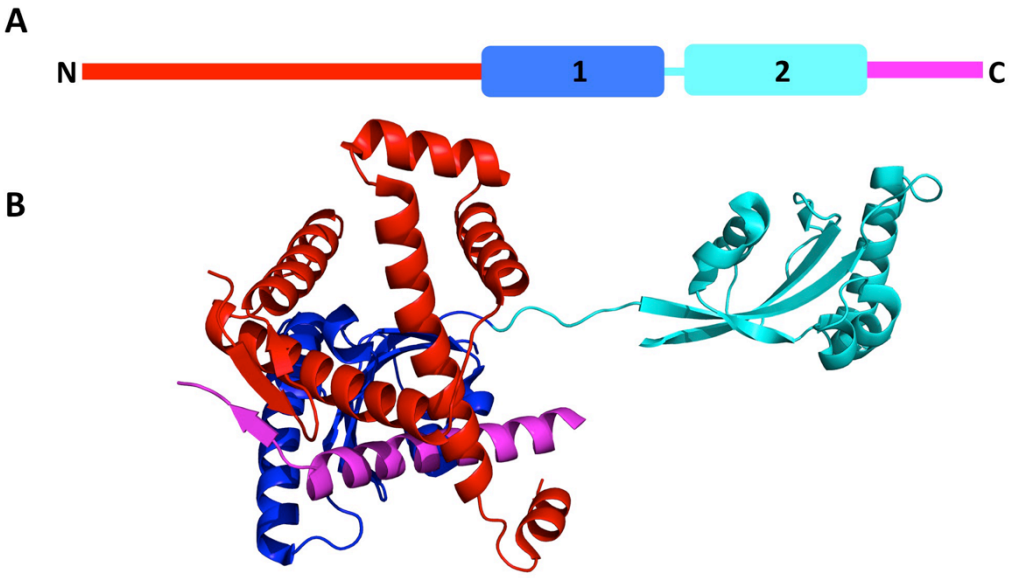


Figure 5-3 Structure of SurA. (A) Schematic diagram showing the modular arrangement of SurA. The N- and C-terminal regions of SurA, which exhibit the chaperone activity, are coloured red and magenta, respectively. The parvulin-like PPIase domains 1 and 2 are coloured blue and cyan, respectively. **(B)** The crystal structure of SurA (PDB ID: 1M5Y [137]) has been coloured using the colour scheme in (A) to indicate the position of its domains.

The chaperone activity of SurA is essential for the transport of OMPs across the periplasm to the BAM complex [205,218]. Therefore deletion of *surA* has a significant effect on the composition of the outer-membrane, which in turn makes the mutant hypersensitive to

stresses such as H_2O_2, bile salts and to some antibiotics [214]. Additionally deletion of *surA* has an effect on the virulence of several pathogens. For example in UPEC, deletion of *surA* led to reduced binding and invasion of bladder epithelial cells as well as a decreased ability to grow intracellularly [219].

Deletion of *surA* has dramatic effects on the assembly of proteins in the outer-membrane. This leads to constitutive expression of the envelope stress (σ^E) response [220], which in turn leads to a change in the transcription of a subset of different genes, with the aim of reducing the level of folding intermediates present in the periplasm [221]. The effects of virulence are therefore likely to be pleotropic and not solely a result of the loss of this periplasmic chaperone, but also owing to more general changes induced by this stress response.

5.3.2 *The FKBPs*

FKBPs are characterised by their ability to bind to FK506, a compound that was isolated from *Streptomyces tsukubasesis*, and shown to have immunosuppressant activity [222]. The first FKBP to be characterised was a 12 kDa protein (FKBP12) that was purified from human T-cells. The enzymatic activity of FKBP12 was shown to be inhibited by both FK506 and rapamycin [223]. This was the first FKBP to be crystallised and analysis of the structure revealed a fold consisting of five β-sheets and a short α-helix. FKBP12 was later crystallised in complex with its inhibitors which revealed the interacting residues in the active site of the protein [224]. The overall fold of the active site and the catalytic residues have been shown to be conserved throughout the FKBP protein family. FKBP12 is the smallest FKBP, with only one PPIase domain; other larger FKBPs have gained additional domains [225]. One of the largest FKBPs identified so far is FKBP77 from *Tritus aestivim*, which encodes three PPIase domains, a tetracopeptide repeat domain which acts as a docking motif for other binding chaperones and a calmodulin binding domain [226]. Acquisition of different functional domains has produced a class of enzymes with a diverse set of substrates, which have been implicated in several essential cellular processes such as protein folding, protein trafficking and protein assembly.

5.3.3 *FkpA*

FkpA is a relatively small FKBP with a molecular weight of 29 kDa. FkpA encodes a C-terminal FKBP PPIase domain and an N-terminal dimerisation domain [227]. This protein has been shown to exist exclusively as a dimer [228]. FkpA has both PPIase and chaperone activity and both of these activities are localised to the C-terminal FKBP domain [228]. However, inhibition of the active site residues with FK506 does not affect chaperone activity, indicating that the two activities originate from different sites within the FKBP domain [228].

FkpA localises to the periplasm and the pro-peptide includes a periplasmic signal sequence [221]. FkpA expression was shown to be mediated by σ^E, indicating its role in the envelope stress response [221]. Deletion of FkpA does not affect cell survival during growth under normal conditions, indicating the high level of functional redundancy amongst the PPIase proteins [227]. However, Horne *et al.* (1997) reported that a *S. typhimurium* Δ*fkpA* mutant showed decreased intracellular survival, suggesting that FkpA may be required during pathogenesis, a time when the bacterium is exposed to several extracellular stresses, which can increase the levels of protein mis-folding [229]. Recently FkpA has been shown to be required for the activation of colicin M, a toxic effector protein secreted by *E. coli* carrying the pColMB plasmid. Colicin M kills sensitive strains of *E. coli* but only following activation by FkpA in the periplasm, therefore Δ*fkpA* mutants are insensitive to colicin M [202,230].

5.3.4 *FklB*

FklB a small (22 kDa) FKBP also known as FKBP22, was first identified and characterised in *E. coli.* However FklB homologues are found in several Gram-negative species [210]. In this initial study, FklB was shown to have PPIase activity and to be inhibitable by FK506 verifying its classification as an FKBP [210]. Further studies have also shown that FklB is strongly inhibited by rapamycin with an IC_{50} of 51 nM for the purified protein from *Y. pseudotuberculosis* [204]. The localisation of this protein has not been categorically determined. It was originally shown to be present in the periplasm [210] along with other PPIases such as FkpA and SurA; however, unlike FkpA and SurA, FklB does not encode a periplasmic signal sequence. Furthermore, cell fractionation experiments report that FklB

is largely associated with the inner membrane fraction [204]. Also it has been shown that FklB interacts with inner-membrane associated thioredoxin, which is further evidence that it may be associated with the inner membrane [231].

The physiological importance of FklB is unknown. Studies of FklB from *Shewanella* indicate that the PPIase activity is highest at 10°C therefore Suzuki and colleagues proposed that this PPIase may be of particular importance during growth at cooler temperatures [232]. Deletion of *fklB* along with other PPIases such as *fkpA, ppiA,* and *ppiD* in a quadruple knockout in *Y. pseudotuberculosis* had little effect under normal growth conditions [204]. However, during stresses such as oxidative stress or in an *in vivo* model of infection, the quadruple mutant showed decreased survival compared with the WT, although the effects were not as dramatic as the effect of *surA* deletion, indicating that SurA is of critical importance to this organism [204].

Studies of the FklB structure have revealed that this protein, like other PPIases, is present exclusively as a dimer [210,233]. Dimerisation is stabilised primarily through hydrophobic interactions in the N-terminal domain, which were elucidated in the crystal structure of the N-terminal dimerisation domain of the *Shewanella* FklB [234]. Analysis of the N and C-terminal domains of FklB has shown that the PPIase activity is found exclusively in the C-terminal domain and that the chaperone activity is found in the N-terminal domain [233]. However the chaperone activity is dependent on the dimerisation of the protein, which is thought to be linked to the formation of the 'V' shape characteristic of the PPIases with chaperone activity [235]. There is currently no crystal structure for full-length FklB or for its C-terminal domain.

5.4 Do FkpA and SurA bind the SA compounds?

5.4.1 *Cloning, expression and purification of SurA and FkpA*

The *surA* gene was amplified by PCR using the primers listed in Table 2-5. The PCR product of the *surA* gene plus restriction sites was cloned into the Strataclone vector (Invitrogen). SurA was cloned into pET-151, p77 and pET28a however over-expression of this protein from these vectors was not successful. The over-expression plasmid used to

produce the SurA protein (TYB1-SurA) from *E. coli* K12 for crystallographic studies by Bitto and McKay [137] was obtained and used for SurA expression in the BL21(DE3) strain. SurA from *E. coli* K12 and O157 are 100% identical. The SurA over-expression plasmid uses the IMPACT expression system (NEB), which utilises a chitin affinity tag. Following protein binding to chitin beads it is eluted using DTT, which cleaves the tag releasing the purified protein [34]. Expression and purification of this protein from this vector was successful and the purity of the eluted protein, corresponding to the band seen at ~ 49 kDa, was sufficient for this investigation (Figure 5-4A).

FkpA was cloned in a manner similar to SurA into the p77 expression vector. FkpA could be purified using the vector in the BL21(DE3) strain (Figure 5-4B). A clear band at approximately 28 kDa, corresponding to size of FkpA, can be seen in the wash fraction, indicating that FkpA did not bind well to the IMAC resin. Purified protein was eluted at higher imidazole concentrations (1 M) and corresponds to a band at ~ 28 kDa on the gel (not clearly visible but indicated in Figure 5-4B). This protein band was identified using mass spectrometry and trypsin digest and FkpA was listed as the hit with the highest MOWSE score. Even following further purification attempts using anionic exchange, the purity of FkpA did not improve significantly. However, the level of purity obtained was sufficient for the studies conducted in this investigation.

5.5 Far-western analysis of the binding of the PPIases to ME0052

To determine whether the PPIases FkpA, SurA and FklB (the purification of which is described below in 5.8.1) were true targets of the SA compounds far-western blotting was performed. One hundred micromoles of the purified PPIase proteins were separated by SDS-PAGE and transferred onto a nitrocellulose membrane. Following the 're-folding' step of the far-western protocol the membrane was incubated with 20 µM biotinylated ME0052. Following several wash steps the presence of biotinylated ME0052 bound to the membrane was detected using a streptavidin-HRP conjugated antibody. The presence of a band on the far-western corresponds only to the lane containing FklB (Figure 5-5A). This blot was replicated and still no binding to FkpA or SurA was detected. An SDS-PAGE gel showing the purity of the protein used in the far-western is shown in Figure 5-5B. Thus the binding of ME0052 to FkpA and SurA could not be confirmed by far-western.

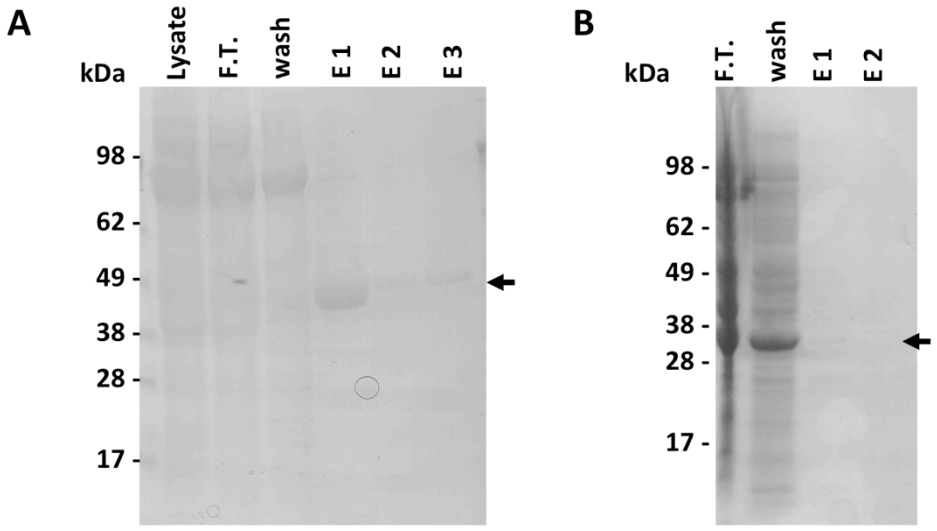


Figure 5-4 Purification of SurA and FkpA. (A) SDS-PAGE gel showing the different stages of purification of SurA expressed from the plasmid used in the crystallographic studies of SurA by Bitto and McKay [137]. The cell lysate prior to purification, the flow through (F.T.) after loading the lysate onto the chitin beads, the wash step to remove non-specifically bound protein and the elution fractions (E) are shown. **(B)** SDS-PAGE gel showing the different purification stages of FkpA. The F.T., 20 mM imidazole wash step and the elution fractions have been shown. The position of the molecular weight markers have been shown (kDa) and the arrow highlights the band corresponding to purified protein.

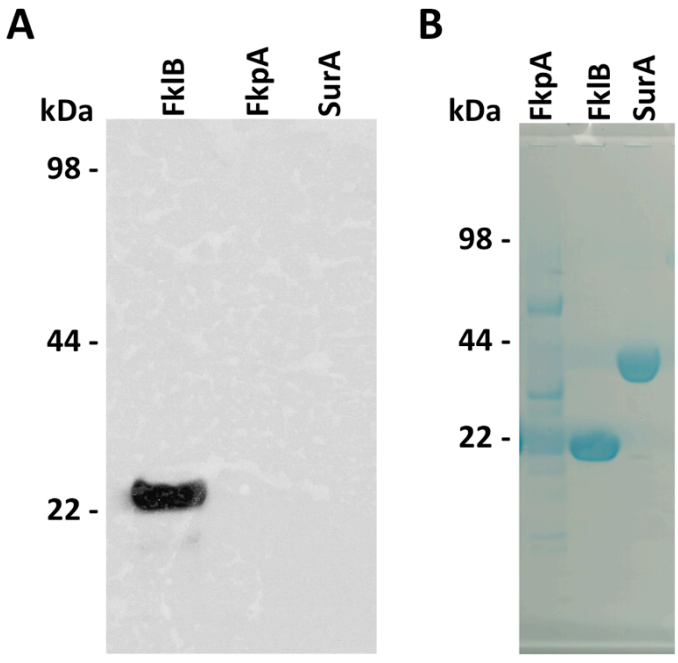


Figure 5-5 Far-western analysis of the binding of ME0052 to the PPIase target proteins. (A) Far-western for purified FklB (*E. coli* O157), FkpA (*E. coli* O157), and SurA (*E. coli* K12) probed with biotinylated ME0052. The corresponding SDS-PAGE gel is shown in **(B)** (the order in which the proteins were loaded differs). The positions of the molecular weight markers are shown (kDa).

5.6 PPIase activity of SurA and FkpA

The enzymatic activities of the PPIases were determined at the Defence Science and Technology Laboratory (DSTL), Porton Down with the help of Dr Isabelle Norville and analysis of the kinetic data was carried out in collaboration with Dr Nic Harmer, University of Exeter. To determine the PPIase activity of SurA from *E. coli* O157 the rate of *cis* to *trans* isomerisation was monitored using a protease coupled PPIase assay (Figure 5-2). This assay confirmed that the purified protein retained its enzymatic activity as the rate of *cis* to *trans* isomerisation was increased in the presence of SurA (Figure 5-6A). The pseudo-first order rate constant was calculated from the rate of the initial 10-50 s of the time-course curve, obtained from reactions performed in the presence of SurA at a range of concentrations (0.1, 0.5, 1 and 2 µM) (Figure 5-6B). From this, the specificity constant (k_{cat}/K_m) was calculated to be 7.76 x 10^3 M^{-1} s^{-1}.

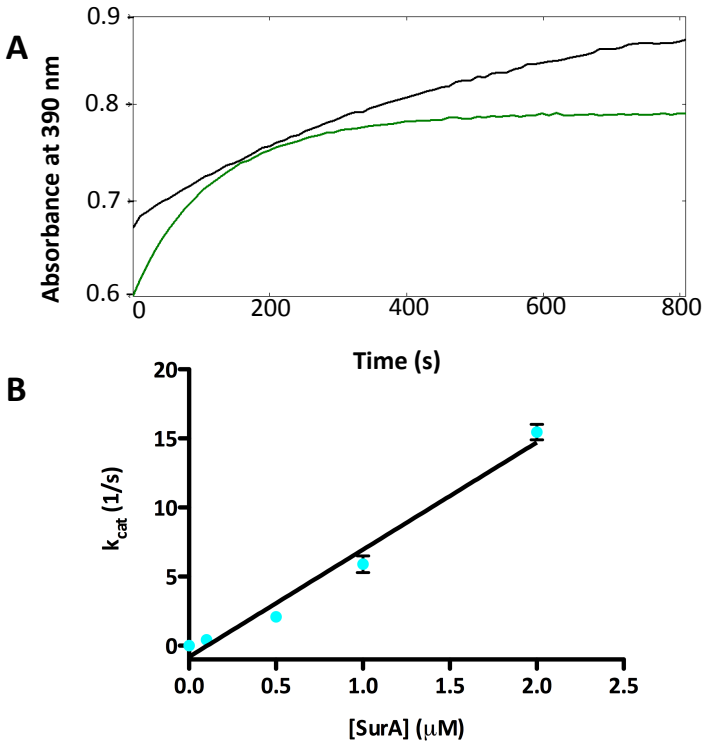


Figure 5-6 First order rate constant of SurA PPIase activity. (A) An example time course curve showing the reaction catalysed by SurA (green line) compared with the uncatalysed reaction (black line). **(B)** The rate constant (k) of the catalysed first-order *cis-trans* conversion upon addition of 0-2 µM of purified SurA. The calculated specificity constant k_{cat}/K_m was 7.76 x 10^3 M^{-1} s^{-1}. The values are means (± standard deviation) from at least three independent experiments.

As for SurA, the PPIase activity of FkpA was also investigated. Purified FkpA retained its PPIase activity (Figure 5-7A). A range of protein concentrations similar to that of SurA was used to determine the catalytic rate of FkpA (Figure 5-7B). The specificity constant of FkpA was calculated to be 9.23 x 10^3 M^{-1} s^{-1}. A higher specificity constant indicates higher enzyme efficiency suggesting that the PPIase activity of FkpA is higher than that of SurA.

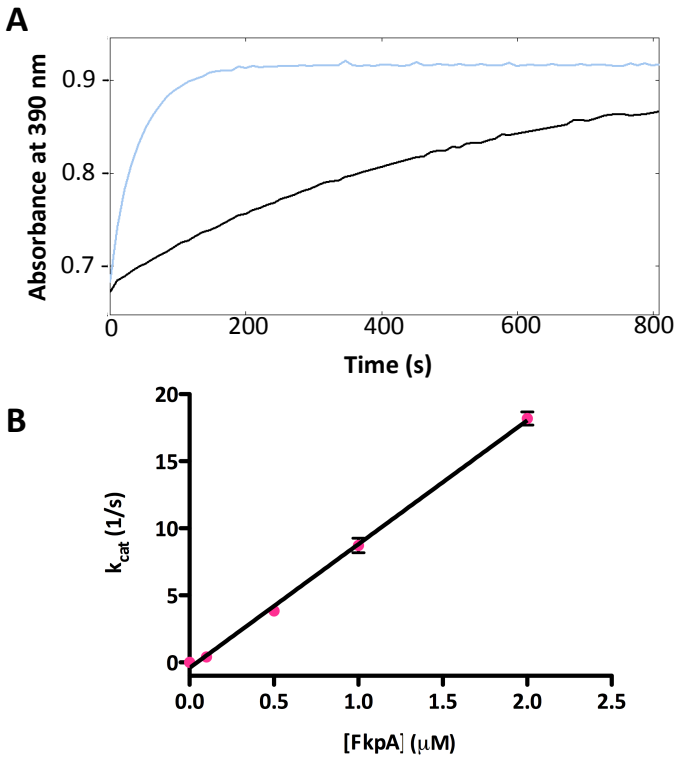


Figure 5-7 Determination of the first order rate constant of the PPIase activity of FkpA. (A) An example time course curve showing the reaction catalysed by the activity of FkpA (blue line) compared with the uncatalysed reaction (black line). **(B)** The rate constant (k) of the catalysed first-order *cis-trans* conversion upon addition of 0-2 µM of purified FkpA. The calculated specificity constant k_{cat}/K_m was 9.23 x 10^3 M^{-1} s^{-1}. The values are means (± standard deviation) from at least three independent experiments.

5.7 Deletion of *surA* in *E. coli* O157

A Δ*surA* deletion mutant was generated in *E. coli* O157:H7 by Dr Dai Wang using allelic exchange, however the phenotype of this mutant was investigated as part of this thesis. To test the effects of *surA* deletion the growth rate of this strain was investigated. The growth rate in both rich (LB) and minimal (MEM-HEPES) media was determined and the doubling rate of WT and Δ*surA* was calculated from triplicate experiments. When grown in LB the doubling rate of WT was 37 ± 3 min, and that of the mutant was 36 ± 2 min, indicating that deletion of *surA* has no fitness cost under these conditions (Figure 5-8A).

In minimal media the doubling rate appeared to be slightly slower for the Δ*surA* mutant when compared with the WT with a rate of 185 min *versus* 168 min, respectively (Figure 5-8B).

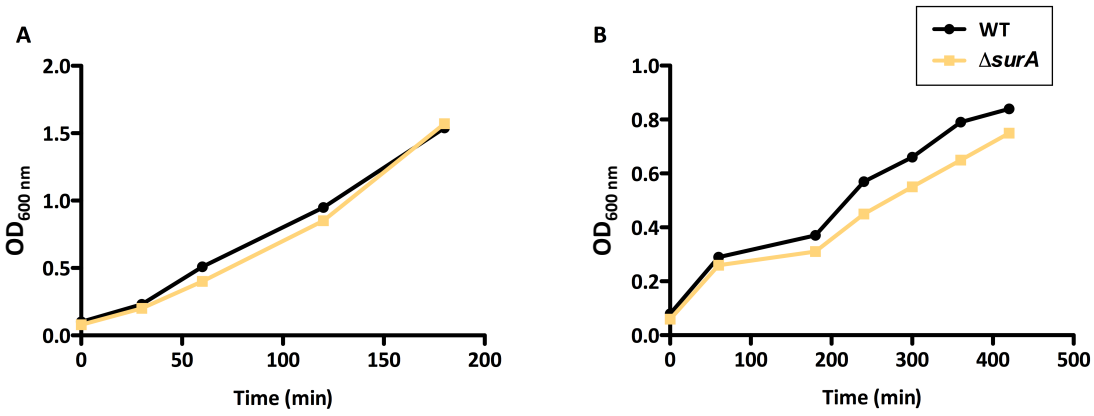


Figure 5-8 Growth rate of WT and the Δ*surA* mutant. (A) The growth of WT (black) and the Δ*surA* mutant (yellow) in LB medium. **(B)** The growth of WT and Δ*surA* in MEM-HEPES.

5.7.1 *ΔsurA is non-motile*

To determine the effects of *surA* deletion on swimming motility, a motility assay was conducted. WT and Δ*surA* were inoculated into the centre of a soft agar plate (0.25% agar, w/v), and incubated overnight at 30°C. This revealed that the Δ*surA* mutant was non-motile, as the bacteria did not migrate in the agar, unlike the WT, which migrated out to a diameter of 4 cm (Figure 5-9).

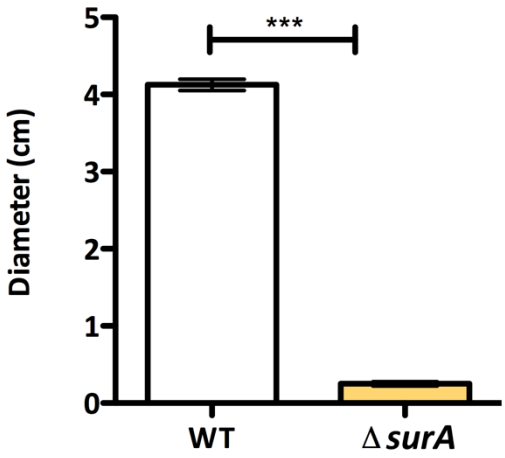


Figure 5-9 Motility assay of the WT and Δ*surA* mutant. Swimming motility of the WT and Δ*surA* mutant was tested in soft agar. The swimming diameter of three triplicate experiments was plotted. (*** indicates a p-value of <0.0001 calculated from a Student's t-test).

5.7.2 *The effect of surA deletion on the expression of the LEE T3SS and flagella*

The expression of the LEE T3SS was monitored using GFP reporter plasmids, where the expression of GFP is under the control of the natural promoters of *LEE1*, *LEE2*, *LEE3*, *tir*, or *rpsM*. To induce LEE T3S, strains were cultured in MEM-HEPES media and the GFP fluorescence was monitored at different time points. Figure 5-10 shows representative graphs for each of the reporter fusions. Overall the level of expression LEE T3SS appeared to be increased in the Δ*surA* mutant. Expression of *LEE1* was increased by 1.7-fold in the Δ*surA* mutant compared with the WT and a Student's t-test revealed that the difference is significant with a p-value of 0.04 (Figure 5-10A & H). There was a 1.5-fold increase in *LEE2* expression in the Δ*surA* mutant representing a significant change in expression levels compared with the WT ($p = 0.04$) (Figure 5-10B & H). The 1.3-fold increase in the expression of *LEE3* seen in the Δ*surA* mutant was similarly significant with a p-value of 0.05 (Figure 5-10C & H). The fold-change in *tir* expression was considerably larger: a 3-fold increase in expression the Δ*surA* mutant ($p = 0.02$) (Figure 5-10D & H). The level of expression of the control gene *rpsM* was not affected in the Δ*surA* mutant (Figure 5-10E & H). The level of FliC expression was monitored using a *fliC* promoter:GFP reporter fusion. The expression of FliC was investigated during growth in MEM-HEPES and LB. Usually growth in MEM-HEPES inhibits flagellar expression, which is clearly demonstrated for the WT as the level of fluorescence remains low throughout; however for the Δ*surA* mutant there was an almost 3-fold increase in the level of FliC expression compared with the WT ($p= 0.01$) (Figure 5-10F & H). The level of *fliC* expression in LB medium was similar for the WT and mutant strain, indicating that the increase in flagellar expression may be specific to growth in minimal medium (Figure 5-10G & H).

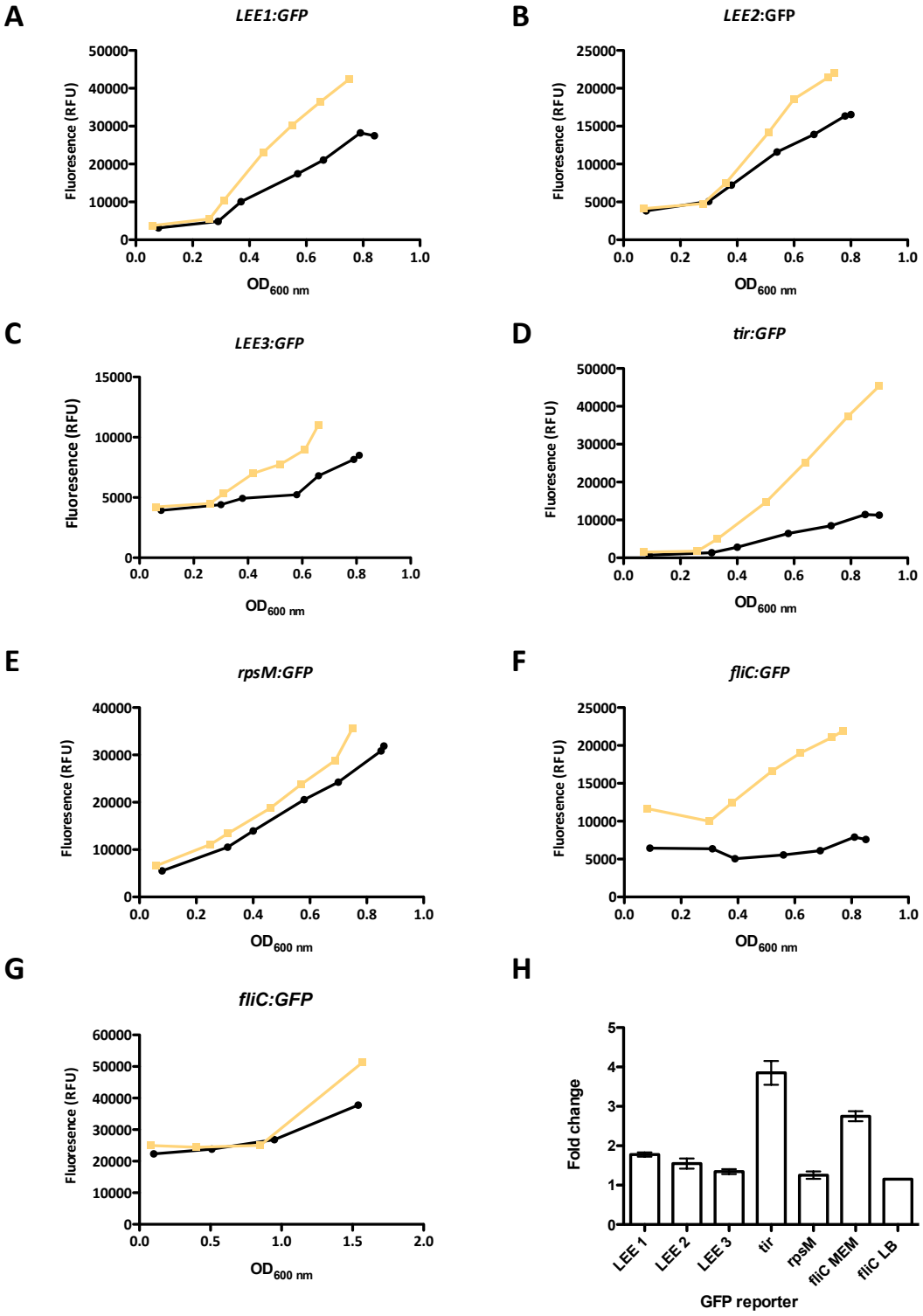


Figure 5-10 Expression of gene:GFP reporter fusions in Δ*surA*. Gene expression in the WT is shown in black and gene expression in the Δ*surA* mutant is shown in yellow. Line graphs show the relative fluorescence plotted against the optical density at 600 nm ($OD_{600\ nm}$) for *LEE1* **(A)**, *LEE*2 **(B)**, *LEE3* **(C)**, t*ir* **(D)**, *rpsM* **(E)** and *fliC* **(F)** grown in MEM-HEPES media and in LB media **(G)**. The data shown are from one growth experiment but the graphs are representative of triplicate experiments. The fold-change of gene expression observed at $OD_{600\ nm}$ = 0.7 is shown in a bar chart **(H)** where the error bars are derived from triplicate experiments.

Expression of the LEE T3SS was further analysed by a secretion assay. WT and Δ*surA* strains were cultured in MEM-HEPES to an $OD_{600\,nm}$ of 0.8. The secreted proteins were then precipitated and analysed by SDS-PAGE. This revealed that the Δ*surA* mutant appeared to secrete more LEE T3S proteins than the WT. The identity of two proteins that appeared significantly increased in the Δ*surA* mutant was determined using tandem mass spectrometry. The identity of protein 1, indicated in Figure 5-11, was determined to be flagellin and protein 2 was identified as EspA with a MOWSE score of 450 [236].

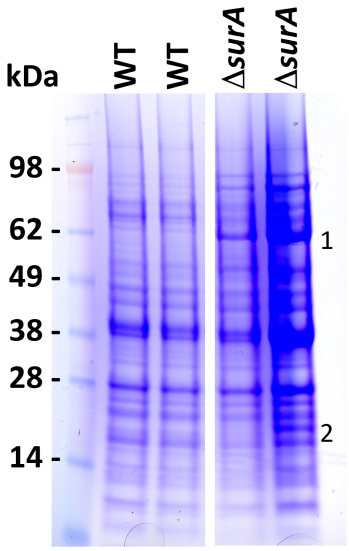


Figure 5-11 Secretion profile of the WT and the Δ*surA* mutant. SDS-PAGE gel showing duplicate experiments for the WT and the mutant. Numbers 1 and 2 correspond to bands that were identified by mass-spectrometry as flagellin and EspA, respectively. The positions of the molecular size makers have been indicated (kDa).

5.7.3 *Attaching and effacing lesion formation*

The ability of the Δ*surA* mutant to form attaching and effacing (A/E) lesions was determined using a cell-adhesion assay. Expression of the LEE T3SS was induced by culturing the cells in MEM to an $OD_{600\,nm}$ of 0.6. At this point bacteria were added to Caco2 epithelial cells with a multiplicity of infection of 10. Bacteria were initially attached to the cells by centrifugation and then unbound bacteria were removed by washing. Cells were fixed and stained 4 h after initial attachment. The formation of A/E lesions can be identified by the condensation of actin around the attached bacteria. From this

experiment the number of WT and mutant cells was equivalent (Figure 5-12A). However the level of actin condensation was significantly reduced in the Δ*surA* mutant (Figure 5-12B) indicating that A/E lesion in this mutant is either delayed or compromised.

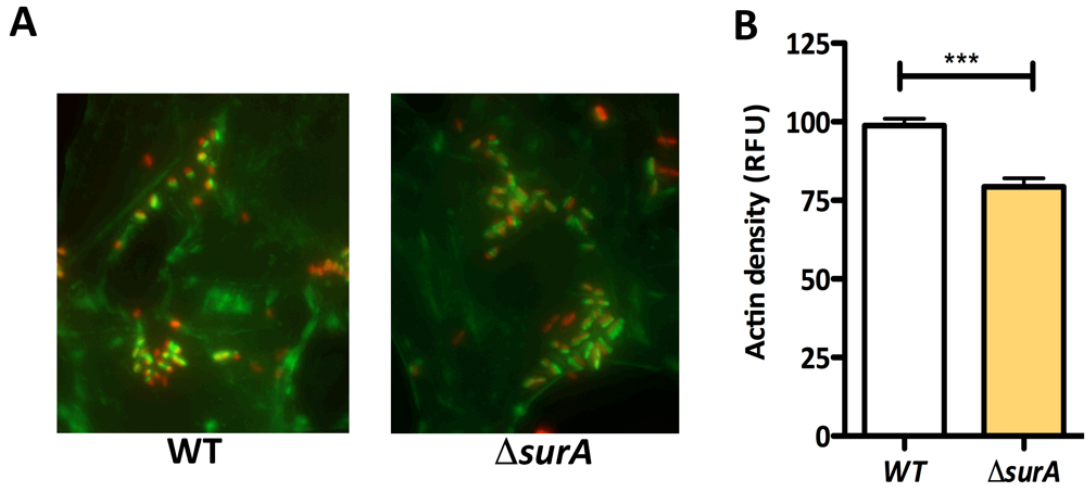


Figure 5-12 Attaching and effacing lesion formation in the Δ*surA* mutant. (A) Immunofluorescence images showing WT and Δ*surA* bacteria expressing RFP from a plasmid (red), the actin of the Caco2 epithelial cells is stained with FITC phalloidin (green). Where the actin has condensed to form an A/E lesion it appears as an area of increased fluorescence. The level of actin condensation has been quantified for the WT and mutant strains **(B)**. The RFU values for the two groups were compared using a Student's t-test and the level of significance is denoted with asterisks (*** denotes $p<0.0001$).

5.8 Structural and functional characterisation of FklB

5.8.1 *Protein overexpression and purification*

The *fklB* gene sequence was amplified from genomic DNA by PCR using the primers listed in Table 2-5 and the PCR product was cloned into the Strataclone vector as in Section 5.4.1. The gene was subcloned from the Strataclone vector by restriction digestion and ligation into the pET28a expression vector. The C-terminal PPIase domain of FklB (FklBc) was similarly cloned into pET-28a. pET28-FklB and pET28-FklBc were then individually transformed into the BL21(DE3) expression strain and 2 L of culture grown in LB was induced with 1 mM IPTG overnight. FklB expressed well and was purified to homogeneity following IMAC and SEC. This can be visualised on the SDS-PAGE gel as a band corresponding to ~ 22 kDa (Figure 5-13A). The C-terminal domain of FklB (FklBc) also purified to homogeneity following IMAC and SEC and a band can be seen on the SDS-

PAGE gel at ~ 14 kDa which corresponds to the mass of the protein. Both FklB and FklBc were put into a variety of commercial crystallisation screens, however no protein crystals were obtained.

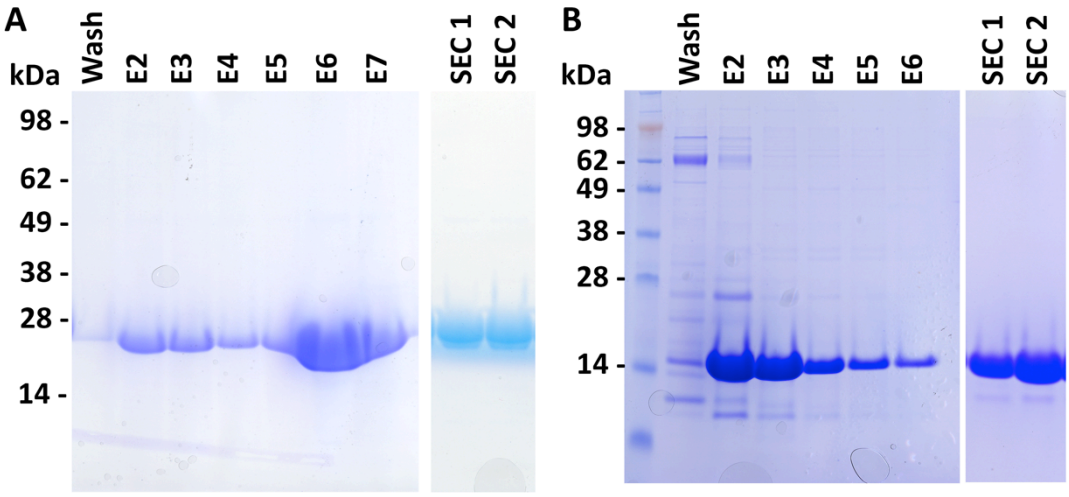


Figure 5-13 Purification of FklB and FklBc. SDS-PAGE gel showing the purification steps for FklB **(A)** and FklBc **(B)**. The 20 mM imidazole wash step, the fractions eluted with 350 mM imidazole (E) and the fractions of the major peak following size exclusion chromatography (SEC) have been shown. The positions of the molecular weight markers have been indicated (kDa).

5.8.2 *Homology modelling of FklB*

There is currently no crystal structure of full-length FklB, however FklB shares a high level of sequence identity with other PPIases. Therefore, a homology model based on the sequence was generated using the modelling server I-TASSER [8]. This generated five models which were then analysed using the program PROCHECK to identify the model that had the best stereochemical quality [9]. The best model was selected and then superimposed onto the crystal structure of the Macrophage infectivity protein (MIP) from *Legionella*, which has a 42% sequence identity to FklB (PDB ID: 1FD9, [237]) to create the FklB dimer (Figure 5-14). The model shows a distinct C-terminal domain corresponding to the catalytic domain, which is joined to the N-terminal dimerisation domain by a long helical region.

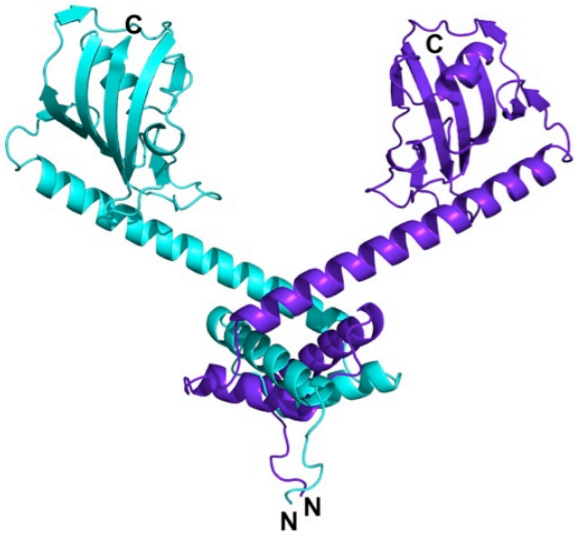


Figure 5-14 FklB homology model. The homology model was built using I-TASSER [151], which predicted the structure of the monomer. A dimer was created using 1FD9 as a template, each monomer is represented in a different colour. The N- and C-termini of each monomer are indicated.

5.9 Determination of the oligomeric state of FklB

To determine the monodispersity of FklB, AUC was performed. Sedimentation velocity experiments confirmed that FklB is predominately present as a single, monodisperse species in solution, evidenced by a single peak in the size distribution analysis (Figure 5-15). A range of protein concentrations was tested (0.2 - 10 mg ml^{-1}) to determine the effects of concentration on the oligomeric state.

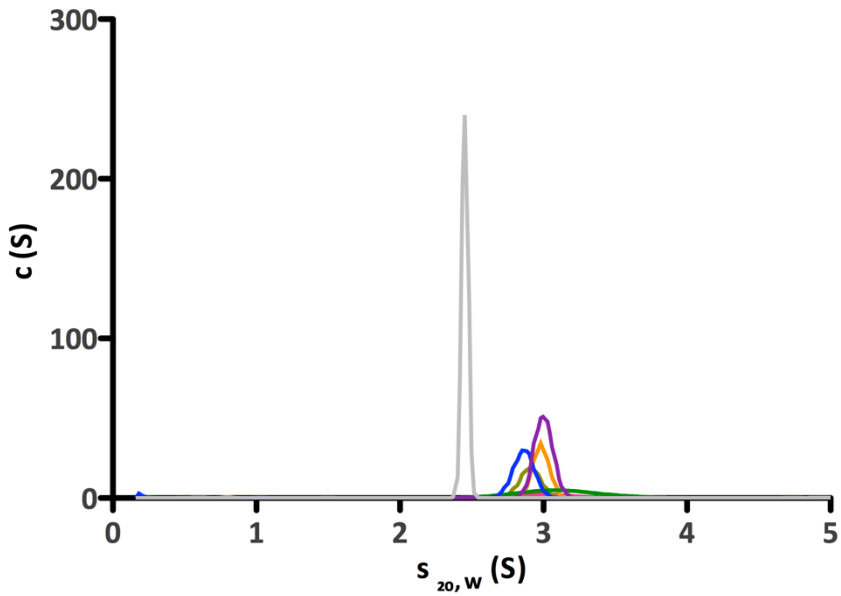


Figure 5-15 Size distribution (c(s)) of FklB determined from SV data. The sedimentation velocity experiment was performed at 4°C and the sedimentation coefficients were converted to their standard values (at 20°C in H_2O). A range of protein concentrations were tested: 0.2 mg ml^{-1}, green; 0.4 mg ml^{-1}, pink; 0.6 mg ml^{-1}, purple; 0.8 mg ml^{-1}, light green; 1 mg ml^{-1}, orange; 5 mg ml^{-1}, blue; 10 mg ml^{-1}, grey.

Plotting the sedimentation coefficient derived from the SV data against the loading concentration of the sample enabled the calculation of the sedimentation coefficient of FklB at infinite dilution (Figure 5-16). This gave $s^0_{20,w}$ = 3.04 ± 0.04 S. The negative slope of the graph indicates the effects of non-ideailty on the system, as the sedimentation coefficients calculated for 5 and 10 mg ml^{-1} are considerably lower than those at low concentrations. Omitting these data points gave $s^0_{20,w}$ = 3.14 ± 0.07 S. Hydrodynamic modelling was used to calculate the sedimentation coefficient of the FklB homology model using SOMO [125], and gave a theoretical sedimentation coefficient of 3.12 S, which corresponds well with the experimental values obtained.

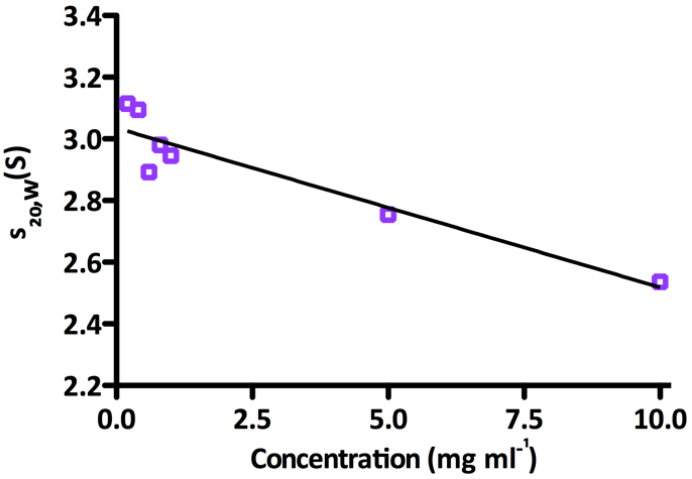


Figure 5-16 Sedimentation coefficient of FklB. The extrapolation to zero concentration of the sedimentation coefficients ($s^0_{20,w}$) obtained by c(s) analysis of SV data for a range of protein concentrations. This gave $s^0_{20,w}$ = 3.04 ± 0.04 S.

To determine the oligomeric state of FklB, sedimentation equilibrium (SE) studies were conducted. SE data fit with a single species model gave a MW of 41.5 ± 2.9 kDa, which indicates the formation of a dimer, predicted from its sequence to have a MW of 50 kDa (Figure 5-17). The MW determined using the single species analysis predicts a lower than expected MW, which may be caused by non-ideality of the protein at higher concentrations. Omitting the values obtained for the higher concentrations gave MW = 47.6± 2.6 kDa. However, at the lowest concentration (0.2 mg ml^{-1}) the MW obtained using a single species model was 35.5 kDa, which is closer to the MW of a monomer. This may indicate the presence of some monomer at lower protein concentrations. Therefore the data were fitted using a monomer-dimer self-association model with non-ideality. The goodness of fit of the model to the data was determined from the chi-squared (χ^2) test, which determines the discrepancies between the model and the experimental data and a value of 0 indicates a perfect fit. Using a monomer-dimer self-association model gave a better fit to the data with a χ^2 of 0.0000712 compared with 11 for the single species model, the fits and residuals are shown in Figure 5-18. The K_d of dimerisation is 7.9 µM.

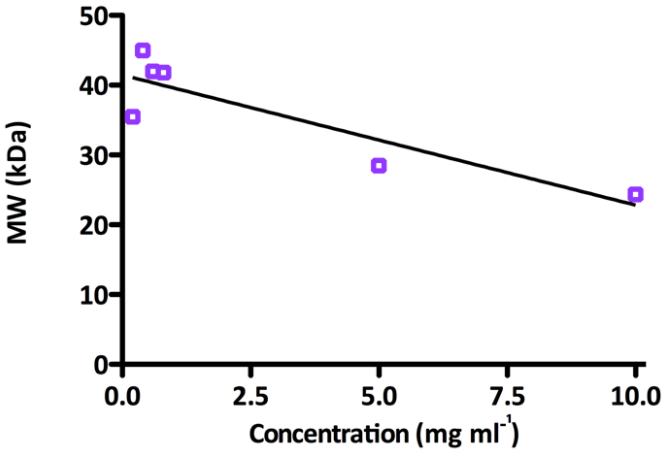


Figure 5-17 Molecular weight of FklB determined from SE data. The molecular weight was determined by fitting the data with a non-interacting single discrete species model for a range of protein concentrations.

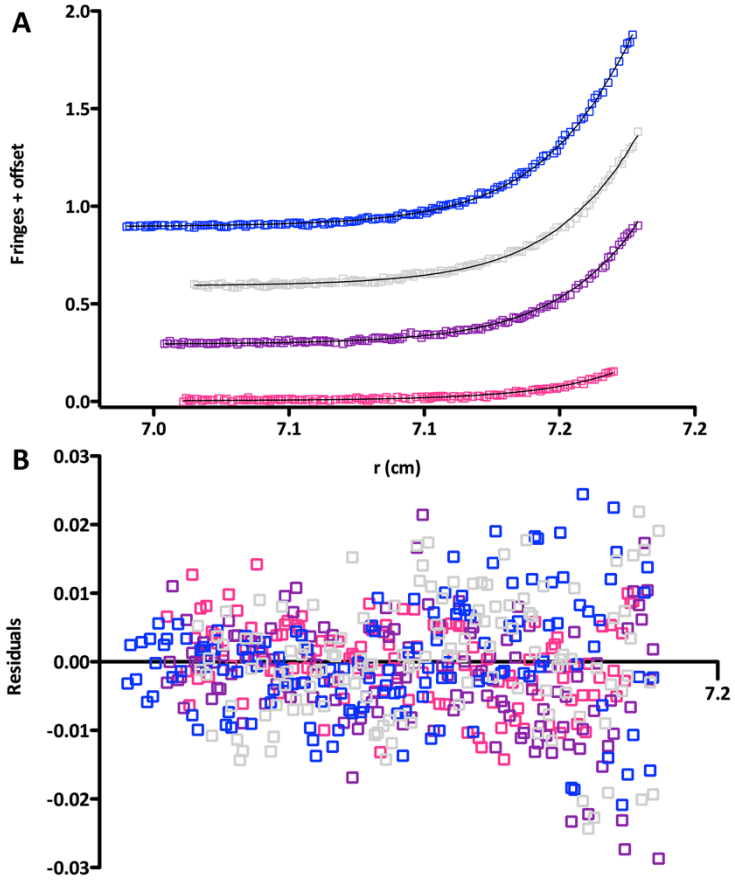


Figure 5-18 FklB sedimentation equilibrium data fitted with a monomer-dimer model. (A) Data were fitted with a monomer-dimer model over a range of concentrations: 0.2 mg ml^{-1}, pink; 0.4 mg ml^{-1}, grey; 0.6 mg ml^{-1}, blue; 0.8 mg ml^{-1}, purple. The residuals for each data set are shown in the corresponding colour in **(B)**. This model gave a dissociation constant (K_d) of 7.9 µM.

5.10 Solution structure of FklB

A monodisperse solution is required for several biophysical techniques, for example X-ray crystallography, and is preferable for others, including SAXS. Since AUC revealed that FklB is monodisperse at concentrations significantly exceeding the K_d of dimerisation and there is currently no crystal structure for this protein, FklB is an ideal candidate for *ab initio* model construction. SAXS data were collected for a sample of FklB at 6 mg ml^{-1} and are shown in Figure 5-19. The data were processed using PRIMUS [113] where in analysis of the Guinier region of the data revealed that the protein was neither aggregated nor affected by inter-particle interference at this concentration (inset Figure 5-19). From this region a radius of gyration (R_g) of 34.5 ± 1.2 Å was determined.

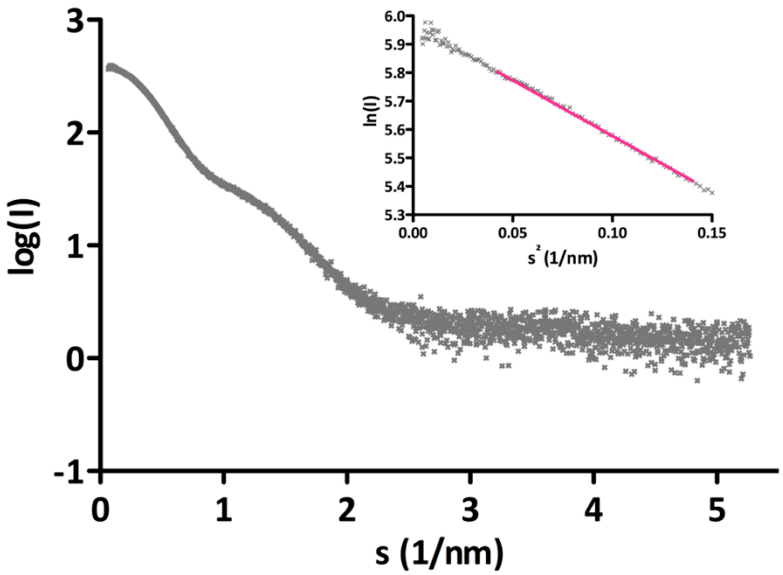


Figure 5-19 SAXS curve with Guinier analysis. The experimental data are shown in grey and the Guinier region of the data is shown inset. From the Guinier region (pink line) an R_g of 34.5 ± 1.2 Å was calculated.

Kratky analysis of scattering data gives information about the folded state of a protein. Figure 5-20A shows the Kratky plot for FklB which displays a predominant peak at low angles, indicative of a folded globular protein. The pair-wise distribution function was calculated using GNOM [116] following indirect Fourier transform of the scattering data. This indicates that the maximum dimension (D_{max}) of the particle is 112 Å. The shape of the *p*(*r*) plot shows a bimodal distribution, which indicates that FklB has a 2-domain structure (Figure 5-20B). Calculation of the particle volume from the data gave a value of

76.72 nm^3. This value can be used as a rough approximation of the molecular mass (MW), since this value usually corresponds to 1.5-2 times the MW (in kDa) [110]. This predicts a MW of between 38.4 and 51.5, which corresponds with the size of dimeric FklB, at 44.4 kDa.

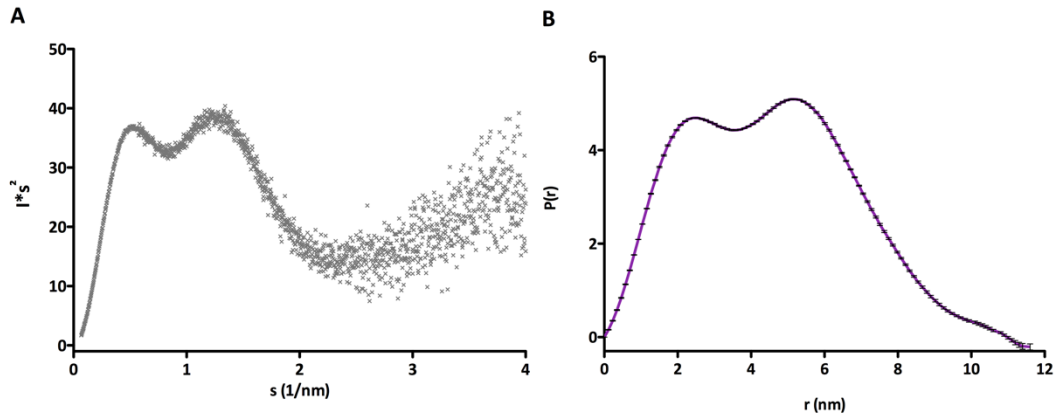


Figure 5-20 Kratky plot and pair-distance p(r) distribution. (A) Kratky plot of SAXS data indicating that the protein is folded but flexible. **(B)** Indirect Fourier transform of the SAXS data using GNOM gives the pair-distance distribution function which indicates that the maximum dimension of the protein is 112 Å.

To compare the fit of the homology model to the SAXS data, a theoretical scattering curve was generated from the model using CRYSOL [186]. The fit of the homology model to the SAXS data is shown in Figure 5-21, and has a χ^2 of 4.16. Despite the high χ^2 value, which indicates a poor fit of the model to the data, the curves generally overlay quite well indicating that the overall shape of the homology model is similar to the shape of the protein in solution.

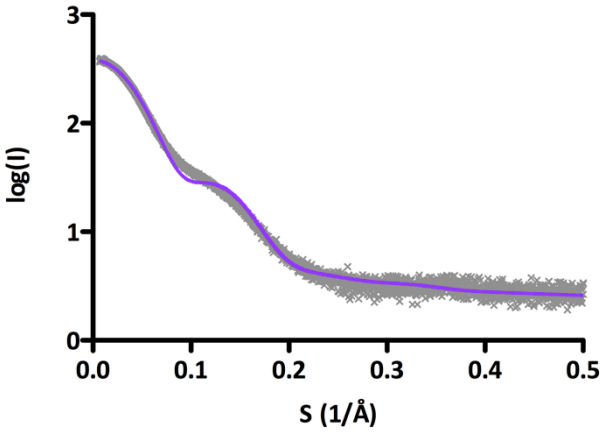


Figure 5-21 Comparison of the scattering curve calculated for the FklB homology model using CRYSOL [186] **with the experimental SAXS data.** The overlay of scattering curves indicates that the homology model has a similar shape to the protein in solution.

An *ab initio* shape reconstruction of FklB from the SAXS data was performed using DAMMIF with P2 symmetry (Figure 5-22A and B). The solution structure was generated from an average of 20 DAMMIF models that were averaged and filtered using DAMAVER [123]. The DAMMIF model has a structure similar to that of the homology model of FklB (Figure 5-22C and D). Both show two distinct C-terminal domains (one per monomer) connected to a single dimerisation domain. The domains of the DAMMIF model corresponding to the C-terminal domains appear more elongated compared with those in the homology model, which may indicate that there is some flexibility in the orientation of these domains. This is consistent with the Kratky plot (Figure 5-20).

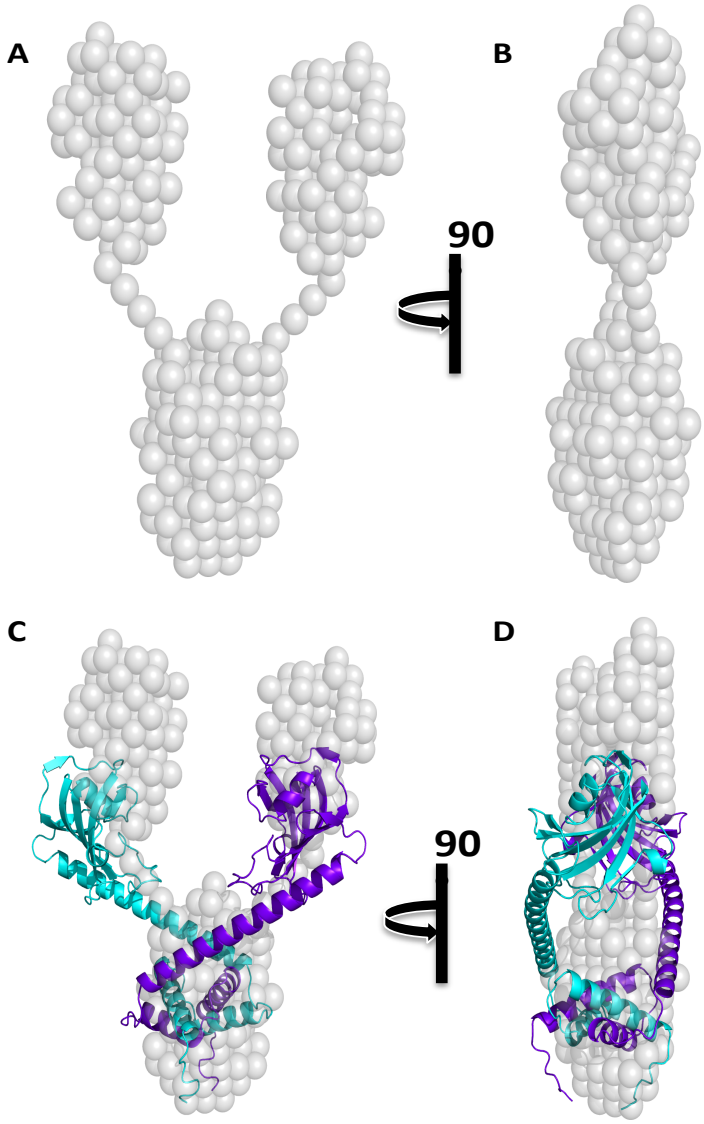


Figure 5-22 *Ab inito* modelling of FklB in solution. The final DAMMIF model was generated following averaging and filtering of 20 *ab initio* models [115,122], which have been shown in orthogonal views **(A)** and **(B)**. The overlay of the DAMMIF model with the homology model is shown in **(C)** and **(D)**.

5.10.1 *Modelling FklB flexibility*

The SAXS data indicate that FklB has some degree of flexibility determined from the Kratky plot. Several programs have been developed that interpret a SAXS curve as an ensemble of structures, such as the EOM [124] and SOMO-DMD [125]. Both programs were tested as a comparison and the results obtained were similar, therefore only the results obtained using SOMO-DMD will be discussed here.

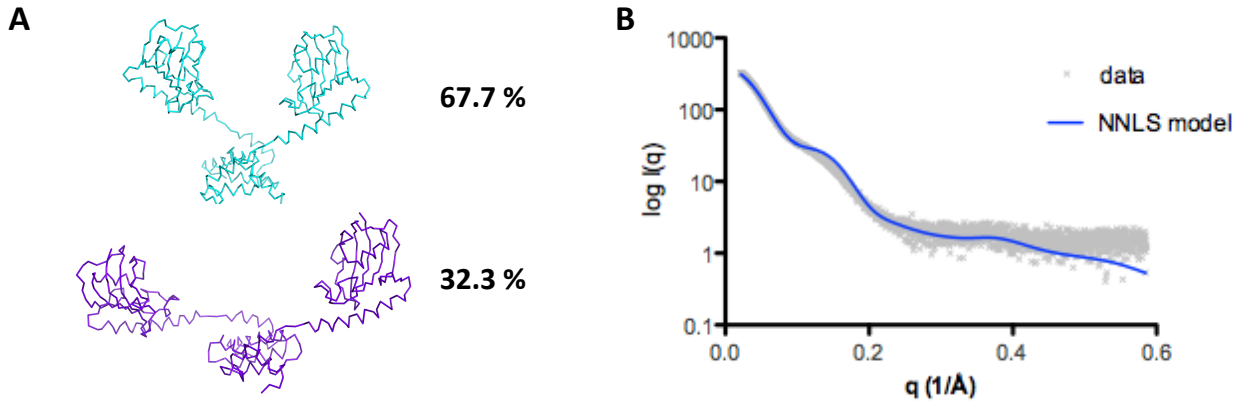


Figure 5-23 Modelling FklB flexibility using SWISS-MODEL homology models. (A) Homology models generated using SWISS-MODEL based on two crystal forms of FkpA with the percentage of scattering that contributed to the fit shown in B. **(B)** NNLS fit of SWISS-MODELs which gave a χ^2 of 3.7 (blue) compared with the experimental data (grey).

As shown in Figure 5-21 the fit of the homology model to the SAXS data was poor (χ^2 = 4.16) which may be because using one model does not reflect the range of conformations seen in solution. The region of FklB that is likely to be flexible was determined from crystal structures of FkpA, which shares 41% sequence similarity with FklB. There are two crystal forms of this protein which, taken together, suggest that the extended helix between the N- and C-terminal domains can bend. Homology models of FklB (Figure 5-23A) based on these structures were generated using SWISS-MODEL [238]. Fitting the data using a combination of the two models gave a better fit with a χ^2 value of 3.7 (Figure 5-23B). SOMO-DMD uses a non-negative least squares method (NNLS) to fit the data, which calculates the percentage of contribution for each of the models. However, the two models may not fully describe the scattering data, since it is likely that a range of confirmations exist between these states in solution. From 50 models generated from the homology model with SOMO-DMD, five models were selected that best describe the SAXS data by NNLS fitting. The chosen models and their percentage contribution determined from NNLS fitting are shown in Figure 5-24A. The fit to the data shown in (Figure 5-24B) gave a χ^2 of 2.6, indicating that using a range of models to describe the SAXS data produced a better fit.

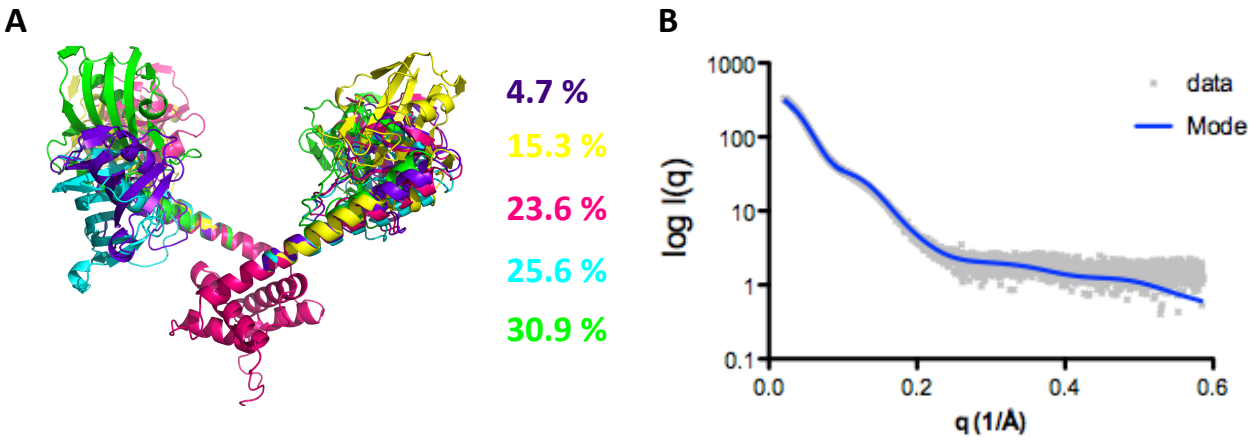


Figure 5-24 Modelling FklB flexibility using SOMO-DMD. (A) FklB models generated using SOMO-DMD based on the FklB homology model described in Section 5.8.2. The models have been superimposed and the colour of the models corresponds to the percentage value, indicating the calculated contribution of that conformation to the scattering data. **(B)** NNLS fit of SOMO-DMD generated models that gave a χ^2 of 2.6 (blue) compared with the experimental data (grey).

5.11 FklB enzyme activity

As with SurA and FkpA, the PPIase activity of FklB was investigated using a protease-coupled enzyme assay. This assay confirmed that the purified protein retained its enzymatic activity as the rate of *cis* to *trans* isomerisation was increased in the presence of FklB (Figure 5-25A). The pseudo first order rate constant was calculated using the enzymatic rate from the initial 10-50 s of the curve from reactions performed in the presence of a range of FklB concentrations (0 to 200 nM) (Figure 5-25B). From this the specificity constant (k_{cat}/K_m) was 0.197 x 10^6 M^{-1} s^{-1}. FklB has a higher specificity constant than FkpA or SurA indicating greater enzyme efficiency.

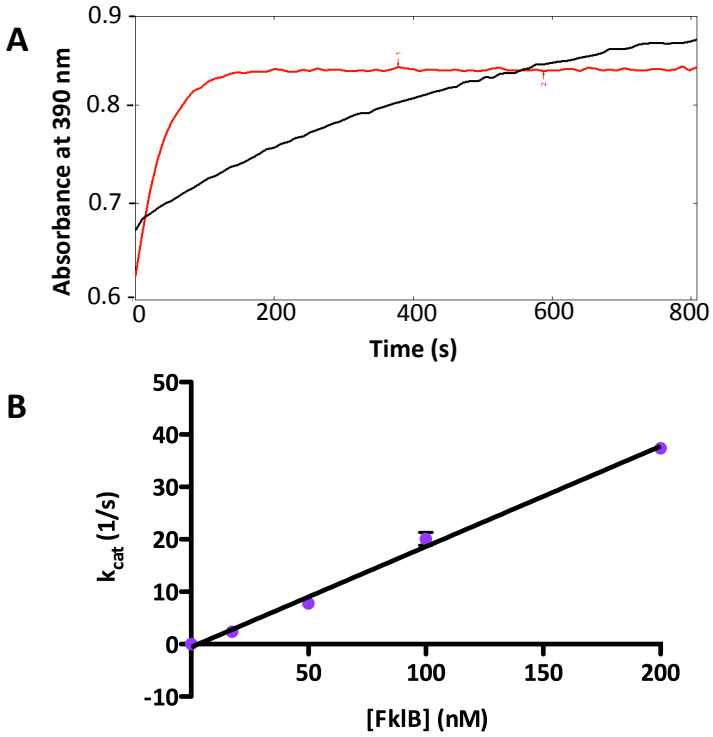


Figure 5-25 Determination of the first order rate constant of the PPIase activity of FklB. (A) An example time course curve showing the reaction catalysed by FklB (red line) compared with the uncatalysed reaction (black line). **(B)** The rate constant (k) of the catalysed first-order *cis-trans* conversion upon addition of 0-200 nM of purified FklB. The calculated specificity constant k_{cat}/K_m was 0.197 x 10^6 M^{-1} s^{-1}. The values are means (± standard deviation) from at least three independent experiments.

5.11.1 *Inhibition of FklB*

All known FKBPs with PPIase activity are inhibited by the immunosuppressant rapamycin [208,239], therefore inhibition of FklB by rapamycin was investigated. FklB activity was inhibited by nanomolar concentrations of rapamycin (Figure 5-26). An accurate K_i could not be calculated from this assay owing to the high affinity of FklB for rapamycin, therefore the binding of rapamycin would need to be investigated by other methods, such as ITC.

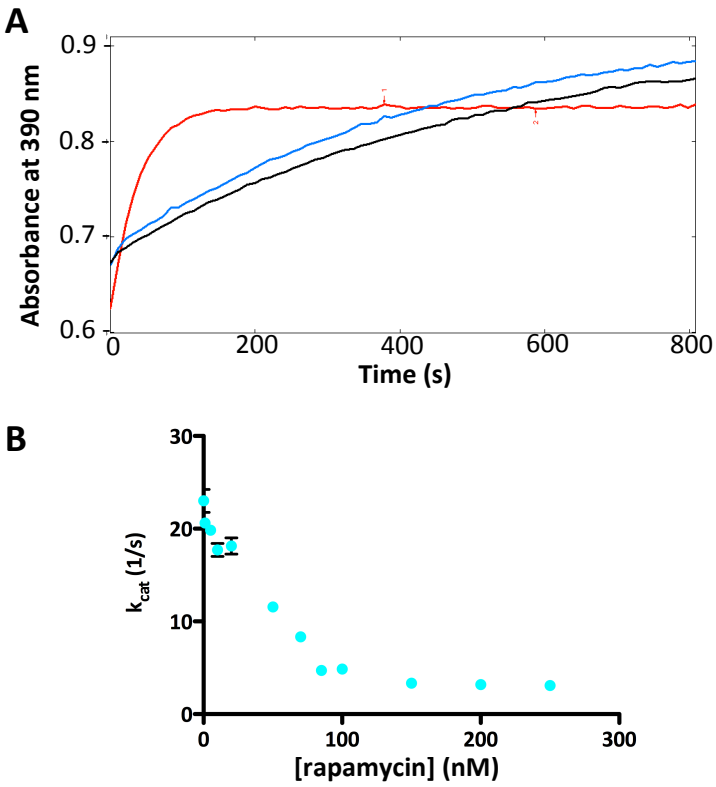


Figure 5-26 Inhibition of FklB with rapamycin. (A) An example time course curve showing the reaction catalysed by FklB (red line) compared with the reaction in the presence of 200 nM rapamycin (blue line) and the uncatalysed reaction (black line). **(B)** 100 nM FklB was inhibited with varying concentrations of rapamycin (0 - 200 nM). The values are means (± standard deviation) from at least three independent experiments.

The effect of ME0052 on FklB activity was also investigated. However, at high concentrations the absorbance of the compound interfered with the assay, therefore a K_i could not be determined, although some inhibition of enzyme activity was observed. For example, the average enzyme rate for 100 nM FklB was 23 ± 1.23 mAbs min^{-1}. However in the presence of 20 µM ME0052 the average rate was 17 ± 0.85 mAbs min^{-1}, which can be seen as a decrease in the steepness of the curve in Figure 5-27. To circumvent the ambiguity of the assay the interaction of ME0052 and FklB was investigated using other methods.

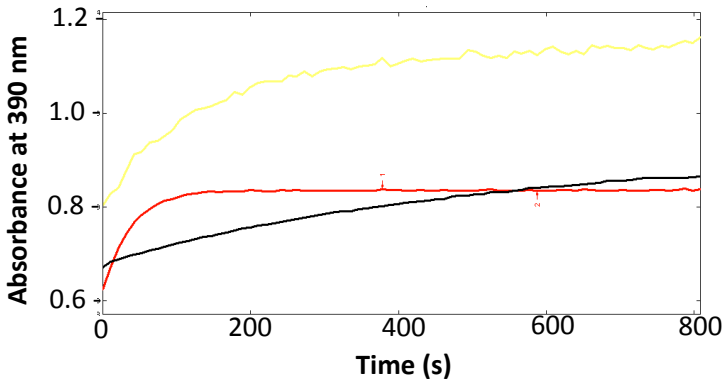


Figure 5-27 Inhibition of FklB by ME0052. An example time course curve showing the reaction catalysed by FklB (red line) compared with the reaction in the presence of 20 µM ME0052 (yellow line) and the uncatalysed reaction (black line).

5.12 Binding of rapamycin and ME0052 to FklB

5.12.1 *^{15}N chemical shift assay*

The binding of ME0052 to FklB required further investigation. First, it remained to be further validated that ME0052 does in fact bind to FklB since far-western analysis can often be misleading, and if so, where it binds. NMR is a powerful technique for investigating protein-ligand interactions. The enzyme assay indicated that ME0052 might be inhibiting FklB activity. Since the C-terminal domain of the protein is responsible for its enzymatic activity; FklBc was labelled with ^{15}N isotope and purified. The use of the C-terminal domain of FklB was preferable to using full-length protein, as initial spectra of the full-length ^{15}N labelled protein were poor owing to the size of FklB and the fact that it forms a dimer. Sharp, well-dispersed peaks in the initial spectra of FklBc indicated that the C-terminal domain was in a folded state. Since ME0052 is soluble only in DMSO, deuterated DMSO (dDMSO) was added as a control to identify whether any resonances were perturbed by its presence. This revealed that several resonances shifted upon addition of dDMSO (Figure 5-28). Since the addition of dDMSO alone induced significant shifts in several peaks, the ^{15}N-HSQC spectrum of FklBc in the presence of 200 µM ME0052 was compared with the spectrum obtained in the presence of dDMSO. There was no observable change in the spectrum upon the addition of ME0052 (Figure 5-29), indicating that FklB does not bind to ME0052.

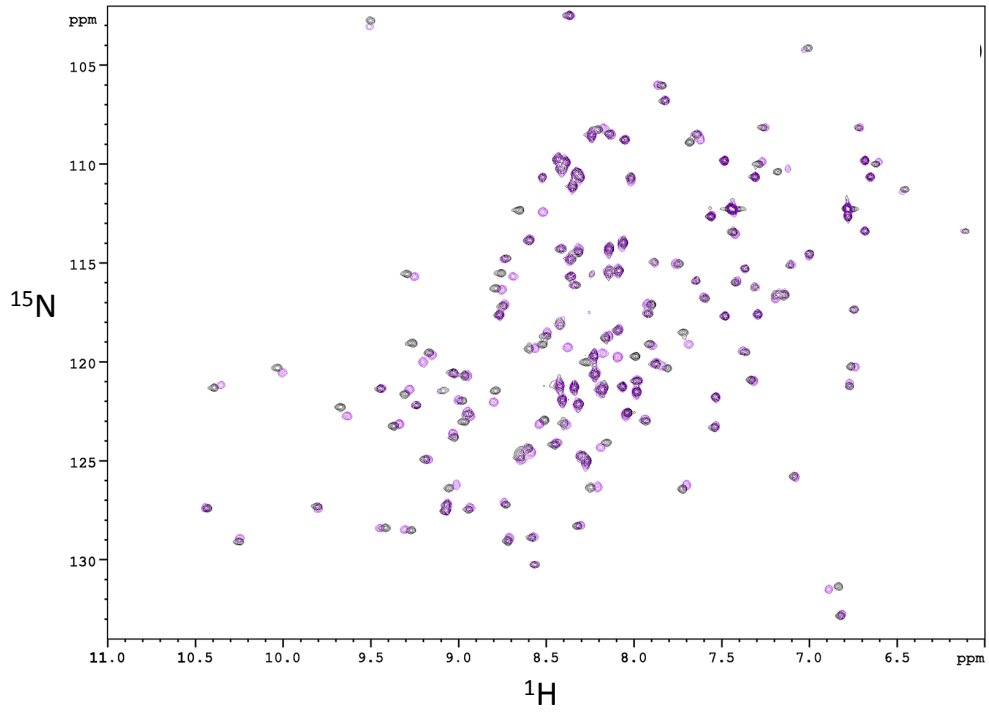


Figure 5-28 ^{1}H-^{15}N HSQC spectra of 100 µM FklBc in the presence and absence of DMSO. The spectrum acquired for FklBc (purple) is overlaid with a spectrum acquired in the presence of 3% DMSO (black).

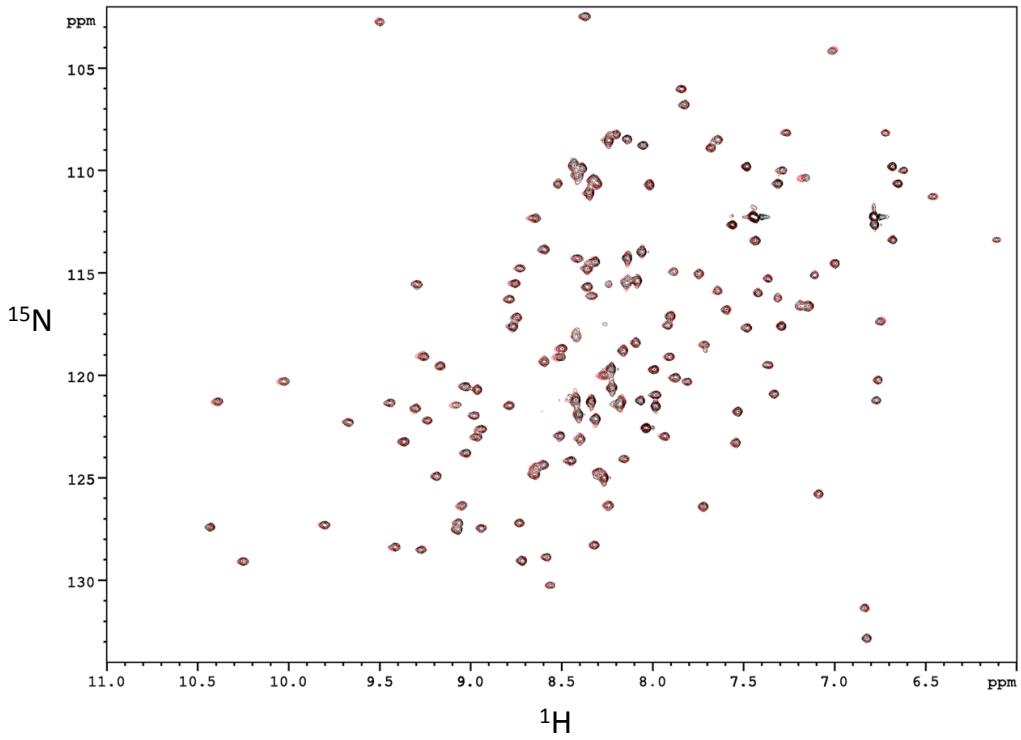


Figure 5-29 ^{1}H-^{15}N HSQC spectra of 100 µM FklBc with and without 200 µM ME0052. The spectrum acquired for FklBc in the presence of ME0052 (red) is overlaid with the spectrum for 100 µM FklBc with 3% DMSO (black).

Since rapamycin inhibits the enzymatic activity of FklB, the binding of this drug to FklBc was investigated using a chemical shift assay. Comparison of the spectra obtained in the presence and absence of rapamycin show chemical shift perturbations for several residues (Figure 5-30), indicating that rapamycin is interacting specifically with FklBc. It is clear from the spectra that the binding is in a fast-exchange regime and most peaks are present in duplicate corresponding to both the bound and unbound states of the protein. This can be explained by a lack of protein saturation by the drug as a result of poor solubility of rapamycin. To determine the binding site of rapamycin on FklBc, ^{1}H-^{15}N 3D NOESY-HSQC and 3D TOCSY-HSQC data were collected. Sequence-specific backbone assignment will be conducted as part of a follow-on project to determine which residues are involved in drug binding.

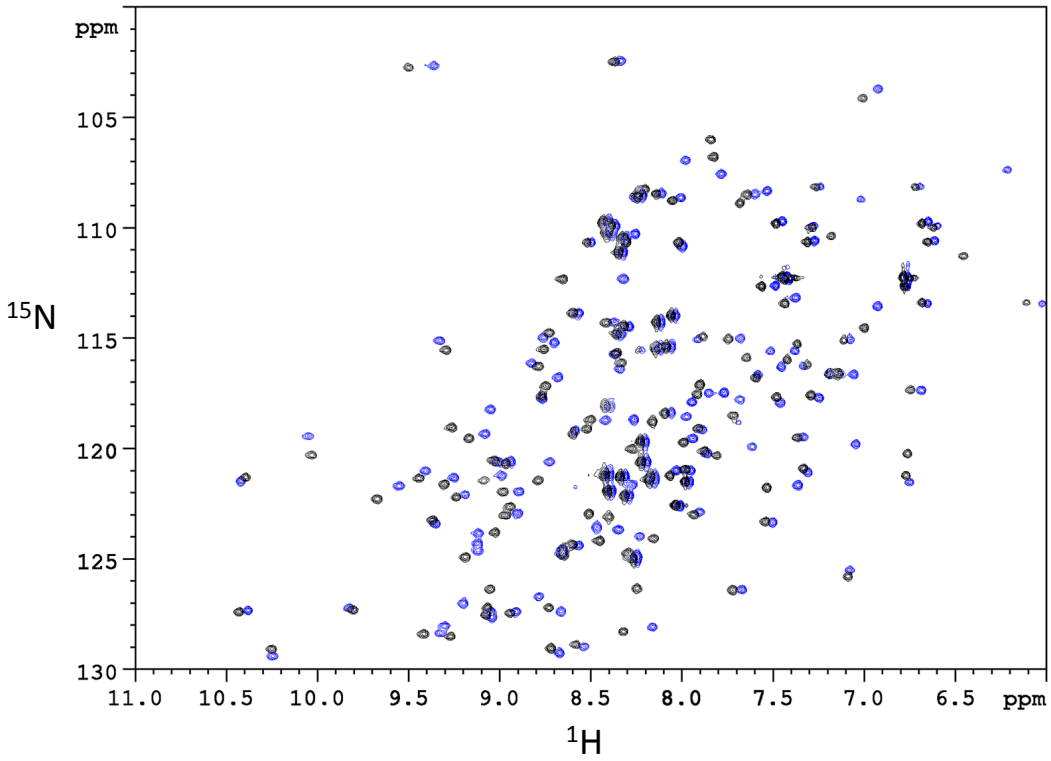


Figure 5-30 ^{1}H-^{15}N HSQC spectrum of 100 µM FklB with and without 20 µM rapamycin. The spectrum acquired for FklBc in the presence of rapamycin (blue) is overlaid with the spectrum for 100 µM FklBc with 3% DMSO (black).

Several FKBPs have been crystallised in complex with rapamycin, and the residues involved in rapamycin binding have been identified. Owing to the high level of homology of FklB with other FKBPs, the residues involved in rapamycin binding can be predicted. Figure 5-31A shows a sequence alignment of FklBc with two FKBPs that have been

crystallised in complex with rapamycin. The sequence similarity between FklBc and the FKBPs is between 39-46% (Figure 5-31). In Figure 5-31 the residues involved in rapamycin binding have been highlighted in cyan and show that all of the required residues are conserved in FklB. The three structures overlay well with an r.m.s.d. less than 0.8 Å (Figure 5-31B) and accordingly the rapamycin binding residues are in a similar orientation (for clarity only the side chains corresponding to FklBc have been shown in Figure 5-31B). Based on the position of rapamycin in the crystal structures 1C9H and 2DG3, the potential binding interactions between FklBc and rapamycin are shown in Figure 5-31C.

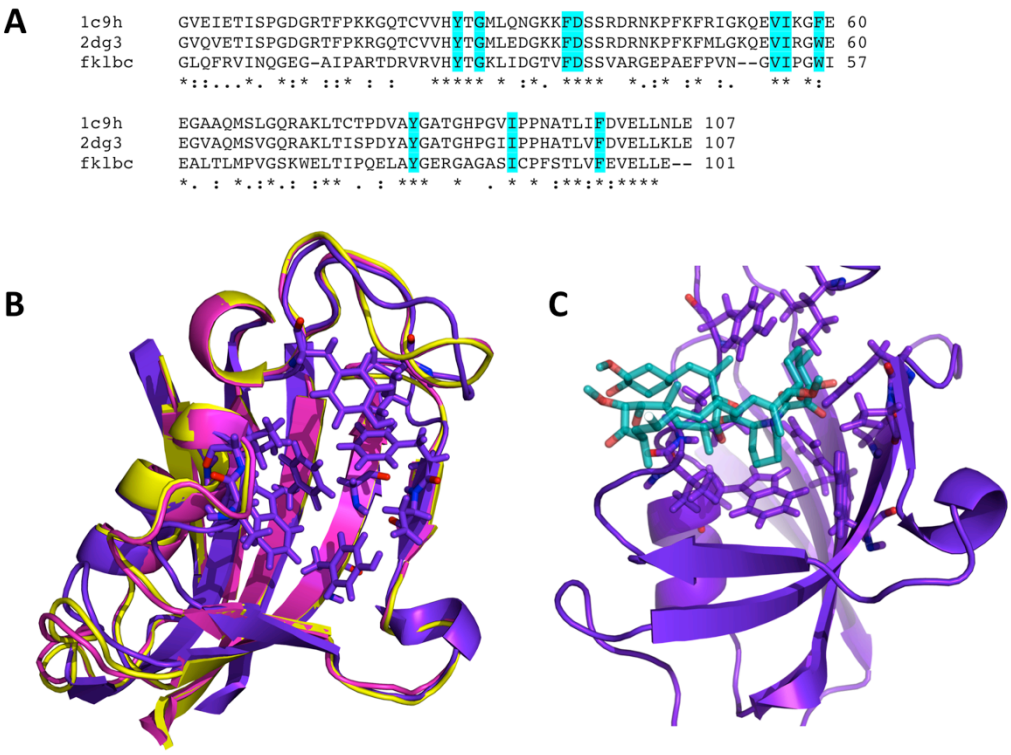


Figure 5-31 Predicted rapamycin binding site in FklBc. **(A)** The residues required for binding rapamycin are conserved in FklB. The sequences of two FKBPs crystallised in complex with rapamycin are shown (1C9H and 2D3G) and the residues involved in forming interactions with rapamycin have been highlighted in cyan. **(B)** Superimposition of FklBc with 1C9H and 2D3G showing the overall structural similarity of the FKBPs. For FklBc the rapamycin binding residues have been shown as stick models. **(C)** A predicted model of the position of rapamycin in the active site and how it would interact with the rapamycin. The position of rapamycin (shown in cyan) in FklBc is based on its position in 1C9H and 2D3G.

5.12.2 *Isothermal titration caliormetry*

Enzyme inhibition data determined that the dissociation constant of FklB binding to rapamycin was in the low nanomolar range but could not accurately be quantified by the assay. Therefore ITC was conducted to determine the binding affinity of FklB to rapamycin. Figure 5-32A shows the binding thermogram from the interactions of rapamycin with FklB. This shows that the binding is exothermic and is of high affinity with a Kd of low micromolar to high nanomolar. The integrated points shown in Figure 5-32C suggest a stoichiometry of binding of 1:0.5, (compound:protein) corresponding to the two catalytic domains present per dimer. No binding site model could be fitted with confidence owing to a low signal to noise ratio. To determine whether the N-terminal domain contributes to the binding interactions with rapamycin, ITC was carried out for the C-terminal domain. FklBc exhibits a very similar binding profile to the full-length protein (Figure 5-32B) and there is no difference in the heats of the reaction, suggesting that the residues involved in rapamycin coordination are all located within the C-terminal. A binding isotherm could not be fitted to the data in this case and the raw data are shown to provide qualitative comparison of the binding.

Since the ^{1}H-^{15}N HSQC chemical shift assay was conducted using FklBc and no binding to ME0052 was detected, it was important to rule out the possibility that ME0052 may be binding to the N-terminal domain of the protein. Therefore, ITC was carried out by titrating ME0052 into FklB. The binding isotherm shows no indication of binding. The insolubility of the ME0052 made the reaction difficult to conduct. However, several attempts to repeat the experiments revealed similar isotherms that did not give any indication that ME0052 was binding to FklB (Figure 5-33).

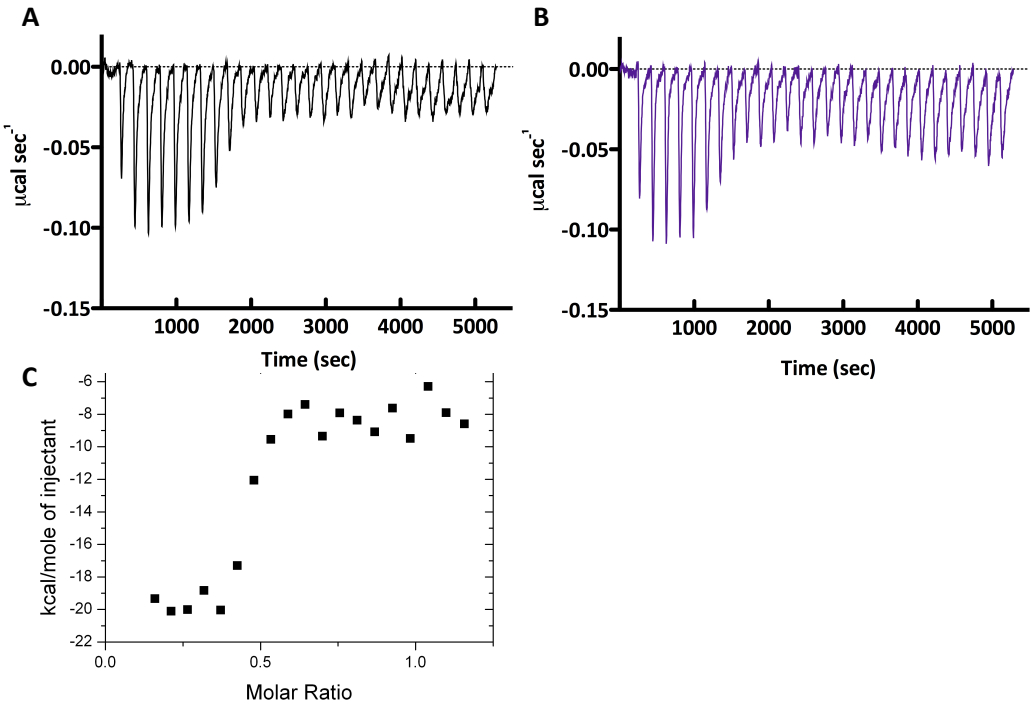


Figure 5-32 Calorimetric titration of FklB or FklBc to rapamycin. Each peak represents the injection of 10 µl of 25 µM FklB (black) **(A)** or FklBc (purple) **(B)** into 3.5 µM rapamycin at 25°C in 25 mM HEPES, pH 8, 50 mM NaCl, 0.5% DMSO. The integrated heats of reaction (A) are shown in **(C)**.

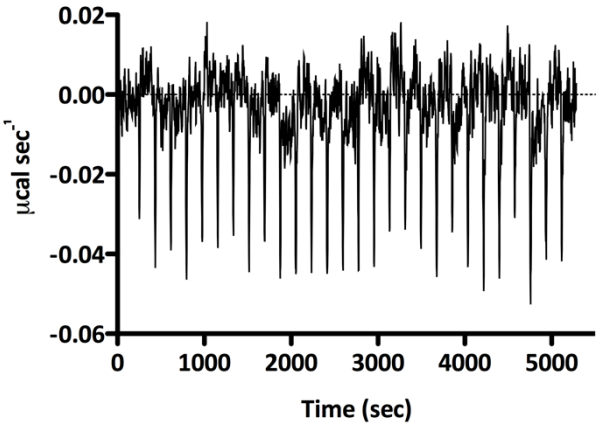


Figure 5-33 Calorimetric titration of ME0052 into FklB. Each peak represents the injection of 10 µl of 200 µM ME0052 into 50 µM FklB at 25°C in 25 mM HEPES, pH 8, 50 mM NaCl, 0.5% DMSO.

5.13 The effects of *fklB* gene deletion on the LEE T3SS

5.13.1 *Construction of O157:H7 ΔfklB mutants*

FklB was deleted in *E. coli* O157:H7 by allelic exchange. The exchange plasmid was constructed by inserting the *fklB* flanking regions into pIB307, which encodes chloramphenicol resistance. The *fklB* flanking regions were synthesised by Genscript and a BamHI sequence was inserted between the left-hand and right-hand flanking regions. A sacB-kan cassette was cloned into the BamHI site to create pIB307-FklB. pIB307-FklB was transformed into *E. coli* O157:H7 by electroporation and transformants were selected based on their ability to grow on both Chl and Kan. Positive colonies were sub-cultured in LB with Chl at 30°C for several days, followed by culturing at 42°C in LB-Kan to promote integration. Cultures were then plated onto LB-Kan plates and checked by PCR for absence of *fklB* using primers that anneal to the 3' and 5' end of *fklB* (621bp). Since the sacB-kan cassette is large it did not amplify in this reaction, confirming that *fklB* had been replaced by the cassette (Figure 5-34). Attempts were made to remove the sacB-kan cassette by allelic exchange, but were unsuccessful.

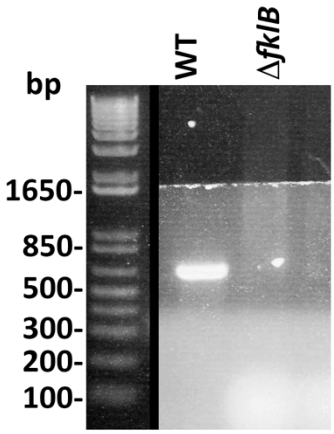


Figure 5-34 Deletion of *fklB*. Agarose gel showing the results of a PCR reaction to amplify *fklB* from genomic DNA from WT and the Δ*fklB* mutant. The size of the *fklB* gene is 621 bp. The position of the DNA ladder has been indicated (bp).

5.14 Does the deletion of *fklB* affect the expression of the LEE T3SS or flagellar system?

As for the Δ*surA* mutant, the expression of the LEE T3SS was monitored using GFP reporter plasmids. Figure 5-35 A-F shows the level of GFP expression for each of the reporters. The fold-change in the level of expression between the WT and deletion mutant is shown in Figure 5-35H. There was no significant difference in the expression of any of the LEE or the *fliC* GFP reporter plasmids indicating that deletion of *fklB* has no effect on these systems. The secretion profile of the Δ*fklB* mutant further confirms that the LEE T3SS in the mutant is comparable with that of the WT (Figure 5-36).

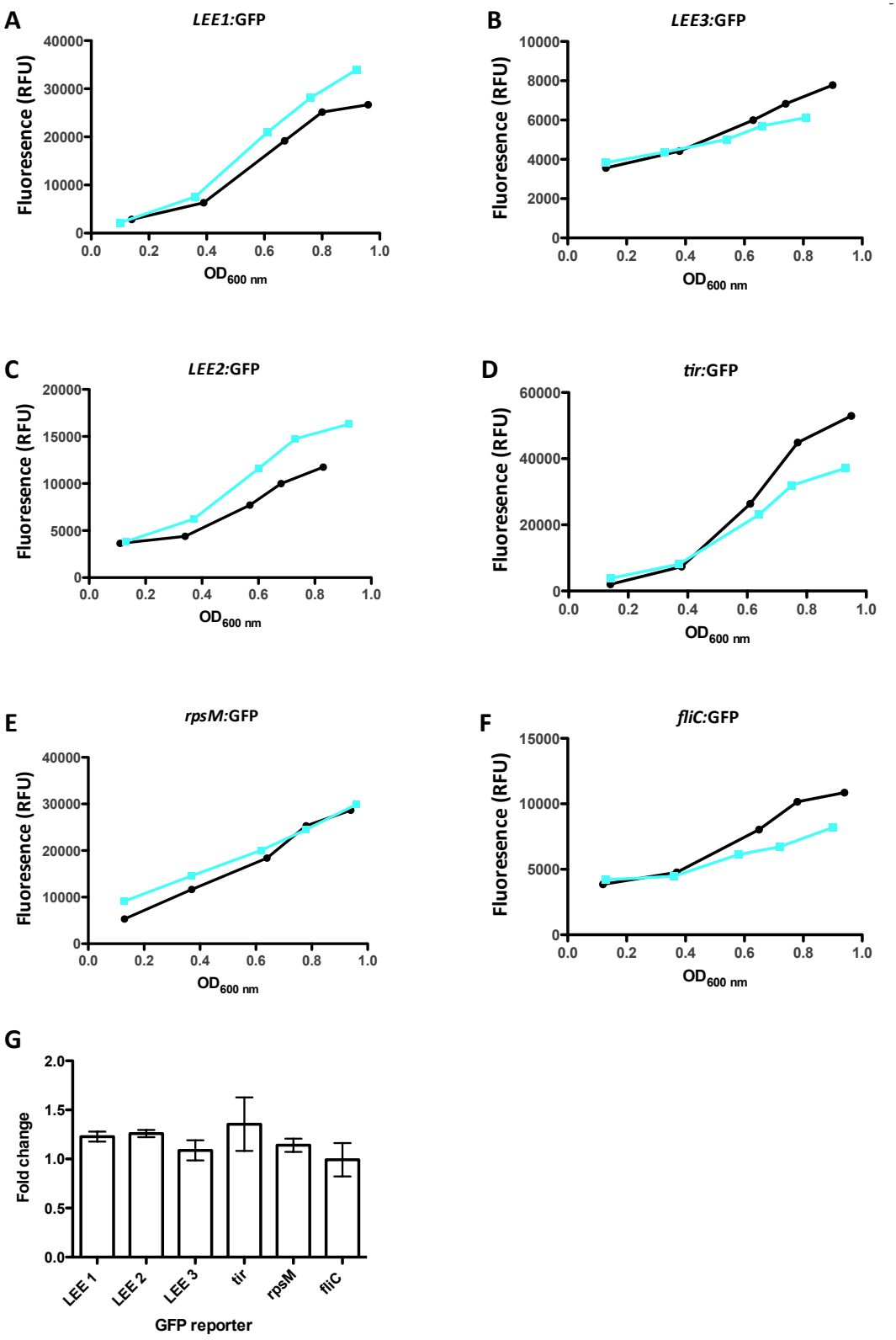


Figure 5-35 Expression of gene:GFP reporter fusions in Δ*fklB*. Gene expression in the WT is shown in black and in the deletion mutant in cyan. Line graphs show the relative fluorescence (plotted against the optical density for LEE1 **(A)**, LEE2 **(B)**, LEE3 **(C)**, Tir **(D)**, RPSM **(E)** and FliC **(F)**. The graphs shown are representative of triplicate experiments. The fold-change of gene expression at an $OD_{600\ nm}$ of 0.7 is shown in a bar chart **(G)** where the error bars are derived from triplicate experiments.

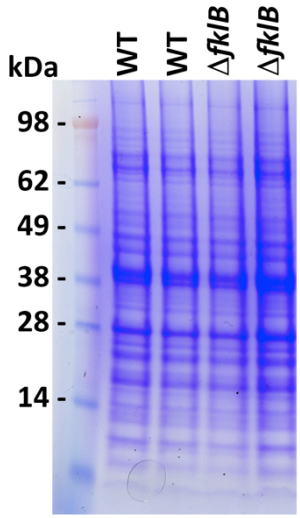


Figure 5-36 Secretion profile of the WT and Δ*fklB* mutant. SDS-PAGE gel showing duplicate secretion assays for the WT and the deletion mutant. The positions of the molecular size makers have been indicated (kDa).

5.15 Discussion

5.15.1 *Validation of binding targets*

The identification of three proteins belonging to the same protein class in the SA compound affinity pull-down assay was an intriguing result. Since the function of these proteins is related, it may follow that they bind to and are inhibited by similar molecules. However, further investigation of the classification of the PPIases revealed that SurA was in fact a parvulin and its structure and composition is unrelated the FKBP family, to which FkpA and FklB belong. As all three proteins catalyse the same reaction, it remained a possibility that the compounds were inhibiting their enzymatic activity.

Initial confirmation of binding was carried out by far-western analysis of the purified proteins. In the case of Tpx and WrbA, the far-western analysis showed binding of ME0052 which was further confirmed using other techniques, thus this technique has proved useful for validating the pull-down targets. However, far-western analysis confirmed the binding of only one (FklB) of the three PPIases to ME0052, indicating that the pull-down assay may have highlighted proteins that were in fact false positives. Consequently the lack of binding seen in the far-western alone was insufficient to

conclude whether these proteins did or did not bind to the compounds, therefore further investigation was performed.

Following purification of the three PPIases the enzyme activity of each of them was confirmed and the specificity constant determined. The specificity constant of SurA determined from this investigation was 7.76 x 10^3 M^{-1} s^{-1} which is lower than the previously reported value of 3.4 x 10^4 M^{-1} s^{-1} [217]. Differences in the measured specificity constant may reflect non-optimal experimental conditions. For FkpA the specificity constant of the protein determined in this study (7.76 x 10^3 M^{-1} s^{-1}) was two orders of magnitude slower than has previously been reported (7.5 x 10^5 M^{-1} s^{-1}) in other studies [240]. FkpA from *E. coli* O157 is identical to the *E. coli* K12 protein, therefore the observed differences further indicate that the experimental setup was not optimal, for example the buffers used may have compromised the protein activity. For FklB the specificity constant is around 10-fold higher than previously reported [204,210]. However, since the kinetics of these enzymes had been studied previously, the aim of the enzyme studies in this investigation was to determine whether the purified proteins retained enzyme activity following purification, indicating that the protein was folded correctly. Furthermore, characterisation of the enzyme rates, although not consistent with the findings of other groups, enabled us to determine whether enzyme activity was inhibited by rapamycin, or by ME0052.

5.15.2 *Characterisation of deletion mutants*

The importance of the periplasmic chaperone SurA has been well documented, as it is reported to be the major chaperone of outer-membrane proteins [215–217]. *surA* has been deleted in several species of Gram-negative bacteria and results in a similar phenotype, namely an altered OM protein profile, an increased sensitivity to several extra-cellular stresses, and a decrease in virulence [204,215]. Although the virulence of *surA* deletion mutants is compromised, the mechanism behind this has not been attributed to the reduction of LEE T3S [204,241]. Since the addition of the SA compounds reduces LEE T3S and increases flagellar expression these two virulence factors were the focus of this investigation.

The expression of the LEE T3SS genes in the Δ*surA* mutant was investigated using GFP:promoter fusions reporting on the expression of the LEE operon. This showed that the expression of LEE1, LEE2, LEE3 and LEE5 (*tir* reporter) was increased compared with the WT. The expression of the control reporter fusion *rpsM* was not affected in the mutant, indicating that there is not a general increase in the level of gene expression. This corresponded to an increase in the level of protein secretion under LEE T3S inducing conditions. The increase in LEE T3S expression is surprising, since deletion of *surA* induces the envelope stress response which acts to reduce the expression of membrane proteins and to increase the expression of folding chaperones and proteases [205]. There are no previous reports showing an increase in LEE T3S following *surA* deletion. Fardini *et al.* (2009) report that deletion of *surA* had no effect on LEE T3S in *Salmonella* [242]. This was also shown to be the case in *Y. pseudotuberculosis* [203]. The reason for the up-regulation of the expression of the LEE T3SS observed in this investigation is unclear and requires further analysis.

Analysis of the ability of Δ*surA* to bind to host cells showed that although the mutant was still capable of attaching to host cells using the LEE T3SS, the efficiency of the process was decreased. This was seen by the reduced density of the actin pedestal formed by the mutant which may imply that the LEE T3SS is compromised.

Another interesting observation was that the expression of *fliC* was increased in the Δ*surA* mutant when grown under LEE T3SS inducing conditions. The reason for this is unclear and there are no reported links between the expression of SurA and flagellar gene expression. When grown in rich media (LB) which would usually promote flagellar production, the expression of *fliC* is equivalent in the WT and the mutant. However, analysis of the motility of the Δ*surA* mutant showed that this mutant was non-motile, indicating that the flagella produced by this mutant may not be functional. However, further investigation of this phenotype is required.

Studies of the Δ*fklB* deletion mutant showed no clear phenotype. The growth rate of this mutant was unaffected. There were no differences in the expression of *LEE* or flagellar genes, which was supported by the finding that there was also no difference in the LEE T3S protein profile. Overall this suggests that the role of FklB in *E. coli* O157 is not essential. There are several PPIase proteins encoded in Gram-negative bacteria and it is

thought that these proteins have overlapping functions [204]. This functional redundancy means that deletion of one of the PPIases is unlikely to have a large effect. This was shown in the study by Obi *et. al* (2011) where a quadruple PPIase mutant, including FklB, in *Y. pseudotuberculosis* did not produce a strong phenotype [204].

5.15.3 *Binding studies*

Rapamycin is an inhibitor of the FKBPs, therefore this was used as a control to determine whether the purified FklB was inhibitable by rapamycin. At the time that the assays were conducted this finding had not been published, however Obi *et al.* reported the inhibition of FklB shortly after these experiments [204]. Ideally, the enzyme assay would have conclusively determined whether ME0052 bound to FklB and inhibited its PPIase activity. However, owing to the absorptive properties of the compounds their presence interfered with the enzyme assay and therefore a full range of concentrations could not be tested. Analysis of the rate of the reaction in the presence of ME0052 appeared to indicate that ME0052 was inhibiting the enzyme activity. However, this evidence was not conclusive, thus it was further investigated using NMR.

^{1}H-^{15}N HSQC of FklBc revealed that ME0052 does not bind to the C-terminal of FklB, since no chemical shifts were seen upon the addition of ME0052 compared with the control. The fact that the addition of DMSO caused several chemical shifts in the HSQC spectrum of the protein indicated that that DMSO may be specifically interacting with the protein. Attempts to solubilise the compounds in other solvents, such as methanol, were unsuccessful. Since the active site of FklB is essentially a hydrophobic pocket, we speculated that DMSO interacts with this region of the protein. If this were the case then DMSO may interfere with the binding of ME0052 to FklBc. To determine whether the compound bound to the full-length protein, ITC was performed. The reaction showed no indication that ME0052 bound to full-length FklB, ruling out the possibility that it interacts with the N-terminal dimerisation domain. Thus it appears that the initial far-western result indicating that FklB bound to the SA compounds was misleading.

NMR studies of the interactions between rapamycin and FklBc showed that addition of the compound induced several chemical shifts. Owing to the low solubility of the compound, binding saturation could not be reached, therefore both the bound and

unbound conformation of the protein was observed in the spectra. Backbone assignment of the spectra based on the structure of the homology model is currently underway as part of a follow-on project. This will reveal which residues are involved in binding to rapamycin. The residues involved in binding rapamycin have been determined for other PPIases and as some of these residues are highly conserved amongst the PPIases it is likely that rapamycin will bind to the corresponding residues in FklB. However, experimental determination of these residues is preferable to homology modelling and is therefore being conducted.

5.15.4 *The structure of FklB*

The oligomeric state of FklB was investigated using AUC. This revealed that FklB was present almost exclusively as a dimer in solution, characterised by a low K_d of dimerisation. Confirmation of the oligomeric state of FklB was an important precursor to further biophysical studies, including SAXS, which further confirmed that FklB forms a dimer in solution. *Ab initio* modelling using the SAXS data gave a model that fitted well with the homology model in terms of overall shape. From this model it was clear that flexibility in the helical region of FklB could allow the C-terminal regions to be present in several different orientations. Crystal structures of similar proteins also indicate that the extended helix connecting the N- and C-terminal domains exhibit some flexibility. The flexibility of this helix was modelled using the SAXS data, which further supported the hypothesis that FklB has a degree of flexibility similar to that of other proteins in this class [243]. The flexibility of the connecting helix has been shown to be important to the chaperone activity of FkpA, therefore this may also be the case for FklB [243,244].

5.15.5 *Conclusions*

Three prolyl *cis*/*trans* isomerases were identified as being putative binding targets of the salicylidene acylhydrazides. The fact that FkpA, SurA and FklB contain similar enzymatic activity was an attractive prospect, which implied that the SA binding site was might be within this enzymatic domain. However, it would appear that these three proteins were false positives from the pull-down assay as no direct confirmation for their binding to the SA compounds was obtained in this investigation. Furthermore, the phenotype observed following deletion of *surA* and *fklB* did not reflect the phenotype seen following

treatment of whole cells with the SA compounds. Deletion of *surA* led to an increase of both LEE T3S and flagellar gene expression under LEE T3S inducing conditions as opposed to a decrease in LEE T3S and an increase in flagellar production. Whereas deletion of *fklB* had no effect on these processes. The role of FklB in other cellular processes has yet to be determined.

6 Is AdhE a target of the salicylidene acylhydrazides?

6.1 Introduction

Another target protein identified in the SA compound affinity pull-down assay was AdhE. Since AdhE is a metabolic enzyme with only tenuous links to virulence, this protein was initially overlooked as playing a role in the action of the SA compounds. However, further investigation of the role of this enzyme in *E. coli* O157 revealed interesting insights into the regulation of the LEE T3SS and flagellar expression.

6.1.1 *The function and regulation of AdhE*

Previous work on AdhE has focused on its role in anaerobic conditions where this multi-functional enzyme is essential for fermentation of glucose, leading to the production of ethanol. The conversion of acetyl-CoA to ethanol is carried out in a two-step reaction performed by the two enzymatic domains of AdhE. The N-terminal domain of AdhE is an acetaldehyde dehydrogenase that converts acetyl-Co A to acetaldehyde, which is then converted to ethanol by the C-terminal alcohol dehydrogenase domain of AdhE (Figure 6-1) [245]. AdhE is also reported to have a third enzymatic function as a pyruvate-formate lyase [246], an enzyme that catalyses the conversion of pyruvate and coenzyme A to formate and acetyl-CoA, however this has recently been questioned [247].

The switch from aerobic to anaerobic respiration that takes place in facultative organisms, such as *E. coli*, leads to the induction of a large number of metabolic enzymes [248]. One of these enzymes is AdhE, the expression of which has been shown to increase by 10- to 20-fold in anaerobic conditions [249,250]. The expression of AdhE is mediated by several factors within the cell, indicating that controlling the expression of this enzyme is important. The promoter region of AdhE contains several binding sites for transcriptional regulators [251]. AdhE expression is repressed by binding of the catabolite repressor activator protein (CRP) to a sequence upstream of the *adhE* gene [250]. Several metabolic enzymes are controlled by CRP since this regulator is responsive to the metabolic status of the cell [252]. Another global regulator, Fis, has been reported to negatively regulate *adhE* expression [251] as has NarL (in the presence of nitrate) [249]. The stress response sigma factor RpoS has been shown to increase the expression of *adhE* upon the bacteria reaching stationary growth phase. However, the trigger for the up-

regulation of AdhE in anaerobic conditions does not appear to be the lack of oxygen *per se*, but the increase in the levels of NADH caused by the loss of oxygen as the terminal electron acceptor [253]. AdhE expression undergoes further control at the post-transcriptional level through the formation of an secondary RNA structure that must first be cleaved by RNase III to expose the ribosome binding site thus allowing translation of AdhE [254].

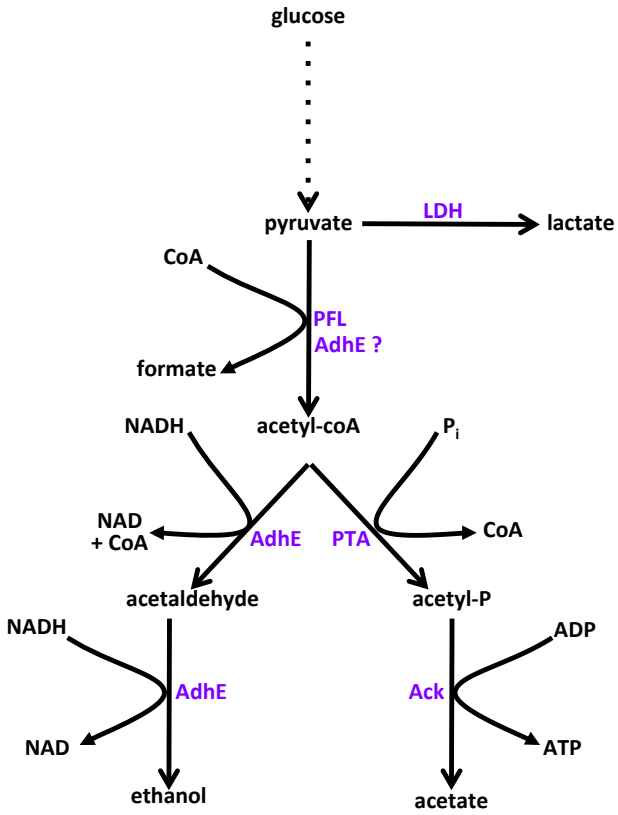


Figure 6-1 The role of AdhE in metabolism. AdhE is involved in the terminal reactions of the glycolytic pathway. The dashed arrow indicates the reactions that take place in glycolysis leading to the production of pyruvate. Pyruvate is then converted to acetyl-coenzyme A (acetyl-CoA) by pyruvate formate lyase (PFL). AdhE may also have this enzymatic function. During glucose fermentation acetyl-CoA can be converted to acetate through the actions of phosphate acetyl transferase (PTA) and acetate kinase (Ack); or acetyl-CoA can be converted to acetaldehyde and ethanol by the bi-functional enzyme AdhE.

Initial studies overlooked the expression of AdhE in aerobic conditions, which although considerably lower than in anaerobic conditions, is still significant as AdhE accounts for 1% of total cellular protein [255]. Furthermore, AdhE is enzymatically active in aerobic conditions, albeit at a slightly reduced rate. Pineda *et al.* showed that inhibition of AdhE had a significant effect on the metabolic flux in *Entamoeba histolytica* in aerobic

conditions. This may suggest that during oxidative respiration AdhE is important in regulating the levels of acetyl-CoA in the cell [256].

6.1.2 *The structure of AdhE*

Structurally, AdhE is interesting. Previous studies have shown that this 96 kDa protein oligomerises to form long filaments which can be visualised by EM (Figure 6-4). This was first reported by Kessler *et al.* who also showed that the arrangement of these filaments was influenced by the presence of the cofactors NAD and Fe^{2+}, which are required for enzyme activity [246]. The physiological role of these filaments, also called spirosomes, is unknown. Since AdhE binds to iron it is susceptible to metal-catalysed oxidation, which significantly reduces the half-life of AdhE in the cell [257]. Furthermore, AdhE has been identified as one of the main targets of oxidative damage in *E. coli* [258]. It has been suggested that under aerobic conditions, AdhE actually acts to protect the cell from oxidative damage since Δ*adhE* mutants have an increased susceptibility to H_2O_2 [255].

This unusual bi-functional enzyme is thought to be the result of evolutionary gene fusion and is found in several other species of Gram-negative bacteria including *Salmonella*, *Yersinia* and *Shigella*. AdhE is also found in Gram-positive bacteria such as *Clostridium* species and in the protozoan parasite *Giardia* (Figure 6-2). Several of the species shown in Figure 6-2 are human pathogens, which potentially highlights AdhE as an attractive drug target. Furthermore, AdhE has a low sequence similarity to human proteins. The acetaldehyde domain of this enzyme has 24% sequence identify to the human acetaldehyde dehydrogenase 1 enzyme and the alcohol dehydrogenase domain is most similar to a human hydroxyacid-oxoacid transhydrogenase (26% identity), further highlighting this protein as a possible drug target.

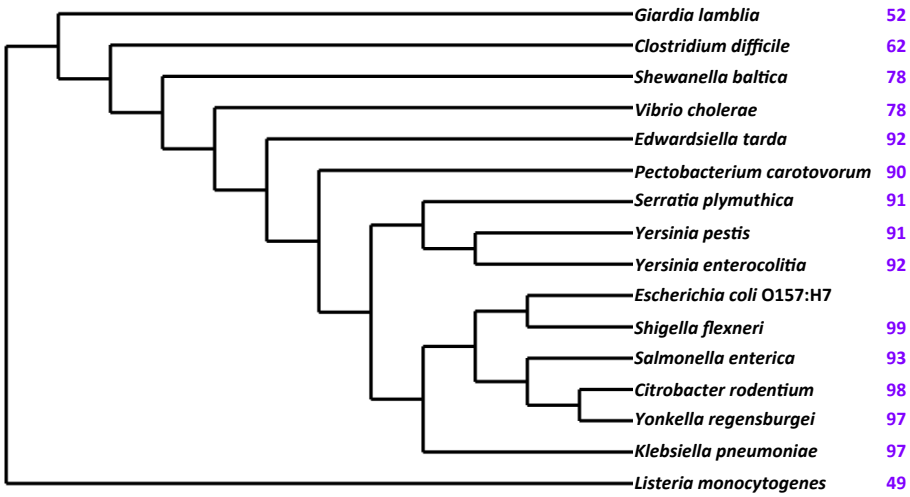


Figure 6-2 AdhE is found in several pathogens. Cladogram generated using Phylogeny.fr [259] based on a BlastP search of the *E. coli* O157:H7 AdhE protein sequence. The numbers in purple indicate the percentage of sequence identity to the *E. coli* protein.

6.1.3 *AdhE and its links to virulence*

As mentioned above, there have been only a few reports linking AdhE with bacterial pathogenicity. Characterisation of transposon mutants in *Salmonella typhimurium* with a diminished capacity to survive murine macrophages mapped one transposon to an insertion in *adhE* [260]. Furthermore, a more recent study showed that deletion of two genes involved in pyruvate metabolism in *Salmonella*, *pfiL* and *adhE*, affected SPI-1-mediated gene expression and infectivity [261]. Both of these reports suggest a link between AdhE and the expression of the LEE T3SS, however the role of AdhE in virulence has yet to be elucidated. As AdhE is a metabolic enzyme it is likely that any links between AdhE and virulence expression relate to metabolic changes within the cell. The metabolic state of the cell has been shown to affect the expression of the LEE T3SS in *P. aeurginonsa* [262].

6.1.4 *Aims of this chapter*

This chapter aimed to investigate whether AdhE contributed to LEE T3S in *E. coli* O157. The effects of deletion of this gene were investigated using several approaches. The data presented in this chapter form the basis of a publication currently submitted for peer review:

- Beckham, K.S.H., Connolly, J.P.R., Ritchie, J.M., Wang, D., Gawthorne, J.A., Tahoun, A., Gally, D.L., Burgess, K., Burchmore, R.J., Smith, B.O., Beatson, S.A., Byron, O., Douce, G.R., and Roe, A.J. The Metabolic Enzyme AdhE Controls the Virulence of Escherichia coli O157:H7. Submitted to Molecular Microbiology, Feb 2014.

6.2 Characterisation of AdhE

6.2.1 *Cloning, expression and purification of AdhE*

The *adhE* gene from *E. coli* was amplified by PCR using gDNA as a template, the restriction sites *Nde1* and *Nco1* were added to the gene sequence using the primers listed in Table 2-5. This was similarly performed for both of the enzymatic domains of AdhE (Figure 6-3A). The PCR product was ligated directly into the Strataclone vector using the Strataclone PCR kit and the ligation mix was transformed into the cells provided. Transformants were checked for the presence of the insert by digestion with *Nde1* and Nco1. Purified plasmid containing the insert was transformed into BL21(DE3) cells for protein over expression. Proteins were purified using IMAC and SEC and the purity of the final product was analysed by SDS-PAGE (Figure 6-3B & C).

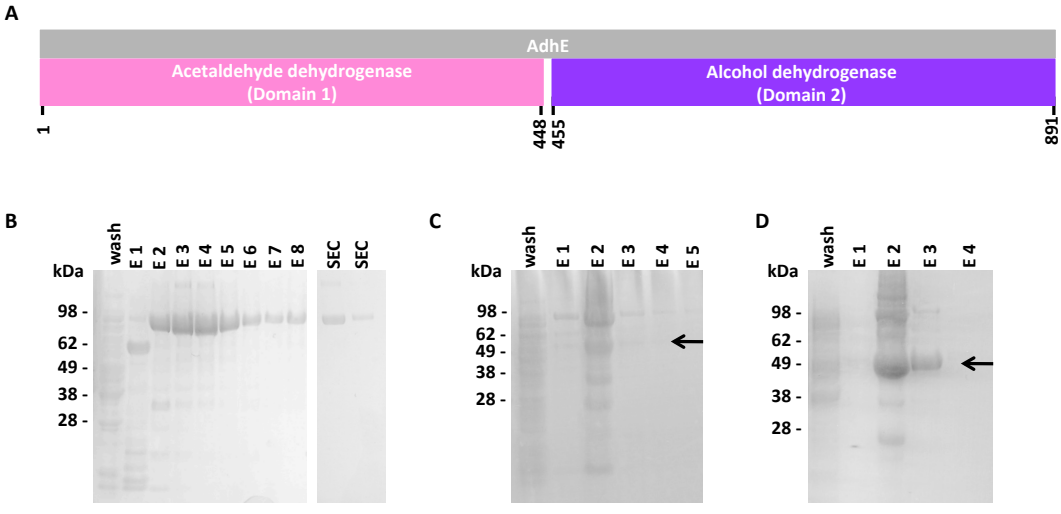


Figure 6-3 Purification of AdhE and its enzymatic domains. (A) The arrangement of AdhE indicating the position of the 2 catalytic domains – domain 1 being the acetaldehyde dehydrogenase domain and domain 2 being the alcohol dehydrogenase domain. Numbers correspond to the amino acid residues of the protein and show the part of the gene inserted into the expression construct. **(B)** Purification of AdhE, domain 1 **(C)** and domain 2 **(D)** shown with different elution fractions following IMAC (E) or size exclusion chromatography (SEC), the wash with 20 mM imidazole is also shown. The position of the molecular weight markers is indicated (kDa). Arrows indicate the protein of interest where it is difficult to see in the image.

The initial aim of protein over-expression and purification was to carry out further biophysical characterisation. However, stabilisation of the protein in solution proved difficult and the protein precipitated. Dialysis into several different buffers did not overcome this issue. Attempts were made to crystallise AdhE, however this was unsucesful. Since the protein is known to be susceptible to degradation through metal catalysed oxidation, structural studies on this protein were largely discontinued.

However, attempts to reproduce and image the spirosomes seen by Kessler *et al*. were successful (Figure 6-4). Electron microscopy was done at the University of Edinburgh with Dr Chris Kennaway. As reported previously, purified AdhE oligomerised to form helical filaments of repeating dimer units [246]. The filaments obtained were heterogeneous in terms of the length and shape and were therefore not ideal for more sophisticated structural analysis, using techniques such as cryo-EM, since the level of resolution that can be obtained depends on the similarity of the averaged molecules. The sample tested retained the his-tag used for purification and this may have affected the oligomerisation of the protein or destabilised the spirosome structure leading to heterogeneity. The addition of cofactors such as NAD and Fe^{2+} appeared to disrupt spirosome formation completely, which may have been caused by degradation of the protein catalysed by the presence of iron.

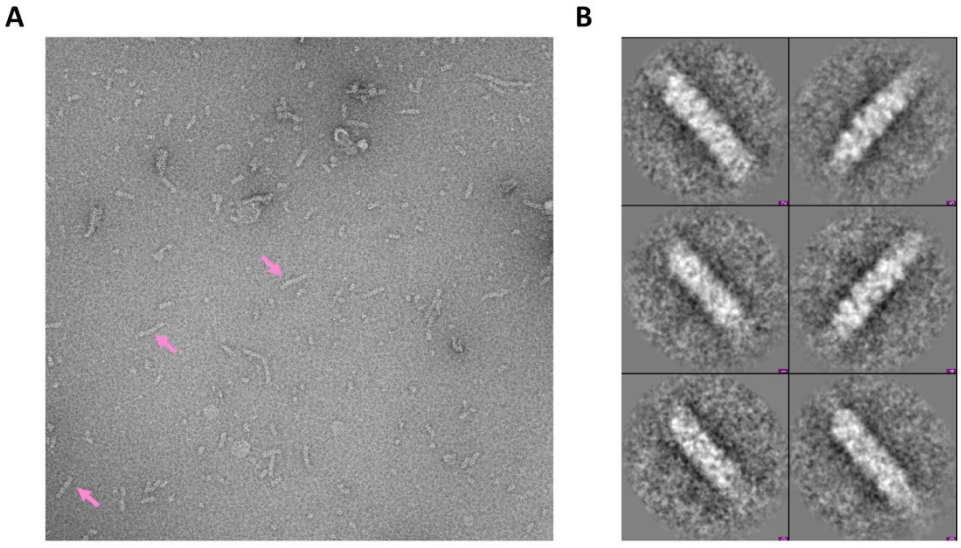


Figure 6-4 Negative stain EM of purified AdhE. (A) Section of a raw micrograph of AdhE stained with 2% uranyl acetate on carbon film at 58,000x magnification. Pink arrows highlight individual spirosomes. **(B)** Image averages of 96 selected filaments following alignment and classification. The length of the filaments is ~150 nm.

6.2.2 *AdhE shows binding to the SA compounds by far-western*

The initial pull-down assay with the SA compounds identified AdhE as a target protein [100]. However in order to determine whether AdhE does in fact bind to the SA compounds a far-western blot was performed using a biotinylated form of ME0052. The far-western was conducted on the purified full-length protein, domain 1 (D1) and domain 2 (D2) to assess whether AdhE binds to the compound and if any differences between compound binding to the two different domains could be detected. A concentration of 100 µM of protein was used in the far-westerns where possible and Figure 6-5A shows a Coomassie stained SDS-PAGE gel of the samples used in the far-western indicating the purity of the proteins. However, D2 could not be concentrated to 100 µM since it appeared to bind to the membrane of the concentrator and in addition would precipitate out of solution at higher concentrations. Therefore a concentration of only 10 µM of D2 was used, i.e. the concentration was 10-fold lower than that of the full-length protein or D1. Far-western analysis confirmed that the full-length protein bound to the compound. Interestingly, there was no signal corresponding to ME0052-bio binding to D1, but there was a strong signal corresponding to D2 binding (Figure 6-5B). This is despite the fact the concentration of D2 was 10-fold lower than that of full-length AdhE and D1. This result suggests the SA compounds bind preferentially to D2 of AdhE. To try to confirm this result using a second binding method, ITC was performed on AdhE and ME0052. However, owing to the instability of the protein and the poor solubility of the compound, this was unsuccessful.

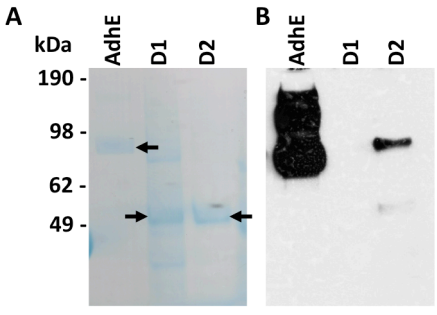


Figure 6-5 Far-western analysis of the binding of AdhE to ME0052. (A) An SDS-PAGE gel showing protein samples used in the far-western **(B).** Arrows correspond to the protein of interest. The position of the protein molecular weight markers has been indicated (kDa).

6.3 Characterisation of Δ*adhE* in *E. coli* O157

6.3.1 *Deletion of adhE*

An *adhE* knockout mutant was generated by allelic exchange by Dr Dai Wang. All further analysis of the Δ*adhE* mutant was conducted as part of this thesis.

The growth rate of the Δ*adhE* mutant was analysed in different media to determine whether deletion of *adhE* had a cost on overall fitness. Δ*adhE* was grown in rich medium (LB), as well as minimal media (MEM-HEPES and DMEM). The doubling times for WT and the Δ*adhE* mutant in the different media were similar: approximately 45 min in LB; 220 min in MEM-HEPES and 170 min in DMEM. There were no observable differences in the growth rate of Δ*adhE* compared with the WT.

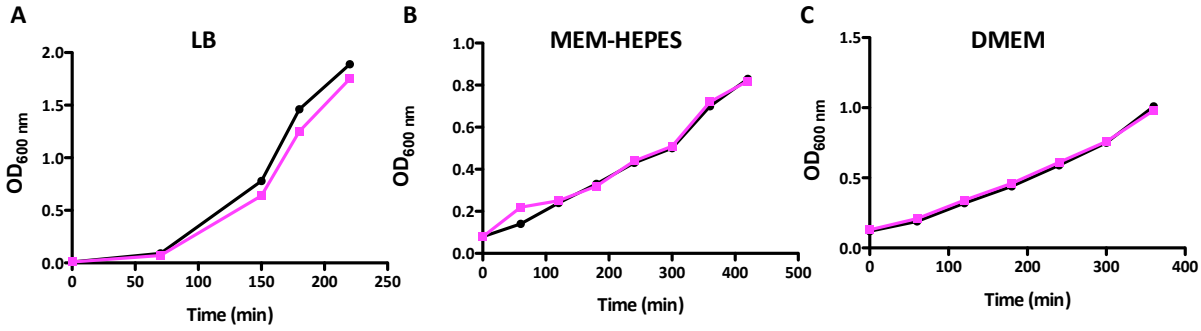


Figure 6-6 Deletion of *adhE* does not affect growth rate. The growth rate of the Δ*adhE* mutant was tested in three different media, LB **(A)**, MEM-HEPES **(B)** and DMEM **(C).** The WT and Δ*adhE* data are indicated in black and pink, respectively. The graph displayed is a representative graph of triplicate experiments.

6.4 Type three secretion is decreased in the Δ*adhE* mutant

Having identified that deletion of *adhE* did not affect the growth rate of *E. coli,* further phenotypic analyses were carried out. The phenotype of WT in the presence of the SA compounds is a decrease in LEE T3S, therefore the secretion of the Δ*adhE* mutant under LEE T3S inducing conditions was tested. To induce LEE T3S in *E. coli* O157, bacteria were cultured in MEM-HEPES and grown to an $OD_{600\,nm}$ of 0.8. Secreted proteins were precipitated and analysed using a combination of SDS-PAGE, tandem mass spectrometry and western blotting. To further investigate the degree of induction of the LEE T3SS, the

levels of gene expression were analysed using GFP-promoter fusions and RNA sequencing analysis.

6.4.1 *Effects on LEE T3SS protein expression*

SDS-PAGE analysis of the secretion profile of the Δ*adhE* mutant showed a clear difference in the secreted proteins produced by this strain. The most obvious change was the presence of a protein of around 60 kDa (labelled 3 in Figure 6-7A). This protein was identified by tandem mass spectrometry as being flagellin, based on a MOWSE score of 1065 [263]. This was an interesting result as the LEE T3SS and flagellar production are cross-regulated, i.e. when one is active the other is down-regulated. Other clear differences were the decreases in the levels of the some of the most predominant secreted proteins, namely Tir (labelled 1 in Figure 6-7A) and EspA (labelled 2 in Figure 6-7A). The secreted proteins were also analysed by western blotting (Figure 6-7B) which confirmed the decrease in the levels of secreted proteins including EspA and Tir. Additionally, analysis of the whole-cell fraction showed a decreased amount of EscJ, a protein that forms part of the basal apparatus of the LEE T3SS. Western blotting showed that the levels of GroEL, a house-keeping protein, did not change between samples indicating that equivalent amounts of cell lysate were being analysed. This also indicated that deletion of *adhE* does not have a global effect on protein expression.

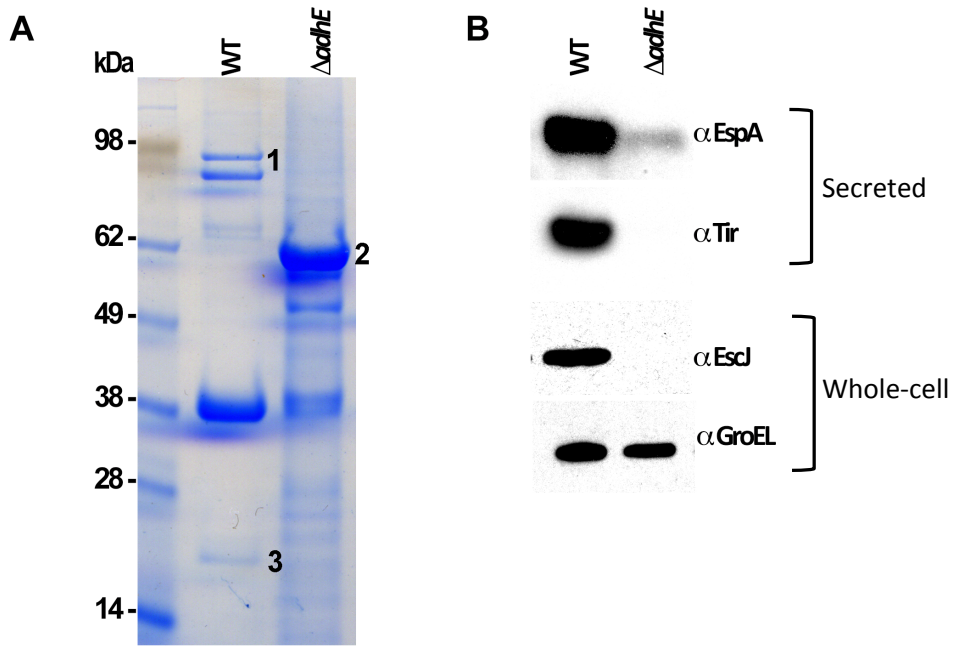


Figure 6-7 Analysis of the secretion profile of the Δ*adhE* mutant. (A) Coomassie stained SDS-PAGE gel of the secreted proteins prepared as in Section 2.4.1 from WT *E. coli O157:H7* and the *ΔadhE* deletion mutant. Numbers indicate proteins that were identified by tandem-MS mass spectrometry as follows; 1, Tir; 2, flagellin; 3, EspA**. (B)** Western blot of secreted proteins (Tir and EspA) and proteins from the cellular fraction (EscJ and GroEL).

6.4.2 *Effects on LEE T3SS gene expression*

Deletion of *adhE* has a dramatic effect on the secretion profile of *E. coli*. To determine the mechanism behind these changes it was first necessary to establish whether the deletion of *adhE* was having an effect on the transcription of the LEE T3SS genes. Gene expression was investigated using two methods, the first of which used a series of translational fusion reporter-plasmids, where the promoter regions of several of the LEE genes have been fused to *gfp*. Therefore the level of GFP fluorescence reports on the activity on the promoter for these genes. The expression of *LEE 1*, *LEE 2*, *LEE 3*, *tir* and *rpsM* was tested using promoter:fusion constructs generated previously [264]. Using this approach no changes in the level of gene expression were detected (Figure 6-8). This was an unexpected result owing to the degree of difference seen in the protein levels. However, translational fusions may not always accurately report the levels of gene transcription in all cases, for example, if a gene is regulated by the presence of a small regulatory RNA

(srRNA). srRNAs generally have an inhibitory effect on gene expression and act by binding to the mRNA encoding target genes, which prevents their translation either by blocking ribosome binding or by decreasing the half-life of the mRNA [265].

Second, owing to potential limitations of the promoter:GFP fusions, RNA sequence analysis was performed as an alternative method to look at the levels of gene expression. RNA sequencing of gene transcripts offers a global outlook on changes in gene expression between strains. RNA sequencing of the WT and the Δ*adhE* mutant enabled the analysis of changes in the transcript levels of all genes of the LEE T3SS. This analysis revealed that although there was a general trend towards a down-regulation of LEE T3SS genes, the fold-changes observed were not significant (Table 6-1). Therefore, the RNA sequencing data are in agreement with the expression data from the promoter fusions. Since the level of LEE T3S protein production, as seen in the secreted protein profile, is greatly decreased but the levels of gene expression remain the same this implies that regulation is occurring post-transcriptionally.

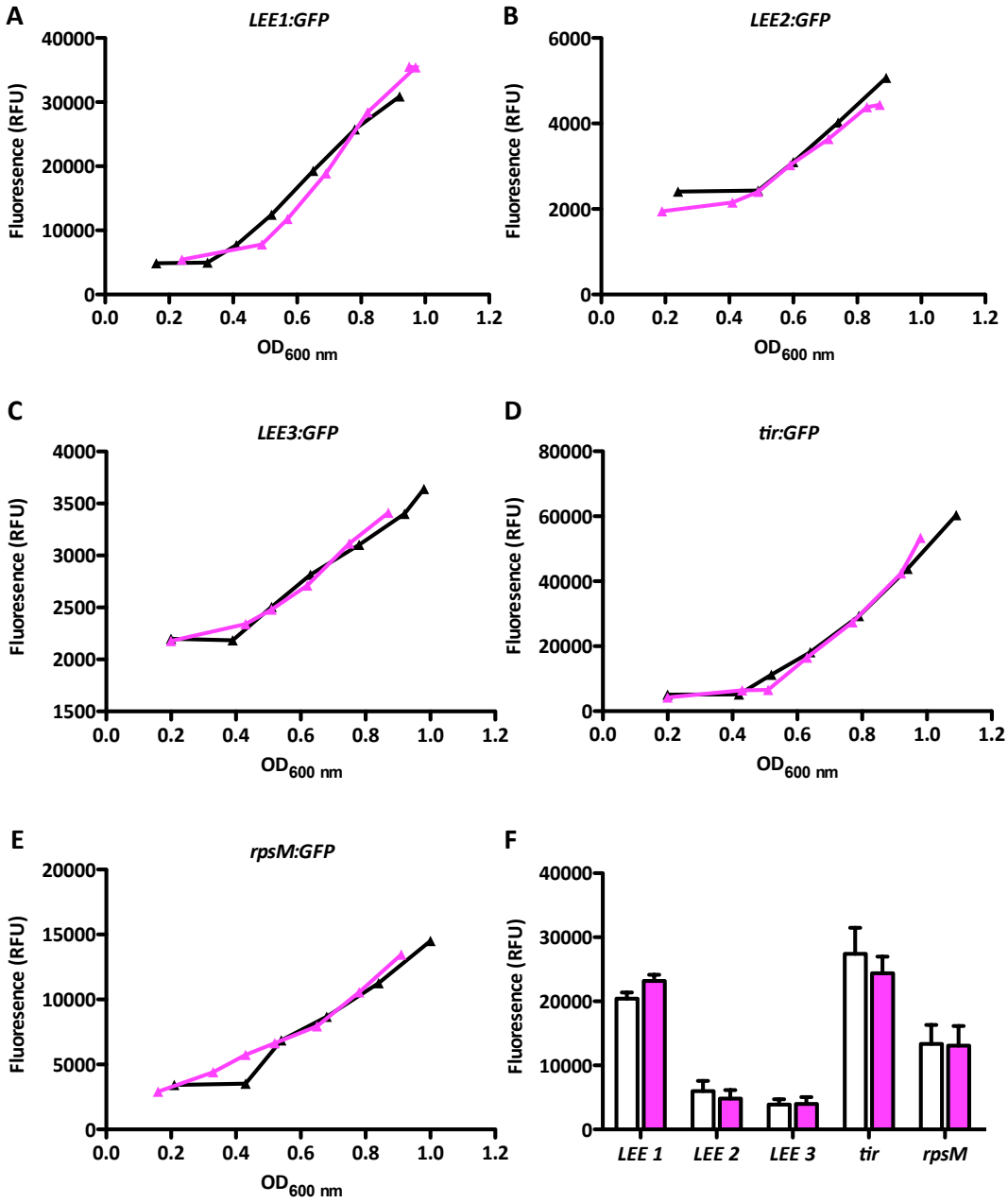


Figure 6-8 Expression of the LEE operons in the Δ*adhE* mutant and the WT. The expression of GFP under the promoter of *LEE1* **(A)**, *LEE2* **(B),** *LEE3* **(C),** *tir* **(D)** and *rpsM* **(E)** monitored over time and compared with the optical density at 600 nm ($OD_{600\ nm}$) for the WT (black) and the Δ*adhE* mutant (pink). The level of fluorescence at $OD_{600\ nm}$ = 0.7 was calculated and compared between samples. A Student's t-test showed there was no significant difference between the levels of expression of the Δ*adhE* mutant compared with the WT. Parts (A)-(E) are representative graphs of triplicate experiments and part **(F)** includes the triplicate data. WT data are in black and data for the Δ*adhE* mutant are in pink.

6.5 The expression of flagella is up-regulated in the Δ*adhE* mutant

6.5.1 *Effects on flagellar protein expression*

As mentioned above, analysis of the secreted protein profile of the Δ*adhE* mutant revealed that there was an increase in the levels of the flagellin produced under LEE T3SS-inducing conditions (Figure 6-7). This was further analysed by western blotting for the FliC protein, which forms the flagellar filament, and for the flagellar gene specific transcription factor σ^{28}. Both were detected only in the secreted proteins and the whole cell proteins, respectively, in the Δ*adhE* mutant.

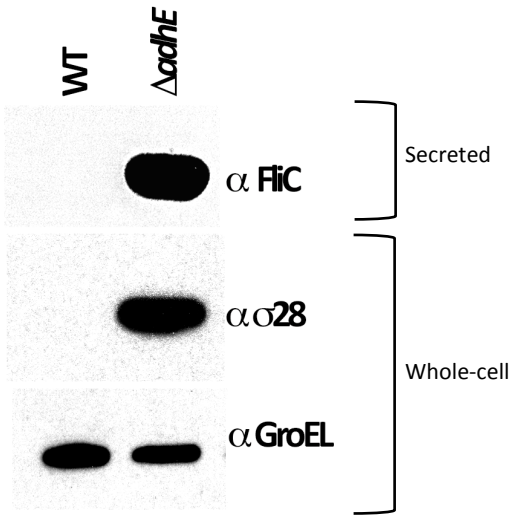


Figure 6-9 Expression of flagellar proteins in the Δ*adhE* mutant. Western blot showing the amount of FliC present in the secreted proteins from WT and the Δ*adhE* mutant and the amount of σ^{28} and GroEL in the whole cell fraction.

Gene name	WT (RPKM)	WT sd	Δ*adhE* (RPKM)	Fold-change (p>0.005)
espF	123.57	155.56	90.12	1.37
Z5102	139.34	158.63	109.35	1.27
escF	177.31	191.86	115.02	1.54
Z5104	886.29	869.38	494.59	1.79
espB	4,564.53	2,511.34	1,851.13	2.47
espD	3,560.95	2,674.13	1,545.96	2.30
espA	4,267.20	3,181.34	1,967.13	2.17
sepL	139.43	121.39	83.53	1.67
escD	33.61	23.62	20.64	1.63
eae	775.99	760.52	664.49	1.17
Z5111	457.86	563.45	478.78	0.96
tir	722.28	539.57	407.76	1.77
Z5113	160.87	145.96	80.47	2.00
Z5114	44.10	59.60	30.66	1.44
Z5115	110.16	124.89	50.69	2.17
sepQ	41.51	40.62	30.71	1.35
Z5117	12.44	14.72	15.62	0.80
Z5118	20.97	19.10	14.91	1.41
escN	45.31	37.27	28.32	1.60
escV	30.54	17.69	13.33	2.29
Z5121	5.72	2.58	1.87	3.06
sepZ	871.51	638.05	445.48	1.96
Z5123	99.65	91.65	55.27	1.80
escJ	36.81	32.69	14.18	2.60
Z5125	34.64	31.53	39.27	0.88
escC	42.39	42.02	33.62	1.26
cesD	43.05	48.28	33.82	1.27
Z5128	46.87	49.85	27.23	1.72
Z5129	84.39	101.72	66.41	1.27
Z5131	5.12	8.13	5.06	1.01
escU	4.25	5.64	4.15	1.02
escT	4.66	7.77	2.99	1.56
escS	7.21	10.68	4.30	1.68
escR	9.60	15.85	8.87	1.08
Z5136	30.68	48.37	16.45	1.87
Z5137	51.61	77.99	39.24	1.32
Z5138	47.71	70.07	35.82	1.33
Z5139	44.93	64.33	32.56	1.38
Z5140	63.58	100.35	72.28	0.88
Z5142	47.61	71.81	27.15	1.75
Z5143	3.27	4.64	2.02	1.62
House-keeping				
groEL	225.73	129.74	179.12	1.2
gapA	793.80	437.90	1,106.77	0.7
dnaJ	18.08	9.01	19.77	0.9
gyrA	30.27	15.53	34.54	0.8
arcA	91.75	29.99	97.03	0.9
rrsA	7,550.23	962.02	6,636.51	1.1
fumC	5.58	1.68	3.89	1.4
rpoS	666.00	136.19	451.89	1.4
recA	76.33	18.68	82.51	0.9

Table 6-1 RNA sequencing data showing the expression of the LEE T3SS genes. The average WT Reads per Kilobase Mapped (RPKM) from four experiments has been shown with the standard deviation (sd) of the data for LEE T3SS genes and house-keeping genes. The fold-change was calculated by dividing the WT RPKM by the *ΔadhE* RPKM values. The levels of significance according to the Levene's t-test indicated a p value of >0.005 and that therefore the change in RPKM value was not significant.

6.5.2 *Effects on flagellar gene expression*

The expression of the flagellar system was analysed using both a *fliC*:GFP fusion and RNA sequencing data in a manner similar to the analysis performed for the LEE T3SS. The *fliC* reporter indicated that the level of *fliC* expression was higher in the Δ*adhE* mutant and the WT when grown in both LB and MEM media (Figure 6-10 A & B). Comparison of the level of gene expression at an $OD_{600\ nm}$ = 0.7 revealed that the differences in the level of *fliC* expression were significant (Figure 6-10C).

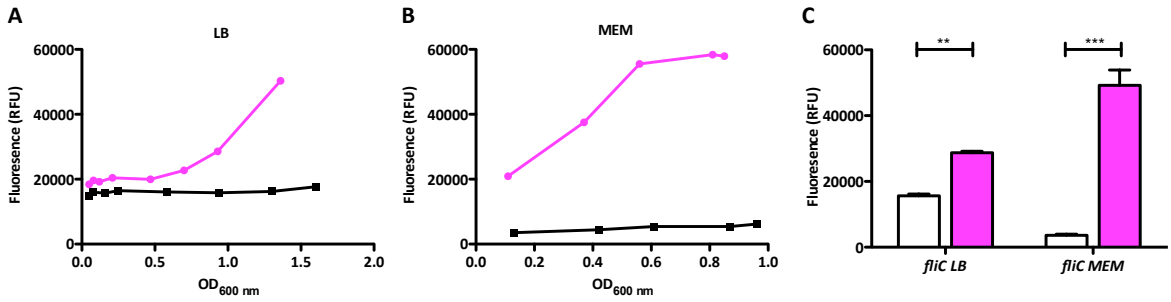


Figure 6-10 Expression of a *fliC* promoter:gfp fusion is increased in the Δ*adhE* mutant. A representative graph showing the amount of fluorescence (RFU) plotted against the $OD_{600\ nm}$ for WT and Δ*adhE* grown in LB **(A)** and MEM **(B)** media. The level of fluorescence at an $OD_{600\ nm}$ = 0.7 from triplicate experiments was calculated and compared between samples **(C).** A Student's t-test showed that there was a significant difference between the levels of expression of the Δ*adhE* mutant compared with the WT. Asterisks denote the degree of significance: ** indicates a p-value between 0.001 to 0.01; *** indicates a p-value below 0.001. WT data are in black and data for the Δ*adhE* mutant are in pink.

The *fliC* reporter showed that the level of gene expression correlated well with the levels of FliC detected by western blotting. However at this point it was not known if all flagellar genes were being up-regulated or whether *fliC* was being targeted specifically. However the increase in the levels of σ^{28} shown by western blotting suggested that several flagellar genes may be affected.

Analysis of the RNA-sequencing data revealed that the expression of all of the flagellar genes was markedly increased in the Δ*adhE* mutant (Table 6-2) 43 of the 48 flagellar genes were shown to be significantly up-regulated compared with the WT and only one gene was shown not to increase significantly, namely *fliR*, a membrane component of the flagellar complex. However, the levels of transcript for this gene are low and therefore the values obtained may not be reliable.

Gene name	WT (RPKM)	WT Sd	ΔAdhE (RPKM)	Fold-change	Significance
Class 1					
flhC	4.00	2.76	60.14	15.04	y
flhD	6.82	6.35	97.18	14.25	y
Class 2					
flgN	0.77	1.03	35.79	46.78	y
flgM	1.05	0.71	55.83	53.43	y
flgA	1.25	0.67	63.56	50.75	y
flgB	-	-	227.05	-	na
flgC	0.10	0.19	207.98	2,189.26	y
flgD	0.26	0.31	256.58	977.45	y
flgE	0.46	0.38	313.34	688.66	y
flgF	0.24	0.28	164.06	690.78	y
flgG	0.31	0.38	132.14	426.26	y
flgH	0.41	0.50	55.03	135.88	y
flgI	0.17	0.20	35.79	213.67	y
flgJ	0.26	0.42	20.07	77.19	y
flgK	0.75	0.43	114.17	152.74	y
flgL	1.59	1.15	215.70	136.09	y
flhA	0.84	0.26	12.12	14.51	y
flhB	0.28	0.19	11.98	42.41	y
fliY	50.78	38.02	194.59	3.83	y
fliZ	0.28	0.33	109.98	389.31	y
fliA	0.05	0.11	188.15	3,583.81	y
fliD	0.37	0.36	158.46	428.27	y
fliS	0.79	0.53	163.80	208.00	y
fliT	1.03	0.15	117.34	113.92	y
fliE	-	-	45.27	-	na
fliF	0.06	0.11	22.49	408.91	y
fliG	0.15	0.31	37.46	245.64	y
fliH	0.52	0.43	34.03	65.13	y
fliI	0.19	0.22	12.31	65.65	y
fliJ	-	-	13.82	-	na
fliK	0.43	0.22	10.58	24.60	y
fliL	0.22	0.45	24.96	112.18	y
fliM	0.18	0.21	28.05	158.03	y
fliN	0.18	0.37	30.44	166.79	y
fliO	0.11	0.21	6.34	60.38	y
fliP	0.10	0.21	5.84	56.98	y
fliQ	-	-	2.46	-	na
fliR	0.64	0.19	1.05	1.63	ns
Class 3					
fliC	2.30	2.19	4,623.80	2,008.17	y
tar	0.53	0.28	235.75	446.92	y
tap	0.49	0.35	243.85	500.21	y
cheW	1.23	1.05	305.96	248.75	y
cheA	0.27	0.30	170.20	624.59	y
motB	0.20	0.23	45.61	233.90	y
motA	0.26	0.52	68.53	266.14	y
cheZ	1.97	0.70	531.63	270.55	y
cheY	2.59	1.84	529.75	204.73	y
cheB	0.75	0.40	118.28	158.23	y
cheR	0.39	0.51	134.23	344.18	y

Table 6-2 RNA sequencing data showing the expression of flagellar genes. The average WT Reads per Kilobase Mapped (RPKM) from four experiments has been shown with the standard deviation (sd) of the data. The fold-change was calculated by dividing the WT RPKM by the Δ*adhE* RPKM values. The levels of significance according to the Levene's t-test have been shown, 'y' indicates a p-value of less than 0.005, 'ns' denotes a non-significant value, and 'na' indicates that a p value could not be calculated as there were no reads detected corresponding to this gene in the WT data. Flagellar genes have been ordered according to their flagellar class.

6.6 Complementation of the Δ*adhE* mutant

To demonstrate that the effects on LEE T3S and flagellar production were the results of *adhE* deletion and not other spontaneous mutations that may have occurred in other genes, the *adhE* gene was complemented. A low-copy number plasmid (pWSK29 background) encoding the *adhE* gene under a constitutive promoter was transformed into Δ*adhE*. A secretion assay with the complement strain showed that the plasmid partially restored the secretion profile of the Δ*adhE* mutant and western blotting for the flagellar protein FliC showed that the complement no longer expressed FliC under LEE T3SS inducing conditions. However, other LEE T3SS proteins could not be detected. This may indicate that the expression of the LEE T3SS had not been restored to its normal levels.

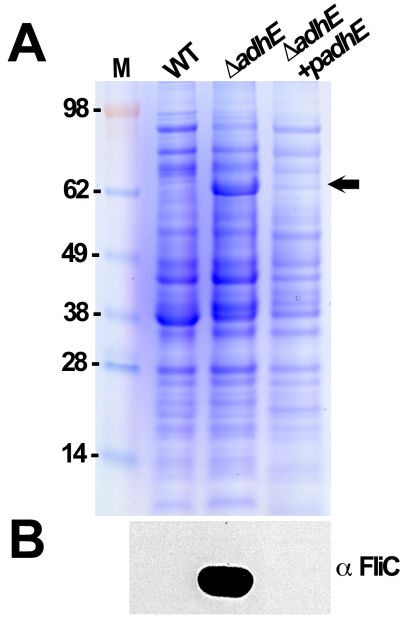


Figure 6-11 Complementation of Δ*adhE*. (A) SDS-PAGE gel showing the LEE T3S secretion profiles of the WT, Δ*adhE* and Δ*adhE* transformed with the complementation plasmid p*adhE*. Bacteria were cultured in MEM-HEPES to an $OD_{600\ nm}$ of 0.8, as described in Section 2.4.1. The arrow indicates the protein band corresponding to FliC which can clearly be seen only for the Δ*adhE* mutant and which is confirmed by western blotting **(B).** The position of the molecular weight markers has been indicated (kDa).

6.7 The Δ*adhE* mutant is non-motile

Since the Δ*adhE* mutant over-expresses all components of the flagellar system, it was hypothesised that this may lead to an increase in motility compared with the WT. Motility assays were carried out on soft agar (0.25%, w/v), which allows bacteria to swim through

the medium using their flagella. Bacteria were grown to $OD_{600\ nm}$ = 0.6 and 5 µl of culture was inoculated into the centre of the plate. After overnight incubation at 30°C the diameter of the growth zone was recorded (Figure 6-12). This showed that the *ΔadhE* mutant did not swim in the agar and was phenotypically comparable to the Δ*fliC* mutant, which lacks the main structural component of the flagellar filament. Swimming assays were carried out in both LB and TB media, however the Δ*adhE* mutant remained non-motile. This was a surprising result because the Δ*adhE* mutant showed an up-regulation in all flagellar genes. This prompted the question of whether the flagella were being properly assembled in the Δ*adhE* mutant.

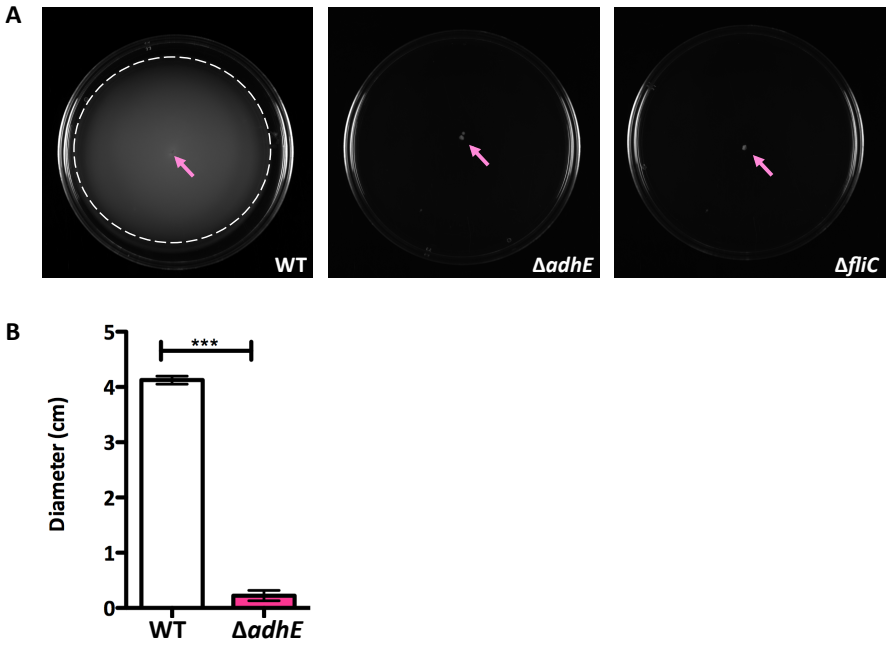


Figure 6-12 Motility assay shows that the Δ*adhE* mutant is non-motile. (A) Swimming assay on semi-solid TB agar (0.25% (w/v)) for the WT, Δa*dhE* mutant and the Δ*fliC* mutant, which has been used as a negative control. Swarm-plates were incubated overnight at 30°C as described in Section 2.4.6. The growth zone of the WT has been encircled for clarity, pink arrows point to the inoculation point / growth zone of the mutants that did not swim. **(B)** The diameters of the growth zone from triplicate experiments for WT and Δ*adhE* were measured. The asterisks indicate the degree of significance as calculated using a Student's t-test and *** indicates a p-value below 0.001.

To directly visualise the flagella, negative stain EM was used as this technique is ideal for looking at cell surface structures. Several micrographs were captured and Figure 6-13 shows representative images. Bacteria were cultured in LB media, to promote flagellar expression, prior to sample preparation. WT bacteria tended to have one flagellum per cell. In contrast the *ΔadhE* mutant had several flagella per cell, which were longer than

those observed on the WT. The length of the flagella was quantified using ImageJ [266] and the mean length of a WT flagellum was 6.1 ± 0.5 µM whereas the *ΔadhE* mutant had flagella of 9.5 ± 0.5 µM in length. A Student's t-test showed these differences to be significant with a p-value of 0.0004. Although the flagella produced were longer than the WT flagella, the *ΔadhE* mutant was in fact able to form flagella. This suggests that the lack of motility does not result from misassembly of the flagellar components.

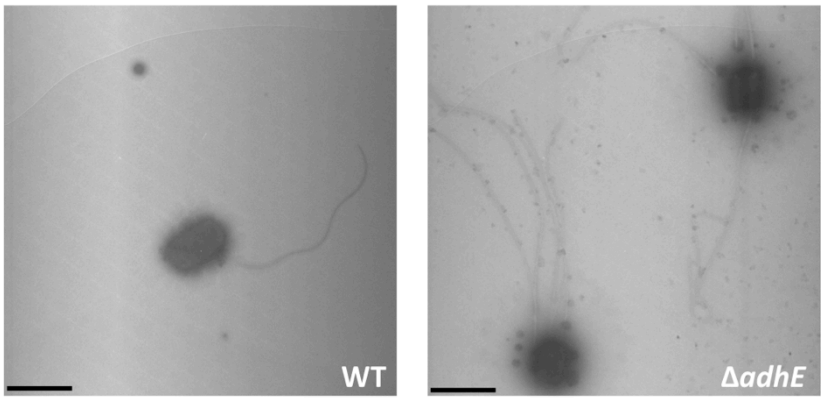


Figure 6-13 Transmission electron micrograph of WT *E. coli* O157:H7 and Δ*adhE* grown in LB medium. Bacteria were cultured in LB at 30°C to and $OD_{600\ nm}$ of 0.8. The cells were pelleted by centrifugation and the LB media removed and in PBS prior to staining with uranyl-acetate. Images were taken at 1840x magnification and the scale bar corresponds to 2 µm.

6.8 Interaction with host cells

Flagella and the LEE T3SS are both known to facilitate bacterial attachment to host cells [58,267]. Flagella are thought to initiate attachment at which point the LEE T3SS is turned on and acts to initiate intimate attachment [58]. Western blotting showed that when grown in MEM media the Δ*adhE* mutant does not express the LEE T3SS but instead expresses flagella. This was also observed by immunofluorescence (Figure 6-14A). However since both virulence factors are important for host cell binding, how the deletion of *adhE* would affect interaction with host cells was unclear, the increased number of flagella may improve binding, or the expression of the LEE T3SS may be switched on in the presence of signals from host cells which were absent in the MEM media. To test this a cell adhesion assay was conducted with the Δ*adhE* mutant to look at the number of bacteria that were attached to host cells and whether these bacteria formed A/E lesions. A/E lesions can be identified by the polymerisation of actin beneath

the bacterial cell. This is shown as an area of stronger fluorescence when compared with the rest of the cell. Additionally, to determine whether the mutant bacteria retained their flagella when attached to host cells, flagella were immunolabelled. Approximately 120 WT bacteria per field of view attached to the host cells, which is considerably higher than the number of cells that were observed to attach for the Δ*adhE* mutant (<3) (Figure 6-14D). A/E lesion formation could be clearly identified beneath the WT cells and the density of actin was considerably greater than the background level, showing a 6-fold increase compared with the mutant (Figure 6-14C).

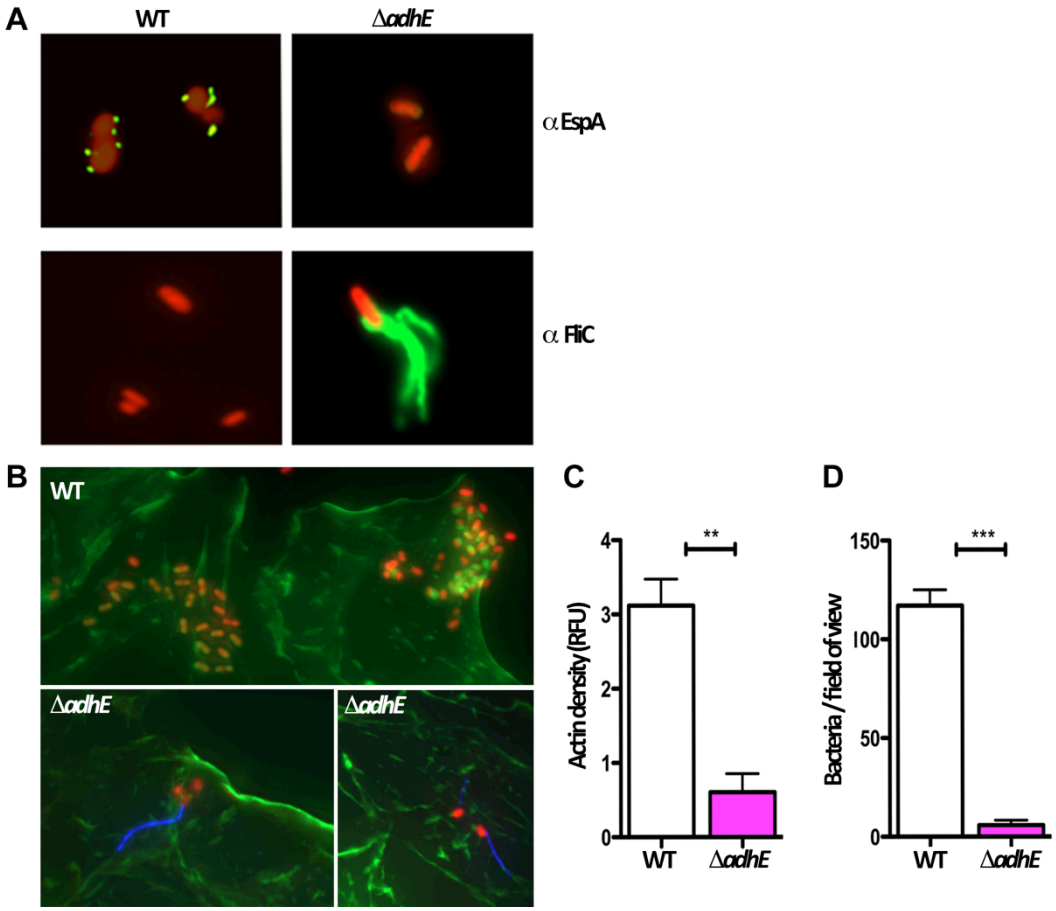


Figure 6-14 Immunofluorescence of WT and Δ*adhE*. (A) Immunofluorescence microscopy of WT *E. coli* O157:H7 and Δ*adhE* bacteria transformed with an RFP expressing plasmid (pRFP, red) probed with anti-EspA and anti-H7 (FliC) antibodies and AlexaFluor-488-conjugated secondary antibodies (green). Bacteria were cultured in MEM-HEPES to an $OD_{600\,nm}$ of 0.8. **(B)** A/E lesion formation by bacteria transformed with pRFP. Host cell actin was visualised using with FITC-phalloidin (green) as described in Section 2.6.3. Flagella were detected using an anti-H7 primary antibody and visualised with a secondary AlexiFluor-647 secondary conjugate antibody (blue). **(C)** A/E formation was quantified using Volocity suite software (Perkin Elmer) by analysing the number of bacteria attached to the host cells and the density of actin beneath attached bacteria (N= 100 for WT and N= 5 for Δ*adhE)* **(D)**. Asterisks indicate the level of significance calculated from a Student's t-test, ** indicates a p value between 0.001 and 0.01; *** indicates a p-value below 0.001.

The phenotypic studies of the Δ*adhE* mutant showed that this strain is unable to activate the LEE T3SS. In contrast the mutant appears to overexpress flagella in all media conditions. The flagella appear to be fully assembled on the surface of the bacteria, however the strain is non-motile. To try to understand the mechanism underlying these phenotypic changes, a combination of approaches was used.

6.9 Metabolic differences in the *ΔadhE* mutant

Initially a global approach to measuring metabolic changes in the Δ*adhE* mutant using liquid chromatography mass spectrometry (LC-MS) was used. This method can accurately detect a large number of metabolites. However, this approach revealed no clear differences in metabolite levels. One limitation of LC-MS is its inability to detect or resolve acetate, acetyl-phosphate and other metabolites of interest involved in the AdhE pathway. Therefore a more focused approach was taken to narrow down the search for metabolites. Assuming that the primary role of AdhE is its enzymatic function we hypothesised that if this function were disrupted, or absent in the case of a knockout mutant, this would directly affect the metabolites produced by this pathway.

6.9.1 *The levels of acetate in the ΔadhE mutant are increased*

To test this hypothesis, the levels of acetate secreted into the media during growth were measured. Acetate can be measured accurately using ^{1}H-NMR spectroscopy. This technique is highly quantitative as there is a linear relationship between the solution concentration and the area under the peak in the 1D-spectra. Therefore by using a standard curve of known acetate concentrations, the amount of acetate present in the media could be quantified (Figure 6-15B). Bacteria were cultured in MEM-HEPES media to $OD_{600\ nm}$ = 0.8, the cells were then removed from the media by centrifugation and the media was filtered. The media content was then analysed by NMR. Un-conditioned medium contained no detectable acetate, which meant that any acetate detected was produced by the growing culture (Figure 6-15A). Quantification of the acetate present in the conditioned media from WT and Δ*adhE* cultures gave values of 4.41 ± 0.04 mM and 5.21 ± 0.05 mM, respectively (Figure 6-15C). Acetate is a weak acid and can therefore

diffuse freely across the cell membrane in its protonated form, thus the extracellular level of acetate relates to the intracellular concentrations. As the undissociated form of acetate is highly membrane permeable and will distribute according to the difference in pH (ΔpH) across the membrane [268], it was possible to estimate the approximate concentration of the intracellular acetate pool to be 35 mM in the WT and 46 mM in the Δ*adhE* mutant (at a ΔpH of 0.88).

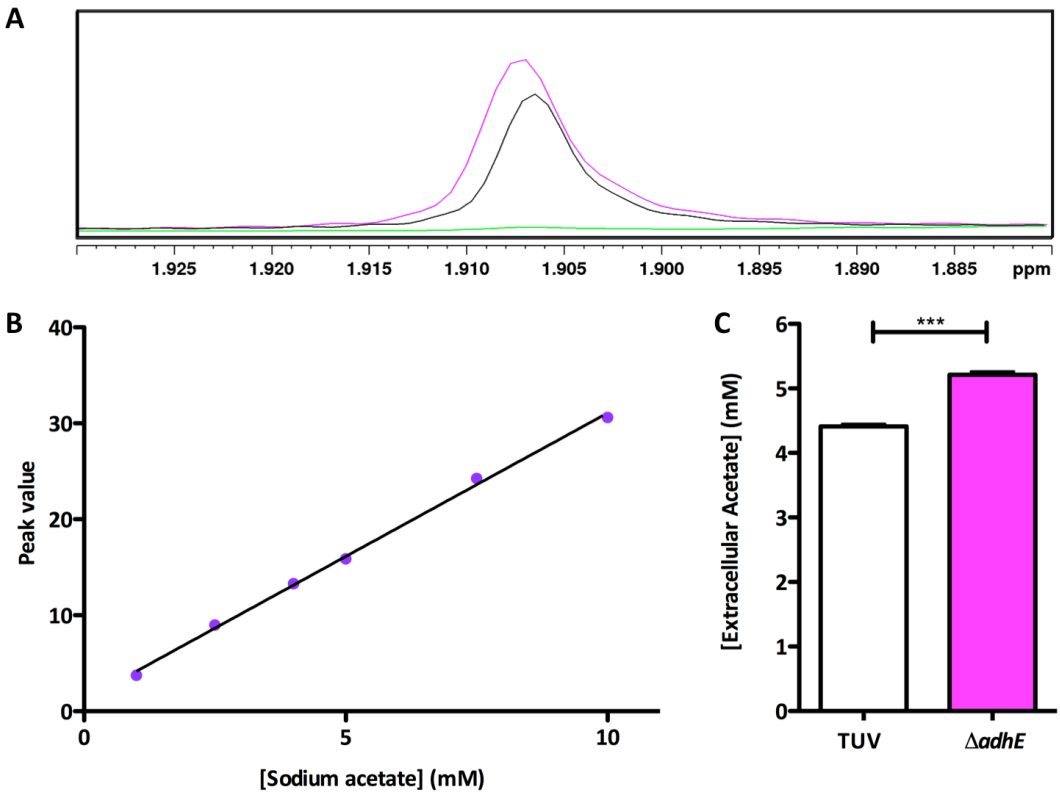


Figure 6-15 Quantification of acetate levels using ^{1}H-NMR spectroscopy. (A) NMR spectra showing the peaks corresponding to acetate at ~1.9 ppm, the spectrum for the media (MEM-HEPES) is shown in green, WT (TUV) conditioned media in black and Δ*adhE* conditioned media in pink. **(B)** Standard curve for a range of concentrations of sodium acetate. **(C)** The levels of acetate present in the conditioned media from WT and Δ*adhE.* The asterisks denote the level of significance following a Student's t-test of the data obtained from triplicate experiments (*** indicates a p-value below 0.001).

6.10 The effects of increased acetate

6.10.1 *The effect of acetate on fliC expression*

Since the Δ*adhE* mutant has increased levels of acetate, could changes in acetate be responsible for the phenotype? There are previous reports implicating SCFAs, such as

acetate, in the up-regulation of flagellar expression by directly stimulating the expression of class II flagellar genes [62]. To test the effects of acetate, the *fliC* promoter fusion was used to report on flagellar expression in MEM-HEPES media supplemented with 20 mM sodium acetate. This showed that under LEE T3SS-inducing conditions the addition of extracellular acetate did not increase *fliC* expression (Figure 6-16A), however this may be indicative of the fact that expression of the LEE T3SS leads to strong suppression of *fliC* expression. However when acetate was added to the Δ*adhE* mutant culture the levels of *fliC* expression increased slightly, by around 1.2-fold (Figure 6-16B). This may indicate that the Δ*adhE* mutant is more sensitive to extracellular acetate concentrations, which may be due to the increased intracellular concentrations when compared with the WT.

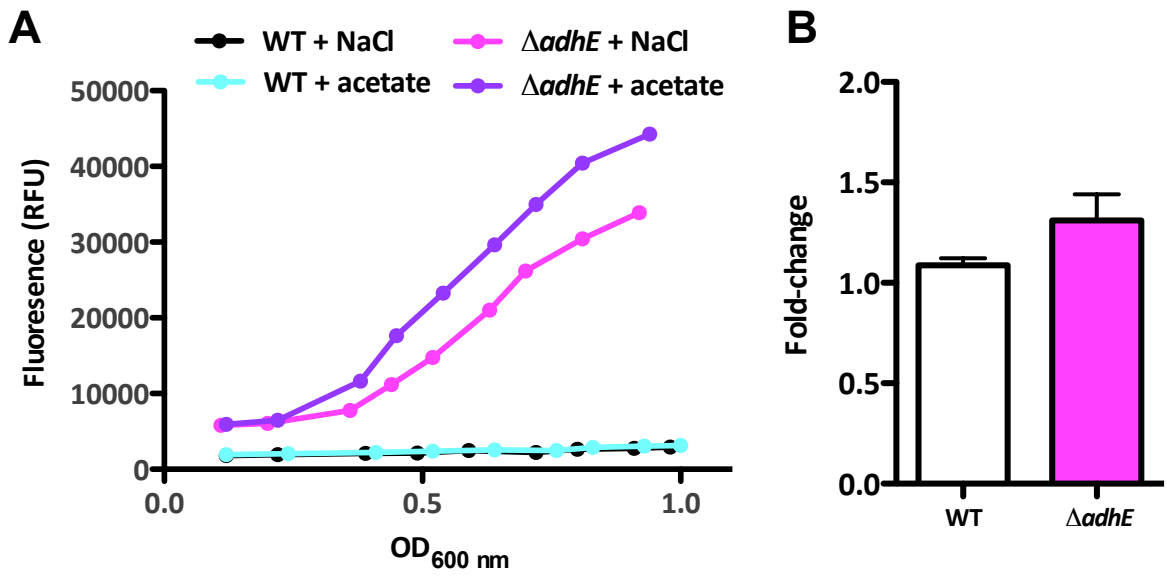


Figure 6-16 The effects of extracellular acetate on *fliC* expression. (A) Expression of the *fliC* promoter:GFP fusion during growth of WT or Δ*adhE* in MEM-HEPES with and without 20 mM sodium acetate. For the control 20 mM NaCl was added. **(B)** The fold-change in *fliC* expression with and without the addition of 20 mM acetate.

6.10.2 *The effect if acetate on LEE expression*

As mentioned above, SCFAs have been implicated in the regulation of virulence factors such as flagella. They have also been shown to reduce EHEC colonisation *in vivo* [269]. The mechanism behind this is not well understood, however since the LEE T3SS is important for EHEC colonisation, an investigation of whether acetate affected LEE T3S was undertaken. Expression of *LEE* genes were monitored using the promoter:GFP fusion constructs described above. WT bacteria were cultured in MEM-HEPES to induce LEE T3S

and the level of GFP produced was monitored throughout growth in the presence of 20 mM NaCl, acting as a control, or 20 mM sodium acetate. This showed that acetate did not affect LEE T3S, which is perhaps not surprising since the Δ*adhE* mutant shows little effect on the level of gene transcription for the LEE T3SS.

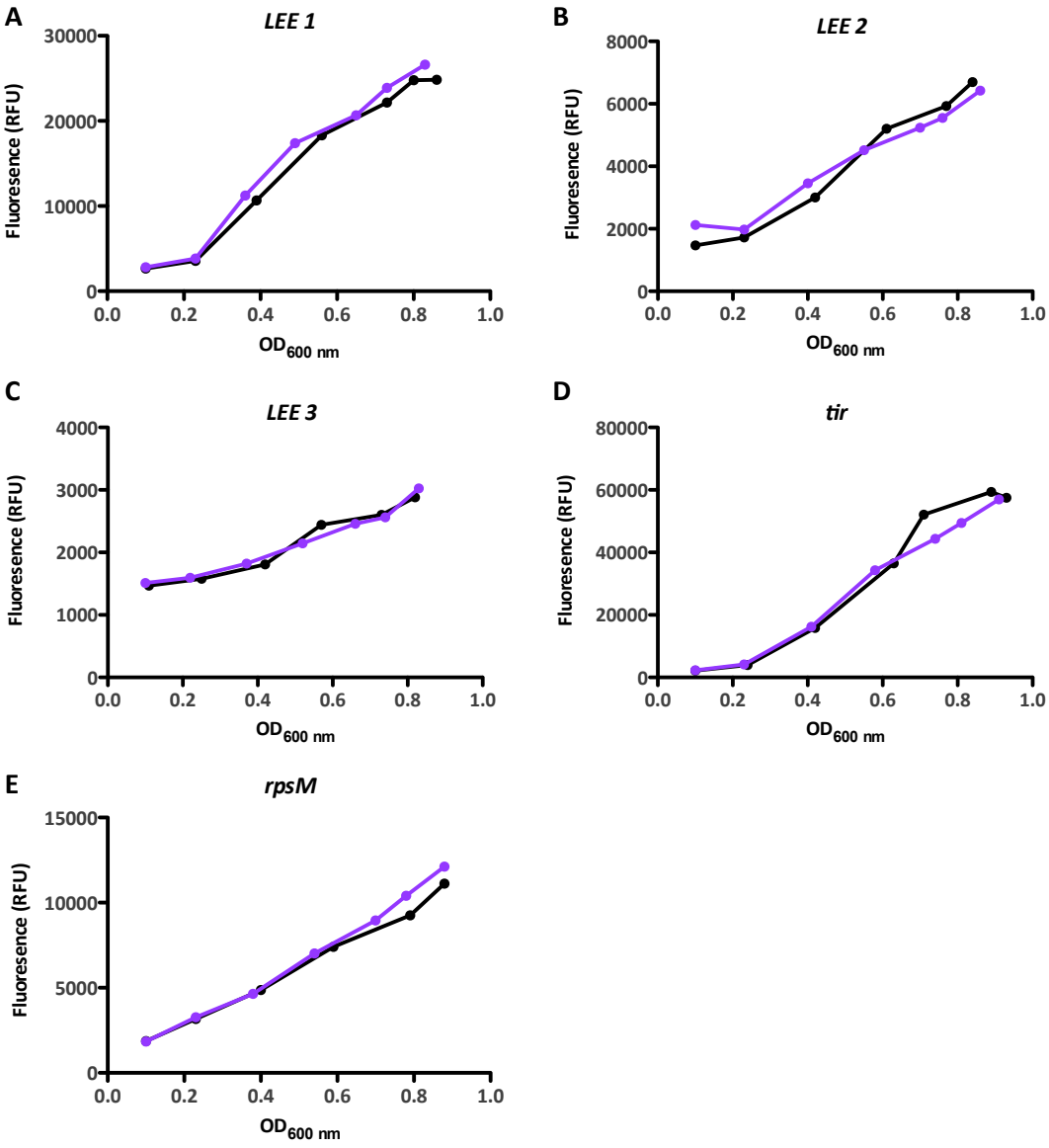


Figure 6-17 The effect of extracellular acetate on the expression of the LEE T3SS. WT bacteria were cultured with either 20 mM NaCl (black) or 20 mM sodium acetate (purple) and the expression of GFP reporting the expression of *LEE 1* **(A)**, *LEE 2* **(B)**, *LEE 3* **(C)**, *tir* **(D)** or *rpsM* **(E)** was monitored.

6.10.3 *Acetate induces protein post-translational modification*

Acetate levels can affect post-translation modifications (PTM) of proteins leading to phosphorylation and acetylation of susceptible residues (Figure 6-18A) [270]. Therefore, it

was hypothesised was that since the Δ*adhE* mutant had increased levels of acetate, the levels of protein PTM may be higher in this background. The levels of protein acetylation were investigated using an anti-acetyl lysine antibody, which indicated that in the Δ*adhE* mutant the levels of acetylated proteins appeared to be higher than observed for the WT (Figure 6-18B).

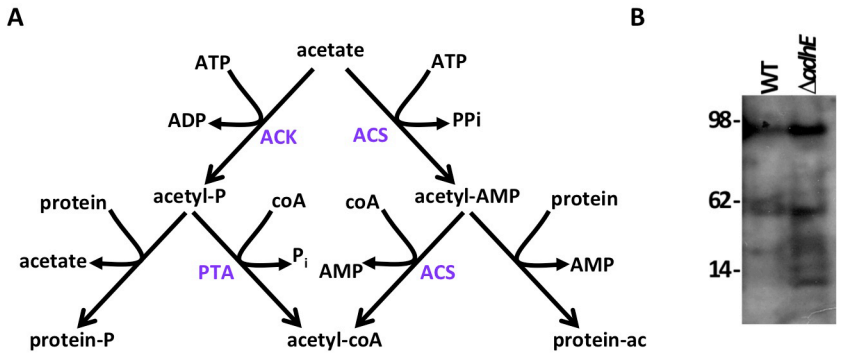


Figure 6-18 Acetate contributes to the post-translational modification of proteins. (A) Acetate can be metabolised to acetyl phosphate (acetyl-P) by acetate kinase (ACK), acetyl-P can then donate its phosphoryl group to target proteins within the cell. Acetate is also metabolised by acetyl-CoA synthase to acetyl-AMP which can then transfer its acetyl group to target proteins within the cell. Figure modified from [270]. **(B)** Anti-acetyl lysine blot of WT and *ΔadhE* cell lysate showing a global increase in protein acetylation in *ΔadhE*.

Acetylation of proteins is increasingly being recognised as having an important role in bacterial cell signalling [271]. For example, several metabolic enzymes are affected by acetylation, which often leads to either an increase or decrease in their activity [272]. Another well-studied protein whose function is affected by acetylation is the chemotactic protein, CheY. CheY is involved in controlling flagellar directionality and therefore the ability of the bacteria to be motile in response to chemotactic signals, such as nutrient levels. The link between CheY acetylation and motility was of interest owing to the fact that the *ΔadhE* mutant is non-motile which may indicate the inappropriate response of chemotactic proteins. Is the hyper-acetylation of CheY responsible for the non-motile phenotype of this mutant?

6.10.4 *Investigation of other metabolic changes in ΔadhE – cAMP*

Deletion of *adhE* resulted in an accumulation of acetate. To test whether other metabolites in the pathway were affected, the amount of intracellular cAMP was

measured by ELISA. WT and Δ*adhE* were grown in LB and MEM-HEPES media to test whether there was a difference in cAMP levels caused by medium composition. Cells were harvested at $OD_{600\,nm}$ = 0.8 and samples were processed as per the manufacturer's instructions. The concentration of intracellular cAMP was determined from a standard curve and the mean concentration of triplicate experiments (Figure 6-19A). This showed that there was no significant difference in cAMP levels between WT and mutant strains and that differences in media composition did not affect the levels. The measurement of cAMP can pose difficulties owing to the fact that it can be secreted into the media as well as being present in the cell. Here, the intracellular levels were measured but the extracellular levels may differ. In either case, the contamination of the samples with intra/extra cellular cAMP is difficult to avoid.

The level of cAMP in the cells has an effect on transcription though binding to cAMP repressor protein (CRP), which subsequently becomes activated. The cAMP-CRP complex has been shown to regulate the transcription of hundreds of genes within bacteria, including metabolic genes, ribosomal encoding genes and also class one flagellar genes [273,274]. CRP levels in WT EHEC and the Δ*adhE* mutant were determined by western blotting, which revealed a clear increase in CRP levels in the Δ*adhE* mutant. This is compared with the control protein, NusA, a translation elongation factor for which the levels do not differ between WT and mutant (Figure 6-19B). Since the levels of cAMP-CRP are known to auto-regulate the expression of CRP, this may indicate that the levels of cAMP are elevated in the mutant although the ELISA failed to detect a significant difference between WT and Δ*adhE* cAMP concentrations. Alternatively, it may be that the up-regulation of CRP is mediated by some other factor. Interestingly, *adhE* expression is up-regulated by cAMP-CRP binding, which may suggest that the metabolic feedback from *adhE* deletion is responsible for increased levels of CRP.

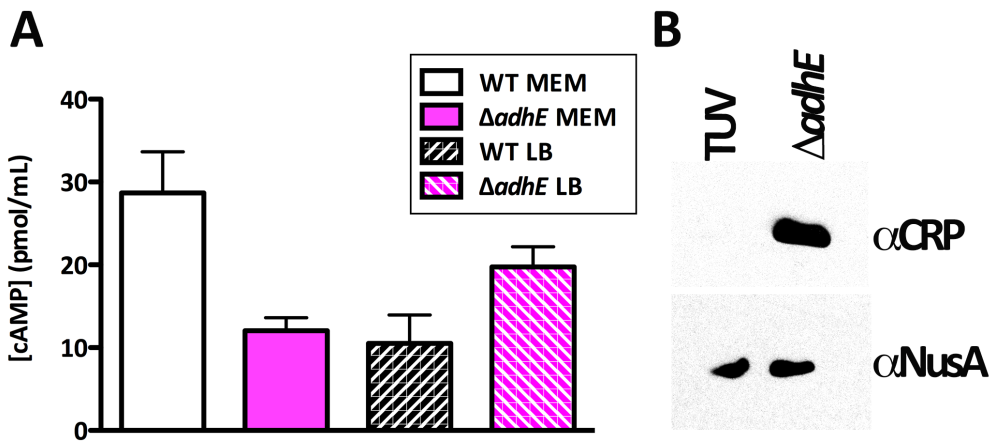


Figure 6-19 Quantification of cAMP levels. (A) The levels of intracellular cAMP were quantified using a cAMP ELISA kit (Enzo Life Sciences, UK) for bacteria grown in MEM-HEPES and LB media. The concentration of cAMP was determined from a standard curve of known cAMP concentrations. The bar chart shows data from triplicate experiments, a Student's t-test showed no significant differences in the mean concentration of cAMP between samples. **(B)** Western blot for the catabolite repressor activator protein (CRP) and for the transcription factor NusA, shown as control for the WT and *ΔadhE* strain.

6.11 Live imaging of bacterial swimming

Since it appeared that the Δ*adhE* mutant had an increased level of acetylation and it is known that CheY acetylation affects the directionality of flagellar movement, the swimming behaviour of this mutant was investigated. Bacteria were cultured in Terrific Broth (TB) at 30°C, conditions that promote motility. Previous work has extensively characterised *E. coli* motility, which consists of directional swimming that is interspersed by brief periods of "tumbling" to facilitate chemotaxis [275]. Transforming the WT and Δ*adhE* mutant with a plasmid encoding red fluorescent protein (RFP) enabled imaging for extended periods and showed both swimming and tumbling for the WT and an absence of motility in the Δ*adhE* mutant. These data suggest that the Δ*adhE* mutant does not display a proper chemotactic response. To observe the arrangement of the flagellar filaments on the bacterial cell surface AlexaFluor-488 dye esters were used to label the bacteria [144]. Thus both the swimming behaviour and flagellar motion could be assessed. A polarised flagellar bundle, required for normal swimming activity, was visualised for WT. In contrast, the Δ*adhE* mutant displayed a "paralysed" phenotype with no observed movement of the bacterium. Moreover, the flagella were dispersed over the bacterium and did not form a distinct bundle (Figure 6-20). These data imply that the

Δ*adhE* mutant is completely non-motile; it can express and assemble flagella but their functionality is disrupted. To investigate whether there was a link between the hyper-acetylation of Δ*adhE* proteins and the dysfunctional flagella, further investigation of the acetylation state of CheY was attempted.

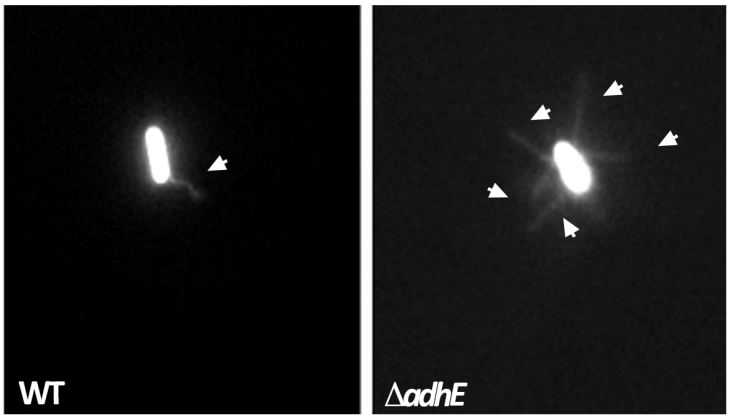


Figure 6-20 Still image from movies of bacterial swimming. Bacteria were labelled with AlexaFluor-488 dyes, which labelled both the bacterium and extracellular appendages such as flagella. For the WT a flagellar bundle can be clearly visualised (indicated by an arrow). The movie shows the bacteria using the flagellar bundle, to swim out of the field of view. The Δ*adhE* mutant did not have a flagellar bundle instead the flagella were dispersed on the surface of the bacterium. The movie shows that the bacterium did not swim, but twitching of the flagella can be observed. The movies are provided on a USB flash drive.

6.12 Investigation of CheY acetylation

6.12.1 *Cloning, overexpression and purification of CheY*

The *cheY* gene from *E. coli* O157 with 5' and 3' restriction sites was amplified by PCR and ligated into a Strataclone vector. The gene was then excised by digestions and ligated into the expression vector pET-28. The ligation transformants were checked for the presence of the *cheY* insert and the correct plasmid was confirmed by sequencing. The plasmid CheY-pET-28 was then transformed into BL21(DE3) cells. CheY was expressed in 2 L of culture following overnight induction with IPTG at 20°C. The protein was then purified using IMAC and SEC following standard protocols (Figure 6-21). The protein was identified using mass spectrometry.

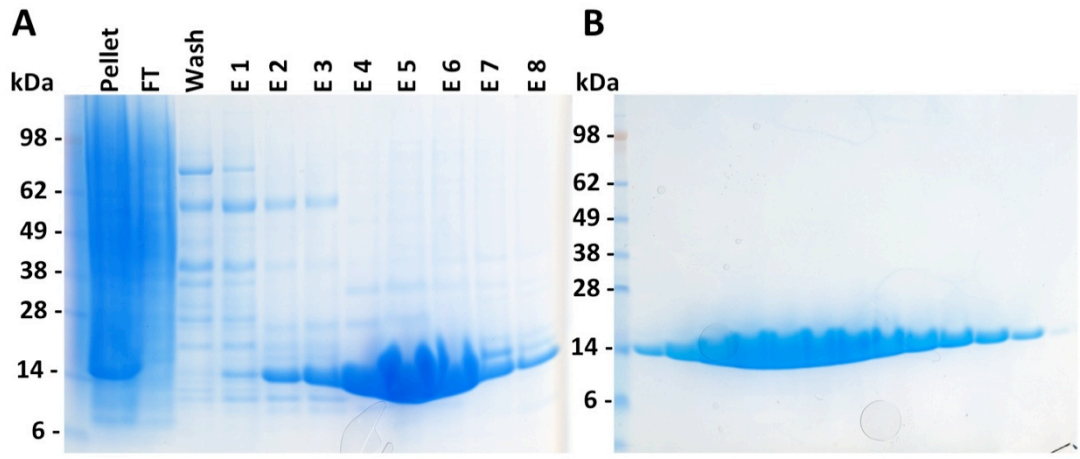


Figure 6-21 Purification of CheY. **(A)** SDS-PAGE gels showing the purification of CheY using IMAC. The pellet fraction, the flow-though (FT) and the 20 mM imidazole wash and the different elution fractions in 500 mM imidazole (E) have been indicated. **(B)** Following IMAC, fractions E4 to E8 were pooled and SEC was carried out using a SW-75 column. The fractions containing CheY are shown. The position of the molecular weight markers are indicated (kDa).

6.12.2 *CheY antibody production and testing*

Purified protein (10 mg ml^{-1}) was sent to BioGenes (Germany) for polyclonal antibody production. Two rabbits were immunised with CheY and the initial bleeds were tested for antibody specificity for CheY. Due to the large amount of non-specific binding the final bleeds underwent a further affinity purification step using immobilised CheY (carried out by BioGenes). The specificity of this antibody was tested against cell lysates and purified protein (Figure 6-22). Although the antibody did bind to purified CheY protein it also bound to a large number of cellular proteins. Therefore accurately determining which band corresponded to CheY was difficult. Since the RNA sequencing data showed that CheY was highly over expressed in the Δ*adhE* mutant this should have been detectable by western blotting. However there was no clear increase in the level of protein at around 14 kDa in the Δ*adhE* mutant. This approach was also tested with a Δ*cheY* strain (from Prof. Judy Armitage, University of Oxford). However this did not reveal a missing band at the correct size. Therefore, since the band corresponding to CheY could not be identified no further work was carried out using this antibody.

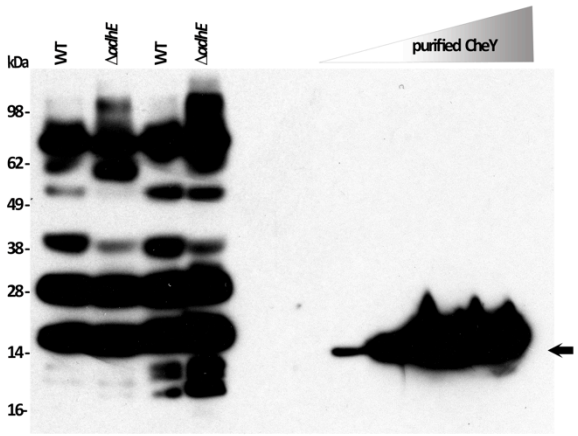


Figure 6-22 Testing of the anti-CheY antibody. Western blot of cell lysates from WT and Δ*adhE* strains in duplicate alongside increasing concentrations of purified CheY. The arrow indicates purified CheY. The position of the molecular weight marker is indicated (kDa).

6.13 TLR5 response to the *ΔadhE* mutant

Characterisation of the Δ*adhE* mutant has shown a phenotype of decreased expression of the LEE T3SS and an increased expression of flagella. Flagellin plays an important role in the activation of the innate immune response raising the possibility that the Δ*adhE* mutant might increase activation of TLR5 signalling. This was assessed using a HEK-293 TLR5 cell line that produces secreted alkaline phosphatase (SEAP) in response to appropriate stimuli. Supernatants from WT, the Δ*adhE* mutant and a Δ*fliC* mutant were added to the reporter cells at a range of dilutions and the extent of TLR5 stimulation determined. This showed a greater than 1000-fold increase in TLR5 stimulation by the Δ*adhE* mutant, compared with WT (Figure 6-23A). The specificity of the reporter was confirmed by addition of supernatants prepared from a Δ*fliC* mutant that produced no detectible TLR5 activity (Figure 6-23A). The effect of adding whole bacteria to the reporter cell line was also tested. This also showed an increase in the amount of TLR5 stimulation for the Δ*adhE* mutant and no stimulation for the Δ*fliC* mutant (Figure 6-23B). There was some detectable stimulation for the WT which may be due to the fact that the bacteria were added to the cells at an $OD_{600\ nm}$ of 0.3 at which point a proportion of the WT cells may still express flagella and be undergoing the switch to LEE T3S.

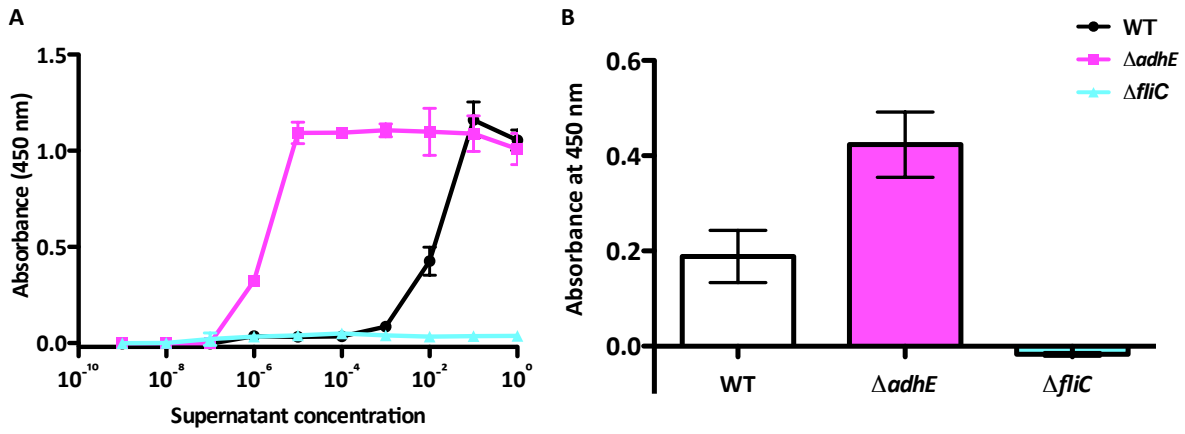


Figure 6-23 The level of TLR5 stimulation in response to the Δ*adhE* mutant differs compared with the WT. **(A)** The TLR5 response generated in response to WT and Δ*adhE* bacteria was tested using HEK-Blue hTLR5 cells. Secreted proteins from WT, Δ*adhE* and Δ*fliC* were added to HEK-Blue hTLR5 reporter cells and the level of NF-κB stimulation measured using Quanti-blue reagent, which reports the levels of SEAP through a change in absorbance at 450 nm. A representative graph from one of four replicate experiments is shown. **(B)** The level of TLR5 stimulation in response to whole bacteria was measured by the addition of bacteria cultured in MEM-HEPES to an $OD_{600\ nm}$ = 0.3 resulting in a multiplicity of infection of 10. Data show the mean and standard deviation of triplicate experiments.

6.14 Is disulfiram an inhibitor of AdhE?

The studies of the Δ*adhE* mutant highlighted AdhE as an attractive drug target. Since inhibition of AdhE would presumably lead to a decrease in the production of the LEE T3SS and an increase of flagellar production, this may be desirable phenotype in the case of an infection. This prompted the search for an existing compound that could be used to specifically target AdhE. The licensed drug disulfiram (DSF) has been used in the treatment of chronic alcoholism for over 60 years and acts by blocking the activity of human acetaldehyde dehydrogenase, one of the functional domains of AdhE.

6.14.1 *Effect of disulfiram on LEE T3S and flagellar gene expression*

To determine if DSF was able to mimic the phenotype observed for the Δ*adhE* mutant, immunoblotting was performed. Addition of 10 μM DSF did not affect growth or expression of cytoplasmic proteins such as GroEL (Figure 6-24). In contrast, addition of DSF strongly repressed secretion of both EspA and Tir and, as in the Δ*adhE* mutant, resulted in an increase in FliC expression (Figure 6-24). These data show that chemical

inhibition of AdhE results in a phenotype similar to that observed upon deletion of *adhE*. This implies that the enzymatic function of AdhE is key to the phenotype that we have described, rather than any regulatory role of the protein itself.

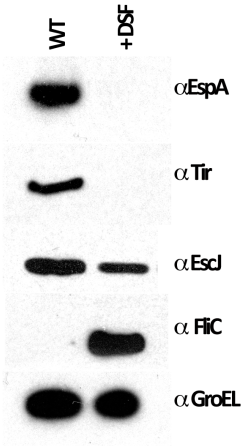


Figure 6-24 The effect of disulfiram on the secretion profile of *E. coli* O157. Western blot showing the effects of the addition of 10 µM disulfiram (DSF) on LEE T3SS proteins (EspA, Tir, EscJ) and flagellar proteins (FliC) on *E. coli* O157 grown in MEM-HEPES medium. GroEL is a loading control for the cellular protein EscJ.

To examine whether the observed changes in protein expression were caused by alterations in gene transcription, initially the GFP reporter fusions were used. WT bacteria were cultured in MEM-HEPES with 10 µM DSF and the expression of *LEE1*, *LEE2*, *LEE3*, *tir*, *rpsM* and *fliC* was reported by the level of GFP fluorescence. As was the case for the Δ*adhE* mutant strain there were no changes in the expression of the LEE T3SS genes, but unlike the mutant the change in *fliC* expression was not significant (Figure 6-25).

The level of gene expression was also investigated using RNA sequencing. Two broad groupings of genes were most prominently affected: those associated with motility (15 genes showed greater than 3-fold increase in expression) and genes associated with the LEE T3SS (5 genes showed greater than 20-fold decrease) (Table 6-3). The RNA sequencing data revealed that there was a significant change in the level of FliC expression, showing a 26-fold increase with DSF treatment, however it is unclear why this difference was not seen in the reporter fusion experiments. The Δ*adhE* mutant showed an increase in transcript levels for all of the flagellar genes, however this was not the case in the DSF treated bacteria. The decreased number of flagellar genes affected may indicate that the concentration of DSF used was not sufficient to completely inhibit AdhE.

Furthermore, DSF is thought to inhibit only D1 of AdhE and it is unknown if D2 can still function if D1 is inhibited, which may alter the metabolite pools within the cell.

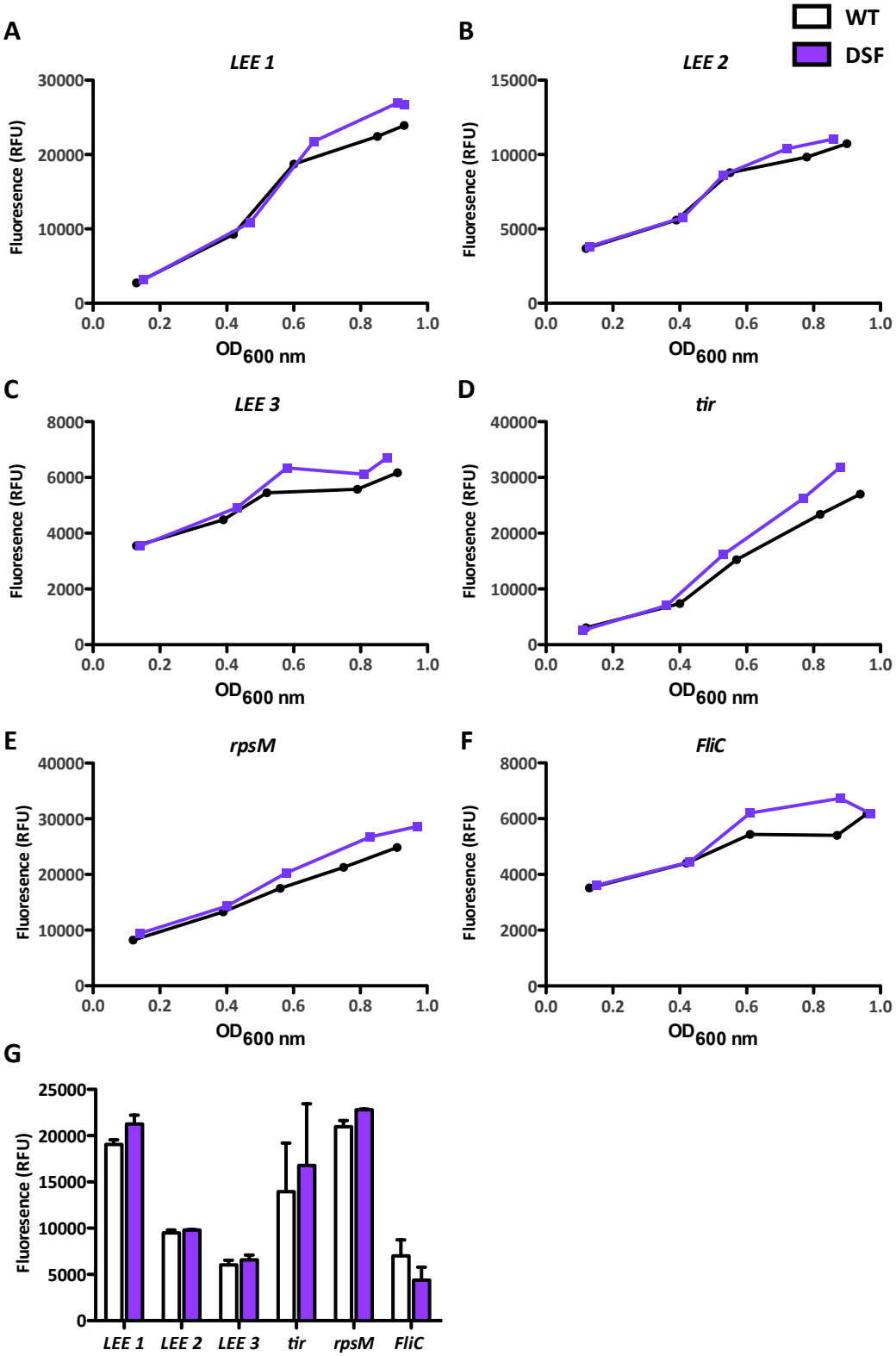


Figure 6-25 GFP:promoter fusions reporting on the expression of the LEE T3SS and flagellar genes in WT and DSF treated bacteria. (A-F) Representative graphs of GFP fluorescence measured during growth in MEM-HEPES, reporting on the expression of *LEE 1* **(A),** *LEE 2* **(B),** *LEE 3* **(C),** *tir* **(D),** *rpsM* **(E) and** *fliC* **(F). (G)** The level of GFP fluorescence at $OD_{600\ nm}$ = 0.7 taken from triplicate experiments. A Student's t-test showed there was no statistical difference between groups.

Gene name	WT (RPKM)	WT Sd	DSF (RPKM)	Fold-change (*p*<0.005)
Flagella				
flhC	11.89	2.76	36.24	3.0
flhD	10.68	6.35	37.72	3.5
cheZ	6.48	0.70	24.41	3.8
cheY	4.79	1.84	15.53	3.2
cheB	3.24	0.40	11.84	3.7
tap	3.37	0.35	23.67	7.0
tar	4.92	0.28	19.97	4.1
cheW	3.62	1.05	17.01	4.7
cheA	3.21	0.30	25.89	8.1
motB	0.64	0.23	6.66	10.4
fliA	0.13	0.11	5.92	46.6
fliC	18.26	2.19	475.61	26.1
fliD	1.80	0.36	12.58	7.0
fliS	1.16	0.53	6.66	5.7
flgK	6.21	0.43	18.50	3.0
LEE T3SS				
espB	21,921.68	25,11.34	435.66	- 50.3
espD	19,956.05	26,74.13	267.76	- 74.5
escD	211.45	23.62	7.40	- 28.6
escV	310.48	17.69	10.36	- 30.0
sepZ	1,398.83	638.05	63.61	- 22.0
Housekeeping				**Fold-change (p>0.005)**
groEL	225.73	129.74	555.64	0.4
gapA	793.80	437.90	1,661.08	0.4
dnaJ	18.08	9.01	49.05	0.3
gyrA	30.27	15.53	52.42	0.5
arcA	91.75	29.99	160.41	0.5
rrsA	7,550.23	962.02	7,266.14	1.0
fumC	5.58	1.68	6.32	0.8
rpoS	666.00	136.19	692.92	0.9
recA	76.33	18.68	100.29	0.7

Table 6-3 RNA sequencing data showing significant changes in *E. coli* treated with disulfiram. The average WT Reads per Kilobase Mapped (RPKM) from four experiments has been shown with the standard deviation (sd) of the data. The fold-change was calculated by dividing the WT RPKM by the disulfiram treated (DSF) RPKM values. Only significant fold-changes in flagellar and LEE T3SS genes have been shown (p <0.005). Several housekeeping genes have been listed and do not show a significant change in the RPKM values (p >0.005). The levels of significance determined using the Levene's t-test are shown.

6.14.2 *Does disulfiram affect the motility of E. coli?*

To determine whether treatment with disulfiram had an effect on motility of the WT bacterium a motility assay in soft agar was carried out. WT bacteria were cultured in LB with and without 10, 100 or 200 µM DSF to an $OD_{600\ nm}$ of 0.6. DSF treated bacteria were

inoculated into 0.25% (w/v) LB agar supplemented with the same concentration of DSF to maintain the presence of DSF throughout bacterial growth. Following overnight incubation at 30°C the growth zone diameter was measured. WT *E. coli* treated with DSF were less motile compared with the untreated WT and the effect of DSF appeared to be dose dependent.

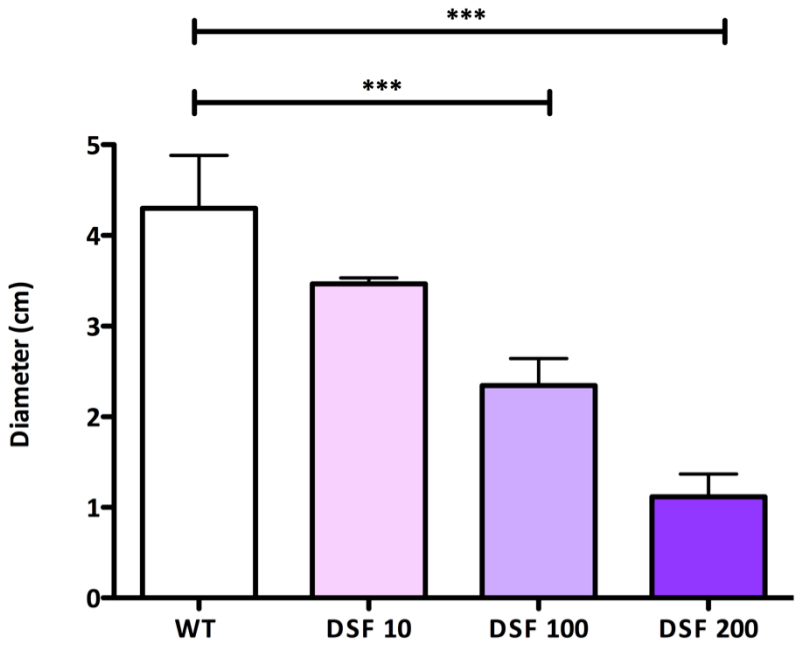


Figure 6-26 Treatment with DSF reduces bacterial motility in soft agar. The diameters of the swimming zone from triplicate experiments of WT and DSF treated bacteria. Bacteria were cultured in 0, 10, 100 or 200 µM DSF to an $OD_{600\ nm}$ of 0.6 before being inoculated into plates containing the equivalent concentration of DSF. The difference between the data sets was determined using a Student's t-test and asterisks indicate the degree of significance, *** indicates a p-value below 0.001.

6.14.3 *Treatment with disulfiram reduces attachment to host cells*

Western blotting analysis showed a significant decrease in the production of the LEE T3SS in the WT treated with DSF. To determine whether the treated bacteria were still able to form A/E lesions upon contact with host cells, a cell adhesion assay was performed (Figure 6-27A). This showed that compared with the WT, there were fewer DSF-treated bacteria attached to host cells (Figure 6-27B). However, the bacteria that were attached showed a level of actin polymerisation similar to that of the WT. Thus, although fewer bacteria were able to attach, the ones that did were able to form normal A/E lesions. This may be due to incomplete inhibition of AdhE at the concentration of DSF used (10 µM) or

because only one of the enzymatic functions of AdhE is being affected by DSF which may result in a less obvious phenotype when compared with the Δ*adhE* mutant.

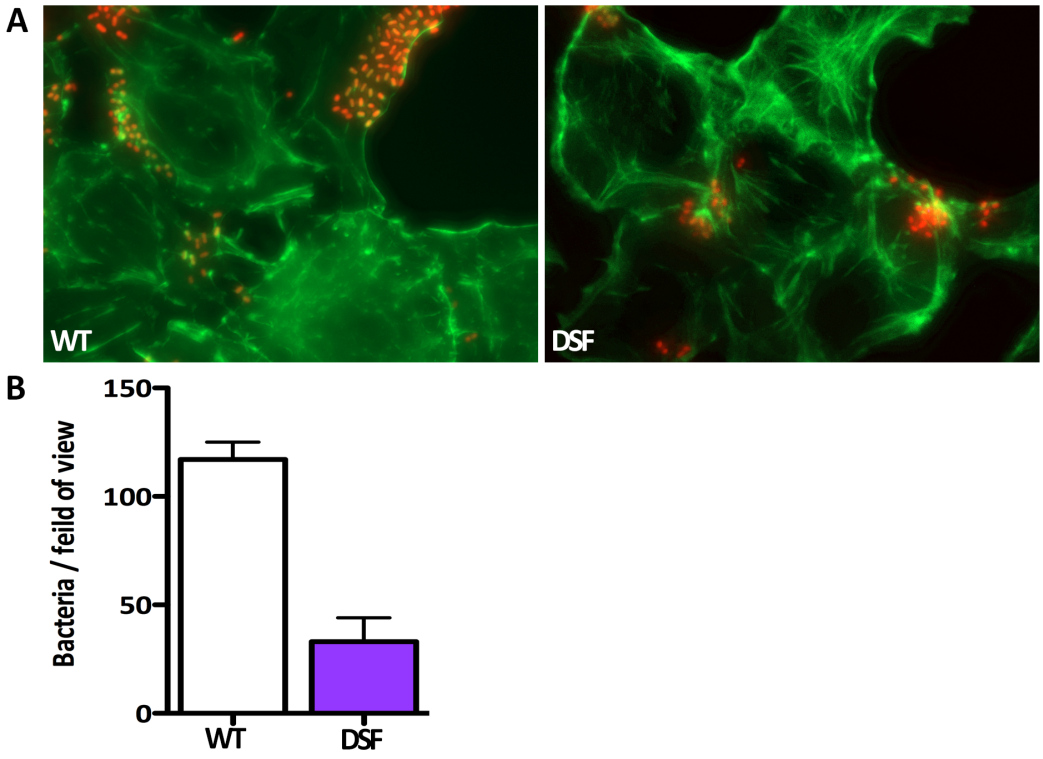


Figure 6-27 Treatment with disulfiram reduces A/E lesion formation. (A) A/E lesion formation by WT and DSF-treated bacteria transformed with pRFP. Host cell actin was visualised using FITC-phalloidin (green). Flagella were detected using an anti-H7 primary antibody and visualised with a secondary AlexaFluor-647 secondary conjugate antibody (blue). **(B)** The number of attached bacteria was quantified and the difference between the means quantified using a Student's t-test that gave a p-value of 0.02.

6.14.4 *Disulfiram reduces C. rodentium colonisation in vivo*

Having demonstrated that addition of DSF affected LEE T3SS and motility gene expression in a manner similar to the Δ*adhE* mutant, we aimed to test *in vivo* if targeting acetaldehyde dehydrogenase activity with DSF could be a general way to reduce the pathogenicity of attaching and effacing enterics. *C. rodentium* is a widely adopted model for LEE T3SS-mediated colonisation and is used as a model of *E. coli* O157 [276]. A BLAST search indicated that AdhE from *C. rodentium* shares 91% identity with the *E. coli* protein. A *lux* positive strain of *C. rodentium* was obtained from Prof. Gad Frankel (Imperial College, London). This enabled the use of an *in vivo* imaging system (IVIS) to track the colonisation of the mice at different time points. Bacteria were cultured with or without 10 μM DSF prior to oral inoculation of BALB/c mice. The mice were given a dose of 1 x 10^9

colony forming units. Colonisation of the small intestine was established at 68 h post inoculation and peaked at 144 h (6 days) (Figure 6-28). The addition of a single dose of DSF reduced colonisation by *C. rodentium* at all time points tested and produced significant differences in colonisation from 96 h onwards. Counting bacterial numbers in the faeces of infected animals further validated these data. Collectively, these data suggest that the addition of DSF can reduce LEE T3SS-mediated colonisation of mice and that this is likely to be a result of inhibition of one of the enzymatic functions of AdhE.

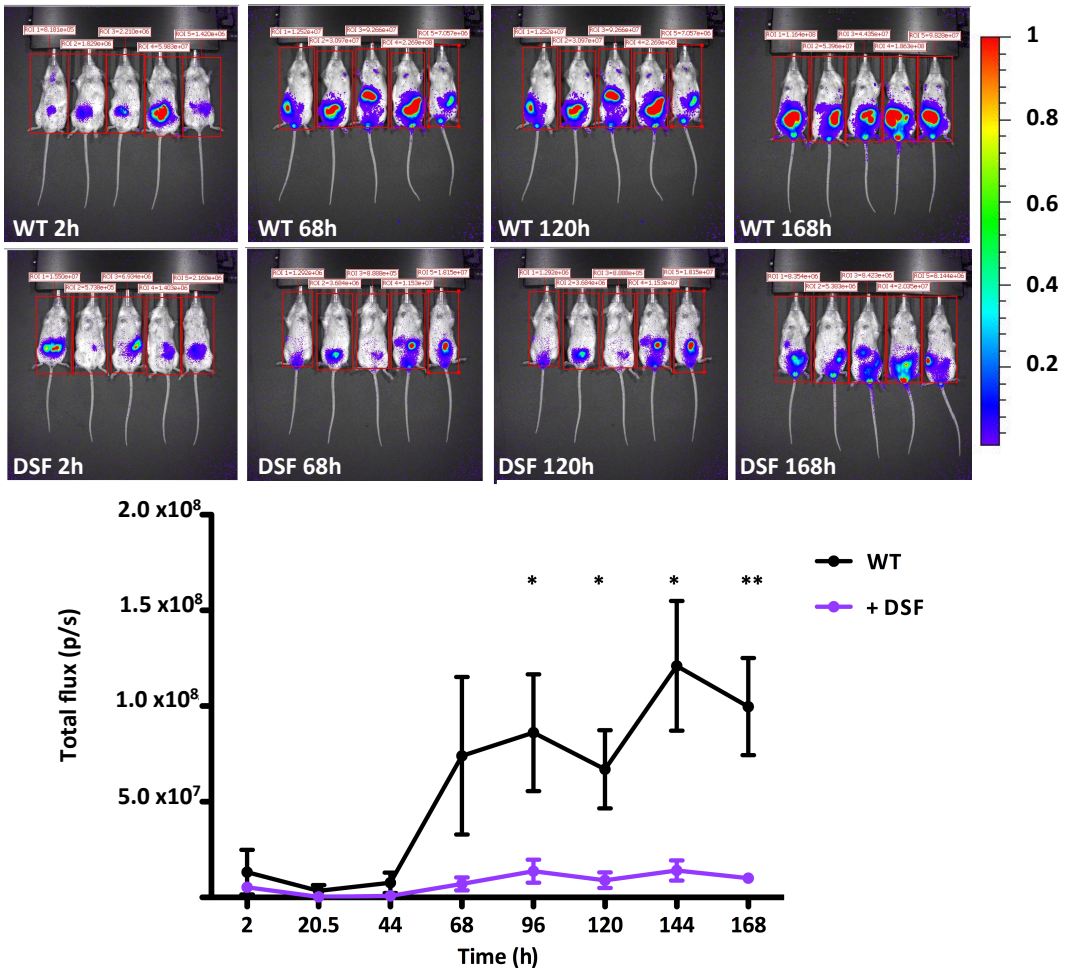


Figure 6-28 Colonisation of BALB/c mice with *C. rodentium* cultured in 10 µM disulfiram. Colonisation of BALB/c mice infected with *lux*-marked *C. rodentium* with and without DSF. Images of the mice in the IVIS have been shown for selected time points. The scale corresponds to the strength of luminescence in photons per second per cm^2 and the red boxes show the region measured for each mouse. The luminescence values for each group of mice are plotted below. Significance values are denoted by an asterisk, * indicates a p value between 0.01 to 0.05; ** indicates a p-value from 0.001 to 0.01 from a Student's t-test.

6.15 Discussion

AdhE was identified as a putative binding target of the salicylidene acylhydrazides by an affinity pull-down assay [100]. AdhE is metabolic enzyme that is responsible for the conversion of acetyl-CoA to ethanol via an acetaldehyde intermediate; this sequential reaction is carried out by the two functional domains of this protein. The role of AdhE in anaerobic conditions has been extensively studied due to interest in its role in conferring ethanol tolerance, an advantageous trait in the bioprocessing of cellulosic biomass to produce ethanol [277]. In aerobic conditions AdhE accounts for 1% of cellular protein, however little is known about the role of this protein in these conditions. Presumably production of a 96 kDa protein is costly in terms of fitness, therefore it would follow that its function must be of some importance to the cell in aerobic respiration.

To further investigate the role of AdhE in *E. coli* O157 a gene knockout was created using allelic exchange. The deletion mutant showed no defect in growth compared with the WT in the three different media tested, therefore indicating that deletion of *adhE* does not affect the fitness of *E. coli* O157 in the conditions tested. This is an important finding in terms of developing anti-virulence strategies as a reduction in survival may increase selective pressure for the development of resistance genes.

6.15.1 *Deletion of adhE mimics the phenotype associated with SA compound treatment*

Since the phenotype observed following treatment with the SA compounds is a decrease in LEE T3S [99] the secretion profile of the Δ*adhE* mutant grown under LEE T3S inducing conditions was analysed. SDS-PAGE analysis revealed a marked change in the proteins produced by the WT compared with the Δ*adhE* mutant. The most prominent change was the appearance of a band corresponding to a 60 kDa protein, which was identified as flagellin by tandem mass-spectrometry. This result was intriguing since the study by Tree *et al.* (2009) as well as showing a decrease in LEE T3S upon the addition of the SA compounds also showed that upon contact with host cells that there was an increased level of flagellar expression [99]. Further analysis of the secreted protein profile revealed that as well as the presence of FliC (which cannot be detected in the WT secreted profile),

the levels of LEE T3SS proteins such as EspA, Tir and EscJ were dramatically reduced. The degree of reduction was higher than that seen for EHEC treated with any of the SA compounds in previous studies.

Transcriptional analysis of the Δ*adhE* mutant did not show a reduction of LEE T3SS gene expression but did show a dramatic increase in flagellar gene expression. Gene transcription was investigated using two approaches; RNA sequencing and GFP reporter plasmids, which report on the activity of the *LEE1*, *LEE2*, *LEE3*, *tir*, *rpsM* and *fliC* promoters. Since there was a clear decrease in the production of LEE T3SS proteins it was presumed that the levels of gene expression would similarly be decreased. However, the GFP reporters showed no decrease in LEE T3SS expression, and initially this was attributed to the fact that plasmid reporter systems may not always accurately report gene expression. This can happen if the regulator binding site has not been included in the promoter region or the secondary structure of the promoter on the plasmid differs from its structure in the chromosome. However, the reporter plasmids have been validated in several previous studies [37]. RNA sequence analysis of the transcript levels of all of the LEE T3SS genes revealed only a minor (non-significant) reduction in LEE T3SS expression when compared with the WT. From these data it would appear that the down-regulation of the LEE T3SS is occurring at the post-transcriptional level. The level of flagellar gene expression shown by both the *fliC* reporter plasmid and the RNA sequencing data demonstrate that expression all of flagellar genes was greatly increased.

The LEE T3S and flagellar systems have been shown to be cross-regulated, enabling temporal separation of these two virulence factors [278]. Flagella are required not only for motility but also for the initial attachment to host cells [58]. However, as flagella are recognised as potent inflammatory signals their expression is down-regulated at which point the LEE T3SS acts to stabilise bacterial attachment to host cells [60]. Therefore, the ability to perform the switch from flagellar expression to LEE T3SS expression is imperative to the infection process. The Δ*adhE* mutant shows over-expression of flagella in all media types throughout all growth phases. This suggests that either the mutant is insensitive to environmental signals, which would usually induce a change in gene expression (for example culturing in MEM-HEPES induces LEE T3SS expression) or that the signals to express flagella override any external influences. The Δ*adhE* mutant is not only in sensitive to signals from culture medium but also to any signals that may arise from

host cells since the mutant showed reduced binding to Caco-2 cells and an inability to form A/E lesions.

6.15.2 *What is the molecular mechanism behind the observed changes in the ΔadhE mutant?*

How does the deletion of *adhE* lead to the over-expression of flagella? Since AdhE is a metabolic enzyme, deletion of this enzyme would presumably affect metabolism, leading to increases in metabolite pools in the pathway prior to the enzyme and decreases in the levels of metabolites after the enzyme. The Δ*adhE* mutant shows increased levels of extracellular acetate, corresponding to an increased level of intracellular acetate that would in fact be around 10-fold higher than the extracellular levels. Acetate has been shown to directly stimulate the expression of class II flagellar genes, which include the gene for the flagellar specific sigma factor (σ^{28}) that increases the expression of flagellar genes in the class 3 operons [62]. SFCAs have been shown to play an important role in the regulation of virulence in enteric bacteria in particular EHEC. In the study by Fukuda *et al.* (2012), acetate produced by B*ifidobacteria* strains conferred protection *in vivo* from a lethal dose of *E. coli* O157 and led to a reduced release of the Shiga toxin (Stx) [269]. Furthermore, relatively small changes in the acetate concentration led to large changes in the levels of Stx released which may suggest that although the measured difference between WT and Δ*adhE* acetate levels was less that 1 mM, this might be enough to induce a significant effect. The mechanism behind the effect of acetate was not determined in this study. However to assess whether acetate was having a direct effect on the LEE T3SS, the expression of the LEE T3SS reporter plasmids was determined in increasing concentrations of acetate. This showed no effect on the level of expression. Although, if the effects of acetate are post-transcriptional a reporter plasmid may not reveal any changes. Furthermore, the level of flagellar expression was not reported in this study; this would be interesting to investigate as the increased levels of acetate may affect flagellar expression.

The level of flagellar gene expression may be further influenced by CRP, which was shown to increase in the Δ*adhE* mutant. CRP is a transcriptional regulator which, when bound to cAMP, can activate or decrease gene transcription. The expression of the flagellar master

regulator, the flhD/C complex, has been shown to be increased by CRP [279]. Thus, in the Δ*adhE* mutant the expression of flagellar genes is being driven (i) by acetate at the class 2 promoter leading to an increase in σ^{28} and (ii) by CRP leading to an increase in the expression of *flhD/C*, both of which increase the expression of all flagellar genes. Through this dual mechanism, the expression of flagella in the Δ*adhE* mutant is greatly up-regulated, perhaps to the extent where the cell is no longer able to respond to other environmental signals. Furthermore, previous work has demonstrated that CRP may also cause repression of the WT LEE T3SS. A CRP binding site has been identified within the promoter of the master regulator, Ler [280]. Therefore increased levels of CRP may also act to repress the LEE T3SS, although RNA sequence analysis does not show a significant decrease in the level of *ler* expression.

How the over-expression of flagellar genes disrupts the translation of LEE T3SS genes remains to be elucidated. However if the regulation is post-translational, it is conceivable that a small regulatory RNA (srRNA) may be inhibiting the expression of LEE T3SS mRNA. In bacteria, examples of srRNAs playing a role in the regulation of physiological processes are becoming increasingly common. srRNAs can affect gene translation via several different mechanisms including recruitment of RNAses leading to mRNA degradation, by preventing ribosome binding or by interfering with the transcription of the mRNA [281]. For example, the *flhD/C* mRNA transcript has been shown to be subject to regulation by several srRNAs which either act to decrease or increase the levels of translation [282]. Therefore, it is probable that the LEE T3SS undergoes similar levels of regulation. For example in *Vibrio harveyi* the expression of the LEE T3SS is limited to mid-cell density owing to the action of two opposing srRNAs preventing its expression at either low or high cell density [283]. However, the role of srRNAs in regulating the LEE T3SS in *E. coli* has yet to be identified.

Since flagella are thought to be important for the initial stages of host cell attachment the effect of over-expressing flagella but not expressing the LEE T3SS on the ability to attach to host cells was unknown. The Δ*adhE* mutant showed a reduced ability to stabilise contact with host cells since following the wash steps used to remove unattached bacteria only a couple of bacteria were attached to the epithelial cells, compared with the ability of the WT bacterium to attach, wherein at least 100 cells remained attached. This suggests that the LEE T3SS is required for stabilising attachment to the host cell under

these conditions. Therefore it seems that expressing only flagella but not the LEE T3SS would be detrimental to *in vivo* colonisation of host tissues.

Furthermore, the inability to switch off flagellar expression in favour of LEE T3SS expression would presumably be detrimental to the disease process, owing to the immunogenic nature of flagellin. Flagellin stimulates a TLR5 response which leads to the production of inflammatory cytokines, leading to inflammation and the infiltration of immune cells [60]. This led to the hypothesis that the Δ*adhE* mutant would stimulate a stronger TLR5 response compared with the WT; this was tested using a TLR5 reporter cell line. The Δ*adhE* mutant secreted more flagellin than the WT and the flagellin correspondingly induced a stronger TLR5 response. A stronger TLR5 response was similarly seen following the addition of whole bacteria to the reporter cells, however some TLR5 activity was still detected. This is probably due to some WT bacteria still expressing flagella when they were added to the reporter cells. The effects of the immune response cannot fully be investigated using a single reporter cell line. To gain a full insight into the effects of increased flagellar expression on the immune system and its ability to clear the infection, a small animal model would need to be used. However, currently there are a limited number of animal models for EHEC infection, most of which require disruption of the host flora or the use of very young animals whose immune systems may not be fully developed.

Another reason for the reduced number of mutant bacteria bound in the A/E lesion assay could be the observation that the flagella of the Δ*adhE* mutant appear to be non-functional and therefore cannot interact with the cells, as required, to initiate contact. The lack of motility is intriguing, since it would seem logical that an increased number of flagella would result in increased motility, as is the case in species such as *Proteus mirabilis* which become hyperflagellated ‘swarmer cells’ with increased motility [284]. This led to the hypothesis that the lack of motility may be caused by dysregulation of the chemotactic response as deletion mutants for chemotactic genes such as *cheY* often have defects in motility. The function of several of the chemotactic proteins is regulated by post-translational modification. The chemotactic protein CheA undergoes phosphorylation and becomes activated which then leads to the phosphorylation of CheY. CheY can also be acetylated and has been shown to contain six acetylation sites,

suggesting that this PTM plays an important role in its function. PTM of CheY leads to 'activation' of the protein which promotes its binding to FliM, a component of the motor switch complex [285]. When active CheY is bound, it causes clockwise rotation of the flagellar rotor and prevents directional swimming. Increased binding of CheY inducing clockwise rotation of flagella is consistent with the non-motile phenotype seen in the Δ*adhE* mutant. This warranted further investigation of the acetylation of CheY in the mutant background. To this end, CheY was cloned into an expression vector and purified for antibody production with the intention of carrying out 2D-gel electrophoresis to differentiate between the differently acetylated species of CheY. However, even following additional affinity purification steps, the antibody was unable to unambiguously detect CheY in cell lysate. The planned analysis could therefore not be performed. However, from the RNA sequence analysis of the Δ*adhE* mutant it was evident that CheY transcript levels were increased >200 fold ($P<0.005$, Table 6-2), which would make it extremely difficult to differentiate between increased acetylation and an increased amount of protein. However the global levels of acetylation appear to be higher in the Δ*adhE* mutant, therefore it would follow that the levels of acetylated CheY would be increased. The altered levels of acetylated CheY are likely to play an important role in the phenotype observed since acetylated CheY has been shown to be more active than non-acetylated CheY by 4-5 orders of magnitude [286].

Figure 6-29 shows the proposed model of how perturbation of the AdhE pathway leads to increased intracellular acetate concentrations, resulting not in only increased expression of the entire flagellar system, but also protein acetylation, including the PTM of chemotactic proteins. This model is consistent with the flagellated but "paralysed" phenotype observed.

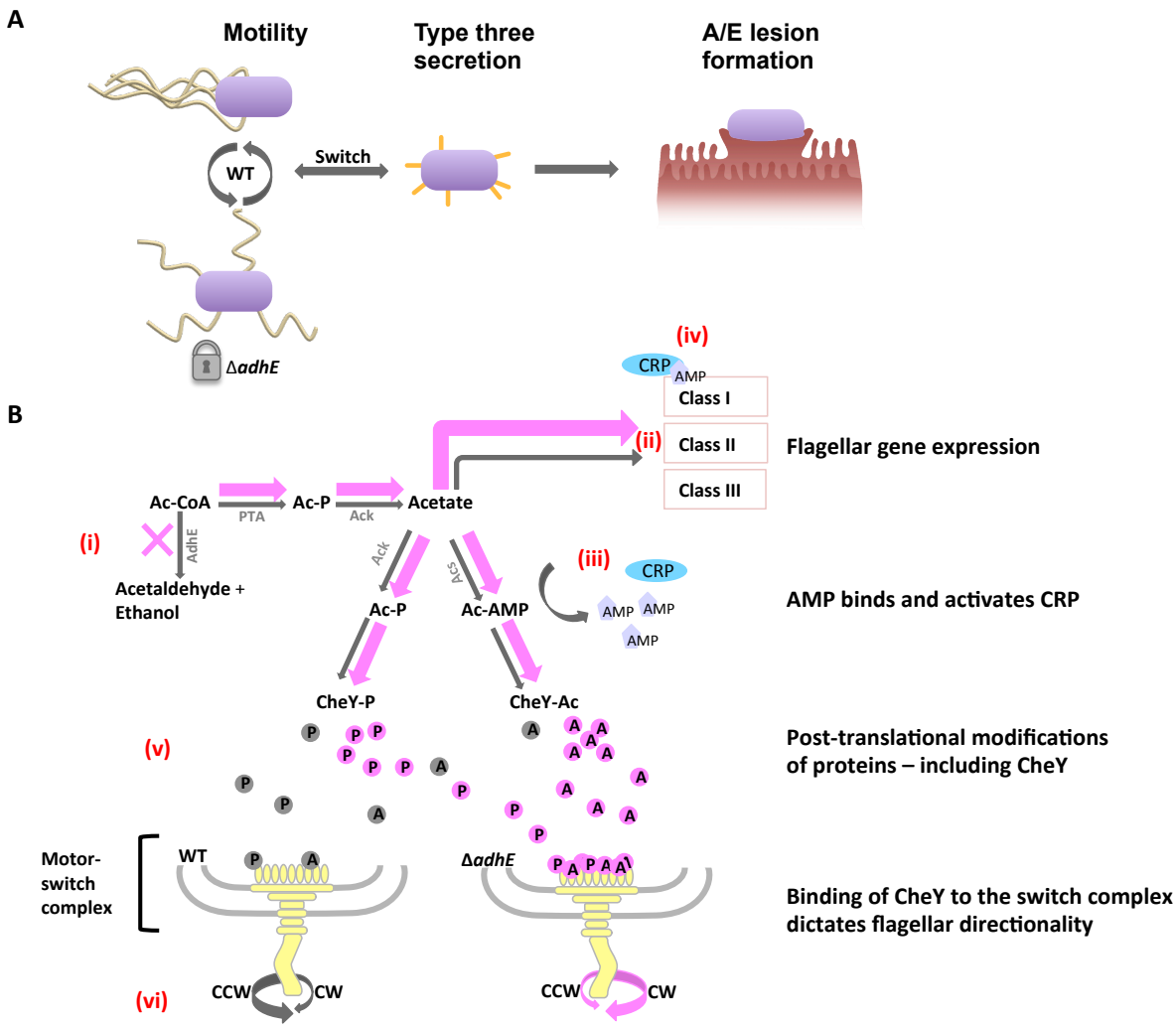


Figure 6-29 Proposed model of the effects of *adhE* deletion in *E. coli* O157. (A) Scheme showing the usual switch from motility, which requires a combination of both clockwise (CW) and counter clockwise (CCW) flagellar rotation, to LEE T3SS and then to A/E lesion formation. However the Δ*adhE* mutant is 'locked' in a flagellated state in which the flagella are biased towards CW rotation therefore reducing motility of the mutant and its ability to switch on LEE T3SS. **(B)** Proposed mechanism underlying the phenotype of the Δ*adhE* mutant. The normal role of AdhE is the generation of acetaldehyde and ethanol (i). Deletion of AdhE results in increases in metabolites in this pathway, (acetyl-CoA, Ac-CoA; acetyl phosphate, Ac-P; acetyl adenosine monophosphate, Ac-AMP) indicated by heavy pink arrows compared with the thinner grey arrows indicating the flux through the pathway in WT cells. Metabolites affected include acetate, known to stimulate expression of class II flagellar genes (ii), and cAMP which binds to cAMP receptor protein (CRP) (iii). In response, CRP activates expression of class I flagellar genes building a cascade of flagellar gene expression (iv). This strong up-regulation of flagella causes the reduction in expression and secretion of LEE T3SS proteins. A further consequence of increased acetate production is increased phosphorylation and acetylation of cytoplasmic proteins, an example being CheY (phosphorylated CheY, CheY-P; acetylated CheY, CheY-Ac). (v). Acetylation of CheY affects binding to the FliM motor switch complex thereby preventing CCW of the flagellar rotor and the formation of the flagellar bundle required for directional swimming seen during CW (vi).

6.15.3 *Inhibition of AdhE with disulfiram*

The licensed drug disulfiram (DSF) has been used in the treatment of chronic alcoholism for over 60 years and acts by blocking the activity of human acetaldehyde dehydrogenase. Therefore it was hypothesised that DSF may inhibit the acetaldehyde dehydrogenase domain of AdhE. To determine whether DSF was able to mimic the phenotype observed for the Δ*adhE* mutant, immunoblotting was performed. Addition of 10 μM DSF did not affect growth or expression of cytoplasmic proteins such as GroEL. In contrast, addition of DSF strongly repressed secretion of both EspA and Tir and, as in the Δ*adhE* mutant, resulted in an increase in FliC expression. These data show that chemical inhibition of AdhE results in a phenotype similar to that observed upon deletion of AdhE. This implies that the enzymatic function of AdhE is key to the phenotype that we have described, rather than any regulatory role of the protein itself. To examine whether the observed changes in protein expression were caused by alterations in gene transcription, RNA-sequencing was performed. Bacteria were cultured in LEE T3S-inducing conditions in the presence or absence of 10 μM DSF. Two broad groupings of genes were most prominently affected: those associated with motility and those associated with the LEE T3SS. This corresponds to changes seen in the Δ*adhE* mutant indicating that chemical inhibition of AdhE induces a phenotype similar to that induced by deletion of the gene, therefore DSF is specifically targeting AdhE.

In vitro work showed that treatment of *E. coli* with 10 μM DSF led to a decrease in LEE T3S and an increase flagellar production, two traits that would presumably be advantageous in terms of treating an infection. In order to test this an *in vivo* study was carried out. While small animal models of EHEC infection have been developed, they usually rely on host flora perturbation with antibiotics or the use of newborn or very young animals. Instead we opted to use a luciferase producing (*lux*+) strain of the mouse pathogen *Citrobacter rodentium*, a widely adopted model for LEE T3SS-mediated colonisation [276]. One treatment of *C. rodentium* with 10 μM DSF lead to a decrease in colonisation *in vivo*, suggesting that DSF has a similar effect on the LEE T3SS of *C. rodentium* as in *E. coli.* However, pre-treatment of the bacteria before inoculation is not the best method for testing the efficacy of a drug. Ideally the host would be infected and then treated with DSF, which would better mimic clinical treatment of an *E. coli* O157:H7

outbreak. Disulfiram in its current form is readily absorbed into the bloodstream, an attribute that is ideal for getting the drug to the liver where it is required in humans suffering from alcoholism. However, in an *E. coli* O157 infection the site of action is the intestine, therefore a different mode of delivery or modification of DSF may improve its activity against enteric pathogens. Interestingly, DSF was reported to possess tuberculostatic properties *in vivo* in the 1950s and recent work has evaluated the molecular mechanism underlying this activity [287]. Repurposing such a proven licensed oral drug could provide a novel treatment to limit EHEC infections in outbreaks through prophylactic treatment of people who have been potentially exposed to the organism.

One limitation of using *C. rodentium* was that, despite it using the same LEE T3SS mediated mechanism of attachment, the bacterium is non-motile and does not express functional flagella. Potentially, this would limit the TLR5 response and reduce the influence of the host immune system on bacterial clearance. In the future, testing a flagellated bacterial strain in a small animal model would reveal the dual effects of inhibiting LEE T3SS and stimulation of host immune responses. A recent paper by Petty *et al*. shows a detailed analysis of the *C. rodentium* genome and shows that despite not producing flagella the strains still possess several of the flagellar genes [288]. This may imply that the proposed mechanism for *E. coli* holds true for *Citrobacter* as well, in that the flagellar genes are still over expressed which leads to a decrease in the expression of LEE T3S genes. However, further work is required to validate the mechanism.

6.15.4 *Conclusion*

AdhE in *E. coli* O157:H7 was identified as a target protein of the salicylidene acylhydrazides, compounds shown to inhibit the expression of the LEE T3S in several species of Gram-negative pathogens. Although this study does not categorically confirm that AdhE is a target of the SA compounds it shows that targeting AdhE leads to a similar phenotype following treatment with the compounds namely, poor expression and function of the LEE T3SS and high expression of the flagellar system. Targeting AdhE in *C. rodentium* with the clinically approved drug disulfiram resulted in a similar phenotype, which was shown to lead to a reduction in colonisation in an *in vivo* model. These findings highlight AdhE as a promising anti-virulence target. Future studies could focus on screening for compounds that specifically inhibit the enzyme activity of AdhE.

7 Concluding remarks

The LEE T3SS is an attractive target for a novel class of anti-bacterial compounds that inhibit the expression of this virulence factor, thereby reducing the ability of the pathogen to cause disease. One such pathogen that relies on this virulence factor to colonise and cause disease in humans is *E. coli* O157:H7. The salicylidene acylhydrazide compounds have been shown to inhibit the expression of this virulence factor; however, the mode of action of these compounds was largely unknown. In order for compounds to be evolved and improved upon their cellular targets must first be identified.

Treatment of *E. coli* O157 with the SA compounds results in a clear phenotype: a decrease in the expression of the LEE T3SS and an increase in the expression of flagellar genes [99]. To investigate the mode of action of the SA compounds, an affinity pull-down assay using an immobilised derivative of the compounds was conducted in order to determine the binding proteins of the compounds. Nineteen putative targets were identified (Figure 1-6) [100]. None of the identified targets had previously been linked to the regulation of the LEE T3SS. Therefore a systematic approach to the investigation of the target proteins was required. The aim of this thesis was to investigate the putative target proteins and determine which of the proteins was responsible for the observed phenotype. A combination of biochemical and biophysical approaches was used to investigate the structure and function of the target proteins with the aim of determining the nature of the interactions between to the compounds and proteins. Alongside this, phenotypic studies of deletion mutants were conducted to determine whether deleting the target genes would replicate the observed phenotype, thus giving insights into which of the targets was responsible for reducing LEE T3S and increasing the production of flagella.

Tpx was the first target protein to be investigated, since previous studies reported some links to virulence in *Salmonella* and *Pseudomonas*, suggesting that this protein may also be important for *E. coli* O157 [171,172]. Structural studies using a combination of X-ray crystallography (Figure 3-7), NMR (Figure 3-15) and AUC (Figure 3-14) validated *yp*Tpx as being a true target of the SA compounds [100,178,179]. The binding site of ME0052 in Tpx was determined and the affinity of binding was determined to be approximately 100 µM [100]. However, the deletion of *tpx* in *E. coli* O157 did not replicate the SA compound phenotype. In fact, the Δ*tpx* mutant showed an increase in LEE T3S and a decrease in flagellar gene expression (Figure 3-20). This was an interesting result as it implied that Tpx might have a role in the regulation of these virulence factors. However, it also implied

that Tpx was not the key target of the SA compounds responsible for decreasing LEE T3S. Further investigation of the role of Tpx in virulence is warranted and development of compounds that target Tpx with higher specificity and affinity could be conducted as follow-on studies from this investigation.

The next target protein to be investigated was WrbA. The role of this redox active enzyme is poorly characterised, therefore investigation of WrbA offered an opportunity to learn more about this protein. Functional studies revealed that ME0052 reduced the enzyme activity of WrbA from *Salmonella* (Figure 4-11). This was a promising result as it indicated specific binding of ME0052 to WrbA. Co-crystallisation of WrbA from *Salmonella* with the compound indicated that ME0052 was binding in the active site of the enzyme (Figure 4-9). Electron density for only half compound could be observed; which may be indicative of the instability of the compound over time or that the concentrations of ME0052 used did not allow full occupancy of the binding site, therefore reducing the quality of the data corresponding to the compound. As with the Δ*tpx* strain the deletion of *wrbA* failed to reproduce the SA compound phenotype, therefore indicating that this protein was not the key target of the SA compounds. However, transcriptomic analysis of the Δ*wrbA* mutant indicated that this protein may have a role in regulating the expression of genes involved in the T2SS thus offering a further avenue of investigation for WrbA [100].

Similarly deletion of the target protein FolX failed to show a decrease in the expression of the LEE T3SS [100]. Since far-western analysis provided further evidence that FolX may bind to the compounds, structural studies of this protein were conducted. Unfortunately no further evidence of the compounds binding to FolX was obtained. However, investigations into how the crystal structure compared with the solution structure of FolX were conducted (Figure 4-3). This study showed that FolX is an octameric protein that is composed of a homodimer of tetramers [197].

Three of the nineteen putative target proteins belonged to the same family of enzymes, the prolyl *cis*/*trans* isomerases. SurA, FkpA and FklB were all investigated in this thesis. Initial studies of the binding of ME0052 to these proteins were conducted using far-western analysis. This indicated that of the three proteins only FklB showed evidence of binding to ME0052 (Figure 5-5). The enzyme activities of the PPIases were investigated, primarily to ensure that the purified proteins were active, but secondly so that inhibition

studies of ME0052 could be performed. However, the interpretation of the effects of ME0052 on the activity of these enzymes was ambiguous as the compounds contributed to the absorbance at the wavelength at which the enzyme assay was conducted. Further studies to investigate the interaction between the compounds were conducted using NMR (Figure 5-30) and ITC (Figure 5-33), both of which showed no evidence of ME0052 binding to FklB. Deletion mutants were obtained for *surA* and *fklB* and were investigated. Deletion of *fklB* had no effect on the expression of the LEE T3SS and the Δ*surA* mutant showed an increased expression of the LEE T3SS (Figure 5-10). From these studies it was concluded that the PPIases were probably false positives from the pull-down assay as no evidence of compound binding could be obtained and deletion knockout studies showed that the expression of the LEE T3SS was not decreased.

The final target to be investigated was AdhE. The role of this enzyme in anaerobic conditions is well characterised but little is known about its role in aerobic conditions. The binding of AdhE to the compounds was shown by far-western analysis, however the poor solubility of the compounds alongside the instability of the protein prevented further biophysical investigation of this protein. Unlike the deletion mutants in the other target proteins the Δ*adhE* mutant showed a decrease in LEE T3S alongside an increase in the production of flagellar proteins (Figure 6-8 and Figure 6-9). This result was exciting as the Δ*adhE* phenotype replicates the phenotype seen following treatment of whole cells with the SA compounds, therefore suggesting that AdhE may be the key target of the SA compounds. However, no concrete biochemical evidence for the binding of the compounds to AdhE was obtained in this study. Nonetheless, the phenotype of the Δ*adhE* mutant alone offers intriguing insights into the regulation of the LEE T3SS and the flagellar system.

Further analysis of changes of gene expression in the Δ*adhE* mutant when compared with the WT revealed that an increase in the expression of all flagellar genes was key to this phenotype (Table 6-2). No significant changes were seen in the expression of the LEE T3SS genes, which was surprising since the protein secretion profile of the Δ*adhE* mutant was devoid of any LEE T3SS proteins. Further investigation into the mechanism driving the increase in flagellar gene expression was conducted. Since AdhE is a metabolic enzyme the hypothesis was that deleting AdhE was perturbing this metabolic pathway by altering the production of metabolites. This hypothesis proved to be correct since the Δ*adhE*

mutant showed a significant increase in the production of acetate, a metabolite produced in the AdhE pathway (Figure 6-15). The next question to answer was: how could acetate influence the production of flagella in *E. coli* O157? Several studies by other groups have previously made the link between SCFAs such as acetate and the expression of flagellar genes [62,289]. The influence of the metabolic state on the expression of virulence factors is an emerging theme in the study of pathogens [262,269,290]. More specifically the role of SCFAs in regulating bacterial gene expression is becoming increasingly evident [62,262,290]. This work highlights AdhE as an attractive drug target as its inhibition results in a desirable phenotype: a decrease in LEE T3S, decreased attachment to host cells and an increase in flagellar production.

To further explore the possibility of more specifically targeting AdhE, the drug disulfiram was tested. This drug inhibits human acetaldehyde dehydrogenase therefore it was hypothesised that it may inhibit that bacterial form the enzyme (one of the enzymatic domains of AdhE) [291]. No biophysical evidence for the interaction between AdhE and disulfiram was obtained in this study; however, treatment of whole cells with disulfiram replicated the phenotype seen in the Δ*adhE* mutant. Disulfiram treatment resulted in a decrease in LEE T3S and an increase in flagellar production (Figure 6-24). *In vivo* studies showed that *t*reatment of *C. rodentium* with disulfiram prior to infection reduced the colonisation of this mouse pathogen (Figure 6-28). Further investigation into the repurposing of disulfiram as a treatment for *E. coli* O157 infections offers an exciting lead for the development of novel antibacterial therapies.

References

1 Sweeney, N. J., Klemm, P., McCormick, B. A., Moller-Nielsen, E., Utley, M., Schembri, M. A., Laux, D. C. and Cohen, P. S. (1996) The *Escherichia coli* K-12 gntP gene allows *E. coli* F-18 to occupy a distinct nutritional niche in the streptomycin-treated mouse large intestine. Infect. Immun. **64**, 3497–503.

2 Russo, T. A. and Johnson, J. R. (2000) Proposal for a new inclusive designation for extraintestinal pathogenic isolates of *Escherichia coli*: ExPEC. J. Infect. Dis. **181**, 1753–4.

3 Bray, J. (1945) Isolation of antigenically homogeneous strains of *Bact. coli Neapolitanum* from summer diarrhoea of infants. J. Pathol. Bacteriol. **57**.

4 Donnenberg, M. S. and Kaper, J. B. (1992) Enteropathogenic *Escherichia coli*. Infect. Immun. **60**, 3953–3961.

5 Robins-Browne, R., Still, C., Miliotis, M., RIchardson, N., Koornhof, H., Freiman, I., Schoub, B., Lecatsas, G. and Hartmand, E. (1980) Summer diarrhoea in African infants and children. Arh. Dis. Child. **55**, 923–928.

6 Giles, C., Sangster, G. and Smith, J. (1948) Epidemic gastro-enteritis of infants in Aberdeen during 1947. Arh. Dis. Child. **24**, 45–53.

7 Riley, L., Remis, R., Helgerson, S., McGee, H., Well, J., Davis, B., Herbert, F., Olcott, E., Johnson, R., Hargrett, N., et al. (1983) Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. N. Engl. J. Med. **308**, 681–685.

8 Karmali, M., Petric, M., Steele, B. and Corazon, L. (1983) Sporadic cases of haemolytic-uraemic syndrome associated with faecal cytotoxin and cytotoxin-producing *Escherichia coli* in stools. Lancet **321**, 619–620.

9 Karmali, M. A., Petric, M., Lim, C., Fleming, P. C. and Steele, B. T. (1983) *Escherichia coli* cytotoxin, haemolytic-uraemic syndrome, and haemorrhagic colitis. Lancet **2**, 1299–1300.

10 Møller-Stray, J., Eriksen, H. M., Bruheim, T., Kapperud, G., Lindstedt, B. a, Skeie, Å., Sunde, M., Urdahl, a M., Øygard, B. and Vold, L. (2012) Two outbreaks of diarrhoea in nurseries in Norway after farm visits, April to May 2009. Eur. Commun. Dis. Bull. **17**, 1–7.

11 Orden, J. A., Cid, D., Ruiz-Santa-Quiteria, J. A., García, S., Martínez, S. and de la Fuente, R. (2002) Verotoxin-producing Escherichia coli (VTEC), enteropathogenic *E. coli* (EPEC) and necrotoxigenic *E. coli* (NTEC) isolated from healthy cattle in Spain. J. Appl. Microbiol. **93**, 29–35.

12 Tilden, J., Young, W., McNamara, A. M., Custer, C., Boesel, B., Lambert-Fair, M. A., Majkowski, J., Vugia, D., Werner, S. B., Hollingsworth, J., et al. (1996) A new route of transmission for *Escherichia coli*: infection from dry fermented salami. Am. J. Public Health **86**, 1142–5.

13 Croxen, M. a, Law, R. J., Scholz, R., Keeney, K. M., Wlodarska, M. and Finlay, B. B. (2013) Recent Advances in Understanding Enteric Pathogenic *Escherichia coli*. Clin. Microbiol. Rev. **26**, 822–80.

14 Buchholz, U., Bernard, H., Werber, D., Böhmer, M. M., Remschmidt, C., Wilking, H., Deleré, Y., an der Heiden, M., Adlhoch, C., Dreesman, J., et al. (2011) German outbreak of *Escherichia coli* O104:H4 associated with sprouts. N. Engl. J. Med. **365**, 1763–1770.

15 Rasko, D., Webster, D., Sahl, J., Bashir, A., Boisen, N., Scheutz, F., Paxinos, E., Sebra, R., Chin, C., Iloppoulos, D., et al. (2011) Origins of the *E. coli* strain causing an outbreak of hemolytic–uremic syndrome in Germany. N. Engl. J. Med. **365**, 709–717.

16 Sheikh, J., Hicks, S., Dall'Agnol, M., Phillips, a D. and Nataro, J. P. (2001) Roles for Fis and YafK in biofilm formation by enteroaggregative *Escherichia coli* . Mol. Microbiol. **41**, 983–97.

17 McDaniel, T. K., Jarvis, K. G., Donnenberg, M. S. and Kaper, J. B. (1995) A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. Proc. Natl. Acad. Sci. U. S. A. **92**, 1664–8.

18 McDaniel, T. K. and Kaper, J. B. (1997) A cloned pathogenicity island from enteropathogenic *Escherichia coli* confers the attaching and effacing phenotype on *E. coli* K-12. Mol. Microbiol. **23**, 399–407.

19 Perna, N. T., Plunkett, G., Burland, V., Mau, B., Glasner, J. D., Rose, D. J., Mayhew, G. F., Evans, P. S., Gregor, J., Kirkpatrick, H. a, et al. (2001) Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature **409**, 529–33.

20 Moon, H. W., Whipp, S. C., Argenzio, R. A., Levine, M. M. and Giannella, R. A. (1983) Attaching and effacing activities of rabbit and human enteropathogenic *Escherichia coli* in pig and rabbit intestines. Infect. Immun. **41**, 1340–51.

21 Jerse, A. E., Yu, J., Tall, B. D. and Kaper, J. B. (1990) A genetic locus of enteropathogenic *Escherichia coli* necessary for the production of attaching and effacing lesions on tissue culture cells. Proc. Natl. Acad. Sci. U. S. A. **87**, 7839–43.

22 Elliott, S. J., Wainwright, L. A., Timothy, K., Jarvis, K. G., Deng, Y., Lai, L., Mcnamara, B. P., Michael, S. and Kaper, J. B. (1998) The complete sequence of the locus of enterocyte effacement (LEE) from enteropathogenic *Escherichia coli*. Mol. Microbiol. **28**, 1–4.

23 Fang, F. C. and Rimsky, S. (2008) New insights into transcriptional regulation by H-NS. Curr. Opin. Microbiol. **11**, 113–120.

24 Mellies, J. L., Larabee, F. J., Zarr, M. a, Horback, K. L., Lorenzen, E. and Mavor, D. (2008) Ler interdomain linker is essential for anti-silencing activity in enteropathogenic *Escherichia coli*. Microbiology **154**, 3624–38.

25 Deng, W., Puente, J. L., Gruenheid, S., Li, Y., Vallance, B. a, Vázquez, A., Barba, J., Ibarra, J. A., O'Donnell, P., Metalnikov, P., et al. (2004) Dissecting virulence: systematic and functional analyses of a pathogenicity island. Proc. Natl. Acad. Sci. U. S. A. **101**, 3597–602.

26 Iyoda, S., Koizumi, N., Satou, H., Lu, Y., Saitoh, T., Ohnishi, M. and Watanabe, H. (2006) The GrlR-GrlA regulatory system coordinately controls the expression of flagellar and LEE-encoded type III protein secretion systems in enterohemorrhagic *Escherichia coli*. J. Bacteriol. **188**, 5682–92.

27 Hughes, D. T., Clarke, M. B., Yamamoto, K., Rasko, D. a and Sperandio, V. (2009) The QseC adrenergic signaling cascade in Enterohemorrhagic *E. coli* (EHEC). PLoS Pathog. **5**, e1000553.

28 Njoroge, J. W., Nguyen, Y., Curtis, M. M., Moreira, C. G. and Sperandio, V. (2012) Virulence meets metabolism: Cra and KdpE gene regulation in enterohemorrhagic *Escherichia coli*. MBio **3**, e00280–12.

29 Zhang, L., Chaudhuri, R. R., Hobman, J. L., Patel, M. D., Antony, C., Sarti, D., Roe, A. J., Vlisidou, I., Shaw, R. K., Falciani, F., et al. (2004) Regulators encoded in the *Escherichia coli* type III secretion system 2 gene cluster influence exression of genes withing the locus of enterocyte effacement in Enterohemorrhagic *E. coli* O157:H7. Infect. Immun. **72**, 7282–7293.

30 Shakhnovich, E. A., Davis, B. M. and Waldor, M. K. (2009) Hfq negatively regulates type III secretion in EHEC and several other pathogens. Mol. Microbiol. **74**, 347–63.

31 Kenny, B., Abe, A., Stein, M. and Finlay, B. B. (1997) Enteropathogenic *Escherichia coli* protein secretion is induced in response to conditions similar to those in the gastrointestinal tract. Infect. Immun. **65**, 2606–12.

32 Mellies, J., Barron, A. M. S. and Carmona, A. M. (2007) Enteropathogenic and enterohemorrhagic *Escherichia coli* virulence gene regulation. Infect. Immun. **75**, 4199–210.

33 Troisfontaines, P. and Cornelis, G. R. (2005) Type III secretion: more systems than you think. Physiology **20**, 326–39.

34 Yip, C. K. and Strynadka, N. C. J. (2006) New structural insights into the bacterial type III secretion system. Trends Biochem. Sci. **31**, 223–30.

35 Deane, J. E., Abrusci, P., Johnson, S. and Lea, S. M. (2010) Timing is everything: the regulation of type III secretion. Cell. Mol. Life Sci. **67**, 1065–75.

36 Zarivach, R., Vuckovic, M., Deng, W., Finlay, B. B. and Strynadka, N. C. J. (2007) Structural analysis of a prototypical ATPase from the type III secretion system. Nat. Struct. Mol. Biol. **14**, 131–7.

37 Tree, J. J., Wolfson, E. B., Wang, D., Roe, A. J. and Gally, D. L. (2009) Controlling injection: regulation of type III secretion in enterohaemorrhagic *Escherichia coli*. Trends Microbiol. **17**, 361–70.

38 Wang, D., Roe, A. J., McAteer, S., Shipston, M. J. and Gally, D. L. (2008) Hierarchal type III secretion of translocators and effectors from *Escherichia coli* O157:H7 requires the carboxy terminus of SepL that binds to Tir. Mol. Microbiol. **69**, 1499–512.

39 Lario, P. I., Pfuetzner, R. a, Frey, E. a, Creagh, L., Haynes, C., Maurelli, A. T. and Strynadka, N. C. J. (2005) Structure and biochemical analysis of a secretin pilot protein. EMBO J. **24**, 1111–21.

40 Wang, Y. A., Yu, X., Yip, C., Strynadka, N. C. and Egelman, E. H. (2006) Structural polymorphism in bacterial EspA filaments revealed by cryo-EM and an improved approach to helical reconstruction. Structure **14**, 1189–96.

41 Veenendaal, A. K. J., Hodgkinson, J. L., Schwarzer, L., Stabat, D., Zenk, S. F. and Blocker, A. J. (2007) The type III secretion system needle tip complex mediates host cell sensing and translocon insertion. Mol. Microbiol. **63**, 1719–30.

42 Wiles, S., Pickard, K. M., Peng, K., MacDonald, T. T. and Frankel, G. (2006) *In vivo* bioluminescence imaging of the murine pathogen *Citrobacter rodentium*. Infect. Immun. **74**, 5391–6.

43 Kenny, B. (2002) Enteropathogenic *Escherichia coli* (EPEC)–a crafty subversive little bug. Mirobiology **148**, 1967–1978.

44 Kalman, D., Weiner, O. D., Goosney, D. L., Sedat, J. W., Finlay, B. B., Abo, A. and Bishop, J. M. (1999) Enteropathogenic *E. coli* acts through WASP and Arp2/3 complex to form actin pedestals. Nat. Cell Biol. **1**, 389–91.

45 Kodama, T., Akeda, Y., Kono, G., Takahashi, A., Imura, K., Iida, T. and Honda, T. (2002) The EspB protein of enterohaemorrhagic *Escherichia coli* interacts directly with alpha-catenin. Cell. Microbiol. **4**, 213–22.

46 Marchès, O., Batchelor, M., Shaw, R. K., Cummings, N., Nagai, T., Sasakawa, C., Carlsson, S. R., Lundmark, R., Cougoule, C., Caron, E., et al. (2006) EspF of Enteropathogenic *Escherichia coli* Binds Sorting Nexin 9. J. Bacteriol. **188**, 3110–3115.

47 Kenny, B. and Jepson, M. (2000) Targeting of an enteropathogenic *Escherichia coli* (EPEC) effector protein to host mitochondria. Cell. Microbiol. **2**, 579–90.

48 Elliott, S. J., Sperandio, V., Girón, J. A., Mellies, J. L., Wainwright, L., Steven, W., Mcdaniel, T. K., Kaper, J. B., Shin, S. and Mellies, J. A. Y. L. (2000) The locus of enterocyte effacement (LEE) -encoded regulator controls expression of both LEE- and non-LEE-encoded virulence factors in Enteropathogenic and Enterohemorrhagic *Escherichia coli*. Infect. Immun. **68**, 6115–6125.

49 Dean, P. and Kenny, B. (2009) The effector repertoire of enteropathogenic *E. coli*: ganging up on the host cell. Curr. Opin. Microbiol. **12**, 101–9.

50 Cornelis, G. R. (2006) The type III secretion injectisome. Nat. Rev. Microbiol. **4**, 811–25.

51 Abby, S. S. and Rocha, E. P. C. (2012) The Non-flagellar type iii secretion system evolved from the bacterial flagellum and diversified into host-cell adapted systems. PLoS Genet. **8**.

52 Chevance, F. F. V and Hughes, K. T. (2008) Coordinating assembly of a bacterial macromolecular machine. Nat. Rev. Microbiol. **6**, 455–465.

53 Erhardt, M., Namba, K. and Hughes, K. T. (2010) Bacterial nanomachines: the flagellum and type III injectisome. Cold Spring Harb. Perspect. Biol. **2**, a000299.

54 Kearns, D. B. (2010) A field guide to bacterial swarming motility. Nat. Rev. Microbiol., Nature Publishing Group **8**, 634–44.

55 Yan, J., Barak, R., Liarzi, O., Shainskaya, A. and Eisenbach, M. (2008) *In vivo* acetylation of CheY, a response regulator in chemotaxis of *Escherichia coli*. J. Mol. Biol. **376**, 1260–71.

56 Barak, R., Prasad, K., Shainskaya, A., Wolfe, A. J. and Eisenbach, M. (2004) Acetylation of the chemotaxis response regulator CheY by acetyl-CoA synthetase purified from *Escherichia coli*. J. Mol. Biol. **342**, 383–401.

57 Dailey, F. E. and Berg, H. C. (1993) Change in direction of flagellar rotation in *Escherichia coli* mediated by acetate kinase. J. Bacteriol. **175**, 3236–9.

58 Mahajan, A., Currie, C. G., Mackie, S., Tree, J., McAteer, S., McKendrick, I., McNeilly, T. N., Roe, A., La Ragione, R. M., Woodward, M. J., et al. (2009) An investigation of the expression and adhesin function of H7 flagella in the interaction of *Escherichia coli* O157 : H7 with bovine intestinal epithelium. Cell. Microbiol. **11**, 121–37.

59 Girón, J. A., Torres, A. G., Freer, E. and Kaper, J. B. (2002) The flagella of enteropathogenic *Escherichia coli* mediate adherence to epithelial cells. Mol. Microbiol. **44**, 361–79.

60 Hayashi, F., Smith, K. D., Ozinsky, A., Hawn, T. R., Yi, E. C., Goodlett, D. R., Eng, J. K., Akira, S., Underhill, D. M. and Aderem, A. (2001) The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. Nature **410**, 1099–103.

61 Smith, T. G. and Hoover, T. R. (2009) Deciphering bacterial flagellar gene regulatory networks in the genomic era. Adv. Appl. Microbiol. 1st ed., pp 257–95, Elsevier Inc.

62 Tobe, T., Nakanishi, N. and Sugimoto, N. (2011) Activation of motility by sensing short-chain fatty acids via two steps in a flagellar gene regulatory cascade in enterohemorrhagic *Escherichia coli*. Infect. Immun. **79**, 1016–24.

63 Johannes, L. and Römer, W. (2010) Shiga toxins--from cell biology to biomedical applications. Nat. Rev. Microbiol., Nature Publishing Group **8**, 105–16.

64 Pacheco, A. R. and Sperandio, V. (2012) Shiga toxin in enterohemorrhagic *E. coli* : regulation and novel anti-virulence strategies. Front. Cell. Infect. Microbiol. **2**, 81.

65 Kaper, J. B., Nataro, J. P. and Mobley, H. L. (2004) Pathogenic *Escherichia coli*. Nat. Rev. Microbiol. **2**, 123–40.

66 Fraser, M., Fujinaga, M., Cherney, M., Melton-Celsa, A., Twiddy, E., O'Brien, A. and James, M. (2004) Structure of Shiga toxin type 2 (Stx2) from *Escherichia coli* O157:H7. J. Biol. Chem. **279**.

67 Endo, Y., Tsurugi, K., Yutsude, T., Y, T., Ogasawara, T. and Igarashi, K. (1988) Site of action of a Vero toxin (VT2) from *Escherichia coli* O157:H7 and of Shiga toxin on eukaryotic ribosomes. Eur. J. Biochem. **50**, 45–50.

68 Lindberg, A., Brown, J. E., Strömberg, N., Westling-Ryd, M., Schultz, J. E. and Karlsson, K. a. (1987) Identification of the carbohydrate receptor for Shiga toxin produced by *Shigella dysenteriae* type 1. J. Biol. Chem. **262**, 1779–85.

69 Uhlich, G. a. (2009) KatP contributes to OxyR-regulated hydrogen peroxide resistance in *Escherichia coli* serotype O157 : H7. Microbiology **155**, 3589–98.

70 Schmidt, H. and Beutin, L. (1995) Molecular analysis of the plasmid-encoded hemolysin of *Escherichia coli* O157:H7 strain EDL 933. Infect. Immun. **63**, 1055–1061.

71 Schmidt, H., Maier, E., Karch, H. and Benz, R. (1996) Pore-forming properties of the plasmid-encoded hemolysin of enterohemorrhagic *Escherichia coli* 0157:H7. Eur. J. Biochem. **601**, 594–601.

72 Aldick, T., Bielaszewska, M., Zhang, W., Brockmeyer, J., Schmidt, H., Friedrich, A. W., Kim, K. S., Schmidt, M. A. and Karch, H. (2007) Hemolysin from Shiga toxin-negative *Escherichia coli* O26 strains injures microvascular endothelium. Microbes Infect. **9**, 282–90.

73 Orth, D., Ehrlenbach, S., Brockmeyer, J., Khan, A. B., Huber, G., Karch, H., Sarg, B., Lindner, H. and Würzner, R. (2010) EspP, a serine protease of enterohemorrhagic *Escherichia coli*, impairs complement activation by cleaving complement factors C3/C3b and C5. Infect. Immun. **78**, 4294–301.

74 Brockmeyer, J., Aldick, T., Soltwisch, J., Zhang, W., Tarr, P. I., Weiss, A., Dreisewerd, K., Müthing, J., Bielaszewska, M. and Karch, H. (2011) Enterohaemorrhagic *Escherichia coli* haemolysin is cleaved and inactivated by serine protease EspPα. Environ. Microbiol. **13**, 1327–41.

75 Zhang, X., McDaniel, a D., Wolf, L. E., Keusch, G. T., Waldor, M. K. and Acheson, D. W. (2000) Quinolone antibiotics induce Shiga toxin-encoding bacteriophages, toxin production, and death in mice. J. Infect. Dis. **181**, 664–70.

76 Trachtman, H., Cnann, A., Christen, E., Gibbs, K., Acheson, D. W. K., Weiss, R., Kaskel, F. J., Spitzer, A. and Hirschman, G. (2013) Effect of an oral Shia toxin-binding agent on diarrhea-assocaited hemolytic uremic syndrome in children. JAMA **290**, 1337–1344.

77 Silberstein, C., Lucero, M. S., Zotta, E., Copeland, D. P., Lingyun, L., Repetto, H. a and Ibarra, C. (2011) A glucosylceramide synthase inhibitor protects rats against the cytotoxic effects of shiga toxin 2. Pediatr. Res. **69**, 390–4.

78 Zhu, C., Yu, J., Yang, Z., Davis, K., Rios, H., Wang, B., Glenn, G. and Boedeker, E. C. (2008) Protection against Shiga toxin-producing *Escherichia coli* infection by transcutaneous immunization with Shiga toxin subunit B. Clin. Vaccine Immunol. **15**, 359–66.

79 Goldwater, P. N. and Bettelheim, K. A. (2012) Treatment of enterohemorrhagic *Escherichia coli* (EHEC) infection and hemolytic uremic syndrome (HUS). BMC Med., BioMed Central Ltd **10**, 12.

80 Tacket, C. O., Sztein, M. B., Losonsky, G., Abe, A., Finlay, B. B., Mcnamara, B. P., George, T., James, S. P., Nataro, J. P., Myron, M., et al. (2000) Role of EspB in experimental human enteropathogenic *Escherichia coli* infection. Infect. Immun. **68**.

81 Donnenberg, M. S., Tacket, C., James, S. P., Losonsky, G., Nataro, J. P., Wasserman, S. S., Kaper, J. B. and Levine, M. M. (1993) Role of the *eaeA* gene in experimental Enteropathogenic *Escherichia coil* infection. J. Clin. Investig. **92**.

82 Wilson, K. H. and Perini, F. (1988) Role of competition for nutrients in suppression of *Clostridium difficile* by the colonic microflora. J. Infect. Dis. **56**.

83 Hannan, T. J., Mysorekar, I. U., Hung, C. S., Isaacson-Schmid, M. L. and Hultgren, S. J. (2010) Early severe inflammatory responses to uropathogenic *E. coli* predispose to chronic and recurrent urinary tract infection. PLoS Pathog. **6**, e1001042.

84 Wu, X. R., Sun, T. T. and Medina, J. J. (1996) *In vitro* binding of type 1-fimbriated *Escherichia coli* to uroplakins Ia and Ib: relation to urinary tract infections. Proc. Natl. Acad. Sci. U. S. A. **93**, 9630–5.

85 Cusumano, C. K., Pinkner, J. S., Han, Z., Greene, S. E., Ford, B. a, Crowley, J. R., Henderson, J. P., Janetka, J. W. and Hultgren, S. J. (2011) Treatment and prevention of urinary tract infection with orally active FimH inhibitors. Sci. Transl. Med. **3**, 109ra115.

86 Viboud, G. I. and Bliska, J. B. (2005) *Yersinia* outer proteins: role in modulation of host cell signaling responses and pathogenesis. Annu. Rev. Microbiol. **59**, 69–89.

87 Linington, R. G., Robertson, M., Gauthier, A., Finlay, B. B., MacMillan, J. B., Molinski, T. F., van Soest, R. and Andersen, R. J. (2002) Caminosides B-D, antimicrobial glycolipids isolated from the marine sponge *Caminus sphaeroconia*. Org. Lett. **4**, 4089–4092.

88 Zhang, Z., Zong, C., Song, G., Lv, G., Chun, Y., Wang, P., Ding, N. and Li, Y. (2010) Total synthesis of caminoside B, a novel antimicrobial glycolipid isolated from the marine sponge Caminus sphaeroconia. Carbohydr. Res., Elsevier Ltd **345**, 750–60.

89 Parmeggiani, A. and Swart, G. (1985) Mechanism of action of kirromycin-like antibiotics. Annu. Rev. Microbiol. **39**, 557–77.

90 Kimura, K., Iwatsuki, M., Nagai, T., Matsumoto, A., Takahashi, Y., Shiomi, K., Omura, S. and Abe, A. (2011) A small-molecule inhibitor of the bacterial type III secretion system protects against in vivo infection with *Citrobacter rodentium*. J. Antibiot. (Tokyo). **64**, 197–203.

91 Iwatsuki, M., Uchida, R., Yoshijima, H., Ui, H., Shiomi, K., Kim, Y., Hirose, T., Sunazuka, T., Abe, A., Tomoda, H., et al. (2008) Guadinomines, Type III secretion system inhibitors, produced by Streptomyces sp. K01-0509. I: taxonomy, fermentation, isolation and biological properties. J. Antibiot. (Tokyo). **61**, 222–9.

92 Pan, N. J., Brady, M. J., Leong, J. M. and Goguen, J. D. (2009) Targeting type III secretion in *Yersinia pestis*. Antimicrob. Agents Chemother. **53**, 385–92.

93 Felise, H. B., Nguyen, H. V, Pfuetzner, R. a, Barry, K. C., Jackson, S. R., Blanc, M.-P., Bronstein, P. a, Kline, T. and Miller, S. I. (2008) An inhibitor of Gram-negative bacterial virulence protein secretion. Cell Host Microbe **4**, 325–36.

94 Kauppi, A. M., Nordfelth, R., Uvell, H., Wolf-watz, H. and Elofsson, M. (2003) Targeting bacterial virulence: Inhibitors of Type III Secretion in *Yersinia*. Chem. Biol. **10**, 241–249.

95 Nordfelth, R., Kauppi, A. M., Norberg, H. A. and Elofsson, M. (2005) Small-molecule inhibitors specifically targeting type III secretion. Infect. Immun. **73**.

96 Wolf, K., Betts, H. J., Chellas-Géry, B., Hower, S., Linton, C. N. and Fields, K. a. (2006) Treatment of Chlamydia trachomatis with a small molecule inhibitor of the Yersinia type III secretion system disrupts progression of the chlamydial developmental cycle. Mol. Microbiol. **61**, 1543–55.

97 Hudson, D. L., Layton, A. N., Field, T. R., Bowen, A. J., Wolf-Watz, H., Elofsson, M., Stevens, M. P. and Galyov, E. E. (2007) The GrlR-GrlA regulatory system coordinately controls the expression of flagellar and LEE-encoded type III protein secretion systems in enterohemorrhagic *Escherichia coli*. Antimicrob. Agents Chemother. **51**, 2631–5.

98 Negrea, A., Bjur, E., Ygberg, S. E., Elofsson, M., Wolf-Watz, H. and Rhen, M. (2007) Salicylidene acylhydrazides that affect type III protein secretion in *Salmonella enterica* serovar typhimurium. Antimicrob. Agents Chemother. **51**, 2867–76.

99 Tree, J. J., Wang, D., McInally, C., Mahajan, A., Layton, A., Houghton, I., Elofsson, M., Stevens, M. P., Gally, D. L. and Roe, A. J. (2009) Characterization of the effects of salicylidene acylhydrazide compounds on type III secretion in *Escherichia coli* O157:H7. Infect. Immun. **77**, 4209–20.

100 Wang, D., Zetterström, C. E., Gabrielsen, M., Beckham, K. S. H., Tree, J. J., Macdonald, S. E., Byron, O., Mitchell, T. J., Gally, D. L., Herzyk, P., et al. (2011)

Identification of bacterial target proteins for the salicylidene acylhydrazide class of virulence-blocking compounds. J. Biol. Chem. **286**, 29922–29931.

101 Bailey, L., Gylfe, A., Sundin, C., Muschiol, S., Elofsson, M., Nordström, P., Henriques-Normark, B., Lugert, R., Waldenström, A., Wolf-Watz, H., et al. (2007) Small molecule inhibitors of type III secretion in *Yersinia* block the *Chlamydia pneumoniae* infection cycle. FEBS Lett. **581**, 587–95.

102 Slepenkin, A., Enquist, P.-A., Hägglund, U., de la Maza, L. M., Elofsson, M. and Peterson, E. M. (2007) Reversal of the antichlamydial activity of putative type III secretion inhibitors by iron. Infect. Immun. **75**, 3478–89.

103 Muschiol, S., Bailey, L., Gylfe, A., Sundin, C., Hultenby, K., Bergström, S., Elofsson, M., Wolf-Watz, H., Normark, S. and Henriques-Normark, B. (2006) A small-molecule inhibitor of type III secretion inhibits different stages of the infectious cycle of *Chlamydia trachomatis*. Proc. Natl. Acad. Sci. U. S. A. **103**, 14566–71.

104 Veenendaal, A. K. J., Sundin, C. and Blocker, A. J. (2009) Small-molecule type III secretion system inhibitors block assembly of the *Shigella* type III secreton. J. Bacteriol. **191**, 563–70.

105 Layton, A. N., Hudson, D. L., Thompson, A., Hinton, J. C. D., Stevens, J. M., Galyov, E. E. and Stevens, M. P. (2010) Salicylidene acylhydrazide-mediated inhibition of type III secretion system-1 in *Salmonella enterica* serovar Typhimurium is associated with iron restriction and can be reversed by free iron. FEMS Microbiol. Lett. **302**, 114–22.

106 Gauthier, A., Robertson, M. L., Lowden, M., Ibarra, A., Puente, J. L., Finlay, B. B., Ibarra, J. A. and Puente, L. (2005) Transcriptional inhibitor of virulence factors in Enteropathogenic *Escherichia coli*. Antimicrob. Agents Chemother. **49**, 4101–4109.

107 Martinez-Argudo, I., Veenendaal, A. K. J., Liu, X., Roehrich, a D., Ronessen, M. C., Franzoni, G., van Rietschoten, K. N., Morimoto, Y. V, Saijo-Hamano, Y., Avison, M. B., et al. (2013) Isolation of *Salmonella* mutants resistant to the inhibitory effect of salicylidene acylhydrazides on flagella-mediated motility. PLoS One **8**, e52179.

108 Leslie, A. G. W. and Powell, H. R. Processing diffraction data with mosflm. Evolving methods for Macromoleuclar Crystallography, 41-51.

109 Evans, P. (2006) Scaling and assessment of data quality. Acta Crystallogr. D. Biol. Crystallogr., **62**, 72–82.

110 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. Q. Rev. Biophys. **40**, 191–285.

111 Emsley, P. and Cowtan, K. (2004) Coot : model-building tools for molecular graphics research papers. Acta Crystallogr. D. Biol. Crystallogr. **60**, 2126–2132.

112 Blanchet, C. E. and Svergun, D. I. (2013) Small-angle X-ray scattering on biological macromolecules and nanocomposites in solution. Annu. Rev. Phys. Chem. **64**, 37–54.

113 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. J. Appl. Crystallogr., **36**, 1277–1282.

114 Koch, M. H. J., Vachette, P. and Svergun, D. I. (2003) Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. Q. Rev. Biophys., pp 147–227.

115 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. J. Appl. Crystallogr. **45**, 342–350.

116 Svergun, D. I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J. Appl. Crystallogr. **25**, 495–503.

117 Jacques, D. a and Trewhella, J. (2010) Small-angle scattering for structural biology-expanding the frontier while avoiding the pitfalls. Protein Sci. **19**, 642–57.

118 Svergun, D. I. and Koch, M. H. J. (2003) Small-angle scattering studies of biological macromolecules in solution. Reports Prog. Phys. **66**, 1735–1782.

119 Doniach, S. (2001) Changes in biomolecular conformation seen by small angle X-ray scattering. Chem. Rev. **101**, 1763–78.

120 Svergun, D. I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophys. J., Elsevier **76**, 2879–86.

121 Chacon, P., Moran, F., Diaz, J. F., Pantos, E. and Andreu, J. M. (1998) Low-resolution structures of proteins in solution retrieved from X-ray scattering with a genetic algorithim. Biophys. J. **74**, 2760–2775.

122 Franke, D. and Svergun, D. I. (2009) DAMMIF, a program for rapid *ab-initio* shape determination in small-angle scattering. J. Appl. Crystallogr. **42**, 342–346.

123 Volkov, V. V. and Svergun, D. I. (2003) Uniqueness of *ab initio shape* determination in small-angle scattering. J. Appl. Crystallogr. **36**, 860–864.

124 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. J. Am. Chem. Soc. **129**, 5656–64.

125 Brookes, E., Demeler, B. and Rocco, M. (2010) Developments in the US-SOMO bead modeling suite: new features in the direct residue-to-bead method, improved grid routines, and influence of accessible surface area screening. Macromol. Biosci. **10**, 746–53.

126 Cole, J. L., Lary, J. W., P Moody, T. and Laue, T. M. (2008) Analytical ultracentrifugation: sedimentation velocity and sedimentation equilibrium. Methods Cell Biol. **84**, 143–79.

127 Laue, T., Shah, B., Ridgeway, T. and Pelletier, S. (1992) Analytical Ultracentrifugation in Biochemistry and Polymer Science (Harding, S., Rowe, A., and Horton, J., eds.), p pp 90–125, The Royal Society of Chemistry, Cambridge.

128 Schuck, P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. Biophys. J. **78**, 1606–19.

129 Lebowitz, J., Lewis, M. S. and Schuck, P. (2002) Modern analytical ultracentrifugation in protein science : A tutorial review 2067–2079.

130 Schuck, P. (2003) On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. Anal. Biochem. **320**, 104–124.

131 Vistica, J., Dam, J., Balbo, A., Yikilmaz, E., Mariuzza, R. A., Rouault, T. A. and Schuck, P. (2004) Sedimentation equilibrium analysis of protein interactions with global implicit mass conservation constraints and systematic noise decomposition. Anal. Biochem. **326**, 234–56.

132 McRorie, D. K. and Voelker, P. J. (1993) Self-associating systems in the analytical ultracentrifuge., Beckman Instruments, Fullerton, CA.

133 Campellone, K. G., Robbins, D. and Leong, J. M. (2004) EspFU is a translocated EHEC effector that interacts with Tir and N-WASP and promotes Nck-independent actin assembly. Protein Sci. **7**, 217–28.

134 Riedel, C. U., Casey, P. G., Mulcahy, H., O'Gara, F., Gahan, C. G. M. and Hill, C. (2007) Construction of p16Slux, a novel vector for improved bioluminescent labeling of Gram-negative bacteria. Appl. Environ. Microbiol. **73**, 7092–5.

135 Wiles, S., Clare, S., Harker, J., Huett, A., Young, D., Dougan, G. and Frankel, G. (2004) Organ specificity, colonization and clearance dynamics *in vivo* following oral challenges with the murine pathogen *Citrobacter rodentium*. Cell. Microbiol. **6**, 963–72.

136 Gawthorne, J. A., Reddick, L. E., Akpunarlieva, S. N., Beckham, K. S. H., Christie, J. M., Alto, N. M., Gabrielsen, M. and Roe, A. J. (2012) Express your LOV: an engineered flavoprotein as a reporter for protein expression and purification. PLoS One **7**, e52962.

137 Bitto, E. and McKay, D. B. (2002) Crystallographic structure of SurA, a molecular chaperone that facilitates folding of outer membrane porins. Structure **10**, 1489–98.

138 Chan, W., Costantino, N., Li, R., Lee, S. C., Su, Q., Melvin, D., Court, D. L. and Liu, P. (2007) A recombineering based approach for high-throughput conditional knockout targeting vector construction. Nucleic Acids Res. **35**, e64.

139 Roe, A. J., Yull, H., Naylor, S. W., Martin, J., Smith, D. G. E., Gally, D. L. and Woodward, M. J. (2003) Heterogeneous surface expression of EspA translocon filaments by *Escherichia coli* O157:H7 Is controlled at the posttranscriptional level. Infect. Immun. **71**, 5900–5909.

140 Emmerson, J. R., Gally, D. L. and Roe, A. J. (2006) Generation of gene deletions and gene replacements in *Escherichia coli* O157:H7 using a temperature sensitive allelic exchange system. Biol. Proced. Online **8**, 153–62.

141 Blomfield, I. C., Vaughn, V., Rest, R. F. and Eisenstein, B. I. (1991) Allelic exchange in Escherichia coli using the *Bacillus subtilis sacB* gene and a temperature-sensitive pSC101 replicon. Mol. Microbiol. **5**, 1447–1457.

142 Wolfe, a J. and Berg, H. C. (1989) Migration of bacteria in semisolid agar. Proc. Natl. Acad. Sci. U. S. A. **86**, 6973–7.

143 Sullivan, M. J., Petty, N. K. and Beatson, S. a. (2011) Easyfig: a genome comparison visualizer. Bioinformatics **27**, 1009–10.

144 Turner, L., Ryu, W. S. and Berg, H. C. (2000) Real-time imaging of fluorescent flagellar filaments real-time imaging of fluorescent flagellar filaments. J. Bacteriol. **182**.

145 Campellone, K. G., Roe, A. J., Løbner-Olesen, A., Murphy, K. C., Magoun, L., Brady, M. J., Donohue-Rolfe, A., Tzipori, S., Gally, D. L., Leong, J. M., et al. (2007) Increased adherence and actin pedestal formation by dam-deficient enterohaemorrhagic *Escherichia coli* O157:H7. Mol. Microbiol. **63**, 1468–81.

146 Laue, T. M. and Shah, B. D. (1992) Computer-aided interpretation of analytical sedimentation data for proteins. In Analytical ultracentrifugation in biochemistry and polymer science (Harding, S. E., Rowe, A. J., and Horton, J. C., eds.), pp 90–125, Society of Chemistry, Cambridge.

147 Solovyova, A. S., Nöllmann, M., Mitchell, T. J. and Byron, O. (2004) The solution structure and oligomerization behavior of two bacterial toxins: pneumolysin and perfringolysin O. Biophys. J. **87**, 540–52.

148 Kozin, M. B. and Svergun, D. I. (2001) Automated matching of high- and low-resolution structural models. In Journal of Applied Crystallography, pp 33–41.

149 Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J. and Laue, E. D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins **59**, 687–96.

150 Mori, S., Abeygunawardana, C., O'Neil Johnson, M. and van Zijl, P. C. M. (1995) Improved sensitivity of HSQC spectra of exchaning protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation. J. Magn. Reson. **108**, 94–98.

151 Zhang, Y. (2008) I-TASSER server for protein 3D structure prediction. BMC Bioinformatics **9**, 40.

152 Laskowski, R. a., MacArthur, M. W., Moss, D. S. and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. **26**, 283–291.

153 Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., et al. (2011) Overview of the CCP4 suite and current developments. Acta Crystallogr. D. Biol. Crystallogr. **67**, 235–42.

154 McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. and Read, R. J. (2007) Phaser crystallographic software. J. Appl. Crystallogr. Crystallogr. **40**, 658–674.

155 Stein, N. (2008) CHAINSAW : a program for mutating pdb files used as templates in molecular replacement. J. Appl. Crystallogr., **41**, 641–643.

156 Murshudov, G. N., Vagin, A A and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D. Biol. Crystallogr. **53**, 240–55.

157 Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., Snoeyink, J., Richardson, J. S., et al. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res. **35**, W375–83.

158 Bernstein, F. C., Koetzle, T. F., Williams, G. J. B. and Meyer, E. F. (1978) The Protein Data Bank: a computer-based archival file for macromolecular structures. Arch. Biochem. Biophys. **185**, 584–591.

159 DeLano, W. (2002) PyMOL : An open-source molecular graphics tool. CCP4 Newsl. Protein Crystallogr. **40**:11.

160 Norville, I. H., Harmer, N. J., Harding, S. V, Fischer, G., Keith, K. E., Brown, K. A., Sarkar-Tyson, M. and Titball, R. W. (2011) A Burkholderia pseudomallei Mip-like protein has rapamycin inhibitable peptidyl-prolyl isomerase activity and has pleiotropic effects on virulence. Infect. Immun.

161 Morrison, J. F. (1969) Kinetics of the reversible inhibition of enzyme-catalysed reactions by tight-binding inhibitors. Biochim. Biophys. Acta **185**, 269–286.

162 Patridge, E. V and Ferry, J. G. (2006) WrbA from *Escherichia coli* and *Archaeoglobus fulgidus* Is an NAD(P)H : Quinone Oxidoreductase. J. Bacteriol. **188**.

163 Seaver, L. C. and Imlay, J. A. (2001) Hydrogen peroxide fluxes and compartmentalization inside growing *Escherichia coli*. J. Bacteriol. **183**, 7182–7189.

164 Mishra, S. and Imlay, J. (2012) Why do bacteria use so many enzymes to scavenge hydrogen peroxide? Arch. Biochem. Biophys. **525**, 145–60.

165 Baker, L. M. S. and Poole, L. B. (2003) Catalytic mechanism of thiol peroxidase from *Escherichia coli*. Sulfenic acid formation and overoxidation of essential CYS61. J. Biol. Chem. **278**, 9203–11.

166 Link, A. J., Robison, K. and Church, G. M. (1997) Comparing the predicted and observed properties of proteins encoded in the genome of *Escherichia coli* K-12. Electrophoresis **18**, 1259–313.

167 Wood, Z. A., Schröder, E., Robin Harris, J. and Poole, L. B. (2003) Structure, mechanism and regulation of peroxiredoxins. Trends Biochem. Sci. **28**, 32–40.

168 Cha, M. K., Kim, H. K. and Kim, I. H. (1995) Thioredoxin-linked "thiol peroxidase'' from periplasmic space of *Escherichia coli*. J. Biol. Chem. **270**, 28635–28641.

169 Tao, K. (2008) Subcellular localization and *in vivo* oxidation-reduction kinetics of thiol peroxidase in *Escherichia coli* . FEMS Microbiol. Lett. **289**, 41–5.

170 Cha, M. K., Kim, W. C., Lim, C. J., Kim, K. and Kim, I. H. (2004) *Escherichia coli* periplasmic thiol peroxidase acts as lipid hydroperoxide peroxidase and the principal antioxidative function during anaerobic growth. J. Biol. Chem. **279**, 8769–78.

171 Horst, S. a, Jaeger, T., Denkel, L. a, Rouf, S. F., Rhen, M. and Bange, F. C. (2010) Thiol peroxidase protects *Salmonella enterica* from hydrogen peroxide stress *in vitro* and facilitates intracellular growth. J. Bacteriol. **192**, 2929–32.

172 Somprasong, N., Jittawuttipoka, T., Duang Nkern, J., Romsang, A., Chaiyen, P., Schweizer, H. P., Vattanaviboon, P. and Mongkolsuk, S. (2012) *Pseudomonas aeruginosa* thiol peroxidase protects against hydrogen peroxide toxicity and displays atypical patterns of gene regulation. J. Bacteriol. **194**, 3904–12.

173 Choi, J., Choi, S., Choi, J., Cha, M. K., Kim, I. H. and Shin, W. (2003) Crystal structure of *Escherichia coli* thiol peroxidase in the oxidized state: insights into intramolecular disulfide formation and substrate binding in atypical 2-Cys peroxiredoxins. J. Biol. Chem. **278**, 49478–86.

174 Choi, J., Choi, S., Chon, J. K., Choi, J., Cha, M. K., Kim, I. H. and Shin, W. (2005) Crystal structure of the C107S/C112S mutant of yeast nuclear 2-Cys peroxiredoxin. Proteins Struct. Funct. Bioinforma. **61**, 1146–9.

175 Alphey, M. S., Bond, C. S., Tetaud, E., Fairlamb, A. H. and Hunter, W. N. (2000) The structure of reduced tryparedoxin peroxidase reveals a decamer and insight into reactivity of 2Cys-peroxiredoxins. J. Mol. Biol. **300**, 903–16.

176 Wood, Z. A., Poole, L. B. and Karplus, P. A. (2003) Peroxiredoxin evolution and the regulation of hydrogen peroxide signaling. Science (80-89). **300**, 650–3.

177 Cao, Z., Roszak, A. W., Gourlay, L. J., Lindsay, J. G. and Isaacs, N. W. (2005) Bovine mitochondrial peroxiredoxin III forms a two-ring catenane. Structure **13**, 1661–4.

178 Gabrielsen, M., Zetterström, C. E., Wang, D., Beckham, K. S. H., Elofsson, M., Isaacs, N. W. and Roe, A. J. (2010) Expression, purification, crystallization and initial X-ray diffraction analysis of thiol peroxidase from *Yersinia pseudotuberculosis*. Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun. **66**, 1606–9.

179 Gabrielsen, M., Beckham, K. S. H., Feher, V. A., Zetterström, C. E., Wang, D., Müller, S., Elofsson, M., Amaro, R. E., Byron, O. and Roe, A. J. (2012) Structural characterisation of Tpx from *Yersinia pseudotuberculosis* reveals insights into the binding of salicylidene acylhydrazide compounds. PLoS One **7**, e32217.

180 Beckham, K. S. H., Byron, O., Roe, A. J. and Gabrielsen, M. (2012) The structure of an orthorhombic crystal form of a “forced reduced” thiol peroxidase reveals lattice formation aided by the presence of the affinity tag. Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun. **68**, 522–6.

181 Larkin, M. a, Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. a, McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., et al. (2007) Clustal W and Clustal X version 2.0. Bioinformatics **23**, 2947–8.

182 Hall, A., Sankaran, B., Poole, L. B. and Karplus, P. A. (2009) Structural changes common to catalysis in the Tpx peroxiredoxin subfamily. J. Mol. Biol. **393**, 867–81.

183 Pan, J. L. and Bardwell, J. C. A. (2006) The origami of thioredoxin-like folds. Protein Sci. **15**, 2217–2227.

184 Krissinel, E. and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. **372**, 774–97.

185 Svergun, D. I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophys. J. **76**, 2879–86.

186 Svergun, D., Barberato, C. and Koch, M. H. J. (1995) CRYSOL – a program to evaluate X-ray solution scattering of biological macromolecules from atomic oordinates. J. Appl. Crystallogr. **28**, 768–773.

187 Petoukhov, M. V and Svergun, D. I. (2005) Global rigid body modeling of macromolecular complexes against small-angle scattering data. Biophys. J. **89**, 1237–50.

188 Hayward, S. and Berendsen, H. J. (1998) Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme. Proteins Struct. Funct. Bioinforma. **30**, 144–54.

189 Sammartano, L. J., Tuveson, R. W. and Davenport, R. (1986) Control of sensitivity to inactivation by H202 and broad-spectrum near-UV radiation by the *Escherichia coli* katF locus. J. Bacteriol. **168**, 13–21.

190 Sperandio, V., Mellies, J. L., Nguyen, W., Shin, S. and Kaper, J. B. (1999) Quorum sensing controls expression of the type III secretion gene transcription and protein

secretion in enterohemorrhagic and enteropathogenic *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. **96**, 15196–201.

191 Haussmann, C., Rohdich, F., Schmidt, E., Bacher, A. and Richter, G. (1998) Biosynthesis of pteridines in *Escherichia coli*. J. Biol. Chem. **273**, 17418–17424.

192 Girgis, H. S., Hottes, A. K. and Tavazoie, S. (2009) Genetic architecture of intrinsic antibiotic susceptibility. PLoS One **4**, e5629.

193 Ploom, T., Haussmann, C., Hof, P., Steinbacher, S., Bacher, a, Richardson, J. and Huber, R. (1999) Crystal structure of 7,8-dihydroneopterin triphosphate epimerase. Structure **7**, 509–16.

194 Yang, W., Ni, L. and Somerville, R. L. (1993) A stationary-phase protein of *Escherichia coli* that affects the mode of association between the trp repressor protein and operator-bearing DNA. Proc. Natl. Acad. Sci. U. S. A. **90**, 5796–800.

195 Gorman, J. and Shapiro, L. (2005) Crystal structures of the tryptophan repressor binding protein WrbA and complexes with flavin mononucleotide. Protein Sci. **14**, 3004–3012.

196 Grandori, R., Khalifah, P., Boice, J. A., Fairman, R., Giovanielli, K. and Carey, J. (1998) Biochemical characterization of WrbA, founding member of a new family of multimeric flavodoxin-like proteins. J. Biol. Chem. **273**, 20960–20966.

197 Gabrielsen, M., Beckham, K. S. H., Cogdell, R. J., Byron, O. and Roe, A. J. (2012) FolX from *Pseudomonas aeruginosa* is octameric in both crystal and solution. FEBS Lett., pp 1160–5.

198 Wolfova, J., Smatanova, I. K., Brynda, J., Mesters, J. R., Lapkouski, M., Kuty, M., Natalello, A., Chatterjee, N., Chern, S.-Y., Ebbel, E., et al. (2009) Structural organization of WrbA in apo- and holoprotein crystals. Biochim. Biophys. Acta, Elsevier B.V. **1794**, 1288–98.

199 Andrade, S. L. A., Patridge, E. V, Ferry, J. G. and Einsle, O. (2007) Crystal structure of the NADH:quinone oxidoreductase WrbA from *Escherichia coli*. J. Bacteriol. **189**, 9101–7.

200 Wallace, A. C., Laskowski, R. A. and Thornton, J. M. (1995) LIGPLOT : a program to generate schematic diagrams of protein-ligand interactions. Protein Eng. **8**, 127–134.

201 Laskowski, R. a and Swindells, M. B. (2011) LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. J. Chem. Inf. Model. **51**, 2778–86.

202 Hullmann, J., Patzer, S. I., Römer, C., Hantke, K. and Braun, V. (2008) Periplasmic chaperone FkpA is essential for imported colicin M toxicity. Mol. Microbiol. **69**, 926–37.

203 Obi, I. R. and Francis, M. S. (2013) Demarcating SurA activities required for outer membrane targeting of *Yersinia pseudotuberculosis* adhesins. Infect. Immun. **81**, 2296–308.

204 Obi, I. R., Nordfelth, R. and Francis, M. S. (2011) Varying dependency of periplasmic peptidylprolyl cis-trans isomerases in promoting *Yersinia pseudotuberculosis* stress tolerance and pathogenicity. Biochem. J. **439**, 321–32.

205 Behrens-Kneip, S. (2010) The role of SurA factor in outer membrane protein transport and virulence. Int. J. Med. Microbiol. **300**, 421–8.

206 Schmid, F. X. (1993) PROLYL ISOMERASE: Enzymatic catalysis of slow protein-folding reactions. Annu. Rev. Biophys. Biomol. Struct. **22**, 123:143.

207 Fischer, G. and Bang, H. (1985) The refoldong of urea-denatured ribonuclease A is catalyzed by peptidyl-proloyl *cis-trans* isomerase. Biochim. Biophys. Acta **828**, 39–42.

208 Fischer, G. (1994) Peptidyl-Prolyl *cis/trans* isomerases and their effectors. Angew. Chemie Int. **33**, 1415–1436.

209 Göthel, S. and Marahiel, M. (1999) Peptidyl-prolyl cis-trans isomerases , a superfamily of ubiquitos folding catalysts. Cell. Mol. Life Sci. **55**, 423–436.

210 Rahfeld, J., Rucknagel, K. P., Stoller, G., Horne, S. M., Schierhorn, A., Young, K. D. and Fischer, G. (1996) Isolation and amino acid sequence of a new 22-kDa FKBP-like peptidyl-prolyl *cis/trans*-isomerase of *Escherichia coli*. J. Biol. Chem. **271**, 22130–22138.

211 Rahfeld, J. U., Rücknagel, K. P., Schelbert, B., Ludwig, B., Hacker, J., Mann, K. and Fischer, G. (1994) Confirmation of the existence of a third family among peptidyl-prolyl *cis/trans* isomerases. Amino acid sequence and recombinant production of parvulin. FEBS Lett. **352**, 180–4.

212 Kühlewein, A. and Voll, G. (2004) Solution structure of Escherichia coli Par10: The prototypic member of the Parvulin family of peptidyl-prolyl *cis/trans* isomerases. Protein Sci. **13**, 2378–2387.

213 Tormo, a, Almirón, M. and Kolter, R. (1990) *surA*, an *Escherichia coli* gene essential for survival in stationary phase. J. Bacteriol. **172**, 4339–47.

214 Rouviere, P. E. and Gross, C. a. (1996) SurA, a periplasmic protein with peptidyl-prolyl isomerase activity, participates in the assembly of outer membrane porins. Genes Dev. **10**, 3170–3182.

215 Lazar, S. W. and Kolter, R. (1996) SurA assists the folding of *Escherichia coli* outer membrane proteins. J. Bacteriol. **178**, 1770–3.

216 Xu, X., Wang, S., Hu, Y.-X. and McKay, D. B. (2007) The periplasmic bacterial molecular chaperone SurA adapts its structure to bind peptides in different

conformations to assert a sequence preference for aromatic residues. J. Mol. Biol. **373**, 367–81.

217 Behrens, S., Maier, R., de Cock, H., Schmid, F. X. and Gross, C. A. (2001) The SurA periplasmic PPIase lacking its parvulin domains functions in vivo and has chaperone activity. EMBO J. **20**, 285–94.

218 Wu, T., Malinverni, J., Ruiz, N., Kim, S., Silhavy, T. J. and Kahne, D. (2005) Identification of a multicomponent complex required for outer membrane biogenesis in *Escherichia coli*. Cell **121**, 235–45.

219 Justice, S. S., Hunstad, D. A., Harper, J. R., Duguay, A. R., Pinkner, J. S., Bann, J., Frieden, C., Silhavy, T. J. and Hultgren, S. J. (2005) Periplasmic peptidyl prolyl cis-trans isomerases are not essential for viability , but SurA is required for pilus biogenesis in *Escherichia coli* . J. Bacteriol. **187**, 7680–7686.

220 Missiakas, D., Betton, J. and Raina, S. (1996) New components of protein folding in extracytoplasmic compartments of *Escherichia coli* SurA, FkpA and Skp/Omp. Mol. Microbiol. **21**, 871–884.

221 Danese, P. N. and Silhavy, T. J. (1997) The sigma(E) and the Cpx signal transduction systems control the synthesis of periplasmic protein-folding enzymes in *Escherichia coli*. Genes Dev. **11**, 1183–1193.

222 Kino, T., Hatanaka, H., Miyata, S., Inamura, N., Nishyama, M., Yajima, T., Goto, T., Okuhara, M., Kohsaka, M., Aoki, H., et al. FK-506, a novel immunosuppressant isolated from a Streptomyces. II: Immunosuppressive effect of FK-506 *in vitro*. J. Antibiot. (Tokyo). **40**, 1256–1265.

223 Harding, M. W., Galat, A., Uehling, D. E. and Schreiber, S. L. (1989) A receptor for the immuno-supressant FK506 is a *cis-trans* peptidyl-prolyl isomerase. Nature **341**, 758–760.

224 Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. and Clardy, J. (1993) Atomic structure of the human immuophillin FKBP-12 complexes with FK506 and rapamycin. J. Mol. Biol. **229**, 105–124.

225 Somarelli, J. and Lee, S. (2008) Structure-based classification of 45 FK506-binding proteins. Proteins Struct. Funct. Bioinforma. **72**, 197–208.

226 Kurek, I., Aviezer, K., Erel, N., Herman, E. and Breiman, A. (1999) The wheat peptidyl prolyl *cis-trans* isomerase FKBP77 is heat induced and developmentally regulated. Plant Physiol. **119**, 693–704.

227 Ramm, K. and Plückthun, A. (2000) The periplasmic *Escherichia coli* peptidylprolyl *cis/trans*-isomerase FkpA. II. Isomerase-independent chaperone activity *in vitro*. J. Biol. Chem. **275**, 17106–13.

228 Ramm, K. and Plückthun, A. (2001) High enzymatic activity and chaperone function are mechanistically related features of the dimeric *E. coli* peptidyl-prolyl-isomerase FkpA. J. Mol. Biol. **310**, 485–98.

229 Horne, S. M., Kottom, T. J., Nolan, L. K. and Young, K. D. (1997) Decreased intracellular survival of an fkpA mutant of *Salmonella typhimurium* Copenhagen. Infect. Immun. **65**, 806–10.

230 Helbig, S., Patzer, S. I., Schiene-Fischer, C., Zeth, K. and Braun, V. (2011) Activation of colicin M by the FkpA prolyl *cis-trans* isomerase/chaperone. J. Biol. Chem. **286**, 6280–90.

231 Kumar, J. K., Tabor, S. and Richardson, C. C. (2004) Proteomic analysis of thioredoxin-targeted proteins in *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. **101**, 3759–64.

232 Suzuki, Y., Haruki, M., Takano, K., Morikawa, M. and Kanaya, S. (2004) Possible involvement of an FKBP family member protein from a psychrotrophic bacterium *Shewanella* sp. SIB1 in cold-adaptation. Eur. J. Biochem. **271**, 1372–1381.

233 Suzuki, Y., Win, O. Y., Koga, Y., Takano, K. and Kanaya, S. (2005) Binding analysis of a psychrotrophic FKBP22 to a folding intermediate of protein using surface plasmon resonance. FEBS Lett. **579**, 5781–4.

234 Budiman, C., Angkawidjaja, C., Motoike, H., Koga, Y., Takano, K. and Kanaya, S. (2011) Crystal structure of N-domain of FKBP22 from *Shewanella* sp. SIB1: dimer dissociation by disruption of Val-Leu knot. Protein Sci. **20**, 1755–64.

235 Suzuki, Y., Takano, K. and Kanaya, S. (2005) Stabilities and activities of the N- and C-domains of FKBP22 from a psychrotrophic bacterium overproduced in *Escherichia coli*. FEBS J. **272**, 632–42.

236 Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V and Mann, M. (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat. Protoc. **1**, 2856–60.

237 Riboldi-tunnicliffe, A., König, B., Jessen, S., Weiss, M. S., Rahfeld, J., Hacker, J., Fischer, G. and Hilgenfeld, R. (2001) Crystal structure of Mip , a prolylisomerase from *Legionella pneumophila*. Nature **8**, 779–783.

238 Arnold, K., Bordoli, L., Kopp, J. and Schwede, T. (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics **22**, 195–201.

239 Koltin, Y., Faucette, L., Bergsma, D. J., Levy, M. a, Cafferkey, R., Koser, P. L., Johnson, R. K. and Livi, G. P. (1991) Rapamycin sensitivity in *Saccharomyces cerevisiae* is mediated by a peptidyl-prolyl *cis-trans* isomerase related to human FK506-binding protein. Mol. Cell. Biol. **11**, 1718–23.

240 Arié, J. P., Sassoon, N. and Betton, J. M. (2001) Chaperone function of FkpA, a heat shock prolyl isomerase, in the periplasm of *Escherichia coli*. Mol. Microbiol. **39**, 199–210.

241 Sklar, J. G., Wu, T., Kahne, D. and Silhavy, T. J. (2007) Defining the roles of the periplasmic chaperones SurA, Skp, and DegP in *Escherichia coli*. Genes Dev. **21**, 2473–84.

242 Fardini, Y., Trotereau, J., Bottreau, E., Souchard, C., Velge, P. and Virlogeux-Payant, I. (2009) Investigation of the role of the BAM complex and SurA chaperone in outer-membrane protein biogenesis and type III secretion system expression in *Salmonella*. Microbiology **155**, 1613–22.

243 Saul, F. A., Arié, J.-P., Vulliez-le Normand, B., Kahn, R., Betton, J.-M. and Bentley, G. A. (2004) Structural and functional studies of FkpA from *Escherichia coli*, a *cis*/*trans* peptidyl-prolyl isomerase with chaperone activity. J. Mol. Biol. **335**, 595–608.

244 Jana, B. and Sau, S. (2012) The helix located between the two domains of a mip-like peptidyl-prolyl *cis-trans* isomerase is crucial for its structure, stability, and protein folding ability. Biochemistry **51**, 7930–9.

245 Clark, D. and Cronan, J. E. (1980) *Escherichia coli* mutants with altered control of alcohol dehydrogenase and nitrate reductase. J. Bacteriol. **141**, 177–83.

246 Kessler, D., Herth, W. and Knappe, J. (1992) Ultrastructure and pyruvate formate-lyase radical quenching property of the multienzymic AdhE protein of *Escherichia coli* . J. Biol. Chem. **267**, 18073–9.

247 Nnyepi, M. R., Peng, Y. and Broderick, J. B. (2007) Inactivation of *E. coli* pyruvate formate-lyase: role of AdhE and small molecules. Arch. Biochem. Biophys. Biophys. **459**, 1–9.

248 Salmon, K., Hung, S., Mekjian, K., Baldi, P., Hatfield, G. W. and Gunsalus, R. P. (2003) Global gene expression profiling in *Escherichia coli K12*. The effects of oxygen availability and FNR. J. Biol. Chem. **278**, 29837–55.

249 Chen, Y. M. and Lin, E. C. (1991) Regulation of the *adhE* gene, which encodes ethanol dehydrogenase in *Escherichia coli*. J. Bacteriol. **173**, 8009–13.

250 Mikulskis, A., Aristarkhov, A. and Lin, E. C. (1997) Regulation of expression of the ethanol dehydrogenase gene (*adhE*) in *Escherichia coli* by catabolite repressor activator protein Cra. J. Bacteriol. **179**, 7129–7134.

251 Membrillo-Hernández, J., Kwon, O., De Wulf, P., Finkel, S. E. and Lin, E. C. (1999) Regulation of *adhE* (encoding ethanol oxidoreductase) by the Fis protein in *Escherichia coli*. J. Bacteriol. **181**, 7390–3.

252 Kolb, A., Busby, S., Buc, H., Garges, S. and Adhya, S. (1993) Transcriptional regulation by cAMP and its receptor protein. Annu. Rev. Biochem. **62**, 749–795.

253 Leonardo, M. R., Dailly, Y. and Clark, D. P. (1996) Role of NAD in regulating the *adhE* gene of *Escherichia coli*. Mol. Microbiol. **178**, 6013–6018.

254 Aristarkhov, A., Mikulskis, A., Belasco, J. G. and Lin, E. C. (1996) Translation of the adhE transcript to produce ethanol dehydrogenase requires RNase III cleavage in *Escherichia coli*. J. Bacteriol. **178**, 4327–4332.

255 Echave, P., Tamarit, J., Cabiscol, E. and Ros, J. (2003) Novel antioxidant role of alcohol dehydrogenase E from *Escherichia coli*. J. Biol. Chem. **278**, 30193–8.

256 Pineda, E., Encalada, R., Olivos-García, A., Néquiz, M., Moreno-Sánchez, R. and Saavedra, E. (2013) The bifunctional aldehyde-alcohol dehydrogenase controls ethanol and acetate production in *Entamoeba histolytica* under aerobic conditions. FEBS Lett. **587**, 178–84.

257 Membrillo-Hernandez, J., Echave, P., Cabiscol, E., Tamarit, J., Ros, J. and Lin, E. C. (2000) Evolution of the *adhE* gene product of *Escherichia coli* from a functional reductase to a dehydrogenase. Genetic and biochemical studies of the mutant proteins. J. Biol. Chem. **275**, 33869–75.

258 Tamarit, J. (1998) Identification of the major oxidatively damaged proteins in *Escherichia coli* cells exposed to oxidative stress. J. Biol. Chem. **273**, 3027–3032.

259 Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.-F., Guindon, S., Lefort, V., Lescot, M., et al. (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. **36**, W465–9.

260 Baumler, A. J., Kusters, J. G., Stojiljkovic, I. and Heffron, F. (1994) *Salmonella typhimurium* loci involved in survival within macrophages. Infect. Immun. **62**.

261 Abernathy, J., Corkill, C., Hinojosa, C., Li, X. and Zhou, H. (2013) Deletions in the pyruvate pathway of *Salmonella Typhimurium* alter SPI1-mediated gene expression and infectivity. J. Anim. Sci. Biotechnol. **4**, 5.

262 Rietsch, A. and Mekalanos, J. J. (2006) Metabolic regulation of type III secretion gene expression in *Pseudomonas aeruginosa*. Mol. Microbiol. **59**, 807–820.

263 Pappin, C., Hojrupt, P. and Bleasbyl, A. J. (1993) Rapid iIdentification of proteins fingerprinting by peptide-inass. Curr. Biol. **3**, 327–332.

264 Roe, A. J., Naylor, S. W., Spears, K. J., Yull, H. M., Dransfield, T. a, Oxford, M., McKendrick, I. J., Porter, M., Woodward, M. J., Smith, D. G. E., et al. (2004) Co-ordinate single-cell expression of LEE4- and LEE5-encoded proteins of *Escherichia coli* O157:H7. Mol. Microbiol. **54**, 337–52.

265 Mandin, P. and Guillier, M. (2013) Expanding control in bacteria: interplay between small RNAs and transcriptional regulators to control gene expression. Curr. Opin. Microbiol., Elsevier Ltd **16**, 125–32.

266 Schneider, C. a, Rasband, W. S. and Eliceiri, K. W. (2012) NIH Image to ImageJ: 25 years of image analysis. Nat. Methods, Nature Publishing Group **9**, 671–675.

267 Nataro, J. P. and Kaper, J. B. (1998) Diarrheagenic *Escherichia coli*. Clin. Microbiol. Rev. **11**, 142–201.

268 Booth, I. A. N. R. (1985) Regulation of Cytoplasmic pH in Bacteria. Microbiol. Rev. **49**, 359–378.

269 Fukuda, S., Toh, H., Hase, K., Oshima, K., Nakanishi, Y., Yoshimura, K., Tobe, T., Clarke, J. M., Topping, D. L., Suzuki, T., et al. (2011) *Bifidobacteria* can protect from enteropathogenic infection through production of acetate. Nature, **469**, 543–7.

270 Ramakrishnan, R., Schuster, M. and Bourret, R. B. (1998) Acetylation at Lys-92 enhances signaling by the chemotaxis response regulator protein CheY. Proc. Natl. Acad. Sci. U. S. A. **95**, 4918–23.

271 Hu, L. I., Lima, B. P. and Wolfe, A. J. (2010) Bacterial protein acetylation: the dawning of a new age. Mol. Microbiol. **77**, 15–21.

272 Wang, Q., Zhang, Y., Yang, C., Xiong, H., Lin, Y., Yao, J., Li, H., Xie, L., Zhao, W., Yao, Y., et al. (2010) Acetylation of metabolic enzymes coordinates carbon source utilization and metabolic flux. Science (80-89). **327**, 1004–7.

273 Soutourina, O. and Kolb, A. (1999) Multiple control of flagellum biosynthesis in *Escherichia coli*: role of H-NS protein and the cyclic AMP-catabolite activator protein complex in transcription of the *flhDC*. J. Bacteriol. **181**, 7500–7508.

274 Gosset, G., Zhang, Z., Nayyar, S., Cuevas, W. A. and Saier, M. H. (2004) Transcriptome analysis of crp-dependent catabolite control of gene expression in *Escherichia coli*. J. Bacteriol. **186**, 3516–3524.

275 Porter, S. L., Wadhams, G. H. and Armitage, J. P. (2011) Signal processing in complex chemotaxis pathways. Nat. Rev. Microbiol. **9**, 153–65.

276 Wiles, S., Clare, S., Harker, J., Huett, A., Young, D., Dougan, G. and Frankel, G. (2004) Organ specificity, colonization and clearance dynamics *in vivo* following oral challenges with the murine pathogen *Citrobacter rodentium*. Cell. Microbiol. **6**, 963–72.

277 Brown, S. D., Guss, A. M., Karpinets, T. V, Parks, J. M., Smolin, N., Yang, S., Land, M. L., Klingeman, D. M., Bhandiwad, A., Rodriguez, M., et al. (2011) Mutant alcohol dehydrogenase leads to improved ethanol tolerance in *Clostridium thermocellum* . Proc. Natl. Acad. Sci. U. S. A. **108**, 13752–7.

278 Saini, S., Slauch, J. M., Aldridge, P. D. and Rao, C. V. (2010) Role of cross talk in regulating the dynamic expression of the flagellar Salmonella pathogenicity island 1 and type 1 fimbrial genes. J. Bacteriol. **192**, 5767–77.

279 Soutourina, O. a and Bertin, P. N. (2003) Regulation cascade of flagellar expression in Gram-negative bacteria. FEMS Microbiol. Rev. **27**, 505–523.

280 Yin, X., Zhu, J., Feng, Y., Chambers, J. R., Gong, J. and Gyles, C. L. (2011) Differential gene expression and adherence of *Escherichia coli* O157:H7 in vitro and in ligated pig intestines. PLoS One **6**, e17424.

281 Sesto, N., Wurtzel, O., Archambaud, C., Sorek, R. and Cossart, P. (2013) The excludon: a new concept in bacterial antisense RNA-mediated gene regulation. Nat. Rev. Microbiol., pp 75–82.

282 De Lay, N. and Gottesman, S. (2012) A complex network of small non-coding RNAs regulate motility in *Escherichia coli*. Mol. Microbiol. **86**, 524–38.

283 Van Kessel, J. C., Rutherford, S. T., Shao, Y., Utria, A. F. and Bassler, B. L. (2013) Individual and combined roles of the master regulators AphA and LuxR in control of the *Vibrio harveyi* quorum-sensing regulon. J. Bacteriol. **195**, 436–43.

284 Fraser, G. M. and Hughes, C. (1999) Swarming motility. Curr. Opin. Microbiol. **2**, 630–5.

285 Barak, R., Abouhamad, W. N. and Eisenbach, M. (1998) Both acetate kinase and acetyl coenzyme A synthetase are involved in acetate-stimulated change in the direction of flagellar rotation in *Escherichia coli*. J. Bacteriol. **180**, 985–988.

286 Barak, R., Welch, M., Yanovsky, a, Oosawa, K. and Eisenbach, M. (1992) Acetyladenylate or its derivative acetylates the chemotaxis protein CheY in vitro and increases its activity at the flagellar switch. Biochemistry **31**, 10099–107.

287 Horita, Y., Takii, T., Yagi, T., Ogawa, K., Fujiwara, N., Inagaki, E., Kremer, L., Sato, Y., Kuroishi, R., Lee, Y., et al. (2012) Antitubercular activity of disulfiram, an antialcoholism drug, against multidrug- and extensively drug-resistant *Mycobacterium tuberculosis* isolates. Antimicrob. Agents Chemother. **56**, 4140–5.

288 Petty, N. K., Feltwell, T., Pickard, D., Clare, S., Toribio, A. L., Fookes, M., Roberts, K., Monson, R., Nair, S., Kingsley, R. a, et al. (2011) *Citrobacter rodentium* is an unstable pathogen showing evidence of significant genomic flux. PLoS Pathog. **7**, e1002018.

289 Wolfe, A. (2005) The Acetate Switch. Microbiol. Mol. Biol. Rev. **69**, 12–50.

290 Fukuda, S., Toh, H., Taylor, T. D., Ohno, H. and Hattori, M. (2012) Acetate-producing *bifidobacteria* protect the hose for enteropathogenic infection via carbohydrate transporters. Gut Microbes **3**, 449–454.

291 Hempel, J., von Bahr-Lindström, H. and Jörnvall, H. (1984) Aldehyde dehydrogenase from human liver. Primary structure of the cytoplasmic isoenzyme. Eur. J. Biochem. **141**, 21–35.

Publications

crystallization communications

Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Expression, purification, crystallization and initial X-ray diffraction analysis of thiol peroxidase from *Yersinia pseudotuberculosis*

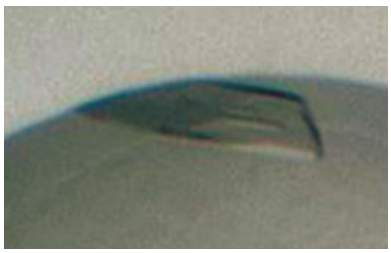
Mads Gabrielsen,[a]*‡ Caroline E. Zetterström,[b]‡ Dai Wang,[a]‡ Katherine S. H. Beckham,[a]‡ Mikael Elofsson,[b] Neil W. Isaacs[c] and Andrew J. Roe[a]

[a]Institute of Infection, Immunity and Inflammation, College of Medical, Veterinary and Life Sciences, Glasgow Biomedical Research Centre, Glasgow G12 8QQ, Scotland, [b]Department of Chemistry, Umeå University, SE-90187 Umeå, Sweden, and [c]Department of Chemistry, University of Glasgow, Glasgow G12 8QQ, UK

‡ These authors contributed equally to this work.

Correspondence e-mail: mads.gabrielsen@glasgow.ac.uk

Received 6 August 2010
Accepted 4 October 2010



© 2010 International Union of Crystallography
All rights reserved

Thiol peroxidase is an atypical 2-Cys peroxiredoxin that reduces alkyl hydroperoxides. Wild-type and C61S mutant protein have been recombinantly expressed in *Escherichia coli* and purified using nickel-affinity chromatography. Initial crystallization trials yielded three crystal forms in three different space groups ($P2_1$, $P6_4$ and $P2_12_12_1$) both in the presence and the absence of DTT.

1. Introduction

Yersinia pseudotuberculosis is a Gram-negative bacterium that is capable of causing a tuberculosis-like disease in humans and animals, as well as being a model organism for *Y. pestis* (Robins-Browne & Hartland, 2003). Thiol peroxidase (Tpx; p20; UniProt accession No. Q66A71) is an atypical 2-Cys peroxiredoxin that uses the redox potential from thioredoxin reductase and thioredoxin I to reduce alkyl hydroperoxides (Baker & Poole, 2003). Tpx was initially suggested to be localized to the periplasm and to be involved in the removal of lipid hydroxyperoxides produced by oxidative stress (Cha *et al.*, 1995). However, recent studies of fractionated cells show that Tpx is cytoplasmic and is released into the periplasm as a response to stress (Tao, 2008).

Tpx contains three cysteines, two of which (Cys61 and Cys95, with Cys95 being the resolving cysteine) form the redox-active pair (Baker & Poole, 2003). Despite the presence of a third cysteine, there is no covalent dimerization in the oxidized state as is observed for most 2-Cys peroxiredoxins (Baker & Poole, 2003). The structure of Tpx from *Escherichia coli* has been elucidated in the oxidized form and the mutated C61S form (Hall *et al.*, 2009; Choi *et al.*, 2003) and shows a sequence identity of approximately 80% to *Y. pseudotuberculosis* Tpx (*yp*Tpx). Here, we describe the crystallization of *yp*Tpx in three different crystal forms grown under different conditions, as well as the crystallization of the catalytically inactive (Baker & Poole, 2003) mutant *yp*TpxC61S.

2. Material and methods

2.1. Cloning

The whole gene encoding *yp*Tpx was amplified from *Y. pseudotuberculosis* YPIII pIB102 genomic DNA using proofreading DNA polymerase (Accuprime Pfx Supermix, Invitrogen) in conjunction with the primer pair *yp*Tpx5 and *yp*Tpx3 (CACCATGACACAGACCGTACATTTC and TTATTTCAGTGCAGCCAGCG, respectively). The amplified product was cloned into the TOPO pET-151 (Invitrogen) expression vector, thereby fusing a hexahistidine tag and TEV cleavage site to the N-terminus of *yp*Tpx.

The mutant *yp*TpxC61S was obtained by site-directed mutagenesis (QuikChange Site-Directed Mutagenesis Kit, Stratagene) using the primer pair mut5 and mut3 (GATACCGGCGTTT**CG**GCCGCCTCGGTACG and CGTACCGAGGCGGC**CG**AAACGCCGGTATC,

respectively). The DNA sequences of the resultant expression vectors (pDW121 for the wild type and pDW121-C61S for the mutant) were confirmed using DNA sequencing.

2.2. Expression and purification of *yp*Tpx

pDW121 or pDW121-C61S was transformed into *E. coli* BL21 (λDE3) cells and grown in 400 ml LB medium containing ampicillin (100 mg l^{-1}). The cells were grown at 310 K until an optical density (OD_{600}) of approximately 0.6 was reached, upon which expression was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 m*M* and the culture was left to grow overnight at 298 K. The cells were harvested and stored at 253 K until required. The cells were suspended in 20 m*M* Tris pH 7.5, 500 m*M* NaCl (buffer *A*) with 10 mg DNAse (Sigma) and lysed in a French press. The lysate was cleared by low-speed centrifugation at 8000*g* and loaded onto a 5 ml Ni^{2+} HisTrap column (GE Healthcare) pre-equilibrated in buffer *A*, washed with five column volumes (CV) of buffer *A* containing 20 m*M* imidazole followed by 3 CV buffer *A* containing 100 m*M* imidazole and finally eluted with 2 CV buffer *A* containing 300 m*M* imidazole. The *yp*Tpx-containing fractions were identified on an SDS–PAGE gel and pooled. The purity based on the SDS–PAGE was deemed to be greater than 95%.

2.3. Crystallization

Purified protein was exchanged into 20 m*M* Tris pH 7.5, 50 m*M* NaCl (buffer *B*) with or without dithiothreitol (DTT) using a PD-10 desalting column (GE Healthcare) before being concentrated to approximately 8 mg ml^{-1} based on the absorbance at 280 nm (using an absorption coefficient of 0.261 M^{-1} cm^{-1}). Initial crystallization screens were set up using sitting-drop vapour diffusion, with drops consisting of 500 nl protein solution and 500 nl reservoir solution. All trays were kept at 293 K. Three crystal forms were observed of wild-type *yp*Tpx (crystal forms 1, 2 and 3) and one of *yp*TpxC61S (crystal form M).

Crystal form 1 appeared in condition No. 46 of the PEG/Ion screen [Hampton Research; 20% polyethylene glycol (PEG) 3350, 0.2 *M* sodium citrate tribasic]. This hexagonal rod grew to dimensions of approximately 150 × 50 × 20 µm within 5 d (Fig. 1*a*). The protein was kept in buffer *B* with the addition of 2 m*M* DTT.

Crystal form 2 appeared in condition No. 3 of the PEG/Ion 2 screen (Hampton Research; 0.1 *M* sodium malonate pH 5, 12% PEG 3350). The protein sample was kept in buffer *B* with 2 m*M* DTT. A single crystal, which appeared within 24 h, with dimensions of 900 × 70 × 50 µm (Fig. 1*b*) was broken into smaller pieces for data collection.

Crystal form 3 appeared in condition No. 24 of the JCSG+ Screen (Molecular Dimensions; 0.2 *M* tripotassium citrate, 20% PEG 3350). The protein was kept in buffer *B* with no additives in the hope of obtaining crystals of the oxidized protein sample. The crystals measured 300 × 80 × 20 µm and appeared within 7 d (Fig. 1*c*).

Crystals of *yp*TpxC61S appeared in condition No. 47 of the PEG/Ion 2 screen (Hampton Research; 12% PEG 3350, 0.05 *M* Na HEPES pH 7.0, 1% tryptone) and belonged to crystal form 2. The protein sample was kept in buffer *B* with no additives. Crystals appeared within 24 h with dimensions of 2000 × 70 × 60 µm and a small fragment was broken off for data collection (Fig. 1*d*).

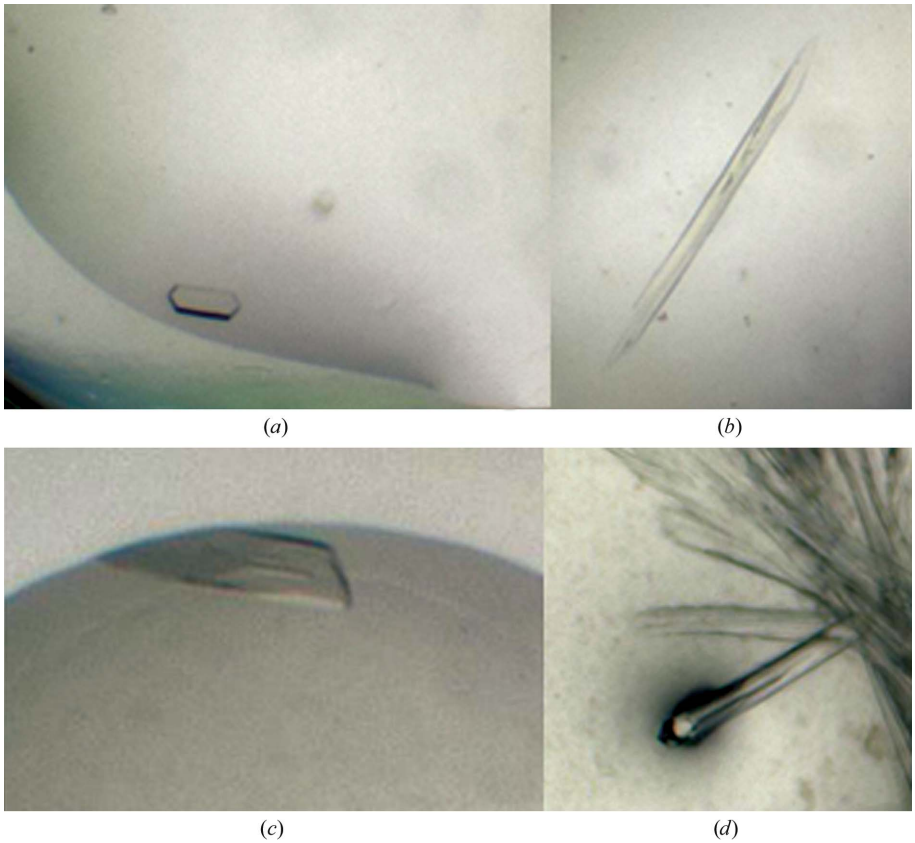


Figure 1
Crystals of *Y. pseudotuberculosis* Tpx obtained using different crystallization conditions. (*a*) Crystal form 1 obtained using 20% polyethylene glycol (PEG) 3350, 0.2 *M* sodium citrate tribasic, (*b*) crystal form 2 obtained using 0.1 *M* sodium malonate pH 5, 12% PEG 3350, (*c*) crystal form 3 obtained using 0.2 *M* tripotassium citrate, 20% PEG 3350 and (*d*) crystals of the mutant protein in crystal form 2 obtained using 1% tryptone, 0.05 *M* Na HEPES pH 7.0, 12% PEG 3350.

2.4. Data collection and processing

Crystals were plunged into a stream of cooled nitrogen gas (110 K; Oxford Cryosystems) with no further cryoprotection. Crystal forms 1–3 and M were transported to the Diamond Light Source for diffraction studies, in which data were collected on an ADSC Q315 CCD detector. Data were collected with 1° oscillations for a total of between 150 and 360 images at a wavelength of 0.9763 Å (Fig. 2).

Data were processed using *MOSFLM* (Leslie, 1992) and scaled and merged using *SCALA* (Evans, 1993) from the *CCP*4 suite of programs (Collaborative Computational Project, Number 4, 1994) or *d*TREK* (Pflugrath, 1999). The space groups were confirmed using *POINTLESS* (Evans, 2006).

3. Results

The crystals were found to grow in three different space groups. Crystal form 1 crystallized in space group $P2_1$ (unit-cell parameters a = 64.86, b = 92.07, c = 85.60 Å, β = 91.41°), crystal form 2 in space

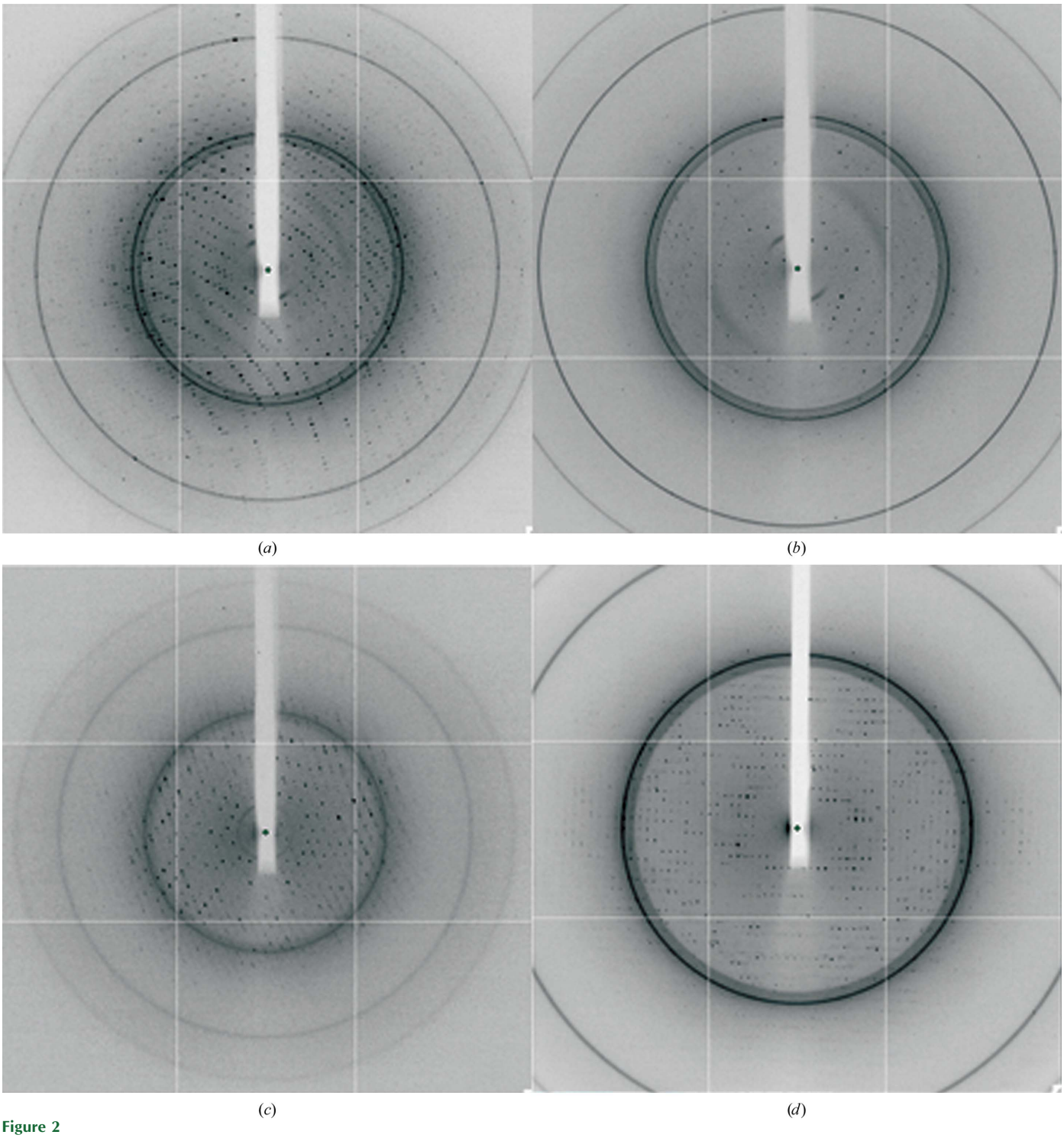


Figure 2
X-ray diffraction of the three crystal forms. (*a*) Crystal form 1, (*b*) crystal form 2, (*c*) crystal form 3 and (*d*) crystals of *yp*TpxC61S.

Table 1
Data-collection and reduction statistics.

Values in parentheses are for the highest resolution bin.

	Crystal form 1	Crystal form 2	Crystal form 3	*yp*TpxC61S
Space group	$P2_1$	$P6_4$	$P2_12_12_1$	$P6_4$
Unit-cell parameters (Å, °)	a = 64.86, b = 92.07, c = 85.60, β = 91.41	a = 65.01, c = 86.28	a = 56.37, b = 63.44, c = 91.39	a = 65.03, c = 85.60
Resolution (Å)	44.67–2.00 (2.11–2.00)	47.15–2.35 (2.48–2.35)	42.14–2.50 (2.59–2.50)	56.32–2.55 (2.69–2.55)
Observed reflections	182651	62950	160225	59770
Unique reflections	64433	8685	11830	6124
Multiplicity	2.8 (2.8)	7.2 (7.4)	13.54 (14.25)	9.8 (9.6)
Completeness (%)	94.9 (91.5)	99.9 (100)	99.9 (100)	90.5 (72.6)
Matthews coefficient (Å^3 Da^{-1})	2.09	2.58	2.01	2.56
Solvent content (%)	41	52	38	52
Monomers in asymmetric unit	6	1	2	1
$R_{\rm meas}$ (%)	15.4 (79.1)	7.7 (78.9)	13.6 (46.6)	10.6 (60.6)
$R_{\rm p.i.m.}$ (%)	8.8 (45.7)	3.0 (28.7)	—	3.5 (18.9)
$\langle I/\sigma(I)\rangle$	5.3 (1.7)	14.6 (2.2)	8.6 (4.0)	11.6 (3.1)
Wilson B factor (Å^2)	20.5	54.8	45.4	55.8

group $P6_2$ or $P6_4$ (unit-cell parameters a = 65.01, c = 86.28 Å) and crystal form 3 in space group $P2_12_12_1$ (unit-cell parameters a = 56.37, b = 63.44, c = 91.39 Å). The data from the best diffracting crystal, crystal form 1, were scaled to 2.00 Å. The data sets from crystal forms 2 and 3 were scaled to 2.35 and 2.5 Å, respectively. The data from the mutant crystal were scaled to 2.55 Å. All relevant diffraction statistics are presented in Table 1.

The molecular weight of *yp*Tpx is 17 592 Da, comprising 167 residues; as the crystallized construct also included the tag and the TEV cleavage site (MHHHHHHGKPIPNPLLGLDSTENLYFQGIDPFT), its full weight and length are 20 375 Da and 200 residues, respectively. Calculated Matthews coefficients (Matthews, 1968) for the data sets (Table 1) suggest that there are six monomers in the asymmetric unit for crystal form 1, one monomer for crystal form 2 and two monomers for crystal form 3, which are in agreement with the molecular-replacement results (see below). As Tpx from *E. coli* has been shown to be a dimer in both solution and crystal structure (Baker & Poole, 2003), it is likely that *yp*Tpx crystallizes with dimers or multiples of dimers in the various crystal forms. The substrate-binding site of the oxidized form of Tpx is made up by both parts of the dimer (Choi *et al.*, 2003).

We have crystallized *yp*Tpx, an atypical 2-Cys peroxiredoxin from *Y. pseudotuberculosis*, and its mutant *yp*TpxC61S in three different space groups; all of the crystal froms diffracted to better than 2.5 Å resolution. As the crystals were obtained with and without a reducing agent present, it is hoped that the protein will have crystallized in both the oxidized and the reduced state. Initial solutions have been found by positioning the known structure of *E. coli* Tpx [PDB codes 1qxh (Choi *et al.*, 2003) and 3hvv (Hall *et al.*, 2009)] by molecular replacement (*Phaser*; McCoy *et al.*, 2007). The solutions had final Z scores of 32.4, 27.3, 21.4 and 26.6 for crystal forms 1, 2, 3 and M, respectively. *Phaser* also identified $P6_4$ as the correct space group for crystal form 2. Work is now being carried out to refine the crystal structures.

This work was carried out with the support of the Diamond Light Source. The work was supported by a grant from the Biotechnology and Biological Sciences Research Council to AJR and MG (BB/G011389/1) and a Medical Research Scotland grant (reference 223 ORG G 0709) to AJR and DW. KSHB is supported by a Wellcome Trust studentship.

References

Baker, L. M. & Poole, L. B. (2003). *J. Biol. Chem.* **278**, 9203–9211.
Cha, M.-K., Kim, H.-K. & Kim, I.-H. (1995). *J. Biol. Chem.* **270**, 28635–28641.
Choi, J., Choi, S., Cha, M.-K., Kim, I.-H. & Shin, W. (2003). *J. Biol. Chem.* **278**, 49478–49486.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Evans, P. R. (1993). *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 114–122. Warrington: Daresbury Laboratory.
Hall, A., Sankaran, B., Poole, L. B. & Karplus, P. A. (2009). *J. Mol. Biol.* **393**, 867–881.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Pflugrath, J. W. (1999). *Acta Cryst.* D**55**, 1718–1725.
Robins-Browne, R. & Hartland, E. (2003). *International Handbook of Foodborne Pathogens*, edited by M. D. Miliotis & J. W. Bier, pp. 323–356. New York: Marcel Dekker.
Tao, K. (2008). *FEMS Microbiol. Lett.* **289**, 41–45.

Supplemental Material can be found at:
http://www.jbc.org/content/suppl/2011/07/01/M111.233858.DC1.html

⌘ *Author's Choice*

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 286, NO. 34, pp. 29922–29931, August 26, 2011
© 2011 by The American Society for Biochemistry and Molecular Biology, Inc. Printed in the U.S.A.

Identification of Bacterial Target Proteins for the Salicylidene Acylhydrazide Class of Virulence-blocking Compounds*[S]

Received for publication, February 22, 2011, and in revised form, June 23, 2011 Published, JBC Papers in Press, July 1, 2011, DOI 10.1074/jbc.M111.233858

Dai Wang[‡1], Caroline E. Zetterström[§1,2], Mads Gabrielsen[‡], Katherine S. H. Beckham[‡3], Jai J. Tree[¶3], Sarah E. Macdonald[‡], Olwyn Byron[‖], Tim J. Mitchell[‡], David L. Gally[¶3], Pawel Herzyk[**], Arvind Mahajan[¶], Hanna Uvell[‡‡4], Richard Burchmore[‡], Brian O. Smith[‖], Mikael Elofsson[§‡‡§§4,5], and Andrew J. Roe[‡6]**

From the [‡]Institute of Infection, Immunity, and Inflammation, [‖]School of Life Sciences, and [**]Institute of Molecular, Cell, and Systems Biology, College of Medical, Veterinary, and Life Sciences, University of Glasgow, Glasgow G12 8QQ, Scotland, United Kingdom, the [§]Department of Chemistry, Linnaeus v, [‡‡]Umeå Centre for Microbial Research and Laboratories for Molecular Infection Medicine Sweden, and [§§]Laboratories for Chemical Biology Umeå, Department of Chemistry, Linnaeus v, University of Umeå, SE-90187 Umeå, Sweden, and the [¶]Zoonotic and Animal Pathogens Research Laboratory, Immunity and Infection Division, The Roslin Institute and R (D)SVS, Chancellor's Building, University of Edinburgh, 49 Little France Crescent, Edinburgh EH16 4SB, Scotland, United Kingdom

A class of anti-virulence compounds, the salicylidene acylhydrazides, has been widely reported to block the function of the type three secretion system of several Gram-negative pathogens by a previously unknown mechanism. In this work we provide the first identification of bacterial proteins that are targeted by this group of compounds. We provide evidence that their mode of action is likely to result from a synergistic effect arising from a perturbation of the function of several conserved proteins. We also examine the contribution of selected target proteins to the pathogenicity of *Yersinia pseudotuberculosis* and to expression of virulence genes in *Escherichia coli* O157.

As the prevalence of antibiotic-resistant strains increases, targeting virulence determinants of pathogenic bacteria has become an attractive alternative to the use of traditional bactericidal antibiotics (1–5). A key feature of this strategy is that the virulence-blocking compounds spare the endogenous microflora and thereby exert less selective pressure, which in turn should reduce development of resistance. One potential target is the bacterial type three secretion system (T3SS),[7] a conserved protein injection organelle that is central to the virulence of many human, animal, and plant pathogens including *Chlamydia* sp., enteropathogenic and enterohemorrhagic *Escherichia coli*, *Pseudomonas aeruginosa*, *Salmonella* sp., *Shigella* sp., and *Yersinia sp.* (6, 7). With the T3SS, the pathogen translocates effector proteins into the cytosol of the host cell and thereby creates a niche that allows bacterial growth. A class of virulence-blocking compounds, the salicylidene acylhydrazides, was originally identified as putative T3SS inhibitors in *Yersinia pseudotuberculosis* (8). In a number of publications the compounds have been shown to be broadly effective in negatively affecting the function of the T3SS in a number of pathogenic bacteria including *Chlamydia* sp. (9–12), *Salmonella typhimurium* (13, 14), *Y. pseudotuberculosis* (15), *Shigella* sp. (16), and *E. coli* O157 (17). Two reports describe activity of the compounds *in vivo* (13, 18) and thus indicate that T3SS inhibitors have the potential to be developed into novel anti-bacterial agents (19, 20). In addition, small molecule inhibitors can be used as chemical probes to study the role of T3SS in bacterial pathogenesis (12, 21, 22). To date, the mechanism of inhibition for the salicylidene acylhydrazides has been unclear with several mechanisms being postulated, including direct effects on the T3SS machinery or on regulatory proteins that affect T3SS expression (15–17). The activity of salicylidene acylhydrazides on *Chlamydia trachomatis* can be reversed by the addition of iron, suggesting a possible link to iron availability in the cell (11).

Although target-based screening can identify compounds that perturb the function of the selected protein, the activity is often lost when the compounds are tested on bacterial cells due to a lack of bacterial cell permeability (23). Phenotypic screens circumvent this challenge and directly provide compounds that are active on the cellular level. The salicylidene acylhydrazides were identified using this strategy, and this is likely the underlying reason for the broad activity spectrum observed for these T3SS inhibitors. The drawback is, however, that the mode of action at the molecular level has to be studied at a later stage. Identification of the cellular targets for the salicylidene acylhydrazides constitutes a crucial step in understanding their true

* This work was generously supported by Medical Research Scotland Grant 223 ORG (to D. W., A. J. R. and R. B.) and Biotechnology and Biological Sciences Research Council (Swindon, United Kingdom) Grant BB/G011389/1 (to A. J. R. and M. G.).

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Figs. S1 and S2, Tables S1–S5, and Protocol S1.

⌘ *Author's Choice*—Final version full access.

The Illumina data have been deposited on the EMBL database (ERP000335), and the MIAME compliant data are deposited on the Gene Expression Omnibus (GEO) at NBCI under the GEO accession number GSE23001.

[1] Both authors contributed equally to this work.

[2] Supported by the Swedish Research Council.

[3] Supported by the Wellcome Trust.

[4] Supported by the Knut and Alice Wallenberg Foundation and Vinnova.

[5] To whom correspondence may be addressed. Tel.: 46-90-786-9328; Fax: 46-90-86-9995; E-mail: Mikael.Elofsson@chem.umu.se.

[6] To whom correspondence may be addressed. Tel.: 44-141-3302980; Fax: 44-141-3304600; E-mail: andrew.roe@glasgow.ac.uk.

[7] The abbreviations used are: T3SS, type three secretion system; AUC, analytical ultracentrifugation; HSQC, heteronuclear single quantum coherence; ecTpx, *E. coli* O157 Tpx; QSAR, quantitative structure-activity relationship; ypTpx, *Y. pseudotuberculosis* Tpx; MEM, minimum Eagle's medium; BisTris, 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol; SNP, single-nucleotide polymorphism.

mode of action. Moreover, such data are essential for the structure-led design and improvement of any potential therapeutic compound. Affinity chromatography, expression-cloning technologies, and microarrays are among the many strategies that can be employed for target deconvolution (24). To help understand the molecular mechanism by which salicylidene acylhydrazides affect the T3SS, we aimed to identify the target proteins bound by this class of novel antibacterial compounds using an affinity reagent strategy (25). In this study we describe synthesis of a T3SS inhibitor affinity reagent and the isolation of putative target proteins from *E. coli* O157. By a series of *in vitro* experiments, we show that the compounds directly and selectively interact with the target proteins WrbA, Tpx, and FolX. The genes encoding the target proteins were deleted individually in both *E. coli* O157 and *Y. pseudotuberculosis*, and we show that the proteins are involved in regulation of T3SS gene expression. Our work provides the first identification of the cellular targets for this group of compounds and convincing evidence that their mode of action is likely to result from a synergistic effect arising from a perturbation of the function of several conserved proteins.

EXPERIMENTAL PROCEDURES

Strains, Plasmids, and Construction of Mutants—Bacterial strains and plasmids used in this study are listed in supplemental Table S1. Mutations in *tpx* were made using QuikChange II (Stratagene) and were confirmed by sequencing. Oligonucleotides (Invitrogen) are listed in supplemental Table S2. *E. coli* O157 strains were cultured in LB, DMEM, or MEM-HEPES media (Sigma) supplemented with standard concentrations of antibiotics (26).

Affinity Chromatography Using Affi-Gel-labeled Salicylidene Acylhydrazides—The stepwise synthesis of ME0055-Aff is described in supplemental Protocol S1. *E. coli* O157 strain TUV93–0 was cultured in MEM-HEPES media at 37 °C to an $A_{600} = 0.8$. Cells were harvested and lysed using a French pressure cell. No subsequent separation steps were used to ensure that cell wall, cell membrane, and cytosolic proteins were represented in the lysate. The bacterial lysate was mixed with PBS-balanced ME0055-Aff beads overnight at 4 °C. The beads were separated by centrifugation at 500 × *g* for 5 min. An aliquot of the supernatant was saved for analysis, and the rest of the supernatant was discarded. The beads were washed 10 times using 10 volumes of PBS and separation by centrifugation at 500 × *g* for 5 min. The beads were mixed gently in the same volume of 20 μM ME0055 and incubated at room temperature for 30 min. Supernatant was collected by centrifugation at 500 × *g* for 5 min. The elution and centrifugation step was repeated with 200 μM ME0055 and 1% acetic acid, respectively. Equal volumes of washes and eluates were loaded onto the protein gels. Samples were visualized using Colloidal Blue Stain (Invitrogen), and bands were excised for subsequent in-gel digestion and analysis (27). Proteins analyzed by peptide mass fingerprinting were given a MOWSE score (28) to indicate the probability of the identification being correct. We set a threshold MOWSE value of 100 or greater, ensuring a significance of $p < 10^{-8}$.

Screening for Protein Secretion and Western Blot Analyses—Eighteen bacterial strains (supplemental Table S1) were cultured overnight in LB and diluted to an $A_{600} \leq 0.05$ in prewarmed MEM-HEPES medium. Cultures were grown at 37 °C to an $A_{600} = 0.8$, and then secreted proteins were extracted by TCA precipitation as described previously (29). Proteins were analyzed by SDS-PAGE, and Western blotting for Tir and EspD was carried out as described previously (29). Strain 430 displayed no reduction in secretion upon the addition of 20 μM ME0052, ME0053, ME0054, or ME0055 to the culture when assessed by Western blotting for Tir or EspD.

Re-sequencing of ZAP 430—Genomic DNA for the ZAP 430 strain was extracted using a QIAamp DNA extraction kit. The sequencing was performed in the Sir Henry Wellcome Functional Genomics Facility, University of Glasgow, using the Illumina Genome Analyzer II platform. The purified genomic DNA was randomly fragmented with a Bioruptor Sonicator (Diagenode Inc., NJ) followed by the ligation of adapters, and genomic library generation was performed according to the standard Illumina protocol. The genomic library at a concentration of 3 pM was then loaded onto a single lane of the Illumina flow cell, and the DNA clusters were generated using the Illumina Single Read Cluster Generation kit Version 4 (Illumina Inc.). Subsequently, 70 cycles of single end sequencing were performed. The raw data containing read sequences and associated qualities were generated with Genome Analyzer Pipeline v.1.5 (Illumina Inc.). Of 15,712,834 70-bp-long reads generated, 14,347,057 were successfully mapped to the *E. coli* O157:H7 strain EC4115 complete genome (GenBank™ accession NC_011353) using Maq software with the default parameter set, which resulted in 179× genome coverage. Subsequently, the SNPs were identified with Maq and filtered with the command "maq.pl SNPfilter -q 40 -w 5 -N 2 -d 3 -D 256 -n 20 -Q 40," where parameters were chosen as suggested for the bacterial genomes (30). The latter identified 1348 SNPs (supplemental Table S3).

Far Western Blotting—The stepwise synthesis of ME0052-Bio and ME0055-Bio are described in supplemental Protocol S1. Target proteins were overexpressed in *E. coli* BL21 (λDE3). Bacterial pellets were lysed using 8 M urea, and the supernatant was collected by centrifugation at 14,000 × *g* for 1 min. The Far Western analysis was performed as described previously (31) using Novex NuPAGE 4–12% BisTris SDS-PAGE gels (Invitrogen), 20 μM ME0052-Bio as the probe, and HRP-conjugated streptavidin (Invitrogen) for the detection of biotin. Protein loading was checked by using an anti-His antibody that showed equivalent levels of overexpressed target protein in each lane (data not shown).

Analyses of Transcript Levels—Transcriptional profiling of *E. coli* O157 strains was carried out essentially as described previously (32). Overnight cultures of *E. coli* TUV-930 or defined mutants were grown in MEM-HEPES (Sigma) supplemented with 250 nM $Fe(NO_3)_2$ and glucose to a final concentration of 0.2% and then diluted to an $A_{600} = 0.1$ in the same media. At an $A_{600} = 0.8$, 15-ml cultures were stabilized in an equal volume of RNAprotect (Qiagen), and RNA was extracted using a Qiagen RNeasy Mini kit. DNA contamination was removed by DNase I treatment (Ambion). Total RNA was assessed for quality and quantified using an Agilent 2100 Bioanalyzer. Synthesis of cDNA and labeling of total RNA was performed using an Amer-

sham Biosciences CyScribe post-labeling kit as per the manufacturer's instructions. The cDNA was hybridized to 70-mer-spotted oligonucleotide arrays containing ORFs from *E. coli* K-12, *E. coli* Sakai VT2, and *E. coli* EDL933 (University of Birmingham) using a Genomic Solutions GeneTac hybridization machine. Hybridized slides were scanned using a Genepix 4000A scanner and GenePix 7.0 software (Axon Instruments, Union City, CA). Data were analyzed using Genespring GX 7.3.1 (Agilent).

Quantitative PCR—Triplicate wild-type and mutant cultures were grown up in MEM-HEPES. RNA extractions were carried out using a Qiagen RNeasy Mini kit, and cDNA synthesis was carried out using a Qiagen QuantiTect™ Reverse Transcription kit. Duplicate quantitative PCRs were carried out using a Qiagen Quantifast™ SYBR® Green PCR kit and Stratagene MX3000, and oligonucleotide primers are listed in supplemental Table S2. All the experiments were performed according to the manufacturer's instructions.

Visualization of Flagella and EspA Filaments—Flagella and EspA filaments were visualized using immunofluorescence microscopy (17). Images were processed and quantified using Volocity suite software (PerkinElmer Life Sciences). Bacterial motility was assayed using soft agar plates as described previously (33).

Purification and Biophysical Studies of Tpx from Y. pseudotuberculosis (ypTpx)—ypTpx and ypTpxC61S were cloned, expressed, and purified as described previously (34). For NMR spectroscopy, ypTpx was expressed in M9 minimal media, with the nitrogen source substituted with $^{15}NH_4Cl$. Experiments to identify chemically-shifted residues were performed using triple-labeled Tpx (^{13}C,^{2}H,^{15}N), produced by culture in algal hydrolysate-based media (Cambridge Isotopes Ltd). For chemical shift perturbation studies, ypTpx NMR samples were prepared to contain 100 μM protein in 13 mM KP_i, pH 7.4, 33 mM NaCl, and 5% D_2O, and ^{15}N FAST-HSQC (35) spectra were recorded at 298 K and 14.1 T using a Bruker AVANCE 600 spectrometer equipped with a 5-mm TCI cryoprobe. Compound was added from a concentrated stock solution in deuterated DMSO (dDMSO), and reference spectra recorded with an equal concentration (0.33% v/v) of dDMSO added. Backbone resonance assignment was achieved using triple resonance, and TROSY (transverse relaxation optimized spectroscopy) triple resonance experiments were recorded from 150 to 500 μM samples at 308 K. Data were processed using AZARA and analyzed using CCPNMR analysis (36).

Analytical Ultracentrifugation (AUC) of Tpx and ME0052—Before AUC analysis, ypTpx and ypTpx C61S were dialyzed against 20 mM Tris, pH 7.4, 50 mM NaCl. Sedimentation velocity experiments were performed using an Optima XL-I analytical ultracentrifuge (Beckman Coulter, Palo Alto, CA) and an An-50 Ti rotor. Data were collected at 49,000 rpm at 4 °C for ypTpx under oxidizing (with 10 mM H_2O_2) and reducing (with 5 mM DTT) conditions and for the C61S mutant. 360 μl of sample (100 μM protein solution with ME0052 at concentrations ranging from 0 to 250 μM) and reference solvent (20 mM Tris, pH 7.4, 50 mM NaCl, 1% v/v DMSO) were loaded into 3-mm path length charcoal-filled epon double-sector centerpieces. Rayleigh interference and absorbance (at 280 and 395 nm) data were collected every 15 min for 25 h (1 replicate of absorbance data in continuous mode with a radial step size of 0.005 cm). A wavelength of 395 nm was used to monitor the sedimentation of ME0052 only, as the protein does not contribute to the signal at this wavelength.

A two-stage analysis of sedimentation velocity data was used to determine the amount of ME0052 bound to oxidized and reduced ypTpx and the C61S mutant. First, the number of species in the system was evaluated using continuous distribution ($c(s)$) analysis, performed with SEDFIT software (37). The sedimentation coefficients ($s^0_{20,w}$) of oxidized and reduced ypTpx and the C61S mutant are 3.04, 2.62, and 2.80 *s*, respectively (data not shown); thus, a peak with this sedimentation coefficient in the $c(s)$ distribution derived from sedimentation velocity data acquired at 395 nm could be attributed to ypTpx bound to ME0052. At ME0052 = 0 μM no peaks were resolved from the data; at higher concentrations two peaks were evident: one at very low *s* (corresponding to unbound ME0052) and the second close to 2.7 *s* (ME0052 bound to protein). The concentration of free and bound compound was determined by fitting the data with a (two species) non-interacting discrete species model in SEDFIT. An extinction coefficient of 11,100 M^{-1} cm^2, determined from the linear dependence of absorbance at 395 nm with ME0052 concentration, was used to calculate the concentration of compound returned by the non-interacting discrete species model. The resultant reduced data were then fitted with a single-site-binding nonlinear regression model using the formula $Y = B_{max} \times X/(K_d + X)$, where Y is the concentration of ME0052 bound to protein (determined by non-interacting discrete species), X is the total concentration of ME0052 (*i.e.* bound and unbound), B_{max} is the concentration of compound required to saturate the protein (*i.e.* at maximal binding), and K_d is the equilibrium dissociation constant or concentration of compound required to reach half-maximal binding. The stoichiometry of interaction between ME0052 and ypTpx was determined by dividing B_{max} by the total protein concentration.

RESULTS

Affinity Chromatography Identifies 16 Putative Salicylidene Acylhydrazide Target Proteins—The two salicylidene acylhydrazides ME0052 and ME0055 were selected for chemical derivatization. ME0052 has been extensively studied and proved to display low toxicity against mammalian cells and activity against T3S in *Y. pseudotuberculosis*, *C. trachomatis*, *Salmonella enterica*, *Shigella flexneri*, and *E. coli* O157:H7 (10, 11, 14–17), and ME0055 was selected based on its low toxicity and high activity against T3S in *E. coli* O157 (17). The Affi-Gel- and biotin-labeled salicylidene acylhydrazide derivatives ME0055-Aff and ME0052-Bio (Fig. 1*A*) were synthesized by combined solid phase and solution phase chemistry. The position for attachment of the linker was based on previous structure-activity data (15, 38, 39). ME0055-Aff was used in affinity chromatography experiments to enrich for inhibitor binding proteins from *E. coli* O157 lysates prepared using a French pressure cell. After incubation, washing, and elution using stepwise increasing concentrations of unlabeled compound, proteins were resolved by SDS-PAGE and visualized with Colloidal Blue Stain (Fig. 1*B*). Bands were selected corresponding to the

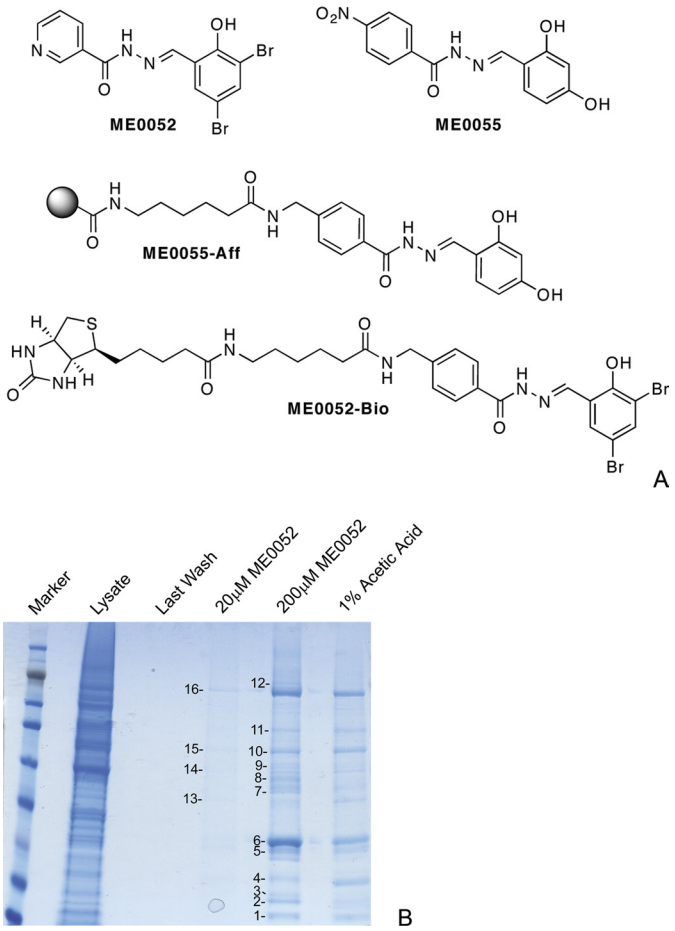


FIGURE 1. *A*, shown are structures of compounds used and visualization of the proteins from the pulldown assay. Structures of the inhibitory compounds, ME0052 and ME0055, and derivatives, ME0055-Aff and ME0052-Bio, used for the identification of target proteins are shown. Supplemental Protocol S1 details the syntheses. ME0055-Aff was used in affinity chromatography experiments to enrich for putative target proteins; ME0055-Bio was used for Far Western experiments. *B*, visualization of the protein fractions from the affinity chromatography procedure is shown. Lysates from *E. coli* O157 were incubated with ME0055-Aff and washed, and then proteins were eluted using 20 and 200 μM unlabeled ME0055. The beads were stripped using 1% acetic acid. Equal volumes of samples from each fraction were separated by SDS-PAGE and then visualized using Colloidal Blue Stain. Proteins were excised, subjected to in-gel trypsin digestion, and analyzed by LC-electrospray ionization tandem MS. Each protein identified from the gel is numbered sequentially and is listed in supplemental Table S4.

lanes derived from the fractions that had been eluted using unlabeled compound but that were less abundant in the lane associated with the acetic acid "stripping" of the beads. These were postulated to be true target proteins and not simply bound to the Affi-Gel bead itself. Bands from the gel were excised to permit in-gel trypsin digestion and analysis by liquid chromatography electrospray ionization tandem mass spectrometry (27). Using this approach, we identified 16 proteins (supplemental Table S4) that were bound by ME0055-Aff.

Identification and Resequencing of an E. coli O157 Strain Insensitive to Salicylidene Acylhydrazides—As an independent screen, a panel of 18 clinical *E. coli* O157 isolates selected to represent a diverse range of phage types were tested for sensitivity against four salicylidene acylhydrazides (ME0052, ME0053, ME0054, and ME0055) that previously proved effective against *E. coli* O157 (17). Western blot analysis of Tir secretion in the presence of 20 μM concentrations of each of the compounds revealed a strain, ZAP 430, that showed no inhibition of Tir secretion upon the addition of any of the four compounds. Illumina re-sequencing of the genome (EMBL, ERP000335) and comparison against EC4115, a strain sensitive to the salicylidene acylhydrazides, demonstrated that ZAP 430 carried numerous mutations (supplemental Table S3). These mutations include the gene encoding WrbA, one of the target proteins identified from the affinity chromatography approach. Previous work has suggested that the salicylidene acylhydrazides may function by directly binding and perturbing the proteins associated with the T3SS basal apparatus (16). ZAP 430 had no mutations in the genes encoding the structural proteins of the T3SS; this insensitive strain contained T3SS proteins identical to both EDL933 and EC4115, reflecting the high level of conservation of these proteins.

ME0052-Bio Binds to Tpx, WrbA, and FolX from E. coli O157—To facilitate studies of the putative binding proteins, we attempted to clone the corresponding genes. From the 16 amplified products, 12 genes were successfully cloned into a pET151 vector and verified by sequencing. Each of the 12 plasmids was tested for protein overexpression, with seven clones producing a protein of the expected size. These seven proteins were (or were encoded by): Tpx, FolX, z2714, z3974, WrbA, SurA, and StcE. Interactions with salicylidene acylhydrazides were tested in Far Western experiments using 20 μM concentrations of the biotin-labeled compound ME0052-Bio as a probe (Fig. 2*A*). From the seven overexpressed proteins, three gave clear signals after Far Western analysis: Tpx, a thiol peroxidase (40), WrbA a NAD(P)H quinine oxidoreductase (41), and FolX, a dihydroneopterin-tri-P-epimerase (42). No signal or a very weak signal was seen corresponding to the other four proteins or for *E. coli* lysates with no overexpressed target protein (Fig. 2*A*). We, therefore, decided to focus on Tpx, WrbA, and FolX for further studies.

ME0052-Bio and ME0055-Bio Bind to WrbA, Tpx, and FolX from Multiple Pathogens—Previous work has demonstrated that salicylidene acylhydrazide compounds inhibit expression and function of the T3SS in numerous pathogens (19). We aimed to test if WrbA, Tpx, and FolX were conserved targets resulting in a common mechanism of inhibition. Cloning and overexpression of the genes encoding WrbA and Tpx were carried out from five different pathogens: *E. coli* O157, *S. typhimurium*, *S. flexneri*, *P. aeruginosa*, and *Y. pseudotuberculosis*. The gene encoding FolX, which is less conserved, was cloned from three pathogens: *E. coli* O157, *P. aeruginosa*, and *S. flexneri*. FolX is not present in *Y. pseudotuberculosis*, but this species does carry a structurally and functionally similar protein, FolB, as do *P. aeruginosa* and *S. typhimurium*. To examine the specificity of the inhibitor-target interaction further, we also cloned and overexpressed FolB from these three species. Using ME0052-Bio as a probe for Far Western experiments, we observed binding with overexpressed WrbA, Tpx, and FolX from all the species tested (Fig. 2, *B–D*). Protein loading was shown to be equivalent for each lane by probing the samples with an anti-His antibody (data not shown). After separation by SDS-PAGE, Tpx was detected as a monomer of 19 kDa as well as a dimer of 38 kDa, which was verified by Tandem MS iden-

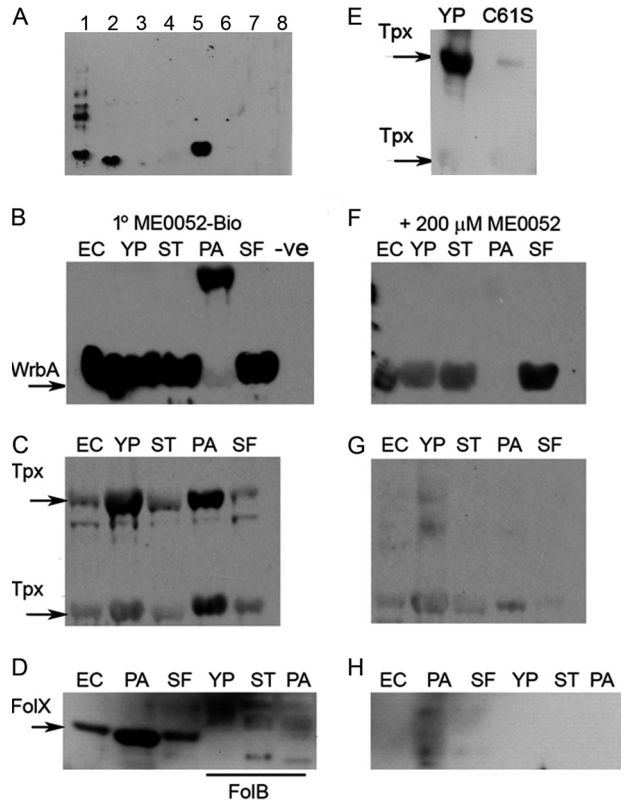


FIGURE 2. **Far Western probing to examine binding of ME0052-Bio to putative target proteins.** *A*, putative target proteins were cloned into pET151 and expressed in *E. coli* BL21 (λDE3) cells before lysis, separation, and transfer to a nitrocellulose membrane. ME0052-Bio was used as a probe, and interactions were detected using HRP-conjugated streptavidin. Proteins probed were (or encoded by) Tpx (*1*), FolX (*2*), z2714 (*3*), z3974 (*4*), WrbA (*5*), SurA (*6*), and StcE (*7*). *Lane 8* contains bacterial lysate with no overexpressed protein as a negative control. The same protocol was used to subsequently probe proteins from *E. coli* O157 (*EC*), *Y. pseudotuberculosis* (*YP*), *S. typhimurium* (*ST*), *P. aeruginosa* (*PA*), and *S. flexneri* (*SF*) or empty vector control (−*ve*). Panels show overexpression of target proteins: WrbA (*B*), Tpx (*C*), FolX and FolB (*D*). Mutation of Tpx Cys-61 (C61S) affected binding of ME0052-Bio (*E*). The addition of 200 μM unlabeled ME0052 (*F–H*) strongly affected the binding of the biotinylated probe.

tification of the excised bands. The presence of both monomer and dimer after SDS-PAGE indicated that Tpx had a stable structure and that heat and detergent were unable to completely denature the protein. The addition of ME0052-Bio resulted in a signal for both dimer and monomer of Tpx (Fig. 2*C*). Experiments were repeated using ME0055-Bio and gave the same results, albeit with a higher level of background signal. No interaction between ME0052-Bio and FolB was detected (Fig. 2*D*), indicating that the compound does not bind this protein when an equivalent amount of protein is used as bait. This is despite the fact that FolB and FolX from *E. coli* display 68% sequence similarity (as determined using ClustalW), and the structures superimpose with a root-mean-square-deviation of 1.7 Å (PDB codes 2O90 and 1B9L, respectively). The addition of 200 μM unlabeled ME0052 (Fig. 2, *F–H*), in competition with ME0052-Bio, resulted in a marked reduction in ME0052-Bio binding to WrbA, Tpx, and FolX irrespective of species.

To investigate the specificity of the interaction between target protein and inhibitor, we focused further on Tpx, a thiol peroxidase previously shown to be part of the oxidative stress defense system that uses reducing equivalents from thioredoxin (Trx1) and thioredoxin reductase (TrxR) to reduce alkyl hydroperoxides (40). Tpx is attractive to study as it is easy to purify, and conditions for protein crystallization have been determined (34, 43). The catalytically important residues of Tpx have been previously identified. Tpx is inactivated by mutation of cysteine 61 that is essential for intramolecular disulfide bond formation (40). The mutant protein species is present only in the structurally reduced form, as it cannot undergo the disulfide bond formation critical for the catalytic cycle of the protein. C61S Tpx formed a less stable dimer, and after SDS-PAGE, the majority of the protein was seen as a monomer (data not shown) rather than as a dimer for wild-type Tpx. The C61S mutation reduced ME0052-Bio binding (Fig. 2*E*).

NMR Chemical Shift Perturbation Shows That ME0052 Binds to Tpx from ypTpx—We validated the interaction between ME0052 and Tpx by NMR chemical shift perturbation. ^{15}N heteronuclear Single Quantum Coherence (HSQC) spectra of 100 μM ypTpx recorded in the absence and presence of excess ME0052 revealed chemical shift changes specific to a small number of cross-peaks originating from particular backbone amides (Fig. 3). The shift changes increased gradually with increasing ligand concentration, indicating that the free and ligand-bound protein were in fast exchange and were saturated by 200 μM ME0052. No shifts were seen with the addition of DMSO alone. 81% of the backbone amide resonances were assigned, and the chemical shift changes mapped to the structure. These data indicate that the compound binds to a specific site on ypTpx, inducing only minor conformational changes.

Identification of Residues Involved in the Binding of ME0052 and Modeling of the Site—Based on the residues shown to be perturbed by NMR, ME0052 was docked into the published structure of *E. coli* O157 Tpx (ecTpx) (PDB code 3HVV (43)) using the program Autodock Vina in PyMOL (44, 45). There is a high sequence identity between Tpx from *E. coli* and *Y. pseudotuberculosis*, comprising 80% overall identity with 100% similarity in the residues affected by binding. As PDB code 3HVV is the structure of a C61S mutant, residue 61 was replaced with a cysteine to mimic the reduced structure. Residues that were shown to have the greatest NMR chemical shifts are colored purple for reference (Fig. 4). The binding site of ME0052 with Tpx forms near the dimer interface and comprises both subunits albeit with an unequal contribution, as can be seen on Fig. 4*A*. The residue with the largest shift is Val-60, adjacent to the catalytically active residue of Tpx.

Determination of Binding Affinity and Stoichiometry—The salicylidene acylhydrazide compounds have a defined absorbance peak at 395 nm. This enabled binding of the compounds to ypTpx in solution to be examined using AUC. The concentration of ME0052 bound to ypTpx was determined for a range of total compound concentrations. These data were used to calculate the dissociation constant (K_d) and the stoichiometry of binding. Using this approach, the K_d for binding of ME0052 to oxidized ypTpx was determined to be 51 and 71 μM to reduced ypTpx (Fig. 5). The K_d for the C61S mutant revealed an approximate 2-fold reduction in binding (93 μM) (Fig. 5). The correlation coefficients (r = 0.8037, 0.9133, and 0.8440 for C61S, reduced, and oxidized ypTpx) of the nonlinear best fits confirmed the strength of dependence between the amount of

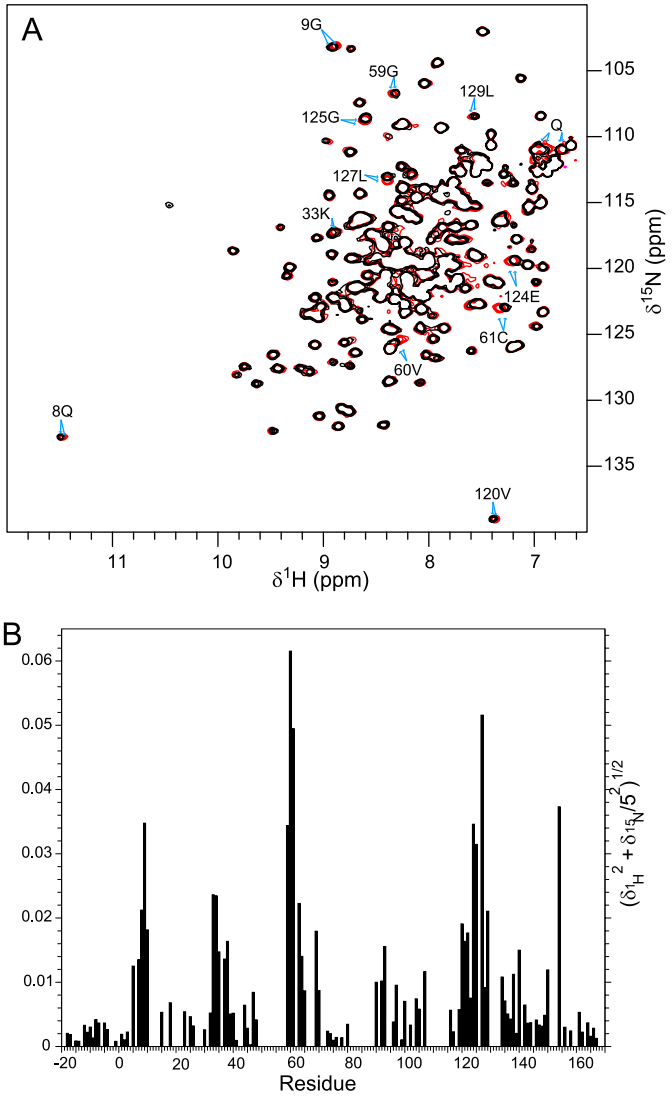


FIGURE 3. **NMR chemical shift perturbation of ypTpx by ME0052.** *A*, ^{15}N HSQC spectra of 100 μM ypTpx in the absence (*black*) and presence (*red*) of 200 μM ME0052. A selection of the most perturbed cross peaks is indicated with their assignments, and a perturbed glutamine side chain is indicated (*Q*). *B*, a histogram of the weighted amide chemical shift perturbation by residue reveals clusters of residues affected by ligand binding.

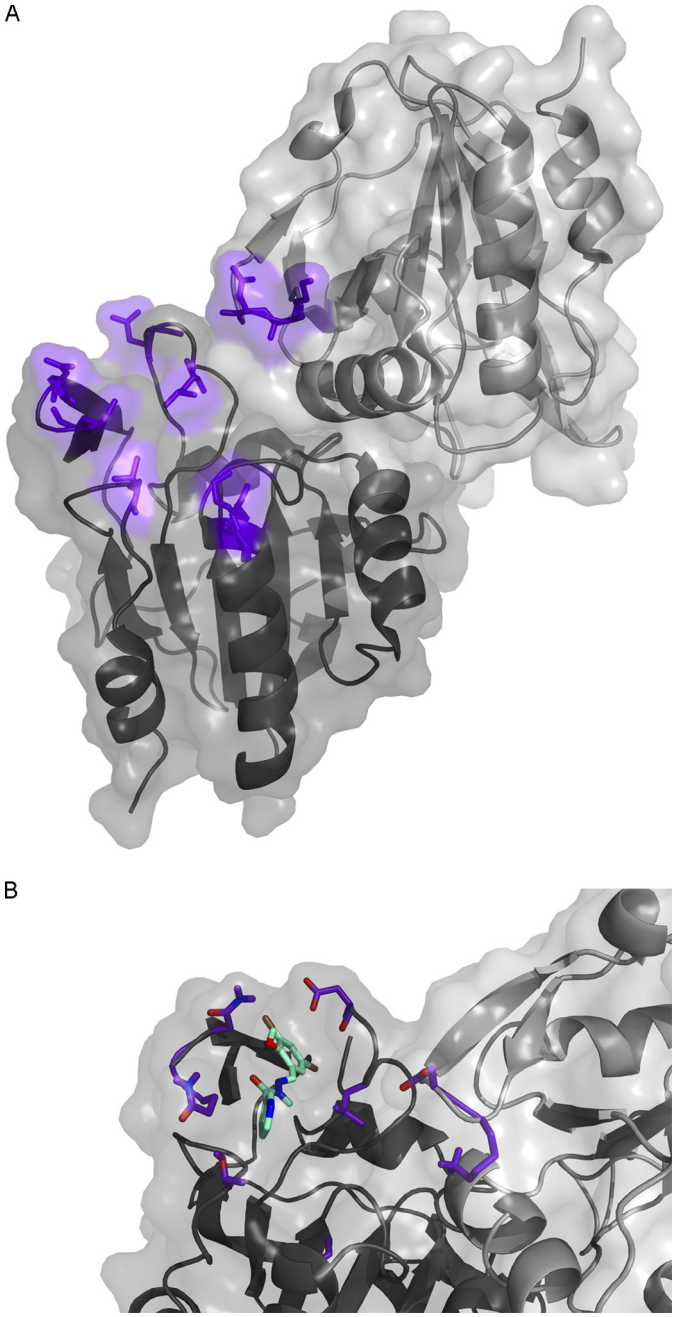


FIGURE 4. **Mapping of residues involved in drug binding and ligand docking.** *A*, shown is a schematic representation of the structure of ecTpx (PDB code 3HVV (43)) colored *purple* by the shifts observed when ME0052 is added to ypTpx. Subunit A is in *dark gray*, and subunit B is in *lighter gray*. *B*, shown is a surface representation of the ecTpx dimer, with the docked compound ME0052 (*cyan*) based on the residues with the highest shift (*purple sticks*) using Autodock Vina. The figures were created in PyMOL (56).

bound drug and free drug. The stoichiometry was determined to be one molecule of ME0052 compound bound per ypTpx dimer.

WrbA, Tpx, and FolX Contribute to Bacterial Gene Regulation—To assess any role of the three target proteins for the expression and function of the T3SS, the corresponding gene encoding each protein was deleted individually in both *E. coli* O157 and *Y. pseudotuberculosis*. In addition, mutagenesis was used to modify ecTpx and introduce a serine residue in place of the cysteine (Cys-61) that is essential for Tpx function. This was achieved by cloning the gene encoding ecTpx into a plasmid (pDW141), site-directed mutagenesis, and then allelic exchange to replace the wild-type copy of the *tpx* gene with the mutant allele. The final strain, *E. coli* Tpx_C61S, was used in subsequent assays to compare the phenotype of a deletion of Tpx with that of a non-functional mutant. The growth rate of all the *E. coli* O157 mutants tested was not affected during culture in MEM-HEPES medium used for the transcriptomic profiling. Using microarrays, the gene expression profile of each *E. coli* O157 mutant was compared with that of the wild type. Transcriptional changes for each mutant are listed in supplemental Table S5.

The most striking finding was that some 27 genes were significantly ($p < 0.05$) affected in all four mutants. These genes could be grouped into six Gene Ontology (GO) categories: cell motility, locomotory behavior, localization of the cell, behavior,

locomotion, and flagellar motility. Nine genes required for flagellar synthesis were significantly (>2-fold change, $p < 0.05$) repressed in all four mutants. The genes encoding the T3SS were also affected, but in contrast to the flagellar genes, these were up-regulated, with the greatest changes seen for genes encoding the basal apparatus of the T3SS, such as *escJ*, *escC*, and *escD* (>2-fold change, $p < 0.05$). Previous work has demonstrated that the expression of flagella and the T3SS of *E. coli* O157 are cross-regulated to allow either motility or attachment as appropriate (44). The strong repression of flagella and stimulation of T3SS expression demonstrates that the three target proteins we have identified are required for normal expression and appropriate regulation of these key virulence factors.

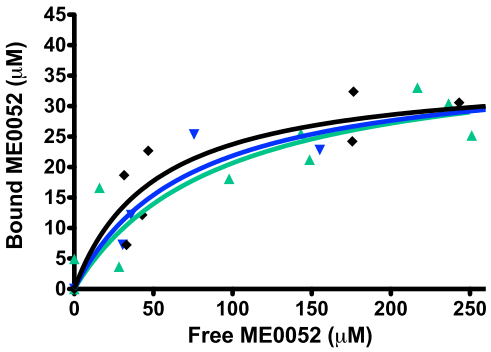


FIGURE 5. **AUC analysis of ME0052 binding to ypTpx.** A line graph shows the binding of ME0052 to purified ypTpx during AUC experiments. The lines represent different redox states for Tpx: oxidized (*black diamonds*), reduced (*blue inverted triangles*) and forced reduced mutant, C61S (*green triangles*). The amount of bound ME0052 was used to calculate the K_d values described in the text.

The transcriptomic data are of particular interest when compared with our previous work that determined the gene expression profile of *E. coli* O157 cultured in the presence or absence of salicylidene acylhydrazide compounds (17). In this earlier study we demonstrated that the addition of salicylidene acylhydrazides resulted in strong repression of the T3SS and up-regulation of flagella; that is, the reciprocal pattern of gene expression compared with the four mutants described here. Some 39 genes repressed by the addition of the inhibitors were up-regulated >1.5-fold in all mutants (15 of which were associated with the T3SS), and conversely, 15 genes stimulated by addition of the inhibitors were repressed in the mutants (13 of which were associated with motility). This pattern of gene expression can be most clearly seen by plotting on a single graph the -fold change of the genes that were significantly ($p < 0.05$) affected in both studies (Fig. 6, *A* and *B*). The C61S mutant displayed a transcriptional profile similar to that of the mutant with the defined deletion. The transcriptomic data were validated by quantitative PCR to examine transcript levels for selected genes; transcript levels for *fliC* confirmed the differences observed in flagellar expression with *escJ* tested to confirm the up-regulation of the T3SS genes. For example, the *E. coli* Tpx_C61S strain showed a 154-fold reduction of *fliC* expres-

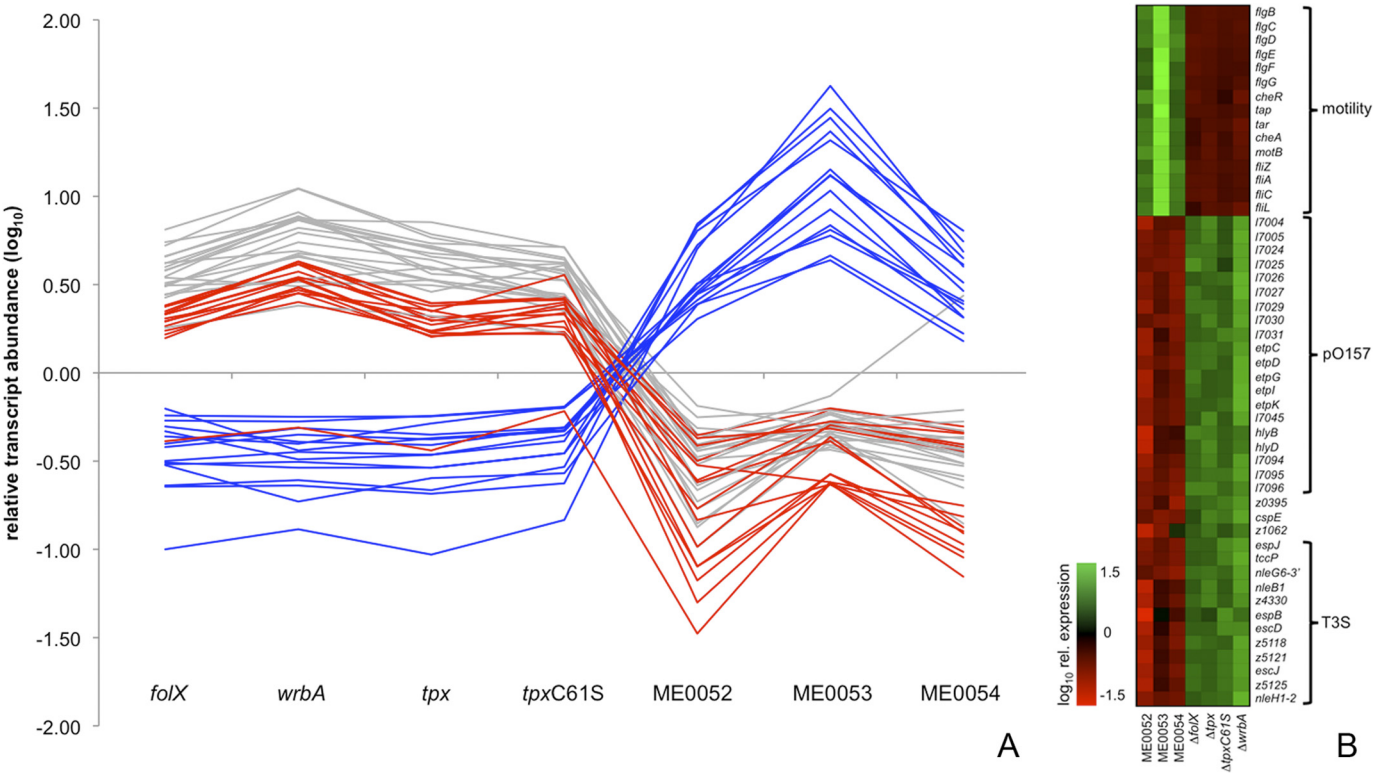


FIGURE 6. **Comparison of gene expression between bacteria cultured with salicylidene acylhydrazides and defined mutants.** Microarray data were filtered for transcripts that showed significant changes in gene expression (>1.5-fold, $p < 0.05$) when treated with salicylidene acylhydrazides (ME0052, ME0053, ME0054) and that also showed >5-fold changes in transcript abundance when compared with the defined mutants. *A*, shown is a graphic representation of gene expression. *Red lines* indicate genes associated with the T3SS, *blue lines* indicate transcripts associated with flagella and motility, and *gray lines* represent other genes. *B*, shown is a heatmap of gene expression. Gene names are indicated to the *right* of the heatmap, and experimental conditions (inhibitor or relevant gene deletion or mutation) are indicated *below*. Coloring of boxes indicates relative expression (*green*, increased; *red*, repressed) as indicated on the *left*. Genes have been grouped according to their function or genetic location in three major classes: *motility*, *pO157*, and *T3SS*.

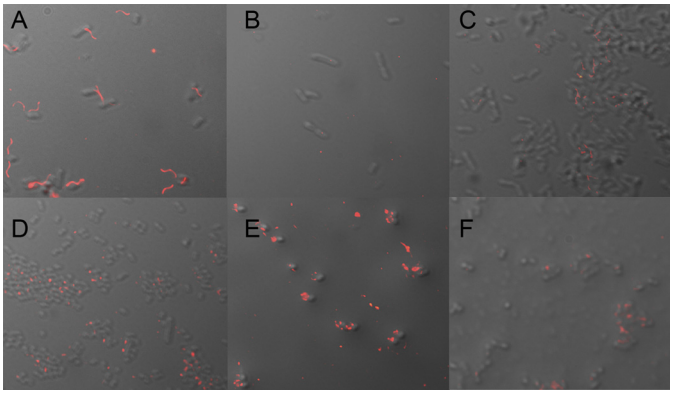


FIGURE 7. **Role of Tpx for appropriate expression of flagella and EspA filaments.** Bacteria were cultured in LB medium (*A–C*) or DMEM (*D–F*) to promote expression of flagella or EspA filaments. Wild-type bacteria (*A*) express flagella when visualized using immunofluorescence microscopy. The Tpx mutant (*B*) is non-motile and aflagellate, a phenotype that could be partially restored by introduction of plasmid-borne *tpx* (*C*). Culture in DMEM results in 45% of wild-type bacteria expressing EspA filaments (*D*). Deletion of *tpx* increased this proportion to 79% of the population (*E*), and complementation by plasmid-borne *tpx* returned this to wild-type levels (*F*). Images are composites created from differential interference contrast and immunofluorescence overlays.

sion and a 3-fold increase in *escJ* transcript level compared with the wild type.

Tpx Regulates Flagellar and T3SS Expression in E. coli O157—To determine whether the transcriptomic data translated into actual phenotypes, analyses of flagellar expression and effects on motility were tested using the *E. coli* O157 ΔTpx mutant. Immunofluorescence microscopy was used to visualize flagellar expression in the wild type, ΔTpx mutant, and in the complemented mutant. A culture of the bacteria was performed in LB medium at 28 °C as this is optimum for expression of flagella. More than 90% of wild-type bacteria expressed flagella in these conditions (Fig. 7*A*) with deletion of *tpx* reducing expression to <1% of bacteria (Fig. 7*B*). Restoration of flagellum expression could be achieved by introduction of *tpx* on a low copy number plasmid but not to wild-type levels (Fig. 7*C*). These data correlated well with the bacterial motility, which was markedly reduced for the ΔTpx mutant (supplemental Fig. S1). Treatment of wild-type *E. coli* O157 with ME0055 results in an increased proportion of bacteria expressing flagella (17). To examine effects on the expression of the T3SS, bacteria were cultured in DMEM. This medium provides an intermediate level of expression in wild-type strains, allowing any increases or decreases in the proportion of bacteria expressing the EspA filaments, the needle-like structures used to deliver effector proteins, to be measured. Some 45% of wild-type bacteria express EspA filaments in DMEM (Fig. 7*D* and supplemental Fig. S1), and deletion of *tpx* increased this proportion to 79% (Fig. 7*E* and supplemental Fig. S1), a result that is consistent with the transcriptomic data supporting the hypothesis that Tpx is involved in the regulation of both flagella and the T3SS, two key virulence factors in *E. coli* O157. Complementation resulted in a phenotype similar to that of wild type (Fig. 7*F* and supplemental Fig. S1).

Effect of Y. pseudotuberculosis Tpx and WrbA on Virulence in Vitro—To examine the contribution of the target proteins to pathogenesis, we created defined deletion strains of ΔTpx and ΔWrbA in *Y. pseudotuberculosis* and analyzed the effect on T3SS and virulence. Analysis of Yop secretion revealed that the secretion profiles of the ΔTpx and ΔWrbA mutants were elevated compared with that of wild-type bacteria (supplemental Fig. S2). This result correlates well with the transcriptomic data that also showed increased expression of genes associated with the T3SS. Secretion from all three strains could be reduced by the addition of 50 μM ME0052 (supplemental Fig. S2), indicating that the overall phenotype associated with the salicylidene acylhydrazide compounds results from affects on multiple target proteins. We also tested the ability of both ΔTpx and ΔWrbA mutants of *Y. pseudotuberculosis* to infect and kill a J774A macrophage-like cell line (10). Macrophage viability was measured by the addition of calcein AM and subsequent fluorescence analysis, with a ΔYopB translocation-deficient mutant as the negative control. Infection with wild-type bacteria resulted in 46 ± 16% viability compared with uninfected cells. The ΔTpx and ΔWrbA mutants were found to be as virulent as the wild-type strain with 42 ± 12 and 54 ± 4% viability, respectively. The addition of 25 μM of ME00052 rescued the macrophages infected with wild-type bacteria or with the ΔTpx and ΔWrbA mutants, restoring viability such that no macrophages were killed. Similarly, the ΔYopB mutant was avirulent, with no macrophages killed by this strain. These data demonstrate that no single target protein appears to be responsible for the phenotype associated with the addition of the compounds.

DISCUSSION

The identity, function, and three-dimensional structure of the target protein of a potential therapeutic compound constitute key information in a drug development process based on structure-based design (45, 46). The salicylidene acylhydrazide class of T3SS inhibitors was originally identified in a phenotypic bacterial reporter-gene screen (8). This strategy provides compounds that are active on the pathogenic organism and circumvents many of the drawbacks experienced in target-based screening based on proteins obtained by genomics (23). The salicylidene acylhydrazides have proved promising as T3SS inhibitors (19), and to further explore the potential of this compound class we have attempted to identify putative target proteins using an affinity reagent approach (47). In this work we have identified 16 proteins bound by the Affi-Gel-labeled compound, ME0055-Aff. These putative protein targets of the salicylidene acylhydrazides include WrbA, an NAD(P)H quinone oxidoreductase, and Tpx, a thiol peroxidase, both of which have been characterized in some detail (40, 41), and FolX, a dihydroneopterin-tri-P-epimerase. In contrast to WrbA and Tpx, the cellular function of FolX has been poorly characterized. How many of these proteins actually contribute to the phenotype associated with the addition of salicylidene acylhydrazide compounds is not clear. It has been established that metabolism is a key factor in the regulation of expression of type III secretion in *P. aeruginosa* (48). We would speculate that if the salicylidene acylhydrazides affect multiple proteins, then there could be a clear perturbation of normal metabolism, resulting in changes in gene expression and protein function. However, some interactions could have no consequence for virulence or result from weak and nonspecific binding. It is also feasible that the unla-

beled compound binds additional, unidentified proteins, as the modification of ME0055 by the addition of the spacer and Affi-Gel bead could affect binding to target proteins by blocking critical functional groups. Given that we successfully confirmed binding for three of these target proteins using Far Western blotting and showed NMR chemical shift perturbations for Tpx using unlabeled ME0052, we would suggest that several of the putative target proteins are likely to be true interacting partners with the salicylidene acylhydrazides. Furthermore, we have demonstrated that WrbA, Tpx, and FolX all contribute to the normal regulation and expression of key virulence factors, specifically the T3SS and flagella, and would suggest that a combination of perturbations to normal protein activity for multiple target proteins is most likely to explain the phenotype observed upon the addition of the salicylidene acylhydrazides.

Quantitative structure-activity relationship (QSAR) models have been computed for focused libraries (38, 39, 49, 50) using a strategy based on statistical molecular design (51). The QSAR models for the salicylidene acylhydrazides were successfully validated with an external test set, and several compounds inhibited virulence *in vitro* (39). The QSAR models were, however, hard to interpret, which is likely the result of the compounds being evaluated in a cell-based assay and would be consistent with the notion that they target multiple proteins. The QSAR models reflect overall properties beneficial for interaction with several proteins and other processes including cell permeability. A mode of action that includes several targets is beneficial from an antibiotic resistance perspective as the pathogen must alter several proteins and pathways to escape the drug. However, further optimization of the salicylidene acylhydrazides against multiple targets is challenging (38, 39, 52), although polypharmacology is increasingly recognized as important in drug discovery and development (53).

Given that the compounds affect the function of the T3SS of several Gram-negative pathogens, the obvious implication is that the different species share common protein targets that result in similar phenotypes. At least three mechanisms of inhibition have been suggested; they are direct effects on the T3SS basal apparatus proteins (16), less-direct effects on proteins that then affect T3SS expression and function (17), and possible changes in iron availability (13). With reference to the latter mechanism, our previous work indicated that for *E. coli* O157:H7, no operons involved in iron metabolism were affected by the addition of the salicylidene acylhydrazides and, therefore, that iron availability was not sufficient to explain the inhibition observed with this strain (17). The three target proteins we have focused on are well conserved among Gram-negative pathogens; we have shown that both Tpx and WrbA from five different pathogens are bound by the salicylidene acylhydrazides and that the compounds bind FolX from three species. Additionally, our identification of the strain that was insensitive to the effects of the salicylidene acylhydrazides showed that this strain had no genetic changes that would affect the proteins that comprise the T3SS. This "insensitive" strain contained a mutant allele of *wrbA*, encoding one of the proposed target proteins. Given the large number of SNPs in this strain, we do not assume or conclude that this particular non-synonymous change results in the "insensitivity," but it is an interesting observation that this target was identified by both approaches. The major conclusion from this aspect of the study was that the T3SS was highly conserved. Other genetic changes affecting cell wall structure, ABC transporters, or efflux pumps could all account for the phenotype observed with this strain. Overall, the weight of evidence in the present study favors a mechanism based on indirect effects that lead to changes in the regulation of the T3SS.

To understand the mechanism by which the salicylidene acylhydrazides affect individual target proteins, we focused on Tpx, a thiol peroxidase previously shown to be part of the oxidative stress defense system (40). One important result was the NMR chemical shift perturbation, as this showed that unlabeled compound binds to the target protein involving specific residues that cluster to a discrete area near the active site and the dimer interface. The binding site was modeled by docking the compound onto ecTpx (PDB code 3HVV), highlighting the residues identified by the chemical shift NMR experiments (Fig. 4).

Determination of both K_d and stoichiometry were achieved using an AUC-based method. This gave a K_d of ~50 μM and a stoichiometry of 1 ME0052 molecule per ypTpx dimer. This result correlates well with our Far Western data that showed preferential binding of ME0052-Bio to the ypTpx dimer compared with the monomer. The C61S mutant forms a less stable dimer, resulting in the majority of the protein migrating as a monomer after SDS-PAGE and less binding of the ME0052-Bio probe. In contrast, when the C61S protein was used for AUC studies, the protein was present as a dimer and gave a K_d approximately twice that of the wild type but with the same stoichiometry as the wild-type protein. The lower K_d observed for the C61S mutant is consistent with the NMR result that identified Cys-61 as being a residue involved in binding of ME0052.

Previous work has shown that Tpx protects against exogenous hydrogen peroxide and is required for survival in macrophages (54). Our data also showed that a *Y. pseudotuberculosis* ΔTpx mutant was as virulent against J774A macrophages as wild-type bacteria and that the mutant remained sensitive to ME0052. This result is in agreement with the *E. coli* transcriptomic data and supports the hypothesis that the salicylidene acylhydrazides act on multiple targets. Although deletion of *tpx* in *E. coli* O157 resulted in clear effects on gene transcription, we were initially surprised to see an up-regulation of genes associated with the T3SS. Our working hypothesis was that the function of the T3SS proteins would be directly inhibited by the addition of the compounds. However, the data from the arrays showed that in *E. coli* O157, Tpx function leads to repression of the T3SS. This can be attributed to the peroxidase activity of Tpx rather than any direct regulatory effect of the protein itself, as the C61S mutant displayed a transcriptional profile similar to that of the mutant with the defined deletion. However, the mechanisms by which Tpx affects gene expression are not clear. As both Tpx and WrbA function to protect against oxidative stress, it does make biological sense for their activity to be linked to activation of motility, allowing the bacteria to evade such stresses. Previous work has shown a clear link between bacterial response to numerous stresses and control of motility (55).

Given that the phenotype associated with addition of the salicylidene acylhydrazide compounds is a repression of T3SS expression, the implication is that these compounds may actually increase activity of Tpx and WrbA, thereby enhancing their repressive effects on the T3SS. The addition of the salicylidene acylhydrazide compounds does not directly affect expression of *tpx* or *wrbA* (17), suggesting that the compounds might directly activate or stabilize the enzymes. The preferential binding to the Tpx dimer coupled with the stoichiometry and NMR chemical shift assignment show that the compounds bind very close to the active site of Tpx. It is possible that the binding of the compound near the active site results in a positive effect on protein activity. Future work will map the binding site of the salicylidene acylhydrazide to more target proteins, allowing any common binding site to be identified, a key step on the route to the development of this class of anti-virulence compounds.

Acknowledgments—We thank Dr. Roger Parton and Dr. Dan Walker for their critical reading of the manuscript. The sequencing of the ZAP 430 strain was carried out in the Sir Henry Wellcome Functional Genomics Facility at the University of Glasgow.

REFERENCES

1. Cegelski, L., Marshall, G. R., Eldridge, G. R., and Hultgren, S. J. (2008) *Nat. Rev. Microbiol.* **6,** 17–27
2. Clatworthy, A. E., Pierson, E., and Hung, D. T. (2007) *Nat. Chem. Biol.* **3,** 541–548
3. Escaich, S. (2008) *Curr. Opin. Chem. Biol.* **12,** 400–408
4. Lynch, S. V., and Wiener-Kronish, J. P. (2008) *Curr. Opin. Crit. Care* **14,** 593–599
5. Rasko, D. A., and Sperandio, V. (2010) *Nat. Rev. Drug Discov.* **9,** 117–128
6. Hueck, C. J. (1998) *Microbiol. Mol. Biol. Rev.* **62,** 379–433
7. Coburn, B., Sekirov, I., and Finlay, B. B. (2007) *Clin. Microbiol. Rev.* **20,** 535–549
8. Kauppi, A. M., Nordfelth, R., Uvell, H., Wolf-Watz, H., and Elofsson, M. (2003) *Chem. Biol.* **10,** 241–249
9. Muschiol, S., Bailey, L., Gylfe, A., Sundin, C., Hultenby, K., Bergström, S., Elofsson, M., Wolf-Watz, H., Normark, S., and Henriques-Normark, B. (2006) *Proc. Natl. Acad. Sci. U.S.A.* **103,** 14566–14571
10. Bailey, L., Gylfe, A., Sundin, C., Muschiol, S., Elofsson, M., Nordström, P., Henriques-Normark, B., Lugert, R., Waldenström, A., Wolf-Watz, H., and Bergström, S. (2007) *FEBS Lett.* **581,** 587–595
11. Slepenkin, A., Enquist, P. A., Hägglund, U., de la Maza, L. M., Elofsson, M., and Peterson, E. M. (2007) *Infect. Immun.* **75,** 3478–3489
12. Wolf, K., Betts, H. J., Chellas-Géry, B., Hower, S., Linton, C. N., and Fields, K. A. (2006) *Mol. Microbiol.* **61,** 1543–1555
13. Hudson, D. L., Layton, A. N., Field, T. R., Bowen, A. J., Wolf-Watz, H., Elofsson, M., Stevens, M. P., and Galyov, E. E. (2007) *Antimicrob. Agents Chemother.* **51,** 2631–2635
14. Negrea, A., Bjur, E., Ygberg, S. E., Elofsson, M., Wolf-Watz, H., and Rhen, M. (2007) *Antimicrob. Agents Chemother.* **51,** 2867–2876
15. Nordfelth, R., Kauppi, A. M., Norberg, H. A., Wolf-Watz, H., and Elofsson, M. (2005) *Infect. Immun.* **73,** 3104–3114
16. Veenendaal, A. K., Sundin, C., and Blocker, A. J. (2009) *J. Bacteriol.* **191,** 563–570
17. Tree, J. J., Wang, D., McInally, C., Mahajan, A., Layton, A., Houghton, I., Elofsson, M., Stevens, M. P., Gally, D. L., and Roe, A. J. (2009) *Infect. Immun.* **77,** 4209–4220
18. Chu, H., Slepenkin, A., Elofsson, M., Keyser, P., de la Maza, L. M., and Peterson, E. M. (2010) *Int. J. Antimicrob. Agents* **36,** 145–150
19. Keyser, P., Elofsson, M., Rosell, S., and Wolf-Watz, H. (2008) *J. Intern. Med.* **264,** 17–29
20. Baron, C. (2010) *Curr. Opin. Microbiol.* **13,** 100–105
21. Puri, A. W., and Bogyo, M. (2009) *ACS Chem. Biol.* **4,** 603–616
22. Hung, D. T., and Rubin, E. J. (2006) *Curr. Opin. Chem. Biol.* **10,** 321–326
23. Payne, D. J., Gwynn, M. N., Holmes, D. J., and Pompliano, D. L. (2007) *Nat. Rev. Drug Discov.* **6,** 29–40
24. Terstappen, G. C., Schlüpen, C., Raggiaschi, R., and Gaviraghi, G. (2007) *Nat. Rev. Drug Discov.* **6,** 891–903
25. Leslie, B. J., and Hergenrother, P. J. (2008) *Chem. Soc. Rev.* **37,** 1347–1360
26. Sambrook, S., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbour Laboratory Press, New York
27. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V., and Mann, M. (2006) *Nat. Protoc.* **1,** 2856–2860
28. Pappin, D. J., Hojrup, P., and Bleasby, A. J. (1993) *Curr. Biol.* **3,** 327–332
29. Roe, A. J., Hoey, D. E., and Gally, D. L. (2003) *Biochem. Soc. Trans.* **31,** 98–103
30. Li, H., Ruan, J., and Durbin, R. (2008) *Genome Res.* **18,** 1851–1858
31. Wu, Y., Li, Q., and Chen, X. Z. (2007) *Nat. Protoc.* **2,** 3278–3284
32. Roe, A. J., Tysall, L., Dransfield, T., Wang, D., Fraser-Pitt, D., Mahajan, A., Constandinou, C., Inglis, N., Downing, A., Talbot, R., Smith, D. G., and Gally, D. L. (2007) *Microbiology* **153,** 1350–1360
33. Lane, M. C., Lockatell, V., Monterosso, G., Lamphier, D., Weinert, J., Hebel, J. R., Johnson, D. E., and Mobley, H. L. (2005) *Infect. Immun.* **73,** 7644–7656
34. Gabrielsen, M., Zetterström, C. E., Wang, D., Beckham, K. S., Elofsson, M., Isaacs, N. W., and Roe, A. J. (2010) *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **66,** 1606–1609
35. Mori, S., Abeygunawardana, C., Johnson, M. O., and van Zijl, P. C. (1995) *J. Magn. Reson. B* **108,** 94–98
36. Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J., and Laue, E. D. (2005) *Proteins* **59,** 687–696
37. Schuck, P. (2000) *Biophys. J.* **78,** 1606–1619
38. Dahlgren, M. K., Oberg, C. T., Wallin, E. A., Janson, P. G., and Elofsson, M. (2010) *Molecules* **15,** 4423–4438
39. Dahlgren, M. K., Zetterström, C. E., Gylfe, S., Linusson, A., and Elofsson, M. (2010) *Bioorg. Med. Chem.* **18,** 2686–2703
40. Baker, L. M., and Poole, L. B. (2003) *J. Biol. Chem.* **278,** 9203–9211
41. Patridge, E. V., and Ferry, J. G. (2006) *J. Bacteriol.* **188,** 3498–3506
42. Ploom, T., Haussmann, C., Hof, P., Steinbacher, S., Bacher, A., Richardson, J., and Huber, R. (1999) *Structure* **7,** 509–516
43. Hall, A., Sankaran, B., Poole, L. B., and Karplus, P. A. (2009) *J. Mol. Biol.* **393,** 867–881
44. Iyoda, S., Koizumi, N., Satou, H., Lu, Y., Saitoh, T., Ohnishi, M., and Watanabe, H. (2006) *J. Bacteriol.* **188,** 5682–5692
45. Simmons, K. J., Chopra, I., and Fishwick, C. W. (2010) *Nat. Rev. Microbiol.* **8,** 501–510
46. Barker, J. J. (2006) *Drug Discov. Today* **11,** 391–404
47. Guiffant, D., Tribouillard, D., Gug, F., Galons, H., Meijer, L., Blondel, M., and Bach, S. (2007) *Biotechnol. J.* **2,** 68–75
48. Rietsch, A., and Mekalanos, J. J. (2006) *Mol. Microbiol.* **59,** 807–820
49. Kauppi, A. M., Andersson, C. D., Norberg, H. A., Sundin, C., Linusson, A., and Elofsson, M. (2007) *Bioorg. Med. Chem.* **15,** 6994–7011
50. Dahlgren, M. K., Kauppi, A. M., Olsson, I. M., Linusson, A., and Elofsson, M. (2007) *J. Med. Chem.* **50,** 6177–6188
51. Linusson, A., Elofsson, M., Andersson, I. E., and Dahlgren, M. K. (2010) *Curr. Med. Chem.* **17,** 2001–2016
52. Hillgren, J. M., Dahlgren, M. K., To, T. M., and Elofsson, M. (2010) *Molecules* **15,** 6019–6034
53. Morphy, R., and Rankovic, Z. (2009) *Curr. Pharm. Des.* **15,** 587–600
54. Horst, S. A., Jaeger, T., Denkel, L. A., Rouf, S. F., Rhen, M., and Bange, F. C. (2010) *J. Bacteriol.* **192,** 2929–2932
55. Nachin, L., Nannmark, U., and Nyström, T. (2005) *J. Bacteriol.* **187,** 6265–6272
56. DeLano, W. L. (2010) *The PyMOL Molecular Graphics System*, Version 1.3r1, Schrodinger, LLC, New York

OPEN ACCESS Freely available online

PLoS one

Structural Characterisation of Tpx from *Yersinia pseudotuberculosis* Reveals Insights into the Binding of Salicylidene Acylhydrazide Compounds

Mads Gabrielsen[1◕], Katherine S. H. Beckham[1◕], Victoria A. Feher[2], Caroline E. Zetterström[3], Dai Wang[1¤], Sylke Müller[4], Mikael Elofsson[3], Rommie E. Amaro[2], Olwyn Byron[5]*, Andrew J. Roe[1]*

1 Institute of Infection, Immunity and Immunology, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, United Kingdom, **2** Departments of Pharmaceutical Sciences, Computer Science, and Chemistry, University of California Irvine, Irvine, California, United States of America, **3** Department of Chemistry and Umeå Centre for Microbial Research, Umeå University, Umeå, Sweden, **4** Wellcome Centre for Molecular Parasitology, Institute of Infection, Immunity and Immunology, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, United Kingdom, **5** School of Life Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, United Kingdom

Abstract

Thiol peroxidase, Tpx, has been shown to be a target protein of the salicylidene acylhydrazide class of antivirulence compounds. In this study we present the crystal structures of Tpx from *Y. pseudotuberculosis* (*yp*Tpx) in the oxidised and reduced states, together with the structure of the C61S mutant. The structures solved are consistent with previously solved atypical 2-Cys thiol peroxidases, including that for "forced" reduced states using the C61S mutant. In addition, by investigating the solution structure of *yp*Tpx using small angle X-ray scattering (SAXS), we have confirmed that reduced state *yp*Tpx in solution is a homodimer. The solution structure also reveals flexibility around the dimer interface. Notably, the conformational changes observed between the redox states at the catalytic triad and at the dimer interface have implications for substrate and inhibitor binding. The structural data were used to model the binding of two salicylidene acylhydrazide compounds to the oxidised structure of *yp*Tpx. Overall, the study provides insights into the binding of the salicylidene acylhydrazides to *yp*Tpx, aiding our long-term strategy to understand the mode of action of this class of compounds.

Citation: Gabrielsen M, Beckham KSH, Feher VA, Zetterström CE, Wang D, et al. (2012) Structural Characterisation of Tpx from *Yersinia pseudotuberculosis* Reveals Insights into the Binding of Salicylidene Acylhydrazide Compounds. PLoS ONE 7(2): e32217. doi:10.1371/journal.pone.0032217

Editor: Petri Kursula, University of Oulu, Finland

Received November 3, 2011; **Accepted** January 25, 2012; **Published** February 27, 2012

Copyright: © 2012 Gabrielsen et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by a grant from the Biotechnology and Biological Sciences Research Council to AJR and MG (BB/G011389/1) and a Medical Research Scotland grant (ref. 223 ORG G 0709) to AJR and DW. KSHB is supported by a Wellcome Trust studentship. CZ and ME are supported by the Swedish Research Council. This work was supported in part by the National Institutes of Health (NIH) through the NIH Director's New Innovator Award Program, 1-DP2-OD007237 to REA. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: Olwyn.Byron@glasgow.ac.uk (OB); Andrew.Roe@glasgow.ac.uk (AJR)

¤ Current address: MRC Centre for Regenerative Medicine Scottish Centre for Regenerative Medicine, The University of Edinburgh, Edinburgh BioQuarter, Edinburgh, United Kingdom

◕ These authors contributed equally to this work.

Introduction

Thiol peroxidase (Tpx, p20, scavengase) is an atypical 2-Cys peroxiredoxin present throughout the eubacteria, including pathogenic strains, such as *Escherichia coli* O157:H7 [1], *Yersinia* sp., *Haemophilus influenzae*, *Streptococcus pneumoniae* and *Helicobacter pylori* [2]. Tpx constitutes part of the bacterial defence system against reactive oxygen species (ROS) and, correspondingly, is upregulated when *E. coli* is exposed to oxidative stress [1]. Tpx functionality specifically relies on the reducing equivalents from thioredoxin (Trx1) and thioredoxin reductase (TrxR) [3]. The catalytic cycle of peroxiredoxin activity consists of three steps; 1) peroxidation, 2) resolution and 3) recycling [4]. Atypical 2-Cys peroxiredoxins are functionally monomeric, in contrast to the typical peroxiredoxins, i.e. the resolving (C_r) and the peroxidasic (C_p) cysteines (C61 and C95, respectively in the case of Tpx) are situated on the same subunit. Structurally, this involves the reduced Tpx encountering a ROS, such as hydrogen peroxide or an alkyl hydroperoxide, and the covalent binding of O^- to C_p. The ROS is released as H_2O resulting in the formation of a disulphide bridge between C_p and C_r. The cycle is completed by a transient interaction with Trx1, ending with two separate cysteine side-chains on Tpx (Figure 1).

Tpx contains three cysteine residues, two of which (C61and C95) form the redox active disulphide bond. The third cysteine (C82) is not involved in the redox activities of Tpx [3], and is not involved in any covalent interactions. Until now, twelve structures of Tpx have been elucidated, from *E. coli* [4,5], *Bacillus subtilis* [6], *Aquifex aeolicus*, *Mycobacterium tuberculosis* [7,8], *H. influenzae*, and *S. pneumoniae*. Most of these structures have been solved in the oxidised state, or in the "forced" reduced state of the C61S (or equivalent) mutant. Two wild-type reduced structures have been solved, one by NMR [6] and one by X-ray crystallography (Structural Genomics Consortium).

Initially presumed to be localised in the periplasm [1], recent work using cross-linking and fractionation studies [9] has shown

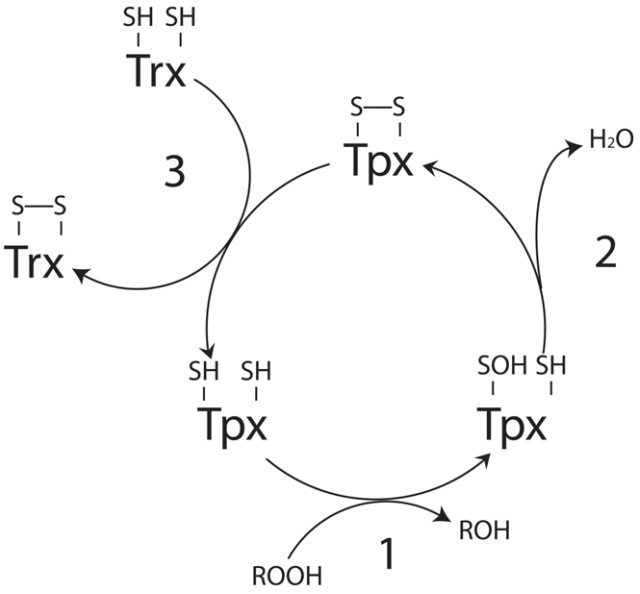


Figure 1. Redox cycle of Tpx. The reduction of ROS by oxidation (step 1) of Tpx producing H_2O (step 2). The oxidised form (disulphide bond) of Tpx is recovered by thioredoxin (Trx), and returned to the reduced form (step 3).
doi:10.1371/journal.pone.0032217.g001

that Tpx is one of several peroxiredoxins in the cytosol of *E. coli*. Tpx has been shown to be important for the survival of *S. typhimurium* in macrophages, where the oxidative burst can be particularly acute [10].

We have recently shown that Tpx is one of several proteins bound by a class of "anti-virulence" compounds, the salicylidene acylhydrazides [11]. These compounds are broadly effective in reducing the expression of the type three secretion system (T3SS) of a range of Gram-negative pathogens including *Chlamydia* spp., *Salmonella enterica* serovar Typhimurium, *Yersinia pseudotuberculosis*, *Shigella flexneri*, and *E. coli* O157 [12]. The T3SS is a critical determinant used by pathogens to modulate host cell processes and facilitate processes such as binding and invasion [13] so compounds that interfere with its expression or function have the potential to become novel anti-infective agents [12,14].

The precise molecular mechanism of action of the salicylidene acylhydrazides is not fully understood, although our identification of multiple binding proteins suggests a synergistic effect arising from a modulation of the activity of several proteins, including Tpx. The binding affinity of the salicylidene acylhydrazide compound ME0052 [N′-(3,5-dibromo-2-hydroxy-benzylidene)-nicotinic acid hydrazide] to Tpx from *Y. pseudotuberculosis* (*yp*Tpx) was measured using analytical ultracentrifugation (AUC) and showed that a catalytically inactive mutant of Tpx displayed a two-fold reduction in binding [11]. In this mutant, cysteine 61 was specifically mutated to a serine residue (C61S). The C61S mutant Tpx is present only in its reduced form and cannot undergo the intramolecular disulphide bond formation critical for the catalytic cycle of the protein.

In the current study, we present the crystal structures of *yp*Tpx in the oxidised and reduced states together with the structure of the C61S mutant. These structural data, combined with our previous NMR chemical shift analysis, allow us to perform detailed molecular modelling of how the salicylidene acylhydrazides bind to target proteins. This work helps our understanding of the mode-of-action of this class of anti-virulence compounds.

Results and Discussion

Purified Tpx is active and bound by the salicylidene acylhydrazides

Tpx can be readily purified using nickel affinity chromatography, thereby facilitating structural and subsequent biochemical studies. The activity of purified *yp*Tpx was tested using a glutamine synthetase (GS) assay, where an active peroxiredoxin protects against ROS [15]. Two micrograms of *yp*Tpx rescued 50% of initial GS activity, with 15 μg of *yp*Tpx raising GS activity to 90% of initial activity, clearly demonstrating that *yp*Tpx was indeed active and reduced H_2O_2 (Figure 2). Previous characterisation of Tpx from *E. coli* (*ec*Tpx) has shown a substrate specificity for alkyl hydroperoxides over H_2O_2, with a K_m of 9 μM for cumene peroxide compared with a K_m>1.7 mM for H_2O_2 [3]. However, despite this lower substrate specificity, analysis of mutants in Salmonella revealed that a Tpx mutant was highly susceptible to exogenous H_2O_2 [10].

Crystal structures

To determine the high-resolution structure of Tpx, recombinant *yp*Tpx in reduced and oxidised state, and the mutants *yp*TpxC61S were crystallised. *yp*Tpx crystallised in three crystal forms, 1, 2 and 3, in three different space groups, $P2_1$, $P6_4$ and $P2_12_12_1$, respectively, as described elsewhere [16]. Crystal forms 1 and 2 grew in conditions containing DTT, and the solved structures were in the reduced state. Crystal form 1 diffracted to 2.00 Å, and the structure comprised three dimers in the asymmetric unit. The six chains superpose well, with root-mean-square-deviations (r.m.s.d.) of less than 0.5 Å. Crystal form 2, diffracting to 2.35 Å, comprises a single subunit in the asymmetric unit, the complete dimer being made up by symmetry operators.

As there are only minor differences between the two reduced structures, with an r.m.s.d. of 0.4 Å over 160 C_α, only the highest resolution structure (space group $P2_1$) will be discussed here. Most residues are accounted for in the electron density, apart from the hexa-histidine tag. The reduced structure refined to R_{work}- and R_{free} -factors of 22.2% and 26.8%, respectively. Refinement statistics for all structures are presented in Table 1.

*yp*Tpx has a regular thioredoxin-like fold: a seven-stranded β-sheet, with β2 and β6 running anti-parallel to the rest, although with an inserted N-terminal β-hairpin (βN1-βN2) (Figure 3) absent

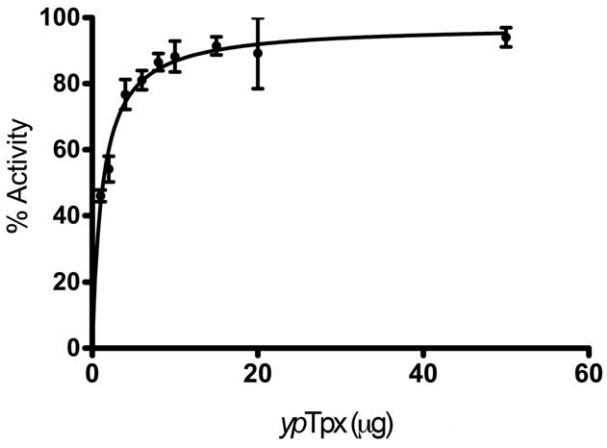


Figure 2. Glutamine synthetase assay. The activity of *yp*Tpx is confirmed by its ability to rescue the activity of glutamine syntethase (to 90% of original activity at 15 μg *yp*Tpx).
doi:10.1371/journal.pone.0032217.g002

Table 1. Refinement statistics for reported structures.

	Reduced Tpx	Reduced Tpx	Oxidised Tpx	Tpx C61S
PDB code	**2XPD**	**3ZRE**	**3ZRD**	**2YJH**
Space group	$P2_1$	$P6_4$	$P2_12_12_1$	$P6_4$
Unit cell (Å)	a = 64.86 b = 92.07 c = 85.60 β = 91.41°	a = 65.01 c = 86.28	a = 56.18 b = 62.56 c = 88.00	a = 65.03 c = 85.60
Resolution (Å)	44.67 – 2.00	56.30 - 2.35	28.09 - 1.74	56.32 - 2.55
Protein residues (atoms)	903 (7449)	165 (1191)	339 (2517)	164 (1188)
Water molecules	490	12	199	4
Ligands	DTT	N/A	N/A	N/A
R_{work} (%)	22.5	23.5	19.5	31.3
R_{free} (%)	26.8	27.3	23.13	33.6
R.m.s.d. for bond lengths (Å)/angles (°)	0.022/1.8	0.02/1.995	0.009/1.080	0.005/0.772
Ramachandran allowed regions (%)	99.8	97.6	100.0	97.6

doi:10.1371/journal.pone.0032217.t001

from other peroxierdoxins. The central sheet is flanked by four α-helices following β3, β4, β5 and β7, and one short 3_{10} helix following β2. The numbering of the β-strands is based upon that of *ec*Tpx [5] to make direct comparisons between all Tpx molecules easier. βN1-βN2 (Figure 3) forms an L-shaped hydrophobic cleft, and it has been speculated that this cleft allows Tpx to accommodate the long fatty acid hydroperoxides [17].

Crystal form 3 captured *yp*Tpx in the oxidised state with an intact intramolecular disulphide bond between Cys61 and Cys95. The crystal belonged to space group $P2_12_12_1$ and diffracted to 1.74 Å. The overall oxidised structure, diffracting to a resolution of 1.74 Å and presenting space group $P2_12_12_1$, is similar to that of the reduced structure, except for some differences that are mostly confined to the region around the active site (see below).

As part of this study, the structure of *yp*TpxC61S was solved to a resolution of 2.55 Å in space group $P6_4$. This structure represents the "forced" reduced form of the protein, as the resolving cysteine has been mutated to a serine, rendering it catalytically inactive [3]. All of our solution data indicate that the mutant structure and the reduced wild type structure are identical, and that the oligomeric states are the same. When superimposed onto the reduced structure, the r.m.s.d. was 0.52 Å over 163 C_α (Figure S1). This fits well with the structural analysis of Hall et al. [4] who used the TpxC61S mutant from *E. coli* to describe the structure of reduced Tpx.

*yp*Tpx crystallised either as a dimer in the asymmetric unit, where the two subunits superposed with an r.m.s.d. of less than 0.2 Å, or the dimer could be created by crystallographic symmetry operators. The dimer interface comprises about 20 residues from each subunit, corresponding to 12% of total surface residues, according to the PISA server [18]. The interfaces are formed mostly by hydrophobic interactions, with a few hydrogen bonds, namely between R110NH1 and three main-chain carboxyl groups on the opposing subunit (G125, P126, A128). There are no salt bridges or covalent bonds between the two dimers in the reduced structure. The dimer interfaces are identical in the structures of reduced Tpx and C61S.

The dimer interface is of similar size in the oxidised state but in addition to the hydrogen bond pattern described above, there are also salt bridges formed between D57 and R93 on opposing subunits (Figure 3C). This is due to the conformational change between the reduced and oxidised states.

Active site

The redox active site of Tpx is made up by C_P and C_R (C61 and C95 in *yp*Tpx, respectively). There is a conformational change between the two states, involving the partial unfolding of helices α1 and α2, and a shift of 8.4 Å for C61 and 5.1 Å for C95, respectively, as presented in Figure 4. The two structures superpose well, in particular the core parts, with an r.m.s.d of 0.7 Å over 135 C_α. Inclusion of the unfolding helices increases the r.m.s.d. to 1.02 Å.

The partial unfolding of α1 opens a cleft in Tpx formed between the loops connecting β1 and α1, α3 and β6, and β7 and α5 on subunit A and connecting β1 and β2, and β4 and α2 on subunit B. When in the reduced state, C61 is orientated into the pocket where it is available for oxidation by H_2O_2 or alkyl peroxides. This cleft makes up the active site of Tpx, and has been described in detail by Hall et al. [4]. In this manuscript they present the fully intact peroxide binding site (a reduced C61S mutant), the locally unfolded binding site (oxidised), and a partially unfolded transitional state (seen only for the double C82, 95S mutant) for Tpx from *E. coli* that shares an identical active site.

Oligomeric state

Peroxiredoxins exhibit a wide variety of oligomeric states, ranging from monomeric (YPrx, [19]), to large decameric or dodecameric assemblies like TryP [20], AhpC [21] and other typical 2-Cys peroxiredoxins, including PrxIII from bovine mitochondria, which forms two concatenated dodecamers [22]. These assemblies are often dependent on redox state, dissociating into homodimers upon oxidation [23]. Previous studies of *ec*Tpx showed that the protein is a homodimer, regardless of the redox state, and despite the lack of any inter-subunit disulphide bond [3]. We analysed the oligomeric state of oxidised and reduced *yp*Tpx, as well as the C61S mutant by AUC and SAXS. Sedimentation velocity (SV) experiments revealed that all three forms of *yp*Tpx were completely monodisperse in solution, as evidenced by a single dominant peak in the c(s) distribution (Figure 5). Infinite dilution sedimentation coefficients ($S^0_{20,w}$) were determined from the concentration dependence of $s_{20,w}$ (obtained from fitting the SV data with a non-interacting discrete species model in SEDFIT [24]) for the oxidised and reduced forms of *yp*Tpx ($S^0_{20,w} = 3.04$

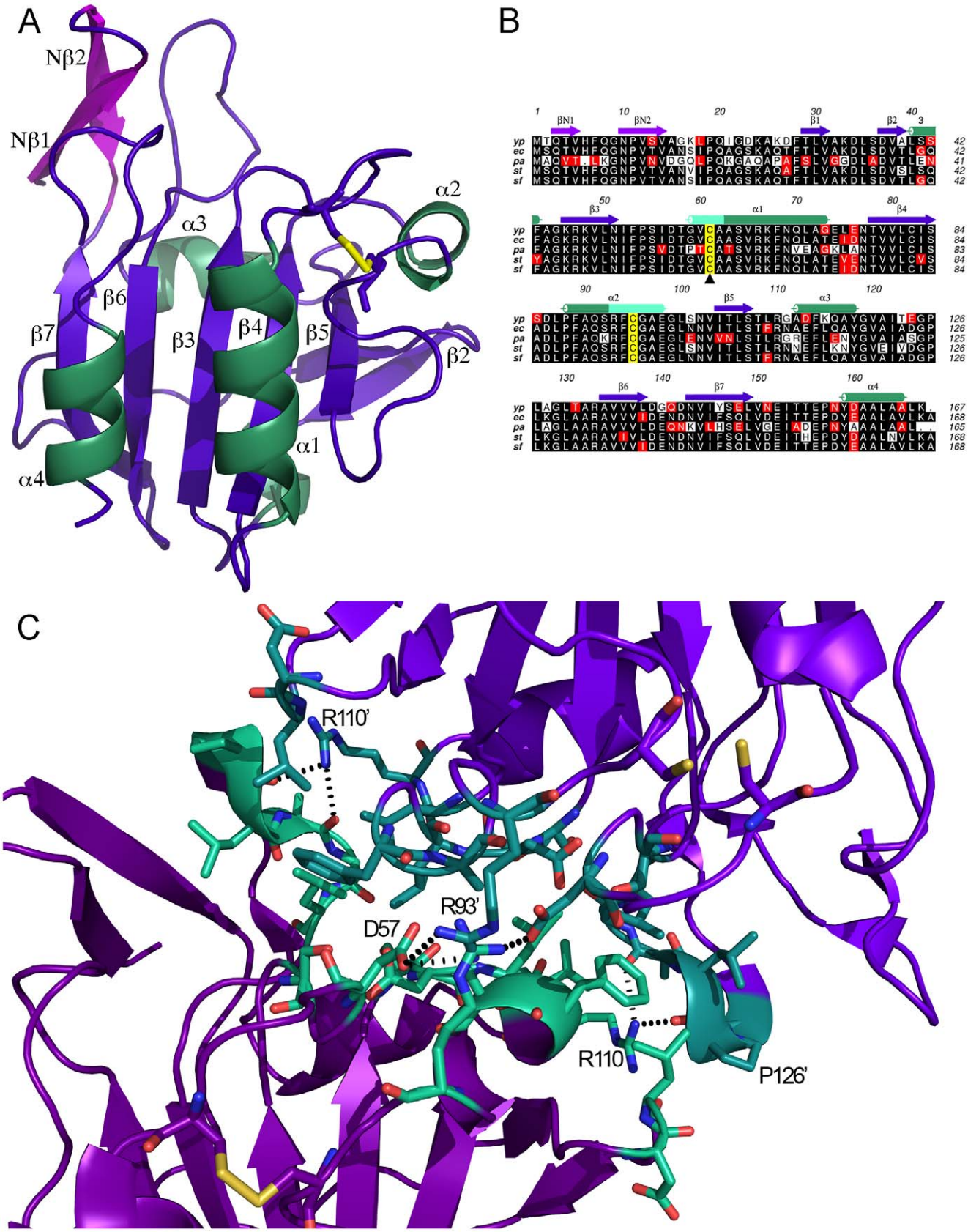


Figure 3. Crystal structures. (A) Cartoon representation of oxidised *yp*Tpx. Strands and loops are purple, helices are green, and the disulphide bond is represented as sticks. The N-terminal hairpin is highlighted in pink. The secondary structure elements are labelled. (B) Sequence alignment of Tpx from a number of pathogens, with the secondary structure based on *yp*Tpx. Black represents identical, and red highly similar residues (based on

an ALSCRIPT algorithm level of 0.7 [53]). The unravelling regions of α1 and α2, caused by the change of redox state, are highlighted by the lighter colour in the secondary structure elements. The reactive C residues are highlighted in yellow, and C61S is marked by a triangle. (C) Close-up of the dimer interface with the residues involved interactions marked as sticks, and salt bridges and hydrogen bonding partners are labelled. Bonds are shown in dashes.
doi:10.1371/journal.pone.0032217.g003

and 2.62 S, respectively). For *yp*Tpx C61S $S^0_{20,w}$ is 2.80 S, suggesting that this mutation does not induce structural instability (i.e. the value is comparable with those determined for reduced and oxidised *yp*Tpx).

Sedimentation equilibrium (SE) data were fitted with the species analysis model in SEDPHAT [25]. From the concentration dependence of the resultant apparent mass of the single species the infinite dilution mass (M^0) was determined to be 41.5±3.4, 38.4±2.9, 39.4±0.4 kDa for oxidised, reduced and C61S *yp*Tpx respectively. The mass of *yp*Tpx dimer, including the tag, calculated from its amino acid sequence is 42,382 Da, which is consistent with the experimentally determined masses. This indicates that *yp*Tpx is present solely as a dimer in solution. It was not possible to fit the SE data with a monomer-dimer (or any other plausible) self-association model, which is further consistent with the complete dimerisation of the protein in the concentration range studied.

Solution structure

The solution structures of *yp*Tpx and the C61S mutant were investigated using SAXS, a powerful method to structurally analyse proteins in solution under more physiologically relevant conditions [26]. Figure 6A shows a SAXS curve for *yp*TpxC61S, representative of the data obtained for *yp*Tpx in both oxidising and reducing conditions. *yp*TpxC61S was obtained at a higher concentration than the other samples, and subsequently produced better scattering data. The D_{max} and R_g of *yp*Tpx and *yp*TpxC61S, obtained by indirect Fourier transform with GNOM [27], were the same (70.5 Å and 24.0±0.2 Å, respectively) indicative that conformational changes induced by disulphide bond formation are too small to be detected by SAXS. Theoretical scattering curves of monomeric and dimeric atomic structures of Tpx were calculated, and again confirm that Tpx is a dimer in solution (Figure 6A). A low-resolution (11 Å) envelope of *yp*TpxC61S (Figure 6B) in solution was generated using the *ab initio* modelling program DAMMIN [28]. The fit of the model to the data is shown in Figure 6A.

The high-resolution structure superimposes well onto the low-resolution envelope (Figure 6B). The D_{max} of the space-fill model of the dimer crystal structure is approximately 68 Å, which agrees with the D_{max} obtained from the SAXS data (70.5 Å), indicating that the low-resolution envelope describes the *yp*Tpx dimer. The differences in the D_{max} values obtained from the two methods are small, and may be explained by the fact that in the crystal structure there is no electron density to account for the two N-terminal residues of Tpx plus the hexa-histidine tag, therefore it has not been included in calculations. However, as these residues were present in the *yp*Tpx studied by SAXS, we would expect the D_{max} value observed in solution by SAXS to exceed that calculated for the incomplete crystal structure.

Rigid body modelling of the oxidised Tpx crystal structure against the SAXS data, using BUNCH [29], based on a single chain, and imposing *P*2 symmetry yielded a model similar to that for the dimeric crystal structure. Comparison of the crystallographic model with the one fitted to the solution data using DYNDOM [30] yielded a rotation angle of 21.4° and a 5 Å translation. This freedom of movement corresponds well with that observed for the structures of Tpx from other different species [4].

Modelling of salicylidene acylhydrazide compounds to *yp*Tpx

We have previously used NMR chemical shift mapping to identify *yp*Tpx amide groups that were shifted upon addition of 200 μM ME0052. The study mapped these shifting residues onto the published TpxC61S structure from *E. coli* (PDB code 3HVV) to show they clustered to a defined region of the protein. Now we have obtained the high-resolution structure of *yp*Tpx itself, allowing us to model the binding of ME0052 and ME0055 to both the oxidised and reduced forms of the protein and examine how this correlates with the NMR data. These two compounds were docked into the receptor structures using MOE Dock, and the 25 best poses determined for each compound were ranked after energy minimisation and dock scoring. Figure 7 shows the lowest energy binding modes for ME0052 (Figure 7A) bound to oxidised *yp*Tpx. Docking using ME0055 gave equivalent binding poses (data not shown). The binding pocket is mostly hydrophobic (Figure 7A) with one hydrogen bond proposed between the ME0052 p-hydroxyl to the *yp*Tpx I153 carbonyl, which fits with the chemical shift change for the neighbouring T154 amide (Figure 7B). The chemical shift data indicated significant shifts in the amides of residues from both subunits, highlighting the importance of the dimer interface for the generation of the binding pocket and compound binding, as illustrated in Figure S2A. The binding site is also consistent with previously published Tpx-substrate models [32].

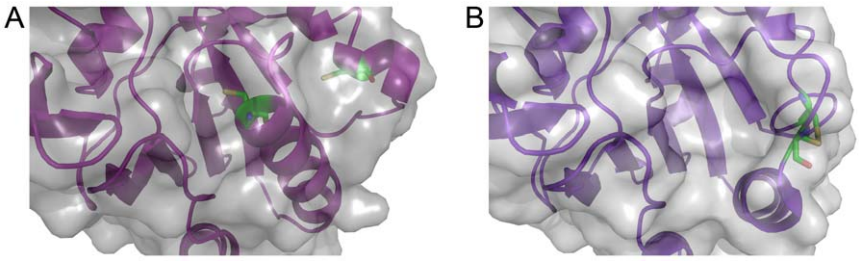


Figure 4. Comparison of the oxidised and reduced active site. (A) Close-up of the active cysteines in the reduced structure. C61 is shown to occlude the active site cleft. (B) Close-up of the reduced structure. The formation of the disulphide bond shifts the helix and opens a cleft, which allows substrate access.
doi:10.1371/journal.pone.0032217.g004

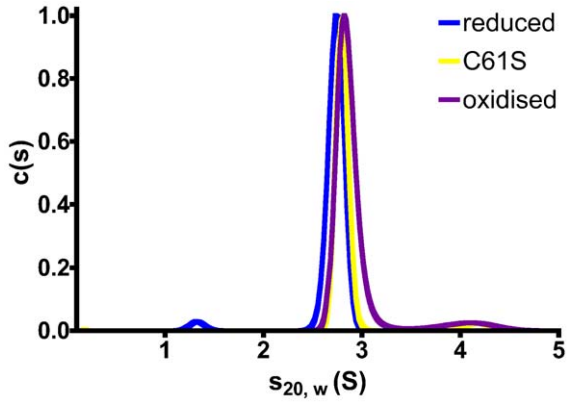


Figure 5. Analytical ultracentrifugation. c(s) distributions derived via SEDFIT from SV data for *yp*Tpx in oxidised (purple) and reduced (blue) states, as well as *yp*TpxC61S (yellow) indicate that the protein is monodisperse (dominated by a single, symmetrical peak) with the same sedimentation coefficient regardless of redox state.
doi:10.1371/journal.pone.0032217.g005

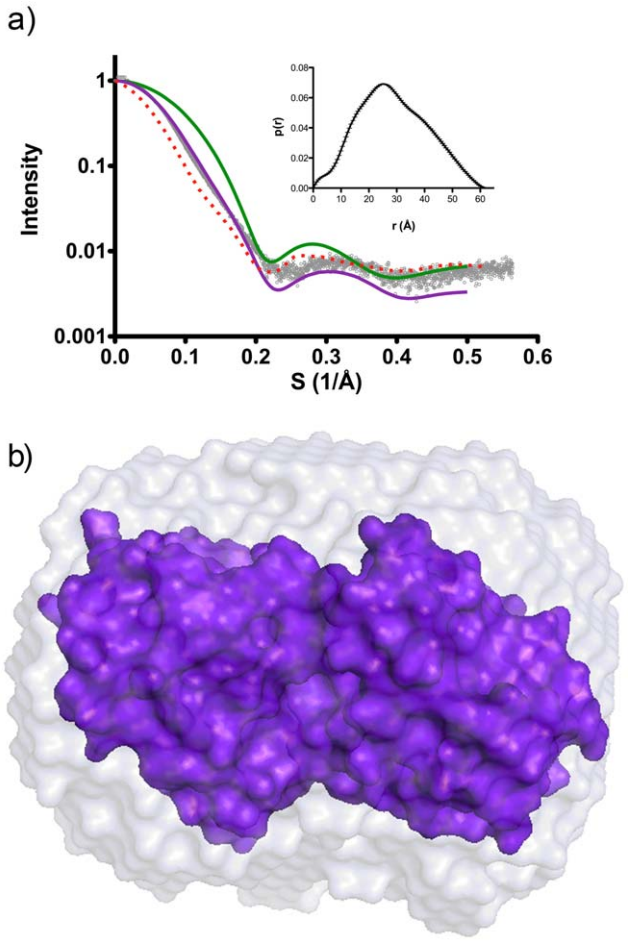


Figure 6. SAXS results. (A) The experimental scattering curve of *yp*TpxC61S (grey), overlaid with the curves for monomeric Tpx (green) and dimeric Tpx (purple) (calculated using CRYSOL [50]) confirming the dimeric solution state of the protein. The fit of the *ab initio* model to the experimental data is shown in red. Inset is the distance distribution function (p(r) *versus* r) of *yp*TpxC61S with error bars. (B) Space-filled crystal structure of *yp*TpxC61S (purple) superposed onto the averaged DAMMIN *ab initio* model (grey) using the program SUPCOMB [49].
doi:10.1371/journal.pone.0032217.g006

During transition to the reduced form, the loop containing G59 to C61 folds into the binding pocket predicted to accommodate the compound. Similarly, the side chain of R133 undergoes a significant conformational change. Collectively, these conformational changes reduce the overall volume of the binding pocket (Figure 7C) and are predicted to affect the binding of ME0052 to the reduced state of *yp*Tpx by inducing steric clashes. In fact, MOE Dock as not able to find a suitable docking pose for the reduced state of *yp*Tpx, presumably because the binding site is occluded by the folded extension of the α1 helix in the reduced state.

Accurately measuring the binding of the salicylidene acylhydrazides to any protein has proved problematic due to the low solubility of the compounds in physiologically relevant solvents. This has prohibited the application of techniques including isothermal calorimetery and surface plasmon resonance that would be default methods to measure binding of ligands to proteins. Previously we have used an AUC-based method to estimate K_d, giving values of 51 and 71 μM for the binding of ME0052 to oxidised and reduced *yp*Tpx, respectively [11]. These data suggest only minor differences in compound binding to *yp*Tpx in the two oxidation states, although we would exercise caution in the interpretation of these data due to the inherent inaccuracy of measuring K_d by this method.

An alternative binding pose, that would accommodate binding of ME0052 or ME0055, to both the oxidised and reduced forms of *yp*Tpx is presented as Figure S2B Although this model would require some conformational accommodation of the binding site, the compounds are predicted to be less buried in the pocket and therefore binding would be largely equivalent irrespective of the oxidation state of Tpx. Such conformational accommodation is plausible given the large backbone fluctuations observed in each state.

Far-western analysis

To test the binding of *yp*Tpx to salicylidene acylhydrazide compounds, far-western blotting was used in which the protein was resolved by SDS-PAGE, transferred to a nitrocellulose membrane and probed using biotinylated ME0052 (ME0052-bio) [33,34]. Interactions were then detected using a StreptAvidin-HRP conjugate in a process similar to that used for a routine western blot. Following SDS-PAGE and staining by Coomassie blue, Tpx could be seen as a monomer of 21 kDa as well as a dimer of 42 kDa (Figure 8). In comparison, the purified C61S mutant was present only as a monomer of 21 kDa. This difference can be attributed to the changes in the strength of the dimer interface, arising from the loss of two salt bridges in the 'forced' reduced C61S mutant, which in turn is more susceptible to the heat and detergent experienced during the far-western blotting procedure. Far-western analysis indicated that ME0052-bio binds with a far higher affinity to *yp*Tpx dimer, as over four times the amount of signal was seen corresponding to the dimer compared with the monomer (see Material and Methods). This finding is particularly stark when the relative proportions of dimer and monomer indicated on the gel are considered: quantification of the monomer-dimer on the SDS-PAGE gel indicated the ratio of these two species was 15:1. This finding confirmed our previous data [11] showing binding of ME0052-bio when tested against *E. coli* overexpressing *ec*Tpx but is more unambiguous as it demonstrates binding to purified protein rather than through probing crude lysates. The caveat with this approach is that the protein must refold before binding the ligand. It could be that the better binding of the ligand to the dimer is simply due to an inherently more stable dimeric structure compared with the dissociated monomer.

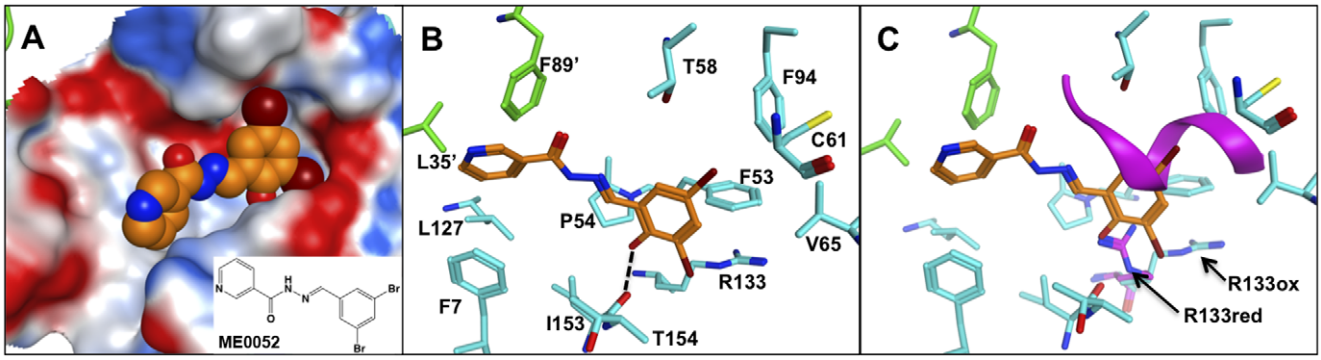


Figure 7. Model for salicylidene acylhyrazide ME0052 binding mode to oxidised *yp*Tpx crystal structure. (A) Binding mode for lowest docked energy ME0052 (atoms represented as CPK orange for carbon, blue for nitrogen and red for bromines) and electrostatic surface for the protein binding site (Connolly surface, sphere radius = 1.4 Å). ChemDraw representation for the ME0052 inhibitor is inset. (B) Detailed view of ME0052 binding mode (stick representation). Oxidised *yp*Tpx residues with atom contacts within 3.5 Å of ME0052 inhibitor are shown (cyan). Residues F89′ and L35′ are contributed by the neighbouring monomer (green). A hydrogen bond (dashed line) is formed between inhibitor and I153 carbonyl. (C) ME0052 binding mode from (B) with reduced *yp*Tpx (protein backbone magenta ribbon) superposed on oxidised *yp*Tpx; shown are α1 helix and the R133 side chain of the reduced form partially occluding the binding pocket of oxidised form.
doi:10.1371/journal.pone.0032217.g007

Therefore, the dimer refolds better than the monomer and gives more signal. However, despite this reservation, the preferential binding of the compound to the *yp*Tpx dimer is consistent with our modelling data and previous NMR studies that indicated both subunits contribute to the binding pocket [11].

Conclusion

In summary, we have solved the high-resolution structure of *yp*Tpx in three forms enabling us to model the binding of salicylidene acylhydrazide compounds. Binding of the compound, ME0052, was found (by far-western blotting) to be markedly stronger to the Tpx dimer compared with the monomer. This is consistent with the surface area of the modelled binding site predominantly comprising one subunit yet also including the dimer interface. The solution structure confirms the oligomeric state of the protein for both redox states.

Overall, the study provides insights into the binding of the salicylidene acylhydrazide compounds to *yp*Tpx, aiding our long-term strategy aiming to understand the mode of action of these compounds. Further studies characterising the role of the conformational flexibility observed around the catalytic triad and dimeric interface in ligand binding may yield additional insights into the binding mechanisms of these compounds and guide efforts to design even more effective inhibitors.

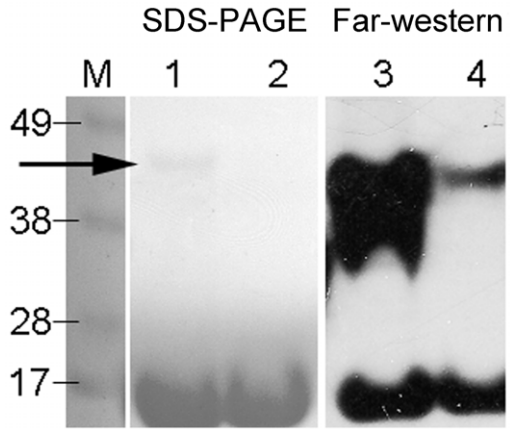


Figure 8. Far-western of purified *yp*Tpx. SDS-PAGE and Coomassie staining of: lane M, molecular weight ladder; lane 1, *yp*Tpx; lane 2, *yp*TpxC61S and (lanes 3 and 4 respectively) far-western of: *yp*Tpx and *yp*TpxC61S probed with ME0052-bio. The arrow highlights the presence of a small amount of dimeric *yp*Tpx on the SDS-PAGE, which can be contrasted with the far stronger signal from the corresponding band on the far-western.
doi:10.1371/journal.pone.0032217.g008

Materials and Methods

Protein expression and purification

*yp*Tpx and *yp*TpxC61S were expressed and purified as described previously [16], and the N-terminal hexa-histidine tag formed part of the expressed protein.

Glutamine synthetase assay

The enzymatic activity of *yp*Tpx was demonstrated by a glutamine synthetase (GS) protection assay [15]. Briefly, 4 µl (6.7 U) of commercially purchased GS (Sigma) was mixed with increasing amounts of *yp*Tpx (1 µg, 2 µg, 4 µg, 6 µg, 8 µg, 10 µg, 15 µg, 20 µg and 50 µg) and 10 µl inactivation solution (50 mM DTT, 25 µM $FeCl_3$), in a final volume of 100 µl (made up in 100 mM HEPES pH 7.4). The mix was incubated for 30 min at room temperature. 2 ml of assay mix (100 mM HEPES, 10 mM KH_2AsO_4, 20 mM NH_2OH, 0.4 mM ADP, 0.5 mM $MnCl_2$, 100 mM glutamine, pH 7.0–7.2) was added to each solution and the incubation continued at 37°C. After 30 min, 1 ml of stop solution (5.5% (w/v) $FeCl_3$, 2% (w/v) TCA, 2.1% (v/v) concentrated HCl) terminated the reaction. Absorbance of the samples was measured at 540 nm.

Protein crystallisation

Purified proteins were dialysed overnight against 20 mM Tris pH 7.5, 50 mM NaCl and kept at a concentration of approximately 8 mg ml^{-1} (based on the absorbance at 280 nm, and a calculated extinction coefficient of 4595 M^{-1} cm^{-1}), for crystallisation studies using crystallisation conditions described previously [16].

Diffraction data collection and structure solution

All diffraction data were collected at Diamond Light Source (Oxfordshire UK), processed with MOSFLM [35] and scaled in SCALA [36], both parts of the CCP4 suite of programs [37,38], or d*TREK [39]. The relevant statistics are published elsewhere

[16]. An improved data set for the oxidised structure of *yp*Tpx was collected and the relevant statistics are found in Table S1.

The structure of *E. coli* Tpx (PDB 3HVV) was used to solve the structures of *yp*Tpx by molecular replacement using PHASER [40] as described previously [16]. Models were refined using REFMAC5 [41] and BUSTER [42], using TLS parameterisation, and inspected, and manipulated when required, in COOT [43], where waters were added. Models were validated in COOT and by the MolProbity server [44]. PDB files were superimposed using LSQMAN [45].

Analytical ultracentrifugation

AUC was carried out in a Beckman Coulter (Palo Alto, CA) Optima XL-I analytical ultracentrifuge. Sedimentation velocity (SV) experiments were performed at 4°C at a rotor speed of 49 k rpm. 360 µl of *yp*Tpx or *yp*TpxC61S in 20 mM Tris pH 7.5, 50 mM NaCl, at four different concentrations between 0.2 and 20 mg ml^{-1}, were loaded into double sector centrepieces. To impose oxidising or reducing conditions, 10 mM H_2O_2 or 5 mM DTT, respectively, were added to the samples. Data were acquired with interference optics; scans were taken every 7 minutes. Data were analysed using SEDFIT [24]. The partial specific volume of *yp*Tpx (0.7407 g ml^{-1}/ 0.7138 g ml^{-1}) and the buffer density (1.00264 g ml^{-1}/ 1.00100 g ml^{-1}) and viscosity (0.015835 P/0.010126 P) at 4°C and 20°C respectively, were calculated using the program SEDNTERP [46]. Sedimentation equilibrium (SE) experiments were performed at 4°C and at rotor speeds of 18 and 24 k rpm. Samples of 80 µl were loaded under the same conditions as for the SV experiments. Scans were taken every 3 h until analysis of the scans with WinMATCH (Jeffrey Lary, University of Connecticut, Storrs, CT, USA) indicated that equilibrium had been reached. SE data were analysed with SEDPHAT [25].

Small angle X-ray scattering (SAXS)

SAXS data were collected on the EMBL ×33 beamline at the DORIS storage ring of the DESY (Deutsches Elektronen Synchrotron) synchrotron (Hamburg, Germany). X-rays were scattered from samples of varying concentrations of *yp*Tpx or *yp*TpxC61S in 20 mM Tris pH 7.5, 100 mM NaCl. Experiments were carried out at 4°C. Data were processed using the program PRIMUS [47]. The distance distribution function and maximum particle dimension (D_{max}) were determined using the program GNOM [27] (Part of ATSAS 2.4 program suite, EMBL Hamburg). *Ab initio* modelling of *yp*Tpx was carried out using the program DAMMIN [28]. Twenty DAMMIN models were generated with an imposed 2-fold (*P2*) symmetry and were merged and averaged using the DAMAVER program suite [48]. The averaged DAMMIN model was superimposed onto the *yp*Tpx dimer crystal structure using SUPCOMB [49]. Theoretical scattering curves of the *yp*Tpx monomer and dimer were generated from the crystal structure coordinates using CRYSOL [50].

Crystal structures were modelled against the solution structure data by rigid body fitting, using BUNCH [29]. In order to ensure maintenance of the correct dimer interface in the reconstruction, distance restraints between interacting interface residues were imposed, namely a maximum distance of 7 Å between D57 and R93, and 4 Å between R110 and G125 (based on data from the high-resolution structure).

Modelling of binding between Tpx and anti-virulence compounds

All modeling was performed using Molecular Operating Environment (MOE) (Chemical Computing Group Inc., Montreal, Canada) software. Modeled ME0052 and ME0055 were subjected to conformational searches for their lowest energy conformations as docking input. The structures of oxidised and reduced *yp*Tpx were subjected to energy minimisation with the MMFF94 force field and the GBSA solvation model prior to docking. MOE Dock (MOE 2010 version 2010.10) was used in the alpha triangle mode; the 25 best poses retained for each compound based on the Affinity dG MOE dock score were further ranked after energy minimisation.

Figures were made using PyMOL (www.pymol.org), ALINE [51], and MOE. All crystal structures have been deposited with the Protein Data Bank (codes presented in Table 1).

Far-western blotting

Samples of *yp*Tpx and *yp*TpxC61S (100 µM) were heated for 10 min at 95°C with LDS loading buffer (NU-PAGE, Invitrogen) and run in MES buffer (NU-PAGE, Invitrogen) on a 4–12% Bis-Tris Novex gel (Invitrogen). The samples were blotted onto a nitrocelluose membrane and the far-western blot was carried out as described previously [11,52]. Briefly, the blot was incubated for 4 h at RT in 5% skimmed milk and was probed with 2 µM ME0052-bio in PBS-Tween for 16 h at 4°C; biotin was detected using HRP-conjugated StreptAvidin (Invitrogen). Band intensities were quantified using ImageJ software (Rasband, http://imagej.nih.gov/ij/, 1997–2011).

Supporting Information

Figure S1 The structure of *yp*TpxC61S superposed onto that for *yp*Tpx in the reduced state with an r.m.s.d. of 0.7 Å indicating that the overall fold of the proteins is highly conserved.
(TIF)

Figure S2 (A) Model for salicylidene acylhydrazide ME0052 (CPK representation) binding to oxidised *yp*Tpx in lowest energy docked conformation (backbone ribbon representation, Connolly surface of sphere radius 1.4 Å). Each subunit of the homodimer is coloured differently (cyan and green ribbon). Spheres on the protein ribbon represent amide groups with largest (blue) and moderate (yellow) chemical shift perturbation as judged by NMR HSQC when ME0052 binds [31]. *Amide sphere for L127 (blue). (B) Alternate binding mode for ME0052 (CPK) to oxidised *yp*Tpx (cyan ribbon) with reduced *yp*Tpx backbone superposed (magenta ribbon). The arrow points to the region of significant backbone and α1 conformational change between oxidised and reduced states.
(TIF)

Table S1 Data collection statistics for the oxidized *yp*Tpx obtained in this study. Values in brackets denote highest resolution shell.
(DOCX)

Acknowledgments

We thank Diamond Light Source for access to beamlines I03 and I04 (proposal numbers MX1229 and MX6638) and the Deutches Elektronen Synchrotron (DESY) for access to beamline ×33 (proposal numbers SAXS-10-172 to SAXS-10-177).

Author Contributions

Conceived and designed the experiments: MG OB AJR. Performed the experiments: CEZ KSHB MG DW. Analyzed the data: SM OB AJR MG VAF REA. Contributed reagents/materials/analysis tools: SM ME. Wrote the paper: MG KSHB VAF OB AJR.

References

1. Cha MK, Kim HK, Kim IH (1995) Thioredoxin-linked "thiol peroxidase" from periplasmic space of *Escherichia coli*. J Biol Chem 270: 28635–28641.
2. Wan XY, Zhou Y, Yan ZY, Wang HL, Hou YD, et al. (1997) Scavengase p20: a novel family of bacterial antioxidant enzymes. FEBS Lett 407: 32–36.
3. Baker LM, Poole LB (2003) Catalytic mechanism of thiol peroxidase from *Escherichia coli*. Sulfenic acid formation and overoxidation of essential CYS61. J Biol Chem 278: 9203–9211.
4. Hall A, Sankaran B, Poole LB, Karplus PA (2009) Structural changes common to catalysis in the Tpx peroxiredoxin subfamily. J Mol Biol 393: 867–881.
5. Choi J, Choi S, Cha MK, Kim IH, Shin W (2003) Crystal structure of *Escherichia coli* thiol peroxidase in the oxidized state: insights into intramolecular disulfide formation and substrate binding in atypical 2-Cys peroxiredoxins. J Biol Chem 278: 49478–49486.
6. Lu J, Yang F, Li Y, Zhang X, Xia B, et al. (2008) Reversible conformational switch revealed by the redox structures of *Bacillus subtilis* thiol peroxidase. Biochem Biophys Res Comm 373: 414–418.
7. Rho BS, Hung LW, Holton JM, Vigil D, Kim SI, et al. (2006) Functional and structural characterization of a thiol peroxidase from *Mycobacterium tuberculosis*. J Mol Biol 361: 850–863.
8. Stehr M, Hecht HJ, Jager T, Flohe L, Singh M (2006) Structure of the inactive variant C60S of *Mycobacterium tuberculosis* thiol peroxidase. Acta Crystallogr D Biol Cryst 62: 563–567.
9. Tao K (2008) Subcellular localization and *in vivo* oxidation-reduction kinetics of thiol peroxidase in *Escherichia coli*. FEMS Microbiol Lett 289: 41–45.
10. Horst SA, Jaeger T, Denkel LA, Rouf SF, Rhen M, et al. (2010) Thiol peroxidase protects *Salmonella enterica* from hydrogen peroxide stress *in vitro* and facilitates intracellular growth. J Bacteriol 192: 2929–2932.
11. Wang D, Zetterstrom CE, Gabrielsen M, Beckham KS, Tree JJ, et al. (2011) Identification of bacterial target proteins for the salicylidene acylhydrazide class of virulence-blocking compounds. J Biol Chem 286: 29922–29931.
12. Baron C (2010) Antivirulence drugs to target bacterial secretion systems. Curr Opin Microbiol 13: 100–105.
13. Hueck CJ (1998) Type III protein secretion systems in bacterial pathogens of animals and plants. Microbiol Mol Biol Rev 62: 379–433.
14. Keyser P, Elofsson M, Rosell S, Wolf-Watz H (2008) Virulence blockers as alternatives to antibiotics: type III secretion inhibitors against Gram-negative bacteria. J Internal Med 264: 17–29.
15. Kim K, Kim IH, Lee KY, Rhee SG, Stadtman ER (1988) The isolation and purification of a specific "protector" protein which inhibits enzyme inactivation by a thiol/Fe(III)/O2 mixed-function oxidation system. J Biol Chem 263: 4704–4711.
16. Gabrielsen M, Zetterstrom CE, Wang D, Beckham KS, Elofsson M, et al. (2010) Expression, purification, crystallization and initial X-ray diffraction analysis of thiol peroxidase from *Yersinia pseudotuberculosis*. Acta Crystallogr F Struct Biol Cryst Commun 66: 1606–1609.
17. Pan JL, Bardwell JC (2006) The origami of thioredoxin-like folds. Protein Sci 15: 2217–2227.
18. Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. J Mol Biol 372: 774–797.
19. Choi J, Choi S, Chon JK, Cha MK, Kim IH, et al. (2005) Crystal structure of the C107S/C112S mutant of yeast nuclear 2-Cys peroxiredoxin. Proteins 61: 1146–1149.
20. Alphey MS, Bond CS, Tetaud E, Fairlamb AH, Hunter WN (2000) The structure of reduced tryparedoxin peroxidase reveals a decamer and insight into reactivity of 2Cys-peroxiredoxins. J Mol Biol 300: 903–916.
21. Wood ZA, Poole LB, Karplus PA (2003) Peroxiredoxin evolution and the regulation of hydrogen peroxide signaling. Science 300: 650–653.
22. Cao Z, Roszak AW, Gourlay LJ, Lindsay JG, Isaacs NW (2005) Bovine mitochondrial peroxiredoxin III forms a two-ring catenane. Structure 13: 1661–1664.
23. Wood ZA, Schroder E, Robin Harris J, Poole LB (2003) Structure, mechanism and regulation of peroxiredoxins. Trends Biochem Sci 28: 32–40.
24. Schuck P (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. Biophys J 78: 1606–1619.
25. Vistica J, Dam J, Balbo A, Yikilmaz E, Mariuzza RA, et al. (2004) Sedimentation equilibrium analysis of protein interactions with global implicit mass conservation constraints and systematic noise decomposition. Anal Biochem 326: 234–256.
26. Putnam CD, Hammel M, Hura GL, Tainer JA (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. Quart Rev Biophys 40: 191–285.
27. Svergun DI (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J App Crystallog 25: 495–503.
28. Svergun DI (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophys J 76: 2879–2886.
29. Petoukhov MV, Svergun DI (2005) Global rigid body modeling of macromolecular complexes against small-angle scattering data. Biophys J 89: 1237–1250.
30. Hayward S, Berendsen HJ (1998) Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme. Proteins 30: 144–154.
31. Wang H, Claveau D, Vaillancourt JP, Roemer T, Meredith TC (2011) High-frequency transposition for determining antibacterial mode of action. Nat Chem Biol 7: 720–729.
32. Karplus PA, Hall A (2007) Structural survey of peroxiredoxins. In: Flohe L, Harris JR, eds. Peroxiredoxin Systems, Structures and Function. New York: Springer. pp 41–60.
33. Nordfelth R, Kauppi AM, Norberg HA, Wolf-Watz H, Elofsson M (2005) Small-molecule inhibitors specifically targeting type III secretion. Infect Immun 73: 3104–3114.
34. Tree JJ, Wang D, McInally C, Mahajan A, Layton A, et al. (2009) Characterization of the effects of salicylidene acylhydrazide compounds on type III secretion in *Escherichia coli* O157:H7. Infect Immun 77: 4209–4220.
35. Leslie AGW (1992) Recent changes to the MOSFLM package for processing film and image plate data. Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography 26.
36. Evans P (2006) Scaling and assessment of data quality. Acta Crystallogr D Biol Crystallogr 62: 72–82.
37. Collaborative Computational Project N (1994) The CCP4 suite: programs for protein prystallography. Acta Crystallogr D Biol Crystallogr 50: 760–763.
38. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, et al. (2011) Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr 67: 235–242.
39. Pflugrath JW (1999) The finer things in X-ray diffraction data collection. Acta Crystallogr D Biol Crystallogr 55: 1718–1725.
40. Mccoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, et al. (2007) Phaser crystallographic software. J App Crystallogr 40: 658–674.
41. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr 53: 240–255.
42. Bricogne G, Blanc E, Brandl M, Flensburg C, Keller P, et al. (2011) BUSTER version 2.8.0. Cambridge, UK: Global Phasing Ltd.
43. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. Acta Crystallogr D Biol Crystallogr 66: 486–501.
44. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, et al. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res 35: W375–383.
45. Kleywegt G, Jones TA (1994) A super position. CCP4/ESF-EACBM Newsletter on Protein Crystallography 31: 9–14.
46. Laue TM, Shah BD, Ridgeway TM, Peletier SL (1992) Computer-aided interpretation of analytical sedimentation data for proteins. In: Harding SE, Rowe AJ, Horton JC, eds. Analytical Ultracentrifugation in Biochemistry and Polymer Science. Cambridge: Society of Chemistry. pp 90–125.
47. Svergun DI, Konarev PV, Volkov VV, Sokolova AV, Koch MHJ (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. J App Crystallog 36: 1277–1282.
48. Volkov VV, Svergun DI (2003) Uniqueness of *ab initio* shape determination in small-angle scattering. J App Crystallog 36: 860–864.
49. Kozin MB, Svergun DI (2001) Automated matching of high- and low-resolution structural models. J App Crystallog 34: 33–41.
50. Svergun DI, Barberato C, Koch MHJ (1995) CRYSOL - a program to evaluate X-ray solution scattering of biological macromoleculse from atomic coordinates. J App Crystallog 28: 768–773.
51. Bond CS, Schüttelkopf AW (2009) ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. Acta Crystallogr D Biol Crystallogr 65: 510–512.
52. Wu Y, Li Q, Chen XZ (2007) Detecting protein-protein interactions by far western blotting. Nat Protoc 2: 3278–3284.
53. Barton GJ (1993) ALSCRIPT a tool to format multiple sequence alignments. Protein Eng 6: 37–44.

PUBLICATIONS

FEBS Letters 586 (2012) 1160–1165

journal homepage: www.FEBSLetters.org

FolX from *Pseudomonas aeruginosa* is octameric in both crystal and solution

Mads Gabrielsen [a,1], Katherine S.H. Beckham [a,1], Richard J. Cogdell [b], Olwyn Byron [c], Andrew J. Roe [a,*]

[a] Institute of Infection, Immunity and Immunology, College of Medical, Veterinary and Life Sciences, Sir Graeme Davies Building, University of Glasgow, Glasgow G12 8QQ, UK
[b] Institute of Molecular, Cell and Systems Biology, College of Medical, Veterinary and Life Sciences, Sir Graeme Davies Building, University of Glasgow, Glasgow G12 8QQ, UK
[c] School of Life Sciences, College of Medical, Veterinary and Life Sciences, Sir Graeme Davies Building, University of Glasgow, Glasgow G12 8QQ, UK

ARTICLE INFO

Article history:
Received 9 February 2012
Revised 5 March 2012
Accepted 14 March 2012
Available online 24 March 2012

Edited by Miguel De la Rosa

Keywords:
Pteridine biosynthesis
X-ray structure
Solution structure

ABSTRACT

FolX encodes an epimerase that forms one step of the tetrahydrofolate biosynthetic pathway, which is of interest as it is an established target for important drugs. Here we report the crystal structure of FolX from the bacterial opportunistic pathogen *Pseudomonas aeruginosa*, as well as a detailed analysis of the protein in solution, using analytical ultracentrifugation (AUC) and small-angle X-ray scattering (SAXS). In combination, these techniques confirm that the protein is an octamer both in the crystal structure, and in solution.

Structured summary of protein interactions:
FolX and **FolX** bind by x-ray crystallography (View interaction)
FolX and **FolX** bind by cosedimentation in solution (View interaction)
FolX and **FolX** bind by x ray scattering (View interaction)

© 2012 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

1. Introduction

The tetrahydrofolate biosynthetic pathway is an established target for important antibiotics including trimethoprim that inhibits the activity of bacterial dihydrofolate reductase [1–3]. However, with the ever-increasing spread of antibiotic resistance in bacteria, there is an urgent need to explore alternative approaches to controlling pathogens. Recent work has elucidated the target proteins of one group of so-called "anti-virulence" compounds, the salicylidene acylhydrazides. In this published study [4], several target proteins were identified including FolX, encoded by a gene only carried by the gammaproteobacteria. FolX has been shown to act as an epimerase (Fig. 1) in conjunction with FolM, a reductase, during the conversion of dihydroneopterin triphosphate to tetrahydrobiopterin [5]. Therefore, FolX and FolM are essential for tetrahydromonapterin synthesis in species such as *Escherichia coli* and *Pseudomonas aeruginosa*.

Here we report the crystal structure of FolX from *P. aeruginosa* as well as a detailed analysis of the protein in solution using analytical ultracentrifugation (AUC) and small-angle X-ray scattering (SAXS). The combination of these techniques confirms that the octameric crystal structure is consistent with the biological state of FolX in solution.

2. Materials and methods

2.1. Protein expression and purification

The gene encoding FolX was amplified from *P. aeruginosa* genomic DNA and cloned into the vector pET-151 (Invitrogen) [4]. The expression and purification was performed following previously published methods [6].

2.2. Crystallisation

Purified FolX was dialysed against 20 mM Tris pH 7.5, 50 mM NaCl and concentrated to approximately 6 mg ml^{-1}, based on the absorbance at 280 nm (A_{280}) and an extinction coefficient of 5960 M cm^{-1} derived from the sequence composition. Screens were set up using commercially available crystallisation kits, using vapour diffusion, with drops consisting of 500 nl protein solution and 500 nl reservoir. Cubes (0.2 × 0.2 × 0.2 mm) appeared in conditions, containing 40% (v/v) 1,2-propanediol, 100 mM HEPES pH 7.5, within a week, at room temperature.

Abbreviations: FolX, 7,8-dihydroneopterin-triphosphate-epimerase; AUC, analytical ultracentrifugation; SAXS, small-angle X-ray scattering; SV, sedimentation velocity; SE, sedimentation equilibrium

* Corresponding author. Address: Sir Graeme Davies Building, 120 University Place, University of Glasgow, Glasgow G12 8QQ, UK. Fax: +44 141 330 4600.
E-mail address: Andrew.roe@glasgow.ac.uk (A.J. Roe).
URL: http://www.gla.ac.uk:443/ibls/staff/staff.php?who=Pee-Pd (A.J. Roe).

[1] These authors contributed equally to this work.

0014-5793/$36.00 © 2012 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.febslet.2012.03.031

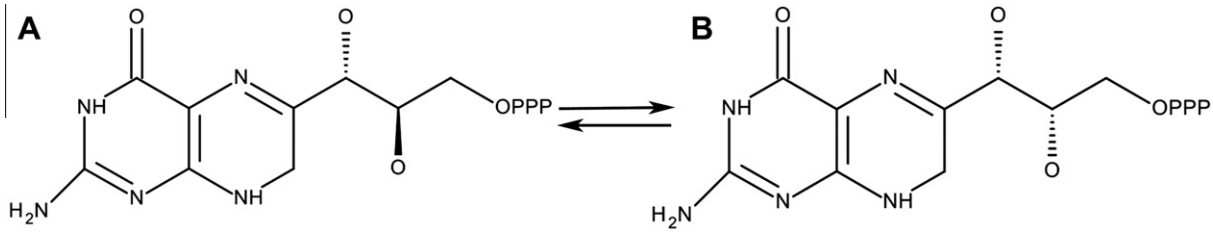


Fig. 1. Chemdraw representation of the epimerase reaction catalysed by FolX, where dihydroneopterin is converted to dihydromonapterin triphosphate.

2.3. Data collection, processing and structure solution

Crystals were flash-frozen in liquid nitrogen, with no further cryo-protection, and brought to Diamond Light Source, station I03. Data were collected on a PILATES 6M detector, at a wavelength of 0.97625 Å. A total of 127° of data were collected using an increment of 0.15°. Data were processed using MOSFLM [7] and scaled in SCALA [8]. The structure was determined using the Balbes molecular replacement server [9], which identified PDB entry 1B9L [10] as the best search model. The solved structure was refined with BUSTER [11], using TLS parameterisation and torsion restraints from the search model, and inspected and altered when required, using COOT [12]. Waters were added using BUSTER. The geometry of the structure was validated by MOLPROBITY [13]. The data showed signs of anisotropy and were processed to 3.0 Å, based on the statistical factors presented in Table 1, although diffraction spots could be detected closer to 2.5 Å.

2.4. Analytical ultracentrifugation

Purified FolX was dialysed against 20 mM Tris pH 7.5, 150 mM NaCl and concentrated to approximately 10 mg ml^{-1}, based on the A_{280}. Analytical ultracentrifugation (AUC) was carried out in a Beckman Coulter (Palo Alto, CA) Optima XL-I analytical ultracentrifuge. Sedimentation velocity (SV) experiments were performed at 4 °C at a rotor speed of 49 krpm. 360 µl of sample, at concentrations of FolX ranging between 0.2 and 10 mg ml^{-1}, were loaded into double sector centrepieces. Data were acquired every 7 min with interference and absorbance optics and were subsequently analysed using SEDFIT [14]. The partial specific volume of FolX (0.737/0.743 g ml^{-1}), the buffer density (1.00677/1.00499 g ml^{-1}) and viscosity (0.0156/0.0102 P) at 4 °C and 20 °C respectively, were all calculated using the program SEDNTERP [15]. Sedimentation equilibrium (SE) experiments were carried out with the same range of FolX concentrations using 90 µl of sample with a rotor speed of 23 krpm. Scans were taken every 3 h until analysis of the scans, using WinMATCH (Jeffrey Lary, University of Connecticut, Storrs, CT, USA), indicated that equilibrium had been reached. SE data were analysed using Origin and were fitted with a tetramer–octamer model with non-ideality. The dissociation constant was calculated using the method described by Solovyova *et al.* [16].

Table 1
Data collection, refinement and model quality statistics for FolX. Values in brackets represent the values of the highest resolution shell. R.m.s.d. is root-mean square deviation.

PDB code	4AEY
Space group	*I*432
Unit Cell (Å)	*a* = 133.98
Resolution (Å)	94.74–3.00 (3.16–3.00)
Observed reflections	103928
Unique reflections	4383
Multiplicity	23.7 (25.1)
Completeness (%)	100.0 (100.0)
R_{meas} (%)	5.1 (141.6)
R_{pim} (%)	1.1 (38.4)
$I/\sigma I$	34.3 (2.9)
Wilson B (Å^2)	115.3
Protein residues/atoms	107/754
Water molecules	8
R_{work} (%)	28.3
R_{free} (%)	31.1
R.m.s.d. for bond lengths (Å)/angles (°)	0.01/1.49
Molprobity score	13.9 [97th percentile]

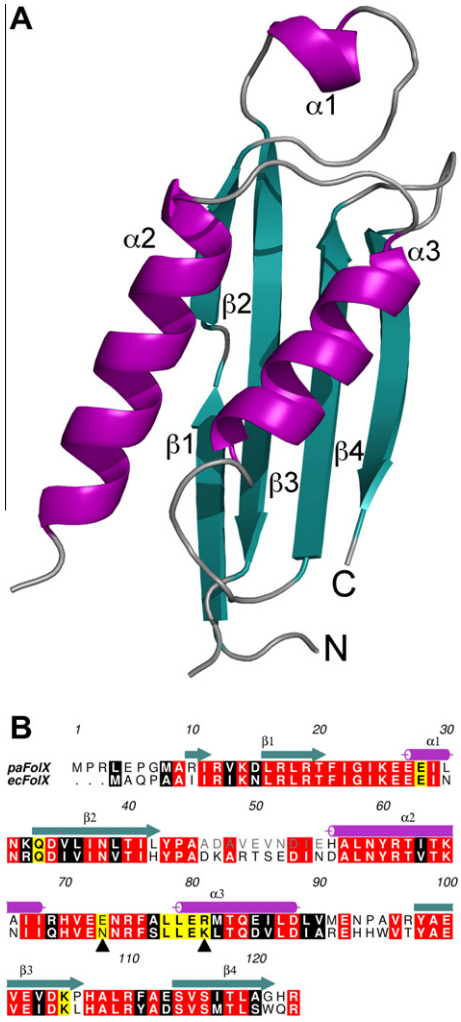


Fig. 2. (A) Cartoon representation of a subunit of FolX. The helices are coloured in purple and the strands are coloured in teal. The N- and C-termini have been labelled. (B) Structural alignment of FolX from *P. aeruginosa* and *E. coli*, with secondary structure elements coloured as above. Highlighted in yellow are the residues involved in the putative active site. Residues in grey indicate residues not modelled in the electron density.

PUBLICATIONS

2.5. Small angle X-ray scattering measurements

SAXS data were collected at the ESRF ID14 EH3 beamline for samples with a range of FolX concentrations of between 0.5 and 6 mg ml^{-1} in 20 mM Tris pH 7.5, 150 mM NaCl. No concentration dependence effects were observed, therefore data from the highest concentration (6 mg ml^{-1}) were processed for further analysis. Initial processing of the data was done using PRIMUS and p(r) analysis was carried out using GNOM [17]. The resolution of the data was calculated from the highest angle at which useable scattering data were recorded [16] *Ab initio* models of FolX were generated from the experimental data using DAMMIF [18]. Twenty DAMMIF models were superimposed and averaged using DAMAVER [19] and the averaged model was superimposed onto the crystal structure using SUPCOMB [20].

Figures were made using ALINE [21] and PyMOL [22].

3. Results and discussion

3.1. Monomeric structure

The monomeric structure of FolX comprised a four-stranded antiparallel sheet, composed of β1 (residues 10–12 and 16–20), β2 (residues 33–42), β3 (residues 98–106) and β4 (residues 114–121) (Fig. 2A). The broken β1 strand, caused by a small kink introduced in the strand, was also observed in the structure of FolX from *E. coli* [10]. A short α-helix (α1, 4 residues) is located in the loop between sheets β1 and β2, whereas α -helices α2 and α3 are nestled against the sheet on the concave side. The structure is not complete, as only residues 6–46 and 55–122 can be observed in the electron density. This disordered region exhibits the greatest sequence disparity compared with *E. coli* FolX (Fig. 2B) although it is located opposite to the active site and is therefore unlikely to affect overall function. There are a number of side chains that cannot be observed in the electron density, and have, accordingly, been cut back to the last ordered atom.

The sequence and structure of FolX are highly conserved throughout the gammaproteobacteria, and the sequence identity and similarity between FolX from *P. aeruginosa* and *E. coli* is 60% and 78% respectively (based on a level of 0.7, using the ALSCRIPT algorithm [23]) (Fig. 2B). The structures of the two homologues superpose with a root-mean square deviation (r.m.s.d.) of 0.99 Å for 100 Cα. The main differences are mostly limited to the loop between α2 and β3. These differences are not localised near the interfaces, or the putative active site.

3.2. Quaternary structure

FolX crystallised in space group *I*432, with a single chain in the asymmetric unit. The Protein Interfaces, Surfaces and Assemblies (PDBePISA) server [24] suggests that the oligomeric state, based

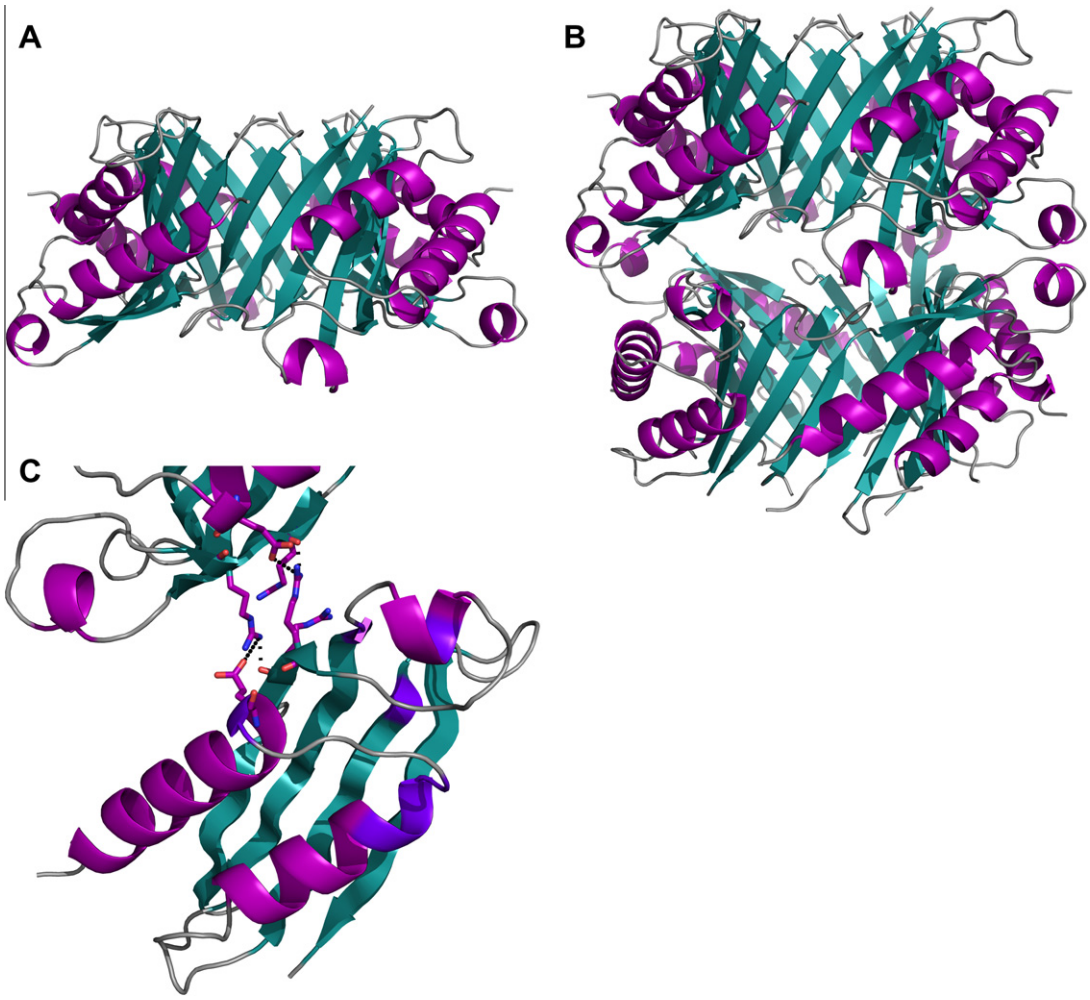


Fig. 3. Oligomerisation of FolX. (A) The tetramer ring formed by FolX dimerises to form an octamer (B). (C) The interface between the tetrameric rings is stabilised by four hydrogen bonds (shown as dashed lines) between residues R19, R17 and E72 (shown in sticks). Active site residues are highlighted in purple.

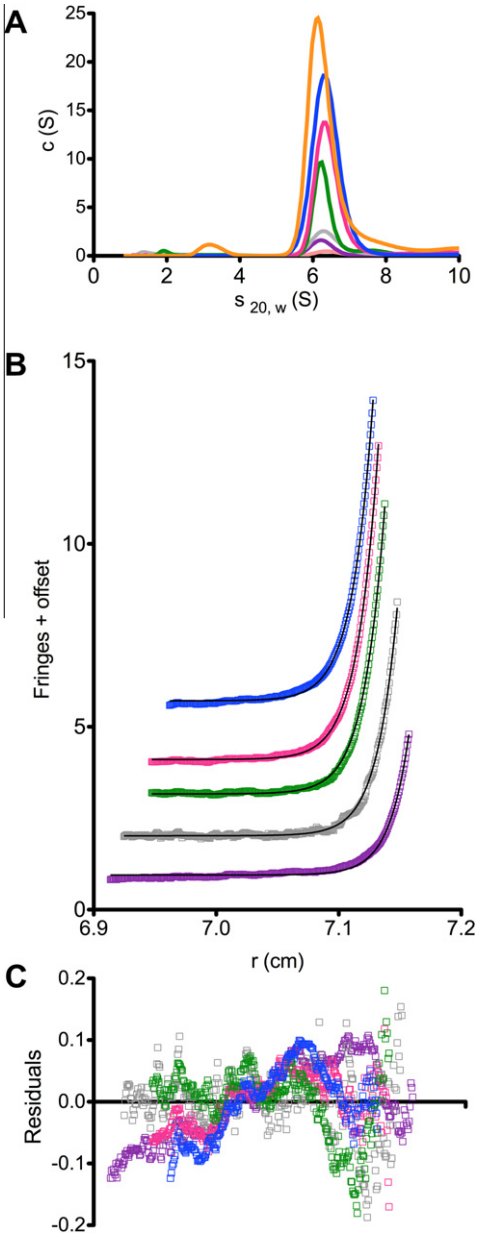


Fig. 4. AUC analysis of FolX. (A) c(S) distributions derived via SEDFIT from SV data for varying concentrations of FolX are dominated by a peak at $s_{20,w} \approx 4$ S. Different concentrations are shown in different colours: 0.2 mg ml^{-1}, light pink; 0.5 mg ml^{-1}, purple; 1 mg ml^{-1}, grey; 2.5 mg ml^{-1}, green; 5 mg ml^{-1}, pink, 7.5 mg ml^{-1}, blue; 10 mg ml^{-1}, orange. (B and C) The global fit to SE data using a non-ideal tetramer–octamer self-association model. The experimental data and fits are shown in (B); the residuals for each fit are plotted in (C). Different protein concentrations are represented by the colour scheme used in (A).

on the crystallographic symmetry, is tetrameric (Fig. 3A), consisting of a circle made up of the convex sides of the sheets of the subunits facing each other. The interfaces between the subunits are made up by strands β4 on one subunit and β1 on the next in the circle. The contacts comprise 20 residues on each strand, accounting for ~18% of the total solvent accessible area of the tetramer. The interface is made up by hydrogen bonds formed between a number of residues, mostly involving main chain nitrogen and carboxyl groups (Supplementary Table 1).

Previous studies have suggested that FolX is an octamer [10,25]. However, when the crystal structure is analysed by PDBePISA, it is predicted to be a tetramer. The only interactions between the tetramers, in what would be a dimer of tetramers (Fig. 3B), involve 7% of the solvent accessible area compared with 18% involved in stabilising the tetramer itself. Analysis of the interface between the tetrameric rings, using PDBePISA, indicates that there are 16 hydrogen bonds and 8 salt bridges connecting the two tetramers (Supplementary Table 2), which may suggest that this interaction is significant despite the relatively small interface area.

In order to try to get conclusive evidence for the quaternary state of FolX, analytical ultracentrifugation was performed. Sedimentation velocity experiments revealed that FolX was present as a single species in solution, as evidenced by a single peak in the concentration distribution of the apparent $s_{20,w}$ (c(s), Fig. 4A). The infinite dilution sedimentation coefficient ($s^0_{20,w}$) of FolX, derived from the concentration dependence of $s_{20,w}$, determined by fitting the data with a non-interacting discrete species model, is 6.09 ± 0.03 S. This corresponds with the value of $s^0_{20,w}$ computed (using SOMO [26]) for the octamer crystal structure (5.97 S) and not with that computed for the tetramer (3.62 S). Sedimentation equilibrium data fitted with a single species model indicated the presence of a species with a mass of 141,500 ± 6659 Da at infinite dilution (M^0). This value is slightly lower than the calculated octamer mass of 143,864 Da, therefore, in order to improve the fit of the model parameters, the effects of non-ideality and the presence of a tetramer–octamer equilibrium were introduced into the data analysis. Addition of non-ideality improved the χ^2 of the global fit from 0.01890 to 0.00316. Extending the model to include a tetramer–octamer equilibrium further improved the fit to a χ^2 of 0.00297 and gave a K_d of 0.887 μM. The fit to the data along with the resultant residuals is shown in Fig. 4B and C.

3.3. Active site

Ploom et al. [10] suggested a putative active site for FolX from *E. coli* involving residues that are mostly conserved between *E. coli* and *P. aeruginosa* with only 2 substitutions (Asn to Glu and Lys to Arg, respectively) (Fig. 2B). Most of these residues are disordered and not present in the electron density, implying that the active site of *P. aeruginosa* FolX is flexible in the absence of substrate, or exhibits multiple conformations. The active site does not appear to be affected by the formation of the octamer (Fig. 3C).

3.4. SAXS structure

To further confirm the oligomeric state of FolX, the solution structure was determined using SAXS. Indirect Fourier transformation of the data using GNOM [17] indicated a D_{max} of 177 Å and an Rg of 44.11 ± 0.6 Å. An *ab initio* model of FolX was generated using DAMMIF [18], imposing *P*4 symmetry, based on the crystal structure, and the fit of the final averaged model to the experimental data is shown in Fig. 5A. We have determined the K_d of FolX to be 0.887 μM, therefore at the protein concentration used in the SAXS study 99.5% (by mass) of the protein would have been fully octamerised. The crystal structure of the octamer was superposed onto the 10.5 Å resolution envelope of FolX in solution, (Fig. 5B). The octamer crystal structure fits well into the envelope confirming that this is the true solution oligomeric state.

Here we present the structure of FolX from *P. aeruginosa*, which is an octamer in both the crystal and in solution (in equilibrium with its tetrameric form). Despite only 7% of the surface area of the two tetramers being involved in the interface these interactions are stabilised by 16 hydrogen bonds and 8 salt bridges. Consequently the association of the two tetrameric rings is strong, with a K_d of 0.887 μM indicating a high affinity association.

The active site of FolX is predicted to comprise residues from two adjacent subunits, which suggests that the tetramer is essential for the activity of the enzyme. The functional need for an

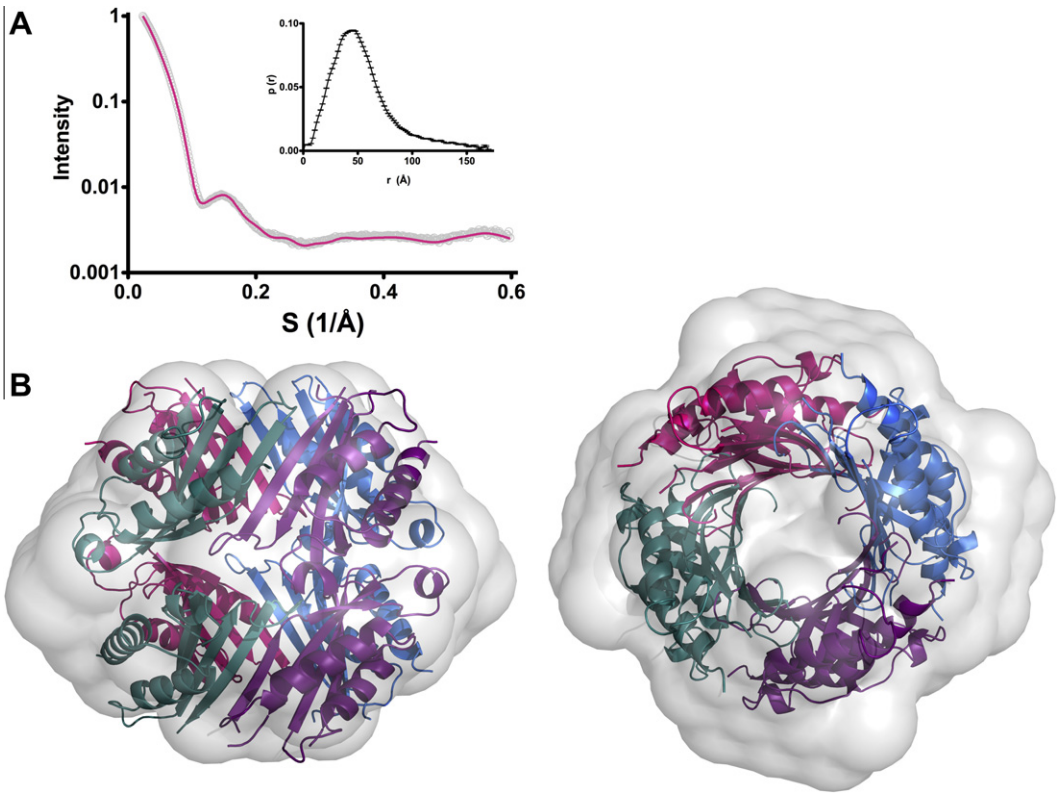


Fig. 5. Solution structure of FolX. (A) The experimental scattering data (grey) with the fit of the DAMMIF model to the data shown in pink. Inset is the pairwise distribution p(r) function of the data. (B) The octamer crystal structure superimposed onto the DAMMIF model shown as side and top view.

octameric form is not apparent from the previously published biochemical data. We speculate that the formation of the octamer may play a role in the stability of FolX, which could be tested by examining the half-life of the protein.

Acknowledgements

The work was supported by a grant from the Biotechnology and Biological Sciences Research Council to M.G., R.J.C. and A.J.R. (BB/G011389/1), a Medical Research Scotland Grant (223 ORG) to A.J.R., and a Wellcome Trust Studentship for K.S.H.B. We thank Diamond Light Source for access to beamline I03 (proposal number MX6638) and the European Synchrotron Radiation Facilities (ESRF) station ID14-EH3 (proposal number MX1254).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.febslet.2012.03.031.

References

[1] Achari, A., Somers, D.O., Champness, J.N., Bryant, P.K., Rosemond, J. and Stammers, D.K. (1997) Crystal structure of the anti-bacterial sulfonamide drug target dihydropteroate synthase. Nat. Struct. Biol. 4, 490–497.
[2] Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M., et al. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269, 496–512.
[3] Spitzy, K.H. (1991) Use of antibacterial chemotherapy. A historical comparison. Infection 19 (Suppl. 1), S52–56.
[4] Wang, D., Zetterstrom, C.E., Gabrielsen, M., Beckham, K.S., Tree, J.J., Macdonald, S.E., Byron, O., Mitchell, T.J., Gally, D.L., Herzyk, P., et al. (2011) Identification of bacterial target proteins for the salicylidene acylhydrazide class of virulence-blocking compounds. J. Biol. Chem. 286, 29922–29931.
[5] Pribat, A., Blaby, I.K., Lara-Nunez, A., Gregory 3rd, J.F., de Crecy-Lagard, V. and Hanson, A.D. (2010) FolX and FolM are essential for tetrahydromonapterin synthesis in *Escherichia coli* and *Pseudomonas aeruginosa*. J. Bacteriol. 192, 475–482.
[6] Gabrielsen, M., Zetterstrom, C.E., Wang, D., Beckham, K.S., Elofsson, M., Isaacs, N.W. and Roe, A.J. (2010) Expression, purification, crystallization and initial X-ray diffraction analysis of thiol peroxidase from *Yersinia pseudotuberculosis*. Acta Crystallogr., Sect. F Struct. Biol. Cryst. Commun. 66, 1606–1609.
[7] Leslie, A.G. (2006) The integration of macromolecular diffraction data. Acta Crystallogr. D. Biol. Crystallogr. 62, 48–57.
[8] Evans, P. (2006) Scaling and assessment of data quality. Acta Crystallogr., D Biol. Crystallogr. 62, 72–82.
[9] Long, F., Vagin, A.A., Young, P. and Murshudov, G.N. (2008) BALBES: a molecular-replacement pipeline. Acta Crystallogr., D Biol. Crystallogr. 64, 125–132.
[10] Ploom, T., Haussmann, C., Hof, P., Steinbacher, S., Bacher, A., Richardson, J. and Huber, R. (1999) Crystal structure of 7,8-dihydroneopterin triphosphate epimerase. Structure 7, 509–516.
[11] Bricogne G, Blanc E, Brandl M, Flensburg C, Keller P, Paciorek W, Roversi P, Sharff A, Vonrhein C & Womack TO (2011) BUSTER version 2.8.0. In. Global Phasing Ltd, Cambridge, UK.
[12] Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. Acta Crystallogr., D Biol. Crystallogr. 66, 486–501.
[13] Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall 3rd, W.B., Snoeyink, J., Richardson, J.S., et al. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res. 35, W375–383.
[14] Schuck, P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. Biophys. J. 78, 1606–1619.
[15] Laue, T.M., Shah, B.D., R Ridgeway, T.M. and Peletier, S.L. (1992) Computer-aided interpretation of analytical sedimentation data for proteins in: Analytical Ultracentrifugation in Biochemistry and Polymer Science (Harding, S.E.R.A. and Horton, J.C., Eds.), pp. 99–125, Society of Chemistry, Cambridge.
[16] Solovyova, A.S., Nöllmann, M., Mitchell, T.J. and Byron, O. (2004) The solution structure and oligomerization behavior of two bacterial toxins: pneumolysin and perfringolysin O. Biophys. J. 87, 540–552.
[17] Svergun, D.I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J. Appl. Crystallogr. 25, 495–503.
[18] Franke, D. and Svergun, D.I. (2009) DAMMIF, a program for rapid *ab initio* shape determination in small-angle scattering. J. Appl. Crystallogr. 42, 342–346.

[19] Volkov, V.V. and Svergun, D.I. (2003) Uniqueness of *ab initio* shape determination in small-angle scattering. J. Appl. Crystallogr. 36, 860–864.
[20] Kozin, M.B. and Svergun, D.I. (2001) Automated matching of high- and low-resolution structural models. J. Appl. Crystallogr. 34, 33–41.
[21] Bond, C.S. and Schüttelkopf, A.W. (2009) ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. Acta Crystallogr., D Biol. Crystallogr. 65, 510–512.
[22] DeLano (2009) The PYMOL molecular graphics system, DeLano Scientific, Palo Alto, CA.
[23] Barton, G.J. (1993) ALSCRIPT a tool to format multiple sequence alignments. Protein Engineer. 6, 37–44.
[24] Krissinel, E. and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. 372, 774–797.
[25] Haussmann, C., Rohdich, F., Schmidt, E., Bacher, A. and Richter, G. (1998) Biosynthesis of pteridines in *Escherichia coli*. Structural and mechanistic similarity of dihydroneopterin-triphosphate epimerase and dihydroneopterin aldolase. J. Biol. Chem. 273, 17418–17424.
[26] Rai, N., Nöllmann, M., Spotorno, B., Tassara, G., Byron, O. and Rocco, M. (2005) SOMO(SOlution MOdeler): differences between X-ray- and NMR-derived bead models suggest a role for side chain flexibility in protein hydrodynamics. Structure 13, 723–734.

structural communications

Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

The structure of an orthorhombic crystal form of a 'forced reduced' thiol peroxidase reveals lattice formation aided by the presence of the affinity tag

Katherine S. H. Beckham,[a] Olwyn Byron,[b] Andrew J. Roe[a]* and Mads Gabrielsen[a]*

[a]Institute of Infection, Immunity and Inflammation, College of Medical, Veterinary and Life Sciences, University of Glasgow, Sir Graeme Davies Building, Glasgow G12 8QQ, Scotland, and [b]School of Life Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Sir Graeme Davies Building, Glasgow G12 8QQ, Scotland

Correspondence e-mail: Andrew.Roe@glasgow.ac.uk, mgabr@chem.gla.uk

Received 9 February 2012
Accepted 16 March 2012

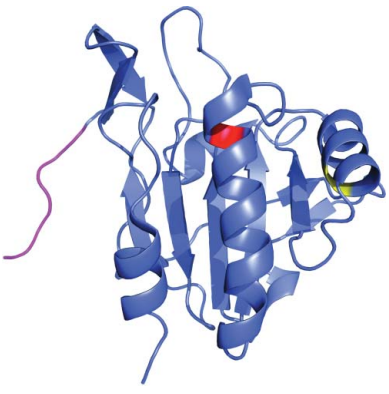
PDB Reference: C61S mutant of Tpx, 4af2.

Thiol peroxidase (Tpx) is an atypical 2-Cys peroxiredoxin, which has been suggested to be important for cell survival and virulence in Gram-negative pathogens. The structure of a catalytically inactive version of this protein in an orthorhombic crystal form has been determined by molecular replacement. Structural alignments revealed that Tpx is conserved. Analysis of the crystal packing shows that the linker region of the affinity tag is important for formation of the crystal lattice.

1. Introduction

Thiol peroxidase (Tpx) is an atypical 2-Cys peroxiredoxin found in both Gram-negative and Gram-positive bacteria (Cha *et al.*, 1995). Tpx is a redox-active protein that reduces alkyl peroxides and hydrogen peroxide through the catalytic recycling of a disulfide bond between cysteines 61 and 95. This cycle results in the production of water, with reduced Tpx being regenerated through interaction with thioredoxin (Baker & Poole, 2003). Tpx makes a major contribution to redox homeostasis in the bacterial cell. For example, *Escherichia coli* Tpx deletion mutants displayed reduced survival after exposure to peroxide (Cha *et al.*, 1995). Furthermore, this protein has been shown to be important for the survival of human pathogens such as *Salmonella typhimurium*, in which the ability to withstand the oxidative burst in the phagosome is essential (Horst *et al.*, 2010). Tpx has also been implicated in the virulence of Gram-negative pathogens as it has been shown to be one of the target proteins of the salicylidene acylhydrazides (Wang *et al.*, 2011). These so-called 'anti-virulence' compounds inhibit the bacterial type 3 secretion system, which is used by pathogens to modulate host-cell pathways and facilitate disease (Baron, 2010).

Here, we describe the crystallization and structure analysis of the catalytically inactive mutant TpxC61S from *E. coli*, which represents a 'forced reduced' structure as described by Hall *et al.* (2009). The crystals diffracted to 1.97 Å resolution and exhibited an orthorhombic form. A brief comparison of this structure with a number of structures of Tpx present in the Protein Data Bank, including a trigonal crystal form of the same mutant from *E. coli* (Hall *et al.*, 2009), is presented. Analysis of the crystal packing of the two TpxC61S structures reveals the importance of parts of the affinity tag in forming crystal contacts in the orthorhombic space group.



2. Materials and methods

2.1. Cloning, expression and purification

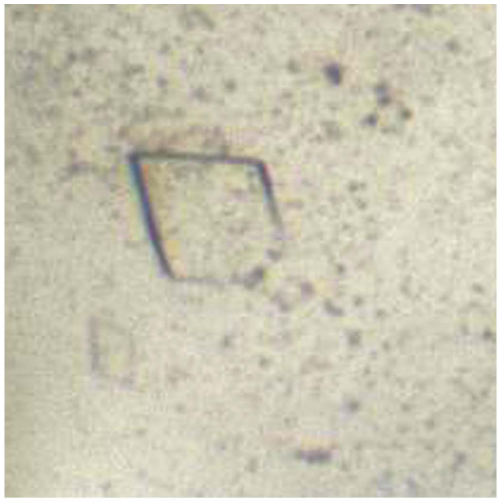
The C61S mutant of Tpx was obtained by site-directed mutagenesis utilizing a QuikChange site-directed mutagenesis kit (Stratagene) with the primer pair ecTpx5/3 (CGTACTGATGCGGCCGAAACACCGGTATC and GATACCGGTGTTTCGGCCGCATCAGTACG); the construct was confirmed by DNA sequencing. The amplified product was cloned into the TOPO pET-151 (Invitrogen) expression vector, which encodes an N-terminal affinity tag consisting of a hexahistidine sequence motif with a TEV cleavage site and a linker region consisting of 25 residues. The resulting construct was transformed into *E. coli* BL21 DE3 (λD3) cells and grown in 1 l LB

© 2012 International Union of Crystallography
All rights reserved

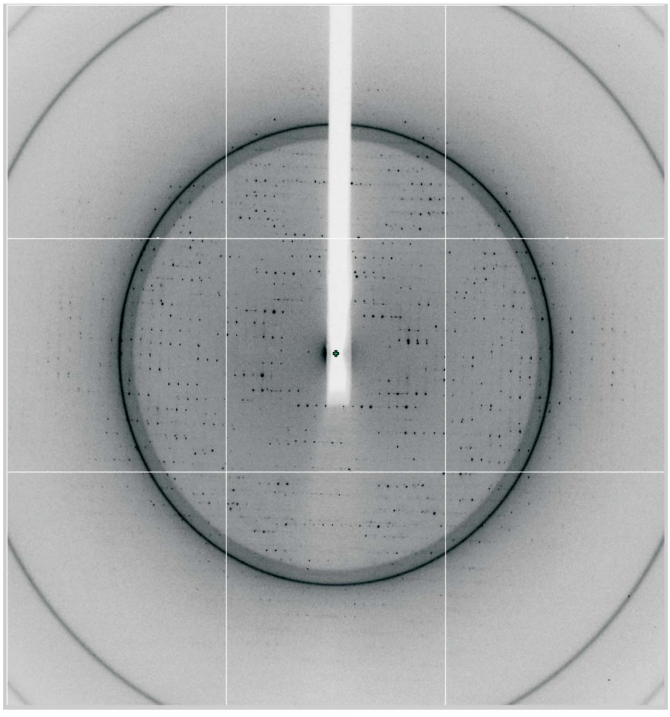
medium containing ampicillin (100 µg ml^{-1}). The protein was purified using immobilized metal-affinity chromatography protocols as described elsewhere (Gabrielsen *et al.*, 2010) and dialysed against 20 m*M* Tris pH 7.5, 50 m*M* NaCl.

2.2. Crystallization

Purified protein with affinity tag at approximately 8 mg ml^{-1} was screened by sitting-drop vapour diffusion using several commercial crystallization screens with 1 µl drops consisting of a 1:1 ratio of protein and reservoir solutions. The crystal trays were incubated at 293 K. Crystals appeared in the JCSG+ Suite (Molecular Dimensions Ltd) after two weeks against a reservoir consisting of 0.2 *M* $MgCl_2$, 0.1 *M* Tris pH 7, 10% polyethylene glycol 8000. The crystal was cryoprotected by a short soak in paraffin oil (Riboldi-Tunnicliffe & Hilgenfeld, 1999) before being flash-cooled in liquid nitrogen.



(*a*)



(*b*)

Figure 1
E. coli TpxC61S crystals and diffraction. (*a*) The orthrombic crystals have dimensions of 0.1 × 0.1 × 0.3 mm. (*b*) The crystals diffracted to 1.97 Å resolution (at the detector edge) and belonged to space group $C222_1$.

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

PDB code	4a2f
Space group	$C222_1$
Unit-cell parameters (Å)	a = 49.37, b = 71.75, c = 121.93
Resolution (Å)	40.67–1.97 (2.08–1.97)
Observed reflections	96980
Unique reflections	15679
Multiplicity	6.2 (6.6)
Completeness (%)	90.2 (100)
R_{meas}† (%)	14.2 (67.5)
$R_{p.i.m.}$‡ (%)	5.7 (26.0)
$\langle I/\sigma(I)\rangle$	7.5 (2.8)
Wilson B factor (Å^2)	19.9
Matthews coefficient (Å^3 Da^{-1})	3.09
Solvent content (%)	60.16
Water O atoms	170
R_{work} (%)	22.93
R_{free} (%)	28.28
R.m.s.d. for bond lengths (Å)/angles (°)	0.009/1.17
Average isotropic thermal parameters (Å^2)	
Main chain	35.34
Side chain	40.95
Water O atoms	37.53
Ramachandran outliers	0
MolProbity clash score	6.57 [94th percentile]

† $R_{meas} = \sum_{hkl}\{N(hkl)/[N(hkl)-1]\}^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$.
‡ $R_{p.i.m.} = \sum_{hkl}\{1/[N(hkl)-1]\}^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$.

2.3. X-ray data measurements, structure solution and refinement

Diffraction data were collected using an ADSC Q315 CCD detector on Diamond Light Source beamline I04 at a wavelength of 0.9763 Å. Data were processed using *MOSFLM* (Leslie, 1992) and scaled and merged using *SCALA* (Evans, 2006) from the *CCP*4 suite (Winn *et al.*, 2011). The structure of TpxC61S was determined using *Phaser*; the existing TpxC61S structure from *E. coli* (PDB entry 3hvv; Hall *et al.*, 2009) was used as a search model for molecular replacement. The structure was refined using *REFMAC*5 (Murshudov *et al.*, 2011) and *BUSTER* (Bricogne *et al.*, 2011). The model was manipulated as required and waters were added using *Coot* (Emsley *et al.*, 2010). The model was validated using *Coot* and *MolProbity* (Chen *et al.*, 2010). *PISA* was used to analyze protein interfaces and crystal contacts (Krissinel & Henrick, 2007) and structural superpositions were performed using the SUPER command in *PyMOL* (Schrödinger; http://www.pymol.org). Figures were produced using *ALINE* (Bond & Schüttelkopf, 2009) and *PyMOL*.

The structure and data were deposited in the Protein Data Bank (Velankar *et al.*, 2012; accession code 4af2).

3. Results and discussion

3.1. Crystallization, data collection and structure determination

The crystals of the C61S mutant of Tpx from *E. coli* were orthorhombic, with dimensions of approximately 0.1 × 0.1 × 0.03 mm (Fig. 1*a*), and belonged to space group $C222_1$, with unit-cell parameters a = 49.37, b = 71.75, c = 121.93 Å. Data were collected to a resolution of 1.97 Å (Fig. 1*b*), and the relevant data-collection statistics are presented in Table 1. The asymmetric unit comprised one subunit with a Matthews coefficient of 3.09 Å^3 Da^{-1}, indicating a solvent content of 60%. The data processed well, with overall R_{meas} and $R_{p.i.m.}$ values of 14.2% and 5.7%, respectively.

The molecular-replacement solution was refined to final R_{work} and R_{free} values of 22.9% and 28.3%, respectively. There are no outliers in the Ramachandran plot and the structure is in the 94th percentile of *MolProbity* clash scores (Chen *et al.*, 2010). All relevant structure-refinement statistics are listed in Table 1.

3.2. Overall structure of TpxC61S

TpxC61S exhibits a thioredoxin-like fold with a seven-stranded β-sheet flanked by four α-helices, with an additional two N-terminal β-strands that are typical for Tpx but are not found in other peroxiredoxin structures (Fig. 2). Owing to the mutation of the active cysteine residue to a serine, the structure is locked in the reduced confirmation. This conformation has an elongated α1 helix, which makes Cys61 available for interactions with H_2O_2 or alkyl peroxides; however, the presence of the mutation to serine prevents these interactions. Following oxidation of Cys61, helix α1 becomes unravelled to form an intramolecular disulfide bond with residue Cys95 (Hall *et al.*, 2009; Gabrielsen *et al.*, 2012). The conformational change observed in response to redox state is shown in Fig. 3(*b*).

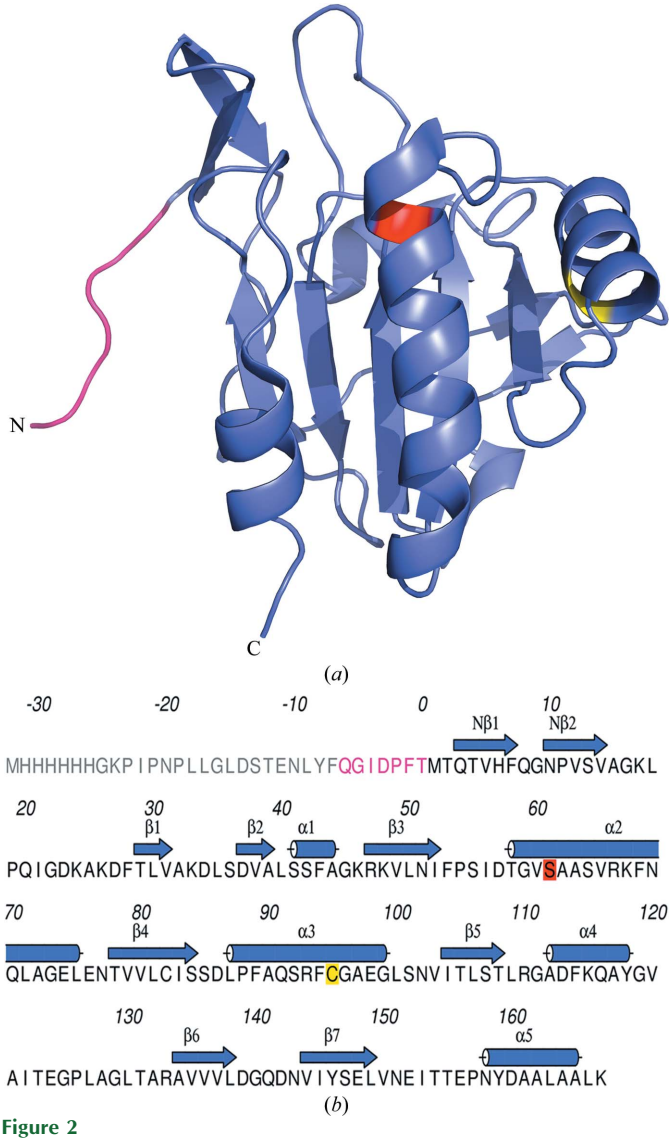


Figure 2
The structure of orthorhombic TpxC61S compared with existing structures. (*a*) The overall structure of TpxC61S is shown with Ser61 (red) and Cys95 (yellow) highlighted. The N- and C-termini of the protein have been labelled, along with the secondary-structure elements. (*b*) The sequence of the purified and crystallized protein with secondary-structure elements indicated above and the catalytic residues highlighted. The affinity tag is shown in grey, with the part of the tag that can be observed in electron density shown in pink.

Table 2
Superposition of orthorhombic TpxC61S with reduced wild-type and 'forced reduced' mutated Tpx structures.

The oxidized Tpx from *E. coli* (PDB entry 3hvs) has been added for completeness.

Species	PDB code
Escherichia coli	3hvv
Yersinia pseudotuberculosis	2xpd
Y. pseudotuberculosis	2yjh
Mycobacterium tuberculosis	1y25
Streptococcus pneumoniae	1psq
E. coli	3hvs

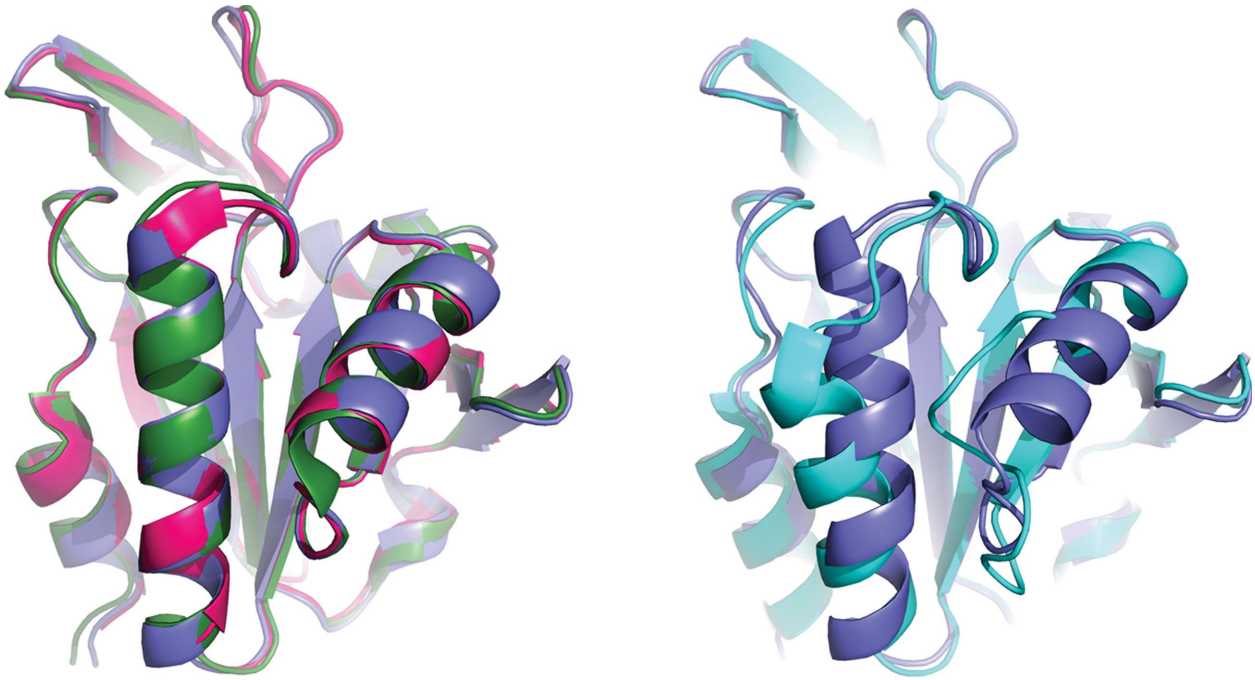
Tpx forms a homodimer in solution (Baker & Poole, 2003; Gabrielsen *et al.*, 2012). In this crystal structure the dimer is formed by a crystallographic symmetry-related molecule. The structure of TpxC61S superposes well onto other crystal structures of reduced Tpx (Table 2). In particular, this mutant superposes with the wild-type reduced protein from *Yersinia pseudotuberculosis* (Gabrielsen *et al.*, 2012) with a root-mean-square deviation (r.m.s.d.) of 0.376 Å, indicating a high degree of similarity (Fig. 3*a*). This confirms the assumption made by Hall *et al.* (2009) that TpxC61S represents a 'forced reduced' structure. When compared with the oxidized *E. coli* structure (PDB entry 3hvs; Hall *et al.*, 2009), TpxC61S superposes with an r.m.s.d. of 0.711 Å. This reflects the local conformational change between the oxidized and the reduced structure mediated by the formation of a disulfide bond between Cys61 and Cys95 (Fig. 3*b*).

3.3. Analysis of the crystal lattice

The orthorhombic form of TpxC61S superposes well onto the existing trigonal TpxC61S structure (PDB entry 3hvv; Hall *et al.*, 2009) with an r.m.s.d. of 0.207 Å, showing that despite the different space groups ($C222_1$ and $P3_121$, respectively) the structure remains essentially the same (Fig. 2*b*). However, when investigating the crystal packing, a noticeable difference between the two crystal forms became apparent. The unit cell of the trigonal form of TpxC61S is much more densly packed, with a solvent content of 42%. In contrast, orthorhombic TpxC61S has a solvent content of 60%. This explains the difference in the crystal contacts formed between the monomeric asymmetric unit and its environment. Whilst the trigonal crystal form has ten neighbouring TpxC61S molecules within 5.0 Å, the orthorhombic form has only six.

Despite the more spacious packing of the protein in $C222_1$, the crystals diffract similarly and are robust. When analyzing the crystal lattice, the linker region of the affinity-tag construct, which is accounted for in the electron density in orthorhombic TpxC61S, is involved in several crystal contacts and thus stabilizes the crystal through facilitating lattice formation (Fig. 4*a*).

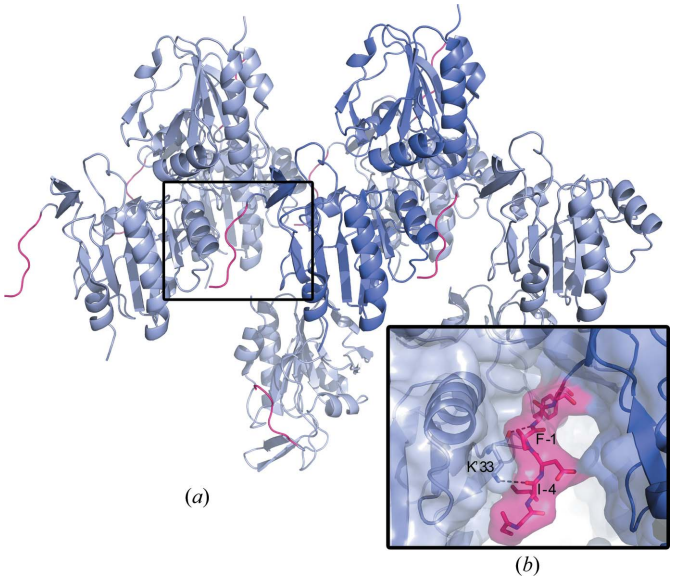
Analysis of the crystal packing in the orthorhombic crystal lattice using the *Protein Interfaces, Surfaces and Assemblies* (*PISA*) server (Krissinel & Henrick, 2007) revealed that 17% of the total surface area of 8349 Å^2 was buried. The largest buried area occurs at the dimerization interface of the protein, which accounts for about half of the buried surface area. The rest of the crystal contacts are formed between the linker region of the affinity tag of one molecule and a neighbouring molecule, as shown in Fig. 4(*b*), with a smaller second crystal contact with another subunit. The area of contact conferred by the linker region is stabilized by two hydrogen bonds between the tag



(*a*) (*b*)

Figure 3
Comparison of TpxC61S with existing Tpx structures. (*a*) Superposition of orthorhombic TpxC61S (slate) with trigonal TpxC61S (PDB entry 3hvv; Hall *et al.*, 2009; pink; *E. coli*) and reduced wild-type Tpx (PDB entry 2xpd; Gabrielsen *et al.*, 2010; green; *Y. pseudotuberculosis*) illustrates the high level of similarity between this structure, the existing TpxC61S structure from *E. coli* and that of reduced wild-type Tpx. (*b*) Orthorhombic TpxC61S (slate) superimposes less well with oxidized wild-type Tpx (cyan; PDB entry 3hvs; Hall *et al.*, 2009) owing to the confirmational change caused by the differing redox states.

(Ile−4 and Phe−1) and a symmetry-related molecule (Lys′33) (Fig. 4*b*). In comparison, the buried surface area between trigonal TpxC61S and symmetry-related neighbours makes up around 10% of the total surface area. However, each subunit makes direct crystal contacts with eight symmetry-related molecules. This highlights the importance of the tag in stabilizing TpxC61S in the orthorhombic form.



(*a*) (*b*)

Figure 4
Crystal packing of TpxC61S. (*a*) The crystal lattice formed in the orthorhombic crystal form of TpxC61S highlights the importance of the linker region of the affinity tag (pink) in forming crystal contacts. The biological dimer formed by crystallographic symmetry is shown in slate, with other molecules in the crystal lattice shown in light blue. (*b*) A detailed view of the crystal contacts between the linker region and a symmetry-related molecule highlighted in the black box shown in (*a*). The residues involved in hydrogen-bond formation have been labelled, namely Lys′33 N^{ζ} to Ile−4 N (3.2 Å) and Lys′33 O to Phe−1 O (3.1 Å). Hydrogen bonds are shown as dashed lines.

4. Conclusions

We have reported the structure of an orthorhombic crystal form of TpxC61S from *E. coli* at 1.97 Å resolution. The overall structure is well conserved amongst known homologues and this mutant form is a good representation of the reduced form of Tpx. The crystal lattice is held in place by the linker region of the affinity tag of one subunit attaching to a neighbouring unit, threading throughout the lattice.

We thank the Diamond Light Source for access to beamline I04 (proposal No. MX1229). This work was supported by a grant from the Biotechnology and Biological Sciences Research Council to MG and AJR (BB/G011389/1), a Medical Research Scotland Grant (223 ORG) to AJR and a Wellcome Trust Studentship for KSHB. We acknowledge Karen McLuskey for discussions and thorough reading of the manuscript.

References

Baker, L. M. & Poole, L. B. (2003). *J. Biol. Chem.* **278**, 9203–9211.
Baron, C. (2010). *Curr. Opin. Microbiol.* **13**, 100–105.
Bond, C. S. & Schüttelkopf, A. W. (2009). *Acta Cryst.* D**65**, 510–512.
Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P., Sharff, A., Smart, O. S., Vonrhein, C. & Womack, T. O. (2011). *BUSTER* v.2.8.0. Cambridge: Global Phasing Ltd.
Cha, M.-K., Kim, H.-K. & Kim, I.-H. (1995). *J. Biol. Chem.* **270**, 28635–28641.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Gabrielsen, M., Beckham, K. S. H., Feher, V. A., Zetterström, C. Z., Wang, D., Müller, S., Elofsson, M., Amaro, R. E., Byron, O. & Roe, A. J. (2012). *PLoS One*, **7**, e32217.
Gabrielsen, M., Zetterström, C. E., Wang, D., Beckham, K. S. H., Elofsson, M., Isaacs, N. W. & Roe, A. J. (2010). *Acta Cryst.* F**66**, 1606–1609.
Hall, A., Sankaran, B., Poole, L. B. & Karplus, P. A. (2009). *J. Mol. Biol.* **393**, 867–881.
Horst, S. A., Jaeger, T., Denkel, L. A., Rouf, S. F., Rhen, M. & Bange, F. C. (2010). *J. Bacteriol.* **192**, 2929–2932.
Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Riboldi-Tunnicliffe, A. & Hilgenfeld, R. (1999). *J. Appl. Cryst.* **32**, 1003–1005.
Velankar, S. *et al.* (2012). *Nucleic Acids Res.* **40**, D445–D452.
Wang, D. *et al.* (2011). *J. Biol. Chem.* **286**, 29922–29931.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.