DOCTOR OF PHILOSOPHY

Structural studies to inform antimicrobial drug discovery and the basis of immunity against T6 effectors

O'Rourke, Patrick

Award date: 2013

Awarding institution: University of Dundee

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 11. Sep. 2020
Structural studies to inform antimicrobial drug discovery and the basis of immunity against T6 effectors

Patrick O'Rourke

2013

University of Dundee
A thesis submitted for the degree of Doctor of Philosophy

Structural studies to inform antimicrobial drug discovery and the basis of immunity against T6 effectors

Patrick E. F. O’Rourke

Supervisor: Professor William N. Hunter

The Wellcome Trust Biocentre
College of Life Sciences
University of Dundee
Scotland

August 2013
CONTENTS

List of Figures III
List of Tables IV
Acknowledgements VI
Declaration VII
Summary VIII
Abbreviations X
Publications XI

Part One – The X-ray crystal structures of potential antimicrobial drug targets

Chapter 1: Introduction

1.1 Structure-based drug discovery and fragment screening 3
1.2.1 The role of isoprenoids in microbial cells 5
1.2.2 The methyl-D-erythritol phosphate (MEP) pathway of isoprenoid-precursor biosynthesis 6
1.2.3 Fosmidomycin, a known drug targeting the MEP pathway 11
1.2.4 Targeting the MEP pathway in Plasmodium falciparum 14
1.2.5 IspF – Known structures, catalytic mechanism and essentiality 15

1.3.1 para-Aminobenzoic acid biosynthesis 19
1.3.2 Anti-folate drug discovery and sulphonamides as antimicrobial agents 21
1.3.3 PabC 22

1.4 Biophysical screening techniques 24
1.5 Aims - Part one 26

Chapter 2: Materials and methods

2.1.1 Reagents, bacterial strains, growth media and equipment 28
2.1.2 Gene cloning and plasmid preparation 29
2.1.3 Protein production 30
2.1.4 Differential scanning fluorimetry and buffer screening 31
2.1.5 Fragment screening by bio-layer interferometry 31
2.1.6 Screening crystallisation conditions 32

2.2.1 Purification of PfIspF 33
2.2.2 Purification of BcIspF 34
2.2.3 Purification of biotinylated BcIspF and EcIspF for BLI 34
2.2.4 Purification PaPabC 36

2.3.1 Crystallisation of PfIspF in complex with CDP 36
2.3.2 Crystallisation of holo-PfIspF 36
2.3.3 Crystallisation of BcIspF in complex with CMP 37
2.3.4 Crystallisation of BcIspF in complex with citrate 37
2.3.5 Crystallisation of PaPabC 38
2.4 Data processing, structure solution and refinement 38

**Chapter 3: IspF; structural analysis and fragment screening**

3.1 *Pf*IspF and *Bc*IspF results
   3.1.1 Structure of IspF 41
   3.1.2 Complexes with CMP and CDP 47
   3.1.3 *Bc*IspF-citrate complex 50
   3.1.4 Co-crystallisation with potential ligands 51

3.2 Structural studies on IspF - Conclusions 52

3.3 IspF fragment screening
   3.3.1 *Bc*IspF for *Bc*IspF 54
   3.3.2 *Ec*IspF screening results 56
   3.3.3 IspF fragment screening - Discussion 59

**Chapter 4: PaPabC; structural analysis**

4.1 PaPabC - Overall structure 62
4.2 PLP cofactor binding site and interactions 65
4.3 PaPabC active site - Comparison with orthologues 67
4.4 PaPabC - Discussion 69

**Part Two - Immunity proteins in the Gram-negative Type VI secretion system**

**Chapter 5: Rap1a and Rap2a - Introduction and methods**

5.1 Type VI secretion systems (T6SS) 77
5.2 Aims - Part two 82
5.3 Rap1a and Rap2a - Crystallisation, molecular replacement and analysis 83

**Chapter 6: Rap1a and Rap2a – Results and discussion**

6.1 Structure of Rap1a 86
6.2 Structure of Rap2a 91
6.3 Structural basis of immunity to T6 effectors - Conclusions 94

**Chapter 7: References** 96

**Appendix 1** (Representative experimental data from the purification of *Bc*IspF) 112
**Appendix 2** (Publications) 115
# LIST OF FIGURES

1.1 An illustrative example of rational drug design informed by X-ray structural studies  
1.2 Isoprenoid compounds  
1.3 Lovastatin inhibits HMG-CoA reductase, a key regulatory enzyme on the mevalonate pathway of isoprenoid-precursor biosynthesis  
1.4 The MEP pathway for isoprenoid-precursor biosynthesis  
1.5 Chemical structure of fosmidomycin  
1.6 Reaction catalysed by IspF  
1.7 Chemical structure of PABA and folate  
1.8 PLP forms a covalent link with the amino group of lysine sidechains  
1.9 Chemical structures of sulphonamide antimicrobials that inhibit folate biosynthesis  
1.10 PabC catalyzes the conversion of 4-amino-4-deoxychorismate to PABA and pyruvate  
1.11 Principles of BLI  
3.1 Sequence alignment of BcIspF and PfIspF  
3.2 Overall structure of IspF  
3.3 Superimposition of position 1 residues in BcIspF and PfIspF  
3.4 CMP binding interactions at the enzyme active site in BcIspF  
3.5 CMP #2 occupies the same site as MECPP  
3.6 Citrate binding to Zn$^{2+}$ in BcIspF  
3.7 Measuring the binding affinities of cytidine nucleotides for BcIspF  
3.8 Fragments identified in the screen against BcIspF
An example of BLI data from fragment screening of EcIspF

Fragments identified in the screen against EcIspF

The PaPabC dimer

Ribbon diagram of the PaPabC monomer

Omit map for PLP and Lys140 from subunit A of PaPabC

Alignment of PaPabC and LpPabC

Proposed catalytic mechanism of PabC

Structural conservation of tyrosine in the PabC active site

A stereoview of the catalytic intermediate docked into the active site of PaPabC

Schematic model of the Type VI secretion apparatus

The T6SS delivers toxic effectors to target cells

Overall structure of Rap1a

Superimposition of YmgD and HdeA on Rap1a

The heterotetrameric Ssp1-Rap1a complex

Overall structure of Rap2a

LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>IspF proteins of known structure</td>
<td>16</td>
</tr>
<tr>
<td>2.1</td>
<td>List of key details for the proteins reported in this thesis (Part 1)</td>
<td>29</td>
</tr>
<tr>
<td>2.2</td>
<td>Oligonucleotide primers for PCR</td>
<td>30</td>
</tr>
<tr>
<td>2.3</td>
<td>BLI fragment screen assay conditions</td>
<td>32</td>
</tr>
<tr>
<td>3.1</td>
<td>Crystallographic statistics for PfIspF</td>
<td>46</td>
</tr>
<tr>
<td>3.2</td>
<td>Crystallographic statistics for BcIspF</td>
<td>47</td>
</tr>
<tr>
<td>4.1</td>
<td>Crystallographic statistics for PaPabC</td>
<td>63</td>
</tr>
</tbody>
</table>
5.1 Components of the T6SS apparatus 79
5.2 List of key details for the proteins reported in this thesis (Part 2) 81
6.1 Crystallographic statistics for Rap1a and Rap2a 87
6.2 Structural similarity of Rap2a to known Tai4-family proteins 93
Acknowledgements

All the members of the Hunter laboratory have been supportive and friendly. In particular, Dr. Thomas Eadsforth has been my post-doctoral mentor and friend throughout my investigations. Sharon Shepherd provided me with instruction and training in protein purification techniques. Dr. Paul Fyfe assisted me during my first few months in the laboratory and provided me with training in the use of the new X-ray diffraction system.

Special acknowledgement is paid here to the contribution of Dr. Justyna Kalinowska-Tłuściek whose work in optimising both expression and crystallisation of the P/JspF protein I have successfully reproduced in my own work.

Thanks must go to Dr. Vincent Rao and Dr. Velupillai Srikanthasasan for sharing with me their unprocessed diffraction data on the Type VI secretion system proteins, Rap1a and Rap2a. Thanks also go to Dr. David Robinson who helped me perform bio-layer interferometry experiments; and to Dr. Mark Agacan for advice on the use of the in-house X-ray generators.

I acknowledge the assistance of all those involved in providing lab-services including wash-up, autoclave and media support.

I am grateful to the Biotechnology and Biological Sciences Research Council and to Pfizer, Inc. for awarding a BBSRC-CASE scholarship in support of my work with Prof. Hunter. I also gratefully acknowledge access to the excellent facilities at the synchrotron sources Diamond Light Source in Oxfordshire and the European Synchrotron Radiation Facility in Grenoble.

I would like to thank my supervisor Prof. Bill Hunter for his warm encouragement, excellent advice and support.

I am especially pleased to acknowledge the kind support of my parents and friends during the course of my PhD studies.

Patrick O’Rourke.
DECLARATION

I hereby declare that the following thesis is based on the result of investigations conducted by myself, and that this thesis is of my own composition. This thesis has not, in whole or in part, been previously presented for a higher degree. Work other than my own is clearly indicated in the text by reference to the relevant researchers or publication.

Patrick E. F. O'Rourke

The work presented in this thesis is the work of the candidate P. O’Rourke. Conditions of the relevant Ordinance and Regulations have been fulfilled.

Prof. W. N. Hunter.
SUMMARY

Work presented in this thesis is in two parts.

**Part one: The X-ray crystal structures of potential antimicrobial drug targets.**

The protein **IspF** (2C-methyl-D-erythritol-2,4-cyclodiphosphate synthase, EC: 4.6.1.12) from two pathogens (*Burkholderia cenocepacia* and *Plasmodium falciparum*) has been investigated. IspF is an enzyme of isoprenoid-precursor biosynthesis and is considered to be a potential drug target. The results of structural and fragment-screening efforts presented here inform early stage drug discovery efforts.

The structure of the **PabC** protein (4-amino-4-deoxychorismate lyase, EC: 4.1.3.38) from the Gram-negative pathogen *Pseudomonas aeruginosa* was also determined. PabC is involved in the production of *para*-aminobenzoic acid on the path to folate. Comparisons with previously solved PabC structures identified a spatially conserved tyrosine residue in the active site and suggest that a re-evaluation of a published mechanism is warranted.

**Part two: Immunity proteins in the Gram-negative Type VI secretion system**

The X-ray crystal structures of the proteins **Rap1a** and **Rap2a** from *Serratia marcescens*, inhibitors of the peptidoglycan amidase toxins secreted by some Gram-negative bacteria employing the Type VI secretion pathway, were determined by molecular replacement and analysed.

**Part one:**

Chapter 1: Introduces isoprenoid-precursor biosynthesis and its importance as a potential source of novel drug targets for infectious diseases. The structure of IspF from the methyl-D-erythritol phosphate (MEP) pathway was sought, in complex with ligands, to inform early-stage drug discovery efforts. The PabC protein is introduced. The significance of *para*-aminobenzoic acid for bacterial folate biosynthesis and the history of anti-folate drug discovery are discussed. Some general background information is provided describing the processes of fragment based screening and structure-based drug discovery.
Chapter 2: Details of materials used and experimental procedures are given.

Chapter 3: (Results and Discussion - IspF structures) High resolution structures of IspF from the pathogens *B. cenocepacia* and *P. falciparum* are presented. The crystal structures of *Bc*IspF with cytidine derivatives were determined by molecular replacement methods, refined and analysed. The active site contained two molecules of cytidine monophosphate (CMP) and a structure was subsequently re-determined from a crystal grown in the presence of a lower CMP concentration. A structure of *Bc*IspF with citrate, a buffer component, bound at the active site is also reported. Despite several concerted attempts, no co-crystal structures could be obtained with novel ligands. The structure of *Pf*IspF is described bound to cytidine diphosphate (CDP). This reproduced earlier studies on this protein in the Hunter lab by Dr. Kalinowska-Tluścik, demonstrating a robust protocol for production of *Pf*IspF crystals. Additionally, *holo*-crystals of *Pf*IspF without any cytidine derivatives present were obtained and the structure determined. A chemical fragment screen was performed against biotinylated *Bc*IspF and the *E. coli* IspF orthologue employing the bio-layer interferometry technique (BLI). Fragments identified are described.

Chapter 4: *PaPabC* structure

The structure of PabC, an essential enzyme from *P. aeruginosa*, is described. This structure demonstrates the presence of a structurally conserved tyrosine residue in the active site that had been overlooked in previous studies.

**Part two:**

Chapter 5: Introduces the Type VI secretion system (T6SS) of Gram-negative bacteria and describes the structural determination of two proteins from *Serratia marcescens*, Rap1a and Rap2a, that confer immunity against certain peptidoglycan amidase effectors of the T6SS.

Chapter 6: The structures of Rap1a and Rap2a are described and compared to previously reported immunity proteins from different bacteria. Rap1a is shown to adopt a unique fold for an immunity protein, similar to that of other periplasmic proteins.
Abbreviations

S.I. units are used throughout with the exception of Å representing $10^{-10}$ m; and temperatures values, which are given in degrees Celsius (°C).

ASA
Accessible surface area

ATP
Adenosine-5'-triphosphate

BAP
Biotin acceptor peptide

Bis-Tris
2-[Bis(2-hydroxyethyl)amino]-2-\((\text{hydroxymethyl})\text{propane-1,3-diol}

BLI
Biolayer interferometry

Cα
Carbon-alpha of an amino acid

CDP
Cytidine 5'-diphosphate

CHES
$N$-cyclohexyl-2-aminoethanesulfonic acid

CMP
Cytidine 5'-monophosphate

CoA
Coenzyme A

DDU
Drug Discovery Unit

DHFR
Dihydrofolate reductase (EC: 1.5.1.3)

DHPS
Dihydropteroate synthase (EC: 2.5.1.15)

DMAPP
Dimethylallyl pyrophosphate

DMSO
Dimethyl sulfoxide

DOXP
1-deoxy-$D$-xylulose 5-phosphate

DSF
Differential scanning fluorimetry

DXR
1-deoxy-$D$-xylulose 5-phosphate reductoisomerase (EC: 1.1.1.267)

DXS
1-deoxy-$D$-xylulose 5-phosphate synthase (EC: 2.2.1.17)

EC
Enzyme Commission number

EDTA
(Ethlenediamine)tetra-acetic acid

ESI-MS
Electrospray ionisation mass spectrometry

ESRF
European Synchrotron Radiation Facility

FPLC
Fast protein liquid chromatography

FPP
Farnesyl pyrophosphate

GPP
Geranyl pyrophosphate

GTP
Guanosine-5'-triphosphate

HEPES
4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid

His-tag
Hexahistidine protein purification tag

HMBPP
$(E)$-4-hydroxy-3-methyl-but-2-enyl pyrophosphate

HMG-CoA
3-hydroxy-3-methylglutaryl-CoA

HPLC
High-performance liquid chromatography
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDI</td>
<td>IPP-isomerase (EC: 5.3.3.2)</td>
</tr>
<tr>
<td>IPP</td>
<td>Isopentenyl pyrophosphate</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl β-D-1-thiogalactopyranoside</td>
</tr>
<tr>
<td>IspD</td>
<td>2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (EC: 2.7.7.60)</td>
</tr>
<tr>
<td>IspE</td>
<td>4-diphosphocytidyl-2C-methyl-D-erythritol kinase (EC: 2.7.1.148)</td>
</tr>
<tr>
<td>IspF</td>
<td>2-C-methyl-D-erythritol-2,4-cyclodiphosphate synthase (EC: 4.6.1.12)</td>
</tr>
<tr>
<td>IspG</td>
<td>HMBPP synthase (EC: 1.17.7.1)</td>
</tr>
<tr>
<td>IspH</td>
<td>HMBPP reductase (EC: 1.17.1.2)</td>
</tr>
<tr>
<td>LB media</td>
<td>Luria-Bertani (lysogeny broth) media</td>
</tr>
<tr>
<td>mA2pm</td>
<td>meso-diaminopimelate</td>
</tr>
<tr>
<td>MALDI-TOF MS</td>
<td>Matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry</td>
</tr>
<tr>
<td>MEcPP</td>
<td>2-C-methyl-D-erythritol 2,4-cyclodiphosphate</td>
</tr>
<tr>
<td>MEP</td>
<td>Methyl-D-erythritol phosphate</td>
</tr>
<tr>
<td>MES</td>
<td>2-(N-morpholino)ethanesulfonic acid</td>
</tr>
<tr>
<td>NADPH</td>
<td>Nicotinamide adenine dinucleotide phosphate</td>
</tr>
<tr>
<td>NCS</td>
<td>Non-crystallographic symmetry</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance spectroscopy</td>
</tr>
<tr>
<td>OD&lt;sub&gt;600&lt;/sub&gt;</td>
<td>Optical density at 600 nm</td>
</tr>
<tr>
<td>PABA</td>
<td>para-aminobenzoic acid</td>
</tr>
<tr>
<td>PabC</td>
<td>4-amino-4-deoxychorismate lyase (EC: 4.1.3.38)</td>
</tr>
<tr>
<td>PAGE</td>
<td>Polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein Data Bank</td>
</tr>
<tr>
<td>PEG</td>
<td>Polyethylene glycol</td>
</tr>
<tr>
<td>PLP</td>
<td>Pyridoxal 5'-phosphate</td>
</tr>
<tr>
<td>PP&lt;sub&gt;1&lt;/sub&gt;</td>
<td>Pyrophosphate</td>
</tr>
<tr>
<td>PVDF</td>
<td>Polyvinylidene difluoride</td>
</tr>
<tr>
<td>R&lt;sup&gt;2&lt;/sup&gt;</td>
<td>Coefficient-of-determination value</td>
</tr>
<tr>
<td>Rap</td>
<td>Resistance associated protein</td>
</tr>
<tr>
<td>RMSD</td>
<td>Root-mean-square deviation</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>Real-time polymerase chain reaction</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium dodecyl sulfate</td>
</tr>
<tr>
<td>SPR</td>
<td>Surface plasmon resonance spectroscopy</td>
</tr>
<tr>
<td>Ssp</td>
<td>Small secreted protein</td>
</tr>
<tr>
<td>T6SS</td>
<td>Type VI secretion system</td>
</tr>
<tr>
<td>T6 effectors</td>
<td>Effector proteins of the Type VI secretion system</td>
</tr>
<tr>
<td>TEV</td>
<td>Tobacco Etch Virus protease</td>
</tr>
<tr>
<td>TLC</td>
<td>Thin layer chromatography</td>
</tr>
<tr>
<td>Tris</td>
<td>Tris(hydroxymethyl)aminomethane</td>
</tr>
<tr>
<td>v/v</td>
<td>Volume per volume</td>
</tr>
<tr>
<td>w/v</td>
<td>Weight per volume</td>
</tr>
</tbody>
</table>
Publications

The work presented in this thesis has resulted in two publications.


PART ONE

THE X-RAY CRYSTAL STRUCTURES OF POTENTIAL ANTIMICROBIAL DRUG TARGETS
CHAPTER 1

INTRODUCTION
Introduction

1.1 Structure-based drug discovery and fragment screening

Modern drug discovery is often a long and costly process. Target-based approaches to antimicrobial drug discovery involve careful selection of proteins demonstrated or predicted to be essential. Crystallographic structural information about the target proteins, especially when structures in complex with ligands are available, can provide valuable information at the early stages of drug discovery; informing target assessment, rationalising structure-activity relationships between ligands and suggesting sites for potential improvements to the lead compounds (Hunter, 2009; Moynie et al., 2013; Verlinde & Hol, 1994; Whittle & Blundell, 1994). In addition, high-quality structures can provide a template for virtual screening efforts (Mpamhanga et al., 2009). The structure-based discovery of drugs in clinical use targeting human immunodeficiency virus (Kempf et al., 1995; Wlodawer & Erickson, 1993), influenza virus (von Itzstein et al., 1993) and kinases involved in cancer development (Bollag et al., 2010; Noble et al., 2004; Tsai et al., 2008) are examples that illustrate the efficacy of this approach (Figure 1.1).

Identifying ligands for target proteins during drug discovery can be challenging, as both high affinity and selective interactions are sought. Libraries of chemical compounds with mass typically less than 300 Da, called 'fragments', can be screened in an attempt to identify starting points for such ligands (Congreve et al., 2008; Scott et al., 2012). Structural studies complement this approach as once the relative binding location and orientation of fragments is elucidated, medicinal chemistry can often provide suggestions to link, fuse or otherwise elaborate on these starting points. Such elaborated ligands often have much higher binding affinities than the original fragments, offering an attractive route to lead compounds; particularly for proteins
where either no or only a limited number of ligands are known (de Kloeg et al., 2009). Alternatively, fragments can be screened by attempting to co-crystallise with the target protein or by soaking into preformed crystals and determining the structure of complexes by high-throughput crystallography (Hartshorn et al., 2005). Owing to the relatively small size and typically weak affinity of fragments, ranging from μM to mM, sensitive biophysical assays, such as surface plasmon resonance (SPR) or nuclear magnetic resonance spectroscopy (NMR), are usually required for screening (Navratilova & Hopkins, 2010; Shuker et al., 1996).

Figure 1.1 An illustrative example of rational drug design informed by X-ray structural studies. The development of zanamivir, a potent neuraminidase inhibitor for the treatment of influenza, is represented in schematic form. Structural determination of the target protein in complex with ligands and lead compounds suggested routes for elaboration of the lead compounds by medicinal chemistry. Iterative cycles of chemistry and structural determination support each other during the drug discovery process, leading to the identification of successively more potent compounds. In this example the crystal structure of the target revealed the presence of two glutamate residues adjacent to the lead compound. Addition of a guanidino group (present in the final inhibitor, zanamivir) to promote interaction with these glutamates increased potency. (Smith et al., 2001; von Itzstein et al., 1993)

In the studies presented in this thesis, the structures of essential proteins from several human pathogens were determined. *Pseudomonas aeruginosa* and *Burkholderia cenocepacia* are Gram-negative bacteria, implicated in hospital-acquired infections of immuno-compromised patients, and *Plasmodium falciparum* is a protozoan parasite - the causative agent of cerebral malaria.
1.2.1 The role of isoprenoids in microbial cells

Isoprenoids, also known as terpenoids, are a large and structurally diverse class of natural products. All isoprenoids derive from the five-carbon universal precursors isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP), isomeric building blocks bearing an unsaturated carbon-carbon double bond that can be linked together to form longer chains known as prenyl groups, such as geranyl- (C10), farnesyl- (C15) and geranylgeranyl (C20) pyrophosphate (Figure 1.2). The attachment of prenyl groups to proteins is a common protein post-translational modification that results in anchoring to lipid membranes. More complex isoprenoid structures, including membrane components such as cholesterol in eukaryotes and hopanoids in bacteria can be formed by cyclisation and further chemical modifications. Certain compounds bearing isoprenoid-derived side chains, such as menaquinone and ubiquinone, are essential to respiratory redox chains for energy generation in cells, where they function as membrane associated mobile electron carriers. Long-chain membrane associated isoprenoid compounds are employed as scaffolds during the assembly of glycan structures. In eukaryotes dolichol pyrophosphate fulfills this role in N-glycosylation, and for many bacteria undecaprenyl pyrophosphate functions in the synthesis of peptidoglycan precursors during assembly of the cell wall (Bouhss et al., 2008; Ershov, 2007).
Figure 1.2 Isoprenoid compounds. The chemical structures of selected isoprenoid compounds are shown. 1: isopentenyl pyrophosphate (IPP). 2: dimethylallyl pyrophosphate (DMAPP). 3: geranyl pyrophosphate (GPP). 4: farnesyl pyrophosphate (FPP). 5: dolichol phosphate. 6: ubiquinone. 7: hopane. A double-headed arrow has been drawn to indicate that the isoprenoid-precursors, 1 and 2, are isomers. (Adapted from Ershov, 2007)

1.2.2 The methyl-D-erythritol phosphate (MEP) pathway of isoprenoid-precursor biosynthesis

Two biosynthetic pathways have been discovered for the production of the universal isoprenoid-precursor molecules IPP and DMAPP. The mevalonate pathway, which was discovered first (Bloch, 1964), starts with the condensation of two molecules of acetyl-CoA to form acetoacetyl-CoA. This is followed by conversion to 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA), which is in turn converted to mevalonic acid; an intermediate metabolite from which the pathway derives its name, by the action of the NADPH-dependent enzyme HMG-CoA reductase (EC: 1.1.1.34). Mevalonic acid is subsequently phosphorylated twice and decarboxylated to form IPP. IPP and DMAPP
are interconverted by Mg$^{2+}$ dependent IPP-isomerase (IDI) enzymes (EC: 5.3.3.2). HMG-CoA reductase is known to be a key regulatory enzyme on the mevalonate pathway (Brown & Goldstein, 1980). Inhibitors of this enzyme such as the statin drugs, for example lovastatin (Alberts et al., 1980), are in clinical use for the treatment of hypercholesterolaemia (Figure 1.3).

![Diagram of lovastatin inhibition](image)

**Figure 1.3** Lovastatin inhibits HMG-CoA reductase, a key regulatory enzyme on the mevalonate pathway of isoprenoid-precursor biosynthesis.

In 1993, Rohmer et al., provided evidence for the existence of a novel, non-mevalonate pathway for isoprenoid-precursor biosynthesis (Figure 1.4). Incorporation of $^{13}$C isotopes into isoprenoid species including hopanoids and ubiquinone was monitored by NMR using labelled test-substrates in several bacteria. The labelling pattern was distinct from that established for the mevalonate pathway, and was consistent with the formation of IPP via a novel pathway starting from pyruvate and a triose phosphate, later shown to be glyceraldehyde-3-phosphate (Rohmer et al., 1996). This contention was tested and confirmed by investigating the isotope labelling pattern.
observed in isoprenoids in the eukaryotic green-alga *Scenedesmus obliquus* following a similar procedure (Schwender *et al.*, 1996). Isoprenoid-precursor biosynthesis in plants was re-investigated in light of the work on green-algae (Lichtenthaler *et al.*, 1997). In that study, isoprenoids extracted from the cytoplasm, sitosterol and stigmasterol, were derived from the classical mevalonate pathway; whereas isoprenoids found exclusively in chloroplasts, phytol, β-carotene, lutein and the isoprenoid side chain of plastocynone-9, exhibited the novel labelling pattern observed previously in bacteria (Rohmer *et al.*, 1993) and algae (Schwender *et al.*, 1996). This suggested that in higher plants the mevalonate pathway operates in the cytoplasm and the non-mevalonate pathway operates in the chloroplasts (Lichtenthaler *et al.*, 1997; reviewed in Lichtenthaler, 1999).

![The MEP pathway of isoprenoid-precursor biosynthesis](image)

**Figure 1.4** The MEP pathway for isoprenoid-precursor biosynthesis. The universal precursors for isoprenoid biosynthesis, IPP and DMAPP, are formed via the MEP pathway. IspF, the fifth enzyme on this pathway, is the subject of studies reported in this thesis.
The non-mevalonate pathway was subsequently re-named the 1-deoxy-D-xylulose-5-phosphate (DOXP) pathway, when it was demonstrated that DOXP is the first metabolite on this pathway downstream of pyruvate and glyceraldehyde-5-phosphate, and the enzyme responsible for its synthesis (DOXP synthase, DXS, EC: 2.2.1.7) was characterised (Sprenger et al., 1997). Since DOXP is also required for the synthesis of thiamine and pyridoxol, the non-mevalonate pathway in now referred to as the 2C-methyl-D-erythritol-4-phosphate (MEP) pathway, named after the first committed metabolite on the pathway to IPP/DMAPP. DOXP is converted to MEP by the enzyme DOXP reductoisomerase (DXR, EC: 1.1.1.267; Takahasi et al., 1998). DXR was shown to require NADPH and divalent cations for activity (Takahasi et al., 1998). Fosmidomycin (Figure 1.5), a low-molecular weight natural product, inhibits DXR (Section 1.2.3). Radiolabelled MEP was prepared (Rohdich et al., 1999) and, following incubation with E. coli cell extracts, thin layer chromatography (TLC) and NMR analyses showed that an enzyme activity transferred a cytidine 5’-phosphate moiety to the phosphate group of MEP, to form 4-diphosphocytidyl-2-C-methylerythritol. A database search for cytidylyltransferases turned up the N-terminal domain of the acs1 gene from Haemophilius influenzae, the enzyme product of which forms CDP-ribitol (Follens et al., 1999). A bioinformatic analysis led to the identification of the gene ygbP, encoding 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (IspD, EC: 2.7.7.60). Enzyme assays demonstrated that IspD indeed has MEP-cytidylyltransferase activity; and when the cytidylated product (CDP-ME) was incubated with chromoplasts of Capsicum annuum it became incorporated into plant isoprenoids (Rohdich et al., 1999). Comparative genomic bioinfomatics identified a gene (ychB) that followed the same distribution as other genes involved in the MEP pathway. The protein encoded by ychB, known as IspE (EC: 2.7.1.148), was shown by TLC and NMR analyses to transfer a phosphate group from ATP to the 2-
hydroxy group of CDP-ME. The product (CDP-MEP) was incorporated into the plant isoprenoids, phytoene and β-carotene (Lüttgen et al., 2000). Previous work (Rohdich et al., 1999) had demonstrated the existence of gene fusions between ygbP and another gene, ygbB. In 2000, Herz et al., sought to demonstrate that the YgbB protein, later renamed IspF, was involved in the MEP pathway. IspF from E. coli was recombinantly expressed, and purified protein was incubated with 13C-labelled CDP-MEP in vitro and demonstrated by TLC and NMR to catalyze the production of 2C-methyl-D-erythritol 2,4-cyclodiphosphate (MEcPP), releasing cytidine 5’-monophosphate (CMP). This reaction was shown to require Mg2+ or Mn2+ ions. MEcPP was incorporated into carotenoids (Herz et al., 2000). Thus, IspF displays 2C-methyl-D-erythritol-2,4-cyclodiphosphate synthase (EC: 4.6.1.12) activity (Section 1.2.5). MEcPP is converted to (E)-4-hydroxy-3-methyl-but-2-enyl pyrophosphate (HMBPP) by the iron-sulfur cluster dependent enzyme, HMBPP synthase (IspG, EC: 1.17.7.1), introducing the carbon-carbon double bond characteristic of isoprenoids (Kollas et al., 2002). The last step on the MEP pathway is performed by HMBPP reductase (IspH, EC: 1.17.1.2), another iron-sulfur cluster dependent enzyme, which produces both IPP and DMAPP from HMBPP (Altincicek et al., 2002; Rohdich et al., 2002; Wolff et al., 2003). While many organisms possess IDI enzymes that interconvert IPP and DMAPP, IDI may be dispensable since IspH produces both precursors (Rohdich et al., 2002) and genes encoding IDI are absent from the genomes of several organisms that use the MEP pathway (Cunningham et al., 2000). In E. coli for example, the gene encoding the sole IDI enzyme is inessential (Hahn et al., 1999). Structural studies have provided insight into the mechanisms of enzymes on the MEP pathway (Hunter, 2007; Zhao et al., 2013).

The MEP pathway of isoprenoid-precursor biosynthesis is widely distributed in various bacteria, algae, and in the chloroplasts of higher plants. Several apicomplexan
Parasites, including *Plasmodium* species responsible for malaria, contain the MEP pathway in an organelle known as an apicoplast (Section 1.2.4). Conversely, the MEP pathway is replaced with the classical mevalonate pathway in animals, fungi, archaea and the cytoplasm of plants (Rohdich *et al.*, 2001). Thus, the MEP pathway contains potential drug-target enzymes whose inhibition may provide broad-spectrum antimicrobial activity, and this is particularly attractive since this pathway is absent in humans.

### 1.2.3 Fosmidomycin, a known drug targeting the MEP pathway

Fosmidomycin, a phosphonic acid bearing an hydroxamate and an aldehyde (Figure 1.5), was first isolated as a natural product from *Streptomyces lavendulae* in 1980 by researchers at Fujisawa Pharamceutical Co. Ltd., Osaka, (Okuhara *et al.*, 1980a). In 1989 it was shown that fosmidomycin exerts its anti-bacterial effects by inhibiting isoprenoid biosynthesis (Shigi, 1989). Fosmidomycin inhibited the production of menaquinones and ubiquinones in growing *E. coli* and inhibited the production of carotenoids and menaquinones in *Micrococcus luteus* (Shigi, 1989). It was further noticed that the bacteria that were susceptible to fosmidomycin were unable to incorporate mevalonate, whereas those that could incorporate mevalonate were not susceptible, however the existence of the MEP pathway, or non-mevalonate pathway, was not known at the time.

Both fosmidomycin and a methylated derivative, FR-900098 (Okuhara *et al.*, 1980b), were shown to potently inhibit recombinant DXR from three different strains of *P. falciparum*, HB3, A2 and Dd2; with *IC*<sub>50</sub> values ranging from 90 - 370 nM (Jomaa *et al.*, 1999). Mice infected with the closely-related *P. vinckei* species were cured following treatment with either compound. Fosmidomycin inhibited the growth of blood-stage *P. falciparum* with an *EC*<sub>50</sub> value of 0.98 µM (Yeh & DeRisi, 2011). High
resolution crystal structures of DXR from *P. falciparum* in complex with fosmidomycin (1.9 Å), FR-900098 (1.9 Å) and inhibitor-free (2.2 Å) were solved by Umeda *et al.* in 2011. Clear electron density was observed for fosmidomycin and FR-900098 in the active site. The hydroxamate group of the inhibitors binds to the active site magnesium cation, with the oxygen atoms in *cis*-configuration.

![Chemical structure of fosmidomycin.](image)

**Figure 1.5 Chemical structure of fosmidomycin.** Fosmidomycin, a phosphonic acid bearing an hydroxamate and aldehyde group, is a natural product with anti-microbial properties. Fosmidomycin inhibits DXR on the MEP pathway, binding to the active site Mg$^{2+}$. Fosmidomycin has been used clinically to manage urinary tract infections; and is now under investigation as a potential anti-malarial drug. (Cassera *et al.*, 2004; Howe *et al.*, 2012; Lell *et al.* 2003; Umeda *et al.*, 2011)

An open-label, uncontrolled trial evaluated fosmidomycin for the treatment of acute uncomplicated *P. falciparum* malaria in subjects aged 18-50 in Gabon and Thailand (Lell *et al.*, 2003). The drug was shown to be a safe and effective anti-malarial, exhibiting rapid parasite and fever clearance - within 44 hours. However, a high number of recrudescent infections were observed following treatment cessation, particularly in Thailand and the authors of the study cautioned against its use as a monotherapy. A follow-up study evaluated the efficacy of a combination treatment of fosmidomycin with clindamycin in children aged 7-14 in Gabon (Borrmann *et al.*, 2004). In contrast to the previous trial, combination therapy with clindamycin resulted in complete parasite elimination and absence of clinical signs of infection, provided treatment lasted more than three days on a 12 hour dosing schedule. Fosmidomycin is
thus an example of an effective anti-malarial and its use in clinical trials chemically validates the MEP pathway as a source of targets for drug discovery (Rohdich et al. 2005; Hunter, 2011).
1.2.4 Targeting the MEP pathway in *P. falciparum*

*P. falciparum* contains an organelle known as an apicoplast which resembles a cyanobacterium and is thought to have arisen from an ancestral secondary endosymbiosis with an alga (Ralph *et al.*, 2004). In 1999, Jomaa *et al.*, reported that DXR localises to the apicoplast in *P. falciparum*. Indeed, all MEP pathway enzymes in *P. falciparum* bear an apicoplast localisation sequence (Ralph *et al.*, 2004) and chemical disruption of apicoplast replication results in parasites that are IPP auxotrophs (Yeh & DeRisi, 2011). The presence of an active MEP pathway in all intra-erythrocytic stages has been experimentally confirmed, and treatment with fosmidomycin depletes intermediates downstream of DXR as well as ubiquinone and dolichol compounds (Cassera *et al.*, 2004). Treatment with fosmidomycin also leads to a loss of protein prenylation, causing various small GTPases such as Rab5 to mislocalise which results in defects in both vesicular trafficking and the integrity of vacuolar membranes (Howe *et al.*, 2013). In 2012 No *et al.* published an example of successful targeting of an isoprenoid-utilising protein downstream of the MEP pathway. A library of bisphosphonates was screened against the intraerythrocytic form of *P. falciparum* and lipophilic derivatives of the drugs risedronate and zoledronate were identified as leads. X-ray crystal structures of the likely target, geranylgeranyl diphosphate synthase (EC: 2.5.1.29), were determined from the *P. vivax* orthologue in complex with these compounds. Such inhibitors suppressed parasitaemia in a murine model infected with *P. chabaudi* (No *et al.*, 2012). *P. falciparum* also synthesises carotenoids that have a likely role in anti-oxidant defense, and inhibition of synthesis by norflurazon suppresses parasite growth in a dose-dependent manner (Tonhosolo *et al.*, 2009). The prospect of developing novel antimalarial agents that exploit the MEP pathway dependence of *P.*
*falciparum*, following the example of fosmidomycin, is particularly attractive (Hunter, 2011; van der Meer & Hirsch, 2012).

### 1.2.5 IspF – Known structures, catalytic mechanism and essentiality

The structure of IspF is known from twelve different species in addition to the two disclosed in this thesis, and thus the overall fold is well established. IspF forms a homotrimer and active sites are located at interfaces between adjacent subunits. Subunits form a fluted β-prism consisting of four β-stands that are oriented approximately parallel to the three-fold axis of symmetry that runs through the center of the molecule; flanked by three α-helices (Kemp et al., 2002). A hydrophobic cavity is present along the axis in most species, but is absent in the IspF from *Thermus thermophilus* (*Tt*IspF, Kishida et al., 2003). The genes encoding IspF and IspD are fused in numerous organisms, an unusual arrangement since these enzymes catalyse non-consecutive steps of the MEP pathway; IspE must first phosphorylate the product of IspD to form the substrate for IspF. Bifunctional IspD-IspF from *Campylobacter jejuni* has been biochemically characterised (Gabrielsen et al., 2004a) and the crystal structure has been determined (Gabrielsen et al., 2004b). The overall structure of IspF is unchanged in the fusion protein. Although IspE associates with the IspD-IspF protein in solution (Gabrielsen et al., 2004b), no rate-enhancement is observed upon complexation (Lherbet et al., 2006), and thus the physiological significance of such an association remains to be clarified.

In 2002, structural studies were published on IspF from *E. coli* (*Ec*IspF) by three groups (Kemp et al., Richard et al., Steinbacher et al., 2002) that provided insight into the catalytic mechanism (Table 1.1). A metal, identified as Zn$^{2+}$ (Kemp et al., 2002), binds at the active site coordinated by the side chains of two histidines and an aspartate. IspF catalyses the production of MEcPP by facilitating an intramolecular nucleophilic
attack on the substrate, CDP-MEP, bridging β-phosphate by the 2-phosphate (Figure 1.6). Zn$^{2+}$ aligns and polarises the β-phosphate for nucleophilic attack and is indispensible for catalysis. The cytidine portion of the substrate is held in a rigid binding pocket, whereas the predominantly apolar binding pocket for the methylerythritol phosphate portion is more flexible.

**Table 1.1 IspF proteins of known structure**

<table>
<thead>
<tr>
<th>Organism name</th>
<th>Representative Protein Data Bank (PDB) code</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Arabidopsis thaliana</em></td>
<td>2PMP (Calisto <em>et al</em>., 2007)</td>
</tr>
<tr>
<td><em>Burkholderia cenocepacia</em> (this thesis)</td>
<td>4C8G (O’Rourke <em>et al</em>., 2013; to be published)</td>
</tr>
<tr>
<td><em>Burkholderia pseudomallei</em></td>
<td>3IEQ (Begley <em>et al</em>., 2011)</td>
</tr>
<tr>
<td><em>Campylobacter jejuni</em> (bifunctional IspDF)</td>
<td>1W55 (Gabrielsen <em>et al</em>., 2004b)</td>
</tr>
<tr>
<td><em>Escherichia coli</em></td>
<td>1U3L (Steinbacher <em>et al</em>., 2002)</td>
</tr>
<tr>
<td><em>Francisella tularensis</em></td>
<td>3RE2 (No associated publication)</td>
</tr>
<tr>
<td><em>Haemophilus influenzae</em></td>
<td>1JN1 (Lehmann <em>et al</em>., 2002)</td>
</tr>
<tr>
<td><em>Mycobacterium smegmatis</em></td>
<td>2UZH (Buetow <em>et al</em>., 2007)</td>
</tr>
<tr>
<td><em>Plasmodium falciparum</em> (this thesis)</td>
<td>4C81 (O’Rourke <em>et al</em>., 2013; to be published)</td>
</tr>
<tr>
<td><em>Plasmodium vivax</em></td>
<td>3B6N (No associated publication)</td>
</tr>
<tr>
<td><em>Salmonella enterica serovar Typhimurium</em></td>
<td>3GHZ (No associated publication)</td>
</tr>
<tr>
<td><em>Shewanella oneidensis</em></td>
<td>1T0A (Ni <em>et al</em>., 2004)</td>
</tr>
<tr>
<td><em>Thermus thermophilus</em></td>
<td>1IV2 (Kishida <em>et al</em>., 2003)</td>
</tr>
<tr>
<td><em>Yersinia pestis</em></td>
<td>3FPI (No associated publication)</td>
</tr>
</tbody>
</table>

In *EcIspF* structures Mg$^{2+}$, positioned by Glu135, fixes the α- and β-phosphate groups of the CDP moiety (Kemp *et al*., 2002; Steinbacher *et al*., 2002). Mg$^{2+}$ is essential for catalysis (Herz *et al*., 2000) and is thought to stabilise the negative charge on the α-phosphate of the CMP leaving group and may help stabilise the transition state complex. The α-phosphate is further positioned by Thr133. Residues in proximity to
the Zn$^{2+}$ site conserved in sequence and structure are expected to contribute to catalysis; these include His34 and Ser35 which form a phosphate binding site for the attacking 2-phosphate group (Steinbacher et al., 2002). Additionally, the His34 side chain positions a loop of ten residues, Pro62 - Ala71, that surround the methylerythritol phosphate moiety. The 2C methyl group makes a favourable hydrophobic contact with highly conserved Ile57 in a structure of EcIspF in complex with CDP-ME, which is a substrate analogue lacking the attacking phosphate. Orientation of the methyl-D-erythritol moiety is achieved by hydrogen bonds from the 3-hydroxy group to main chain carbonyl groups of Phe61 and Ile57 and the product, MEcPP, binds to His34, Ser35, Ile57, Phe61 and Asp63 (Steinbacher et al., 2002). In TtIspF a lysine side chain, Lys132, binds to the bridging β-phosphate further polarising the substrate, however this interaction is not conserved in EcIspF, but is replaced by a water positioned by Thr132. Some of the interactions described here are presented in the corresponding results section (Chapter 3, Figure 3.3 and 3.4).
The essentiality of IspF for growth and survival of *E. coli* has been established (Campbell & Brown, 2002; Campos et al., 2001; Freiberg et al., 2001; Sauret-Güeto et al., 2003). Deletion of the *ispF* gene, formerly referred to as *ygbB*, does not support growth and is fatal (Freiberg et al., 2001). When IspF expression was impaired by deletion of *ispF* and replacement with a copy at a distant and conditionally expressed locus, thereby depleting isoprenoid-precursors, *E. coli* exhibited an aberrant filamentous phenotype indicative of disrupted peptidoglycan synthesis (Section 1.2.1). Additionally, heightened sensitivity to the cidal effect of fosmidomycin and various inhibitors of cell wall biosynthesis was observed (Campbell & Brown, 2002). The essentiality of IspF has been demonstrated in other organisms including, *Bacillus subtilis* (Campbell & Brown, 2002) and *Mycobacterium tuberculosis* (Buetow et al., 2007).
1.3.1 *para*-Aminobenzoic acid biosynthesis

*para*-Aminobenzoic acid (also known as 4-aminobenzoic acid, PABA) is synthesised by many microorganisms and plants since it is required to form folate (Figure 1.7). Folate, also known as vitamin B9, is a generic name for a family of structurally related compounds; essential cofactors in several enzymes. Folates typically function as single carbon unit donors in catalysed reactions (Reviewed, for example, by Benkovic, 1980; Gisondi et al., 2007; Lucock, 2000). Folate is required for the formation of thymidylate and methionine and as such is essential for nucleic acid synthesis, cell growth and division. Unlike plants and most microorganisms, humans do not synthesise their own folate and so it must be obtained in the diet. Folate deficiency can have several injurious effects including certain forms of anaemia and, during pregnancy, it is associated with a range of complications and adverse outcomes including a higher risk of neural tube defects in children (Kirke et al., 1993; Lucock, 2000; Varela-Moreiras et al., 2009).

![Chemical structure of a) para-aminobenzoic acid (PABA). b) folic acid (mono-L-glutamyl), also known as vitamin B9. The para-aminobenzoic acid derived moiety, highlighted in red, links a pterin moiety to L-glutamic acid.](image)

*Figure 1.7 Chemical structure of a) para-aminobenzoic acid (PABA). b) folic acid (mono-L-glutamyl), also known as vitamin B9. The para-aminobenzoic acid derived moiety, highlighted in red, links a pterin moiety to L-glutamic acid.*

The enzyme dihydropteroate synthase (DHPS, EC: 2.5.1.15) links PABA to a GTP-derived pterin moiety, 7,8-dihydro-6-hydroxymethylpterin-pyrophosphate. The product of this reaction, 7,8-dihydropteroate, is subsequently glutamylated and reduced
to form the intracellular pool of available folate (Lucock, 2000; Richey and Brown, 1969). Thus, a PABA-derived moiety has a structural role in the folate molecule (Figure 1.7). In *E. coli* and many other bacteria, PABA is synthesised in two steps starting from chorismate (Nichols *et al.*, 1989). The first step produces 4-amino-4-deoxychorismate (Anderson *et al.*, 1991) by transferring ammonia, derived from glutamine, to chorismate. This reaction is performed by aminodeoxychorismate synthase (EC: 2.6.1.85), a heterodimer of proteins encoded by the *pabA* and *pabB* genes (Parsons *et al.*, 2002). 4-amino-4-deoxychorismate is then converted to PABA with the loss of pyruvic acid (Figure 1.10). 4-amino-4-deoxychorismate lyase (PabC, EC: 4.1.3.38), a pyridoxyl 5'-phosphate (PLP) - dependent enzyme (Figure 1.8), catalyzes this second step (Green & Nichols, 1991; Nichols *et al.*, 1989). In the Gram-positive bacterium *Bacillus subtilis* a third chemical step is required for PABA biosynthesis. Chorismate is first converted to 2-amino-2-deoxyisochorismate before production of 4-amino-4-deoxychorismate, which is then converted to PABA and pyruvate by PabC (Schadt *et al.*, 2009).

![Figure 1.8 PLP forms a covalent link with the amino group of lysine sidechains.](image)

The ε-amino group of a lysine sidechain, typically in the enzyme active-site acts as a nucleophile, attacking PLP to form an enzyme-PLP adduct (Schiff base); with water as a leaving group. This lysine-linked PLP is referred to as the internal aldimine form. Curly arrows indicate the flow of electrons.
1.3.2 Anti-folate drug discovery and sulphonamides as antimicrobial agents

Since PABA is required for folate production and by extension a range of physiological processes in many microorganisms, PABA biosynthesis presents itself as a potential target for antimicrobial drug development. Targeting PABA biosynthesis directly has been chemically validated as a viable approach to identifying antimicrobial compounds, as exemplified by the natural products *atrop*-abyssomicin C (Keller *et al*., 2007; Riedlinger *et al*., 2004) and rubreserine (Camara *et al*., 2012) that both inhibit aminodeoxychorismate synthase and show potent activity. In 2004, Bulloch *et al.* demonstrated the mechanism though which the known anti-bacterial agent (6S)-6-fluoroshikimate targets PABA biosynthesis; it is converted to 2-fluorochorismate *in vivo* by chorismate synthase (EC: 4.2.3.5), which proceeds to irreversibly inhibit aminodeoxychorismate synthase.

Several anti-infective medicines in both current clinical use and historically have achieved their salutary effect through inhibition of folate metabolism in microbes. A large number of sulphonamide drugs were discovered during the 1930s, starting with prontosil - shown to be a pro-drug of the active agent, sulfanilamide (Domagk, 1947). Sulphonamide antimicrobials inhibit DHPS, thereby depleting folate levels (Brown, 1962; Woods, 1940). This is due to their structural similarity to the PABA molecule (Figure 1.9). Prior to the introduction of penicillin, 'sulfa drugs' and their derivatives were highly successful; for example the sulfone drug dapsone was widely used to treat leprosy (Barr, 2010; Shepard, 1969).

Dihydrofolate reductase (DHFR, EC: 1.5.1.3) is another enzyme of folate metabolism that has been targeted for therapeutic benefit. Several DHFR inhibitors, such as trimethoprim and pyrimethamine are used clinically to manage a range of conditions including urinary tract infections, bacterial pneumonia and protozoal
infections. Indeed, DHFR inhibitors are often given in combination with sulphonamide drugs for synergistic effect (Bushby & Hitchings, 1968; Masters et al., 2003); well-established drug combinations include co-trimoxazole (a combination of trimethoprim with sulfamethoxazole) and pyrimethamine with sulfadiazine or sulfadoxine (World Health Organization, List of Essential Medicines, 2011).

![Chemical structures of sulphonamide antimicrobials that inhibit folate biosynthesis.](image)

**Figure 1.9** Chemical structures of sulphonamide antimicrobials that inhibit folate biosynthesis.

Structural studies of the enzymes involved in PABA and folate metabolism can provide significant insight into the mechanism of the reaction catalysed (Nakai et al., 2000, Parsons et al., 2002, Schnell et al., 2004), the molecular nature of drug-resistance (Vanichthanankul et al., 2011; Yun et al., 2012) and can guide early-stage drug discovery initiatives (Chhabra et al., 2012; Eadsforth et al., 2012; Morgan et al., 2011).

### 1.3.3 PabC

4-Amino-4-deoxychorismate lyase, encoded by the pabC gene, catalyzes the formation of PABA from 4-amino-4-deoxychorismate (Figure 1.10) as part of the biosynthetic route to folate (Green & Nichols, 1991; Ye et al., 1990).
Figure 1.10 PabC catalyzes the conversion of 4-amino-4-deoxychorismate to PABA and pyruvate.

Inhibitors of microbial PabC would reduce PABA levels and so deplete the supply of folic acid available to the microorganism. Provided that the microbe cannot obtain sufficient PABA or folate from the cellular environment or by other means these cells would be expected to cease growth and die. This is what is observed in pabC knockouts in *E. coli* (Green *et al.*, 1992), *Pseudomonas aeruginosa* (Hoang *et al.*, 1998) and *Acinetobacter baylyi* (de Berardinis *et al.*, 2008). Combining novel PabC inhibitors with sulphonamides or with other anti-folates such as the dihydrofolate reductase inhibitor trimethoprim, could produce a synergistic antimicrobial effect. New inhibitors of PabC would allow such a hypothesis to be tested.

X-ray crystal structures of PabC from *E. coli* (PDB code: 1ETO; Nakai *et al.*, 2000), *Thermus thermophilus* (2ZGI; Padmanabhan *et al.*, 2009) and *Legionella pneumophila* (3LUL) have been determined. PabC displays a similar overall fold to branched-chain amino acid transferases and D-aminoc acid transferases (Eliot & Kirsch, 2004; Jhee *et al.*, 2000; Nakai *et al.*, 2000). These aminotransferases have been investigated to identify potential inhibitors in support of research for treatments of neurodegenerative disease (Castell *et al.*, 2010; Hu *et al.*, 2006; Lepore *et al.*, 2010; Soper & Manning, 1981). Moreover, since PLP-dependent enzymes provide opportunities for the design of mechanism-based suicide inhibitors (Eliot & Kirsch, 2004) this enzyme family appears an attractive one from the perspective of structure and mechanism-based drug discovery. For example, the anti-trypanosomal medicine
eflornithine (α-difluoromethylornithine) used in the treatment of African sleeping sickness is a suicide inhibitor of the PLP-dependant enzyme ornithine decarboxylase (EC: 4.1.1.17), forming an adduct with the active site cofactor (Grishin et al., 1999).

The absence of the PabC enzyme in humans and its essentiality in various microbes suggests that its inhibition offers the possibility of new therapies targeting a range of microbial infections and structural studies provide useful data to assess the potential of this protein for such early stage drug discovery (Hunter, 2009; O'Rourke et al., 2011).

1.4 Biophysical screening techniques

Bio-layer interferometry (BLI) is an optical technique that can be used to perform label-free fragment screening against an immobilised protein target. The sensors used in the assay are fiber optic tips with a streptavidin-coated protein binding surface and protein target can be biotinylated to facilitate binding to the sensor (Figure 1.11). Following protein loading and capture by the streptavidin surface, the sensors are immersed in wells containing analytes, typically chemical fragments and control compounds, and a binding response is recorded in real-time. A sensor records the interference pattern produced by white light reflected from two layers in the sensor, an internal reference layer and the protein binding surface, or 'bio-layer'. Binding of ligand to the bio-layer increases the optical thickness and shifts the interference pattern, resulting in an apparent shift in wavelength of the reflected light (Wartchow et al., 2011).
Figure 1.11 Principles of BLI. A protein target is immobilised on the biolayer and subjected to incident white light. An optical interference pattern is recorded from light returning from the biolayer and compared to that returning from an internal reference layer. An increase in mass at the biolayer, due to the binding of a chemical fragment for example, will increase the optical thickness, shifting the interference pattern, which results in an apparent shift in wavelength ($\Delta\lambda$) of the reflected light (Wartchow et al., 2011). The OctetRed BLI system is shown inset. Images taken from the ForteBio website: http://www.fortebio.com/bli_technology.html

In differential scanning fluorimetry (DSF), protein unfolding in response to increasing temperature can be measured by monitoring the fluorescence signal of a dye with an affinity for hydrophobic residues (Pantoliano et al., 2001). The dye will be excluded from the hydrophobic interior of the protein and will not fluoresce until the protein exposes hydrophobic residues as it unfolds. The assay is typically performed using an RT-PCR machine to incrementally raise the temperature. Melt curves can be fit to the fluorescence data and the temperature corresponding to the transition midpoint ($T_m$) during unfolding can be calculated. The stabilisation of a protein by ligand or buffer can be inferred from a positive shift in $T_m$ (Ericsson et al., 2006; Niesen et al., 2007). Screening of buffers using differential scanning fluorimetry can also identify conditions to pursue or avoid during crystallisation (Ericsson et al., 2006; Reinhard et al., 2013).
Both of these methods were used in these studies; buffer screening with differential scanning fluorimetry (Section 2.1.4) and screening of fragment libraries with BLI (Section 2.1.5).

1.5 Aims - Part one

This study sought to determine the crystal structures of potential drug targets from human pathogens to inform early stage drug discovery. Studies of recombinant IspF, an essential enzyme on the MEP pathway; from \textit{B. cenocepacia} and \textit{P. falciparum} were pursued. Fragment screening was carried out to identify novel ligands. The suitability of the enzyme from these organisms as a platform for ligand co-crystallisation studies was assessed. Additionally, the structure of PabC from \textit{P. aeruginosa}, an essential enzyme required for the synthesis of PABA, was determined to provide insight into catalytic mechanism and to facilitate studies of this enzyme in antimicrobial drug discovery.
CHAPTER 2

MATERIALS AND METHODS
Materials and methods

2.1.1 Reagents, bacterial strains, growth media and equipment

All chemical reagents used were of the highest possible purity available and were sourced from VWR and Sigma-Aldrich unless otherwise stated. *E. coli* strains were sourced from Life Technologies, Promega and Stratagene. PCR amplifications of DNA were performed using a programmable thermal cycler (GeneAmp 2700, Applied Biosystems). Oligonucleotide primers for PCR were sourced from Life Technologies. DNA ligase and restriction enzymes were sourced from Fermentas, New England Biolabs and Promega. All sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed using NuPAGE Bis-Tris pre-cast gels sourced from Life Technologies.

Plasmid preparations and DNA gel-extraction were performed using Qiagen kits (QIAprep spin miniprep, QIAquick gel extraction). DNA electrophoresis was performed in gels prepared to contain 10% w/v agarose (VWR) and 0.0002% w/v ethidium bromide (10 mg mL\(^{-1}\) stock, Electran, VWR BDH Prolabo). DNA sequencing was performed by DNA Sequencing & Services (University of Dundee) using Applied Biosystems Big-Dye Ver 3.1 chemistry on an Applied Biosystems model 3730 automated capillary DNA sequencer. All LB bacterial growth media used in these studies were prepared and autoclaved by the Media Services staff (University of Dundee). Shake-flask bacterial culture was performed using either an incubation shaker (Multitron, Infors-HT) or a room temperature orbital shaker (KS501 Digital, IKA). Both electrospray ionisation mass spectrometry (ESI-MS), and matrix-assisted laser desorption/ionisation - time-of-flight mass spectrometry (MALDITOF MS) were performed by the Proteomics and Mass Spectrometry Facility (FingerPrints, University of Dundee). In particular, MALDI-TOF MS analyses were performed by Samantha Kosto on an AB Sciex Voyager DE-STR system using a sinapinic acid matrix. Samples
for MALDI-TOF MS were diluted in de-ionised water and loaded onto the matrix in a concentration range between 2-5 pmol.

Table 2.1 List of key details for the proteins reported in this thesis (Part 1)

<table>
<thead>
<tr>
<th></th>
<th>BcIspF</th>
<th>PfIspF</th>
<th>PaPabC</th>
</tr>
</thead>
<tbody>
<tr>
<td>UniProt code</td>
<td>B4EC22</td>
<td>P62368</td>
<td>Q9HZN6</td>
</tr>
<tr>
<td>Enzyme nomenclature</td>
<td>2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase</td>
<td>2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase</td>
<td>4-amino-4-deoxychorismate lyase</td>
</tr>
<tr>
<td>E.C. number</td>
<td>4.6.1.12</td>
<td>4.6.1.12</td>
<td>4.1.3.38</td>
</tr>
<tr>
<td>Source organism</td>
<td>Burkholderia cenocepa, strain J2315</td>
<td>Plasmodium falciparum, isolate 3D7</td>
<td>Pseudomonas aeruginosa, strain PAO1</td>
</tr>
<tr>
<td>Number of amino acids</td>
<td>182</td>
<td>184</td>
<td>292</td>
</tr>
<tr>
<td>Molecular mass (Da)</td>
<td>19328.8</td>
<td>20545.5</td>
<td>32291.9</td>
</tr>
<tr>
<td>Theoretical isoelectric point</td>
<td>6.4</td>
<td>9.0</td>
<td>6.0</td>
</tr>
<tr>
<td>Theoretical extinction coefficient (ε)</td>
<td>5960 (Bradford assay was used to estimate concentration)</td>
<td>11920</td>
<td>33710</td>
</tr>
<tr>
<td>PDB codes</td>
<td>4C8G (CMP complex), 4C8E (2 x CMP complex), 4C81 (citrate complex)</td>
<td>4C81 (CDP complex), 4C82 (holo-structure)</td>
<td>2Y4R (PLP complex)</td>
</tr>
</tbody>
</table>

2.1.2 Gene cloning and plasmid preparation

The genes encoding the proteins described in this thesis were amplified from genomic DNA by PCR (Table 2.2), with the exception of EcIspF which was provided by Dr. Thomas Eadsforth and PfIspF which was purchased as a synthetic gene (below). PCR products were blunt-end ligated into a TOPO cloning vector (Life Technologies), then ligated into an expression vector, pET15b-TEV, that encoded the protein-of-interest bearing an N-terminal hexahistidine purification tag (His-tag) linked via a tobacco-etch virus (TEV) protease cleavage site.
Table 2.2 Oligonucleotide primers for PCR

<table>
<thead>
<tr>
<th></th>
<th>Primer Sequence</th>
<th>Strain</th>
</tr>
</thead>
<tbody>
<tr>
<td>BcIspF</td>
<td>5'- CAT ATG GAC TTC AGA ATC GGA CAA GG -3'</td>
<td>J2315</td>
</tr>
<tr>
<td></td>
<td>5'- GGA CCT CAG CCG CCC TGC TTC ACC -3'</td>
<td></td>
</tr>
<tr>
<td>PaPabC</td>
<td>5'- CAT ATG CTG GAC TGG GTC -3'</td>
<td>PAO1</td>
</tr>
<tr>
<td></td>
<td>5'- GGA TCC TCA GAA ATC CAG GTC G -3'</td>
<td></td>
</tr>
</tbody>
</table>

The gene encoding IspF from *P. falciparum* isolate 3D7, GenBank accession number: XP_001349603, was synthesised (Genscript, Boston) with codons optimised for recombinant expression in *E. coli*. *NdeI* and *BamHI* restriction sites were included at the 5' and 3' ends of the gene. The sequence encoding residues 1-59, a predicted apicoplast targeting sequence, was excluded and the codon for Cys60 was replaced with one for serine. The synthetic gene was provided by Dr. Scott Cameron in a pUC57 shuttle vector and was subcloned into the pET15b-TEV vector.

The genes encoding *BcIspF* and *EcIspF* were also ligated into a pET15b-TEV vector that had been further modified to append an N-terminal biotin acceptor peptide to the protein in addition to the protease-cleavable His-tag. These vectors were used in the production of biotinylated IspF proteins for fragment screening.

### 2.1.3 Protein production

Recombinant production was performed in *E. coli* BL-21 (DE3) Gold for all proteins. Strains were transformed with expression plasmids by heat-shock. Transformants were selected following outgrowth at 37 °C, on LB agar plates supplemented with the antibiotics chloramphenicol and/or carbenicillin. 1 L shake-flask cultures were grown in LB at 37 °C and monitored by UV spectrophotometry (BioPhotometer, Eppendorf). When the optical density at 600 nm (OD$_{600}$) reached 0.6, expression was induced with
0.1 mM IPTG - with the exception of EcIspF, which was induced at an OD$_{600}$ of 1.0. Culturing was continued for a further 16 hours at 18 °C, or 22 °C in the case of BcIspF, before cells were harvested.

### 2.1.4 Differential scanning fluorimetry and buffer screening

All assays reported here were performed in 96-well plates (Thermo-Fast, Thermo Scientific) on a Stratagene-Mx3005P RT-PCR machine. SYPRO Orange (Life Technologies) was used as the fluorescent dye. Buffer screening was carried out for all proteins reported here; with the exception of the IspF samples, which were found to give either irreproducible (PfIspF) or uninterpretable (BcIspF) responses during assessment.

### 2.1.5 Fragment screening by bio-layer interferometry

A fragment screen using the Drug Discovery Unit (DDU, University of Dundee) fragment library was performed by BLI (Octet Red system, ForteBio) on two enzymes; BcIspF and EcIspF. BcIspF was screened first against the original fragment set (652 commercially available compounds). EcIspF was screened later against a revised version of the fragment library that had been changed to remove non-specific binders and to add novel fragments with different scaffolds (Table 2.3; Dr. David Robinson, personal communication).
Table 2.3 BLI fragment screen assay conditions

<table>
<thead>
<tr>
<th></th>
<th>Original DDU fragment library (652 fragments at 200 µM in 100% DMSO)</th>
<th>Revised DDU fragment library (658 fragments at 200 µM in 100% DMSO)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Biotinylated-BcIspF</strong></td>
<td>Buffer: 20 mM NaCl, 100 µM MgCl₂, 100 µg ZnCl₂, 20 mM Tris-HCl, pH 7.5</td>
<td>Buffer: 50 mM NaCl, 100 µM MgCl₂, 100 µM ZnCl₂, 50 mM Tris-HCl, pH 7.7</td>
</tr>
<tr>
<td><strong>Biotinylated-EcIspF</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Both BcIspF and EcIspF were loaded onto biosensors (SuperStreptavidin, ForteBio) at 1 µM and screened at 25 °C, alongside double-referenced biosensors (one set referenced with assay buffer, and one set with blocked streptavidin). Known IspF ligands, CMP and CDP, were used as positive controls. Fragment screens were conducted after immobilising target proteins for ten minutes, blocking unoccupied binding sites with the biotin analogue biocytin at 10 µg mL⁻¹ for one minute, followed by surface washing and equilibration of the biosensors in assay buffer for five minutes. Coefficient-of-determination ($R^2$) values were calculated for sensorgram fitting curves and used as a measure of reliability throughout the screens.

### 2.1.6 Screening for crystallisation conditions

An automated liquid-handling system (Phoenix, Art Robbins) was used in screening for crystallisation conditions in combination with commercially available screens (JCSG-plus, Molecular Dimensions; AmSO₄ suite, PEGs suite, Classics suite, Qiagen). All samples were filtered through a 0.1 µm PVDF filter (Ultrafree, Millipore) immediately prior to screening. The protein samples were dispensed (100 nL drops) into 96-well sitting-drop plates (MRC-2, Molecular Dimensions). Reservoirs were dispensed in a 1:1 or 1:2 ratio with the protein sample (100 nL protein + 100 nL reservoir, or, 100 nL...
protein + 200 nL reservoir). Plates were then sealed and stored at room temperature in a benchtop protein crystal drop imager (Minstrel DT system, Rigaku).

Crystallisation plates were inspected at regular intervals for up to one month. Productive conditions identified in initial screens were optimised by scaling-up to larger volumes (1-2 µl) in hanging-drop vapour diffusion trays (24-well Linbro plates, Hampton Research). The crystallisation landscape around identified conditions was investigated by varying the concentrations of various reservoir components and/or the protein concentration, pH and temperature (20 °C or 4 °C).

2.2.1 Purification of PfIspF

Cells were harvested by centrifugation at 3000 g for 30 minutes at 18 °C and re-suspended in 100 mM KCl, 100 mM L-arginine, 50 mM CHES, pH 9.5 (buffer A). DNaseI and a protease inhibitor tablet (EDTA-free, Roche) were added prior to cell lysis by passage through a French pressure cell. The cell lysate was clarified by ultracentrifugation at 40000 g for 30 minutes at 18 °C (Avanti centrifuge, Beckmann) and supernatant was syringe-filtered through a pre-filter followed by a 0.2 µm filter (Sartorius).

*Pf*IspF precipitates during purification at temperatures below 18 °C (Dr. Kalinowska-Tluścik, personal communication). Therefore all procedures described were performed at room temperature unless otherwise stated. The cell lysate was loaded onto a nickel ion affinity chromatography column (5 mL HisTrap HP, GE Healthcare) with the assistance of an FPLC system (Äkta Explorer, GE Healthcare). A linear concentration gradient of imidazole, from 0 - 1M, was applied to elute *Pf*IspF (which eluted at 200 mM imidazole). Protein sample was then dialysed against fresh buffer A. The purification tag (His-tag) was cleaved with His-tagged TEV protease at 20 °C for 16 hours. The protease-treated sample was re-applied to a nickel ion affinity
chromatography column, which eluted the cleaved protein while removing the TEV protease and remaining uncleaved PfIspF. Fractions were analysed by SDS-PAGE and those containing PfIspF were pooled. The protein was further purified on a size-exclusion chromatography column (Superdex 200 26/60, GE Healthcare) equilibrated with buffer A. This column had previously been calibrated (by Sharon Shepherd) with molecular weight standards; thyroglobulin (670 kDa), γ-globulin (158 kDa), ovalbumin (44 kDa), myoglobin (17 kDa), and vitamin B₁₂ (1.35 kDa); (Bio-Rad). The high level of PfIspF purity was confirmed by SDS-PAGE and MALDI-TOF MS. The sample was concentrated to 6 mg mL⁻¹ using a centrifugal filter (Vivaspin 20, Sartorius) to provide a stock solution for crystallisation. Protein concentration was estimated spectrophotometrically; a theoretical extinction coefficient, ε: 11920 M⁻¹ cm⁻¹ at 280 nm, was calculated by the ProtParam server (Gasteiger et al., 2005).

### 2.2.2 Purification of BcIspF

BcIspF was purified in the manner described for PfIspF, except that buffer B (100 mM NaCl, 100 mM Tris-HCl, pH 7.5) was used throughout the process and the His-tag was not cleaved off the protein. Further, the marked temperature sensitivity observed for PfIspF was not seen for BcIspF and samples were kept either on ice or at 4 °C throughout.

### 2.2.3 Purification of biotinylated BcIspF and EcIspF for BLI

Cultures containing BcIspF were harvested by centrifugation and re-suspended in buffer D (150 mM NaCl, 50 mM Tris-HCl, pH 8.0). To this was added one protease inhibitor tablet (EDTA-free, Roche) and DNAse I. The re-suspended cells were lysed and the high-speed supernatant was recovered as described for PfIspF. The supernatant was syringe-filtered and loaded on to a nickel ion affinity chromatography column. The
protein was eluted from the column with a linear gradient of imidazole. Protein sample was dialysed into fresh buffer D then treated with TEV protease (16 hours at 4 °C). A sample was taken for MALDI-TOF MS before and after treatment with protease.

Protease-treated protein was passed through a second nickel ion affinity chromatography column and protein cleaved from its purification tag, was collected in the flow-through. This cleaved protein was dialysed into biotinylation buffer (10 mM Tris-HCl, 200 mM KCl, 5 mM MgCl₂). Following dialysis, the protein was prepared for biotinylation by being diluted to <10 µM in biotinylation buffer. D-biotin was added to a final concentration of 100 µM, and the biotinylation reaction was started with the addition of 500 µM ATP and 1 µM of the biotin ligase, BirA (His-tagged, provided by Dr. Vincent Rao). The reaction was carried out for four hours at 20 °C. The protein eluted in the flow-through of a third nickel ion affinity chromatography column, removing the BirA enzyme. Biotinylated-BcIspF was concentrated and exchanged into a buffer for size-exclusion chromatography (20 mM NaCl, 20 mM Tris-HCl, pH 7.5) by centrifugal filtration, to remove residual biotin and ATP. Sample was loaded on to a size-exclusion chromatography column (Superdex 200 26/60, GE Healthcare). The peak was collected and a sample was sent for MALDI-TOF MS analysis. Protein, diluted to 5 mg mL⁻¹ in size-exclusion chromatography buffer and supplemented with 10% glycerol, was flash-cooled in liquid nitrogen and stored at -80 °C for later use on the Octet Red BLI system. EcIspF was purified and biotinylated as described above for BcIspF, except that the buffer condition used throughout purification was different (50 mM NaCl, 50 mM Tris-HCl, pH 7.7), and that the His-tag was kept.
2.2.4 Purification of PaPabC

PaPabC was purified in the manner described for PfIspF, except that a different buffer was used throughout the process (500 mM NaCl, 100 mM HEPES, pH 7.5) and that the His-tag was kept. Samples were kept either on ice or at 4 °C throughout. Sample purity and mass were confirmed by electrospray ionisation mass spectrometry.

2.3.1 Crystallisation of PfIspF in complex with CDP

A stock solution of PfIspF was prepared at 6 mg mL\(^{-1}\) in buffer A supplemented with 2 mM MgCl\(_2\). CDP disodium salt was added to a final concentration of 2 mM. Crystals were obtained by mixing protein (2 µl) with reservoir solution (2 µl) at 20 °C. The reservoir solution consisted of 1.8-2.5 M (NH\(_4\))\(_2\)SO\(_4\), 5 mM ZnCl\(_2\) and 100 mM Bis-Tris, pH 5.5. Prisms (0.15 x 0.15 x 0.15 mm) appeared within 2-4 days. Crystals were cryo-protected with a saturated sucrose solution and data were collected at the European Synchrotron Radiation Facility (ESRF, Grenoble).

2.3.2 Crystallisation of holo-PfIspF

PfIspF was prepared as before except that fosmidomycin, an hydroxamate compound (Figure 1.5), was added to the protein sample (20 mM) prior to crystallisation. Crystals of similar morphology were obtained as described above and a dataset was collected under a gaseous nitrogen cryostream at \(-173\) °C on an in-house rotating-anode X-ray source (MicroMax 007 HF, Rigaku) with an image plate detector (RAXIS IV\(^{++}\)).

Despite the presence of fosmidomycin, a putative ligand, no interpretable electron density was observed in the resulting maps (Results, Chapter 3), producing a holo-PfIspF structure.
2.3.3 Crystallisation of BcIspF in complex with CMP

BcIspF was prepared at 5 mg mL\(^{-1}\) in buffer B supplemented with 2 mM MgCl\(_2\), to which was added CMP disodium salt to a final concentration of 20 mM. A second sample was prepared at 10 mg mL\(^{-1}\) in buffer C (100 mM NaCl, 2 mM MgCl\(_2\), 100 mM sodium formate, pH 5.0) to which CMP was also added (2 mM final concentration).

The BcIspF sample in buffer B was crystallised by mixing (2 \(\mu\)L) with reservoir (2 \(\mu\)L) containing 25% w/v PEG 3350, 2.2 M (NH\(_4\))\(_2\)SO\(_4\), 0.1 M sodium formate, pH 5.0, and 1% v/v dioxane. Orthorhombic blocks reached maximum size within ten days. A crystal of approximate dimensions 0.15 x 0.10 x 0.10 mm was used for data collection. Buffer B supplemented with 20% v/v glycerol was used as a cryoprotectant.

The second BcIspF sample, in buffer C, was crystallised by mixing (2 \(\mu\)L) with reservoir (4 \(\mu\)L) containing 5% w/v PEG 1000, 33% v/v ethanol, 0.1 M Na\(_2\)HPO\(_4\), pH 4.0, and 2% v/v dioxane. Prisms grew within six days. A crystal of approximate dimensions 0.25 x 0.10 x 0.10 mm was used for data collection. A mixture of buffer C with reservoir (1:2), replacing PEG 1000 with PEG 4000 at 25% w/v, was used as a cryoprotectant.

2.3.4 Crystallisation of BcIspF in complex with citrate

BcIspF at 10mg mL\(^{-1}\) in buffer C without MgCl\(_2\) present was crystallised by mixing (1.5 \(\mu\)L) with reservoir (3 \(\mu\)L) containing 5% w/v PEG 1000, 36% v/v ethanol, 0.1 M Na\(_2\)HPO\(_4\) and 0.1 M citric acid, pH 4.2. Prisms grew within 4 days, belonging to a different crystal form. A crystal of approximate dimensions 0.4 x 0.10 x 0.10 mm was used for data collection. A mixture of sample buffer with reservoir (1:2), replacing PEG 1000 with PEG 4000 at 25% w/v, was used as a cryoprotectant. Datasets for
BcIspF crystals were collected on the in-house X-ray system as described for PfIspF (Section 2.3.2).

2.3.5 Crystallisation of PaPabC

To PaPabC at 33 mg mL\(^{-1}\) in sample buffer (100 mM HEPES, 500 mM NaCl pH 7.5), PLP and PABA were added (final concentration: 0.1 mM and 10 mM respectively). PLP increased the thermal stability of PaPabC in a dose-dependent manner; the optimum concentration was determined by DSF. PaPabC was crystallised by mixing sample (1 \(\mu\)L) with reservoir (1 \(\mu\)L) containing 10% w/v PEG 400, 1.8 M (NH\(_4\))\(_2\)SO\(_4\) and 100 mM MES, pH 6.5. Clumps of pale yellow, orthorhombic blocks grew over one week. Mother liquor supplemented with 25% v/v glycerol was used as a cryo-protectant. A crystal fragment (0.3 \(\times\) 0.2 \(\times\) 0.1 mm) was flash-cooled in liquid nitrogen and sent for data collection at a synchrotron-radiation source (ESRF).

2.4 Data processing, structure solution and refinement

Data were indexed and integrated using either MOSFLM (Leslie, 2006; Battye et al., 2011) or XDS (Kabsch, 2010) and scaled using SCALA (Evans, 2006) with 5% of data kept for \(R_{\text{free}}\) calculations. All structures were solved by molecular replacement. Search models for molecular replacement used coordinates from orthologues in E. coli (PDB code: 1GX1; \(~33\%\) identity, Kemp et al., 2002) and B. pseudomallei (PDB code: 3F0G; 63\%, Begley et al., 2011) for PfIspF and BcIspF respectively, edited to remove all non-protein atoms. The search model for PaPabC was prepared by generating a poly-alanine model of the E. coli PabC structure (35\% identity, Nakai et al., 2000) in CHAINSAW (Stein, 2008). Molecular replacement calculations were performed using
either PHASER (McCoy et al., 2007; Collaborative Computational Project, Number 4, 1994) or MOLREP (Vagin and Teplyakov, 1997).

Refinements were performed in REFMAC5 (Murshudov et al., 2011) and models were inspected and manipulated in COOT (Emsley et al., 2010). Strict non-crystallographic symmetry (NCS) restraints were employed in the early stages of refinement and then removed towards the end. Greater weighting was given to geometric parameters than default values during refinement of medium resolution structures of BcIspF. The validation tools in COOT and the MolProbity server (Chen et al., 2010) were used to assess model geometry. Comparisons with structures in the PDB were performed using the DALI server (Holm & Rosenström, 2010). AREAIMOL (Collaborative Computational Project, Number 4, 1994) and the Protein Interfaces, Surfaces & Assemblies server, (Krissinel & Henrick, 2007) were used to calculate surface and interface areas. Figures of proteins were prepared using PyMOL molecular graphics software (DeLano, 2002).
CHAPTER 3

IspF; STRUCTURAL ANALYSIS and

FRAGMENT SCREENING
3.1 \textit{Pf}IspF and \textit{Bc}IspF results

3.1.1 Structure of IspF

Structures were determined of \textit{Bc}IspF in complex with either CMP or citrate, and of \textit{Pf}IspF in complex with CDP, as well as a \textit{holo-Pf}IspF structure (where \\textit{holo-} refers to the protein in complex with the Zn$^{2+}$ cofactor, without any other ligands present in the active site). All structures are at a resolution of 2.0 Å or better. Crystallographic statistics are presented in Tables 3.1 and 3.2. Sequence alignments and structural overlays of IspF from \textit{P. falciparum}, \textit{B. cenocepacia}, \textit{B. pseudomallei} and \textit{E. coli} were inspected to inform on conservation of residues and interactions, and an alignment of \textit{Bc}IspF with \textit{Pf}IspF is shown in Figure 3.1.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{alignment.png}
\caption{Sequence alignment of \textit{Bc}IspF and \textit{Pf}IspF. The alignment has been annotated with elements of secondary structure. Only the \textit{Bc}IspF sequence has been numbered. Strictly conserved residues between the two species are highlighted in black. Residues that bind Zn$^{2+}$ are marked with a yellow star, those that bind at position 1 with a grey disk and those at position 2 with an orange triangle. Residues that bind MEcPP in \textit{Ec}IspF (Steinbacher et al., 2002) are marked with a magenta triangle. A conserved glutamate that binds either Mg$^{2+}$ or Mn$^{2+}$ in \textit{Ec}IspF structures is highlighted with a blue star. This figure was prepared using \textit{ALINE} (Bond & Schüttelkopf, 2009).}
\end{figure}
The calculated masses of monomeric \( BcIspF \) and \( PfIspF \) are 19.3 kDa and 20.5 kDa respectively. \( IspF \) forms a homotrimer, both in the crystal structures and in solution; size-exclusion chromatography returns an estimate of 57 kDa for each. \( IspF \) from both species adopts the established fold reported from other organisms; a core region containing four \( \beta \)-strands, flanked by \( \alpha \)-helices (Section 1.2.5, Figure 3.2). The accessible surface area (ASA) of an \( IspF \) subunit averages out as approximately 8130 Å\(^2\) for \( BcIspF \) and 8680 Å\(^2\) for \( PfIspF \); for comparison the value for \( EcIspF \) is 7900 Å\(^2\).

A search for structurally-related proteins in the PDB using the DALI server (Holm & Rosenström, 2010) showed that \( PfIspF \) aligns 143 residues with an RMSD of 0.8 Å, and 151 residues with an RMSD of 1.2 Å with \( IspF \) from \( P. \) \textit{vivax} and \( A. \) \textit{thaliana} respectively. \( BcIspF \) aligns 151 residues with an RMSD of 0.4 Å to \( IspF \) from \( B. \) \textit{pseudomallei}. In \( BcIspF \) structures the asymmetric unit consists of three subunits, chain A, B, C; related by a three-fold non-crystallographic axis of symmetry. NCS is high with least squares overlays of \( C_\alpha \) positions in the range 0.2 - 0.3 Å; nonetheless some differences are present and these are described below. One subunit is present in the asymmetric unit of \( PfIspF \) and the trimer is generated by the symmetry operations (-x+y, -x, z) and (y, y-x, z). A least squares superimposition of \( C_\alpha \) positions from both \( PfIspF \) structures matches all 156 residues with an RMSD of 0.3 Å. Since there are no pronounced differences between the CDP bound and \textit{holo-} \( PfIspF \) structures, only the former is discussed. In 2007, a \textit{holo-}structure of \( IspF \) to a resolution of 2.3 Å from \( P. \) \textit{vivax}, a closely related species to \( P. \) \textit{falciparum}, was deposited with the PDB without an associated publication (code: 3B6N). Although \( Zn^{2+} \) has been modelled at the active site, a water molecule most likely occupies the metal binding site instead, or the site is only partially occupied, since the corresponding \( B \)-value is 2.6 times greater than the average \( B \)-value of the protein atoms and the side chain of one of the two histidines that
coordinate the metal in other structures is in an unsuitable orientation. In contrast, in *Pf*IspF the *B*-value for Zn$^{2+}$ is within 30% of the average *B*. Further, *Pf*IspF provides a higher resolution structure from a *Plasmodium* species.

![Image](image1.png)

**Figure 3.2 Overall structure of IspF.** IspF forms a homotrimer. a) Axial view. b) Side view. The figure has been drawn using the coordinates for *Pf*IspF in complex with CDP; a trimer has been generated from symmetry-related subunits. Subunits (displayed as ribbons) are coloured blue, orange and green respectively. Zn$^{2+}$ cations are displayed as grey spheres, sulfate and CDP are displayed as sticks. c) and d) respectively, similarly orientated. The location of unmodelled loops is highlighted with arrows. Electron density is absent for a loop between α2 and α3 that covers the active site in both *Pf*IspF structures and the *Be*IspF complex with a single CMP in the active
site, spanning Asp144 - Asn151 and Phe63 - Ala73 respectively. In contrast, these loop residues could be modelled in the BcIspF structure where two molecules of CMP are present. The loop residues have elevated B-factors, 2.4 times greater than the average B-value for the entire the protein. The same loop could only be modelled for chain A in the BcIspF-citrate complex. These observations are consistent with the inference that the loop that covers the active site is flexible (Kemp et al., 2002; Steinbacher et al., 2002). Such flexibility has been speculated to have a role in catalysis (Lehmann et al., 2002). PflspF contains a 19 residue sequence at the N-terminus, Lys76 - Glu94, that does not align with sequences from other organisms (Figure 3.1). Residues Tyr82 - Phe97, which overlap with this region, are unmodelled in PflspF because corresponding electron density is absent. The highly conserved His115 and Ser116 that are located on a six residue loop above the Zn$^{2+}$ binding site (Figure 3.1) are also flexible and have elevated B-factors in all structures except in one chain of a BcIspF-CMP structure where they are unmodelled.

The active site is located in a cleft formed by adjacent subunits (Section 1.2.5). Zn$^{2+}$, which plays a functional role in catalysis is present in all structures and coordinated by the side chains of Asp10, His12 and His44 in BcIspF. In EcIspF a highly conserved glutamate, Glu135, helps coordinate either a Mg$^{2+}$ or Mn$^{2+}$ together with the bridging phosphates of the substrate or CDP (Kemp et al., 2002; Steinbacher et al., 2002). However, no metal is seen at this site in any structure reported here.

A hydrophobic cavity is present at the trimeric core of IspF. In BcIspF the cavity is lined by the side chains of Tyr9, Val11, Ile101 and Tyr141, and the floor of the cavity is formed by a ring of hydrophobic residues; Val3 and Leu155 from all three subunits. In PflspF the side chains of Phe230 form the floor of the cavity. Anions such as sulfate, phosphate or pyrophosphate bind at the entrance to this cavity in several IspF structures, for example phosphate is observed in IspF from Arabidopsis thaliana (PDB...
In the structure of \textit{BcIspF} crystallised with (NH$_4$)$_2$SO$_4$ as a precipitant, a sulfate is coordinated by Arg144. In \textit{PfIspF}, a sulfate at a similar site is coordinated by the main chain amides of Val220 from all three subunits. In structures of \textit{BcIspF} in which Na$_2$HPO$_4$ was the precipitant, Tyr141 amide groups bind a phosphate in a similar fashion. Isoprenoid-species such as geranyl- or farnesyl-pyrophosphate have been observed in the cavity that is present in IspF structures from other species, and a possible role in feedback regulation has been suggested (Kemp \textit{et al.}, 2005; Ni \textit{et al.}, 2004; Sgraja \textit{et al.}, 2005). A molecule of di(hydroxyethyl)ether; a likely decomposition product or impurity of PEG 3350 used in crystallisation, occupies the cavity in one of the \textit{BcIspF} structures (Table 3.2), possibly mimicking a prenyl chain. A solvated Mg$^{2+}$ is present further down the cavity in one of the \textit{BcIspF} structures, forming interactions similar to IspF from \textit{B. pseudomallei} (Begley \textit{et al.}, 2011). A shell of water molecules surround Mg$^{2+}$ in an octahedral coordination geometry. These waters are bound by Glu149, and waters in the second shell bind to the hydroxyl groups of Tyr9 and Tyr141. A Mg$^{2+}$ at a similar position is coordinated directly by the side chains of the conserved Glu149 from all three subunits in IspF from \textit{Salmonella enterica} serovar Typhimurium (PDB code: 3GHZ) and \textit{E. coli} (Ramsden \textit{et al.}, 2009).
### Table 3.1 Crystallographic statistics for PfIspF

<table>
<thead>
<tr>
<th>Structure</th>
<th>PfIspF - CDP complex</th>
<th>holo - PfIspF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>H3</td>
<td>H3</td>
</tr>
<tr>
<td>Unit cell dimensions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a, b, c) (Å)</td>
<td>84.9, 84.9, 101.0</td>
<td>84.2, 84.2, 102.5</td>
</tr>
<tr>
<td>Resolution range* (Å)</td>
<td>41.63-1.60 (1.69-1.60)</td>
<td>29.71-2.00 (2.11-2.00)</td>
</tr>
<tr>
<td>No. reflections</td>
<td>181781 (16968)</td>
<td>41802 (5780)</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>35607 (5194)</td>
<td>18068 (2622)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.4 (99.3)</td>
<td>98.6 (97.0)</td>
</tr>
<tr>
<td>(R_{merge}^b) (%)</td>
<td>9.3 (42.0)</td>
<td>9.4 (48.2)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>5.1 (3.3)</td>
<td>2.3 (2.2)</td>
</tr>
<tr>
<td>(&lt;I&gt;/\sigma(I)&gt;^c</td>
<td>11.1 (2.6)</td>
<td>5.3 (2.0)</td>
</tr>
<tr>
<td>Wilson B (Å^2)</td>
<td>24.4</td>
<td>32.0</td>
</tr>
<tr>
<td>(R_{work}/R_{free}^d)</td>
<td>14.4, 16.5</td>
<td>19.4, 22.6</td>
</tr>
<tr>
<td>Number of residues / waters / ligands</td>
<td>156 residues, 85 waters, 1 CDP, 2 Zn(^{2+}), 3 Cl(^{-}), 1 SO(_4)(^{2-})</td>
<td>156 residues, 94 waters, 2 Zn(^{2+}), 1 SO(_4)(^{2-})</td>
</tr>
<tr>
<td>RMSD bond lengths (Å) / bond angles* (°)</td>
<td>0.005 / 1.000</td>
<td>0.019 / 2.087</td>
</tr>
<tr>
<td>Mean B-factors (Å^2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Main chain, side chain</td>
<td>25.1, 33.8</td>
<td>34.2, 39.5</td>
</tr>
<tr>
<td>Waters</td>
<td>35.2</td>
<td>47.8</td>
</tr>
<tr>
<td>Ligands</td>
<td>31.5 (CDP)</td>
<td>44.5 (Zn(^{2+}), Cl(^{-}), SO(_4)(^{2-}))</td>
</tr>
<tr>
<td>35.8 (Zn(^{2+}), Cl(^{-}), SO(_4)(^{2-}))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ramachandran plot</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favoured regions (%)</td>
<td>98.6</td>
<td>95.3</td>
</tr>
<tr>
<td>Allowed regions (%)</td>
<td>0.7</td>
<td>4.0</td>
</tr>
<tr>
<td>Outliers</td>
<td>1 outlier: Asp79</td>
<td>1 outlier: Asp79</td>
</tr>
</tbody>
</table>

\(^a\) Values in parentheses refer to the highest resolution shell. \(^b\) \(R_{merge} = \frac{\sum_{hkl} \sum_{i} |I(hkl)| - <I(hkl)>|}{\sum_{hkl} \sum_{i} I(hkl)}\); where \(I(hkl)\) is the intensity of the \(i\)th measurement of reflection \(hkl\) and <I(hkl)> is the mean value of I(hkl) for all \(i\) measurements. \(^c\) \(R_{work} = \frac{\sum_{hkl} |F_o| - |F_c|}{\sum_{hkl} |F_o|}\), where \(F_o\) is the observed structure factor and \(F_c\) is the calculated structure factor. \(^d\) \(R_{free}\) is the same as \(R_{work}\) except calculated with a subset, 5%, of data that are excluded from the refinement calculations. \(^e\) (Engh & Huber, 1991). These considerations apply to all tables of crystallographic statistics presented in this thesis.
Table 3.2 Crystallographic statistics for BcIspF

<table>
<thead>
<tr>
<th>Structure</th>
<th>BcIspF - 1 x CMP complex</th>
<th>BcIspF - 2 x CMP complex</th>
<th>BcIspF - citrate complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>C2</td>
<td>C2</td>
<td>C2</td>
</tr>
<tr>
<td>Unit cell dimensions $a$, $b$, $c$ (Å), $β$</td>
<td>97.2, 88.6, 72.4, 103.9°</td>
<td>97.9, 89.3, 73.0, 104.2°</td>
<td>131.7, 52.5, 72.1, 94.6°</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>29.84-2.00 (2.10-2.00)</td>
<td>19.32-1.90 (2.00-1.90)</td>
<td>40.90-2.00 (2.10-2.00)</td>
</tr>
<tr>
<td>No. reflections</td>
<td>98451 (13335)</td>
<td>164120 (24760)</td>
<td>119715 (15957)</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>38409 (5579)</td>
<td>47918 (6972)</td>
<td>33460 (4696)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>96.0 (95.4)</td>
<td>99.9 (100.0)</td>
<td>99.4 (96.1)</td>
</tr>
<tr>
<td>$R_{merge}$ (%)</td>
<td>11.0 (32.2)</td>
<td>13.8 (32.3)</td>
<td>3.2 (7.8)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>2.6 (2.4)</td>
<td>3.4 (3.6)</td>
<td>3.6 (3.4)</td>
</tr>
<tr>
<td>$&lt; I/σ(I)&gt;$</td>
<td>5.9 (2.7)</td>
<td>6.2 (3.1)</td>
<td>23.9 (12.8)</td>
</tr>
<tr>
<td>Wilson $B$ (Å$^2$)</td>
<td>16.5</td>
<td>16.3</td>
<td>18.7</td>
</tr>
<tr>
<td>$R_{merge}/R_{free}$ (%)</td>
<td>23.7, 28.9</td>
<td>20.4, 25.1</td>
<td>15.6, 19.8</td>
</tr>
<tr>
<td>Number of residues / waters / ligands</td>
<td>458 residues, 173 waters, 3 CMP, 3 Zn$^{2+}$, 1 Mg$^{2+}$, 1 PO$_4^{3-}$</td>
<td>483 residues, 420 waters, 6 CMP, 3 Zn$^{2+}$, 1 SO$_4^{2-}$, 1 di(hydroxyethyl)ether</td>
<td>460 residue, 320 waters, 3 citrate, 3 Zn$^{2+}$, 1 PO$_4^{3-}$</td>
</tr>
<tr>
<td>RMSD bond lengths (Å) / bond angles (°)</td>
<td>0.020 2.198</td>
<td>0.011 1.526</td>
<td>0.020 2.101</td>
</tr>
<tr>
<td>Mean $B$-factors (Å$^2$)</td>
<td>Main chain, side chain</td>
<td>13.6, 15.3</td>
<td>18.7, 20.4</td>
</tr>
<tr>
<td></td>
<td>Waters</td>
<td>16.5</td>
<td>29.9</td>
</tr>
<tr>
<td></td>
<td>Ligands</td>
<td>15.1 (CMP)</td>
<td>18.2 (CMP, #1)</td>
</tr>
<tr>
<td></td>
<td>24.8 (Zn$^{2+}$, Mg$^{2+}$, PO$_4^{3-}$)</td>
<td>28.7 (CMP, #2)</td>
<td>26.1 (Zn$^{2+}$, PO$_4^{3-}$)</td>
</tr>
<tr>
<td></td>
<td>Ramachandran plot</td>
<td>96.9</td>
<td>97.8</td>
</tr>
<tr>
<td></td>
<td>Favourite regions (%)</td>
<td>2.2</td>
<td>2.2</td>
</tr>
<tr>
<td></td>
<td>Allowed regions (%)</td>
<td>4 outliers: Ser37A/C, Tyr29B, Gly17C</td>
<td>No outliers</td>
</tr>
</tbody>
</table>

3.1.2 Complexes with CMP and CDP

Residues from both subunits that form the active site position cytidine nucleotides though hydrogen bonding (Figures 3.1 and 3.3). In BcIspF, carbonyl groups from Ala102 and cis-Pro105 bind to the N4 base amine, and the amide groups of Leu107 and Ala108 bind to the N3 and O2 atoms of the cytosine respectively. Ribose hydroxyl O2’ is held by Asp58 from the adjacent subunit. Gly60 amide, also from the adjacent subunit, binds the O3’ hydroxyl, however this interaction is not conserved in PflspF. Thr135 amide and side chain hydroxyl bind to an oxygen on the α-phosphate.
Lys134NZ binds the CMP α-phosphate and may play a role in catalysis by helping to position the transition-state intermediate or, in conjunction with Zn$^{2+}$, align the substrate for nucleophilic attack; but this interaction is not observed in PfIspF. This interaction in BcIspF is reminiscent of that previously reported for the active site Lys132 in TtIspF (Kishida et al., 2003), and an overlay (not shown) confirms this, except that in TtIspF Lys132NZ binds the β-phosphate. Lys106 side chain is situated above the cytosine and provides van der Waals contacts.

Un预料edly, a second molecule of CMP was observed in the active site in a structure of BcIspF co-crystallised in the presence of 10 mM ligand (Section 2.3.3). The previously described cytidine-binding site is referred as position 1, and this new site as position 2 (Figure 3.4). Residues that bind CMP at position 2 (CMP #2) are from the same subunit that binds the active site Zn$^{2+}$. The CMP #2 phosphate associates directly with the active site Zn$^{2+}$. Ser64 carbonyl binds cytosine O2 thorough a water
molecule. Phe70 and Ala73 carbonyl groups position a water molecule that binds cytosine N3. Ala73 carbonyl also binds directly to cytosine N4, together with Ser37 side chain OG. Asp65 binds to ribose O2'. His36 main chain amide and carbonyl position waters that each bind to the phosphate. Overlay with the EcIspF-product complex shows that CMP #2 binds in the same location as MEcPP (Figure 3.5; Steinbacher et al., 2002). Additionally, citrate may mimic the MEcPP product, or transition state since it binds directly, and indirectly through solvent, to residues conserved with EcIspF that are known to bind MEcPP (Section 3.1.3). Residues that bind MEcPP are conserved in PfIspF as His115, Ser116, Ile138, Phe142 and Asp144 (Figures 3.1 and 3.5).

Figure 3.4  CMP binding interactions at the enzyme active site in BcIspF. Interactions between BcIspF and CMP, and the coordination of Zn$^{2+}$ are highlighted with black and red dashed lines respectively. The established nucleotide binding site is labelled CMP "1"; the novel CMP-binding site is labelled CMP "2". Solvent molecules are depicted as red spheres. There are some minor differences in the residues involved in binding interactions with CMP #2 in active sites related by NCS (not shown). Adjacent subunits are coloured yellow and cyan respectively. Electron density is displayed as green chicken-wire corresponding to an |F$_{o}$-F$_{c}$| map (omit map) for CMP #2 contoured at 2.5 $\sigma$. 

49
**Figure 3.5** CMP #2 occupies the same site as MEcPP. a) The binding interactions between EcIspF (PDB code: 1JY8) and the reaction products CMP and MEcPP. Relevant residues are discussed in the text and marked in Figure 3.1. b) An overlay of EcIspF product complex and BcIspF. CMP #2 binds in a similar site as MEcPP and also associates with Zn\(^{2+}\). Further, the loop surrounding the ligands (EcIspF: Phe61-Ser73; BcIspF: Phe63-Ser75) adopts a strikingly different conformation.

### 3.1.3 BcIspF-citrate complex

Attempts to produce holo-BcIspF crystals for ligand soaking experiments identified conditions with citrate in the crystallisation reservoir (Section 2.3.4). However holo-BcIspF crystals had not been formed, instead electron density consistent with citrate was present, blocking the three active sites. Citrate was well defined in one site and modelled at full-occupancy, but poorly defined in the remaining two sites. The well-defined citrate binds to the active site Zn\(^{2+}\) via both a carboxylate and hydroxyl. In contrast, citrate is present at lower occupancy, 0.75, in the remaining sites and adopts a different pose, with only a carboxylate binding to Zn\(^{2+}\). The carboxylates of the well-defined citrate also interact with the main chain amides of His36 and Ser37, and side chains of Asp65 and Lys134 (Figure 3.6). An additional solvent mediated interaction is formed with the side chain of Glu137 (not shown). This well resolved citrate helped to order the loop covering the active site in this subunit and an overlay (not shown) demonstrated that these residues adopt the conformation observed in the EcIspF-MEcPP complex (Steinbacher *et al.*, 2002).
Figure 3.6 Citrate binding to Zn\(^{2+}\) in BcIspF. A well resolved citrate binds to Zn\(^{2+}\) and makes interactions in the active site through carboxyl groups. An omit map is displayed for citrate in green chicken wire contoured at 3\(\sigma\). Residues are shown as sticks and a grey sphere represents Zn\(^{2+}\). Interactions are marked with dashed lines. Solvent mediated interactions are not shown.

3.1.4 Co-crystallisation with potential ligands

Crystallisation of PfIspF and BcIspF in the presence of fosmidomycin, benzohydroxamic acid, acetohydroxamic acid or pyrithione was attempted. A dataset was collected to resolution of 2.0 Å from a crystal of BcIspF that formed in the presence of 10 mM pyrithione. The structure demonstrated that pyrithione leaches Zn\(^{2+}\) from the active site (not shown) and once both Zn\(^{2+}\) and pyrithione were known to be absent the structure was not refined further. Crystals of BcIspF were prepared in the presence of both 2 mM CMP and stoichiometric amounts (0.5 µM) of pyrithione. Nine such crystals were screened for diffraction quality and a data set was collected from one to a resolution of 2.1 Å and refined (\(R_{work}\) 20.8%, \(R_{free}\) 24.6%). Electron density was present for both CMP and Zn\(^{2+}\), however no pyrithione was visible in the resulting maps. Similarly, structures of IspF determined in the presence the other compounds, showed no density for the putative ligands and were not subsequently refined.

A fragment screen against BcIspF, described in Section 3.3.1., identified six compounds that gave a reliable binding response. Co-crystallisation and ligand soaking
experiments were performed with three of the fragments identified (fragments a-c, Figure 3.8). When fragments were present at 1 mM, crystals were only observed to form in the presence of 1 mM CMP. Higher concentrations of fragments resulted in precipitation; fragment e undergoing phase-separation. A crystal of BcIspF that formed in the presence of 1 mM CMP and 1 mM of fragment a was collected and the structure determined (not shown). Electron density maps indicated that CMP was present but no density for fragment a was identifiable. Another dataset was collected from a crystal that had been co-crystallised with 1 mM CMP present and subsequently soaked for three hours with a cocktail of fragments a-c. Fragments were dissolved in DMSO and reached a final concentration of 5 mM in the crystallisation drop. In this crystal structure also, CMP was present but no ligand density was identified. Longer soaking times resulted in crystal damage or poor diffraction.

3.2 Structural studies on IspF - Conclusions

The structures of IspF from two pathogens have been determined, demonstrating similarity to known structures from other species. These proteins have been assessed for suitability as platforms for structure-based drug discovery. However, since numerous attempts to obtain the structure of protein-ligand complexes failed, it is concluded that neither BcIspF nor PfIspF are promising candidates for such studies; at least using the crystallisation conditions reported here, all of which are at high ionic strength. Future studies may employ EcIspF or BpIspF, for which successful co-crystallisation with small molecule ligands have been reported (Crane et al., 2006; Ramsden et al., 2009; Begley et al., 2011), in concert with efforts to co-crystallise with the fragments identified in this thesis (Section 3.3). As observed previously, the cytidine-binding pocket is rigid in both structures, whereas the pocket that binds the methylerythritol-phosphate portion of the substrate is relatively flexible.
The unexpected observation of a second molecule of CMP in the active site of BcIspF, as well as the presence of citrate in a related structure may inform on ligand design. Novel ligands addressing the MEcPP pocket could conceivably be designed to recapitulate the binding interactions observed for CMP #2, although loop flexibility and the requirement to select out a single conformation from an ensemble may introduce unfavourable entropic costs to binding (Zürcher & Diederich, 2008). Potential ligands could be generated, for example, by extending out from a core cyclohexa-1,5-dien-1-amine fragment, replacing the water molecules in Figure 3.4 that bind to Ala73, Phe70 and Ser64 with hydrogen bonding groups, and strengthening the interaction with Asp65 by replacing the ribose O2' hydroxyl with a positively charged group such as a primary amine. A screen of six generic Zn$^{2+}$-binding compounds identified L-tryptophan hydroxamate as one of the compounds with the highest affinity for IspF known to date ($K_d$: 2 µM), suggesting that addition of a hydroxamate group may anchor ligands in the active site (Ramsden et al., 2009). Additionally, any compounds targeting position 2 that might be developed could be linked with CMP analogues that target position 1 to produce compounds that may block both sites, mimicking the substrate or transition state.
3.3 IspF fragment screening

3.3.1 BcIspF screening results

Dissociation constant values ($K_d$) recorded for CMP and CDP were 18 µM and 3.7 µM respectively using the BLI technique (Figure 3.7). These measurements provided a positive control for the fragment screen.

![Figure 3.7 Measuring the binding affinities of cytidine nucleotides for BcIspF.](image)

The response of BcIspF to CMP and CDP was recorded by BLI, acting as a positive control for fragment screening (CMP was measured in duplicate). The dose-dependent binding curves were generated using statistically smoothed data derived from the raw sensorgram traces with interpolation. $K_d$ values were calculated using this smoothed data. $R^2$-values indicate the goodness-of-fit and reliability of the data.

Seven compounds (a-g) were identified in the screen against BcIspF (Figure 3.8). One of the fragments (g) is now considered to be a non-specific binder since it has been flagged in several unrelated screens (Dr. David Robinson). Fragments a-e were selected for purchase. However co-crystal structures with BcIspF could not be determined despite several efforts (Section 3.1.4).
Figure 3.8  **Fragments identified in the screen against BcIspF.** The chemical structures and identification codes of seven compounds (a-g) that were identified as potential IspF binders are shown.
3.3.2 EcIspF screening results

CMP gave a positive control signal ($K_d$: 40 µM, $R^2$: 0.96). In 2009, Ramsden et al. reported a $K_d$ value of 900 µM, however experimental differences complicate direct comparison. The previously reported $K_d$ value was arrived at using an SPR assay with EcIspF bound to sensors via amine-coupling. Additionally, a different buffer was used; 2 mM MgCl, 50 mM Na$_2$HPO$_4$, pH 7 (Ramsden et al., 2009). Nonetheless, the $R^2$ value reported in the current study provides reliability. The content of the DDU fragment library changed between the screen on BcIspF and subsequently on EcIspF (Section 2.1.5). 31 compounds from the revised library were identified that gave a double-referenced binding response to EcIspF, corresponding to 4.7 % of all the compounds in the fragment library. 28 of these fragments were taken forward to generate dose-response curves (Figure 3.9). The three excluded compounds, DDD00100335, DDD00100377 and DDD00100546, were not considered reliable since they produced the large binding responses characteristic of non-specific aggregation. Consideration of the reliability and goodness-of-fit statistical scores, including $R^2$-values as well as direct inspection of the raw sensorgrams produced for each of the 28 fragments allowed the identification of the most reliable hits (Figure 3.10). These fragments have $K_d$ values ranging from approximately 30 - 700 µM.
a) BLI Sensorgram for EcIspF dose-dependent response to fragment j

![BLI Sensorgram](image)

b) Dose-dependent binding curve for j generated from sensorogram

![Dose-dependent binding curve](image)

Figure 3.9 An example of BLI data from fragment screening of EcIspF. a) The binding response to the fragment (inset) is shown. Vertical displacement represents binding response in nm at increasing concentrations (2, 6.2, 18.5, 55.5, 166.7 & 500 µM). b) A dose-dependent binding curve generated for the fragment using statistically smoothed data derived from the raw sensorgram traces with interpolation. The maximum shift in wavelength (after 60 seconds) is plotted for each concentration. Readings from the two lowest concentration points were discarded owing to weak and noisy signal. $K_d$ values were calculated using this smoothed data.
Figure 3.10 Fragments identified in the screen against EcIspF. The chemical structures of the compounds (h-n) identified in the screen against EcIspF with the highest reliability scores and best quality sensograms are shown. In addition to the compounds displayed, fragment e (Figure 3.8) was re-identified in this screen ($K_d$: 75 µM, $R^2$: 0.82). $K_d$ values have been calculated from a global fit of the BLI sensogram data and $R^2$-values represent the reliability of the fit.
### 3.3.3 IspF fragment screening - Discussion

A total of 13 unique chemical fragments were identified that gave reliable binding responses in the two BLI fragment screens reported here. Of the fragments identified for BcIspF from the original DDU fragment library, only fragment c and fragment e (Figure 3.8) were present in the revised library that was screened against EcIspF. Fragment e, was re-identified in the later screen against EcIspF. Curiously, while fragment c was not re-identified in the later screen, fragment h (Figure 3.10) differs from fragment c by a single chlorine atom.

Computational docking and modelling of the identified fragments was considered unsuitable owing to their small size, limited number of hydrogen bonding groups and the relatively large, open active site of IspF. For example, EcIspF has an active site volume of 1530 Å³, according to the CASTp volume calculation server set to a probe radius of 1.4 Å (Dundas et al., 2006).

Six of the fragments contain either a thiazole or an oxazole (fragment c in Figure 3.8; fragments h, j, k, l and n in Figure 3.10). Thiazole-bearing fragments have been reported to bind IspF previously (Begley et al., 2011). Such a fragment was observed to bind in a pocket outside the active site in a co-crystal structure with IspF from the related Burkholderia species, B. pseudomallei (Begley et al., 2011). In that crystal structure an arginine residue, Arg69, forms a cation-π interaction with the thiazole ring, however Arg69 is not conserved in BcIspF and EcIspF - replaced by Ala69 and Ala67 respectively. The thiazole- and oxazole-bearing fragments reported here may bind to the active site Zn²⁺ cation. Most of the fragments identified contain metal binding groups, planar groups and can make hydrophobic contacts, similar to CDP-MEP, the substrate of IspF.
In fragment screening it is considered good practice to validate initial hits by counter-screening with a complementary biophysical technique (Scott et al., 2012). Assessment and counter-screening of fragment hits using either isothermal titration calorimetry or NMR, accompanied by attempts at co-crystallisation with EcIspF for structural determination would be ideal, but this has not been pursued due to time-constraints. Future work could build from the results of the initial fragment screens reported here.

A convenient enzyme kinetics assay for IspF is unavailable; however unambiguous demonstration of enzyme activity and determination of inhibition constants will require such an assay. Fluorescent IspF substrate-analogues have been designed, which may facilitate a fluorescence-displacement approach (Crane et al., 2006). Alternatively, Narayanasamy et al. described a novel IspF kinetics assay in 2010; which involves coupling CMP release to four enzymes (nucleotide monophosphate kinase, pyruvate kinase, pyruvate oxidase and horseradish peroxidase); which alter the emission wavelength of a fluorescent dye. Future work could also direct effort to developing a simpler kinetics assay (Narayanasamy et al., 2010).
CHAPTER 4

$P_aP_{abC};$ STRUCTURAL ANALYSIS
4.1 *PaPabC* - Overall structure

A single species, ~50 kDa, commensurate with a *PaPabC* homodimer was observed during size-exclusion chromatography. A dimer was also observed in the X-ray crystal structure. Two polypeptide chains, labelled A and B, constitute the asymmetric unit (Figure 4.1), and crystallographic statistics are presented in Table 4.1.

![The *PaPabC* dimer.](image)

The subunits display a high degree of NCS, with an RMSD of 0.4 Å following least squares fitting of 270 Ca positions. It is therefore only necessary to detail one subunit and one active site. The subunit is constructed from two domains (Figure 4.2). The N-terminal domain (domain I) is smaller than the C-terminal domain (domain II). Domain I comprises residues 1–107, which contains three α-helices, α1–α3, and a four-stranded anti-parallel β-sheet in order β1–β4–β3–β2. Domain II comprises residues 112–271 that form five α-helices, α4–α8, one short two-stranded parallel β-sheet, β5–β6, and one short two-stranded anti-parallel β-sheet, β7–β8. Residues 108–111 link the two domains. The PLP cofactor is covalently bound to Lys140 in domain II and the active site is found in a cleft formed between the two domains (Figures 4.1, 4.2, 4.3).
### Table 4.1 Crystallographic statistics for PaPabC

<table>
<thead>
<tr>
<th><strong>PDB code</strong></th>
<th><strong>2Y4R</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Space group</strong></td>
<td>$P2_12_12_1$</td>
</tr>
<tr>
<td><strong>Unit cell dimensions</strong></td>
<td></td>
</tr>
<tr>
<td>$a$, $b$, $c$ (Å)</td>
<td>40.8, 66.8, 202.7</td>
</tr>
<tr>
<td><strong>Resolution range</strong> (Å)</td>
<td>30.00 - 1.75 (1.84 - 1.75)</td>
</tr>
<tr>
<td><strong>No. reflections</strong></td>
<td>209336 (26399)</td>
</tr>
<tr>
<td><strong>Unique reflections</strong></td>
<td>56255 (8090)</td>
</tr>
<tr>
<td><strong>Completeness (%)</strong></td>
<td>98.9 (98.5)</td>
</tr>
<tr>
<td><strong>$R_{merge}$</strong></td>
<td>7.8 (45.1)</td>
</tr>
<tr>
<td><strong>Multiplicity</strong></td>
<td>3.7 (3.3)</td>
</tr>
<tr>
<td>$&lt;I/\sigma(I)&gt;$</td>
<td>10.7 (2.5)</td>
</tr>
<tr>
<td><strong>Wilson $B$ (Å$^2$)</strong></td>
<td>18.4</td>
</tr>
<tr>
<td><strong>$R_{work}$/$R_{free}$</strong></td>
<td>16.7, 21.5</td>
</tr>
<tr>
<td><strong>Number of residues / waters / ligands</strong></td>
<td>538 residues, 474 waters, 2 PLP, 3 Cl$^-$, 2 SO$_4^{2-}$, 2 glycerol, 3 di(hydroxyethyl)ether, 5 tetraethylene glycol, 3 1,2-ethanediol</td>
</tr>
<tr>
<td><strong>RMSD bond lengths</strong> (Å) / bond angles (°)</td>
<td>0.025 / 2.100</td>
</tr>
<tr>
<td><strong>Mean $B$-factors (Å$^2$)</strong></td>
<td></td>
</tr>
</tbody>
</table>
| Main chain, side chain | Chain A: 15.1, 18.2  
| Chain B: 18.5, 20.7 |
| Waters | 30.8 |
| Ligands | (PLP) Chain A: 15.8, Chain B: 21.5  
| 36.3 (other ligands) |
| **Ramachandran plot** | 
| Favoured regions (%) | 93.7 |
| Allowed regions (%) | 5.7 |
| Outliers | 3 outliers: Arg90A & B, Leu89B |
Figure 4.2 Ribbon diagram of the PaPabC monomer. Secondary structural elements have been numbered and the position of PLP is indicated. β-strands are shown in purple, α-helix in yellow.

Sixty residues from each subunit, provided by both domains, contribute to dimer formation. These residues participate in around 30 hydrogen-bonding associations and eight salt bridge interactions to stabilise the quaternary structure. Most of the residues that contribute to the dimer interface are found in two patches; residues Ala10 - Gly25 centred on α1 and residues Arg127 - Ala152 centered on α5. The approximate dimensions of the dimer are $80 \times 40 \times 40$ Å and this displays a solvent accessible surface area of approximately $21000$ Å$^2$. The contact area between the two subunits covers approximately $2050$ Å$^2$, about 15% of the surface of a subunit. These values are similar to those observed for other PabC proteins; for example the interface of the E. coli protein (Nakai et al., 2000) covers an area of $2130$ Å$^2$.

The structures of PabC from three organisms, E. coli, L. pneumophila and T. thermophilus, display a high degree of structural conservation with PaPabC. Least
squares fitting of the Cα positions of single subunits results in RMSD values between 1.3 - 2.0 Å, with a coverage in the range of 81 - 92% of the structures. The sequence identity shared with PaPabC, falls in the range 26 - 35%.

4.2 PLP cofactor binding site and interactions

The active site of PaPabC is positioned between the two domains of the monomer close to the dimer interface (Figure 4.1). Here, the PLP cofactor is covalently linked to Lys140 and oriented by interactions with key residues, a selection of which are depicted in Figure 4.3. The PLP phosphate accepts hydrogen bonds donated by a cluster of main chain amide groups from Val200, Met201 and Ser237, and the side chain of Arg46. The side chain of Arg46 is positioned by interactions with the side chain of Glu28, which in turn accepts a hydrogen bond from His43 ND1. Two well ordered water molecules mediate hydrogen bonding networks linking the phosphate group to Asn178 OD1 and Arg202 N on one side, and the main chain O and N groups of Thr29 on the other. The cofactor ring is sandwiched between Val197 and the main chain of a tripeptide segment comprising Val175-Phe176-Ser177. A hydrogen bond between Asn236 ND2 and the carbonyl of Phe176, in conjunction with bonds formed between the amide groups and side chain of Glu161, determine the conformation of the tripeptide that forms a tight turn between β9 and β10 covering part of the cofactor. The PLP methyl group makes van der Waals interactions with Gln147 CG and is positioned 3.2 Å distant from the carbonyl oxygen of Val175. The geometry of this latter contact is compatible with the presence of a C-H⋯O hydrogen bond (Leonard et al., 1995). The side chain of Gln147 is held in place by a hydrogen bond with Arg144 NE. Bonds formed with the carbonyl groups of Leu139 and His141 position the side chain of Arg144. There are also three water molecules, which form a network of hydrogen bonds linking the side chain of Arg144 to other sections of the enzyme (not shown).
addition the positive dipole from α6 - the closest residue on this helix is Gly199 - provides an attractive force to interact with the PLP phosphate. The side chain of Glu173 displays a degree of conformational freedom with two rotamers being observed. One rotamer places the side chain to accept a hydrogen bond donated from PLP N1 (Figure 4.3). The alternative rotamer places the carboxylate at a distance of 3.7 Å from N1. Tyr92 from the partner subunit forms a hydrogen bond to PLP O3. The cofactor is also likely to possess an intra-molecular hydrogen bond involving O3 and NZ of Lys140. The side chain of Phe27, Val197 and aliphatic components of Arg46 and Arg144 surround the side chain of Lys140 (not shown).

![Image](image.png)

**Figure 4.3 Omit map for PLP and Lys140 from subunit A of PaPabC.** (Fo-Fc) difference density contoured at 3 σ (green chicken wire). Selected hydrogen bonding associations between the protein, two water molecules (red spheres) and PLP are depicted as dashed lines. An asterisk marks Tyr92 contributed from subunit B. The side chain of Glu173 displays two rotamers with only one shown.

No structures of PabC-ternary complexes have yet been published to inform on aspects of substrate recognition and enzyme activity. It was considered important to attempt structural determination of PaPabC in complex with a number of compounds including PABA, pyruvate, phenylpyruvate, thiomersal and sulfanilamide. For example, the structure reported here was crystallised in the presence of 10 mM PABA. However, no electron density compatible with binding of any such ligand in the
substrate-binding pocket has been identified. Therefore, the position of PLP and Lys140 NZ, together with the conservation of amino acid sequence and structures was judged to give an indication of where substrate binds; as well as which residues are important for substrate recognition and catalytic function. Consideration of an amino acid sequence alignment and structural overlay of *Pa*PabC with three other structures in the PDB proved particularly informative (Discussion) when combined with molecular modelling and docking of the covalent intermediate formed during catalysis.

4.3 *Pa*PabC active site – Comparision with orthologues

Overlays of *Pa*PabC with the structures of three orthologues were generated, together with sequence alignments, for comparison of active site residues. The contributions of 24 residues within 5 Å of PLP to binding the cofactor (Section 4.2) and organising the active site are summarised in Figure 4.4. Seven of these residues are strictly conserved in PabC from *P. aeruginosa*, *E. coli*, *L. pneumophila* and *T. thermophilus* (Phe27, Thr29, His43, Arg46, Gly199, Lys140 and Arg202 in *Pa*PabC). With one exception these residues contribute to the active site using the chemical properties of their side chains. The exception is Gly199 where any alteration on Cα would result in a steric clash with the phosphate group of PLP. Another seven residues contribute to the active site primarily through the hydrogen bonding capacity of the main chain (Val175, Phe176, Ser177, Leu139, His141, Val200 and Met201). Phe176 is not conserved in any of the three other sequences but the remainders are conserved in at least one other orthologue. Ser237 interacts with the PLP phosphate using main chain and side chain groups. This residue is only different in the *E. coli* sequence where it is replaced by an alanine and an additional water molecule binds nearby helping to satisfy hydrogen-bonding capacity. A similar observation is made when considering *Pa*PabC Asn236. This residue is conserved in EcPabC and LpPabC but is changed to glycine in TIPabC
where a water molecule is observed to replace the side chain. Asn178 is strictly conserved in EcPabC and LpPabC but changed to serine in TtPabC. The side chain here helps to form one side of the cofactor-binding site. Val197 is conserved in EcPabC and conservatively substituted as leucine in the other two sequences. The residue at this position in the active site forms van der Waals interactions with the cofactor. In PaPabC Arg144 NE donates hydrogen bonds to Gln147 OE1 and the carbonyl of His141; interactions that help to create one side of the cofactor-binding site and to fix the position of Val197. These residues and the interactions they form are conserved in EcPabC but not in LpPabC or TtPabC. Sequence and structure alignments indicate that PaPabC Arg144 is replaced by tyrosine in both LpPabC and TtPabC, and PaPabC Gln147 replaced by an alanine in LpPabC and a tyrosine in TtPabC (Discussion).

Figure 4.4 Alignment of PaPabC and LpPabC. Residues encased in black are strictly conserved. Red stars indicate the tyrosine residues at position I (PaPabC Y92), and position II (LpPabC Y144). Triangles mark residues that are discussed in the text. Red triangles identify Lys140 and residues that interact directly with the Lys140-PLP adduct; yellow triangles mark residues that contribute to the organisation of the active site or that participate in solvent mediated interactions between the protein and the cofactor.

There are three glutamates among the 24 residues discussed above (Glu28, Glu161 and Glu173). Glu28 helps to position Arg46 to bind the cofactor phosphate. Both residues are strictly conserved in LpPabC. A conservative substitution of the glutamate to threonine occurs in both EcPabc and TtPabC, and the interactions to support the binding of arginine to the cofactor are preserved. Glu161 helps to configure
a tripeptide segment to cover part of the cofactor (not shown) and this feature is conserved since a glutamate is present in the other structures (an aspartate occupies this position in \(LpPabC\)). The carboxylate side chain of Glu173 can interact directly with the cofactor by accepting a hydrogen bond donated from N1. In a similar fashion to what is observed for Glu161, such interactions occur in the orthologues with the residue either strictly conserved, or as in \(TTPabC\), replaced by aspartate.

### 4.4 \(PaPabC\) - Discussion

A catalytic mechanism, based on previous studies of PLP-dependent enzymes, was proposed for PabC prior to any structural data being available (Green et al., 1992). Following structure determination of \(EcPabC\), a model of substrate binding was prepared to inform an assessment of the mechanism (Nakai et al., 2000). The contribution of PLP was considered the same, that is, following covalent linkage the enzyme-cofactor adduct undergoes a nucleophilic attack by the substrate amino group to produce an external aldimine. PLP, which contains a conjugated electronic structure linked to the pyridine-ring nitrogen N1, acts as an electron sink; weakening the bond between the \(\alpha\)-proton and C4 of the substrate (Figure 1.10). The PLP-substrate species then undergoes \(\alpha\)-proton elimination, to the \(\varepsilon\)-amino group of a lysine to yield a quinonoid intermediate. In \(EcPabC\) Lys159 forms the covalent link to the cofactor and is likely responsible for \(\alpha\)-proton abstraction. Nakai et al., further suggested that Thr28 of \(EcPabC\) supplies a proton, converting the methylene group of the substrate to a methyl group as pyruvate is released. PABA is released and the Schiff-base form of the enzyme is recovered as PLP reacts with Lys159 (Figure 4.5).
Figure 4.5 Proposed catalytic mechanism of PabC. Curly arrows indicate the flow of electrons. The amino group of the substrate, 4-amino-4-deoxychorismate, attacks the Schiff-base to form an external aldimine. The displaced lysine is proposed to abstract the α-proton from the substrate at C4, producing a quinonoid intermediate. Aromatisation of the substrate ring leads to the loss of pyruvate. Concomitant with the loss of pyruvate a suitably placed threonine is proposed to act as a proton donor to facilitate conversion of methylene to methyl. Regeneration of the internal aldimine following nucleophilic attack of lysine displaces PABA (highlighted in a yellow box). (Mechanism described in Nakai et al., 2000)

Three residues described as contributing to the mechanism of action in EcPabC, Thr28, Lys97 and Lys159 (Nakai et al., 2000), are strictly conserved in PaPabC as Thr29, Lys80 and Lys140. They are also conserved in LpPabC and TtPabC with the conservative change of Lys80 to arginine in the thermophilic orthologue (Figure 4.4). Lys140 of PaPabC is suitably placed to support α-proton abstraction and although Thr29 may engage in the proposed proton shuffling with both the olefin moiety of the substrate and Lys80, an alternative residue (Tyr90) might fulfil this role. Tyr90 may have been overlooked because Nakai et al., had access to only a single structure of PabC. Alignments based on sequences alone would not identify a tyrosine that is conserved in the three-dimensional structure.
In *EcPabC* and *PaPabC* tyrosine hydroxyl groups (Tyr109 and Tyr92 respectively), from the partner subunit, donate a hydrogen bond to PLP O3 (Figures 4.3, 4.5) in an interaction that influences the electronic structure of the cofactor (Nakai *et al.*, 2000). The sequence alignment of PabC orthologues suggests a degree of variation in the identity of the amino acid that occupies the position of this tyrosine (Figure 4.4). In contrast, the structural overlays are striking since they clearly identify a spatially conserved tyrosine in relation to the cofactor (Figure 4.6). In *EcPabC* and *PaPabC*, Tyr109 and Tyr92 respectively, are contributed from the partner subunit (position I). However, in *LpPabC* and *TtPabC*, Tyr144 and Tyr130 respectively, are contributed from the same subunit (position II).

A conservative group of 129 putative PabC sequences were retrieved from the UniProt database, aligned using *MUSCLE* (Edgar, 2004) and then inspected following pairwise alignment with *PaPabC*. This was done to investigate the distribution of tyrosine at these two positions (I & II); seeking to address the hypothesis that if tyrosine is present at position I then it is likely absent from position II and *vice versa*. This hypothesis holds true for 125 out of the 129 sequences. Two sequences, which share less than 30% sequence identity, have a tyrosine at both positions. Analyses of the sequences suggests that PabC can be classified into two groups depending on whether
this active site tyrosine is provided by the same subunit which primarily forms the active site or from a partner subunit.

The high degree of conservation of an active site tyrosine in three-dimensional structures of PabC suggests a functional role. One possibility is that this residue could act as a proton donor for the conversion of methylene to methyl that completes the reaction. Such a conversion would have little influence on the cleavage of the C-O bond of the substrate since the driving force for that reaction would be production of an aromatic six-carbon ring system. A model of the catalytic intermediate (4-amino-4-deoxychorismate linked to PLP, Figures 4.5, 4.7) was specifically prepared to investigate the potential role of this active site tyrosine and to re-evaluate the olefinconversion process that completes the reaction (calculations were performed by Dr. Thomas Eadsforth).

![Figure 4.7](Image)

**Figure 4.7 A stereoview of the catalytic intermediate docked into the active site of PaPabC.** Protein components are coloured as in Figure 4.3. The C atoms of the modelled intermediate (1) are in black. Black dashed lines represent potential hydrogen bonding interactions, with a separation of 3.5 Å or less, and the red dashed line between Thr29 OG1 and the sp² hybridised C indicates the proximity of functional groups likely involved in catalysis.

The model was prepared using the PRODRG server (Schüttelkopf & van Aalten, 2004) and placed in the active site with COOT using the position of PLP as a guide. Docking in the active site was performed using ICM pro (MolSoft L.L.C.) with the centre of the ligand-binding site defined by a cavity that contained the residues within 5
Å of the ligand. The top ten docking poses, as scored by ICM pro gave similar orientations in the active site and one was selected for display.

The model suggests that Tyr112 donates a hydrogen bond to the C1 carboxylate while, with contributions from the partner subunit, the main chain amide of Tyr92 and hydroxyl group of Tyr22 donate hydrogen bonds to the other carboxylate of the intermediate (Figure 4.7). Tyr92 contributes to the positioning and electronic structure of the cofactor by virtue of the hydrogen bond involving the hydroxyl group with the PLP carbonyl (Figure 4.3). For example, it has been suggested that a tyrosine at a similar position in the PLP-dependent enzyme aspartate aminotransferase affects the $pK_a$ of the Schiff base; and this has been investigated by generating Tyr-Phe mutants (Goldberg et al., 1991; Inoue et al., 1991). The bulk of the tyrosine side chain serves to create a hydrophobic wall on one side of the active site. Although placed to interact with the catalytic intermediate this residue does not appear to be involved in donating a proton to convert methylene to methyl (Figure 4.7). Rather, Thr29, as first postulated by Nakai et al., appears to fulfil such a role with the hydroxyl group placed about 3.5 Å from the olefin (Figure 4.7). Lys80 NZ is around 4.3 Å distant from the $sp^2$ hybridised C and it is possible that an alternative rotamer could bring this functional group into proximity; providing a proton itself or activating the Thr29 hydroxyl for proton donation. Thr29 is strictly conserved in 123 out of the 129 PabC sequences and conservatively substituted by serine in another five sequences.

To further investigate the structure-activity relationship of PabC, and to assess the potential of this enzyme as a target for structure-based drug discovery will require the powerful combination of steady-state kinetic analysis with site-directed mutagenesis studies, alongside screening of chemical libraries. In particular, the construction of Tyr92-Phe, Thr29-Ala, Lys80-Met and Glu173-Gln mutants will permit assessment of the contribution of these residues to catalysis (Goldberg et al., 1991; Inoue et al., 1991).
Following production of the substrate from commercially available chorismate and glutamine by recombinant PabA-PabB, the kinetics of PabC can be assayed by determining the production of PABA upon addition of the enzyme. This would involve taking aliquots at various time-points, incubating in 1 M HCl to inactivate the enzymes, and then quantifying PABA concentrations using HPLC (He et al., 2004; and references therein). UV-visible spectroscopy could be performed in parallel to gain insight into the electronic structure of the PLP cofactor (Fasella, 1967). Substrate-analogues containing Michael acceptors or halogens could also be considered during inhibitor design and testing, since these may potentially form covalent adducts with the cofactor (Eliot & Kirsch, 2004). It will be imperative to extend from this work and to elucidate a means whereby structural data on enzyme-ligand complexes can be obtained.
PART TWO

IMMUNITY PROTEINS IN THE GRAM-NEGATIVE TYPE VI SECRETION SYSTEM
CHAPTER 5

Rap1a and Rap2a -

INTRODUCTION AND METHODS
Introduction

Preamble: Dr. Vincent Rao and Dr. Velupillai Srikannathasan performed the purification, crystallisation and X-ray data collection of Rap1a and Rap2a respectively. My contribution to this project involved the processing of datasets, determining the structures by molecular replacement and analysing the results. This work, performed as part of a team, has been accepted for publication and this thesis draws on this source extensively. The relevant publication is referred to throughout as (Srikannathasan et al., 2013).

5.1 Type VI secretion systems (T6SS)

Specialised secretion systems are key to bacterial fitness, survival and pathogenesis. They contribute a myriad of roles in the processes that influence growth, colonisation, attack and defense as bacteria interact with each other and with eukaryotic organisms (Filloux, 2011; Gerlach & Hensel, 2007). The recently identified Type VI secretion system (T6SS), present in about 25% of Gram-negative bacteria for which genome sequences are available (Boyer et al., 2009; Pukatzki et al., 2006), can be used to target bacterial and eukaryotic cells, and is thus important for both inter-bacterial competition and pathogenesis (Burtnick et al., 2011; de Pace et al., 2010; Jani & Cotter, 2010). Anti-bacterial T6SSs mediate efficient killing of competitors by direct injection of toxic, antagonistic effector proteins into target cells (Hood et al., 2010; MacIntyre et al., 2010; Murdoch et al., 2011; Schwarz et al., 2010).

T6SS gene clusters encode the core components of a secretion machine capable of membrane perforation (Table 5.1). This is a multi-protein needle-like assembly (Figure 5.1) -resembling the contractile tail of a bacteriophage - that delivers effectors across three envelope layers in a single step (Basler et al., 2012; Bönemann et al., 2010; Cascales & Cambillau, 2012; Silverman et al., 2012). The clusters also encode accessory and post-translational regulatory components. Some T6SS-secreted effector proteins are also encoded within these large gene clusters. In the case of anti-bacterial
T6SSs, effectors are always encoded adjacent to specific, cognate immunity proteins. Immunity proteins bear appropriate signals to direct their localisation to the compartment in which the toxic effectors act; for example, a Sec-signal sequence guides localisation to the periplasm. The provision of a cognate immunity protein provides protection against attack from sister cells (Coulthurst, 2013).

![Schematic model of the Type VI secretion apparatus](image)

**Figure 5.1 Schematic model of the Type VI secretion apparatus.** The components of the T6SS assemble to form a cell-puncturing needle that resembles an inverted bacteriophage tail. Hcp protein forms a membrane-spanning tube with an internal diameter of ~40 Å (Mougous et al., 2006), and VgrG protein punctures membranes of target cells. Inner and outer membranes are shown as green rectangles, and a grey palisade represents the peptidoglycan layer. Certain immunity proteins such as Rap, shown as a light pink disk, localise to the periplasm via the Sec transporter, coloured cyan. Effector proteins of the T6SS (T6 effectors, red disk), such as the peptidoglycan amidase Ssp1 are thought to be secreted into target cells by transit through the Hcp tube.

Different catalytic activities are associated with T6SS effectors. These include actin cross-linking and ADP-ribosylation that disrupt the cytoskeleton of mammalian and amoebal cells (Pukatzki et al., 2007; Rosales-Reyes et al., 2012; Suarez et al., 2010), and phospholipases that degrade phosphatidylethanolamine, the major component of the bacterial membrane (Russell et al., 2013). The best-characterised effectors are peptidoglycan hydrolases, which exhibit potent anti-bacterial activity (Russell et al., 2012; Russell et al., 2011). These enzymes degrade peptidoglycan, the
heteropolymer that occupies the periplasmic space imparting mechanical strength to the cell wall and helping to maintain the shape of Gram-negative bacteria.

**Table 5.1 Components of the T6SS apparatus**
(Cascales and Cambillau, 2012; and references therein)

<table>
<thead>
<tr>
<th>Components of the T6SS</th>
<th>Localisation of component</th>
<th>Homologous proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>TssA</td>
<td>cytosol (putative)</td>
<td>n/a</td>
</tr>
<tr>
<td>TssB</td>
<td>associates with outer membrane</td>
<td>bacteriophage sheath</td>
</tr>
<tr>
<td>TssC</td>
<td>associates with outer membrane</td>
<td>bacteriophage sheath</td>
</tr>
<tr>
<td>TssD/Hcp</td>
<td>forms a hexameric ring, secretion tube</td>
<td>bacteriophage tail gp19</td>
</tr>
<tr>
<td>TssE</td>
<td>not known</td>
<td>bacteriophage wedge gp25</td>
</tr>
<tr>
<td>TssF</td>
<td>not known</td>
<td>n/a</td>
</tr>
<tr>
<td>TssG</td>
<td>not known</td>
<td>n/a</td>
</tr>
<tr>
<td>TssH/ClpV</td>
<td>cytosol, AAA⁺ ATPase</td>
<td>Hsp100/Clp AAA⁺ ATPase</td>
</tr>
<tr>
<td>TssI/VgrG</td>
<td>forms trimers, cell puncturing device</td>
<td>bacteriophage tail spike gp27-gp5</td>
</tr>
<tr>
<td>TssJ/SciN</td>
<td>outer membrane lipoprotein</td>
<td>transthyretin</td>
</tr>
<tr>
<td>TssK</td>
<td>cytosol (putative)</td>
<td>n/a</td>
</tr>
<tr>
<td>TssL</td>
<td>inner membrane</td>
<td>T4bSS IcmH/DotU protein</td>
</tr>
<tr>
<td>TssM</td>
<td>inner membrane</td>
<td>T4bSS IcmF protein</td>
</tr>
</tbody>
</table>

Peptidoglycan hydrolases constitute a large enzyme family, which displays a rich diversity in terms of structure, mechanism and specificity (Vollmer *et al.*, 2008). There are enzymes specific for every glycosidic and amide bond in peptidoglycan. Such diversity is exploited to regulate bacterial cell growth, division and daughter cell separation, and of particular interest here, to provide bactericidal properties that can be exploited in niche competition. Several classes of peptidoglycan hydrolase have been identified as T6SS effector proteins. These are termed Tse proteins. Tse3 of *P. aeruginosa* is a muramidase, cleaving the glycan backbone (Russell *et al.*, 2011), and the C-terminal domain of *Vibrio cholerae* VgrG-3, a component at the tip of the secretion apparatus (Figure 5.1), is suggested to have lysozyme-like muramidase activity (Brooks *et al.*, 2013). A diverse group of T6SS-secreted peptidoglycan amidases, which cleave the peptide cross-links, has been described (Russell *et al.*, 2012). Within this superfamily, four distantly-related families with distinct cleavage specificities were defined. Family 1 (‘Tae1’), which includes Tse1 of *P. aeruginosa*,
hydrolyse peptide crosslinks at the \(\gamma\)-D-glutamyl-\textit{meso}-diaminopimelate DL-bond; representatives of Families 2 and 3 (‘Tae2’ and ‘Tae3’) hydrolyse DD-crosslinks between D-\(m\)A\textsubscript{2}pm (\textit{meso}-diaminopimelate) and D-alanine, and a representative of Family 4, Tae4 from \textit{Salmonella enterica} serovar Typhimurium, also hydrolysies the \(\gamma\)-D-glutamyl-\(m\)A\textsubscript{2}pm DL-bond. Structures of Tse1 from \textit{P. aeruginosa}, and Tae4 from \textit{S. Typhimurium} and \textit{Enterobacter cloacae} (Benz et al., 2012; Chou et al., 2012; Ding et al., 2012; Zhang et al., 2013) place these T6-secreted effectors in the NlpC/P60 family of endopeptidases, amidases and acyltransferases; named after the New lipoprotein C from \textit{Escherichia coli} and a 60 kDa extracellular protein from \textit{Listeria monocytogenes} (Anantharaman & Aravind, 2003). The opportunistic pathogen \textit{Serratia marcescens} has recently been shown to utilise T6-dependent secretion of two Family 4 amidases, Ssp1 and Ssp2, to mediate anti-bacterial activity (English et al., 2012).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig52.png}
\caption{\textbf{The T6SS delivers toxic effectors to target cells.} A schematic close-up is depicted of two adjacent bacteria during an attack by one (dark green) using the T6SS. \textbf{a)} T6 effectors are translocated from the attacking bacterium into a susceptible target cell, inflicting damage in the periplasmic space or in the cytoplasm. \textbf{b)} A resistant cell, such as a sibling of the attacking strain, neutralises T6 effectors through the action of specific immunity proteins. In the example shown, the peptidoglycan amidase 'Tse1' is bound and inhibited by the immunity protein 'Tsi1' (crystal structures inset; Ding et al., 2012). This figure has been adapted with permission from Coulthurst, 2013.}
\end{figure}
Table 5.2 List of key details for the proteins reported in this thesis (Part 2)

<table>
<thead>
<tr>
<th>Source organism</th>
<th>Rap1a</th>
<th>Rap2a</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Source organism</strong></td>
<td><strong>Serratia marcescens, strain Db10</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Number of residues</strong></td>
<td>101 (mature protein)</td>
<td>100 (mature protein)</td>
</tr>
<tr>
<td><strong>Signal sequence omitted</strong></td>
<td>Yes; 26 residues at N-terminus</td>
<td>Yes; 24 residues at N-terminus</td>
</tr>
<tr>
<td><strong>Molecular mass (Da)</strong></td>
<td>11486.0</td>
<td>11106.4</td>
</tr>
<tr>
<td><strong>Theoretical isoelectric point</strong></td>
<td>6.2</td>
<td>5.1</td>
</tr>
<tr>
<td><strong>PDB codes</strong></td>
<td>3ZFI (1.98 Å resolution)</td>
<td>3ZIB (1.90 Å resolution)</td>
</tr>
<tr>
<td><strong>Position of disulfide bonds</strong></td>
<td>Cys78-Cys122</td>
<td>Cys42-Cys102</td>
</tr>
</tbody>
</table>

Bacteria that secrete potent peptidoglycan hydrolase effectors, using the T6SS to attack competitors, could generate a deleterious effect on their own population. To cope with this potential for 'friendly fire' damage, such bacteria also possess cognate immunity or resistance proteins located in the periplasm (Figure 5.2). These immunity proteins bind their cognate effectors with low nanomolar affinity to neutralise them in a highly specific manner, (English et al., 2012). Four distinct families of putative immunity proteins are associated with four Tae amidase families (Tai1-Tai4; Russell et al., 2012). In *S. marcescens*, the resistance associated proteins, Rap1a and Rap2a, neutralise Ssp1 and Ssp2, respectively (English et al., 2012). The crystal structures of Rap1a and Rap2a are reported in this thesis (Table 5.2). Rap1a and Rap2a have only marginal sequence identity and are scarcely alignable. Additionally two other Rap proteins, Rap1b and Rap2b are, together with Ssp1 and Ssp2, encoded in the same locus within the T6SS gene cluster. Structures of Rap1b and Rap2b have revealed a novel α-helix fold and dimeric assembly (English et al., 2012), which was later observed in the Tai4 protein from *E. cloacae* (*EcTai4*) and *S. Typhimurium* (*STTai4*, Zhang et al., 2013). This fold is a template for some T6SS immunity proteins, called Tsi proteins, but not all. For example, analysis of the Tse1-Tsi1 effector-immunity protein combination found in *P. aeruginosa* revealed Tsi1 as an all-β protein (Benz et al., 2012;
Ding et al., 2012; Shang et al., 2012), whereas Tsi2 in the same organism, the immunity protein associated with the cytoplasmic effector Tse2, has a helical fold distinct again (Li et al., 2012). That effector immunity defense systems based on distinct folds have evolved is perhaps to be expected, given the strong evolutionary pressure applied by multifarious secreted effectors. Indeed such pressure may even have contributed to the acquisition of effector-immunity pairs encoded outside of T6SS operons - for example all three Tse/Tsi pairs of P. aeruginosa.

5.2 Aims - Part two

Structures of Rap1a and Rap2a from S. marcescens were sought to gain insight into their fold and function, and to allow comparison with immunity proteins of known structure.
Methods

5.3 Rap1a and Rap2a - Crystallisation, molecular replacement and analysis

Rap1a was dialysed against 150 mM sodium chloride, 25 mM Tris-HCl, pH 7.5, prior to crystallisation trials. The sitting drop vapor diffusion method was used with 0.2 µL drops in a 1:1 ratio of protein at 9 mg mL\(^{-1}\) with reservoir at 20 °C. A slender orthorhombic crystal of Rap1a, approximate dimensions 150 x 35 x 35 µm, grew within one month using a reservoir of 25% PEG 3350 and 100 mM Bis-Tris, pH 5.5. Monoclinic blocks of Rap2a were grown by combining protein at a concentration of 13.5 mg mL\(^{-1}\) in 100 mM sodium phosphate, pH 6.4, with a reservoir of 25% PEG 1000 and 0.1 M MES, pH 6.5. These crystals attained a maximum dimension of 200 µm within five days. Crystals were soaked briefly in mother liquor adjusted to contain 20 % (v/v) glycerol as a cryo-protectant prior to being flash-cooled in liquid nitrogen then used for diffraction experiments.

Datasets for Rap1a and Rap2a were collected at Diamond Light Source, indexed and integrated in iMOSFLM (Battye et al., 2011) and scaled with SCALA (Evans, 2006). The structures were solved by molecular replacement in PHASER (McCoy et al., 2007). The starting model for Rap1a was taken from the structure of an Ssp1-Rap1a complex that had been solved by Dr. Velupillai Srikannathasan (PDB code: 4BI8; Srikannathasan et al., 2013); while that for Rap2a was a poly-alanine model prepared from the structure of the related Rap protein, Rap2b (PDB code: 4B6I; English et al., 2012). All non-protein atoms were removed from the models and initial B-factors were adjusted to match the Wilson B-values obtained during scaling. Following successful molecular replacement the models were subject to several rounds of rigid-body refinement followed by real-space refinement in REFMAC5 (Murshudov et al., 2011),
interspersed with inspection of the electron density maps and model improvement in COOT (Emsley et al., 2010). Strict NCS restraints and van der Waals contact distance restraints were applied from the start of refinement. NCS restraints were dropped at a late stage in the refinement following addition of water molecules to the models. Side chains with more than one conformation visible in the density maps were modelled as rotamers with partial occupancies. Model quality was validated using the MolProbity server (Chen et al., 2010) and the validation tools found in COOT. Coordinates and structure factors for Rap1a and Rap2a have been deposited with the PDB, under codes 3ZFI and 3ZIB respectively.
CHAPTER 6

Rap1a and Rap2a –

RESULTS AND DISCUSSION
6.1 Structure of Rap1a

The structure of Rap1a was determined to 1.98 Å resolution and the amino acid sequence and assigned secondary structure is depicted in Figure 6.1. Crystallographic statistics are presented in Table 6.1.

Figure 6.1 Overall structure of Rap1a. a) The sequence of mature Rap1a. The 26-residue N-terminal signal sequence is not shown. The assigned α-helical secondary structure is shown and numbered. Residues involved in disulfide bond formation are coloured in yellow and those that contribute to the dimer interface are encased in blue. b) Cartoon representation of the Rap1a dimer with labelled helices, Nt and Ct mark the N and C-terminal positions. The disulfides formed between Cys78 and Cys122 are shown as yellow sticks. c) Residues and hydrogen bonds (dashed lines) at the dimer interface. C atoms are grey and cyan to distinguish the subunits.

The Rap1a subunit displays a compact globular structure constructed from five α-helices that assemble to form the highly stable, symmetric dimer that constitutes the asymmetric unit (Figure 6.1). This is consistent with the size-exclusion chromatography data, which identified that only a dimer was observed in solution (not shown). The NCS is highly conserved with an RMSD for a least squares overlay of 96 Ca positions of 0.6 Å. A disulfide bond is formed between Cys78 on the α2-α3 loop and Cys122 at the C-terminal region. This interaction appears crucial to stabilising the
subunit fold since it helps position $\alpha_2$, $\alpha_3$ and $\alpha_5$ close to each other and these segments of secondary structure provide the side chains that form the hydrophobic core of the subunit. During initial recombinant expression tests it was noted that soluble protein was produced only in *E. coli* Rosetta-gami (DE3) cells; a strain that promotes the formation of disulfide bonds in the cytoplasm and so mimics what might occur in the oxidative environment of the periplasm. This suggests the covalent bond is necessary for correct folding to occur, and for stability of the Rap1a fold and dimeric quaternary structure.

Table 6.1 Crystallographic statistics for Rap1a and Rap2a

<table>
<thead>
<tr>
<th>Structure / PDB code</th>
<th>Rap1a / 3ZFI</th>
<th>Rap2a / 3ZJB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>$C22_1$</td>
<td>$P2_1$</td>
</tr>
<tr>
<td>Unit cell dimensions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a$, $b$, $c$ (Å)</td>
<td>82.7, 93.0, 51.3</td>
<td>39.7, 81.4, 58.5</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>46.50-1.98 (2.09-1.98)</td>
<td>40.67-1.90 (2.00-1.90)</td>
</tr>
<tr>
<td>No. reflections</td>
<td>63106 (8849)</td>
<td>214423 (31351)</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>14089 (2003)</td>
<td>29067 (4190)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.9 (100.0)</td>
<td>99.4 (98.9)</td>
</tr>
<tr>
<td>$R_{merge}$</td>
<td>6.6 (48.6)</td>
<td>9.7 (51.6)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>4.5 (4.4)</td>
<td>7.4 (7.5)</td>
</tr>
<tr>
<td>$&lt;I/\sigma(I)&gt;$</td>
<td>13.4 (2.7)</td>
<td>15.3 (4.0)</td>
</tr>
<tr>
<td>Wilson $B$ (Å$^2$)</td>
<td>31.5</td>
<td>21.7</td>
</tr>
<tr>
<td>$R_{work}/R_{free}$</td>
<td>19.2, 23.4</td>
<td>18.3, 23.4</td>
</tr>
<tr>
<td>Number of residues / waters</td>
<td>185, 57</td>
<td>377, 137</td>
</tr>
<tr>
<td>RMSD bond lengths (Å) / bond angles (°)</td>
<td>0.018 / 1.925</td>
<td>0.019 / 1.745</td>
</tr>
<tr>
<td>Mean $B$-factors (Å$^2$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Main chain, side chain</td>
<td>Chain A 34.3, 40.9</td>
<td>Chain A 19.7, 23.7</td>
</tr>
<tr>
<td></td>
<td>Chain B 35.2, 40.0</td>
<td>Chain B 23.0, 28.3</td>
</tr>
<tr>
<td></td>
<td>38.4</td>
<td>Chain C 22.4, 27.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chain D 18.9, 24.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30.6</td>
</tr>
<tr>
<td>Ramachandran plot</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favoured regions (%)</td>
<td>97.0</td>
<td>99.0</td>
</tr>
<tr>
<td>Allowed regions (%)</td>
<td>3.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Outliers</td>
<td>No outliers</td>
<td>No outliers</td>
</tr>
</tbody>
</table>
Interactions involving residues on α2 make the major contribution to dimer formation. Each subunit contributes a surface area of $1130 \text{ Å}^2$ to the dimer interface, which is 20% of the solvent accessible surface area (ASA, $5700 \text{ Å}^2$) of a subunit. Such a percentage of surface area is indicative of a stable association. Nine residues from each subunit form a network of hydrogen bonds using both main chain and side chain functional groups (Ser36, Asn40, Leu41, Glu59, Tyr64, Asp70, Lys75, Arg107 and Glu113; Figure 6.1). There are also indirect hydrogen bonding interactions via well-ordered water molecules that link a number of side chains and main chains (not shown). Hydrophobic interactions that contribute to the stability of the dimer mainly involve the aliphatic side chains of Val39, Ile63, Leu66, Val68 and Ala71 but also Tyr80 (not shown).

Overlays were attempted with the structures of known immunity proteins, including Rap2a (Section 6.2), Rap1b, Rap2b from *S. marcescens* (English *et al*., 2012), and the Tai4 proteins (Zhang *et al*., 2013), however these proteins were all dissimilar from Rap1a. Thus Rap1a adopts a unique fold for an immunity protein. A search using the *DALI* server identified two proteins, YmgD and HdeA from *E. coli*, with some degree of structural similarity to Rap1a. Both are acidic, α-helical periplasmic proteins, and overlays in *COOT* and *PyMOL* show that almost all secondary structural elements match. YmgD and HdeA contain 89 and 90 amino acids respectively, discounting residues in periplasmic localisation sequences - for comparison, mature Rap1a contains 101 residues.

YmgD (UniProt: P0AB46) shares 13% sequence identity with Rap1a and is a functionally uncharacterised homodimeric protein that becomes up regulated in response to alkaline-stress (pH $> 8.0$, Maurer *et al*., 2005). The structure has been determined by NMR (PDB codes: 2LRV and 2LRM, Wang *et al*., unpublished) and aligns 81 Cα positions to Rap1a with an RMSD of 3.4 Å (Figure 6.2).
Figure 6.2 Superimposition of YmgD and HdeA on Rap1a.  

a) Superimposition of YmgD on Rap1a.  Rap1a is coloured cyan and YmgD is coloured magenta.  
b) Superimposition of HdeA on Rap1a.  HdeA is coloured orange.  Disulfides are displayed as sticks in yellow, protein backbones are displayed as Cα traces.

HdeA (UniProt: P0AES9), which has marginal sequence identity with Rap1a (~5%), is a periplasmic acid-stress chaperone.  HdeA forms a well-ordered, inactive homodimer at neutral pH, but under acidic conditions (pH < 3) it dissociates into partially unfolded monomers and binds to other periplasmic proteins; acting as a chaperone that prevents aggregation (Foit et al., 2013; Hong et al., 2005).  The crystal structure has been determined by two different groups (PDB codes: 1DJ8, Gajiwala & Burley, 2000; and 1BG8, Yang et al., 1998) and aligns 72 Cα positions with Rap1a, with an RMSD of 2.9 Å (Figure 6.2).  Disulfide bonds are present at a similar position in both YmgD (Cys36-87) and HdeA (Cys39-87), but different from that in Rap1a (Cys78-122).  Such similarity between Rap1a and these periplasmic chaperones suggests the possibility of an evolutionary link from a common ancestor.

In the structure of the Ssp1-Rap1a complex (Srikannathasan et al., 2013) there are two molecules in the asymmetric unit, one Ssp1 and one Rap1a, and a crystallographic two-fold axis generates a heterotetramer (Figure 6.3).  The overall
dimensions of this assembly are approximately 90 x 50 x 50 Å. The solvent accessible surface area of Ssp1 is 5840 Å², and about 16% of this is occluded when the complex with Rap1a is formed (Srikannathasan et al., 2013).

![Image](image.png)

**Figure 6.3 The heterotetrameric Ssp1-Rap1a complex.** Ssp1 is shown as van der Waals surface in orange, and the Rap1a subunits as grey and cyan ribbons. A dashed line running through the centre of the figure indicates the position of the crystallographic axis of symmetry. Disulfide bonds in Rap1a are displayed as yellow sticks. An arrow indicates the location of the active site in Ssp1 subunit 1. Rap1a sterically blocks the Ssp1 active site; Rap1a α3 is highlighted. This figure has been adapted from Srikannathasan et al., 2013.

Both α3 and the α3-α4 loop of Rap1a are directly positioned to block the Ssp1 active site (Figure 6.3). Gln84 of Rap1a forms a hydrogen bond with Nδ of the catalytic Ssp1 His133 and is a clear marker of the steric block provided by the immunity protein. An overlay (not shown) of the two Rap1a structures, alone and in complex with Ssp1, gives an RMSD of 0.4 Å for 94 Cα atoms indicating that the overall structure is highly similar with no major conformational differences. Therefore, Rap1a is a pre-formed binding partner for Ssp1 under the right redox state. However, there is a localised effect following complex formation. In Rap1a, the segment linking α2 and α3, residues 74 to 84 shows high average B-factors, 76.4 Å², compared to the overall average B-factor,
31.5 Å$^2$, and relatively diffuse electron density. This region contributes to Ssp1-Rap1a complex formation and becomes well ordered with an average $B$-factor of 15.6 Å$^2$ compared to the overall $B$-factor of 22.6 Å$^2$ for the Rap1a component of the complex (Srikannathasan et al., 2013).

### 6.2 Structure of Rap2a

Rap2a, like Rap1a, is predicted to be localised in the periplasm and could only be produced in soluble recombinant form using the *E. coli* Rosetta-gami (DE3) strain. It is also a stable dimer in solution as shown by size-exclusion chromatography (English et al., 2012). The structure was determined at 1.9 Å resolution with four molecules in the asymmetric unit arranged as two dimers (Table 6.1). These four molecules, subunits A:B and C:D, are similar overall with RMSD values between superimposed Ca atoms ranging from 0.6 Å - 0.7 Å. Minor deviations from NCS are present in a five-residue loop between $\alpha 2$ and $\alpha 3$, Gly67-Leu71, and indicate some conformational freedom in that part of the molecule. The ASA of a Rap2a subunit averages out at approximately 6110 Å$^2$. Each dimer covers an area that is approximately 20 % of the ASA.

The Rap2a subunit displays a compact globular structure of five $\alpha$-helices with an extended loop linking $\alpha 3$ to $\alpha 4$ (Figure 6.4). A disulfide bond between Cys42 and Cys102 links $\alpha 1$ with $\alpha 4$ and interactions from these elements of secondary structure help to create the helical bundle fold and in particular to align $\alpha 2$. Together $\alpha 1$, $\alpha 2$, the $\alpha 2$-$\alpha 3$ loop and $\alpha 4$ form the dimer interface, giving rise to a two-fold NCS axis, and here a combination of hydrogen bonding, salt bridge and van der Waals interactions serve to stabilise the association (not shown). Main chain hydrogen-bonding contributions come from the amides of Ser48, Ala49, Met97, Thr98 and Met99 on both chains. Side chain contributions come from the hydroxyl groups of Tyr28 and Tyr47,
and carboxylates of Glu51, Asp55 and Asp104. There are in addition several solvent mediated contacts that serve to link functional groups on partner subunits (not shown). The side chains of Leu39, Ile43, Tyr47, Val52, Met99 and Ile103 are involved in van der Waals interactions with the partner subunit to stabilise the dimer.

Figure 6.4 Overall structure of Rap2a. a) The amino acid sequence with assigned secondary structure and helices numbered. Residues involved in disulfide bond formation are coloured in yellow and residues involved in subunit-subunit interactions are encased in blue. b) Cartoon representation of the Rap1a dimer with subunits coloured brown and green. The disulfides, N and C termini are labelled. c) Cartoon representation of Rap2b (magenta; English et al., 2012) and EcTai4 (grey; Zhang et al., 2013) superimposed on Rap2a (green). Selected residues are displayed as sticks. In EcTai4, Asp54 and Asp47 contribute to the hydrogen-bonding network at the dimer interface and are absolutely conserved in Rap2b, while in Rap2a Asp47 is replaced by Ser48. EcTai4 Arg40 and Glu74, conserved in Rap2b, have a major role in binding with the T6 effector EcTae4 (Zhang et al., 2013) - these residues are replaced by Val41 and Thr75 in Rap2a.

Despite a low level of sequence conservation, the structural similarities of five T6SS immunity proteins (Rap1b, Rap2a, Rap2b, EcTai4 and STTai4) indicate an orthologous sub-set. This structurally defined set is consistent with the designation of a ‘Tai4’ family of immunity proteins cognate to Tae4 effectors (Russell et al., 2012). Pairwise comparisons indicate a range of sequence identities from 18 to 38%, and
matching between 90 and 92 Ca positions gives an RMSD range of 1.2 to 1.7 Å, indicative of a close structural similarity (Table 6.2). Such similarity is exemplified by superimposition of Rap2a, Rap2b and EcTai4 (Figure 6.4). Strikingly, the extended or protruding loop structure is structurally conserved between these species, despite high variability in the amino acid sequence (English et al., 2012; Zhang et al., 2013). Parts of the EcTai4 subunit involved in interaction with an effector, identified by Val41 and Thr75 of Rap2a, are also well conserved in terms of three-dimensional structure (Figure 6.4).

<table>
<thead>
<tr>
<th>Structure (PDB code)</th>
<th>STTai4 (4HFF)</th>
<th>Rap2b (4B6I)</th>
<th>EcTai4 (4HFK)</th>
<th>Rap1b (4AX2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>% sequence identity to Rap2a</td>
<td>38</td>
<td>23</td>
<td>18</td>
<td>21</td>
</tr>
<tr>
<td>No. Ca / RMSD (Å)</td>
<td>92 / 1.3</td>
<td>92 / 1.2</td>
<td>90 / 1.6</td>
<td>91 / 1.7</td>
</tr>
</tbody>
</table>

This sub-set of immunity proteins also displays similar dimer structures and the Cys42 - Cys102 disulfide bond in Rap2a is conserved in the other two S. marcescens proteins, Rap1b and Rab2b, highlighting the importance of this covalent interaction to the creation of the subunit fold and stable quaternary structure. In the structure of EcTai4 and STTai4 there are conserved cysteines that match to the disulfide forming residues in the Rap proteins however they are in a reduced form. In STTai4 for example the Cys48 SG Cys108 SG distance is 3.6 Å and the electron density unambiguously defines reduced cysteine residues. This difference in the redox states may simply reflect distinct experimental conditions.
6.3 Structural basis of immunity to T6 effectors - Conclusions

Rap2a has a similar fold to that seen previously for the Tai4 family of immunity proteins including Rap1b, Rap2b, EcTai4 and STTai4. In contrast, Rap1a adopts a novel fold for an immunity protein. These two structurally distinct families of Tai4 immunity proteins are built upon similar underlying principles; namely stable, dimeric small α-helical bundles. *S. marcescens* Rap1a and Rap2a appear to rely on the formation of disulfide linkages for folding and activity. Intriguingly, the unexpected structural similarity of Rap1a to the periplasmic chaperones HdeA and YmgD permits speculation that these proteins share a common evolutionary ancestor.

The Type VI associated endopeptidase effectors are highly basic proteins as exemplified by Ssp1 and Ssp2, with predicted pI values of 9.1 and 9.3 respectively. The immunity proteins are acidic; the pI values for the four Rap proteins fall in the range 5.1 to 6.3 after omitting the signal peptides. Such complementarity of charge contributes to the high affinity interactions that support complex formation, despite distinctive structures and variation in sequence, and may generate long-range electrostatic attraction to assist correct binding.
CHAPTER 7

REFERENCES
References


Anantharaman V and Aravind L (2003) Evolutionary history, structural features and biochemical diversity of the NlpC/P60 superfamily of enzymes. Genome Biol. 4: R11


Brooks TM, Unterweger D, Bachmann V, Kostiuk B and Pukatzki S (2013) Lytic activity of the \textit{Vibrio cholerae} type VI secretion toxin VgrG-3 is inhibited by the antitoxin TsaB. J. Biol. Chem. 288: 7618-7625


Calisto BM, Perez-Gil J, Bergua M, Querol-Audi J, Fita I and Imperial S (2007) Biosynthesis of isoprenoids in plants: structure of the 2C-methyl-D-erythritol 2,4-
cyclophospho diphosphate synthase from Arabidopsis thaliana. Comparison with the bacterial enzymes. Protein Sci. 16: 2082-2088


Crane CM, Kaiser J, Ramsden NL, Lauw S, Rohdich F, Eisenreich W, Hunter WN,


Li M, Le Trong I, Carl MA, Larson ET, Chou S, De Leon JA, Dove SL, Stenkamp RE
and Mougous JD (2012) Structural basis for type VI secretion effector recognition by a
cognate immunity protein. PLoS Pathog. 8: e1002613

Lichtenthaler HK, Schwender J, Disch A and Rohmer M (1997) Biosynthesis of
isoprenoids in higher plant chloroplasts proceeds via a mevalonate-independent
pathway. FEBS Lett. 400: 271-274


Lucock M (2000) Folic acid: nutritional biochemistry, molecular biology, and role in

Lüttgen H, Rohdich F, Herz S, Wungsintaweekul J, Hecht S, Schuhr CA, Fellermeier
M, Sagner S, Zenk MH, Bacher A and Eisenreich W (2000) Biosynthesis of terpenoids:
YchB protein of Escherichia coli phosphorylates the 2-hydroxy group of 4-

type VI secretion system displays antimicrobial properties. Proc. Natl. Acad. Sci. 107:
19520-19524

Masters PA, O’Bryan TA, Zurlo J, Miller DQ and Joshi N (2003) Trimethoprim-

regulates genes for flagellar motility, catabolism, and oxidative stress in Escherichia

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC and Read RJ

Morgan RE, Batot GO, Dement JM, Rao VA, Eadsforth TC and Hunter WN (2011)
Crystal structures of Burkholderia cenocepacia dihydropteroate synthase in the apo-
form and complexed with the product 7,8-dihydropteroate. BMC Struct. Biol. 11: 21

Mougous JD, Cuff ME, Raunser S, Shen A, Zhou M, Gifford CA, Goodman AL,
virulence locus of Pseudomonas aeruginosa encodes a protein secretion apparatus.
Science 312: 1526-1530

Moynie L, Schnell R, McMahon SA, Sandalova T, Boukerou WA, Schmidberger JW,
and Schneider G (2013) The AEROPATH project targeting Pseudomonas aeruginosa:
crystallographic studies for assessment of potential targets in early-stage drug

Mpamhanga CP, Spinks D, Tulloch LB, Shanks EJ, Robinson DA, Collie IT, Fairlamb
three binding modes: novel and selective pteridine reductase 1 inhibitors derived from
fragment hits discovered by virtual screening. J. Med. Chem. 52: 4454-4465


O'Rourke PEF, Eadsforth TC, Fyfe PK, Shepherd SM and Hunter WN (2011) *Pseudomonas aeruginosa* 4-amino-4-deoxychorismate lyase: spatial conservation of an active site tyrosine and classification of two types of enzyme. PLoS One 6: e24158


Schwarz S, Hood RD and Mougous JD (2010) What is type VI secretion doing in all those bugs? Trends Microbiol. 18: 531-537


APPENDIX 1

Representative experimental data from the purification of $Bc$IspF
**Imobilised metal ion (Ni\(^{2+}\))-affinity chromatograph for BclspF-His\(_6\)**

- **1:** Protein marker
- **2:** Flow-through
- **3:** Total insoluble protein
- **4:** Total soluble protein
- **5:** Histidine-rich proteins
- **6 & 7:** BclspF-His\(_6\) (200 mM imidazole peak)

*SeeBlue Plus2, pre-stained protein standard, Life Technologies*

**Flow-through**

- **Endogenous E. coli histidine-rich proteins**
- **50 mM imidazole**
- **200 mM imidazole**
Size-exclusion chromatograph for BclspF-His$_6$

Elution volume at peak maximum: 217 ml
Corresponding mass according to column calibration: 52 kDa
Predicted mass (ProtParam): 57 kDa for trimer

Yield: 50 mg of BclspF from 2 L of LB$_{carb}$ culture

SDS-PAGE; fractions centred on the peak

Only those fractions near the centre of the main protein peak were taken forward for crystallisation trials.
BcIspF-His$_6$

MALDI-TOF MS

Experimental mass: 19324.58 Da
Predicted mass: 19328.8 Da
APPENDIX 2

Publications
**Pseudomonas aeruginosa** 4-Amino-4-Deoxychorismate Lyase: Spatial Conservation of an Active Site Tyrosine and Classification of Two Types of Enzyme

Patrick E. F. O’Rourke, Thomas C. Eadsforth, Paul K. Fyfe, Sharon M. Shepherd, William N. Hunter*

Division of Biological Chemistry and Drug Discovery, College of Life Sciences, University of Dundee, Dundee, United Kingdom

---

**Abstract**

4-Amino-4-deoxychorismate lyase (PabC) catalyzes the formation of 4-amino-4-deoxychorismate, and release of pyruvate, during folate biosynthesis. This is an essential activity for the growth of Gram-negative bacteria, including important pathogens such as *Pseudomonas aeruginosa*. A high-resolution (1.75 Å) crystal structure of PabC from *P. aeruginosa* has been determined, and sequence-structure comparisons with orthologous structures are reported. Residues around the pyridoxal 5'-phosphate cofactor are highly conserved adding support to aspects of a mechanism generic for enzymes carrying that cofactor. However, we suggest that PabC can be classified into two groups depending upon whether an active site and structurally conserved tyrosine is provided from the polypeptide that mainly forms an active site or from the partner subunit in the dimeric assembly. We considered that the conserved tyrosine might indicate a direct role in catalysis: that of providing a proton to reduce the olefin moiety of substrate as pyruvate is released. A threonine had previously been suggested to fulfill such a role prior to our observation of the structurally conserved tyrosine. We have been unable to elucidate an experimentally determined structure of PabC in complex with ligands to inform on mechanism and substrate specificity. Therefore we constructed a computational model of the catalytic intermediate docked into the enzyme active site. The model suggests that the conserved tyrosine helps to create a hydrophobic wall on one side of the active site that provides important interactions to bind the catalytic intermediate. However, this residue does not appear to participate in interactions with the C atom that undergoes an sp² to sp³ conversion as pyruvate is produced. The model and our comparisons rather support the hypothesis that an active site threonine hydroxyl contributes a proton used in the reduction of the substrate methylene to pyruvate methyl in the final stage of the mechanism.

---

**Introduction**

Aminodeoxychorismate lyase catalyzes the formation of para-aminobenzoate (PABA) from 4-amino-4-deoxychorismate (Figure 1) as part of the biosynthetic route to folate [1][2]. The enzyme, encoded by the pabC gene, is labeled as PabC.

In *Escherichia coli* and many other bacteria, PABA is synthesized in two steps starting from chorismate [3]. The first step produces 4-amino-4-deoxychorismate [4] by transferring ammonia, derived from glutamine, to chorismate. This reaction is performed by PABA synthase, a heterodimer of proteins encoded by the pabA and pabB genes [5]. 4-amino-4-deoxychorismate is then converted to PABA with the loss of pyruvic acid (Figure 1). The pyridoxyl 5'-phosphate (PLP) - dependent PabC catalyzes this second step [2][3][6][7]. In the Gram-positive *Bacillus subtilis* a third chemical step is required for PABA biosynthesis. Chorismate is first converted to 2-amino-2-deoxyisochorismate before production of 4-amino-4-deoxychorismate, which is then converted to PABA and pyruvate by PabC [8].

PABA is required for the biosynthesis of folic acid and by extension also of essential metabolites such as thymidylate and methionine. In addition, PABA is required for biosynthesis of antibiotics such as cefadroxil [9] and certain *Streptomyces* even possess additional copies of pabC, located in antibiotic production operons [10].

The inhibition of PabC would reduce PABA levels and so deplete the supply of folic acid available to the microorganism. Provided that the microbe cannot obtain sufficient PABA from the cellular environment or by other means these cells would be expected to cease growth and die. This is what is observed in pabC knockouts in *E. coli* [7], *Pseudomonas aeruginosa* [11], Helicobacter pylori [12] and Acinetobacter baylyi [13]. A few anti-microbial agents exert their effect through inhibition of folate biosynthesis. Sulphonamides for example deplete bacterial intracellular folate levels by inhibiting PABA synthase and dihydropterotate synthetase [14]. Combining novel PabC inhibitors with sulphonamides or with other anti-folates such as the dihydrofolate reductase inhibitor trimethoprim, could produce a synergistic anti-microbial effect and new inhibitors of PabC would allow such a hypothesis to be tested.

X-ray crystal structures of PabC from *E. coli* [15], *Thermus thermophilus* [16] and *Legionella pneumophila* [Protein Data Bank...
tobacco-etch virus (TEV) protease cleavage site. The resulting vector (Novagen) that produces a hexahistidine-tag linked via a vector (Invitrogen) then ligated into a modified pET15b expression site in bold)

Materials and Methods

Cloning, protein purification and analytical ultracentrifugation

The P. aeruginosa pabC gene (locus tag: PA2964) was amplified from genomic DNA (American Type Culture Collection 47065, strain PAO1) by PCR using the following primers:

**PaPabC forward primer**

5'-CAT ATG CTG GAC TGG GTC GAC-3' (NdeI restriction site in bold)

**PaPabC reverse primer**

5'-GGA TCG TCA GAA ATC CAG GTC G-3' (BamHI restriction site in bold)

The PCR product was blunt-end ligated into a TOPO cloning vector (Invitrogen) then ligated into a modified pET15b expression vector (Novagen) that produces a hexahistidine-tag linked via a tobacco-etch virus (TEV) protease cleavage site. The resulting plasmid was heat-shock transformed into E. coli BL-21 (DE-3) GOLD cells (Novagen). Cultures were grown at 37°C in 2 L of Luria-Bertani medium supplemented with 100 μg mL⁻¹ carbenicillin. Gene expression was induced, when the OD₆₀₀ reached 0.6, by addition of 1 mM IPTG and growth continued for a further 16 hours at 18°C. Cells were harvested by centrifugation (3000 x g at 4°C for 30 minutes), resuspended in 50 mM Tris-HCl, 200 mM NaCl, pH 7.5, and lysed by passage through a French pressure cell. Cell lysate was clarified by ultracentrifugation (40000 x g at 4°C for 30 minutes) and the supernatant was passed through a 0.2 μm filter.

The lysate was loaded onto an immobilised metal ion affinity chromatography column (5 mL HisTrap HP, GE Healthcare) with the assistance of an FPLC system (Acta Explorer, GE Healthcare). The hexahistidine-tagged protein was eluted from the column by a linear concentration gradient of imidazole and fractions were analysed by SDS-PAGE. Selected fractions were pooled and dialysed into 100 mM HEPES, 500 mM NaCl pH 7.5. Attempts to cleave the hexa-histidine tag by addition of TEV protease had produced large amounts of precipitate so it was decided to leave the purification tag in place. The protein was further purified by size exclusion gel chromatography (Superdex 200 26/60, GE Healthcare). This column had previously been calibrated with molecular weight standards, blue dextran (>2,000 kDa), thyroglobulin (669 kDa), ferritin (440 kDa), aldolase (158 kDa), conalbumin (75 kDa), ovalbumin (43 kDa), carbonic anhydrase (29.5 kDa), ribonuclease A (13.7 kDa) and aprotinin (6.5 kDa); (GE Healthcare; data not shown). Protein purity was assessed by SDS-PAGE, matrix assisted laser desorption/ionisation – time of flight mass spectrometry (MALDI-TOF MS) and electrospray ionisation mass spectrometry.

The protein concentration was determined spectrophotometrically using a theoretical extinction coefficient of 33,460 M⁻¹ cm⁻¹ (ProtParam, [26]).

Crystallization, X-ray data collection, structure solution and refinement

PaPabC was crystallized at 20°C by the hanging drop vapour diffusion method. The enzyme was concentrated to 33 mg mL⁻¹ in 100 mM HEPES, 500 mM NaCl pH 7.5. Stock solutions of PLP (100 mM in DMSO) and PABA (200 mM in deionized water) were prepared and added to the PabC solution to give final concentrations of 0.1 mM PLP and 10 mM PABA. The mixture was left to incubate for 3 hrs then crystallization drops were assembled by mixing 1 μL of the protein:ligand mixture with 1 μL of reservoir containing 10% (w/v) PEG 400, 1.8 M ammonium sulfate and 100 mM MES pH 6.5. Clumps of pale yellow,

![Image](https://example.com/image.png)
orthorhombic blocks grew over one week and a fragment of approximate dimensions 0.3 mm × 0.2 mm × 0.1 mm was used for data collection. Addition of cryoprotectant (25% v/v glycerol) was required to allow crystals to be mounted in a gaseous nitrogen cryostream at −173°C for diffraction experiments. Crystals were initially characterized using a copper anode X-ray generator (Rigaku 007) and an image plate detector (RAXIS IV++). The crystal was stored in liquid nitrogen for data collection at the European Synchrotron Radiation Facility (ESRF, Grenoble, France).

Diffraction data were collected on beamline ID29 at ESRF using a Quantum4 CCD detector. It proved beneficial to use the automated crystal annealing facility established on the station. Data were indexed and integrated using MOSFLM [27], and scaled with SCALA [28]. Crystallographic statistics are presented in Table 1. The search model for molecular replacement was prepared by generating a poly-alanine model of the E. coli PabC structure (PDB code 1et0, [15]) from which all waters and non-protein molecules had been removed. This model shares 35% sequence identity with PabC. Molecular replacement calculations were performed using MOLREP [29].

Refinement was performed in REFMAC5 [30][31] and the model was inspected and manipulated in COOT [32]. Strict non-crystallographic symmetry restraints were employed in the early stages of the refinement and then removed. The presence of the PLP cofactor covalently bound to Lys140 was confirmed by calculating an omit electron density map (Figure 2). The N-terminal histidine tag was not visible in electron density maps. A number of ligands (tetaethylene glycol, 1,2-ethanediol, di(hydroxyethyl)ether, glycerol, chloride and sulfate ions) were assigned on the basis of the difference density and chemical environment, and refined satisfactorily. Both tetaethylene glycol and ethylene glycol are likely decomposition products of PEG400 used in crystallization, impurities or represent partially ordered molecules. MOLPROBITY [33] was used to assess model geometry, AREAIMOL [31] and the Protein Interfaces, Surfaces and Assemblies, EMBL server [http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html; [34]] were used to calculate surface and interface areas. Figure 1 was prepared with CHEMDRAW (Adept Scientific), Figure 5 with ALINE [35] and the remainder using PyMOL [36] and ILLUSTRATOR (Adobe Systems Inc.).

Amino acid sequence alignments and analysis

A search of the UniProt database with the Enzyme Commission number for PabC (4-amino-4-deoxychorismate lyase; 4.1.3.38) identified 368 entries. These were clustered such that sequences

<table>
<thead>
<tr>
<th>Table 1. Crystallographic statistics.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Space group</strong></td>
</tr>
<tr>
<td><strong>Unit cell dimensions:</strong> a, b, c (Å)</td>
</tr>
<tr>
<td><strong>Resolution range (Å)</strong></td>
</tr>
<tr>
<td><strong>Unique reflections</strong></td>
</tr>
<tr>
<td><strong>Completeness (%)</strong></td>
</tr>
<tr>
<td><strong>&lt;I/σ(I)&gt;</strong></td>
</tr>
<tr>
<td><strong>Multiplicity</strong></td>
</tr>
<tr>
<td><strong>Rmerge</strong></td>
</tr>
<tr>
<td><strong>No. of reflections</strong></td>
</tr>
<tr>
<td>Rcryst, Rfree</td>
</tr>
<tr>
<td><strong>Protein residues</strong></td>
</tr>
<tr>
<td><strong>Ligands</strong></td>
</tr>
</tbody>
</table>

r.m.s. deviations from ideal geometry

| bond lengths (Å) | 0.025 |
| bond angles (˚) | 2.1 |

From Wilson Plot

Mean B values

Chain A, main chain atoms: 15.1
Chain A, side chain atoms: 18.2
Chain B, main chain atoms: 18.5
Chain B, side chain atoms: 20.7
Waters: 30.8
PLP: Chain A 15.8 Chain B 21.5
Other ligands 36.3

Ramachandran favoured/allowed (%) 93.7/99.4 (3 outliers: Arg90, Chains A, B Leu89, Chain B)

*Values in parentheses refer to the highest resolution shell (1.84–1.75 Å).

Rmerge = ∑|I(hkl)−<I(hkl)>|/∑I(hkl), where I(hkl) is the intensity of the i-th measurement of reflection hkl and <I(hkl)> is the mean value of I(hkl) for all i measurements.

Rfree = Σ|Fobs|−|Fcalc|/Σ|Fobs|, where Fobs is the observed structure factor and Fcalc is the calculated structure factor.

Ramachandran is the same as Rcryst except calculated with a subset, 5%, of data that are excluded from refinement calculations.

doi:10.1371/journal.pone.0024158.t001
with a 90% or greater similarity to each other were grouped and a representative sequence identified. This reduced the list to 197 entries. Sequences consisting of more than 400 amino acids were removed and this further reduced the list to 181 entries. These were aligned using MUSCLE [37] and then inspected following pairwise alignment with PaPabC. A conservative group of 129 putative PabC sequences was then identified following the removal of sequences that were fragmentary, labeled as other proteins or which displayed a sequence identity of less than or equal to 20%.

Molecular modeling of the catalytic intermediate

A model of the substrate, 4-amino 4-deoxychorismate, was prepared using PRODRG [38] and placed into the active site using the molecular graphics program COOT [32] with the position of the bound pyridoxyl 5’ phosphate as a guide to allow formation of the catalytic intermediate. The active site was prepared for docking of this ligand using ICM pro (MolSoft L.L.C; http://www.molsoft.com/index.html) with the centre of the ligand-binding site defined by a cavity that contained the residues within 5 Å of the 4-amino 4-deoxychorismate. The top ten docking poses, as scored by ICM Pro software gave similar orientations in the active site and one was selected for discussion and display.

Results and Discussion

Overall structure

Size exclusion gel filtration, carried out during protein purification, indicated a single species of approximate mass 50 kDa that is commensurate with a homodimeric species. In the crystal structure two polypeptide chains, labelled A and chain B together constitute the asymmetric unit in PaPabC (Figure 3), and the crystals have an estimated bulk solvent content of 45%. The polypeptide chains form a dimer (Figure 3) and, as described below, such oligomerization is required to create the enzyme active site. The presence of a PaPabC dimer is consistent with previous studies since all PabC structures solved to date are dimeric or in the case of T. thermophilus PabC, a dimer-of-dimers is observed [16].

The PaPabC subunits display a high degree of non-crystallographic symmetry. The rmsd following a least-squares fit of 270 Ca positions is 0.4 Å. It is therefore only necessary to detail one subunit and one active site.

The PaPabC subunit is constructed from two domains (Figure 4). The N-terminal domain (domain I) is smaller than the C-terminal domain (domain II). Domain I comprises residues 1–107, which
contains three α-helices (α1–3) and a four-stranded anti-parallel β-sheet in order β1-β4-β3-β2. Domain II comprises residues 112–271 that form five α-helices (α4–6), one short two-stranded parallel β-sheet (β5–β6) and one short two-stranded anti-parallel β-sheet (β7–β8). Residues 108–111 link the two domains. The PLP cofactor is covalently bound to Lys140 that is in domain II. The active site is formed in a cleft formed between the two domains (Figure 3).

Sixty residues from each subunit, provided by both domains, contribute to dimer formation. These residues participate in around 30 hydrogen-bonding associations and eight salt bridge interactions to stabilize the quaternary structure. Most of the residues that contribute to the dimer interface are found in two patches; residues Ala10 to Gly25 centred on α1 and residues Arg127 to Ala152 centered on α5.

The approximate dimensions of the dimer, which constitutes the asymmetric unit, are 80 Å×40 Å×40 Å and this displays a soluble accessible surface area of approximately 21,000 Å². The contact area between the two subunits occupies approximately 2,050 Å², about 15% of the surface of a subunit. These values are similar to those observed for other PabC proteins, for example the interface of the E. coli protein (PDB code 1e70) covers an area of 2,130 Å².

The structures of PabC from three organisms (E. coli, L. pneumophila and T. thermophilus) display a high degree of structural conservation with PabC. Least-squares fitting of the Cz positions of single subunits results in RMSD values of between 1.27 and 1.95 Å, with a coverage in the range of 81 to 92% of the structures. The sequence identity shared with PabC, falls in the range 26% to 35%.

The active site and mechanistic considerations

The active site of PabC is positioned between the two domains of the monomer close to the dimer interface (Figure 3). Here, the PLP cofactor is covalently linked to Lys140, bound and oriented by interactions with key residues, a selection of which are depicted in Figure 2. The PLP phosphate accepts hydrogen bonds donated by a cluster of main chain amide groups from Val200, Met201 and Ser237, and the side chain of Arg46. The side chain of Arg46 is itself positioned by interactions with the side chain of Glu28, which in turn accepts a hydrogen bond from His43 ND1. Two well ordered water molecules mediate hydrogen bonding networks linking the phosphate group to Asn178 OD1 and Arg202 N on one side, the main chain O and N groups of Thr29 on the other. The cofactor ring is sandwiched between Val197 and the main chain of a tripeptide segment comprising Val175-Phe176-Ser177. A hydrogen bond between Asn236 ND2 and the carbonyl of Phe176 in conjunction with hydrogen bonds formed between the amide groups and the side chain of Glu161 helps to determine the conformation of the tripeptide that forms a tight turn, between β9 and β10, that covers part of the cofactor.

The PLP methyl group makes van der Waals interactions with Gln147 CG and is positioned 3.2 Å distant from the carbonyl oxygen of Val175. The geometry of this latter contact is compatible with the presence of a C-H→O hydrogen bond [39].

The side chain of Gln147 is held in place by a hydrogen bond with Arg144 NE. The arginine side chain is itself positioned by hydrogen bonds formed with the carbonyl groups of Leu139 and His141. There are also three water molecules (data not shown), which form a complex network of hydrogen bonds linking the side chain of Arg144 to other sections of the enzyme. In addition the positive dipole from the α6, the closest residue on this helix is Gly199, provides an attractive force to interact with the PLP phosphate. The side chain of Glu173 displays a degree of conformational freedom with two rotamers being observed. One rotamer places the side chain to accept a hydrogen bond donated from PLP N1 (Figure 2). The alternative rotamer places the carboxylate at a distance of 3.7 Å from N1 (data not shown). Tyr92 from the partner subunit forms a hydrogen bond to PLP O3. The cofactor is also likely to posses an intra-molecular hydrogen bond involving O3 and NZ of Lys140. The side chain of Phe27, Val197 and aliphatic components of Arg46 and Arg144 surround the side chain of Lys140.

No structures of PabC-ternary complexes have yet been published to inform on aspects of substrate recognition and enzyme activity. Given the importance of such complexes we judged it important to get such data and attempted to determine structures of PaPabC in complex with a number of compounds, for example, the sample used for structure determination was crystallized in the presence of 10 mM PABA. However, we have been unable to identify electron density compatible with binding of any such ligand in the substrate-binding pocket. We judged therefore that the position of PLP and Lys140 NZ, together with the conservation of amino acid sequence and structures give an indication of where substrate binds and which residues are important for substrate recognition and catalytic function. Consideration of an amino acid sequence alignment and structural overlay of PaPabC with three other structures in the PDB proved particularly informative. To this we added molecular modelling and docking of the covalent intermediate formed during catalysis.

The contributions of 24 residues, within about 5 Å of PLP, to binding the cofactor and organising the active site are described above and summarized in Figure 5. Seven of these residues are strictly conserved in PabC of P. aeruginosa, E. coli, L. pneumophila and T. thermophilus (Phe27, Thr29, His43, Arg46, Gly199, Lys140 and Arg202 in PaPabC). With one exception these residues contribute to the active site using the chemical properties of their side chains. The exception is Gly199 where any alteration on Cz would result in a steric clash with the phosphate group of PLP. Another seven residues contribute to the active site primarily through the hydrogen bonding capacity of the main chain (Val175, Phe176, Ser177, Leu139, His141, Val200 and Met201). Phe176 is not conserved in any of the three other sequences but the remainder are conserved in at least one other orthologue. Ser237 interacts with the PLP phosphate using main chain and side chain groups. This residue is only different in the E. coli sequence where an alanine is observed and an additional water molecule binds nearby helping to satisfy hydrogen-bonding capacity (data not shown). A similar observation is made when considering PaPabC Asn236. The residue is conserved in EcPabC and LpPabC but changed to glycine in TpPabC where a water molecule is observed to replace the side chain (data not shown). Asn178 is strictly conserved in EcPabC and LpPabC but changed to serine in TpPabC. The side chain here simply helps to form one side of the cofactor-binding site. Val197 is conserved in EcPabC and conservatively substituted as leucine in the other two sequences. The residue at this position in the active site forms van der Waals interactions with the cofactor. In PaPabC Arg144 NE donates hydrogen bonds to Gln147 OE1 and the carbonyl of His141; interactions that help to create one side of the cofactor-binding site and to fix the position of Val197. These residues and the interactions they form appear to be conserved in EcPabC but not in LpPabC or TpPabC. Sequence and structure alignments (not shown) indicate that PaPabC Arg144 is replaced by tyrosine in both LpPabC and TpPabC and PaPabC Gln147 replaced by an alanine in LpPabC and a tyrosine in TpPabC. This is an intriguing observation to be discussed below.

There are three glutamates amongst the 24 residues discussed above, Glu28, Glu161 and Glu173. Glu28 helps to position Arg16 to bind the cofactor phosphate. Both residues are strictly
A catalytic mechanism, based on previous studies of PLP-dependent enzymes, was proposed for PabC prior to any structural data being available [10]. Following structure determination of EcPabC, a model of substrate binding was prepared to inform an assessment of the mechanism [15]. The contribution of PLP was considered the same, that is, following covalent linkage the enzyme-cofactor adduct undergoes a nucleophilic attack by the substrate amino group to produce a reactive external aldimine. This species then undergoes α-proton elimination, at C4, to the ε-amino group of a lysine to yield a reaction intermediate. In EcPabC Lys159 forms the covalent link to the cofactor and likely responsible for α-proton abstraction. Nakai et al., [15] further

A catalytic mechanism, based on previous studies of PLP-dependent enzymes, was proposed for PabC prior to any structural data being available [10]. Following structure determination of EcPabC, a model of substrate binding was prepared to inform an assessment of the mechanism [15]. The contribution of PLP was considered the same, that is, following covalent linkage the enzyme-cofactor adduct undergoes a nucleophilic attack by the substrate amino group to produce a reactive external aldimine. This species then undergoes α-proton elimination, at C4, to the ε-amino group of a lysine to yield a reaction intermediate. In EcPabC Lys159 forms the covalent link to the cofactor and likely responsible for α-proton abstraction. Nakai et al., [15] further
suggested that Thr28 of EcPabC supplies a proton to reduce the olefin moiety of the substrate as pyruvate is released. To recover the Schiff-base form of the enzyme, PABA is released and PLP reacts with Lys159.

Three residues described as contributing to the mechanism of action in EcPabC, Thr28, Lys97 and Lys159 [15], are strictly conserved in PaPabC as Thr29, Lys80 and Lys140. They are also conserved in LpPabC and TtPabC with the conservative change of Lys80 to arginine in the thermophilic orthologue (Figure 5). Lys140 of PaPabC is suitably placed to support α-proton abstraction and although Thr29 may engage in the proposed proton shuffling with both the olefin moiety of the substrate and Lys80 we noted an alternative residue that might fulfil this role. This residue may have been overlooked because Nakai et al., [15] had access to only a single structure of PabC and alignments based on sequences alone would not identify a tyrosine that is conserved in three-dimensions.

In EcPabC and PaPabC tyrosine OH groups (Tyr109 and Tyr92 respectively), from the partner subunit, donate a hydrogen bond to PLP O3 (Figures 2, 6); an interaction that influences the electronic structure of the cofactor [15]. The sequence alignment of PabC orthologues suggests a degree of variation in the identity of the amino acid that occupies the position of this tyrosine (Figure 5). The structural overlays are striking since they clearly identify a conserved tyrosine in relation to the cofactor (Figure 6). In EcPabC and PaPabC, Tyr109 and Tyr92 respectively, are contributed from the partner subunit. We call this position I. However, in LpPabC and TtPabC, Tyr144 and Tyr130 respectively, are contributed from the same subunit, position II.

We isolated 129 putative PabC sequences to investigate the distribution of tyrosine at these two positions seeking to address the hypothesis that if tyrosine is present at position I then it is likely absent from position II and vice versa. This hypothesis holds true for 125 out of the 129 sequences (data not shown). Two sequences, which share less than 30% sequence identity, have a tyrosine at both positions. Analyses of the sequences suggests that PabC can be classified into two groups depending on whether this active site tyrosine is provided by the same subunit which primarily forms the active site or from a partner subunit.

Furthermore the high degree of conservation of an active site tyrosine in three-dimensional structures of PabC suggests a functional role. One possibility is that this residue could act as a proton donor for the methylene to methyl reduction that completes the reaction. Such a reduction would have little influence on the cleavage of the C-O bond of the substrate since the driving force for that reaction would be production of a conjugated aromatic six-carbon ring system. Our model of the catalytic intermediate (Figure 7) was specifically created to investigate the potential role of this active site tyrosine and to re-evaluate the reduction process that completes the reaction.

The model suggests that Tyr112 donates a hydrogen bond to the C1 carboxylate whilst, with contributions from the partner subunit, the main chain amide of Tyr92 and hydroxyl group of Tyr22 donate hydrogen bonds to the other carboxylate of the intermediate (Figure 7). Tyr92 contributes to the positioning and electronic structure of the cofactor by virtue of the hydrogen bond involving the hydroxyl group with the PLP carbonyl (Figure 2). The bulk of the tyrosine side chain serves to create a hydrophobic wall on one side of the active site. Although placed to interact with the catalytic intermediate this residue does not appear to be involved in donating a proton to convert methylene to methyl as part of the catalytic cycle (Figure 7). Rather, Thr29, as first postulated by Nakai et al., [15] appears to fulfil such a role with the hydroxyl group placed about 3.5 Å from the olefin (Figure 7). Lys80 NZ is around 4.3 Å distant from the sp² hybridized C and it is possible that an alternative rotamer could bring this functional group into play with respect to either providing a proton itself or in activating the Thr29 hydroxyl for proton donation. We note that Thr29 is strictly conserved in 123 out of the 129 PabC sequences and conservatively substituted by serine in another five sequences. The single sequence in which we could not identify conservation in terms of a threonine or serine corresponding to Thr29 is the putative PabC from Pantonella vagans. This is a poorly conserved sequence sharing approximately 30% identity.

To further investigate the structure-activity relationship of PabC, and to assess the potential of this enzyme as a target for structure-based drug discovery will require the powerful combination of steady-state kinetic analysis with site-directed mutagenesis studies, and to carry out screening of chemical libraries. It will be imperative to extend from our work and to elucidate a means whereby structural data on enzyme-ligand complexes can be obtained.
Acknowledgments

We thank our colleague Mark Agacan for discussions, staff at ESRF for access to facilities and excellent support. Coordinates and structure factors are deposited with the PDB under accession code 2y4r.

References


Author Contributions

Conceived and designed the experiments: PEFO TCE WNH. Performed the experiments: PEFO TCE PKF SMS. Analyzed the data: PEFO TCE PKF WNH. Contributed reagents/materials/analysis tools: SMS. Wrote the paper: PEFO TCE WNH.
Structural basis for type VI secreted peptidoglycan 
DL-endopeptidase function, specificity and 
neutralization in Serratia marcescens

Velupillai Srikannathasan, a Grant English, b Nhat Khai Bui, c Katharina Trunk, b Patrick E. F. O’Rourke, a Vincenzo A. Rao, a Waldemar Vollmer, c Sarah J. Coulthurst b and William N. Hunter a*

aDivision of Biological Chemistry and Drug Discovery, College of Life Sciences, University of Dundee, Dundee DD1 5EH, Scotland,
bDivision of Molecular Microbiology, College of Life Sciences, University of Dundee, Dundee DD1 5EH, Scotland, and cCentre for Bacterial Cell Biology, Institute for Cell and Molecular Biosciences, Newcastle University, Newcastle upon Tyne NE2 4HH, England

‡Current address: Helmholtz Institute for Pharmaceutical Research Saarland, Department of Microbial Natural Products, Saarland University, 66123 Saarbrücken, Germany.

Correspondence e-mail: s.j.coulthurst@dundee.ac.uk, w.n.hunter@dundee.ac.uk

Some Gram-negative bacteria target their competitors by exploiting the type VI secretion system to extrude toxic effector proteins. To prevent self-harm, these bacteria also produce highly specific immunity proteins that neutralize these antagonistic effectors. Here, the peptidoglycan endopeptidase specificity of two type VI secretion-system-associated effectors from Serratia marcescens is characterized. These small secreted proteins, Ssp1 and Ssp2, cleave between γ,δ-glutamic acid and l-meso-diaminopimelic acid with different specificities. Ssp2 degrades the acceptor part of cross-linked tetratetrapeptides. Ssp1 displays greater promiscuity and cleaves monomeric tripeptides, tetrapeptides and pentapeptides and dimeric tetratetra and tetrapenta muropeptides on both the acceptor and donor strands. Functional assays confirm the identity of a catalytic cysteine in these endopeptidases and crystal structures provide information on the structure–activity relationships of Ssp1 and, by comparison, of related effectors. Functional assays also reveal that neutralization of these effectors by their cognate immunity proteins, which are called resistance-associated proteins (Raps), contributes an essential role to cell fitness. The structures of two immunity proteins, Rap1a and Rap2a, responsible for the neutralization of Ssp1 and Ssp2-like endopeptidases, respectively, revealed two distinct folds, with that of Rap1a not having previously been observed. The structure of the Ssp1–Rap1a complex revealed a tightly bound heteromeric assembly with two effector molecules flanking a Rap1a dimer. A highly effective steric block of the Ssp1 active site forms the basis of effector neutralization. Comparisons with Ssp2–Rap2a orthologues suggest that the specificity of these immunity proteins for neutralizing effectors is fold-dependent and that in cases where the fold is conserved sequence differences contribute to the specificity of effector–immunity protein interactions.

1. Introduction

Specialized secretion systems are key to bacterial fitness, survival and pathogenesis. They perform a myriad of roles in the processes that influence growth, colonization, attack and defence as bacteria interact with each other and with eukaryotic organisms (Filloux, 2011; Gerlach & Hensel, 2007). The recently identified type VI secretion system (T6SS), which is present in about 25% of Gram-negative bacteria for which genome sequences are available (Boyer et al., 2009), can be used to target bacterial and eukaryotic cells and is thus important for both inter-bacterial competition and pathogenesis (Burtnick et al., 2011; de Pace et al., 2010; Jani &
from pimelate) and Acta Cryst. like muramidase activity (Brooks et al., 2011; Hood et al., 2010; MacIntyre et al., 2010; Schwarz et al., 2010).

T6SS gene clusters encode the core components of a secretion machine capable of membrane perforation. This is a multi-protein needle-like assembly, resembling the contractile bacteriophage tail, that delivers effectors across three envelope layers in a single step (Bönemann et al., 2010; Cascales & Cambillau, 2012; Silverman et al., 2012). The clusters also encode accessory and post-translational regulatory components. Some T6SS-secreted effector proteins are also encoded within these large gene clusters. In the case of antibacterial T6SSs, effectors are always encoded adjacent to specific cognate immunity proteins. Immunity proteins bear appropriate signals to direct their localization to the compartment in which the toxic effectors act; for example, a Sec signal sequence guides localization to the periplasm. The provision of a cognate immunity protein provides protection against attack from sister cells (Coulthurst, 2013).

Different catalytic activities are associated with T6SS effectors. These include actin cross-linking and ADP-ribosylation, which disrupt the cytoskeletons of mammalian and amoebal cells (Pukatzki et al., 2007; Rosales-Reyes et al., 2012), and phospholipases, which degrade phosphatidyl-ethanolamine, the major component of the bacterial membrane (Russell et al., 2013). The best-characterized effectors are peptidoglycan hydrolases, which exhibit potent antibacterial activity (Russell et al., 2011, 2012). These enzymes degrade peptidoglycan, the heteropolymer that occupies the periplasmic space, imparts mechanical strength to the cell wall and helps to maintain the shape of Gram-negative bacteria.

Peptidoglycan hydrolases constitute a large enzyme family which displays a rich diversity in terms of structure, mechanism and specificity (Vollmer et al., 2008). There are enzymes specific for every glycosidic and amide bond in peptidoglycan. Such diversity is exploited to regulate bacterial cell growth, division and daughter-cell separation and, of particular interest here, to provide bactericidal properties that can be exploited in niche competition. Several classes of peptidoglycan hydrolases have been identified as T6SS effector proteins. These are termed Tse proteins. Tse3 from Pseudomonas aeruginosa is a muramidase, cleaving the glycan backbone (Russell et al., 2011), and the C-terminal domain of Vibrio cholerae VgrG-3 has been suggested to have lysozyme-like muramidase activity (Brooks et al., 2013). A diverse group of T6SS-secreted peptidoglycan amidases which cleave peptide cross-links has been described (Russell et al., 2012). Within this superfamily, four distantly related families with distinct cleavage specificities were defined. Family 1 (Tae1), which includes Tse1 from P. aeruginosa, hydrolyses peptide cross-links at the γ-d-glutamyl-meso-diaminopimelate peptide cross-link, representatives of families 2 and 3 (Tae2 and Tae3) hydrolyse peptide cross-links between d-meso-diaminopimelate and d-alanine, and a representative of family 4, Tae4 from Salmonella enterica serovar Typhimurium, also hydrolyses the γ-d-glutamyl-meso-diaminopimelate DL-bond. Structures of Tse1 from P. aeruginosa and of Tae4 from Salmonella Typhimurium and Enterobacter cloacae (Benz et al., 2012; Chou et al., 2012; Ding et al., 2012; Zhang et al., 2013) place these T6-secreted effectors in the NlpC/P60 family of endopeptidases, amidases and acyltransferases (named after the new lipoprotein C from Escherichia coli) and a 60 kDa extracellular protein from Listeria monocytogenes; Anantharaman & Aravind, 2003). The opportunistic pathogen Serratia marcescens has recently been shown to utilize T6-dependent secretion of two family 4 amidases, Ssp1 and Ssp2, to mediate antibacterial activity (English et al., 2012).

Bacteria that secrete potent peptidoglycan hydrolase effectors using the T6SS to attack competitors could generate a deleterious effect on their own population. To cope with this potential for friendly-fire damage, such bacteria also possess cognate immunity or resistance proteins located in the periplasm. These immunity proteins bind their cognate effectors with low nanomolar affinity to neutralize them in a highly specific manner (English et al., 2012). Four distinct families of putative immunity proteins are associated with the four Tae amidase families (Tai1–Tai4; Russell et al., 2012). In S. marcescens, the resistance-associated proteins Rap1a and Rap2a neutralize Ssp1 and Ssp2, respectively (English et al., 2012). Additionally, two other Rap proteins, Rap1b and Rap2b, are encoded together with Ssp1 and Ssp2 in the same locus within the T6SS gene cluster. The structures of Rap1b and Rap2b revealed a novel α-helix fold and a dimeric assembly (English et al., 2012), which was later observed in the Tai4 proteins from E. cloacae (EcTai4) and Salmonella Typhimurium (STTai4; Zhang et al., 2013). This fold is a template for some T6SS immunity proteins, called Tsi proteins, but not all. For example, analysis of the Tse1/Tsi1 effector/immunity protein combination found in P. aeruginosa revealed Tsi1 to be an all-β protein (Benz et al., 2012; Ding et al., 2012; Shang et al., 2012), whereas Tsi2, the immunity protein associated with the cytoplasmic effector Tse2, in the same organism has a helical fold that is distinct again (Li et al., 2012). That effector immunity defence systems based on distinct folds have evolved is perhaps to be expected, given the strong evolutionary pressure applied by multifarious secreted effectors. Indeed, such pressure may even have contributed to the acquisition of effector–immunity pairs encoded outside T6SS operons (for example, all three Tse/Tsi pairs in P. aeruginosa).

Here, we assess the peptidoglycan hydrolase specificity of Ssp1 and Ssp2 and the protective role of cognate immunity proteins. We define the specificity of the enzymes and confirm using site-directed mutagenesis that an in vivo cell-killing mechanism is directly attributable to their catalytic activity. We report crystallographic analyses of S. marcescens Ssp1, a disabled mutant (Ssp1-C50A), Rap1a, Rap2a and the heterotetrameric Ssp1–Rap1a complex. Our data provide information on the enzyme mechanism, aspects of substrate specificity, the structural classification of Ssp1 and Rap proteins, including the identification of a novel immunity protein fold, and the molecular details of how an effector
is neutralized by its cognate immunity protein, and suggest generic features related to function that allow the classification of these proteins into distinct groups. Finally, we consider diversity within the Tae4 family of effectors and their immunity proteins and how this may explain the presence of multiple homologues within the same organism.

2. Materials and methods

2.1. Recombinant protein production and effector–immunity protein complex formation

Recombinant Ssp1 and Ssp2 were expressed in E. coli BL21 (DE3), and Rap1a and Rap2a, minus their N-terminal periplasmic targeting sequences, were expressed in E. coli Rosetta-gami (DE3) and purified in high yield using established protocols (English et al., 2012). The predicted catalytic cysteine of Ssp1 and Ssp2 (Cys50) was mutated to an alanine using the QuikChange II Site-Directed Mutagenesis Kit (Stratagene) and the altered proteins were purified as for the wild-type samples. For co-expression of the Ssp1–Rap1a and Ssp2–Rap2a complexes, full-length mature Rap1a (residues 25–127) and Rap2a (residues 25–131), without signal peptides, were cloned using NdeI and XhoI restriction sites at the second multiple-cloning site of the co-expression vector pACYCDuet-1 (Novagen), and full-length Ssp1 (residues 1–163) and Ssp2 (residues 1–158) were subsequently cloned using NcoI and BamHI restriction sites at the first multiple-cloning site. This introduces an N-terminal His6 tag into the expressed product. Both the Ssp1–Rap1a and Ssp2–Rap2a complexes were produced in E. coli Rosetta-gami (DE3) and were purified by immobilized metal-ion affinity chromatography (English et al., 2012) and size-exclusion gel-filtration chromatography with a Superdex 75 10/300 GL column (GE Healthcare). The samples were concentrated to 13.5 mg ml\(^{-1}\) by centrifugation (10 000 molecular-weight cutoff, Amicon) and dialyzed into 10 mM sodium phosphate pH 6.4. A high level of purity of greater than 95% was confirmed by SDS–PAGE. Size-exclusion chromatography was also used to investigate the association of cognate Ssp-C50A mutant–Rap combinations, with the proteins being mixed in equimolar amounts prior to separation, as described by English et al. (2012).

2.2. Peptidoglycan-cleavage assay

Purified peptidoglycan sacculi (300 μg) from E. coli D456, consisting mainly of tetrapeptides and pentapeptides (Chou et al., 2012), were incubated with either Ssp1 or Ssp2 (100 μg ml\(^{-1}\)) or, as a control, no enzyme in 300 μl 20 mM sodium phosphate pH 4.8 for 4 h at 310 K. The samples were incubated with 40 μl 100 mg ml\(^{-1}\) of the muramidase Cellosyl (kindly provided by Höchst AG, Frankfurt, Germany) for 16 h at 310 K to convert the residual peptidoglycan and solubilized fragments into muropeptides. The sample was boiled for 10 min and insoluble material was removed by centrifugation. The muropeptides were reduced with sodium borohydride and analyzed by high-pressure liquid chromatography using established methods (Glauner, 1988; Chou et al., 2012; Russell et al., 2012). The profile/retention times were compared with previous samples for which mass-spectrometric analyses with fragmentation had been carried out and which had identified that the linear disaccharide hexapeptide elutes before the disaccharide dipeptide, whereas the branched disaccharide hexapeptide elutes after the disaccharide dipeptide.

2.3. Phenotypic assays to characterize the C50A mutants of Ssp1 and Ssp2

These assays were performed using the bacterial strains, plasmids and protocols described previously (English et al., 2012). The QuikChange II Site-Directed Mutagenesis Kit (Stratagene) was used to generate C50A (Cys50 to Ala) mutants of Ssp1 and Ssp2 in the existing plasmids pSC152 and pSC138 (for periplasmic expression in E. coli) and pSC539 and pSC541 (for complementation of the cognate mutation in S. marcescens). Microscopic analysis, antibacterial co-culture (competition) assays and immunodetection of secreted Ssp2 were performed as described in the previous study. In brief, the format of the antibacterial co-culture assays was that the attacker and target strains were mixed in a 1:1 ratio, co-cultured on solid Luria–Bertani (LB) media for 7.5 h at 303 K and the surviving target cells (in this case a streptomycin-resistant version of the Δrap2a, ΔclpV mutant) were enumerated by serial dilution and viable counts on streptomycin-containing media. ClpV is an ATPase that is essential for the type VI secretion system to function and so deletion provides an appropriate control. Statistical significance testing was performed using ANOVA followed by Dunnett’s post-test (GraphPad Prism software). For the detection of Ssp1 and Ssp2 levels in solid-grown E. coli or S. marcescens, cells were removed from the surface of agar plates, resuspended in liquid media, boiled and the total cell extract was separated by SDSPAGE followed by anti-Ssp1 or anti-Ssp2 immunoblotting as described previously (English et al., 2012).

2.4. Bioinformatic analyses

Multiple Ssp/Tae4 homologues were identified from public databases as part of previous studies (English et al., 2012; Russell et al., 2012). Multiple sequence alignments were generated using MUSCLE (Edgar, 2004), and Jalview (Waterhouse et al., 2009) was used to visualize the alignment and to calculate the resulting tree (using neighbour-joining construction and the BLOSUM62 distance matrix). Genomic analyses using publicly available databases (http://www.ncbi.nlm.nih.gov/ and http://www.sanger.ac.uk/resources) allowed the identification of adjacent encoded candidate immunity proteins, which were then used as bait to interrogate the S. marcescens Db11 genome and determine the Rap protein to which each was most closely related.

2.5. Crystallographic analyses

2.5.1. Crystal growth and data collection. For crystallization trials, Rap1a was dialyzed against 25 mM Tris–HCl,
Table 1
Crystallographic statistics.
Values in parentheses are for the highest resolution shell.

<table>
<thead>
<tr>
<th>Structure/PDB code</th>
<th>Ssp1/4bi3</th>
<th>Ssp1-C50A/4bi4</th>
<th>Rap1a/3zf2</th>
<th>Rap2a/3zib</th>
<th>Ssp1–Rap1a/4bi8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P2₁2₁2₁</td>
<td>P2₁2₁2₁</td>
<td>C222</td>
<td>P2₁</td>
<td>P2₁2₁2₁</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.5418</td>
<td>1.5418</td>
<td>0.97950</td>
<td>0.96110</td>
<td>1.5418</td>
</tr>
<tr>
<td>Unit-cell parameters a (Å)</td>
<td>56.83</td>
<td>56.79</td>
<td>82.65</td>
<td>39.65</td>
<td>68.47</td>
</tr>
<tr>
<td></td>
<td>65.26</td>
<td>64.59</td>
<td>93.00</td>
<td>81.35</td>
<td>68.47</td>
</tr>
<tr>
<td>b (Å)</td>
<td>97.50</td>
<td>97.87</td>
<td>51.26</td>
<td>58.45</td>
<td>92.52</td>
</tr>
<tr>
<td>c (Å)</td>
<td>R (°)</td>
<td>91.54</td>
<td>90.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>48.75–1.85</td>
<td>48.93–2.21</td>
<td>46.50–1.98</td>
<td>40.67–1.90</td>
<td>49.92–2.00</td>
</tr>
<tr>
<td>No. of reflections</td>
<td>503341 (41592)</td>
<td>139598 (9390)</td>
<td>63106 (8849)</td>
<td>214423 (31351)</td>
<td>493039 (32290)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.4 (100.0)</td>
<td>99.9 (98.5)</td>
<td>99.9 (100.0)</td>
<td>99.4 (98.9)</td>
<td>99.9 (99.1)</td>
</tr>
<tr>
<td>No. of reflections</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>completeness (%)</td>
<td>99.4 (100.0)</td>
<td>99.9 (98.5)</td>
<td>99.9 (100.0)</td>
<td>99.4 (98.9)</td>
<td>99.9 (99.1)</td>
</tr>
<tr>
<td>Rmerge (%)</td>
<td>7.4 (23.9)</td>
<td>7.5 (26.3)</td>
<td>6.6 (48.6)</td>
<td>9.7 (51.6)</td>
<td>10.9 (58.7)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>16.0 (9.5)</td>
<td>7.5 (7.1)</td>
<td>4.5 (4.4)</td>
<td>7.4 (7.5)</td>
<td>28.2 (25.2)</td>
</tr>
<tr>
<td>Wilson B (Å²)</td>
<td>10.7</td>
<td>18.7</td>
<td>31.5</td>
<td>21.7</td>
<td>18.8</td>
</tr>
<tr>
<td>Rmerge²/Rmerge (%)</td>
<td>20.4/24.0</td>
<td>19.2/24.6</td>
<td>19.2/23.4</td>
<td>18.3/23.4</td>
<td>17.9/22.7</td>
</tr>
<tr>
<td>No. of residues</td>
<td>326</td>
<td>326</td>
<td>185</td>
<td>377</td>
<td>256</td>
</tr>
<tr>
<td>No. of waters</td>
<td>300</td>
<td>372</td>
<td>57</td>
<td>137</td>
<td>192</td>
</tr>
<tr>
<td>No. of ligands</td>
<td>3 K⁺, 5 SO₄⁻, 1 glycerol</td>
<td>5.13 – 0.21</td>
<td>0.15 – 0.15</td>
<td>0.15 – 0.16</td>
<td></td>
</tr>
<tr>
<td>R.m.s.d. bond lengths† (Å)</td>
<td>0.006</td>
<td>0.008</td>
<td>0.018</td>
<td>0.019</td>
<td>0.006</td>
</tr>
<tr>
<td>R.m.s.d. angles †/ ‡ (°)</td>
<td>0.967</td>
<td>1.301</td>
<td>1.925</td>
<td>1.745</td>
<td>1.020</td>
</tr>
<tr>
<td>Average B factors (Å²)</td>
<td>Chain A</td>
<td>1.27</td>
<td>17.7</td>
<td>34.3</td>
<td>19.7</td>
</tr>
<tr>
<td></td>
<td>Chain B</td>
<td>11.2</td>
<td>17.9</td>
<td>35.2</td>
<td>23.0</td>
</tr>
<tr>
<td></td>
<td>Chain C</td>
<td>22.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chain D</td>
<td>18.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waters</td>
<td>15.2</td>
<td>25.1</td>
<td>34.8</td>
<td>30.6</td>
<td>27.0</td>
</tr>
<tr>
<td>K⁺</td>
<td>22.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SO₄⁻</td>
<td>12.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glycerol</td>
<td>35.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ramachandran plot analysis</td>
<td>Favored regions</td>
<td>97.1</td>
<td>97.0</td>
<td>97.0</td>
<td>99.0</td>
</tr>
<tr>
<td></td>
<td>Allowed regions</td>
<td>2.9</td>
<td>3.0</td>
<td>3.0</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>Outliers</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

† Rmerge = Σ(|I(hkl)| − |Fobs(hkl)|)/Σ|I(hkl)|, where I(hkl) is the intensity of the hth measurement of reflection hkl and |Fobs(hkl)| is the mean value of |I(hkl)| for all i measurements. † Rmerge = Σ(|Fobs(i) − |Fcalc(i)|)/Σ|Fobs(i)|, where Fcalc is the observed structure factor and Fobs is the calculated structure factor. ‡ Rfree is the same as Rmerge except calculated with a subset (5%) of data that were excluded from the refinement calculations. † Diffraction-component precision index (Crucickshank, 1998). †† Engh & Huber (1991).

150 mM sodium chloride pH 7.5 and all other samples were in 100 mM sodium phosphate pH 6.4. The sitting-drop vapour-diffusion method was used with 0.2 µl drops with a 1:1 ratio of protein solution to reservoir solution at 293 K. Several commercially available screens were used in 96-well plates with a Phoenix Liquid Handling System (Rigaku, Art Robbins Instruments) to scout out initial conditions, which were then optimized.

Crystals of Ssp1 were obtained by combining protein solution at a concentration of 10 mg ml⁻¹ with reservoir solution consisting of 0.2 M potassium sulfate, 20% PEG 3350. Orthorhombic block crystals grew to a maximum dimension of approximately 350 µm over 5 d. The Ssp1-C50A mutant (10 mg ml⁻¹) gave isomorphous crystals (maximum dimension of 250 mm) in 2 d using reservoir solution consisting of 0.1 M sodium citrate pH 5.5, 20% PEG 3000. The Ssp1–Rap1a complex at 13.5 mg ml⁻¹ formed clusters of plate-like crystals using a reservoir solution consisting of 12.5% PEG 1000, 12.5% PEG 3350, 12.5% MPD. These crystals attained a maximum size of 200 µm within 3 d. A single-crystal fragment was removed from the cluster for diffraction measurements. Monoclinic blocks of Rap2a were grown by combining a protein concentration of 13.5 mg ml⁻¹ with a reservoir solution consisting of 25% PEG 1000, 0.1 M MES pH 6.5. These crystals attained a maximum dimension of 200 µm within 5 d. A slender orthorhombic crystal of Rap1a with approximate dimensions of 150 × 35 × 35 µm was observed after about one month using a reservoir solution consisting of 25% PEG 3350, 100 mB bis-tris pH 5.5. We were unable to obtain crystals of the Ssp2–Rap2a complex.

All crystals were soaked briefly in mother liquor adjusted to contain 20% (v/v) glycerol as a cryoprotectant prior to flash-freezing in liquid nitrogen and use in diffraction experiments. Data for Ssp1, the Ssp1-C50A mutant and the Ssp1–Rap1a complex were measured at 100 K using a Rigaku MicroMax-007 rotating-anode X-ray generator (Cu Kα) coupled to a Saturn 944 CCD detector and were processed using XDS (Kabsch, 2010) or HKL-3000 (Minor et al., 2006). The Rap1a and Rap2a data were collected on beamlines I04 and I03 of the Diamond Light Source, respectively, and were indexed and integrated in iMosflm (Battye et al., 2011). Data sets were analyzed and scaled with POINTLESS and SCALA (Evans, 2006) from the CCP4 program suite (Winn et al., 2011).

The isomorphous Ssp1 and Ssp1-C50A crystals contained two polyopeptides in the asymmetric unit with an estimated solvent content of 50% and a V_M of 2.46 Å³ Da⁻¹. The Ssp1–Rap1a complex crystal presented a heterodimer in the asymmetric unit with an estimated solvent content of 40% and a V_M of 2.06 Å³ Da⁻¹. Rap2a crystallized with four molecules in the asymmetric unit, an estimated solvent content of 40% and a V_M of 2.06 Å³ Da⁻¹, whilst Rap1a displayed two molecules in the asymmetric unit with an estimated solvent content of 45% and a V_M of 2.06 Å³ Da⁻¹.

2.5.2. Structure determination and refinement. The Ssp1 structure was solved targeting the single-wavelength anomalous scattering properties of sulfur (Micossi et al., 2002). Auto-Rickshaw, the EMBL Hamburg automated structure-determination platform (Panjikar et al., 2005), was used. The heavy-atom structure-factor contributions were estimated in SHELXC (Sheldrick, 2010; Sheldrick et al., 2001) and the
maximum resolution for substructure determination and initial phase calculation was set to 2.30 Å. All 14 heavy atoms (sulfurs) were found using SHELXD (Schneider & Sheldrick, 2002). The correct hand for the substructure was determined using ABS (Hao, 2004) and SHELXE (Sheldrick, 2002) and the initial phases produced a CC (correlation coefficient) of 0.32. The initial phases were improved by density modification (Terwilliger, 2003) prior to phase extension to 1.85 Å resolution. The CC improved to 0.63, resulting in an electron-density map with excellent quality. Almost 80% of the model was constructed in ARP/wARP (Langer et al., 2008). The initial model consisted of two polypeptides of 295 residues in total, with $R_{\text{work}}$ and $R_{\text{free}}$ values of 29.7 and 31.4%, respectively. Subsequent model building extended this to 326 residues, with the $R_{\text{work}}$ and $R_{\text{free}}$ values improving to 20.7 and 24.1%, respectively.

A molecule of Ssp1 was used to solve the structures of the Ssp1–Rap1a complex and the Ssp1-C50A mutant by molecular replacement (autoMR; Winn et al., 2011). The structures of Rap1a and Rap2a were solved by molecular replacement (Phaser; McCoy et al., 2007) using the structure of Rap1a from the Ssp1–Rap1a complex and Rap2b (PDB entry 4b6i; English et al., 2012), respectively. In the latter case the sequence identity shared by the search model and the target structure is only about 20%.

All structures were refined in an iterative process combining REFMAC5 (Murshudov et al., 2011) with electron-density and difference-density map inspections and model manipulations in Coot (Emsley et al., 2010). For those structures with multiple copies in the asymmetric unit, tight NCS (noncrystallographic symmetry) restraints were imposed which were gradually released during refinement. Translation/libration/screw analysis (TLS) refinements were applied with the appropriate groups determined using the TLSMD server (Painter & Merritt, 2006). Water molecules, and in the case of the Ssp1 structure also potassium and sulfate ions, were added.

---

**Figure 1**

The specificity of Ssp1 and Ssp2. (a) Peptidoglycan from *E. coli* strain D456 was incubated with Ssp1, Ssp2 or no enzyme (control) followed by digestion with the muramidase Cellosyl and analysis of the resulting muropeptides by HPLC. Both Ssp1 and Ssp2 cleaved non-cross-linked (tri, tetra and penta) and cross-linked (tetradeca and tetrapenta) muropeptides between d-glutamate and meso-diaminopimelic acid ($m$-A$_2$pm), resulting in the disaccharide dipeptide (di) product. Ssp2 also produced a small amount of the disaccharide hexapeptide (hexa) product. (b) Proposed structures of the muropeptides separated in (a). The arrows indicate the cleavage sites of d-Glu-$m$-A$_2$pm endopeptidases in peptidoglycan. GlcNAc, N-acetylglucosamine; MurNAc(r), N-acetylmuramitol; $m$-A$_2$pm, meso-diaminopimelic acid.
during the refinement process. Where appropriate, dual rotamer side-chain conformations were also included. Refinements were terminated when there were no significant changes in the $R_{\text{work}}$ and $R_{\text{free}}$ values and inspection of the difference density maps suggested that no further corrections or additions were justified.

*MolProbity* (Chen et al., 2010) was used to investigate the model geometry. Secondary-structure and surface-interaction analyses were performed using *DSSP* (Kabsch & Sander, 1983) and *PISA* (Krissinel & Henrick, 2007), respectively. Figures were prepared using *ALINE* (Bond & Schütteklkopf, 2009) and *PyMOL* (Schrödinger). The *DALI* server was used to search the PDB for structural homologues, whilst superpositions were calculated using *DaliLite* (Holm & Park, 2000).

**Figure 2**
The Ssp1 and Ssp2 C50A mutants are inactive. (a) Growth of *E. coli* MG1655 transformed with plasmids expressing OmpA$_{sp}$-Ssp1 (sp-Ssp1; pSC152), OmpA$_{sp}$-Ssp1(C50A) [sp-Ssp1(C50A); pSC548], OmpA$_{sp}$-Ssp2 (sp-Ssp2; pSC138) or OmpA$_{sp}$-Ssp2(C50A) [sp-Ssp2(C50A); pSC549] from an arabinose-inducible promoter, or with the empty vector (VC, pBAD18-Kn), on LB or M9 medium containing 0.2% arabinose. (b) Recovery of a sensitive Δrap2a mutant as the target strain following co-culture with the different attacking strains indicated, expressed relative to recovery of target when co-cultured with the wild-type strain. Attacking strains are wild-type *S. marcescens* Db10 (WT), mutant lacking ClpV (ΔclpV), mutant lacking Ssp2 (Δssp2), each carrying either the vector control plasmid (+VC; pSUPROM) or a plasmid expressing wild-type Ssp2 (+Ssp2; pSC541) or the C50A mutant of Ssp2 (+Ssp2(C50A); pSC1230). (c) Phenotypes of wild-type *S. marcescens* Db10 and selected single and double mutants carrying plasmids expressing variants of Ssp1 or Ssp2 following growth on solid medium. For each strain, representative images of the morphology of a culture spot (left; scale bar 1 mm), single colonies (middle; scale bar 1 mm) and individual cells [right; scale bar 2 μm in (i) or 5 μm in (ii)] are shown. Plasmids direct the expression of wild-type Ssp1 (+Ssp1; pSC539), the C50A mutant of Ssp1 [+Ssp1(C50A); pSC1229], wild-type Ssp2 (+Ssp2; pSC541), the C50A mutant of Ssp2 [+Ssp2(C50A); pSC1230] or represent the vector control (+VC, pSUPROM). Growth was for 48 h on MM (i) or 24 h on LB (ii). (d) Immunoblot analysis of levels of Ssp1 or Ssp2 proteins from *E. coli* MG1655 (top two panels) or *S. marcescens* (bottom two panels) grown on solid medium. Strains and plasmids are as in (a)–(c), except that wild-type OmpA$_{sp}$-Ssp1 and OmpA$_{sp}$-Ssp2 were co-expressed with their cognate immunity proteins to allow the strains to grow [sp-Ssp1(WT)Rap1a; pSC160 or sp-Ssp2(WT)Rap2a; pSC144].
Relevant crystallographic statistics and geometric details of the refined models were extracted from the programs used in the analyses and are reported in Table 1.

3. Results and discussion

3.1. Activity of Ssp1 and Ssp2 and neutralization by cognate Rap partners

We previously identified *S. marcescens* Ssp1 and Ssp2 as periplasmic acting antibacterial T6SS effectors (English et al., 2012) and noted their membership, on the basis of sequence similarity, of the family 4 amidases (Tae4 proteins) proposed by Russell et al. (2012). We aimed to experimentally determine the enzymatic activity of each protein. The incubation of peptidoglycan sacculi with purified Ssp1 or Ssp2, followed by muramidase digestion and high-pressure liquid-chromatography analysis, revealed the disaccharide dipeptide as the single major product. This indicates that both enzymes cleave the amide bond in peptidoglycan between isoglutamic acid and *meso*-diaminopimelic acid. Ssp1 quantitatively hydrolyzed monomeric tripeptides, tetrapeptides and pentapeptides, as well as dimeric tetratetrapeptides and tetrapentapeptides, on both the acceptor and the donor side (Fig. 1). In the cross-linked peptides, the ε-amino group of the *meso*-diaminopimelic acid residue at position 3 on the acceptor side is connected by an amide bond to the α-carboxylic group of d-alanine at position 4 on the donor side. Although Ssp2 was active against these muropeptides, the cleavage of monomeric pentapeptide and dimeric tetratetrapeptide was less efficient and the partially cleaved disaccharide hexapeptide product was detected (Fig. 1a). This behaviour is similar to the family 4 amidase Tae4 from *Salmonella* Typhimurium (STTae4, also known as Tae4<sup>TM</sup>; Russell et al., 2012). Hence, Ssp2 and STTae4 preferentially cleave tetrapeptides in the acceptor part of cross-linked peptides. This is distinct from *P. aeruginosa* Tse1, which preferentially cleaves pentapeptides in the donor part of cross-linked peptides (Chou et al., 2012; Russell et al., 2012). These data indicate that Ssp1 and Ssp2 are distinctive peptidoglycan *dl*-endopeptidases in terms of specificity, with the former being more promiscuous regarding the chemical structure that it is able to recognize and then cleave.

Established functional assays were utilized to examine the effect of mutating the catalytic cysteine (Cys50) of Ssp1 and Ssp2. Heterologous expression targeting Ssp1 and Ssp2 to the periplasm of *E. coli* MG1655 prevents growth on M9 minimal medium and, in the case of Ssp2, also on LB medium. As shown in Fig. 2(a), Ssp1-C50A and Ssp2-C50A were no longer toxic to *E. coli*. In *S. marcescens* itself, a mutant lacking the immunity protein, /C1rap2a, is susceptible to self-killing by Ssp2 injected by the wild-type strain. This is shown in two ways. Firstly, co-culture of a wild-type attacker strain with a /C1rap2a target strain results in the death of the latter. When the attacker lacks a functional T6SS (/C1clpV) or Ssp2 (/C1ssp2 mutant), increased recovery of the target strain is observed. Complementation of the /C1ssp2 mutant attacker by expression of wild-type Ssp2 from a plasmid restores the killing activity, with the recovery of the target reduced below the levels observed with the wild-type attacker. However, the expression of Ssp2-C50A was unable to complement the /C1ssp2 mutant and restore the killing.
activity (Fig. 2b). Secondly, a single Δrap2a mutant exhibits fitness and morphological defects when grown on solid medium and these defects are alleviated when ssp2 is also deleted in a Δrap2αΔssp2 mutant. Plasmid-mediated expression of Ssp2 in the Δrap2αΔssp2 mutant re-induced the fitness and morphological defects caused by self-killing; however, expression of the Ssp2-C50A mutant had no negative effect (Fig. 2c). Similarly, expression of Ssp1, but not Ssp1-C50A, in a Δrap1αΔssp1 mutant induced mild morphological defects (Fig. 2c). In each case, an equivalent level of stable expression of Ssp1-C50A and Ssp2-C50A compared with wild-type Ssp1 and Ssp2 was confirmed by immunoblotting (Fig. 2d).

3.2. Catalytic inactivation of Ssp1 and Ssp2 does not affect their secretion or immunity protein binding

It was necessary to confirm that the above observations were not affected by any influence of the catalytic C50A mutation on immunity protein binding or secretion of the effectors. Each of the Ssp C50A mutants was mixed with an equimolar amount of the cognate Rap protein, followed by size-exclusion chromatography analysis to monitor complex formation. In each case, a single species of apparent molecular weight ~55 kDa was observed (Supplementary Fig. S1a). This behaviour exactly matches that of the wild-type proteins (English et al., 2012) and indicates that complexes consisting of two Ssp and two Rap polypeptides are formed. It was also demonstrated that the Ssp2-C50A mutant protein was secreted by the T6SS of S. marcescens as efficiently as the wild-type protein (Supplementary Fig. S1b). We conclude that mutation of the catalytic cysteine in Ssp1 and Ssp2 abolishes their bacterioclytic effect because the Δdl-endopeptidase activity is compromised. Furthermore, we can conclude that the Ssp–Rap interaction does not depend on the catalytic cysteine and that recognition of the Ssp effectors by the T6SS machinery is also independent of their enzymatic activity.

3.3. Structure of Ssp1

To gain a detailed insight into the structure and activity of the peptidoglycan endopeptidase effector Ssp1, we applied crystallographic methods. Isomorphous crystal structures of Ssp1 (at 1.85 Å resolution) and the active-site Ssp1-C50A mutant (at 2.2 Å resolution) reveal the overall enzyme structure and details of the active site. The protein is a monomer in solution but crystallized with two molecules in the asymmetric unit, each consisting of residues 1–163. The root-mean-square deviations (r.m.s.d.s) between least-squares superimposed Cα atoms gives an average of 0.3 Å when comparing the four polypeptides in the crystal structures. The copies in the asymmetric unit are therefore judged to be essentially identical and neither mutagenesis of the catalytic Cys50 nor crystal lattice packing interactions induces any major structural changes.

The primary, secondary and tertiary structure of Ssp1 (Fig. 3) classifies it into the NlpC/P60 cysteine peptidase superfamily (Anantharaman & Aravind, 2003). The overall dimensions of the bilobal structure are about 35 × 40 × 50 Å. An N-terminal subdomain comprises residues 1–95 and is formed by six α-helices (α1–α6) and, on the surface of the protein, a β-hairpin-like loop (β1, turn, β2). A short peptide segment links α6 to β3, which represents the start of the C-terminal subdomain. This subdomain is dominated by a four-stranded antiparallel β-sheet with order β3–β7–β4–β5, which is flanked on one side by α7 and β6 and on the other by α6. A disulfide bond is observed between Cys146 at the C-terminal end of β6 and Cys150 located on the loop linking β6 to β7. A cleft, the substrate-binding site, is created between the subdomains by residues in the α1–α2 and α4–α5 links, the C-terminal segment of β6 and the loop leading to β7. The disulfide linkage appears important to create one side of the active site and perhaps also contributes to the stability of the C-terminal subdomain fold.

The family 4 amidase effectors hydrolyze the amide bond between D-Glu and mAmDpm. They are similar to CHAP (cysteine, histidine-dependent amidohydrolases/peptidases) family members, a subset of the NlpC/P60 superfamily with a strictly conserved cysteine–histidine catalytic dyad (Fyfe et al., 2008). In the case of Ssp1, Cys50 is located in the N-terminal subdomain and His133 in the C-terminal subdomain (Fig. 3b). Cys50 donates a hydrogen bond to a water molecule, which in turn interacts with His133, Asn48 and Tyr129 (Fig. 4). Asp135 accepts a hydrogen bond from His133 and appears to complete a catalytic triad reminiscent of that observed in typical cysteine peptidases. The mechanism of such cysteine-dependent proteases, especially papain, is well documented (Alphey & Hunter, 2006 and references therein). The Cys–His bonds are shown as dashed lines and selected residues are presented as sticks with C-atom positions coloured orange, N blue, O red and S yellow.

Figure 4
The Ssp1 catalytic centre. The protein is shown as a semi-transparent van der Waals surface coloured according to atom type (N, blue; O, red; S, yellow; C, white). A water (blue sphere) participates in four hydrogen-bonding interactions with Asn48, Cys50, Tyr129 and His133. Hydrogen bonds are shown as dashed lines and selected residues are presented as sticks with C-atom positions coloured orange, N blue, O red and S yellow.
catalytic dyad forms a thiolate–imidazolium pair that is oriented by a hydrogen bond between the histidine and an acidic residue. In this case the residues are Cys50, His133 and Asp135. The generation of a thiolate Cys50 following proton abstraction by His133 would support nucleophilic attack at the γ-d-glutamyl-MAp amide bond to form an acyl thioester. Hydrolysis of the thioester, exploiting an activated water molecule, then releases the products. The main-chain amides of the catalytic cysteine and Thr49 create an oxyanion hole that may support thiolate attack on the amide linkage by stabilizing the tetrahedral intermediate formed prior to the formation of an acyl-enzyme complex. In this there are striking parallels to the reaction catalyzed by nicotinamidase (Fyfe et al., 2009).

Ssp1 shares about 20% sequence identity with Ssp2, but we do not have a structure of Ssp2 for comparative purposes to address the observation of distinct substrate specificities. However, the closest structural relative of Ssp1 is the bacterio-
ytic effector Tae4 from E. cloacae and S. typhimurium (PDB entries 4hfk and 4hff, respectively; Zhang et al., 2013), with Z-scores of 17 and 16 and r.m.s.d.s of 2.2 and 2.1 Å for the least-squares fit of 138 and 137 Cα positions, respectively. Ssp1 shares about 15% sequence identity with these proteins but, despite such a low sequence similarity, the r.m.s.d. values and overlay confirm that these structures share the same fold (Supplementary Fig. S2). On the basis that Ssp2 shares about 50% sequence identity with Tae4 and the same substrate specificity, we suggest that it provides a suitable surrogate structure for comparative purposes.

An overlay of the Ssp1 and Tae4 structures confirms similarities in the protein fold and the relative positions of residues linked to catalytic activity. However, the overlay also indicates noteworthy differences (Supplementary Figs. S2 and S3). The α1–α2 loop in Ssp1, consisting of about 15 amino acids, lines one side of the active site then extends away from the catalytic core. Adjacent to this loop is a turn carrying the Cys146 and Cys150 disulfide. The turn and the polypeptide that follows, which lead directly into β7, line the active site. In EcTae4, the disulfide is conserved (Cys137 and Cys141), as is the structure just after the turn. An eight-residue segment extends from the turn and curls over to narrow the active-site cleft. In STTae4 the conserved residues Cys135 and Cys139 are in a reduced conformation, for reasons that are not made clear, and the polypeptide chain at the periphery of the active site is flexible and disordered, as shown by inflated thermal parameters and a lack of electron density. Alignment of the Ssp1 and Ssp2 amino-acid sequences (Fig. 3a) indicates that there is a truncation of four residues in the region of the disulfide-containing loop and this may reduce the size of the active-site cleft of Ssp2 compared with Ssp1. The alignment also suggests that Asp135 of Ssp1, assigned as a component of the catalytic triad, is not conserved and corresponds to Thr133 in Ssp2. In STTae4 and EcTAE4 a threonine is also present at this position (Supplementary Fig. S4). A superposition of Ssp1 and EcTAE4 indicates a difference of Ser148 replaced by Asp139, with the carboxylic acid side chain positioned to fill the role of Asp135 of Ssp1 in the catalytic triad (Supplementary Fig. S3). The aspartate is conserved in Ssp2, STTae4 and EcTAE4 (Supplementary Fig. S4).

The peptidoglycan hydrolase assay data indicate that Ssp2 preferentially targets the acceptor component of the peptidoglycan tetratetra cross-link rather than donor part and in this it is similar to Tae4 (Russell et al., 2012). Ssp1 cleaves the acceptor and donor stem of cross-linked and non-cross-linked peptidoglycan. The differences evident from structural comparisons suggest that a distinctive and more open substrate-binding site in Ssp1 compared with those observed in Tae4 structures or implied in Ssp2 might explain the promiscuous endopeptidase activity of Ssp1.
3.4. Structure of Rap1a

Next, we sought to reveal molecular details of the immunity proteins Rap1a and Rap2a, which neutralize Ssp1 and Ssp2 toxicity, respectively (English et al., 2012). The structure of Rap1a was determined to about 2.0 Å resolution (Table 1) and the amino-acid sequence with assigned secondary structure is depicted in Fig. 5(a). A search for structural orthologues failed to identify anything of relevance in the PDB. We therefore conclude that the Rap1a fold has not been observed previously, making it a unique member of the T6SS immunity family of proteins. Several other candidate immunity proteins identified in other organisms share significant sequence identity with Rap1a (see below). Therefore, we propose that this group of proteins be referred to as the ‘Tai4a’ immunity proteins, to reflect the fact that whilst they are immunity proteins to Tae4 effectors, they are distinct from the main family of Tai4 proteins, which includes Rap1b, Rap2a and Rap2b.

The Rap1a subunit displays a compact globular structure constructed from five α-helices that assemble to form the highly stable symmetric dimer that constitutes the asymmetric unit (Fig. 5). This is consistent with the size-exclusion chromatography data, which identified that only a dimer was observed in solution (Supplementary Fig. S1). The NCS is highly conserved, with an r.m.s.d. of 0.6 Å for a least-squares overlay of 96 Cα positions. A disulfide bond is formed between Cys78 (in the α2–α3 loop) and Cys122 in the C-terminal region. This interaction appears to be crucial to stabilizing the subunit fold since it helps position α2, α3 and α5 close to each other and these segments of secondary structure provide the side chains that form the hydrophobic core of the subunit. During initial recombinant expression tests it was noted that soluble protein was only produced in E. coli Rosetta-gami (DE3) cells, a strain that promotes the formation of disulfide bonds in the cytoplasm and so mimics what might occur in the oxidative environment of the periplasm. This suggests that the covalent bond is necessary for correct folding to occur and for stability of the Rap1a fold and dimeric quaternary structure.

Interactions involving residues in α2 make the major contribution to dimer formation. Each subunit contributes a surface area of 1130 Å² to the dimer interface, which is 20% of the solvent-accessible surface area (ASA) of the subunit (5700 Å²). Such a percentage of surface area is indicative of a stable association (Krissinel & Henrick, 2007). Eight residues from each subunit form a network of hydrogen bonds using both main-chain and side-chain functional groups (Ser36, Asn40, Leu41, Glu59, Tyr64, Asp70, Lys75, Arg107 and Glu113, Fig. 5c). There are also indirect hydrogen-bonding interactions via well ordered water molecules that link a number of side chains and main chains. Hydrophobic interactions that contribute to the stability of the dimer mainly involve the aliphatic side chains of Val39, Ile63, Leu66, Val68 and Ala71, but also Tyr80.

3.5. Structure of Rap2a

Rap2a, like Rap1a, is predicted to be localized in the periplasm and could only be produced in soluble recombinant form using the E. coli Rosetta-gami (DE3) strain. It is also a stable dimer in solution as shown by size-exclusion chromatography (English et al., 2012). The structure was determined at 1.9 Å resolution with four molecules, arranged as two dimers, in the asymmetric unit. These four molecules, subunits A–B and C–D, are similar...
overall, with r.m.s.d.s between superimposed Cα atoms ranging from 0.6 Å (subunits A and B, residues Thr27–Gln123) to 0.7 Å (subunits C, residues 27–127, and D, residues 27–126). Minor deviations from NCS are observed in a five-residue loop between α2 and α3 (Gly67–Leu71) and indicate some conformational freedom in this part of the molecule. The ASA of a Rap2a subunit averages at approximately 6110 Å²; the range is from 6140 Å² for subunit C to 6090 Å² for subunit D. Dimer formation occludes an area that is approximately 20% of the ASA, which is indicative of a stable association (Krisinel & Henrick, 2007).

The Rap2a subunit displays a compact globular structure of five α-helices with an extended loop linking α3 to α4 (Figs. 6a and 6b). A disulfide bond between Cys42 and Cys102 links α1 to α4 and interactions of other side-chain groups on these elements of secondary structure help to create the helical bundle fold and in particular to align α2. Together, α1, α2, the α2–α3 loop and α4 form the dimer interface, giving rise to a twofold NCS axis, and a combination of hydrogen-bonding, salt-bridge and van der Waals interactions serve to stabilize the association. Main-chain hydrogen-bonding contributions come from the amides of Ser48, Ala49, Met97, Thr98 and Met99 on both chains. Side-chain contributions come from the hydroxyl groups of Tyr28 and Tyr47 and the carboxylates of Glu51, Asp55 and Asp104. Several solvent-mediated contacts also serve to link functional groups on partner subunits (not shown). The side chains of Leu39, Ile43, Tyr47, Val52, Met99 and Ile103 are involved in van der Waals interactions with the partner subunit to stabilize the dimer.

Despite a low level of sequence conservation, the structural similarity of five T6SS immunity proteins (Rap1b, Rap2a, Rap2b, EcTai4 and STTai4) indicates an orthologous subset. This structurally defined set is consistent with the designation of a ‘Tai4’ family of immunity proteins cognate to the Tae4 effectors (Russell et al., 2012). Pairwise comparisons indicate a range of sequence identities from 15 to 35%, and matching between 93 and 124 Cα positions gives an r.m.s.d. range of 1.2–1.7 Å indicative of close structural similarity. The structural overlay is exemplified by superposition of Rap2a, Rap2b and EcTai4 (Fig. 6c). Of note is the conservation of an extended or protruding loop structure which is highly variable in terms of amino-acid sequence (see Fig. 7 in English et al., 2012). Parts of the EcTai4 subunit involved in interaction with an effector, identified by the positions of Val41 and Thr75 in Rap2a, are also well conserved in terms of overall structure (Fig. 6c).

This subset of immunity proteins also display similar dimer structures and the Cys42–Cys102 disulfide bond in Rap2a is conserved in Rap1b and Rab2b. The covalent interaction appears to be important for the creation of the subunit fold

---

**Figure 7**

The Ssp1–Rap1a heterotetramer complex. (a) Ssp1 is shown as an orange van der Waals surface and the Rap1a subunits are shown as cyan and grey ribbons. The disulfide bonds in Rap1a are presented as yellow sticks. Blue, red and purple circles mark the three distinct areas (I–III) of interaction between the effector and immunity proteins. (b) Schematic diagram showing the hydrogen-bonding and salt-bridge interaction network involved in complex formation. Black lines mark interacting residues. Residue positions in the protein structures are given by labelling the appropriate secondary structure. The exceptions are Glu84 in Rap1a, which occurs just prior to α3 in this structure, and Asn48 and Asp130 in Ssp1, which are between α4 and α5 and at the C-terminal segment of β4, respectively. (c) Glu84 directly interacts with the catalytic His133 and blocks the active site. Grey, apo Ssp1; dark orange, Ssp1 from the Ssp1–Rap1a complex; cyan, Rap1a from the Ssp1–Rap1a complex. Broken arrows indicate the adjustments of Asn48 and Tyr129 when comparing free and complexed Ssp1.
and the stable quaternary structure. In the structures of EcTai4 and STTai4 there are conserved cysteines that match the disulfide-forming residues in the Rap proteins; however, they are in a reduced form. In STTai4, for example, the distance between the SG atoms of Cys48 and Cys108 is 3.6 Å and the electron density unambiguously defines reduced cysteine residues. This difference in the redox states may simply reflect distinct experimental conditions.

3.6. Inhibition of Ssp1 by Rap1a

We next sought to delineate the molecular basis of how the S. marcescens Rap proteins neutralize their cognate Ssp amidase/endopeptidase effectors and what features engendered the exquisite specificity noted in the effector–resistance protein combinations. We previously proposed that the complexes exist as Sp2–Rap2 heterotetramers based on biochemical analyses (English et al., 2012), a conclusion that was subsequently confirmed by the structures of related Tae4–Tai4 complexes (Zhang et al., 2013). The similarities of Sp2 and Rap2a to Tae4 and Tai4 in terms of peptidoglycan hydrolase activity, sequence and structure suggested that the mode of effector inhibition is similar and this will be discussed below. In contrast, Sp1 displays differences in activity and structure from Tae4 and is biologically distinct from Sp2. In addition, the structure of Rap1a presents a new fold distinct from the Tai4 proteins; hence, it was important to elucidate the structure of an Sp1–Rap1a complex.

Following purification using a co-expression strategy, we determined the structure of the Sp1–Rap1a complex at about 2.0 Å resolution. There are two molecules in the asymmetric unit (one Sp1 and one Rap1a) and a crystallographic twofold axis given by the symmetry operation x, y, −y, z + 2/3 generates a heterotetramer (Fig. 7a). The overall dimensions of this heteromeric assembly are about 90 × 50 × 50 Å and its molecular weight is approximately 60 kDa, consistent with that observed by size-exclusion chromatography during purification of the complex (Supplementary Fig. S1). The model contains residues 1–163 of Sp1 and residues 30–123 of Rap1a. The solvent-accessible surface area of Sp1 is 5840 Å² and about 16% of this (960 Å²) is occluded when the complex with Rap1a is formed.

Rap1a interacts with Sp1 using residues in α1, α3, α4, α5 and the α4–α5 loop (Fig. 7). In particular, the α4–α5 loop and α4 of Rap1a are directly positioned to block the Sp1 active site. An extensive network of hydrogen bonds, van der Waals forces and water-mediated hydrogen bonds are present at the Sp1–Rap1a interface. These interactions occur primarily in two areas (areas I and II in Fig. 7a). A total of 13 Sp1 residues (Asp16, Tyr17, Ser18, Tyr22, Ala25, Asp37, Ala47, Asn48, Arg53, Asp130, His133, His149 and Tyr151) and 12 Rap1a residues (Tyr50, Lys58, Arg62, Ser81, Gln82, Gln84, Thr85, Val86, Thr87, G1u90, G1u94 and Arg119) contribute to the network of interactions and the details are presented in a schematic form in Fig. 7(b). Gln84 of Rap1a forms a hydrogen bond to the N² atom of the catalytic His133 of Sp1 and is a clear marker of the steric block provided by the immunity protein (Figs. 7b and 7c). Arg119, Arg66, Lys59, Glu94 and Glu94 of Rap1a form salt-bridge interactions with Asp37, Asp16, G1u40, Arg53 and Arg53 of Sp1, respectively. Tyr50 in α1 of Rap1a interacts with Tyr17 from Sp1 through a hydrogen bond and π-stacking. A small interface (100 Å²) is created where Asp36 from Rap1a and His149 from Sp1 interact (area III in Fig. 7).

An overlay (not shown) of the two Rap1a structures, alone and in complex with Sp1, gives an r.m.s.d. of 0.4 Å for 94 Cα atoms, indicating a high similarity in overall structure with no major conformational differences. Rap1a therefore appears to be a highly stable pre-formed binding partner for Sp1 when in the correct redox state. There is, however, a localized effect following complex formation. In Rap1a, the segment linking α2 and α3, residues 74–84, shows high average B factors (temperature factors: 76.4 Å²) compared with the overall average B factor (31.5 Å²) and relatively diffuse electron density. This region contributes to Sp1–Rap1a complex formation and becomes well ordered, with an average B factor of 15.6 Å² compared with the overall B factor of 22.6 Å² for the Rap1a component of the complex. The overall structure of Sp1 is also well retained between the free and complexed states, with an r.m.s.d. of 0.8 Å for the least-squares fit of 162 Cα atoms. There are localized adjustments which appear to support complex formation. In the α1–α2 loop of Sp1, Tyr17 is relocated by about 7 Å to interact with Tyr50 and Arg62 of Rap1a. In addition, Asn48 and Tyr129 of Sp1 move 1.8 and 3.5 Å outwards and away from the catalytic centre (Fig. 7c).

The Tae4–Tai4 complex from E. cloacae also forms a heterotetrameric assembly, with the Tai4 dimer forming the central segment and a Tae4 effector at either end (Zhang et al., 2013). The complex is stabilized by extensive noncovalent interactions formed between the effector and both Tai4 subunits. The protruding loop (Fig. 6c) of one Tai4 subunit binds in the active site of the effector and helices α2 and α3 of the other subunit are positioned to interact with residues in the N-terminal subdomain. This Tae4–Tai4 complex is likely to be a good model for the Sp2–Rap2a complex. The close structural relationship, in terms of fold, for both the effector and the immunity proteins (see, for example, Fig. 6) would suggest that Rap2a would interact with Sp2 in a similar fashion and that the selectivity of immunity proteins with the Rap2a fold for or against other enzymes would be determined by variation in the side chains. Such differences are likely to involve residues in the loops linking α2 to α3 and α3 to α4 in the immunity proteins, which interact with the α3–α4 section, the β4–β5 loop and the N-terminal section of β7 in some Tae4 proteins. These parts of the proteins are poorly conserved in terms of sequence identity (Supplementary Figs. S4 and S5).

However, the complex assembly that leads to inhibition is completely distinct between Sp1 and the other Tae4 systems. Rap1a, which has a completely different fold to the conventional Tai4 proteins, does not possess a protruding loop, but in the complex a helix is positioned to block the effector active site and the interactions that stabilize the complex involve residues in the C-terminal subdomain of the effector rather than the N-terminal subdomain (Supplementary Fig. S6).
3.7. Co-evolved diversity within the Tae4/Sp and Tai4/Rap families

Finally, we considered the question of why *S. marcescens* has two Tae4 homologues (Ssp1 and Ssp2) and four Rap-family proteins rather than just one amidase effect–immunity combination. Ssp1 and Ssp2 are not redundant since they elicit different biological consequences in a target cell. For example, as shown here and previously (Fig. 2; English et al., 2012), an immunity mutant suffering Ssp2-mediated killing is highly unfit and displays striking cell filamentation, whereas an immunity mutant suffering Ssp1-mediated killing is less disabled and shows round enlarged cells. The difference between the Ssp1-mediated and Ssp2-mediated effects may be related to the apparent modulation of substrate preference observed in the *in vitro* enzyme assays, perhaps reflecting differences in the extent or pattern of cell-wall damage (Fig. 1).

![Figure 8](image)

**Figure 8**

Subgroups of Tae4 family proteins and their associated immunity proteins. Neighbour-joining tree calculated from a multiple sequence alignment of Ssp1, Ssp2 and other Tae4 homologues from different bacterial species (a full alignment is shown in Supplementary Fig. S6). For each Tae4 protein, the adjacently encoded candidate immunity protein was identified by genomic analysis and the *S. marcescens* Rap protein to which it was most closely related was determined. Tae4 homologues with associated immunity proteins similar to Rap1a (Tai4a) are shown by red circles, whereas those with immunity proteins of the Rap1b/Rap2a/Rap2b (Tai4) type are shown by squares (in light blue, green or dark blue for most similarity to Rap1, Rap2a or Rap2b, respectively). Apart from Ssp1 and Ssp2 from *S. marcescens*, the Tae4 homologues are labelled by organism and their identities are as follows (UniProt or genomic identifiers): Actinobacter baumannii, BOVVE3_ACBIS; Agrobacterium tumefaciens, Atu4347; Burkholderia cenocepacia, Been_4308; Cronobacter sakazakii, ESA_03935; Enterobacter cloacae, ECL_01542; Enterobacter hormaechei, F5RYK9_9ENTR; Erwinia amylovora, EAMY_3018; Erwinia tasmannensis, ETA_06210; Escherichia coli, ECEG_03250; Pantoea sp., S7A_11480; Pseudomonas syringae, Psyr_4040; Salmonella enterica serovar Newport, SNSL254_A0303; S. enterica serovar Typhi, STY0307; Serratia odorifera, DME4R6_SEROD; *A. tumefaciens* and *S. marcescens* Tae4 homologues for which structures have previously been reported (Zhang et al., 2013) and the open square indicates that this candidate immunity protein showed only weak similarity to Rap2a. Details of the Tae4 homologues and associated immunity proteins are given in Supplementary Table S1.

This study has unexpectedly revealed that the cognate immunity protein for Ssp1, Rap1a, has a distinct structure compared with the other Rap and Tai4 proteins studied to date (Fig. 5).

A comparison of the amino-acid sequences of Ssp1, Ssp2 and Tae4 homologues showed that Ssp1 and several other Tae4 homologues form a distinct grouping (Fig. 8). This is consistent with several structural differences being observed in Ssp1 compared with *Es* Tae4 (Supplementary Fig. S2). When the adjacently encoded known or putative immunity protein was identified for all of these effectors and compared with the four *S. marcescens* Rap proteins, a pattern became evident. The Ssp1-like proteins all co-occur with immunity proteins of the Rap1a type. The other Tae4 homologues are less separated from each other, but there is a clustering of Tae4 proteins sharing related immunity proteins. So, for example, the three Tae4 proteins whose immunity proteins are most closely related to Rap1b all cluster together (Fig. 8) and the Tae4 homologues most similar to Ssp2 all have associated Rap2a-like immunity proteins (Fig. 8). Hence, effector and immunity proteins appear to have co-evolved within the Tae4 and Tai4 family. In particular, the Ssp1-like proteins appear to form a subgroup distinct enough to utilize a structurally unrelated immunity protein (‘Tai4a’), exemplified by Rap1a. The reason for this divergence is unclear, although it is consistent with the distinct biological phenotypes associated with the two effectors. Having both Ssp1 and Ssp2 is likely to confer an evolutionary advantage on the secreting organism, perhaps with each being more efficient against different target species or under different growth conditions than the other. Additionally, having Ssp1 may allow attack on a close relative with Ssp2/Rap2a and *vice versa*, maximizing the ability to distinguish ‘self’ from competitors.

This analysis strongly suggests that Rap1b and Rap2b are not ‘inactive’ immunity proteins; rather, they are likely to provide protection against incoming Tae4 proteins from other T6SS-elaborating bacterial species. Rap1b is most closely related to immunity proteins associated with Tae4 proteins (Fig. 8) and thus would be expected to bind an effector of this type and not Ssp1- or Ssp2-like Tae4 proteins. Similarly, Rap2b may neutralize Tae4 proteins related to *Es* Tae4 (Fig. 8). This idea would also predict that neither Rap1b nor Rap2b would bind Ssp1 or Ssp2, and indeed we have shown previously that they do not (English et al., 2012). Other bacteria can also be observed to possess ‘extra’ Tae4 proteins compared with their Tae4 complement (this study and Russell et al., 2012). For example, *Cronobacter sakazakii* has ‘orphan’ Rap1a, Rap2a-like and Rap2b-like proteins (ESA_03939, ESA_03933 and ESA_03932) in addition to its Ssp1–Rap1a pairing. Thus, one could envisage an ‘arms race’ in which attackers can gain an advantage by acquiring a Tae4 protein of a different subgroup whilst targets can counter by acquiring an immunity protein of the matching type.

4. Concluding remarks

Our structural and activity data, and comparisons with related systems, reveal that the T6SS family 4 endopeptidase effectors...
are likely to share the same enzyme mechanism but fall into two functional categories in terms of overall structure and substrate processing. One displays narrow specificity and one is more promiscuous. Some of the immunity proteins encoded within the T6SS gene clusters, those cognate to Tae4/family 4 effectors, can be placed in two protein-fold families which we suggest be termed (i) Tai4, as described previously and including Rap2a, EcTai4 and STTai4, and (ii) Tai4a, including Rap1a and described for the first time here. Although distinct in terms of structure, the two families are built upon similar underlying principles, namely stable dimeric small α-helical bundles. These T6SS-associated immunity proteins and indeed the effectors in *S. marcescens* appear to rely on the formation of disulphide linkages for folding and activity.

Although the mode of effector inhibition is conserved in the two families, namely a steric block of the active site within a heterotetrameric complex, our study reveals that a very different structure can be used to accomplish effector neutralization. The inhibitory action of immunity proteins is therefore highly specific for effector proteins, even though some effectors have closely related enzyme activities. The type VI-associated endopeptidase effectors are highly basic proteins, as exemplified by Ssp1 and Ssp2, with predicted pI values of 9.1 and 9.3, respectively. These enzymes act on peptidoglycan and process a substrate in the vicinity of two acidic carboxylates, the δ-Glu residues. The immunity proteins are acidic; the pI values for the four Rap proteins fall in the range 5.1–6.3 after omitting the signal peptides. Such acidic carboxylates, the

References

research papers


Supplementary Material

Structural basis for Type VI secretion peptidoglycan DL-endopeptidase function, specificity and neutralization in *Serratia marcescens*.

Velupillai Srikannathasan, Grant English, Nhat Khai Bui, Katharina Trunk, Patrick E. F. O’Rourke, Vincenzo A. Rao, Waldemar Vollmer, Sarah J. Coulthurst* and William N. Hunter*

Supplementary Table S1. Ssp/Tae4 family proteins and corresponding adjacently-encoded immunity proteins.

<table>
<thead>
<tr>
<th>Organism</th>
<th>Ssp/Tae4 protein</th>
<th>Adjacent Immunity Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>UNIPROT identifier or name</td>
<td>Genomic identifier</td>
</tr>
<tr>
<td><em>Agrobacterium tumefaciens</em> C58</td>
<td>Q7CPU9_AGRT5</td>
<td>Atu4347</td>
</tr>
<tr>
<td><em>Burkholderia cenocepacia</em> AU1054</td>
<td>Q1BN86_BURCA</td>
<td>Becn_4030</td>
</tr>
<tr>
<td><em>Cronobacter sakazakii</em> ATCC BAA-894</td>
<td>A7MQ14_CROS8</td>
<td>ESA_03935</td>
</tr>
<tr>
<td><em>Enterobacter cloacae</em> ATCC 13047</td>
<td>D5C6F6_ENTCC</td>
<td>ECL_01542</td>
</tr>
<tr>
<td><em>Enterobacter hormaechai</em> ATCC 49162</td>
<td>F5RYK9_9ENTR</td>
<td>F5RYK8_9ENTR</td>
</tr>
<tr>
<td><em>Erwinia amylovora</em> CFBP1430</td>
<td>D4I0Q7_ERWAC</td>
<td>EAMY_3018</td>
</tr>
<tr>
<td><em>Erwinia tasmaniensis</em> F5/99</td>
<td>B2VH84_ERWT9</td>
<td>ETA_06210</td>
</tr>
<tr>
<td><em>Escherichia coli</em> B354</td>
<td>D6J6Z7_ECOLX</td>
<td>ECEG_03250</td>
</tr>
<tr>
<td><em>Pantoea</em> sp. Sc1</td>
<td>H8DNR2_9ENTR</td>
<td>S7A_11480</td>
</tr>
<tr>
<td><em>Pseudomonas syringae</em> pv. syringae B728a</td>
<td>Q4ZP52_PSEU2</td>
<td>Pyr_4040</td>
</tr>
<tr>
<td><em>Salmonella Newport</em> SL254</td>
<td>B4SV53_SALNS</td>
<td>SNSL254_A0303</td>
</tr>
<tr>
<td><em>Salmonella Typhi</em> CT18</td>
<td>Q8Z963_SALTI</td>
<td>STY0307</td>
</tr>
<tr>
<td><em>Salmonella Typhimurium</em> LT2</td>
<td>Q93IS4_SALTY</td>
<td>STM0277</td>
</tr>
<tr>
<td><em>Serratia odorifera</em> DSM 4582</td>
<td>D4E4R6_SEROD</td>
<td></td>
</tr>
<tr>
<td><em>Acinetobacter baumannii</em> SDF</td>
<td>B0VVE3_ACIBS</td>
<td>p2ABSDF0033</td>
</tr>
<tr>
<td><em>Serratia marcescens</em> Db10</td>
<td>Ssp1</td>
<td>SMA2261</td>
</tr>
<tr>
<td><em>Serratia marcescens</em> Db10</td>
<td>Ssp2</td>
<td>SMA2264</td>
</tr>
</tbody>
</table>

Note that this is not the same protein as the one identified as a Tae4 protein in this organism by Russell et al., 2012; that protein, ESA_03932, is an orphan Tae4 protein of the Rap2b type not immediately adjacent to Tae4.

² STY0306 (SciQ) is approximately twice the size of other Rap proteins, resembling a fusion of two adjacent Rap1a proteins.

³ p2ABSDF0034 is shows only very weak sequence similarity with Rap2a.
Supplementary Figure S1. Immunity protein binding and secretion is not impaired in Ssp1 and Ssp2 C50A mutants. (A) Size exclusion chromatography analysis of complex formation between Ssp1 (C50A) and Rap1a, top, or Ssp2 (C50A) and Rap2a, bottom. 10 nmol of the protein indicated, or of each protein in the case of the mixtures, was separated on a calibrated Superdex 75 10/300 GL column. (B) Immunoblot detection of Ssp2 in the secreted fraction of the strains indicated: wild type S. marcescens Db10 [WT]; mutant lacking Ssp2 [Δssp2]; Δssp2 mutant carrying plasmids expressing wild type Ssp2 [+Ssp2(WT); pSC541] or the C50A mutant of Ssp2 [+Ssp2(C50A); pSC1230]; and a Type VI secretion system mutant [ΔT6SS] expressing wild type Ssp2.
Supplementary Figure S2. Superimposition of Ssp1 (orange ribbon) and EcTae4 (green ribbon). The catalytic triads (histidine, cysteine and aspartate) are shown as sticks and divergent regions are shown in the boxes.

Supplementary Figure S3. Superimposition of the catalytic residues of Ssp1, PaTse1 and EcTae4. The color code is N blue, O red, S yellow then C positions for Ssp1 orange, PaTse1 grey and EcTae4 green.
Supplementary Figure S4. Sequence alignment of Ssp1 and selected homologues. A. Structure-based sequence alignment highlights the conserved secondary-structure content (orange cylinders and sheets) in this group of four endopeptidase effectors. Residues involved in disulfide bond formation are coloured yellow. The catalytic histidine and cysteine residues are marked in cyan. The alignment was generated using ClustalW and the figure was prepared using ALINE (Bond & Schüttelkopf, 2009). B. An alignment of S. marcescens Rap2a with Tai4 from E. cloacae and S. Typhimurium. Sequence numbering and secondary structure has been drawn based on the Rap2a crystal structure (without signal peptide - brown). Strictly conserved residues in all four sequences are encased in black, conserved in two or three in shades of grey with the conserved cysteines involved in disulfide bond formation highlighted in yellow.
Supplementary Figure S5. Multiple sequence alignment of Ssp1 and Ssp2 with other Tae4 proteins. Alignment was generated using MUSCLE (Edgar, 2004) and visualised using Jalview (Waterhouse et al., 2009). Red arrows show the amino acid residues forming the catalytic triad. The Ssp1-like proteins associated with Rap1a-type immunity proteins are also indicated. The Tae4 homologues are labelled by organism and their identities are as follows (UNIPROT identifier, genomic identifier): *Acinetobacter baumannii* (B0VVE3_ACIBS, p2ABSDF0033), *Agrobacterium tumefaciens* (Q7CUP8_AGRT5, Atu4347), *Burkholderia cenocepacia* (Q1BN86_BURCA, Bcen_4030), *Cronobacter sakazakii* (A7MQ14_CROS8, ESA_03935), *Enterobacter cloacae* (D5C6F6_ENTCC, ECL_01542), *Enterobacter hormaechei* (F5RYK9_9ENTR), *Erwinia amylovora* (D4I0Q7_ERWAC, EAMY_3018), *Erwinia tasmaniensis* (B2VH84_ERWT9, ETA_06210), *Escherichia coli* (D6J6Z7_ECOLX, ECEG_03250), *Pantoea* sp. (*H8DNR2_9ENTR, S7A_11480), *Pseudomonas syringae* (Q4ZP52_PSEU2, Psyr_4040), *Salmonella enterica* serovar Newport (B4SV53_SALNS, SNSL254_A0303), *Salmonella enterica* serovar Typhimurium (Q93IS4_SALTY, STM0277), *Salmonella enterica* serovar Typhi (Q8Z963_SALTI, STY0307), *Serratia odorifera* (D4E4R6_SEROD). See also Supplementary Table S1.
Supplementary Figure S6. Distinct placement of the immunity proteins inhibits different peptidoglycan endopeptidases. Four polypeptides are shown in ribbon style; Ssp1 is colored orange, EcTae4 is green, one subunit of EcTai4 is grey and one subunit of Rap1a is cyan. A least-squares overly of Ssp1 and Tac4 was carried out and this then highlights the different positions that the immunity proteins occupy when they bind and inhibit their cognate effector.