PHD

Structural studies on bacterial toxins

Holbourn, Kenneth Paul

Award date:
2006

Awarding institution:
University of Bath

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: 10. Jan. 2020
Structural studies on bacterial toxins

(Volume 1 of 3)

Kenneth Paul Holbourn

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Department of Biology and Biochemistry

September 2006

Copyright
Attention is drawn to the fact that copyright of this thesis rests with its author. The copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without the prior written consent of the author.

This thesis may be available for consultation within the University library and may be photocopied or lent to other libraries for the purpose of consultation.

Kenneth Holbourn
# Contents list

Contents list .......................... I  
List of figures ........................ VI  
List of tables .......................... VIII  
Acknowledgements ....................... IX  
Abstract ............................... X  
List of Abbreviations ................... XI

1. An introduction to ADP-ribosylating toxins .......................... 1 
   1.1 An introduction to ADP-ribosylating toxins ......................... 1  
   1.2 Cellular properties of ADPRTs .................................. 6  
   1.3 Structural analysis of the NAD binding site between ADPRTs .... 10  
      1.3.1 The ADPRT core fold .................................... 11  
      1.3.2 The ARTT motif ....................................... 13  
      1.3.3 The STS motif ....................................... 14  
      1.3.4 The key Arg/His residue ................................ 16  
      1.3.5 The Tyr-X_{10}-Tyr motif ................................ 17  
      1.3.6 The PN loop ......................................... 18  
      1.3.7 The α-3 motif ....................................... 21  
   1.4 Catalytic model and mechanism .................................... 22  
   1.5 Summary ................................................. 24

2. The structure of the C3bot-RalA complex at 2.66Å ...................... 26  
   2.1 Aims of structural studies on the C3bot-RalA complex .............. 26  
   2.2 The C3bot exoenzyme from *Clostridium botulinum* ................. 26  
      2.2.1 The origins of the C3bot exoenzyme ........................ 26  
      2.2.2 Related C3-like exoenzymes ................................ 27  
      2.2.3 Biological role of C3bot ................................ 27  
      2.2.4 The structure of C3bot .................................. 31  
   2.3 Interactions between C3bot and RalA ................................ 32
2.4 Methods

2.4.1 Expression of C3bot 38
2.4.2 Purification of C3bot 38
2.4.3 Expression of RalA 39
2.4.4 Purification of RalA 39
2.4.5 Formation of the C3bot-RalA complex 40
2.4.6 Crystallisation of the C3bot-RalA complex 40
2.4.7 Data collection and refinement 40

2.5 Results and discussion 44

2.5.1 Experimental results 44
2.5.2 Overall structure of the C3bot-RalA complex 48
2.5.3 The C3bot-RalA interface 52
  2.5.3.1 The role of the ARTT and PN loops 54
2.5.4 Implications for other C3 exoenzymes 58
2.5.5 Implications for other GTPases 61
2.5.6 An alternative structure 62

2.6 Conclusions 63

3. The structure of the C3bot-RalA-NAD complex 64

3.1 Aims of obtaining an NAD complex 64

3.2 NAD binding and C3bot-RalA interactions 64

3.3 Methods 66
  3.3.1 Expression and purification of C3bot and RalA 66
  3.3.2 Crystallisation 66
  3.3.3 Soaking NAD into C3bot-RalA crystals 66
  3.3.4 Data collection and refinement 66

3.4 Results and discussions 69
  3.4.1 Experimental results 69
  3.4.2 Overall structure of the C3bot-RalA-NAD complex 7
  3.4.3 The NAD binding site 73
  3.4.4 Electrostatic charge around the active site 76
List of Figures

Figure 1.1. Schematic of the ADP ribosylation reaction 1
Figure 1.2 Representative members of the 4 classes of bacterial ADPRT 2
Figure 1.3 The NAD binding site of ADPRTs 12
Figure 1.4 The conserved ADPRT binding motifs 20
Figure 2.1 The Rho GTPase cycle 29
Figure 2.2 The structure of C3bot 31
Figure 2.3 The structure of RalA 34
Figure 2.4 The role of Ral in the cell 35
Figure 2.5 A diffraction pattern from the C3bot-RalA complex 42
Figure 2.6 Purification of C3bot and RalA 44
Figure 2.7 Native gel of the 2 competing crystal forms 45
Figure 2.8 Crystals of the C3bot-RalA complex 46
Figure 2.9 Electron density surrounding the GDP molecule 47
Figure 2.10 Ramachandran plot of the refined structure of C3bot-RalA 49
Figure 2.11 The structure of the C3bot-RalA complex 50
Figure 2.12 The GDP binding site of RalA 51
Figure 2.13 The interface between RalA and C3bot 53
Figure 2.14 Comparison of the ARTT loops in C3bot structures 57
Figure 2.15A Sequence alignment of the C3 family 59
Figure 2.15B Residues of C3bot and C3stau around the proposed interface 60
Figure 2.16. Sequence alignment of the GTPase family 62
Figure 3.1 Electron density in the active site of C3bot 70
Figure 3.2 Ramachandran plot of the refined structure of C3bot-RalA-NAD 71
Figure 3.3 Structure of the C3bot-RalA-NAD complex 72
Figure 3.4. A detailed view of the interface between C3bot and RalA 73
Figure 3.5 The NAD binding site highlighting some of the important interactions 74
Figure 3.6 Differences between wild type and RalA bound C3bot-NAD 75
Figure 3.7 Electrostatic surface picture of the C3bot active site 76
Figure 4.1 Purification of RalA mutants
Figure 5.1 Crystals of the different C3bot-RalA complexes
Figure 5.2 The alternative structure of a C3bot-RalA complex
Figure 5.3 Ramachandran plot of the alternative C3bot-RalA complex
Figure 5.4 A detailed view of the switch II interface
Figure 5.5 A detailed view of the crystal interface
Figure 5.6 Sequence alignment of the C3exoenzymes
Figure 5.7 Sequence alignment of RalA and related GTPases
Figure 6.1 The structure of RhoA
Figure 6.2 Potential crystals of the C3bot-RhoA complex
Figure 7.1 Insoluble purification of SpyA
Figure 7.2 Pure SpyA and dimerised SpyA
Figure 7.3 Sequence alignment of SpyA with related ADPRTs
Figure 7.4 The homology model of SpyA
Figure 7.5 A close-up view of the active site of the SpyA model
Figure 8.1 Surface models of C3bot & RalA
Figure 8.2 Surface electrostatic charge depiction of the Switch II interface
Figure 8.3 B-factor plots of C3bot and RalA
List of Tables

Table 1.1 Summary of the ADPRTs that have had their 3D structures determined, their targets and physiological effects 3
Table 2.1 Details of crystallographic data 43
Table 2.2 Potential hydrogen bonds between C3bot and RalA 54
Table 2.3 Water mediated hydrogen bonds between the C3bot and RalA 54
Table 2.4 Contacts in the C3bot-RalA structure 55
Table 3.1 Details of crystallographic data 68
Table 5.1 Table of crystallographic statistics 93
Table 5.2 Potential hydrogen bonds between C3bot and RalA 95
Table 5.3 Contacts between C3bot and RalA 95
Table 5.4 Potential hydrogen bonds between C3bot and RalA 99
Table 5.5 Contacts between C3bot and RalA 99
Table 8.1 Protein-protein interaction server results 129
Acknowledgments

There are many people I would like to thank for their help, support or assistance over the past 3 years. My friends and family have all been very supportive of me and special thanks to my lovely wife, Susan. Without her support it would have been much, much harder.

All of my co-workers in lab 0.34 deserve thanks for the help they've given me. Especially to Matt who managed to teach me the basics of UNIX and Steve who reminded how the machines in the wet lab actually worked.

The staff at the SRS Daresbury and EMBL Hamburg are also due thanks for their help, assistance and beam time at the synchrotrons.

Finally I would like to thanks Ravi and Cliff for giving me the chance to work on these projects and for their supervision and through the University and HPA the funding for my three years in Bath.
Abstract

This work is primarily on ADP-ribosylating toxins. ADP-ribosylating toxins (ADPRTs) are a family of toxins that catalyse the hydrolysis of NAD and the transfer of the ADP-ribose moiety onto a target. This family includes many notorious killers, responsible for thousands of deaths annually including: cholera, enterotoxic Escherichia coli, whooping cough, diphtheria and a plethora of Clostridial binary toxins. One of their members, the C3bot exoenzyme from Clostridium botulinum, has been extensively used as a cellular tool to study and elucidate the functions of small GTPases that it targets. These studies have focused on the well documented specificity of C3bot for the mammalian GTPase Rho. Recent work, however, has highlighted a novel interaction between C3bot and another small mammalian GTPase, RalA. It is this novel interaction between C3bot and RalA and the biochemical consequences of this interaction that have been investigated.

In this work, we present an overview of structural motifs that differentiate the different classes of bacterial ADPRTs in relation to their function. We also present the structure of the C3bot with the small GTPase RalA in two different forms, and in the case of one form, with C3bot in its active (NAD bound) and inactive form. Attempts to determine the structure of C3bot with its primary target RhoA and the purification of a related ADPRT from Streptococcus pyogenes, SpyA, are also described. This also includes an analysis of the differing forms of the complex for an insight into the biologically relevant form.

A brief overview of X-ray crystallography and experimental work performed on a family of cysteine proteases that are produced by pathogenic bacteria such as Yersinia pestis, E. coli O157:H7, Haemophilus somnus and the plant pathogen Psuedomonas syringae are included in the appendices.
List of Abbreviations

ADP   Adenine diphosphate
ADPRT ADP-ribosylating toxin
A/E   Attaching and effacing
AFX   Forkhead transcription factor
ARTT  ADP-ribosyl turn-turn
ATP   Adenine triphosphate
AMoRe Automated molecular replacement
ARF6  ADP-ribosylation activation factor 6
ART   ADP-ribosyltransferases
Avr   Avirulence proteins
BisTris Bis(2-hydroxyethyl)imino tris(hydroxymethyl)methane
CCD   Charge coupled device
CT    Cholera toxin
DT    Diphtheria toxin
DTT   Dithiothreitol
EDIN  Epidermal differentiation inhibitor
EDTA  Ethylenediaminetetraacetic acid
eEF2  Elongation factor 2
Efa1  EHEC factor for adherence
EHEC  Enteropathogenic E. coli
EPEC  Enterohemorrhagic E. coli
GAP   GTPase activating protein
GDP   Guanidine diphosphate
GDI   Guanine nucleotide dissociation inhibitor
GEF   Guanine nucleotide exchange factor
GppNHp S'-Guanylylimidodiphosphate
GTP   Guanidine triphosphate
GST   Glutathione S-transferase
HB-EGF Epidermal like growth factor growth factor precursor
Hepes 4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid
His₆ Hexa-histidine
Hop Hrp outer protein
Hrp Hypersensitivity and pathogenicity genes
HUS Haemolytic uremic syndrome
HW-IGBPs High weight - immunoglobulin binding proteins
IGBPs Immunoglobulin binding proteins
IPTG isopropyl-β-D-thiogalactopyranoside
IR Isomorphous replacement
LB Luria broth
LEE Locus of enterocyte effacement
LIC Ligation independent cloning
LSQ Least squares
LT E. coli heat labile enterotoxin
MAD Multiple wavelength anomalous scattering
MPD 2-Methyl-2, 4-pentanediol
MBP Maltose binding protein
MES 2-Morpholinoethanesulfonic acid
MIR Multiple isomorphous replacement
MR Molecular replacement
NAD Nicotinamide dinucleotide
NF-κB Nuclear factor κB
NMN Nicotinamide mononucleotide moiety
PAETA Pseudomonas aeruginosa exotoxin A
PARP Poly ADP-ribose polymerases
PCR Polymerase chain reaction
PEG Polyethylene glycol
PLD1 Phospholipase D1
PN Phosphate nicotinamide
PT Pertussis toxin
RMSD Root mean square deviation
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDS</td>
<td>Sodium dodecyl (lauryl) sulfate</td>
</tr>
<tr>
<td>SDS-Page</td>
<td>Sodium dodecyl (lauryl) sulfate-polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>SPN</td>
<td>Streptococcus pyogenes NAD glycohydrolase</td>
</tr>
<tr>
<td>Syc</td>
<td>Specific YOP chaperones</td>
</tr>
<tr>
<td>TB</td>
<td>Terrific broth</td>
</tr>
<tr>
<td>TCF</td>
<td>Ternary complex factor</td>
</tr>
<tr>
<td>Tris</td>
<td>Tris(hydroxymethyl)methane</td>
</tr>
<tr>
<td>TTSS</td>
<td>Type III secretion system</td>
</tr>
<tr>
<td>Vip2</td>
<td>Vegetative insecticidal protein</td>
</tr>
<tr>
<td>Yop</td>
<td>Yersinia outer proteins</td>
</tr>
<tr>
<td>Ysc</td>
<td>Yersinia secretion genes</td>
</tr>
</tbody>
</table>
1.1 Introduction to ADP-ribosylating toxins

Pathogenic bacteria are known to possess an arsenal of toxins and effectors that assist them in targeting and killing their host cells. The ADP-ribosylating toxins (ADPRTs) are a large family of dangerous and potentially lethal toxins. Examples of these toxins can be found in a diverse range of bacterial pathogens and they are the principal causative agents in many serious diseases including cholera, whooping cough and diphtheria. ADPRTs, as the name would suggest, break NAD into its component parts (nicotinamide and ADP-ribose) before selectively linking the ADP-ribose moiety to their protein target, as simple schematic of this is shown in figure 1.1.

Figure 1.1 Schematic of the ADP ribosylation reaction

Figure 1.1 A generalised mechanism of ADP-ribosylation: NAD is bound to the toxin and the catalytic glutamate forms a hydrogen bond with the 2'-OH of the ribose. This hydrogen bond stabilises the active intermediate and leaves the N-glycosidic bond vulnerable to nucleophilic attack from the target. This results in ADP-ribose being covalently bonded to the target.
In the majority of these toxins the targets are key regulators of cellular function and interference in their activity, caused by ADP-ribosylation, leads to serious deregulation of key cellular processes and in most cases, eventual cell death.

This large family of toxins has been extensively studied with many structures of individual members determined. These include: diphtheria toxin (1TOX) (Bell & Eisenberg, 1996), pseudomonas exotoxin A (1AER) (Li et al, 1996), pertussis toxin (1PRT) (Stein et al, 1994), cholera toxin (1XTC) (Zhang et al, 1995), E.coli heat labile enterotoxin (1LTS) (Sixma et al, 1993), iota toxin (1GIQ) (Tsuge et al, 2003), vegetative insecticidal protein (1QS1) (Han et al, 1999) and the C3-like toxins, C3bot (1G24) (Han et al, 2001) and C3stau (1OJZ) (Evans et al, 2003). These structures and extensive cellular and functional research performed over the last 20 years have provided an enormous insight into the function of these toxins and an understanding on their effects on host cells. These data are summarised in table 1.1.

The ADPRT family can be split into four groups on the basis of their domain organisation and the nature of their target. The 3D structure of a representative member of each group is shown in figure 1.2. The most well known toxins: cholera, pertussis and the E.coli enterotoxin are members of the AB$_5$ family which target small regulatory G-proteins. The enzymatically active A subunit is situated on a scaffold made of a pentamer of B-subunits (Cassel & Pfeuffer, 1978; Gill, 1976; Locht & Keith, 1986; Streatfield et al, 1992; Tamura et al, 1982; Zhang et al, 1995). Diphtheria and Pseudomonas exotoxin A ribosylate a diphthamide residue on elongation factor 2. Both are large multi-domain proteins with receptor binding, transmembrane targeting and protease-resistant catalytic domains (Allured et al, 1986; Collier, 1975; Hwang et al, 1987; Morris et al, 1985; Sandvig & Olsnes, 1980; Wilson & Collier, 1992). The third group are the actin-targeting AB binary toxins that, unlike the more common AB$_5$ binary toxins, comprise of two domains, an active catalytic domain and a cell-binding domain. This group includes a wide range of clostridial toxins including C2 toxin from *Clostridium botulinum*, *Clostridium perfringens* iota toxin, *Clostridium spiroforme* toxin, *Clostridium difficile* toxin and

Table 1.1 Summary of the ADPRTs that have had their 3D structures determined, their targets and physiological effects.

<table>
<thead>
<tr>
<th>Toxin</th>
<th>Organism</th>
<th>PDB ID</th>
<th>Class</th>
<th>Target</th>
<th>Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pertussis toxin</td>
<td><em>Bordetella pertussis</em></td>
<td>1PRT</td>
<td>AB3</td>
<td>Cysteine on Gα, Gβ and Gγ</td>
<td>Uncoupling of effectors from the adenylate cyclase pathway</td>
</tr>
<tr>
<td>Cholera toxin</td>
<td><em>Vibrio cholerae</em></td>
<td>1XTC</td>
<td>AB3</td>
<td>Arg on Gα</td>
<td>Trapping of G-protein in GTP bound states and uncontrolled upregulation of adenylate cyclase</td>
</tr>
<tr>
<td>E.coli heat labile enterotoxin</td>
<td><em>Escherichia coli</em></td>
<td>1LTS</td>
<td>AB3</td>
<td>Arg on Gα</td>
<td>Trapping of G-protein in GTP bound states and uncontrolled upregulation of adenylate cyclase</td>
</tr>
<tr>
<td>Diphtheria toxin</td>
<td><em>Corynebacterium diptheriae</em></td>
<td>1TOX</td>
<td>AB</td>
<td>Diphthamide on EF2</td>
<td>Inhibition of protein synthesis</td>
</tr>
<tr>
<td>Pseudomonas exotoxin A</td>
<td><em>Pseudomonas aeruginosa</em></td>
<td>1AER</td>
<td>AB</td>
<td>Diphthamide on EF2</td>
<td>Inhibition of protein synthesis</td>
</tr>
<tr>
<td>VIP2</td>
<td><em>Bacillus cereus</em></td>
<td>1QS1</td>
<td>AB binary toxin</td>
<td>Arg177 on Actin</td>
<td>Prevent actin polymerisation</td>
</tr>
<tr>
<td>Iota toxin</td>
<td><em>Clostridium perfringens</em></td>
<td>1GIQ</td>
<td>AB binary toxin</td>
<td>Arg177 on Actin</td>
<td>Prevent actin polymerisation</td>
</tr>
<tr>
<td>C3bot</td>
<td><em>Clostridium botulinum</em></td>
<td>1G24</td>
<td>Single polypeptide</td>
<td>Asn41 on Rho A-C</td>
<td>Trap Rho GTPase in GDP bound state and leads to dissagregation of actin cytoskeleton</td>
</tr>
<tr>
<td>C3stau</td>
<td><em>Staphylococcus aureus</em></td>
<td>1OJZ</td>
<td>Single polypeptide</td>
<td>Asn41 on Rho A-C, RhoE and Rnd3</td>
<td>Trap Rho GTPase in GDP bound state and leads to dissagregation of actin cytoskeleton</td>
</tr>
<tr>
<td>Ecto-ART2</td>
<td><em>Rat</em></td>
<td>1OG1</td>
<td>Single polypeptide</td>
<td>Arg residue on integrins</td>
<td>Cell regulation and a role in apoptosis</td>
</tr>
</tbody>
</table>
The final group are the small single domain C3 exoenzymes that have an unknown role in bacterial pathogenesis, but are widely used as tools in cellular signalling work, and are characterised by C3bot from *Clostridium botulinum*. This group also includes

Several ADPRT structures determined to date have been elucidated in the presence of bound NAD molecule or non-hydrolysable NAD analogues and these have allowed a detailed understanding of NAD binding. These structures combined with biochemical results have also suggested a possible catalytic mechanism. The current understanding of the mechanism of catalysis is that NAD is bound by the ADPRT in a manner that orients the glycosidic bond to render it amenable to hydrolysis. The "catalytic glutamate residue", that all known ADPRTs possess, forms a hydrogen bond with the 2'-OH of the ribose ring and this can be seen for all 3 classes in Figures 1.3A-C. This interaction stabilises a positively charged oxocarbenium ion intermediate which is then attacked by a nucleophile, either an activated water molecule in the case of auto-hydrolysis that many ADPRTs can demonstrate, or the protein substrate. A simplified version of this mechanism is shown in figure 1.1. While the toxin-substrate recognition process has still not been fully understood, through biochemical and mutagenic analysis there are some elements of protein-recognition that are known. In the case of cholera, pertussis, the E.coli enterotoxin and the diphtheria family of toxins, an "active site loop" has been shown to be essential for substrate binding (Han & Tainer, 2002). An example of such a detailed interaction has been provided by the recently determined crystal structure of elongation factor 2 (eEF2) and Pseudomonas aeruginosa exotoxin A (PAETA) complex (Jorgensen et al, 2005) in which the active site loop (L4 in PAETA) plays an important role. Likewise, the active site loop has been shown to be essential for activity in cholera toxin, E.coli enterotoxin and the distantly related ExoS and T toxins (Bell & Eisenberg, 1996). The recent structure of the complex between the activated form of cholera toxin (CT) and a human ADP-ribosylation activation factor 6 (ARF6) may suggest the mechanism of this active site loop involvement. In the activated ARF6-CT complex, the binding of ARF6 causes an allosteric change on the CT toxin that results in the active site loop forming an extended ‘knob’ near the
ADP-ribosyl turn-turn (ARTT) loop altering the structure of the CT active site into a more suitable one for substrate binding and ADP-ribosylation. In the C3 and Iota-like toxins the details of substrate recognition are less clear though all of them possess an aromatic residue on the ARTT loop that has been found to be critical for substrate binding (Wilde & Aktories, 2001). The conserved Q/E residue that is found two residues upstream of the catalytic glutamate in the ADPRTs may also play a role in substrate specificity. The Rho binding toxins all possess a Q-x-E motif whilst the actin binding toxins possess an E-x-E motif as shown in the sequence alignment in figure 2.4A. The change between Q and E has also been demonstrated to have substrate altering properties in the eukaryotic Ecto-ART proteins (Ritter et al, 2003). However, as only one toxin-substrate complex has been determined so far, the exact mechanism and the process of protein-protein recognition still remain much of a mystery.

1.2 Cellular properties of ADPRTs

The bacterial ADPRTs are all thought to play important roles in bacterial pathogenesis acting as key virulence factors in many diseases. The disruption caused by the ADPRTs varies considerably between the four classes, but all rely on the ADP-ribosylation of key regulatory proteins in the host cells to disrupt cell signalling and interfere with downstream regulatory and structural processes.

The AB₅ proteins (pertussis toxin (PT), E.coli heat labile enterotoxin (LT) and cholera toxin (CT)) all ribosylate a small subsection of the G-protein family. In all three toxins, the A₁ catalytic domain sits on top of a doughnut shaped pentamer of binding domains that comprise the cell binding and translocation apparatus (Finkelstein et al, 1987; Gill et al, 1981; Sixma et al, 1991; Stein et al, 1994; Zhang et al, 1995). In the bacterial cell, this hetero-hexamer is assembled in the bacterium before being transported across the membrane via the type II secretion apparatus (Sandkvist et al, 2000). Once secreted into the lumen of the gut the B-pentamer recognises the GM-1 ganglioside on the host cell surfaces inducing endocytosis and
translocation into the cytosol. Trafficking and processing of the full holotoxin in the host cell is a tremendously complex process and the description is outside the scope of this review. For the toxin to become active however, the catalytic domain must undergo proteolytic cleavage of the disulphide linked A1-A2 domain before becoming fully active (Gill, 1976; Kassis et al, 1982). This also results in the now separate A1 domain being released from the A2-Bs complex. Even then these toxins are not fully functional and require activation by host cell proteins to become fully active. In the case of cholera toxin, these ADP-ribosylation activation factors (ARFs) come from the host and are small GTPases that bind the cholera toxin in their GTP-bound state. Both CT and LT target the Gₛₒₐ the stimulatory G-protein of the adenylate cyclase system. ADP-ribosylation of this causes the G-protein to be maintained in its activated GTP bound state (Galloway et al, 1987) and leads to a massive upregulation of adenylate cyclase and subsequent increase in the amount of cytosolic cyclic AMP (Gill, 1976; Gill & Richardson, 1980). This eventually leads to a major loss of fluids and ions from the affected intestinal cells and gives rise to the severe diarrhoea and fluid loss associated with both cholera and enterotoxigenic E.coli pathogenesis (Field et al, 1989; Peterson & Ochoa, 1989).

Pertussis toxin (PT) is one of the primary virulence agents produced by Bordetella pertussis, the major causative agent of whooping cough. Pertussis toxin ADP-ribosylates an exposed cysteine residue on several small heteromeric G-proteins. The most prominent examples being Gₒₒₐ, Gₒₒₒ, and Gₒₒₒa (Katada & Ui, 1982). This results in uncoupling of the G-proteins from their effectors and an unregulated increase in adenylate cyclase activity and an increase in cyclic AMP (West et al, 1985). As many cells possess PT receptors the physiological effects of PT pathogenesis vary greatly from one cell type to another.

Diphtheria toxin (DT) and Pseudomonas aeruginosa exotoxin A (PAETA) are both examples of elongation factor 2 (eEF2) ribosylating toxins (Wilson & Collier, 1992). Upon cell entry they both specifically ribosylate an exposed histidine that has been modified by the addition of a diphthamide side-group (Van Ness et al, 1980). The
ADP-ribosylation interrupts the function of eEF2 in the host cell interfering with protein synthesis which results in profound physiological changes and ultimately cell death (Collier, 1975; Iglewski et al, 1977). The events leading up to this point are well understood; and appear to rely on the action of the receptor binding and transmembrane targeting domains. PAETA binds to the α2-macroglobulin receptor on the cell surface and induces receptor-mediated endocytosis becoming internalised into endosomes where the low pH creates a conformational change in the toxin leaving it open to furin protease cleavage that removes the binding domain. The catalytic domain then undergoes retrograde transport to the endoplasmic reticulum, translocates into the cytoplasm and can enzymatically ribosylate eEF2. DT by contrast binds to the epidermal growth factor like growth factor precursor (HB-EGF) and is cleaved on the cell surface before uptake through receptor mediated endocytosis. Once in the early endosome the DT catalytic fragment is not processed and penetrates the membrane of the endosome to pass directly into the host cell cytoplasm where it can ribosylate eEF2.

Iota toxin, from Clostridium perfringens (Stiles & Wilkins, 1986), and Vip2 from Bacillus cereus (Han et al, 1999) are both actin ADPRTs, each ribosylating actin at an exposed arginine, Arg177 (Aktories, 1990). The ADP-ribosylation prevents actin polymerisation by capping the exposed ends of the actin filaments which leads to cell rounding and eventual cell death as the actin cytoskeleton breaks down (Aktories & Hall, 1989). This class of actin modifying binary toxins also includes Clostridium botulinum C2 toxin (Aktories et al, 1986), Clostridium spiroforme toxin (Simpson et al, 1989) and Clostridium difficile toxin components cdtA and cdtB (Popoff & Boquet, 1988). The domain structure of Iota and Vip2 is also of interest as the 2 domains resemble one another closely. The second domain is responsible for cell binding and lacks catalytic activity suggesting that the binary toxins may have arisen from gene duplication of an original ADPRT ancestor (Han et al, 1999). The toxins do not bind cells as complete A-B units. Instead proteolytically activated B monomers bind to cell surface receptors as homo-heptamers. These homo-heptamers then bind the A domains and are taken into cells via endocytosis. Once inside acidic
endosomes the low pH activates the translocation function of the B domain heptamers and they translocate the catalytic A domains across the endosomal membrane into the cytoplasm where they can act to ribosylate actin and bring about cell death (Barth et al, 2004).

The C3 exoenzymes are characterised by C3bot first identified from Clostridium botulinum (Aktories et al, 1987) but also include representatives from Clostridium limosum (C3lim) (Just et al, 1992), Bacillus cereus (C3cer) (Bohmer et al, 1996) and Staphylococcus aureus (C3EDIN, C3Stau2) (Wilde et al, 2002c; Yamaguchi et al, 2001). This family of ADPRTs selectively ribosylates the small GTPases, Rho A, B and C (Chardin et al, 1989) at an exposed Arg 41 (Sekine et al, 1989). This reaction is highly specific to only those substrates, except in the case of C3stau2 which has a slightly broader specificity that includes RhoE and Rnd3 (Etienne-Manneville & Hall, 2002; Wilde et al, 2002b). The ADP-ribosylation prevents Rho moving into its active GTP bound state and leads to a loss of control in the downstream pathways controlled by the Rho GTPases and resulting in loss of control of the cell cytoskeleton and eventual cell death (Wilde & Aktories, 2001). Although these affects are in seen in vitro, the role of C3bot and its related ADPRTs in pathogenesis is not yet known as they lack any cell translocation or binding domains. C3stau, however, has been found in some clinical isolates and both C3bot and C3stau2 have been shown to prevent wound healing in vivo (Aepfelbacher et al, 1997; Czech et al, 2001) suggesting that they may have some role in pathogenesis.

In addition to the selection of bacterial ADPRTs there are ADP-ribosyltransferases (ART) present in eukaryotic organisms. Eukaryotic ADP-ribosylation can be of two forms: (1) poly-ADP ribosylation that is mediated by poly ADP-ribose polymerases (PARPs) and catalyses the transfer of multiple ADP-ribose moieties onto a substrate; and (2) mono-ADP ribosylation that catalyses the transfer of a single ADP-ribose moiety onto a target and is mediated by Ecto-ADP ribosyltransferases (Ecto-ARTs) (Schreiber et al, 2006). The PARP superfamily plays a role in the repair of DNA strand breaks and modulation of chromatin (Schreiber et al, 2006). The structure of
the catalytic domains of chicken PARP-1 and mouse PARP-2 however did
demonstrate structural homology to the active site of diphtheria toxin (Ruf et al, 1996).
The 5 Ecto-ARTs found in mammalian systems, named Ecto-ART1-5, are located in
the extracellular space of mammalian tissues and play a role in cell adhesion and the
immune system. They are closer in structure to the C3-like ADPRTs than to the
PARP family (Mueller-Dieckmann et al, 2002).

1.3 Structural analysis of the NAD binding site between ADPRTs

It has previously been demonstrated that all the ADPRTs, both bacterial and the
eukaryotic Ecto-ARTs, share a common active site and NAD binding motif
(Domenighini & Rappuoli, 1996; Han & Tainer, 2002). With the determination of
several ADPRT structures an examination of the active sites and interactions that are
necessary for NAD binding and ribosyl transfer is possible. The structural analysis of
several ADPRTs has led to them being classified into two groups that share a similar
active site architecture but lack sequence homology. The ‘DT’ group is based on the
active site and NAD binding features of Diptheria toxin (Bell & Eisenberg, 1996)
and also includes Pseudomonas exotoxin A (Allured et al, 1986) and the mammalian
poly-ADP ribosylating proteins (PARP’s) (Marsischky et al, 1995). The ‘CT’ group
is based upon the NAD binding observed in cholera toxin (Zhang et al, 1995) and
includes: the E.coli heat labile toxin (Sixma et al, 1991); pertussis toxin (Stein et al,
1994); C3bot (Han et al, 2001); VIP2 (Han et al, 1999); iota toxin (Tsuge et al, 2003)
and the Ecto-ART (Ritter et al, 2003) family from eukaryotes (Domenighini &
Rappuoli, 1996). With the determination of more ADPRT structures it is now clear
that the CT group should be divided further into those toxins that possess an active
site loop involved in substrate binding (Bell & Eisenberg, 1996) and those that
instead have an α-helix forming part of the NAD binding cleft (Han et al, 2001).

There are five key structural features that have been identified in the ADPRTs
(Domenighini & Rappuoli, 1996). These are: (a) the Q/E-X-E motif centred on the
catalytic glutamate and the glutamate/glutamine responsible for the ribosyltransferase
activity; (b) The *Arom-H/R* motif that contains either a histidine or an arginine that contributes to the NAD binding and maintains the structure of the active site cleft; (c) The *Aromatic-Hydrophobic-S-T-S* motif on a β-strand that stabilises the NAD binding; (d) the *Y-X_{10}-Y* motif in DT and PAETA that fulfils the role of the STS motif in the bacterial ADPRT; and (e) the phosphate nicotinamide (PN) loop that contains *A/G-x-R-x-I* motif and is found in the Iota-like binary toxins and C3bot type ADPRTs. The PN loop is a flexible loop above the NAD binding site that creates a more compact binding site. It also brings into play an essential arginine residue which positions the NAD in a conformation more suitable for the cleavage of the nicotinamide N-glycosidic bond in these toxins (Menetrey *et al*, 2002). The conserved sequences and physical position of these features are shown in Figures 1.4A and B.

### 1.3.1 The ADPRT core fold

All the ADPRTs, both in the DT and the CT family possess a near identical mixed α/β core structure of approximately 100 residues even though there is little sequence homology among many of them. This core structure has the approximate dimensions of 35Å x 40Å x 55Å and possesses the NAD binding site which supports both NAD glycohydrolytic and ribosyltransferase activities. The core is constructed from 2 perpendicular β-sheets with a variable number of α-helices attached to it both above and below the frame of β-sheets. The NAD binding site is positioned in a cleft made between the β-framework and either an α-helix in the case of C3bot, C3stau, Vip2, iota and mammalian Ecto-ART or a variable length active site loop in pertussis, cholera, *E.coli* heat labile toxin, diphtheria and exotoxin A. The latter is thought to be involved in EF2 or G-protein recognition (Bell & Eisenberg, 1996).
Figure 1.3 The NAD binding site of ADPRTs

A

B

C

Glu 148

Arg 7

Phe 183 Arg 186

Glu 214

Tyr 65

His 21

Tyr 54

Glu 112

Active site loop

Glu 214

Tyr 79 Asn 87 Arg 91
**Figure 1.3A.** Schematic view of the active site cleft of the DT class of toxins highlighting the key catalytic residues and mode of NAD binding (Bell & Eisenberg, 1996). This illustrates the stacking of the nicotinamide ring between the conserved tyrosines, the binding of the conserved His to the O2 of the adenine ribose and the carbonyl oxygen of Y54 and the binding of the conserved catalytic glutamate to the O2 of the nicotinamide ribose. **1.3B:** Schematic view of the NAD binding cleft of the cholera-like class of toxins showing the intramolecular interactions around the active site and the key features (Zhang et al, 1995). The binding of R7 to the oxygens of the NAD when in the active state instead of the carbonyls of R54 and S61 in the inactive form can clearly be seen. An arginine from the active site loop, in its active form, is also involved in binding the phosphate-oxygens and the ribose ring of the adenine.; **1.3C:** The important residues and bonds formed around the NAD binding site by the 4 motifs found in the α-3 type Toxins (Han et al, 2001). The α-3 asparagine and arginine bind the phosphate oxygen holding the NAD in a compact state. This is the role also undertaken by the conserved arginines of the PN-loop and Arg/His motif. The catalytic glutamate and its stabilising bond from the tyrosine of the α-3 motif are also shown.

**1.3.2 The ARTT motif**

The ARTT loop contains the key catalytic glutamate responsible for the catalysis and transfer of the ribose moiety and the Q/E-X-E motif that is found in all members of the CT group (Domenighini & Rappuoli, 1996). The ARTT loop is of variable length comprised of two sharp turns (turn 1 and turn 2) connecting either two β-sheets as in the ARTT loop of iota toxin (Tsuge et al, 2003) and C3bot (Han et al, 2001) that connects β5 and β6. Or a longer loop connecting an α-helix to a β-sheet as in cholera toxin, pertussis toxin and E.coli heat labile toxin. The ARTT loop in the different types of ADPRT can be seen in figure 1.4B. In all the ADPRTs that are members of the CT group there is also a Glu or Gln residue two residues from the catalytic glutamate residue (Han & Tainer, 2002). This second residue is vital for the ribosyltransferase activity of the ADPRT but not necessarily the NAD glycohydrolase activity (Nagahama et al, 2000; Wilde et al, 2002c). In Ecto-ART the equivalent residue, Gln 187, has been implicated in changing the substrate from cell surface to cytoplasmic substrates (Ritter et al, 2003). This substrate selectivity has also been observed through mutational analysis of C3lim where mutation from Gln to Glu altered the ADP-ribosylation target from asparagine to arginine (Vogelsgesang & Aktories, 2006). This Gln/Glu residue may play a role in substrate...
selection. As can be seen from the sequence alignment in figure 1.4A the actin and G-protein modifying proteins that ribosylate an exposed arginine possess a glutamate residue. In the Rho GTPase ribosylating proteins that ribosylate an exposed asparagine it is a conserved glutamine residue.

Another important residue on the ARTT loop is the aromatic group situated on the centre of the loop between turn 1 and turn 2 that is found in C3bot, C3stau, Iota, Vip2 and Ecto-ART. In C3bot this has been shown to be essential for Rho substrate binding (Wilde et al, 2002c) and in the proposed model of C3bot-RhoA recognition (Han et al, 2001) it is a vital determinant of substrate recognition and binding. This mode of action could apply to the other closely related ADPRTs. This aromatic group, proposed to be important in substrate binding, is only present in the actin and Rho GTPase-modifying toxins and Ecto-ART family. The C3, Iota-like and Ecto-ART enzymes have also substituted the active site loop with an α-helix at the NAD binding cleft. In CT and DT the active site loop is involved with substrate selectivity and binding and has been seen to become disordered and undergo conformational change upon binding of NAD in both DT and PAETA (Bell & Eisenberg, 1996; Li et al, 1996).

1.3.3 The STS motif

The STS motif forms part of the β-sheet that comprises part of the NAD binding cavity and follows the pattern: Aromatic-Hydrophobic-S-T-S (Domenighini & Rappuoli, 1996). The STS motif acts as an anchor to hold the NAD binding site together. In C3bot the contribution of the STS motif (Ser 174, Thr 175, Ser 176) is well understood (Han & Tainer). Ser 174, the first ‘S’ of the motif forms hydrogen bond with the catalytic glutamate and a tyrosine residue (Tyr 79) beneath the cleft to hold the glutamate in the correct position to catalyse the cleavage of NAD. This is also seen in the Ecto-ART2 structure with the STS (Ser 147) forming a hydrogen bond with the catalytic glutamate (Glu 189) (Mueller-Dieckmann et al, 2002). Similar interactions are also observed in C3stau2 (Evans et al, 2003) and Vip2 (Han
et al, 1999). Mutation of this serine residue in the *C. botulinum* C2 toxin (Barth *et al*, 1998b) eliminates the transferase activity. However, in diphtheria and iota toxins mutation of this serine residue reduces activity but does not entirely abolish the glycohydrolytic activity (Bell & Eisenberg, 1996; Nagahama *et al*, 2000) suggesting that while the serine of this motif plays an important role in stabilising the catalytic glutamate it is not essential in all ADPRTs. The threonine residue forms additional hydrogen bonds with perpendicular β-strands to stiffen the active site. Ser 176 in C3bot, the second ‘S’ of the motif, forms hydrogen bonds with the loop immediately following the STS β-sheet and also with the glutamine residue of the Gln/Glu-x-Glu motif keeping the ARTT loop and the glutamine in the correct orientation for the transferase reaction (Han *et al*, 2001). In C3stau2, the second serine is replaced with a glutamine residue that binds to the nicotinamide of NAD directly. This is different to other ADPRTs that possess serine where the ‘S’ has a role in forming intramolecular bonds with the catalytic glutamatic acid. Because of this, the STS motif is thought to have a less important role in C3stau than in other related C3-like enzymes (Evans *et al*, 2003).

In the DT group, the STS motif is partially lost (diphtheria) or entirely lost (exotoxin A). In diphtheria toxin the STS motif is replaced by an YTS motif, but both the T and S residues are in similar positions as they are in the other CT toxins and are likely to play a similar role. The tyrosine residue (Tyr 54) is crucial to the diphtheria activity and is one of the two conserved tyrosines essential for NAD binding through aromatic ring π-orbital stacking (Carroll & Collier, 1984; Carroll & Collier, 1987; Carroll & Collier, 1988). This is also the case in exotoxin A, with Tyr470, though it lacks the serine or threonine residue of the YST motif. This ring stacking stabilises the bound NAD and plays a similar role to that of the STS motif in other toxins; that of stabilising and maintaining the structure of the active site (Bell & Eisenberg, 1996; Li *et al*, 1996).
1.3.4 The key Arg/His residue

The conserved Arg/His residue (Domenighini & Rappuoli, 1996) is comprised of an aromatic residue followed by the Arg/His and has been found in all the ADPRTs to date. In the DT family, the motif is Tyr-His while the members of the CT family also include a Val/Leu before the aromatic residue and all have arginine not histidine. In the ADPRTs, the purpose of the Arg/His motif is NAD binding and maintaining the structure of the active site rather than actual involvement in either glycohydrolase or the transferase reaction. Though not directly involved in the catalysis, the presence of the Arg/His motif has been shown to be vital from mutagenesis studies on C3stau2 (Wilde et al., 2002c), LT (Lobet et al., 1991), PT (Burnette et al., 1988; Cieplak et al., 1988), CT (Burnette et al., 1991) and iota toxin (Tsuge et al., 2003) and it has been shown that loss of this arginine abolishes transferase activity either severely (C3stau2) or completely (Iota, LT, CT, and PT) reduces the hydrolase activity as well. The exact role that this Arg/His residue plays in NAD binding varies between the toxins depending on whether they contain an active site loop (e.g. DT, PAETA, CT, LT and PT) or an α-3 helix (e.g. C3-like and Iota like binary toxins). In the “active site loop group” the residue, His21/440 in diptheria/PAETA, Arg 7 in LT and CT and Arg 9 in PT, does not play an important part of binding NAD but instead supports key parts of the active site to position them in the correct orientation to hydrolyse the NAD. In DT, the histidine forms a hydrogen bond (Bell & Eisenberg, 1996) with one of the hydroxyl groups on the adenine ribose ring and more importantly forms a bond with the backbone carbonyl of one of the tyrosine pair in DT (Tyr 54). Through this bond the tyrosine is orientated into the correct orientation to bind the NAD, and can be seen in Figure 1.3A. Until the recent structure of the NAD-bound cholera toxin (O'Neal et al., 2005) a similar active site stabilisation was thought to occur in LT and the related CT and PT toxins (Sixma et al., 1993). In the absence of NAD the catalytic Arg 7 forms hydrogen bonds with Ser 61 of the STS motif and the main chain carbonyl of Arg 54, an arginine that forms electrostatic bonds with both Glu 110 and Glu 112. This network of bonds plays an important role in stabilising the active site. However when the cholera toxin is activated by an ARF
protein the active site loop undergoes a large conformational shift. This results in Arg 54 being unavailable to interact with Arg 7 and has a slight effect on the position Arg 7 enabling it to bind an oxygen from each of the NAD phosphates rather than stabilising Ser 61 as shown in Figure 1.3B. Thus, Arg 54 acts in a manner similar to that seen in the α-3 toxins by binding directly to the NAD phosphates and positioning the NAD in a suitable conformation for hydrolysis. In DT and LT the conserved His/Arg occupy the same spatial position and interact with the backbone carbonyls of analogous residues, Tyr 54 in DT and Ser 61 of LT, that are the first residues of the YST/STS motifs. They also support a network of interactions that maintain the structure of the active site. Upon activation however, the role of the catalytic arginine in CT reverts to the manner seen in the α-3 toxins by binding directly to the NAD. In the α-3 toxins the arginine forms hydrogen bonds with the phosphates of the NAD positioning them in a more compact manner than is found in the “active site loop” ADPRTs. These hydrogen bonds serve two purposes: firstly to improve the binding of NAD to the toxin and secondly to hold the phosphates in a position where they can interact with the nicotinamide amide group NN7 of the nicotinamide mononucleotide moiety (NMN) (Tsuge et al., 2003). This can then assume a ring like conformation that is prevented from moving due to stacking interactions with the aromatic residue on the PN loop (Phe 349 in Iota toxin, Phe 183 in C3bot) and withdraws electrons from the nicotinamide ring amide increasing the susceptibility of the N-glycosidic bond to cleavage (Ritter et al., 2003). Thus the role of the Arg/His in the α3-helix toxins is very different to their structural role in the DT toxins and the NAD-free CT toxins where the Arg/His holds the active site in the correct manner to facilitate NAD binding.

1.3.5 The Tyr-X_{10}-Tyr motif

The toxins of the DT group (diphtheria and exotoxin A) are different from those in CT group in several respects. The first difference is the absence of the Gln/Glu-x-Glu motif (both diphtheria and exotoxin A possess only the catalytic glutamate) and the second is the lack of the STS motif that the toxins of the CT group possess. Instead
the DT group possess a pair of tyrosines that stack above and below the plane of the NAD moiety and contribute to binding via π-orbital interactions as shown in figure 1.3A. The orbital stacking is of vital importance here and has been shown by mutagenesis studies of PAETA with Tyr 470 Phe/Tyr 481 Phe mutants still possessing enzymatic activity (Lukac & Collier, 1988). Aromatic ring stacking may explain the slightly different conformation that NAD possesses in the PAETA and diphtheria toxin structures compared to the structures of ADPRTs from the CT group (Bell & Eisenberg, 1996; Dolan et al, 2000). The aromatic stacking also places the NAD molecule in a position suitable to interact with the catalytic glutamate with the anomeric carbon of ribose exposed to solvent available for nucleophilic attack. Many of the CT group from the binary and C3-like families that possess the PN-loop have an aromatic residue that stack against the nicotinamide ring in a similar manner to the first tyrosine of the Tyr-Xio-Tyr pair.

1.3.6 The PN loop

The PN loop was first identified in the C3bot-NAD structure (Menetrey et al, 2002) and forms an essential part of the NAD binding site apparatus. The PN loop is a flexible loop that occurs ten residues after the STS motif connecting strands β3- β4 and which undergoes a large movement upon NAD binding becoming more ordered in the process. In C3bot the PN loop has two residues that contributes to the binding of NAD; Arg 186 which forms a hydrogen bond with one of the phosphate groups of NAD and Phe 183 which stacks against the nicotinamide ring of NAD as can be seen in figure 1.3C. Mutational analysis of Arg 186 has revealed that it is essential to NAD binding (Menetrey et al, 2002). In the case of Phe 183, the stacking is similar to that of Tyr 65 from diphtheria toxin, Phe 160 from Ecto-ART and Tyr 481 from exotoxin A. In C3stau, the PN loop is present and retains the critical Arg residue (Arg 150) but the aromatic residue is replaced by a leucine. In both iota and Vip2, the PN loop is intact with an Arg residue (Arg 352 and Arg 400 respectively) that forms bonds with an NAD phosphate molecule and an aromatic residue (Phe 349 and Phe 397 respectively) that stacks directly above the nicotinamide ring. In a similar manner to
C3bot, the important nature of the PN loop has been confirmed in iota toxin with mutants of Arg 352 and Phe 349 showing no activity or strongly diminished activity (Tsuge et al., 2003). The PN-loop is found in the α-3 toxins, both the C3-like and Iota-like, and there are no analogous arginines or aromatic residues present among the members of the CT group that possess active site loops. The DT group possesses a conserved tyrosine in a similar position to the aromatic residue of the PN loop’s $\textit{Arom-X}_2$-$R$ motif, but lacks the conserved arginine that is involved in NAD binding.
Figure 1.4 The conserved ADPRT binding motifs

<table>
<thead>
<tr>
<th>Motifs</th>
<th>STS</th>
<th>Arg/His</th>
<th>ARTT loop</th>
<th>PN Loop</th>
<th>α-3 Helix</th>
</tr>
</thead>
<tbody>
<tr>
<td>C3bot</td>
<td>YIESTSLM</td>
<td>LFRGD</td>
<td>ISAHAGCLEMMLLP</td>
<td>QAG-EPPI</td>
<td>-X,-I NYLPSN</td>
</tr>
<tr>
<td>C3stau</td>
<td>YSSSTQLV</td>
<td>VYRLL</td>
<td>LTAMPGQYELVLLP</td>
<td>AG-EPPI</td>
<td>-X,-I TPPLSA</td>
</tr>
<tr>
<td>C3cer</td>
<td>YISTSLIL</td>
<td>VYRGD</td>
<td>LTAMQGYELLPLL</td>
<td>G-AKT-PV</td>
<td>-X,-Y EMFAMH</td>
</tr>
<tr>
<td>C3lim</td>
<td>YISTSLV</td>
<td>LFRGD</td>
<td>ISTEKGGLEVLLP</td>
<td>AG-EPPI</td>
<td>-X,-I GPIAN</td>
</tr>
<tr>
<td>Iota toxin</td>
<td>FISTSIG</td>
<td>VYRRS</td>
<td>IPGYAGYEVLLN</td>
<td>AGAK-KKI</td>
<td>-X,-I NYLPSN</td>
</tr>
<tr>
<td>Vip2</td>
<td>YMSSSSL</td>
<td>VYRCW</td>
<td>AIGEASEKILLLD</td>
<td>AGKS-KKI</td>
<td>-X,-I NYLPSQ</td>
</tr>
<tr>
<td>Difficile</td>
<td>FISTSIG</td>
<td>VYRRS</td>
<td>IPGYAGYEVLLN</td>
<td>AGAK-KKI</td>
<td>-X,-I NYLPSN</td>
</tr>
<tr>
<td>C2 toxin</td>
<td>FSSTSLK</td>
<td>AYRRV</td>
<td>FSGEDEQEILLNN</td>
<td>SYSKS-RPI</td>
<td>-X,-I SYLNNN</td>
</tr>
<tr>
<td>Spiroforme</td>
<td>FVSTSIG</td>
<td>VYRRS</td>
<td>IPGYAGYEVLLN</td>
<td>AGAK-KKI</td>
<td>-X,-I NYLPSN</td>
</tr>
<tr>
<td>EctoART</td>
<td>FTSSSSLS</td>
<td>VYRT</td>
<td>FSEPDQEEVLIP</td>
<td>-X,-I NYLPSN</td>
<td></td>
</tr>
<tr>
<td>Cholera</td>
<td>YVSTSSIS</td>
<td>LYRAD</td>
<td>YSPHPDEQEVSL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E.coli LT</td>
<td>YVSTSSLS</td>
<td>LYRAD</td>
<td>YSPHPYEQEVSL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pertussis</td>
<td>FVSTSSS</td>
<td>VYRYD</td>
<td>GALATGSEYLH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diphtheria</td>
<td>GFYSTDN</td>
<td>SYHTG</td>
<td>FAEGSSVRYINN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exotoxin A</td>
<td>GFYIAGD</td>
<td>GYHTG</td>
<td>PEEEGRLSTILG</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1.4A. Sequence alignment of the three classes of ADPRT highlighting the conserved residues that make up each of the motifs. The conserved residues in each motif are shaded in the same colours used in figure 1.4B. Ribbon diagrams of a diphtheria-like (Bell & Eisenberg 1996), cholera-like (Zhang et al 1995) and α-3 (Han et al, 2001) toxins highlighting the important motifs in each molecule. The glutamate containing ARTT loop is highlighted blue with the STS and Arg/His motif in purple and yellow respectively. The active site-loops are shown in red as is the α-3 helix. In green is the PN-loop from the α-3 toxins and the Ty-X<sup>10</sup>-Tyr motif for the DT toxins. In all cases figures were generated using Molscript (Esnouf, 1997).
1.3.7 The α-3 motif

Within the ADPRT family, the actin-binding enzymes and C3-like exoenzymes all lack a 15-residue active site loop that is found in the cholera and diphtheria-like toxins (Bell & Eisenberg, 1996; Sixma et al., 1991). In CT and LT this loop is implicated in G protein binding and occurs around residues 45-58. (O'Neal et al., 2005). In DT and PAETA this active site loop is called L4 and comprises residues 39-48 and 483-490 respectively. In its place the iota and C3-toxins (Evans et al, 2003; Han et al, 2001; Han et al, 1999; Tsuge et al, 2003) have an α-helix that packs tightly against the NAD cleft forming a more compact binding site. Amongst these ADPRTs there are three important residues that appear on this α-helix and are conserved amongst nearly all of the C3-like and iota-like ADPRTs. These are: (a) a tyrosine residue that interacts with the ‘S’ of the STS motif and the catalytic glutamate through hydrogen bonds (b) an asparagine residue and (c) an arginine residue that form part of the adenine ring binding pocket. These three residues form another motif specific to the C3bot-like ADPRTs that possess the α-3 helix instead of the active site loop including the EctoART-2 (Ritter et al, 2003). From sequence alignment the related C3cer and C3lim also seem to have the signature sequence Y-X_{6-7}-N-X_{2}-L-R except in the case of iota and Clostridium difficile toxin where the arginine residue is replaced with an isoleucine. Tyr 79 in C3bot (Han et al, 2001) and Tyr 78 in EctoART (Mueller-Dieckmann et al, 2002) can be seen to form charge interactions with the serine of the STS motif (Ser 174 and Ser 146 respectively) and the catalytic glutamate (Glu 214 and Glu 189 respectively). Mutational analysis on the equivalent tyrosine (Tyr 246) in iota toxin (Tsuge et al, 2003) showed that loss of the tyrosine resulted in reduced glycohydrolytic and transferase activities. This Y-S-E network of interactions may act as a stabilising influence on the catalytic glutamate residue to ensure that it is positioned correctly to bind the NAD molecule and in a suitable charge state to stabilise the positively charged oxocarbenium transition state intermediate. The arginine and asparagine residues are involved in binding to the adenine end of NAD with binding observed in C3bot (Arg 91- hydrophobic packing with the adenine ring and Asn 87 binding to one of the phosphates), C3stau (Arg 48 -
binding to the adenine ring), Vip2 (Arg 315) and Ecto-ART2 (Asn 87 and Arg 91). In iota toxin although the arginine at position 259 is replaced with an Ile, a downstream asparagine residue (Asn 255) has been shown to bind to the phosphates at the adenine end of the molecule. This bond contributed by the α-3 helix can be observed in figure 1.3C. Mutation of this asparagine to alanine caused a significant drop in enzymatic activity (Tsuge et al, 2003). This binding to the adenine moiety may help NAD binding and assist in holding the ADP-ribose+ after cleavage of the N-glycosidic bond until the transferase reaction can take place. The interactions between the Asn and the phosphates may contribute to positioning the phosphates closer to the NMN ring in a similar manner to the arginines of the Arg/His motif and PN-loop.

1.4 Catalytic model and mechanism

The conserved nature of the NAD binding site and key catalytic residues would suggest a common catalytic mechanism shared between all members of the ADPRT family. The nucleophilic attack occurs at the anomeric carbon of the nicotinamide ribose and results in the cleavage of N-glycosidic bond separating the ADP-ribose moiety from the nicotinamide ring. The manner in which this occurs is still not precisely understood. Mechanisms have been put forward for $S_{N1}$ and $S_{N2}$ type reactions though recent biochemical data would indicate that the reaction is of an $S_{N1}$ type.

The $S_{N2}$ reaction was first suggested for diphtheria toxin (Bell & Eisenberg, 1996) and has also been put forward for Vip2 (Han et al, 1999), C3bot (Han et al, 2001), pertussis toxin (Locht & Antoine, 1995) and Ecto-ART2 (Ritter et al, 2003). In the $S_{N2}$ reaction the attacking nucleophile that may be the substrate arginine, diphthamidine or asparagine depending on the toxin involved or even water in autohydrolysis. This nucleophile is then deprotonated by either the conserved glutamate in diphtheria toxin (Bell & Eisenberg, 1996; Wilson et al, 1990), or the downstream Gln/Glu-X-Glu in the suggested mechanisms for Iota, C3bot and Vip2 or in pertussis toxin by a catalytic His 35 (Locht & Antoine, 1995). This activated nucleophile then
attacks the anomeric carbon of the ribose ring which due to the conformation of the NAD, has been exposed to the solvent forming a pentacoordinate oxocarbenium transition state intermediate. In the CT group this intermediate is partially stabilised by the catalytic glutamate forming a hydrogen bond with the O2 atom on the nicotinamide ribose. This makes the ring more electronegative which stabalises the positively charged oxocarbenium ion before the N-glycosidic bond is cleaved completing the transfer of the ADP-ribose$^+$ moiety on to the substrate. In the "α3-helix" toxins, the nucleophilic attack may also be aided by the interactions between the nicotinamide amide group and phosphate oxygens that withdraw electrons from the nicotinamide ring making the N-glycosidic bond even more attractive to the attacking nucleophile.

An S$_{N}1$ reaction has also been proposed for ADPRT catalysis (Parikh & Schramm, 2004), that has an isolated positively charged oxocarbenium intermediate stabilised in a pentacoordinate state with direct stabilising electrostatic interactions from the catalytic glutamate and serine hydroxyl group. This oxocarbenium intermediate comes about after the breaking of the glycohydrolytic bond. The serine residue in this case is conserved and found in the STS motif. In the S$_{N}1$ reaction mechanism the ADP-ribose$^+$ stays bound to the protein until attack from the nucleophilic substrate. Of the two mechanisms the S$_{N}1$ reaction seems to be favoured by the available biochemical data, especially the uncoupling of glycohydrolytic and transferase reactions. This was achieved through mutagenesis of the Glu/Gln residue (Glu 338 in iota or Gln 212 in C3bot). The DT group only possesses the catalytic glutamate that in the S$_{N}1$ reaction is required to stabilise the oxocarbenium intermediate. However, members of his group possess the active site loop that is thought to be involved in substrate binding and could position and activate the diphthamide group for nucleophilic attack on the stabilised oxocarbenium intermediate. Recently, the idea of a compulsory order S$_{N}1$ reaction with the NAD binding being required for substrate binding has been challenged by new findings that demonstrate the interaction between PAETA and eEF2 follows a random order reaction mechanism and that either NAD or eEF2 can bind and the reaction takes place via a random third order
SN1 reaction (Armstrong et al, 2002). This has also been observed through kinetic isotope exchange studies (Parikh & Schramm, 2004) which supports a random order SN1 type reaction. A more recent development supporting the available biochemical data is the determination of the structure of the eEF2-TAD-PAETA complex that captures the ADPRT reaction at an intermediate stage (Jorgensen et al, 2005). Though this complex supports eEF2 and PAETA react via an SN1 mechanism, the exact nature of the reaction is still unknown. In the complex structure the distance between the diphthamide N3 and the anomeric carbon was much larger than 2.8Å as suggested by the kinetic isotope exchange work inferring that a crucial step in the mechanism has still not been captured.

1.5 Summary

The conserved ADPRT core structure with a network of perpendicular β-strands is present in all of the ADPRTs that have had their structure determined to date. This is also true of the catalytic glutamate that is found in both the DT and the CT families. This residue occurs in an equivalent position in all the structures determined to date where it is required to stabilise the positively charged oxocarbenium ion transition state. The conserved structural features and conserved areas of sequence are highlighted in figure 1.4A and 1.4B.

The division between the CT group and DT group is not as clear as once thought. Although members of the DT group lack the Q/E-x-E motif, and in the case of PAETA the STS motif, they possess a unique Y-X10-Y motif and, importantly, contain an active site loop that is responsible for substrate recognition and binding. This "active site loop" makes them similar in both structure and function to the members of the CT group that have active site loops. The CT group should be split into two groups; (a) the ‘active site loop’ toxins including cholera toxin, pertussis toxin, E.coli heat labile toxin and (b) the ‘α-3’ toxins that possess an α-helix beside the active site rather than a loop. These include the Iota-like binary toxins, the C3-like toxins and the eukaryotic Ecto-ART family. The α3-toxins differentiate themselves
from the ‘active site loop’ toxins in the PN-loop and the α-3 helix as well as the behaviour of the Arg/His motif.

The recent kinetic and biochemical studies and eEF2-PAETA complex, (Jorgensen et al, 2005) have confirmed that for the DT group, the reaction is a third order S_N1 reaction. It seems likely that with the similarity in the mode of transferase action that the other ADPRTs will follow a similar reaction mechanism relying on the “active site loop” or the Gln/Glu-X-E motif to differentiate between their substrates. The “active site loop” or the Gln/Glu-X-E motif can also select the nature of the activated nucleophile that attacks the ADP-ribose^+ moiety from toxin to toxin. This is supported by the findings that actin, eEF2, G_{ai}, G_{as} and the Rho GTPases all have target residues that are analogous to diphthamide and a nearby aspartate/glutamate residue similar to one of the key ADPRT binding residues seen in the eEF2-PAETA complex (Asp 696 of EF2). This aspartate interacts with the 2-OH of ADP-ribose in a manner similar to that of the catalytic glutamate of the ADPRT and this could be essential to the mechanism and form part of the mechanism that has not yet been captured in a crystallographic structure. This could be part of a conserved recognition site for all the ADPRT substrates. But to confirm that the CT and α-3 toxins do have similar mechanisms further structural studies with their substrates must be performed.
2. The structure of the C3bot-RalA complex at 2.66Å

2.1. Aims of structural studies on the C3bot-RalA complex

The C3bot ADPRT from *Clostridium botulinum* has high specificity to the RhoA-C small low molecular weight GTPases. It has been shown recently however that C3bot, along with the related C3lim and C3cer also interact with the small low molecular weight GTPase RalA. This interaction does not result in the ribosylation of RalA and displays several other unusual properties. We aim to determine the structure of a complex between C3bot and RalA to allow us to understand how this interaction comes about and how this gives rise to the observed biochemical results.

2.2.1 The C3bot exoenzyme from *Clostridium botulinum*

*Clostridium botulinum* produces a large number of toxins including the binary ADPRT C2 toxin (Aktories *et al*, 1986) and the seven neurotoxic botulinum toxins A-G. Early research indicated that the *Clostridium botulinum* neurotoxins C1 and D also possessed cytotoxic ADPRT activity (Ohashi & Narumiya, 1987) targeting small 21-26kDa GTPases. However this was proved to be incorrect as a low molecular weight ADP-ribosyltransferase distinct from the C1 and D neurotoxins (Rosener *et al*, 1987) or the binary C2 toxin (Aktories *et al*, 1987) was identified and titled C3bot. Studies on C3bot identified a group of small 21-24 kDa GTPases as targets for the ribosyltransferase activity (Rosener *et al*, 1987; Aktories & Frevert, 1987) and this GTPase was confirmed as the Rho GTPase (Narumiya *et al*, 1988; Kikuchi *et al*, 1988). Like the binary C2 toxin (Aktories *et al*, 1986) and Iota toxin (Stiles & Wilkins, 1986) the C3bot ADPRT had NAD glycohydrolytic activity as well as a ribosyltransferase activity. The $K_M$ of these interactions was determined as a 2μM for the ADP-ribosylation (Aktories *et al*, 1988) and 60μM for the glycohydrolase activity (Aktories *et al*, 1989).
Substrate studies of C3bot confirmed that RhoA-C are the substrates of C3bot and not the closely related GTPases Rac, Ras or Cdc42 (Aktories & Hall, 1989; Sekine et al, 1989). The ribosylation on Rho was also identified as an asparagine residue (Asn 41) lying in the important switch I region of the GTPase as the residue that is targeted by the transferase reaction (Sekine et al, 1989).

2.2.2 Related C3-like exoenzymes

There are now 5 other C3-like ADPRTs identified from other bacterial species. These are C3cer from Bacillus cereus (Just et al, 1995), C3lim from Clostridium limosum (Just et al, 1992) and the 3 isoforms isolated from Staphylococcus aureus, C3stau1-3. The C3stau exoenzymes may also be referred to as C3EDIN (Epidermal differentiation inhibitor) (Czech et al, 2001; Wilde et al, 2001). C3lim is about 63% identical to C3bot while C3cer and C3stau1-3 share ~30% identity with C3bot.

In the case of C3cer and C3lim it is apparent that they function in a similar manner to C3bot and ADP-ribosylate the same residue as C3bot on Rho (Just et al, 1992). C3stau has demonstrated the same effects on RhoA-C but also extends its substrate specificity to the related GTPases RhoE and Rnd3 (Wilde et al, 2002c).

2.2.3 Biological role of C3bot

The biological targets of the C3 family of ADPRTs are the small low molecular weight Rho GTPases. The Rho GTPase family is a large family comprising of more than twenty related proteins. This includes the Rho proteins (Rho A, B, C, D, E, G, F, G and H), Rac proteins (Rac 1, 1b, 2, 3), Rnd 1-3 and Cdc42 (Wilde & Aktories, 2001; Wennnerberg & Der, 2004). These GTPases function as molecular switches in eukaryotic cells cycling between their inactive GDP bound state and their active GTP bound state (Wilde & Aktories, 2001). Cycling between the inactive and active forms is tightly controlled by 3 main classes of regulating proteins: (1) guanine
nucleotide dissociation inhibitors (GDIs) that sequester inactive GTPase in the cytosol away from the plasma membrane; (2) guanine nucleotide exchange factors (GEFs) that initiate the exchange of GDP to GTP; and (3) GTPase activating proteins (GAPs) that stimulate the catalytic activity of the GTPases and hydrolyse GTP back to GDP. While there have been only been 3 different GDIs identified, a large number of GAPs (over 70) and GEFs (over 60) have been identified due to the broad range of cellular processes that the Rho GTPases are implicated in. A simplified view of the cell cycle is shown below in figure 2.1. In their active GTP bound state the Rho GTPases interact with effector molecules to pass on their downstream function. There have been over 50 of these effector molecules identified for Rho proteins and these can be scaffold proteins, serine/threonine kinases, tyrosine kinases, lipid kinases, lipases or oxidases (Jaffe & Hall, 2005). The interaction with these effector molecules centres around 2 switch regions that are involved in the nucleotide binding. Both switch regions undergo dramatic conformational shifts upon GDP/GTP transfer. These switch regions are common to all the small-GTPases and are involved in the binding of nucleotides and the co-ordination of Mg\(^{2+}\) that has been found to be essential for GTPase activity. GEFs are also thought to bind to the switch I region during GDP/GTP exchange and are involved in stabilising the switch region during nucleotide transfer (Shimizu et al, 2000). Asparagine 41, the target for the C3 family of ADPRTs lies on switch I and ribosylation of this asparagines residue results in GEFs unable to bind to RhoA and permanent inactivation of RhoA as it is unable to facilitate the transfer of GDP to GTP. Ribosylation however does have a minimal effect on nucleotide binding or GTPase activity, either intrinsic or GAP stimulated. It also does not prevent the Rho protein interaction with effectors although this requires the Rho to be in the activated state prior to ribosylation (Genth et al, 2003a; Genth et al, 2003b).
Figure 2.1 The Rho GTPase cycle

Figure 2.1. The Rho GTPase cycle. Rho starts off (1) sequestered in the cytosol bound to the Rho guanidine dissociation inhibitor (Rho-GDI). It is then bound still in its inactive, GDP, form to the membrane (2) and then with assistance from a guanine exchange factor (GEF) the GDP is exchanged for GTP and Rho is active (3) and able to interact with effectors. It then interacts with a GTPase activating protein that stimulates the GTPase activity of Rho and hydrolyses GTP to GDP inactivating Rho (4). C3bot ribosylates asparagines 41 on switch I of Rho preventing GEFs binding and trapping Rho in its GDP bound inactive state.

This interrupts the essential functions of Rho proteins within the cell. These functions include: regulation of actin cytoskeleton and stress fibres, regulation of NADPH oxidase, cell cycle control, secretion, apoptosis, transcriptional activation, cell adhesion, cell transformation and transcriptional activation (Wilde & Aktories, 2001; Etienne-Manneville & Hall, 2002; Aktories et al, 2004; Jaffe & Hall, 2005).

The most prominent and well studied effect of this is to disrupt the ability of Rho proteins to maintain the cellular cytoskeleton. This results in well characterised cell rounding, redistribution of actin filaments and the breakdown of stress fibres (Chardin et al, 1989). Other well studied processes that are inhibited by ribosylation
by C3bot are related to the immune system and immune response and may represent a pathogenic purpose to C3 activity. Areas of the immune system affected include: cell adhesion (Nemoto et al, 1996); lymphocyte cytotoxicity (Lang et al, 1992); and the migration and invasion of both leukocytes and lymphocytes (Stam et al, 1998; Laudanna et al, 1996).

The great specificity of C3bot and the related C3s has been used as a pharmacological tool in many experiments studying the various effects of the Rho proteins on the cell cycle, skeleton and gene transcription. However in these studies the lack of any translocation or receptor binding domains on C3bot results in these experiments relying on either microinjection (Paterson et al, 1990) or the construction of chimeric toxins. The approach with chimeric toxins has proved successful and in all cases improved the potency of the C3 exoenzyme involved by several hundredfold. The reported chimeric toxins have involved fusing a C3 toxin to the binding and translocation domain of either diphtheria toxin (Aullo et al, 1993) or to the binding and translocation domain of Clostridial C2 binary toxin (Barth et al, 1998a). More recently new techniques have been used that rely on attaching a short peptide chain to the C-terminus of C3bot to encourage uptake into target cells. A short sequence of the HIV Tat gene (Park et al, 2003) or a short proline or arginine rich sequence (Winton et al, 2002) have both been used in this way.

The lack of a translocation or binding domain has left the exact role of C3bot in pathogenicity largely unknown. As despite the wide ranging effects that C3bot action has on cells there is no obvious method for C3bot to gain entry into its host cells. By examining the behaviour of other pathogens there are two possibilities that may explain the single domain nature of C3bot: (1) More pathogenic bacteria than first though have now been observed to invade host cells. If this is the case for Clostridial species then C3bot could be secreted directly into the cell cytosol (Wilde & Aktories, 2001). (2) It has been observed in Streptococcus pyogenes that the cholesterol dependant streptolysin O can form a pore to translocate S. pyogenes NAD-glycohydrolase (SPN) into host cells (Madden et al, 2001). Pore forming proteins are
common to many C3 producing bacteria that may exhibit a similar mechanism. This method of entry is also fairly reminiscent of the type III secretion system used by many species of pathogens.

2.2.4 The structure of C3bot

C3bot is a 251 amino acid protein though the first 40 residues have been identified as a general secretory protein signal peptide. This signal sequence possesses a series of positively charged amino acids at the extreme N-terminus (Lys2, Arg5 and Arg6), a hydrophobic central region (Ile8 to Ile26) and a proteolytic cleavage site between Lys40 and Ala41 in keeping with the general features of a general secretory protein (Popoff et al, 1991; von Heijne, 1985).

Figure 2.2 The structure of C3bot

![Figure 2.2. The structure of C3bot [PDB code 1G24] (Han et al, 2001) showing the mixed α/β core and highlighting the active site and the 2 important loop regions.](image)
C3bot also shares an active site with the other members of the C3 family and a common 3-dimensional structure. The structure of C3bot showed a central mixed α/β core with a 5 stranded anti-parallel β-sheet stacking perpendicular against a 3 stranded anti-parallel β-sheet and is shown above in figure 3.2. The 3 stranded sheet is flanked by four consecutive α-helices below the active site and a fifth α-helix stacks against the 5 stranded sheet above the active site. The active site of C3bot is similar to that of the other bacterial ADPRTs (Holbourn et al, 2006) and it possesses the 5 motifs common to the family of α-3 ADPRTs: (1) an ARTT loop containing the catalytic glutamate (Glu 214) and substrate determining glutamine (Gln 212) two residues upstream. In C3lim mutation of this glutamine to glutamate has been shown to change the specificity from asparagine to arginine (Vogelsgesang & Aktories, 2006). (2) An STS (Ser 174-Thr175-Ser176) motif to stabilise the active site and form a network of hydrogen bonds between the catalytic glutamate and Tyr79 of the α-3 helix. (3) The PN loop (residues 182-186) that moves in a “crab-claw” motion to close the active site around the bound NAD. The PN loop contains a phenylalanine (Phe183) that stabilise the nicotinamide with stacking interactions and an arginine (Arg 186) that binds to the phosphate oxygen (Evans et al, 2003) (Menetrey et al, 2002). (4) The conserved arginine (Arg128) from the conserved Arg/His motif found in all of the ADPRT family forms another stabilising interaction with one of the phosphate oxygens. A third arginine (Arg91) on the (5) α-3 helix stacks against the adenine ring. Helix α-3 also possesses tyrosine (Tyr79) and an aspargine (Asn87) that forms another bond with another of the phosphate oxygens.

2.3 Interactions between C3bot and RalA

While C3bot specifically binds to and ADP-ribosylates Rho -A,-B,-C GTPases it has recently been demonstrated that it also binds Ras-like GTPase RalA. This occurs without ribosylation of RalA and inhibits the ADP-ribosylation of Rho by C3bot (Wilde et al, 2002a). C3lim and C3cer were also found to bind to RalA in a manner that does not result in the ribosylation of RalA and inhibits the activity of the C3 exoenzyme involved, though to a lesser extent than the inhibition observed with
C3bot. C3stau2 although sharing 35% identity with C3bot does not bind, and is not inhibited by, RaIA.

Ral (Ras-like) GTPases were originally identified in simian B-cell line cDNA library due to their similarity to Ras and others in the GTPase family (Chardin et al, 1989). Ral occurs in two isoforms, A and B, that have a high sequence identity of ~80%. Related proteins in the Ras superfamily include several of the Ras proto-oncogenes including H-, N-, and K-Ras that share ~58% identify to Ral (Feig, 2003) and 33% identity to RhoA (Wilde et al, 2002a).

RalA like the other GTPases functions as a molecular switch by cycling between a GTP and GDP bound form. RaIA is inactive in the GDP bound form and active in the GTP bound form and the binding and exchange of GTP is regulated by guanine nucleotide exchange factors (GEFs), guanine nucleotide dissociation inhibitors (GDI's) and GTPase activating proteins (GAPs). In the case of RaIA these include a subset of specific GEFs and GAPs specific to RaI alone (Feig, 2003). RaIA binds to GDP very strongly and requires the assistance of GEFs to remove GDP and to allow GTP to bind in its place. The GTP-bound RaIA is then active and through post-translationally modified residues with palmitoyl, farnesyl or geranylgeranyl groups interacts with downstream effectors. The GDIs prevent the dissociation of GDP and act as a negative control on the activation of RaIA. GAPs are required as the GTPase activity of the small GTPases is quite slow and the GAPs speed up the process in returning the active GTP-bound form to the inactive GDP-bound form (Takai et al, 2001). RaIA, like the related Ras, Rho and Rac GTPases are regulated by GTP/GDP transfer. This exchange occurs around the nucleotide-binding region that is made up of two sequences defined as switch I and switch II. These switches have been identified in RaIA; switch I residues 40-48 and switch II 70-78 (Fukai et al, 2003). Both switch regions undergo dramatic conformational shifts upon GDP/GTP transfer. These switch regions are common to all the small-GTPases and are involved in the binding of nucleotides and the co-ordination of Mg$^{2+}$ and are shown below in figure 2.3. GEFs are also thought to bind to the switch I region during GDP/GTP exchange.
and are involved in stabilising the switch region during nucleotide transfer (Shimizu et al, 2000).

**Figure 2.3 The structure of RalA**

![Figure 2.3 The structure of RalA highlighting the important switch I and switch II regions and the GDP binding site [PDB code 2BOV] (Holbourn et al, 2005).](image)

RalA like all GTPases has many functions in the cell and can mediate a plethora of different cellular processes. RalA is primarily activated by Ras. Ras binds to and activates a family of Ral specific GEFs including the Ral-GDP disassociation stimulator (Ral-GDS) and transfers them to where they may interact with Ral at the plasma membrane (Feig, 2003). Ras may also act on another protein, PI3K, which enhances the Ral GEFs intrinsic GEF activity to switch on Ral (Tian et al, 2002; Feig, 2003). Additionally Ral may be activated directly by intracellular secondary messengers such as Ca$^{2+}$ (Hofer et al, 1998). Ral plays a role in many important pathways, Some of these diverse signal processes are illustrated below in figure 2.4.
Figure 2.4 The role of Ral in the cell

One of the most well studied roles of Ral is in vesicle trafficking in the plasma membrane (Nakashima et al, 1999; Brymora et al, 2001). This involves cycling between the control of both secretion through the exocyst complex and indirect control of receptor mediated endocytosis through regulation of RalBP1 and through stimulation of phospholipase D1 (PLD1) (Jiang et al, 1995; Feig, 2003). This includes a regulatory role in neurotransmitter release through the control of vesicle secretion (Polzin et al, 2002).
It also plays a substantial role through several pathways in control of the cytoskeleton and alterations in cell morphology such as the formation of filopodia and the cytoskeletal changes associated with cell migration (Ohta et al, 1999; Takai et al, 2001). These are controlled by either direct interaction of Ral with filamin or the exocyst complex (Feig, 2003) or via RalBLP1s GAP activity on Cdc42 and Rac GTPases (Jullien-Flores et al, 1995).

A still not fully understood role of Ral is the control of gene transcription. It has been shown to activate a host of transcription factors via interaction with an unknown effector. These transcription factors include nuclear factor-κB (Henry et al, 2000), a tyrosine kinase (Src) (Feig, 2003), the ternary complex factor (TCF) (Wolthuis et al, 1997) and the forkhead transcription factor (AFX) (Kops et al, 1999).

Finally it has been seen to play a role in cell proliferation, promotion of oncogenesis and Ras-mediated cell transformation (Urano et al, 1996; Feig et al, 1996). Its contribution to cell proliferation occurs partly through the activation of PLD1. (Lu et al, 2000) Its role in cell proliferation has now been widened to include cell differentiation as well. (Verheijen et al, 1999).

The interaction between C3bot and RalA is surprising given the well documented specificity of C3bot for RhoA-C and its use as a pharmacological tool in studying the Rho GTPases and their role in the cell. There are several major questions raised by the observed interaction. The interaction although resulting in inhibition of ribosylation of RhoA, increases the rate of NAD-glycohydrolysis five fold (Wilde et al, 2002a) and does not result in the ribosylation of RalA. Another ADP-ribosylating enzyme, ExoS, the Pseudomonas auruginosa exoenzyme, is known to ribosylates RalA at Arg52- the equivalent residue to Asn41 on RhoA (Fraylick et al, 2002). Further initial studies demonstrated that a mutation of the aromatic residue on the ARTT loop of C3bot (Phe209) had no effect on the interaction with RalA suggesting that this binding occurred in a novel manner. This binding is not unique to C3bot as both the related C3 exoenzymes C3lim and C3cer interact with RalA and suffer
inhibition of their ribosylation activity though less than the inhibition shown for the C3bot-RalA interactions, though in the case of C3lim the interaction has been shown to be stronger (K_d ~12nM) (Wilde et al, 2002a) than that of the C3bot-RalA interaction. (K_d ~60nM) (Pautsch et al, 2005). This inhibition is also specific to the RalA GTPase as there is no similar inhibition reported for the related Rac, Ras or CD42 small GTPases and has several potentially important functional consequences for RalA. The first of these is that the interaction between C3bot and RalA seems to prevent the activation of PLD1 by RalA which would have serious affect on the ability of RalA to maintain its normal cellular function (Wilde et al, 2002a). Secondly, RalA, like many small GTPases, is glucosylated by Clostridium sordelli lethal toxin at Thr46 on the switch I region (Genth et al, 1996) and interaction with C3bot prevents this interaction giving a clue as to where the interaction between C3bot and RalA may take place. There are also conflicting reports of the nucleotide dependant state of the interaction with it being reported as GDP/GTP independent (Wilde et al, 2002a) and GDP dependant (Pautsch et al, 2005).

We have determined the structure of complex between C3bot-RalA at 2.66Å (Holbourn et al, 2005). This complex answers many of the questions raised by the biochemical studies and accounts for the broad C3 specificity (all C3 exoenzymes apart from C3stau) but narrow GTPase specificity (only RalA).
2.4 Methods

The expression and purification was based on methods previously described by Evans et al (Evans et al, 2004) and myself (Holbourn et al, 2005).

2.4.1 Expression of C3bot

A C3bot-MBP fusion construct in BL21(DE3) cells was kindly provided by Dr. J.M. Sutton (Health Protection Agency UK). A 10ml starter culture of the C3bot-MBP construct was grown overnight at 37°C in Terrific Broth supplemented with 35mg/l chloramphenicol and 0.5% (w/v) glucose. The overnight culture was diluted 1:50 in fresh media and grown for four hours until an OD$_{600}$ of ~0.8 was reached. Protein expression was induced with a final concentration of 500μM isopropyl-β-D-thiogalactopyranoside (IPTG) for four hours at 25°C. Cells were harvested by centrifugation (Beckman Avanti J-25).

2.4.2 Purification of C3bot

Cell pellet was resuspended in 20mM Hepes pH 7.3, 20mM NaCl, 1mM Ethylenediaminetetraacetic acid (EDTA), 1mM PMSF, 1mg/ml lysozyme and stirred on ice for 45 minutes. The cells were lysed by sonication (Soniprep 150) and the cell lysate was clarified by centrifugation (Beckman Avanti J-25).

The clarified cell lysate was applied to an SP-sepharose column (GE Healthcare) equilibrated with 20mM Hepes pH 7.3 and was eluted on an ascending NaCl gradient as a single peak. Peak fractions were pooled with 1U Factor Xa (New England Biolabs) /100μg protein overnight at room temperature. The cleaved C3bot was reapplied to an SP-sepharose column (GE Healthcare) equilibrated with 20mM Hepes pH 7.3 and was eluted with an ascending NaCl gradient. The protein was concentrated in a 10kDa centrifugal concentrator (Amicon) and applied to a Superdex 200 gel filtration column (GE Healthcare) equilibrated in 20mM Hepes pH 7.3,
120mM NaCl. The purified protein was concentrated down to 10mg/ml in a 10kDa centrifugal concentrator (Amicon) and stored for crystallisation.

2.4.3 Expression of RalA

The human *RalA* gene clone (as a pGEX-2T fusion) was a kind gift of Dr Hiroshi Koide, Department of Stem Cell Biology, Graduate School of Medical Science, Kanazawa University, Japan. The plasmid was then transferred to the host strain JM109 for expression. The clone was inoculated into 100ml of Terrific Broth supplemented with 0.5% glucose and 100μg/ml ampicillin and grown overnight at 37°C. The culture was diluted 1:50 into 1l of fresh medium and grown for 4 hours at 37°C until an OD$_{600}$ of ~0.8 was reached. Expression was induced by addition of IPTG to a final concentration of 1mM and grown for a further 4 hours at 25°C. Cells were harvested by centrifugation (Beckman Avanti J-25).

2.4.4 Purification of RalA

Cell pellets were resuspended in 50mM BisTris, 0.5M NaCl pH7.0 1mg/ml lysozyme and left on ice for 45 minutes. The cells were lysed by sonication (Soniprep 150) and the cell debris removed by centrifugation (Beckman Avanti J-25).

The clarified cell lysate was applied to a GST-trap column (GE Healthcare) equilibrated with 50mM BisTris pH 7.0, 0.5M NaCl, and was eluted with 20mM Tris pH 8.0, 0.2M NaCl, 10mM reduced glutathione as a single peak. The peak was dialysed into 20mM Hepes pH 7.3, 150mM NaCl, 5mM MgCl$_2$, 2.5mM CaCl$_2$ overnight at 4°C. It was then treated with 1U thrombin (New England Biolabs) /100μg protein overnight at 4°C. RalA was separated from the GST fusion tag by reapplication to the GST-trap column. The RalA protein was concentrated in a 10kDa centrifugal concentrator (Amicon) and applied to a Superdex 200 gel filtration column (GE Healthcare) equilibrated with 20mM Hepes pH 7.3, 20mM NaCl, 5mM
MgCl₂, 1mM EDTA. The purified sample was further concentrated to 10mg/ml in a 10kDa centrifugal concentrator (Amicon) and stored for crystallisation.

**2.4.5 Formation of the C3bot-RalA complex**

Pure C3bot and RalA were incubated together at 4°C overnight in a range of ratios ranging from 2:1 to 1:2. The formation of a complex after incubation was observed on a continuous 15% acrylamide, 30mM histidine-30mM Mes pH 5.9 native gel. In all cases the gel was pre-electrolysed for 45 minutes at 120V. Samples were loaded as a mixture of 2µl methyl green and 2µl of the C3bot-RalA.

**2.4.6 Crystallisation of the C3bot-RalA complex**

For crystallisation RalA and C3bot were incubated together overnight in a ratio of 2:1 giving a working concentration of 6.67mg/ml RalA and 3.33mg/ml C3bot. Crystallisation trials were performed as hanging drops using 2µl of protein with 2µl of reservoir solution in the drop, giving a final protein concentration of 3.3mg/ml RalA and 1.67mg/ml C3bot in the crystallisation drops. Structure screens I and II (Hamptons Research) were used to generate initial crystallisation conditions. “Whisker-like” crystals were observed in condition 19 of structure screen II (Hamptons Research) (0.1M Hepes pH 7.5, 10% PEG8000 and 8% ethylene glycol).

Fine screening around this condition led to narrow crystals appearing over the course of ~3 days in 0.1M Hepes pH 7.3, 12-15% ethylene glycol, 14-18% PEG8000 at 16°C.

**2.4.7 Data collection and refinement**

Diffraction data to 2.66Å were collected from a single cryo-cooled (100K) crystal with 25% ethylene glycol on station PX9.6 of the synchrotron radiation source at Daresbury (UK) using an ADSC Quantum 4 CCD detector and one of the diffraction
images is shown below in figure 2.5. The data were processed and scaled using the HKL2000 package (Otwinowski, 1993a; Otwinowski, 1993b) (Otwinowski, 1993a). The symmetry and systematic absences were consistent with the P2_12_12 space group (unit cell dimensions \(a=56.63\AA\), \(b=90.84\AA\) and \(c=100.40\AA\)) with one C3bot1-RalA complex per asymmetric unit and the crystals containing ~54% solvent. Data reduction was carried out using the program TRUNCATE (CCP4, 1994).

The structure of the C3bot1-RalA complex was determined by molecular replacement method with MOLREP (Vagin & Teplyakov, 2000) using PDB entries 1G24 for native C3bot1 (Han et al, 2001) and 1UAD for RalA from the RalA-Sec5 complex (Fukai et al, 2003) as search models. Initial refinement of the structure was carried out using REFMAC5 (Murshudov, 1997) which improved the model and subsequent rounds of refinement was carried out using the CNS suite (Brünger et al, 1998) and the model building program O (Jones et al, 1991). In each data set 5% of reflections were kept aside for \(R_{\text{free}}\) calculation (Brünger, 1992). Water molecules were added using the WATERPICK module of CNS. The clear continuous density observed (in both the 2mFo-nFc and mFo-nFc SIGMA-A weighted maps) in the GTP binding pocket of RalA during the final stages of the refinement was interpreted as GDP. The final model contains 3010 protein atoms, 214 water molecules and 24 ligand atoms, has an R-factor of 21.8 and an R-free value of 27.9%. The details of the data processing and refinement statistics are presented in table 2.1. The stereochemical quality of the finished model was checked using PROCHECK (Laskowski et al, 1993).
Figure 2.5 A diffraction pattern from the C3bot-RalA complex

Figure 2.5. Diffraction pattern from a cryo-cooled crystal of the complex between C3bot-RalA at station 9.6 at the SRS Daresbury. The resolution range is shown by a series of concentric circles.
Table 2.1 Details of crystallographic data

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution range (Å)</td>
<td>50-2.66</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.87</td>
</tr>
<tr>
<td>Exposure time (s)</td>
<td>20</td>
</tr>
<tr>
<td>Number of images</td>
<td>150</td>
</tr>
<tr>
<td>Rotation per image (°)</td>
<td>1</td>
</tr>
<tr>
<td>Number of reflections measured</td>
<td>242,860</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>15,556</td>
</tr>
<tr>
<td>I/σ(I) (outer shell(^a))</td>
<td>14.6 (5.8)</td>
</tr>
<tr>
<td>Completeness (outer shell(^a)) %</td>
<td>99.8 (99.9)</td>
</tr>
<tr>
<td>(R_{\text{symm}}) (outer shell(^a)) (%)</td>
<td>11.1 (26.0)</td>
</tr>
<tr>
<td>(R_{\text{cryst}}) (%)</td>
<td>20.6</td>
</tr>
<tr>
<td>(R_{\text{free}}) (%)</td>
<td>27.4</td>
</tr>
<tr>
<td>Average temperature factor (Å(^2))</td>
<td></td>
</tr>
<tr>
<td>RalA main-chain (side-chain)</td>
<td>20.4 (21.9)</td>
</tr>
<tr>
<td>C3 main-chain (side-chain)</td>
<td>21.9 (22.8)</td>
</tr>
<tr>
<td>Ligand</td>
<td>20.0</td>
</tr>
<tr>
<td>Solvent</td>
<td>25.6</td>
</tr>
<tr>
<td>Mg(^{2+}) ion</td>
<td>19.3</td>
</tr>
<tr>
<td>RMSD from ideal values:</td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.006</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.1</td>
</tr>
</tbody>
</table>

\(^{a}\) Outer shell is 2.76Å – 2.66Å

\(^{b}\) \(R_{\text{symm}}=\frac{\sum_h \Sigma |I(h)-\langle I(h)\rangle|}{\sum_h \Sigma I(h)}\), where \(I(h)\) and \(\langle I(h)\rangle\) are the \(i\)th and the mean measurements of the intensity of reflection \(h\), respectively.

\(^{c}\) \(R_{\text{cryst}}=\frac{\sum_h |F_o-F_c|}{\Sigma_h F_o}\), where \(F_o\) and \(F_c\) are the observed and calculated structure factor amplitudes of reflection \(h\), respectively.

\(^{d}\) \(R_{\text{free}}\) is equal to \(R_{\text{cryst}}\) for a randomly selected 917 (5.9%) reflections not used in the refinement.
2.5 Results and discussion

2.5.1 Experimental results

The expression and purification of C3bot and RalA was fairly straightforward as a protocol was already available and needed little modification to yield reasonable amounts of pure homogenous protein, ~12mg/l C3bot and ~3mg/l RalA. SDS-page gels illustrating the various stages of purification are shown below in figures 2.6A and 2.6B.

Figure 2.6 Purification of C3bot and RalA

A

B

Figure 2.6. A The purification of RalA showing the purification of the initial 45kDa RalA-GST fusion protein and the post cleavage the purification of the 24kDa RalA. B Illustrates the purity of C3bot at the various stages throughout purification. The initial 60kDa band is the C3bot-MBP fusion construct. The MBP is then cleaved off and the C3bot isolated with a second ion exchange step and further purified with gel filtration.
To assess the formation of the complex a native gels were used rather than gel filtration as there was a limited amount of material available. On the native gels the C3bot band appears as a clear band while the RalA band is never present. The complex is seen as a third band near the top of the gel and can be seen in figure 2.7.

**Figure 2.7 Native gel of the 2 competing crystal forms**

![Native gel of the 2 competing crystal forms](image)

In the course of crystallisation trials the initial structure screens gave two possible conditions: Condition 6 from Hamptons structure screen 1 (0.1M Na acetate trihydrate, 8% PEG 4000); and condition 19 from structure screen 1 (0.1M Hepes pH 7.5, 10% PEG8000 and 8% ethylene glycol). Prior to data collection samples of both crystals were taken, crushed and applied to a native gel as shown below in figure 2.7. The results of this gel suggested that the first crystal form in 0.1M Na acetate
trihydrate, 8% PEG 4000 was a new crystallisation condition for C3bot. The second condition was that of the C3bot-RalA complex. Optimisation screens around this gave final crystallisation conditions of 0.1M Hepes pH 7.3, 12-15% ethylene glycol, 14-18% PEG8000. These crystals appeared as long thin hollow needles as shown in figure 2.8.

**Figure 2.8 Crystals of the C3bot-RalA complex**

Data to 2.66Å were collected from a single cryo-cooled (100K) crystal on an ADSC-CCD detector on station PX9.6 at the Synchrotron radiation source at Daresbury, UK. The data were processed & scaled using the HKL2000 package (Aepfelbacher et al, 1997). Initial determination of cryo-conditions using 25% ethylene glycol proved satisfactory and 25% ethylene glycol was used as the cryo-protectant in all diffraction studies. The structure of the NAD-free wild type C3bot (PDB code 1G24) (Han et al, 2001) and RalA from the complex with the Sec5 (PDB code 1UAD) (Fukai et al, 2003) were used as search models for molecular replacement using MOLREP (Vagin
MOLREP gave many possible solutions; the final molecular replacement solution had an R-factor of 44.7% and a correlation coefficient of 54.1%. The MOLREP solution chosen was the top solution as the second solution found had an R-factor of 54.6% and a correlation coefficient of 26.5%. Initial refinement using REFMAC (CCP4, 1994) dramatically improved the model (R_crys 29.7, R_free 38.1, FOM 81.3). Subsequent rounds of refinement and model building were performed using CNS (Brünger et al., 1998) and O (Jones et al., 1991). During the course of refinement a large stretch of contiguous difference density was observed in the nucleotide binding site of the RalA molecule and this was modelled as GDP and a single magnesium ion as shown below in figure 2.9.

**Figure 2.9 Electron density surrounding the GDP molecule**

![Figure 2.9. Stereo view of the contiguous 2Fo-Fc density at 3 sigma in the nucleotide binding site with a molecule of GDP modelled into it. The Mg²⁺ ion is shown as a green sphere. The PDB file for GDP and the torsion and parameter files were supplied by the Hicup webserver (Kleywegt & Jones, 1998)](image)

The N and C termini of the RalA molecule were visible but residues 41-44 of the C3bot N-termini were removed due to insufficient density, residues 181-184 on C3bot were disordered and modelled on the basis of weak density with the side chains of residues 181 and 182 and 54 modelled with 0 occupancy due to weak density, residue 210 on C3bot was also modelled with 0 occupancy due to insufficient density. Analysis of the Ramachandran plot (Ψ-Φ), shown below in figure 2.10, showed 86.5% of residues lie within the most favourable region with 0.3% in disallowed regions (Ala184 on C3bot). The details of the data processing and
refinement statistics are presented in table 2.1. The stereochemical quality of the finished model was checked using PROCHECK (Laskowski et al, 1993). The final refined structure of C3bot1-RalA complex at 2.66Å resolution has a conventional R-factor \((R_{\text{cryst}})\) of 20.64% and a free R-factor \((R_{\text{free}})\) of 27.36% and is shown below in figure 2.11. The final model of the complex contains 3010 protein atoms, 260 water molecules and one GDP molecule, with an Mg\(^{2+}\) ion, bound to the nucleotide binding site of RalA (Holbourn et al, 2005).

2.5.2 Overall structure of the C3bot-RalA complex.

The overall topology of C3bot in the complex is identical to that of previously determined structures of native C3bot molecule (Evans et al, 2004; Han et al, 2001; Menetrey et al, 2002). C3bot consists of a mixed \(\alpha/\beta\) core. A five stranded mixed \(\beta\)-sheet \((\beta 1, \beta 2, \beta 4, \beta 7, \beta 8)\) stacks perpendicularly against a three stranded anti-parallel \(\beta\)-sheet \((\beta 3, \beta 5, \beta 6)\) with four helices \((\alpha 1, \alpha 2, \alpha 3, \alpha 4)\) arrayed around the three stranded sheet plus a fifth helix \((\alpha 5)\) flanking the five stranded sheet. A least squares superposition of the C3bot with uncomplexed C3bot (PDB code 1G24; 24) yields a root mean square deviation of 0.79 Å for 207 C\(_{\alpha}\) atoms. The largest difference between the structures concerns the conformation of the ARTT loop. There is also a change in conformation for residues 232-236 between strands \(\beta 7\) and \(\beta 8\) on the surface of the molecule.
Figure 2.10 A Ramachandran plot of the refined structure of the C3bot-RalA complex. One residue lies in the disallowed region (Ala184 of C3bot) and 86.5% of the residues lie within the most favourable region. The Ramachandran plot was calculated with PROCHECK (Laskowski et al, 1993).
Figure 2.11. The structure of the C3bot-RalA complex shown in cartoon form. RalA is shown with yellow helices and green strands and switch I and II are highlighted. The bound GDP molecule is shown bound into the active site and the bound Mg\(^{2+}\) ion is shown in purple. C3bot is shown with red helices and blue strands and the PN loop and ARTT loop at the active site are highlighted.

RalA is constructed from a six stranded mixed \(\beta\)-sheet sandwiched between five \(\alpha\)-helices; three helices (\(\alpha_2, \alpha_3, \alpha_4\)) below the plane of the \(\beta\)-sheet, two helices (\(\alpha_1, \alpha_5\)) above. RalA, like the other related low molecular weight GTPases, possesses two switch regions that form part of the GDP binding pocket and change conformation on GTP or GDP binding. In RalA, the switch regions comprise residues 40-48 (switch I, between helix \(\alpha_1\) and strand \(\beta_2\)) and 70-78 (switch II, between helix \(\alpha_2\) and strand \(\beta_3\)). These regions, which display a wide range of conformations among known GTPase structures, are well ordered in the C3bot-RalA complex structure. The GDP binding site has a total buried surface area of 808.9\(\AA^2\), 9.25\% of the RalA-GDP total surface area (8743.8\(\AA^2\)). The binding of GDP to RalA in the C3bot1-RalA complex is
somewhat different to that of GppNHp (a non-hydrolysable GTP homologue) to RalA in the RalA-Sec5 complex (Fukai et al, 2003). It follows the normal mode of GDP binding as seen in other small G-proteins (Takai et al, 2001) and is shown below in figure 2.12.

**Figure 2.12** The GDP binding site of RalA

![Figure 2.12](image)

**Figure 2.12.** A stereo view of the GDP binding site of RalA highlighting the bonds made between RalA and the bound GDP. The magnesium ion is shown in green and water molecules are show in turquoise. The Mg$^{2+}$ ion has a square pyramidal geometry through Ser28 of RalA and water mediated hydrogen bonds. The guanine ring is stabilised by the conserved NKXD motif with Asn127, Lys128 and Asp130 forming the hydrogen bonds.

The guanine base is recognized by residues from the consensus sequence NKXD (loop between α4 and β5) with Lys128 and Asp130 making direct interaction with the nucleotide. In addition, part of the loop located between helix α5 and strand β6 acts as a recognition site for the guanine base in both structures. The loop (also known as P-loop), connecting strand β1 and helix α1, houses the α- and β- phosphate groups. Binding of GDP requires coordination by the Mg$^{2+}$ ion (square pyramidal geometry) that binds the β- phosphate oxygen. The other coordinating atoms for the
Mg\textsuperscript{2+} are provided by three water molecules and the OG atom of Ser 28 from the α1 helix.

Superposition of RalA molecule from the C3bot-RalA complex with the RalA-Sec5 complex (PDB code 1UAD) (Fukai et al, 2003) shows a root mean square deviation of 1.2 Å for 172 Cα atoms, although this value is biased by different conformation of the switch I and switch II regions (root mean square deviations of 2.6Å and 2.2Å respectively). The striking conformational differences between the RalA switch regions between the Sec5 and the C3bot complexes, in particular switch I, can be attributed to the different nucleotides bound, the GTP analogue, GppNHp, in the Sec5 complex, and GDP in our C3bot complex. Exchange of GDP to GTP would be unlikely to cause any changes in the RalA-C3bot structure, however, as the switch regions are not involved in binding of C3bot1 and previous studies have shown the binding to be nucleotide independent (Wilde et al, 2002a).

2.5.3 The C3bot-RalA interface.

Sixteen residues (9 from C3bot1 and 7 from RalA which are well ordered and clearly defined in the electron density) provide all the interactions at the binding interface as shown below in figure 2.13. The contact residues on C3bot are positioned on helices α1, α2 and the ARTT loop. The contacts on the RalA molecule are distributed on helix α4 and strand β6. This orientation of the complex is notable as RalA binds C3bot at the same position that RhoA is presumed to bind (at the ARTT loop) but in a different orientation. The ribosylation target of C3bot in RhoA (Asn41), and the equivalent residue in RalA (Arg52), lies close to switch I. But in the C3bot-RalA complex, the switch regions I and II and the GDP binding pocket are on the opposite side of RalA to the binding interface. The interaction of RalA with the ARTT loop of C3bot implies that RalA inhibits RhoA ribosylation by blocking the RhoA binding site.
Figure 2.13 The interface between RalA and C3bot

There are a total of 22 interactions at the interface between 14 RalA atoms (3 main-chain and 11 side-chain atoms) and 14 C3bot atoms (6 main-chain and 8 side-chain atoms). All of the residues involved in the interface are detailed in tables 2.2-2.4. Some $642.3 \text{ Å}^2$ of the total accessible surface area is buried at the interface upon complex formation (compared to $925 \text{ Å}^2$ in Sec5-RalA complex, 28), equivalent to 3.5% of the total surface area of the C3bot-RalA complex.
Table 2.2 Potential hydrogen bonds between C3bot and RalA

<table>
<thead>
<tr>
<th>C3bot residue</th>
<th>B-factor (Å²)</th>
<th>RalA residue</th>
<th>B-factor (Å²)</th>
<th>Distance D...A (Å)</th>
<th>Angle D-H...A (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asp204 O</td>
<td>20.65</td>
<td>Lys143 NZ</td>
<td>21.80</td>
<td>2.57</td>
<td>160.7</td>
</tr>
<tr>
<td>Ser207 O</td>
<td>33.10</td>
<td>Lys143 NZ</td>
<td>21.80</td>
<td>2.56</td>
<td>160.0</td>
</tr>
<tr>
<td>Lys56 NZ</td>
<td>32.30</td>
<td>Glu147 OE2</td>
<td>45.03</td>
<td>2.84</td>
<td>156.4</td>
</tr>
<tr>
<td>Lys73 NZ</td>
<td>22.40</td>
<td>Glu140 OE1</td>
<td>25.38</td>
<td>2.91</td>
<td>170.3</td>
</tr>
<tr>
<td>Tyr63 OH</td>
<td>27.19</td>
<td>Glu140 OE2</td>
<td>24.88</td>
<td>2.73</td>
<td>175.3</td>
</tr>
<tr>
<td>Lys81 NZ</td>
<td>22.4</td>
<td>Glu155 OE1</td>
<td>20.05</td>
<td>2.57</td>
<td>163.0</td>
</tr>
</tbody>
</table>

All contacts shorter than 3.6Å with D-H...A angle > 120° are shown (D, hydrogen bond donor; A, hydrogen bond acceptor). Hydrogen bond parameters were calculated with the program HBPLUS (McDonald & Thornton, 1994).

Table 2.3 Water mediated hydrogen bonds between the C3bot and RalA

<table>
<thead>
<tr>
<th>Water</th>
<th>C3bot residue</th>
<th>Distance (Å)</th>
<th>RalA residue</th>
<th>Distance (Å)</th>
<th>B-factor (Å²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>54</td>
<td>Lys81 NZ</td>
<td>2.88</td>
<td>Ser129 O</td>
<td>2.57</td>
<td>22.99</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Glu132 OE1</td>
<td>2.96</td>
<td></td>
</tr>
<tr>
<td>213</td>
<td>Pro205 O</td>
<td>2.64</td>
<td>Glu140 O</td>
<td>2.61</td>
<td>24.70</td>
</tr>
</tbody>
</table>

Hydrogen bond parameters were calculated with the program HBPLUS (McDonald & Thornton, 1994).

2.5.3.1 The role of the ARTT and PN loops

The ARTT loop (residues 207-214) in the C3bot has been established by structural and mutational studies to be the key region responsible for ADP-ribose transfer. It contains residues Gln212, and Glu214, conserved across all C3 exoenzymes, which are essential for the transfer of ADP-ribose onto Asn41 of RhoA (Menetrey et al, 2002; Sekine et al, 1989). The ARTT loop also contains Phe209 that has been implicated in RhoA binding by mutational studies with a P209A C3bot mutant that lost the ability to bind to, and ribosylat RhoA (Wilde et al, 2002a). This mutant however was found to retain the ability to bind RalA, suggesting that RalA did not interact directly with the ARTT loop. In our structure, this loop is well ordered apart from residues Ala210 and Gly211 and in agreement with the work performed by Wilde (Wilde et al, 2002a), Phe209 does not interact directly with RalA.
Table 2.4 Contacts in the C3bot-RalA structure

<table>
<thead>
<tr>
<th>RalA residue</th>
<th>C3bot residue</th>
<th>Distance (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Val 139 CG2</td>
<td>Val 77 CG1</td>
<td>3.61</td>
</tr>
<tr>
<td>Glu 140 CA</td>
<td>Ile 206 CG2</td>
<td>3.98</td>
</tr>
<tr>
<td>Glu 140 CB</td>
<td>Ile 206 CG2</td>
<td>3.57</td>
</tr>
<tr>
<td>Glu 140 CD</td>
<td>Tyr 63 OH</td>
<td>3.47</td>
</tr>
<tr>
<td></td>
<td>Lys 73 NZ</td>
<td>3.81</td>
</tr>
<tr>
<td>Glu 140 OE2</td>
<td>Tyr 63 CZ</td>
<td>3.40</td>
</tr>
<tr>
<td>Lys 143 CD</td>
<td>Pro 205 C</td>
<td>3.93</td>
</tr>
<tr>
<td></td>
<td>Pro 205 O</td>
<td>3.49</td>
</tr>
<tr>
<td></td>
<td>Asp 204 O</td>
<td>3.71</td>
</tr>
<tr>
<td></td>
<td>Ser 207 O</td>
<td>3.60</td>
</tr>
<tr>
<td>Lys 143 CE</td>
<td>Ser 207 O</td>
<td>3.44</td>
</tr>
<tr>
<td>Lys 143 NZ</td>
<td>Asp 204 C</td>
<td>3.70</td>
</tr>
<tr>
<td></td>
<td>Ser 207 C</td>
<td>3.43</td>
</tr>
</tbody>
</table>

Contact distances are the maximum allowed values of C-C, 4.1 Å; C-N, 3.1 Å; C-O, 3.7 Å; O-O, 3.3 Å; O-N, 3.4 Å; N-N, 3.4 Å; C-S, 4.1 Å; O-S, 3.7 Å; N-S, 3.8 Å. Superscript numbers in parentheses represent the number of contacts made by the indicated C3bot residue. The contact distances were calculated using the program CONTACT (CCP4, 1994).

The interface is centred on the ARTT loop of C3bot with Lys143 of RalA resting in a groove of the ARTT loop, forming hydrogen bonds with the main chain carbonyls of Asp204 and Ser207. The main chain carbonyl oxygen of Pro205 in the ARTT loop also forms a water-mediated interaction with the main chain carbonyl oxygen of Glu140 on helix α4 of RalA. The interface is anchored at both ends, however, by three salt bridges (referred to as hydrogen bonds due to their distance and angle, but could be partly electrostatic in nature). The two lysine residues, Lys73 and Lys81, which lie on helix α2 of C3bot, anchor the bottom half of the interface forming hydrogen bonds with Glu140 and Glu155 of RalA respectively, and also water mediated interactions with Glu132 and the carbonyl oxygen of Ser129. Lys56 from C3bot and Glu147 from RalA form another ion pair and anchor the top end of the interface.

A comparison of the RalA-C3bot complex with the native (Han et al, 2001) and NAD-bound (Menetrey et al, 2002) C3bot structures shows that part of the ARTT loop undergoes a significant conformational change on RalA binding (root mean square deviation for 6 Cα atoms of 2.94 Å between residues Ile206 and Gly211).
different to that seen for NAD binding. The dynamic change in the conformation of the ARTT loop is shown in figure 2.14. The C3bot residues Ser207 and Ala208 are slightly twisted in the RalA bound structure so that the carbonyl oxygen of Ser207 is positioned optimally for binding to Lys143 of RalA. The side chain of Ser207 is placed on the inside of the loop (aligning with Ala210 in the native structure), effectively shortening the loop by a residue and placing Ala208 (in the complex) in the same place as Phe209 (in the native structure). In the native structure, the Phe209 ring protrudes into the solvent, but the effective shortening of the ARTT loop in the complex allows it to rotate and face away from the interface, preventing it from disrupting the bond between Lys56 (RalA) and Glu147 (C3bot). The shift of this hydrophobic residue may also facilitate the interaction of Lys143 (RalA) and Asp204 (C3bot) and explains why the F209A mutation had no effect on RalA binding (Wilde et al., 2002a). The alternate conformation of the ARTT loop also appears to indicate that Ala210 is shifted so that it points inwards, towards the core of C3bot. The apparent kink in the loop at Ala210 and Gly211, although unclear in the density, is necessary to bring the sequence back in line with the native structure as the position of Gln212 is similar in both structures (displacement for \(C_a\) of 0.39Å).

Although the ARTT loop undergoes a large shift on RalA binding, the catalytic residue, Glu214, is in the same position and available for NAD hydrolysis. As the mechanism of NAD hydrolysis is not fully understood, how the binding of RalA increases its rate (Wilde et al., 2002a) is not clear, although this is similar to the increase in NAD hydrolysis observed on binding mutant RhoA with ADP-ribose acceptor residue removed (Wilde et al., 2000). However, it is possible that either the novel conformation of the ARTT loop or a difference in its mobility may contribute towards these observations.
Figure 2.14 Comparison of the ARTT loops in C3bot structures.

Figure 2.14. Stereo view of the ARTT loops of wild type C3bot (silver) and the RalA bound C3bot (gold). The key catalytic residues (Gln 2121 and Glu214) match up exactly but there are large changes in other parts of the ARTT loop. Ser207 moves into the centre of the ARTT loop displacing Ala210 and Phe209 undergoes a large shift away the side of the ARTT loop that interacts with RalA.

The previously determined C3bot-NAD structure has shown that the PN loop (residues 177-186), connecting strands β3 and β4, rearranges to enclose the nicotinamide moiety of the NAD upon binding (Menetrey et al, 2002), and is therefore thought to play a key role in NAD hydrolysis. The loss of Arg186 (in the PN loop) has been shown to abolish NAD binding and prevent ribosylation of RhoA (Menetrey et al, 2002). NAD is not bound to the C3bot-RalA complex and the PN loop, which is 10Å away from the nearest RalA residue, adopts a similar conformation to the free C3bot structure and is not involved in RalA binding. It has been suggested previously that changes in the structure of the PN loop may account for the increased rate of glycohydrolysis observed upon C3bot binding to RalA (Wilde et al, 2002a), but this is not supported by our structure. It would be supposed that the PN loop would change conformation on NAD binding of the C3bot-RalA complex, but from comparison with the C3bot-NAD complex, it seems very unlikely that this change in conformation would cause it to interact with the RalA.
2.5.4.1 Implications for other C3 exoenzymes

The structure of C3bot-RalA complex highlights structural elements relevant for the recognition of C3 exoenzymes by RalA. It is known that RalA inhibits the ribosylation of RhoA by C3cer and C3lim, as well as C3bot. This inhibition is to a lesser extent than C3botl (63% for C3lim and 24% for C3cer, measured as percentage ribosylated RhoA compared to C3bot after 30 mins) suggesting a lower binding affinity for RalA (Wilde et al, 2002a). Although the binding constant between C3bot and RalA was not measured by Wilde, the binding strength between C3lim-RalA was estimated 12nm through fluorescence quenching experiments (Wilde et al, 2002a). The closely related C3 exoenzyme C3stau2 from Staphylococcus aureus, which shares 35% similarity with C3bot, shows no binding to RalA. A comparison of the sequence alignment of the related C3 exoenzymes; C3bot, C3lim, C3cer offers an explanation for these differences and is shown below in figure 2.15A.

Of the six residues in C3bot involved in direct RalA binding, Lys56, Tyr63, Lys81 and Asp204 are not conserved in C3lim (although it is the backbone carbonyl of Asp204 which binds RalA), offering an obvious explanation for the difference in binding strength. However, the substitutions are conservative and the ability of C3lim to bind RalA suggests that they still offer a functional binding face. C3cer appears to have a lower affinity for RalA than C3bot and C3lim, which is consistent with its further sequence differences. C3cer and C3stau2 have two extra residues in the ARTT loop that could interfere with the binding.
**Figure 2.15A Sequence alignment of the C3 family**

<table>
<thead>
<tr>
<th>C3botl</th>
<th>41 50 60 70 80 90 100</th>
</tr>
</thead>
<tbody>
<tr>
<td>C3lim</td>
<td>1 10 20 30 40 60 40</td>
</tr>
<tr>
<td>C3cer</td>
<td>1 10 20 30 40 50 50</td>
</tr>
<tr>
<td>C3stau2</td>
<td>1 10 20 30 40 50 50</td>
</tr>
<tr>
<td>C3botl</td>
<td>80 80 90 90 80 90</td>
</tr>
<tr>
<td>C3lim</td>
<td>110 110 120 120 120</td>
</tr>
<tr>
<td>C3cer</td>
<td>70 70 80 80 90 90</td>
</tr>
<tr>
<td>C3stau2</td>
<td>180 210 210 210 210</td>
</tr>
<tr>
<td>C3botl</td>
<td>150 160 160 160 160</td>
</tr>
<tr>
<td>C3lim</td>
<td>150 160 170 170 170</td>
</tr>
<tr>
<td>C3cer</td>
<td>150 160 160 170 170</td>
</tr>
<tr>
<td>C3stau2</td>
<td>150 160 160 170 170</td>
</tr>
</tbody>
</table>

Figure 2.15A. Sequence alignment of the C3 family highlighting the residues at the C3bot-RalA interface. The alignment was performed and shaded by T-Coffee (Notredame et al., 2000). Residues involved in forming bonds with RalA are highlighted in grey whilst the key catalytic residues are highlighted in turquoise.

A superposition of C3stau2 (Evans et al., 2003) with the C3bot-RalA complex reveals that the extra length of the ARRT loop may be the cause of its inability to bind RalA. The two additional residues in the ARTT loop of C3stau2, Lys170 and Glu171, both extend the ARTT loop forward into the RalA interface region as can be seen in figure 2.15B.

In particular, the exchange of Pro205 (C3bot) to Glu171 (C3stau2) would obscure the hydrogen bond between Asp168 (equivalent to Asp204 in C3bot) and Lys143 (RalA) and the water mediated interaction between Glu140 (RalA) and the backbone at this position (Pro205 in C3bot). The equivalent residues in the C3cer are smaller (both aspartate) and are presumably less detrimental to RalA binding.
Figure 2.15B. Comparison of the interface region of C3bot (gold) and C3stau (silver) highlighting the residues that form bonds at the interface in the C3bot and their equivalents in C3stau. The extra residues in the ARTT loop of C3stau (Lys170 and Glu171) block access to the ARTT loop and would hinder RalA interacting with the ARTT loop.

This comparison of the C3 exoenzymes therefore also allows speculation as to which residues in the interface are most crucial to the interaction. It is interesting to note that the residues that are conserved between C3bot and C3lim (Pro205, Ser207 and Lys73 in C3bot) and the presumably conserved interaction of the backbone oxygen at Asp204 in C3bot, are clustered at the centre of the interface. It is also two of these interactions (of Asp204 and Pro205 in C3bot) that may be disrupted in C3stau2 by the extra length of the ARTT loop. These central interactions would therefore appear to be the key residues in the interface, held in position to differing extents by the outer interactions. Of the two outer interactions, the bonds formed by the two lysines of C3bot (Lys56 and Lys81) would appear to promote a more ideal binding face than the equivalent two arginines in C3lim, as judged by the biochemical data (Wilde et al, 2002a).
2.5.4.2 Implications for other GTPases

The well documented specificity of C3bot for ribosylation of RhoA-C via the ARTT loop, and our findings that RalA binds at the same site, lead to the question of whether RhoA or other GTPases could bind C3bot in the same orientation as RalA. The sequence alignment of RalA with Ras, RhoA and the Rho-related Rac and Cdc42 in figure 2.16 shows that not all the RalA binding residues are conserved across these GTPases. Of the four direct binding residues involved in the RalA interface, only Glu 147 is conserved in all five GTPases. In particular, Ras replaces Glu140 and Glu147 of RalA with the positively charged arginines, disallowing interactions with Lys73, Tyr63 and Lys56 of C3bot. These sequence differences in Ras would prevent half of the direct binding interactions available to RalA. Also, RhoA, Cdc42 and Rac all show a conserved change of Glu132 (RalA) to an arginine, which may affect the water-mediated interaction with Lys81 of C3bot1. Similarly, this residue is replaced by an alanine in Ras.

Surprisingly, RhoA shows the greatest degree of similarity to RalA at these residues, even though RhoA is ribosylated by C3bot at Asn41. As this ribosylation occurs on the opposite side of RhoA from the equivalent RalA binding face, it is presumed that the residues around the ribosylation site provide a better interface. These comparisons suggest that the interaction between RalA and C3bot is not related to the highly specific ribosylating activity that C3bot possesses for RhoA, but is more dependent on specific RalA residues, not present in RhoA. In particular, Lys143 of RalA, which makes two of the central interactions of the interface and sits in a groove of the ARTT loop, is not conserved in any of the other GTPases. It is this residue and Glu140 which form the central interactions presumed to be common between C3bot1 and C3lim. Additionally, Lys143 and Glu140 bury 94Å² and 90Å² solvent-accessible surface area, respectively, on complex formation, making them ideal candidates for anchoring residues (Rajamani et al, 2004).
Figure 2.16 Sequence alignment of the GTPase family highlighting the residues at the C3bot-RalA interface. The alignment was performed and shaded by T-Coffee (Notredame et al. 2000). Residues involved in forming bonds with C3bot are highlighted in red.

2.5.5 An alternative structure

The publication of an alternative structure for a C3bot-RalA complex (Pautsch et al., 2005) and our own findings presented in chapter 5 have now raised some questions about the biological relevance of this model. In addition upon re-examination a symmetry related molecule positioned C3bot and RalA oriented in such a manner that the alternative complex centred on the switch II loop of RalA was formed, suggesting that this complex centred on the ARTT loop may come about from symmetry related contacts.
2.6 Conclusions

We have determined the structure of the C3bot-RalA complex at 2.66Å. The complex occurs around helix α4 on RalA and the ARTT loop of C3bot. The interaction centred on the active site of C3bot readily explains the observed inhibition of Rho ribosylation. Sequence analysis can also offer an explanation for the Ral specific nature of this in the GTPase family and the broader specificity in the C3 family. However, upon re-examination it can be seen that there are symmetry related molecules that can form the complex seen in chapter 5 and the structure published by Pautsch and co-workers (Pautsch et al, 2005). The consequences of this and an examination of the merits of each of the relative complexes are presented in chapter 8.
3. The structure of the C3bot-RalA-NAD complex

3.1 Aims of obtaining an NAD complex

It was recently shown that C3bot binds to RalA in such a manner that inhibits the ribosylation of Rho and shows several other interesting properties (Wilde et al, 2002a). One of the reported properties associated with this interaction was the 5 fold increase observed in the rate of NAD glycohydrolysis (Wilde et al, 2002a). Our aim was to determine the structure of the C3bot-RalA complex with NAD soaked into the active site and examine the structure for any changes that may explain the increased glycohydrolytic rate.

3.2 NAD binding and C3bot-RalA interactions

C3bot is an ADP-ribosyltransferase that catalyses the hydrolysis of NAD and the transfer of the ADP moiety onto Asn41 of RhoA-C (Aktories & Frevert, 1987; Sekine et al, 1989). It has also been shown to be able to auto-hydrolyse NAD without any substrate present and the K_M for the glycohydrolase reaction is ~60µM (Aktories & Hall, 1989).

Upon NAD binding the active site of the C3bot molecule has been seen to undergo a ‘crab claw’ movement involving the PN loop and other elements of the C3bot secondary structure moving by as much as 10Å to enclose the NAD fully in the active site (Evans et al, 2004; Menetrey et al, 2002).

On of the unusual biochemical properties associated with the interaction between C3bot and RalA was an increase in the rate of glycohydrolysis by approximately 5 fold (Wilde et al, 2002a). As the complex between C3bot and RalA centered around the switch II region has been shown by Pautsch and co-workers to have no effect upon NAD binding (Pautsch et al, 2005) the increased rate of glycohydrolysis must
come about from changes in the active site when the complex between C3bot and RalA forms around the ARTT loop of C3bot (Holbourn et al, 2005).
3.3 Methods

3.3.1 Expression and purification of C3bot and RalA

C3bot and RalA were purified exactly as described before.

3.3.2 Crystallisation

For crystallisation RalA and C3bot were incubated together overnight in a ratio of 2:1 giving a working concentration of 6.67mg/ml RalA and 3.33mg/ml C3bot. Crystallisation was performed in 0.1M Hepes pH 7.3, 12-15% ethylene glycol, 14-18% PEG8000 at 16°C. Thin needle like crystals appeared over the course of ~3 days.

3.3.3 Soaking NAD into C3bot-RalA crystals

A solution of 5mM NAD in 10% ethylene glycol 14% PEG 8000, 0.1M Hepes pH 7.3, 100mM NaCl was used for soaking the C3bot-RalA crystals. 2μl of the NAD solution was added to the crystal containing drops and then incubated at room temperature for ~3 hours before being flash frozen in a stream of liquid nitrogen and data collection.

3.3.4 Data collection and refinement

Diffraction data to 2.47Å were collected from a single cryo-cooled (100K) crystal with 25% ethylene glycol on station X13 of the synchrotron radiation source at EMBL Hamburg (Germany) using an ADSC Quantum 4 CCD detector. The data were processed and scaled using the HKL2000 package (Otwinowski, 1993a; Otwinowski, 1993b). The symmetry and systematic absences were consistent with the P2_12_1 space group (unit cell dimensions a=56.52Å, b=92.30Å and c=102.36Å) with one C3bot-RalA complex per asymmetric unit and the crystals containing ~54%
solvent. Data reduction was carried out using the program TRUNCATE (CCP4, 1994).

The structure of the C3bot-RalA complex was determined by molecular replacement method with AmoRe (Navaza, 1994) using the original C3bot-RalA complex (PDB code 2BOV) (Holbourn et al, 2005) as a search model. Initial refinement of the structure was carried out using REFMAC5 (Murshudov, 1997) which improved the model and subsequent rounds of refinement was carried out using the CNS suite (Brünger et al, 1998) and the model building program O (Jones et al, 1991). In each data set 5% of reflections were kept aside for $R_{free}$ calculation (Brünger, 1992). Water molecules were added using the WATERPICK module of CNS. The clear continuous density observed (in both the 2Fo-Fc and Fo-Fc SIGMA-A weighted maps) in the GTP binding pocket of RalA during the final stages of the refinement was interpreted as GDP. Clear continuous density around the NAD binding pocket was interpreted as an NAD molecule. The final model contains 3010 protein atoms, 90 water molecules, one NAD molecule, one GDP molecule and one Mg$^{2+}$ ion, has an R-factor of 25.1% and an R-free value of 28.7%. The details of the data processing and refinement statistics are presented in table 3.1. The stereochemical quality of the finished model was checked using PROCHECK (Laskowski et al, 1993).
Table 3.1. Details of crystallographic data

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution range (Å)</td>
<td>50-2.47 A</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.81</td>
</tr>
<tr>
<td>Exposure time (s)</td>
<td>30</td>
</tr>
<tr>
<td>Number of images</td>
<td>150</td>
</tr>
<tr>
<td>Rotation per image (°)</td>
<td>1</td>
</tr>
<tr>
<td>Number of reflections measured</td>
<td>271926</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>19844</td>
</tr>
<tr>
<td>I/σ(I) (outer shell)</td>
<td>9.8 (1.7)</td>
</tr>
<tr>
<td>Completeness (outer shell) (%)</td>
<td>86.7 (64.8)</td>
</tr>
<tr>
<td>R_{symm} (outer shell) (%)</td>
<td>13.3 (35.3)</td>
</tr>
<tr>
<td>R_{cryst} (%)</td>
<td>25.1</td>
</tr>
<tr>
<td>R_{free} (%)</td>
<td>28.7</td>
</tr>
<tr>
<td>Average temperature factor (Å²):</td>
<td></td>
</tr>
<tr>
<td>RalA main-chain (side-chain)</td>
<td>37.1 (37.5)</td>
</tr>
<tr>
<td>C3 main-chain (side-chain)</td>
<td>38.1 (38.5)</td>
</tr>
<tr>
<td>GDP</td>
<td>29.2</td>
</tr>
<tr>
<td>NAD</td>
<td>38.4</td>
</tr>
<tr>
<td>Solvent</td>
<td>37.6</td>
</tr>
<tr>
<td>Mg^{2+} ion</td>
<td>28.8</td>
</tr>
<tr>
<td>RMSD from ideal values</td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.007</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.550</td>
</tr>
</tbody>
</table>

*Outer shell is 2.56Å – 2.47Å

b\( R_{symm} = \frac{\Sigma_i |I_i(h) - \langle I_i(h)\rangle|}{\Sigma_i |I_i(h)|} \), where \( I_i(h) \) and \( I(h) \) are the \( i \)th and the mean measurements of the intensity of reflection \( h \), respectively.

c\( R_{cryst} = \frac{\Sigma_h |F_o(h) - F_c(h)|}{\Sigma_h F_o(h)} \), where \( F_o \) and \( F_c \) are the observed and calculated structure factor amplitudes of reflection \( h \), respectively.

d\( R_{free} \) is equal to \( R_{cryst} \) for a randomly selected 886 (4.4%) reflections not used in the refinement.
3.4 Results and discussions

3.4.1 Experimental results

Crystals of the complex between C3bot and RalA were obtained as described previously in chapter 2 (Holbourn et al, 2005). The needle-like morphology of the crystals suggested that they were crystals of the complex that forms around the ARTT loop of C3bot rather than the alternative complex that is centred on the switch II region of RalA and described in chapter 5 (Pautsch et al, 2005). Two different methods of soaking NAD into C3bot crystals have been reported and both were used (Menetrey et al, 2002) (Pautsch et al, 2005). Both methods involved soaking the NAD directly into the crystal prior to flash freezing for cryogenic data collection with 25% ethylene glycol as a cryo-protectant. One involved using a relatively high concentration of NAD (20mM) for a short soak (1 hour) (Menetrey et al, 2002) while the other that involved a lower NAD concentration (5mM) but a longer soak (~3 hours) (Pautsch et al, 2005) gave a better data set.

Data to 2.47Å were collected from a single cryo-cooled (100K) crystal on an ADSC-CCD detector on station X13 at the Synchrotron radiation source at the EMBL Hamburg. The data were processed & scaled using the HKL2000 package (Otwinowski, 1993a). The structure of the original C3bot-RalA complex (PDB code 2BOV) (Holbourn et al, 2005) was used as a search model for molecular replacement using the molecular replacement program AMoRe (Navaza, 1994). The top solution from AMoRe was initially refined with REFMAC5 (Murshudov, 1997) and gave a drastic improvement in the model ($R_{crys}$ 25.7%, $R_{free}$ 30.7, FOM 0.76). Subsequent rounds of refinement and model building were performed using CNS (Brünger et al, 1998) and the molecular graphics program O (Jones et al, 1991). During the course of refinement a large stretch of contiguous difference density was observed in the nucleotide binding site of the RalA molecule and this was modelled as GDP and a single magnesium ion. A second patch of contiguous difference density was observed.
in the active site of the C3bot molecule. An NAD molecule was modelled into this density as shown in figure 3.1 and led to a drop in $R_{crys}$ and $R_{free}$.

**Figure 3.1 Electron density in the active site of C3bot**

![Figure 3.1](image)

**Figure 3.1.** A stereo view of the 2fo-fe density at 3 SIGMA-A in the active site of the C3bot molecule with a molecule of NAD fitted into the difference density.

Analysis of the Ramachandran plot ($\Psi$-$\Phi$) showed 88.4% of residues lie within the most favourable region. The details of the data processing and refinement statistics are presented in table 3.1. The stereochemical quality of the finished model was checked using PROCHECK (Laskowski et al, 1993) as shown in figure 3.2. The final refined structure of C3bot-RalA-NAD complex at 2.47Å resolution has a conventional R-factor ($R_{crys}$) of 25.1% and a free R-factor ($R_{free}$) of 28.7% and is shown below in figure 3.3. The final model of the complex contains 3010 protein atoms, 90 water molecules, one NAD molecule in the active site of C3bot and one GDP molecule, with an Mg$^{2+}$ ion, bound to the nucleotide binding site of RalA.
Figure 3.2. A Ramachandran plot of the refined structure of the C3bot-RalA-NAD structure. No residues lie in the disallowed regions and 88.4% of residues lie within the most favourable region. The Ramachandran plot was calculated with PROCHECK (Laskowski et al, 1993).
Figure 3.3 The structure of the C3bot-RalA-NAD complex. RalA is shown here with yellow α-helices and green β-strands and a GDP molecule shown in the nucleotide binding site in ball-and-stick. C3bot is shown here with blue β-strands and red α-helices with NAD in the active site in ball-and-stick. The position of the PN and ARTT loops are also highlighted.

3.4.2 Overall structure of the C3bot-RalA-NAD complex

The overall structure of the complex shown in figure 3.3 maintains the same structure as that of the original complex (Holbourn et al, 2005). The C3bot and RalA are oriented towards each other with the interface between them taking place primarily between helix α-4 of RalA and the ARTT loop and helix α-2 of C3bot. The interface between C3bot and RalA maintains the same conformation as described previously in chapter 2 and is shown below in figure 3.4 (Holbourn et al, 2005).
A least squares superposition of the original complex onto the NAD-bound complex reveals very little difference with an RMSD of 0.2Å over 172 Cα atoms of RalA and 0.4Å over 207 Cα atoms of C3bot.

3.4.3 The NAD binding site

The largest differences of C3bot occur around the PN loop that undergoes a dramatic ‘crab-claw’ movement to bring Phe183 and Arg186 closer to NAD to stabilise the ring of the nicotinamide moiety and the oxygens of one of the phosphate groups (Menetrey et al, 2002) (Holbourn et al, 2006). The NAD binding site possesses the
same interactions between C3bot and NAD as observed in the wild-type structure. The important interactions are highlighted below in figure 3.5.

**Figure 3.5 The NAD binding site highlighting some of the important interactions**

![Figure 3.5](image)

The three arginines that have been shown to be important to NAD binding: Arg91 from the α-3 motif; Arg128 from the conserved Arg/His motif; and Arg186 from the PN loop are highlighted. Also shown are Tyr79 and Asn87 on the α-3 helix. Asn87 binds directly to one of the β-phosphate oxygens whilst Tyr79 forms an interaction with Ser174 of the STS loop and through Ser174 with the catalytic glutamate, Glu214 (Menetrey *et al.*, 2002; Evans *et al.*, 2003; Holbourn *et al.*, 2006).

When the structure of the RalA bound C3bot-NAD is superimposed onto the wild type C3bot-NAD, shown below in figure 3.6A and B, there are two main areas of difference.
Figure 3.6 Differences between wild type and RalA bound C3bot-NAD

A. The ARTT loop of the C3bot-RalA-NAD structure. The inversion of Ser207 and Ala208 introduces additional hydrogen bonds within the ARTT loop that may act to stiffen the ARTT loop.

B. A close-up view of the PN loop of the wild-type structure in gold (PDB code 1GZF) (Menetrey et al., 2002) and the current structure in silver. The loop undergoes an inversion with Val180 and Ser181 reversing direction and Gln182 undergoing a slight shift that places it too far from Gln212 to form an interaction.

When compared to the wild type the ARTT loop of the RalA bound C3bot is more rigid due to the unique conformation of the ARTT loop seen in the C3bot-RalA
structure. The inversion that takes place about Ser207 to place Ser207 in the centre of the ARTT loop (Holbourn et al, 2005) allows it to form possible ionic interactions with the carbonyl oxygens of Phe209, Ala208 and Asp204. In the wild type with Ala210 in the centre of the ARTT loop and the carbonyl oxygen of Ala208 facing outwards these interactions would be lost. The increase in NAD glycohydrolysis could then come about from the “stiffness” that these interactions give to the otherwise floppy ARTT loop. The interaction with only the carbonyl oxygen would also support the mutagenic data that a Phe209Ala mutant had no effect on the properties of the C3bot-RalA interaction (Wilde et al, 2002a).

The PN loop also undergoes a change in the C3bot-RalA-NAD structure. While the critical interactions involving Phe183 and Arg186 are maintained in both the wild type and RalA-bound structures there are differences in the rest of the ARTT loop. The most pronounced differences are with Val180 and Ser181 that switch their solvent/active site facings. In the case of the RalA-bound PN loop the Val180 now faces outwards into the solvent whilst Ser181 now faces into towards the active site, although it is too far from the NAD molecule to form direct interactions. The other change is in the position of Gln182. In the RalA-bound structure Gln182 is displaced slightly with respect to the wild type and thus is positioned too far from the ARTT loop to stabilise Gln212. Mutagenesis studies using a Q182A mutant found virtually no change in NAD binding ($K_D$ 75μM ±15 compared to 60μM ±6 for the wild type) (Menetrey et al, 2002) but a doubling of the rate of glycohydrolysis (Menetrey et al, 2002).

3.4.4 Electrostatic changes around the active site

Whilst there are some small changes observed in the ARTT loop and the PN loop they do not fully explain the increased glycohydrolytic rate increase. A large change in the electrostatic nature of the active site may be responsible for the increased rate of NAD hydrolysis. The electrostatic surface of C3bot and C3bot-NAD complexed with RalA are shown below in figure 3.7A and B.
Figure 3.7 Electrostatic surface picture of the C3bot active site

A

B

Figure 3.7A. The electrostatic surface of C3bot with NAD. B. The electrostatic surface of C3bot-NAD when in complex withRalA. The electrostatic surfaces were generated using APBS tools in PYMOL (Baker et al, 2001).
As can be seen in figure 3.7 the electrostatic surface of C3bot changes dramatically upon complexation with RalA. In the uncomplexed C3bot the surface charge is predominantly positive, as would be expected with C3bot’s high iso-electric point, with small areas of neutral or negative charge. In figure 3.7A there are two patches of negative charge. One of these lies around the active site E214 and the exposed 2’OH of NAD. The second patch of negative density however is at the beginning of the ARTT loop adjacent to D204 and S207 and is the small patch of negative density that interacts with K143 of RalA in the interface centred on the ARTT loop. When complexed with RalA there is a drastic shift in the overall electrostatic potential of the complex with face of the C3bot shifting towards a more neutral charge and large changes around the active site including the formerly neutral helix α3 becoming slightly negatively charged. This large change in electrostatic potential may be the cause of the increased rate of NAD hydrolysis.

3.5 Conclusions

We have determined the structure of the C3bot-RalA-NAD complex. This retains all of the features of the already determined C3bot-RalA complex but a close examination of the ARTT loop and PN loop may offer an insight into the increased rate of glycohydrolysis seen for the C3bot-RalA interactions. This may come about from changes in the PN loop that involve the loss of an interaction between Gln212 and Gln182 and additional hydrogen bonds to give a more stable ARTT loop. However an examination of the electrostatic potential of the surface of the molecules illustrates a large change in the surface charge that may also play a role in the increase in NAD hydrolysis.

As in the original complex described in chapter 2 this complex also possesses a symmetry related interface identical to the ‘alternative’ interface centred on the Switch II region of RalA. However the changes noted around the PN loop and ARTT loop in the structure presented here seem specific to our structure as no alterations were reported in the NAD-bound complex reported by Pautsch and co-workers (Pautsch et al, 2005).
4. Mutagenesis of residues central to the C3bot-RalA interface

4.1 Aims of mutagenesis

Analysis of the structure between C3bot and RalA illustrated a network of key ion pairs at the interface that seemed to be important for the formation of the C3bot-RalA complex. Our aims were to perform site-directed mutagenesis on 5 of the key residues from RalA and assess the impact that these mutations had on the binding kinetics.

4.2 Interactions between C3bot and RalA

C3bot was long thought to be specific to the Rho family of GTPases and RhoA-C in particular (Aktories et al, 1989; Braun et al, 1989; Chardin et al, 1989). The recent work demonstrating its ability to interact with the Ras-like GTPase, RalA, in a novel manner that did not result in ribosylation of RalA was intriguing (Wilde et al, 2002a).

The initial structure of C3bot-RalA (Holbourn et al, 2005) demonstrated an interface that involved helix \( \alpha-4 \) of RalA and the ARTT loop and helix \( \alpha-2 \) of C3bot. This interface lying at the active site of C3bot suggested a possible mechanism for the RalA induced inhibition of RhoA ribosylation by direct competition in which the binding of RalA physically blocks the access of RhoA to the C3bot active site.
4.3 Methods

There were several key residues of RalA assessed as being important to the formation of the C3bot-RalA complex. These were the residues selected for site directed mutagenesis. The mutagenesis involved using the Stratagene quick change mutagenesis kit and a KOD polymerase based method.

4.3.1 Mutagenesis using the Stratagene quick change mutagenesis kit

4.3.1.1 Mutagenesis

The human RalA gene clone (as a pGEX-2T fusion) was a kind gift of Dr Hiroshi Koide, Department of Stem Cell Biology, Graduate School of Medical Science, Kanazawa University, Japan. The following primers were designed to induce the point mutations. In each case the altered nucleotides are underlined.

Mutant primer E132A
GGT TGG TAA CAA ATC AGA TTT AGC AGA TAA AAG ACA GGT TTC TGT
Mutant primer E140A
AAA GAC AGG TTT CTG TAG CAG AGG CAA AAA ACA GAG C
Mutant primer E140Q
GAT AAA AAG ACA GGT TTC TGT ACA AGA GGC AAA AAA CAG AGC CTG
Mutant primer K143A
GGT TTC TGT AGA AGA GGC AGC AAA CAG AGC TGA GCAGTG
Mutant primer K143R
GGT TTC TGT AGA AGA GGC AAA GAA CAG AGC TGA GCAGTG
Mutant primer E147A
GGC AAA AAA CAG AGC CTG CAC AGT GGA A TG TTA ACT ACG
Mutant primer E155A
GG AAT GTT AAC TAC GTG GCA ACA TCT GCT AAA ACA CG
The PCR was performed using the Stratagene quick change mutagenesis PCR system and the following protocol: 95°C 1 minute, 30 cycles of 95°C 1 minute, 55°C 1 minute, 65°C 12 minutes, followed by a 12 minute hold at 65°C and storage at 4°C. The resulting DNA was then digested for 1 hour at 37°C with DpnI (Stratagene) and transformed directly through heat shocking into competent XL-10 cells (Stratagene) and grown overnight at 37°C on LB agar plates supplemented with 100μg/ml ampicillin. Single colonies were then selected and used to inoculate 10ml of luria broth supplemented with 100μg/ml ampicillin. Plasmid DNA was extracted using a mini-prep wizard (Promega) and sent for T7 forward and reverse sequencing (MWG Biotech).

4.3.1.2 Expression of RalA K143A

Following sequencing the RalA K143A plasmid was transformed via heat shock into BL21(D3)codon+ cells (Novagen). A successful colony was inoculated into 100ml of Terrific Broth supplemented with 0.5% glucose and 100μg/ml ampicillin and grown overnight at 37°C. The culture was diluted 1:50 into 1l of fresh medium and grown for 4 hours at 37°C until an OD$_{600}$ of ~0.8 was reached. Expression was induced by addition of IPTG to a final concentration of 1mM and grown for a further 4 hours at 25°C. Cells were harvested by centrifugation (Beckman Avanti J-25).

4.3.1.3 Purification of RalA K143A

RalA K143A was purified following the protocol used for wild type RalA.

4.3.2 Mutagenesis using a KOD polymerase based method

As only 1 mutation out of 7 had been successful it was decided a different mutagenesis protocol would be used.
4.3.2.1 Mutagenesis

The following forward and reverse primers were used (site directed mutants are underlined).

E132A forward primer
GGT TGG TAA CAA ATC AGA TTT AGC AGA TAA AAG ACA GGT TTC TGT
E132A reverse primer
ACA GAA ACC TGT CTT TTA TCT GCT AAA TCT GAT TTG TTA CCA ACC
E140A forward primer
AAA GAC AGG TTT CTG TAG CAG AGG CAA AAA ACA GAG C
E140A reverse primer
GCT CTG TTT TTT GCC TCT GCT ACA GAA ACC TGT CTT T
E140Q forward primer
GAT AAA AGA CAG GTT TCT GTA CAA GAG GCA AAA AAC AGA GCT G
E140Q reverse primer
CAG CTC TGT TTT TTG CCT CT TGT GT A CAG AAA CCT GTC TTT TAT C
K143A forward primer
GGT TTC TGT AGA AGA GGC AGC AAA CAG AGC TGA GCA GTG
K143A reverse primer
CAC TGC TCA GCT CTG TTT GCT GCC TCT TCT ACA GAA ACC
K143R forward primer
GGT TTC TGT AGA AGA GGC AAG GAA CAG AGC TGA GCA GTG
K143 reverse primer
CAC TGC TCA GCT CTG TTC CTT GCC TCT TCT ACA GAA ACC
E147A forward primer
AGA GGC AAA AAA CAG AGC TGC ACA GTG GAA TGT TAA CTA CG
E147A reverse primer
CGTA GTT AAC ATT CCA CTG TGC AGC TCT GTT TTT TGC CTC T
E155A forward primer
GGA ATG TTA ACT ACG TGG CAA CAT CTG CTA AAA CAC G
E155A reverse primer

CGT GTT TTA GCA GAT GTT GCC ACG TAG TTA ACA TTC C

The PCR was performed using KOD polymerase (Novagen) PCR system and the following protocol: 95°C 1 minute, 18 cycles of 95°C 30 seconds, 60°C 50 seconds, 68°C 8 minutes, followed by a 7 minute hold at 68°C and storage at 4°C. The resulting DNA was then digested for 1 hour at 37°C with DpnI (Promega) and transformed directly through heat shocking into BL21(DE3)codon+ cells (Novagen) and grown overnight at 37°C on LB agar plates supplemented with 100µg/ml ampicillin. Single colonies were then selected and used to inoculate 10ml of luria broth supplemented with 100µg/ml ampicillin. Plasmid DNA was extracted using a mini-prep wizard (Promega) and sent for T7 forward and reverse sequencing (MWG Biotech).

4.3.2.2 Expression and purification of RalA mutants

The expression and purification of the RalA mutants, E132A, E140A and K143A followed the same protocol as used for wild type RalA.
4.4 Results and Discussion

4.4.1 Experimental results

The choice of residues for mutation was based on the key residues identified at the interface between C3bot and RalA (Holbourn et al., 2005). The 5 residues that have side chains involved in hydrogen bond formation were all selected. Additional mutants of Lys143 and Glu140 were designed as they both bury ~90Å in the interface and are involved in numerous hydrogen bonds and seem to be key residues involved in the formation of the interface.

Generation of the mutants using the Stratagene quick change mutagenesis kit proved more difficult than expected as in the initial mutagenesis experiments only 1 of the 7 mutations was successful (K143A) as the rest failed at the PCR step or when sequencing results were analysed showed no mutation present.

After several failed attempts to generate mutations using the Stratagene kit it was decided to alter the protocol to avoid using the Stratagene kit. KOD polymerase (Novagen) was used as a fast, high fidelity polymerase and the cells were treated with DpnI from Promega rather than Stratagene. After digestion the plasmids were then transformed directly into BL21(DE3)codon+ expression cells, avoiding the use of the XL-10 cells (Stratagene). This alternative protocol successfully managed to generate 3 more mutations; E132A, E140A and E155A, although E155A was never expressed on a large scale for purification.

Purification of the mutants was straight forward as all of them expressed as a soluble GST-fusion protein and were purified in the same manner as the wild type RalA to a high degree of purity as shown below in figure 4.1. Though yields of all the mutants produced was lower than the wildtype RalA, less than 2mg/l of culture.
4.4.2 Effects of the mutations on the formation of a C3bot-RalA complex

The mutants were going to be used alongside the wild type RalA for a kinetic analysis of the binding between C3bot and RalA. However, as the mutants were being prepared Pautsch and co-workers published their work that included extensive mutagenesis results and an alternative structure of the C3bot-RalA complex, though very similar to that of the one described in chapter 5 (Pautsch et al, 2005).

Figure 4.1. Purification of RalA mutants

Figure 4.1. SDS-page Gel illustrating the purity of two of the RalA mutants, E140A and K143A, following GST affinity and gel filtration chromatography.

It was reported that the binding constant between C3bot and RalA was fairly strong ($K_D \sim 60\text{nM}$ (Isothermal calorimetry) $K_D \sim 35\text{nM}$ (fluorescence titration)) (Pautsch et al, 2005) though weaker than the measured interaction between C3lim and RalA ($K_D \sim 12\text{nM}$) (Wilde et al, 2002a). Mutational analysis also indicated that mutation of many of the C3bot residues at the interface of the original complex presented in chapter 2 had no effect on the inhibition of RhoA or precipitation of C3bot by GST-RalA. This was demonstrated with several C3bot mutants: K73L; K81E and P205A.
(Pautsch et al, 2005). Two of these mutations, K73L and K81E are the partners for the RalA residues Glu147, Glu155 and Glu140. The results for the third mutation, P205A, were unsurprising as Lys143 of RalA interacts with the man chain carbonyl of Pro205. Mutations of the two most heavily buried residues at the interface (K143E and E140A of RalA) were also analysed and shown to have no effect on the inhibition of RhoA ribosylation or precipitation of C3bot on a GST pull down assay (Pautsch et al, 2005).

Taken together the mutagenic analysis of the residues around the original C3bot-RalA interface (Holbourn et al, 2005) suggests that it may not be the true biological dimer though the structure of the alternative complex, both the published structure [PDB code 2A78] (Pautsch et al, 2005) and the one presented in chapter 5 still do not answer all of the biochemical questions raised by the interaction between C3bot and RalA (Wilde et al, 2002a).

4.5 Conclusions

We have managed to purify 3 mutant variants of RalA (E140A, K143A and E132A) and clone an additional mutant (E155A). However due to the work published by Pautsch and co-workers (Pautsch et al, 2005) we have not used these mutants in an activity study. The results of mutagenic work carried out by Pautsch and co-workers (Pautsch et al, 2005) however do support the complex being centred on the Switch II loop of RalA and, when taken with the late observation that the Switch II interface is present in the original structure presented in chapter, suggests that the Switch II interface the true biological interface rather than the one centred on the ARTT loop.
5. The structure of the alternative C3bot-RalA complex

5.1 Aims of determining the structure of the alternative complex

A second crystal form of the C3bot-RalA complex gave rise to an alternative form of the complex that has 2 C3bot and 2 RalA in the asymmetric unit and a different binding interface to the one already described. By analysis of the structure of this alternative complex it was hoped that some of the unanswered questions about the C3bot-RalA interaction would become clearer.

5.2 An alternative arrangement of the C3bot-RalA complex

C3bot and the related C3lim and C3cer have been found to bind to RalA in a manner that does not result in the ribosylation of RalA and inhibits the activity of the C3 exoenzyme involved (Wilde et al, 2002a). This interaction is strong both for C3bot and C3lim with $K_D$ 60nM (Pautsch et al, 2005) and 12nM (Wilde et al, 2002a) respectively. The interaction between C3bot and RalA, although resulting in inhibition of ribosylation of RhoA, does not result in the ribosylation of RalA and increases the rate of NAD-glycohydrolysis by C3bot five fold (Wilde et al, 2002a). This interaction has also been reported as being nucleotide independent (Wilde et al, 2002a) or GDP dependant (Pautsch et al, 2005).

The structure of the C3bot-RalA complex has now been determined in two alternative complexes (Holbourn et al, 2005; Pautsch et al, 2005). Both complexes are formed from one C3bot molecule binding to one RalA molecule; the differences lie in the orientation and positioning of the molecules. In the original complex (Holbourn et al, 2005) the C3bot is bound around the ARTT loop and the RalA at helix a4. In the second complex the C3bot is bound at helix 3 and 4 distant from the C3bot active site and at the RalA switch I region (Pautsch et al, 2005). However neither complex answers the entire set of questions raised by the biochemical information.
5.3 Methods

5.3.1 Protein Expression and purification

The C3bot and RalA were purified as described previously.

5.3.2 Crystallisation

Crystals resembling square plates were observed after using C3bot-RalA solutions that had been stored for greater than 1 month at 4°C. C3bot-RalA crystals were grown by the hanging drop vapour diffusion method in 0.1M Hepes pH 7.3, 12-15% ethylene glycol, 14-18% PEG8000 at 16°C over 3-4 days with protein concentrations 6.67mg/ml RalA, 3.33mg/mlC3bot.

5.3.4 Data Collection

Data to 2.4Å were collected from a single cryo-cooled (100K) crystal on an ADSC-CCD detector on station 9.6 at the Synchrotron radiation source at Daresbury, UK. The data were processed and scaled using the HKL2000 package (Otwinowski, 1993a; Otwinowski, 1993b). The symmetry and systematic absences were consistent with a P2₁ space group and a unit cell with dimensions a=91.7, b=56.7, c=90.8Å, β=124°. The Matthews coefficient indicated that there were two C3bot and two RalA molecules in the asymmetric unit.

Initial phases were found with PHASER (McCoy et al, 2005) using the C3bot and RalA molecules from the original complex as search models (PDB code 2BOV). PHASER gave two solutions and the first solution was used for refinement (R-factor 48.2%, log likelihood gain +709, Z score +26.8 for the first solution and R-factor 48.3%, log likelihood gain +560, Z score +20.3). Initial refinement of the structure was carried out using REFMAC5 (Murshudov, 1997) which improved the model and subsequent rounds of refinement were carried out using the CNS suite (Brünger et al,
1998) and the model building program O (Jones et al, 1991). In each data set 5% of reflections were kept aside for $R_{\text{free}}$ calculation (Brunger, 1992). Water molecules were added using the WATERPICK module of CNS (Brünger et al, 1998). The continuous density observed in the nucleotide binding pocket (in both the 2Fo-Fc and Fo-Fc SIGMA-A weighted maps) during the final stages of the refinement was interpreted as an Mg$^{2+}$ ion and a GDP molecule. The stereochemical quality of the finished model was checked using PROCHECK (Laskowski et al, 1993).
5.4 Results and discussion

5.4.1 Experimental methods

The expression and purification of the C3bot and RalA were identical to that previously described and the solutions used for crystallisation of this 'alternative' complex were from the same protein batch that gave rise to the complex described previously (Holbourn et al, 2005).

Crystallisation in the same conditions as that used to generate the original complex, 0.1M Hepes pH 7.3, 12-15% ethylene glycol, 14-18% PEG8000, but using protein stocks that had been stored for longer than 1 month at 4°C gave rise to a different crystal morphology. These large plate-like crystals were easily distinguishable from the expected needle-like crystals obtained before as can be seen in figure 5.1A and B.

Figure 5.1 Crystals of the different C3bot-RalA complexes

A B

Figure 5.1A. Crystals of the original C3bot-RalA complex grown in 0.1M Hepes pH 7.3, 12-15% ethylene glycol, 14-18% PEG8000. 5.1B. The square crystals of the alternative C3bot-RalA crystal also grown in 0.1M Hepes pH 7.3, 12-15% ethylene glycol, 14-18% PEG8000.

The final model contains 2 molecules of RalA and 2 molecules of C3bot, 5996 protein atoms, 337 water molecules and 58 ligand atoms, has an R-factor of 23.3% and an R-free value of 31.0% and is shown in figure 5.2. A molecule of GDP and an Mg\(^{2+}\) ion was identified and modelled bound to each of the RalA molecules in the
nucleotide binding site. Residue 42 on the first RalA molecule and a further 10 residues on the second RalA molecule including residues 178-181 had their side chains modelled with occupancy 0 due to weak density. 4 Residues on the first C3bot molecule and 11 residues on the second C3bot were disordered and modelled on the basis of weak density with their side chains modelled with 0 occupancy due to weak density. Analysis of the Ramachandran plot (Φ-Ψ) in figure 6.3 showed 85.3% of residues lie within the most favourable region with 4 residues in disallowed regions (A184 on both C3bot molecules and K235 on the 2nd C3bot molecule and A48 on the 1st RalA molecule). The details of the crystallographic data are presented in table 5.1.

Figure 5.2 The alternative structure of a C3bot-RalA complex

Figure 5.2. The alternative C3bot-RalA structure. This possesses 2 C3bot and 2 RalA within the asymmetric unit. The two interfaces; the switch II interface and crystal interface are highlighted with boxes.
Figure 5.3 Ramachandran plot of the alternative C3bot-RalA complex

![Ramachandran plot](image)

**Plot statistics**

- Residues in most favoured regions: 563 (95.3%)
- Residues in additional allowed regions: 87 (13.2%)
- Residues in generously allowed regions: 6 (0.9%)
- Residues in disallowed regions: 14 (2.1%)

- Number of non-glycine and non-proline residues: 660 (100.0%)
- Number of real residues (excl. Gly and Pro): 236
- Number of glycine residues (shown as triangles): 47
- Number of proline residues: 18
- Total number of residues: 961

Based on an analysis of 11 structures of resolution of at least 2.0 Å with an R-factor no greater than 30%, a good quality model would be expected to have over 95% in the most favoured regions.

**Figure 5.3** A Ramachandran plot of the refined structure of the C3bot-RalA structure. Four residues lie in the disallowed regions (Ala184 on each C3bot, Ala 84 on RalA '1' and Lys 235 on C3bot '2') and 85.3% of residues lie within the most favourable region. The Ramachandran plot was calculated with PROCHECK (Laskowski et al, 1993).
Table 5.1 Table of Crystallographic statistics

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution range (Å)</td>
<td>50-2.40</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.98</td>
</tr>
<tr>
<td>Exposure time (s)</td>
<td>20</td>
</tr>
<tr>
<td>Number of images</td>
<td>120</td>
</tr>
<tr>
<td>Rotation per image (°)</td>
<td>1</td>
</tr>
<tr>
<td>Number of reflections measured</td>
<td>205496</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>30637</td>
</tr>
<tr>
<td>(I/\sigma(I)) (outer shell(^a))</td>
<td>17.82 (3.0)</td>
</tr>
<tr>
<td>Completeness (outer shell(^b)) (%)</td>
<td>85.8 (83.8)</td>
</tr>
<tr>
<td>(R_{symm}) (outer shell(^b)) (%)</td>
<td>5.4 (29.1)</td>
</tr>
<tr>
<td>(R_{crys}) (%)</td>
<td>23.3</td>
</tr>
<tr>
<td>(R_{free}) (%)</td>
<td>31.0</td>
</tr>
<tr>
<td>Average temperature factor (Å(^2))</td>
<td></td>
</tr>
<tr>
<td>RalA main-chain (side-chain)</td>
<td>37.0 (38.3)</td>
</tr>
<tr>
<td>RalA main-chain (side-chain)</td>
<td>40.2 (40.9)</td>
</tr>
<tr>
<td>C3 main-chain (side-chain)</td>
<td>37.2 (38.2)</td>
</tr>
<tr>
<td>C3 main-chain (side-chain)</td>
<td>49.8 (50.6)</td>
</tr>
<tr>
<td>Ligand</td>
<td>33.8</td>
</tr>
<tr>
<td>Solvent</td>
<td>42.7</td>
</tr>
<tr>
<td>RMSD from ideal values:</td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.007</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.247</td>
</tr>
</tbody>
</table>

\(^a\) Outer shell is 2.49Å – 2.40Å

\(^b\) \(R_{symm}=\Sigma_{n}\left|l(i) - \langle l(h)\rangle\right|/\Sigma_{n}\langle l(h)\rangle\), where \(l(i)\) and \(l(h)\) are the \(i\)th and the mean measurements of the intensity of reflection \(h\), respectively.

\(^c\) \(R_{crys}=\Sigma_{n}\left|F_o - F_c\right|/\Sigma_{n}F_o\), where \(F_o\) and \(F_c\) are the observed and calculated structure factor amplitudes of reflection \(h\), respectively.

\(^d\) \(R_{free}\) is equal to \(R_{crys}\) for a randomly selected 1526 (5.0%) reflections not used in the refinement.

5.4.2 Analysis of the alternative complex

The 2 C3bot molecules (C3bot 1 and C3bot 2) are both consistent with previously determined structures of C3bot with a mixed α/β core that possesses a five stranded mixed β-sheet (β1, β2, β4, β7, β8) stacking perpendicularly against a three stranded anti-parallel β-sheet (β3, β5, β6) with four helices (α1, α2, α3, α4) arrayed around the
sheets and a fifth helix (α5) flanking the five stranded sheet. Superposition with the C3bot from the original C3bot-RalA complex presented in chapter 2 yields a root mean square deviation of 0.6Å for 207 Cα atoms. The largest differences occur around the ARTT loop, 1.0Å between residues 204-212, and on a short loop region connecting strands β7 and β8; 1.9Å between residues 232-235.

The two RalA molecules (RalA 1 and RalA 2) are both consistent with the previously determined structures of RalA and other related GTPase molecules composed of a six stranded mixed β-sheet sandwiched between two helices (α1, α5) above the plane of the β-sheet, and three helices (α2, α3, α4) below. A superposition of RalA 1 with RalA from the original C3bot-RalA complex (Holbourn et al, 2005) reveals little difference in the position of the atoms with a root mean square deviation of 0.3Å for all of the Cα atoms.

5.4.3 C3bot-RalA interfaces

The asymmetric unit possesses two interfaces between C3bot and RalA molecules; the first is the interface first observed by Pautsch and co-workers between C3bot “1” and RalA “1” (Pautsch et al, 2005). This interface has a total buried surface area of 1663Å² and involves helices α3 and α4 from C3bot and helix α3 and the switch II region from RalA. This interface will hereafter be referred to as the “switch II interface”. The second interface is smaller and is observed between the α4 helix of RalA and the C-terminus of C3bot and can be seen twice in the unit cell: between C3bot “1” and RalA “2” and between C3bot “2” and RalA “1” (total buried surface area of 710.4Å²). This will be referred to as the “crystal interface”.

5.4.4 The switch II interface

The main interface between C3bot 1 and RalA 1 involves parts of helix 3 and 4 of C3bot fitting into a groove of RalA between helix 2 and 3 and making contact with the important switch II region (figure 5.4). This mode of binding and the residues
involved around the interface are consistent with the interface reported by Pautsch and coworkers and listed in tables 5.2 and 5.3. The interface contains a total of 11 hydrogen bonds and 81 van der Waal interactions involving 15 residues from C3bot and 14 residues from RalA.

Table 5.2 Potential hydrogen bonds between C3bot and RalA

<table>
<thead>
<tr>
<th>C3bot residue</th>
<th>RalA residue</th>
<th>Distance (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gln 92 ND2</td>
<td>Glu 99 O</td>
<td>3.19</td>
</tr>
<tr>
<td>Asn 93 O</td>
<td>Gly 23 N</td>
<td>3.12</td>
</tr>
<tr>
<td>Ile 97 N</td>
<td>Gly 71 O</td>
<td>3.10</td>
</tr>
<tr>
<td>Asn 98 OD1</td>
<td>Tyr 82 OH</td>
<td>2.44</td>
</tr>
<tr>
<td>Asn 98 O</td>
<td>Arg 79 NH1</td>
<td>2.52</td>
</tr>
<tr>
<td>Gly 99 O</td>
<td>Gln 110 OE1</td>
<td>3.12</td>
</tr>
<tr>
<td>Phe 100 O</td>
<td>Arg 79 NH2</td>
<td>3.09</td>
</tr>
<tr>
<td>Phe100 O</td>
<td>Gln 110 OE1</td>
<td>3.28</td>
</tr>
<tr>
<td>Asp 112 OD2</td>
<td>Tyr 75 OH</td>
<td>3.09</td>
</tr>
<tr>
<td>Arg 219 NH1</td>
<td>Gln 72 O</td>
<td>2.99</td>
</tr>
<tr>
<td>Arg 219 NH2</td>
<td>Gln 72 O</td>
<td>3.26</td>
</tr>
</tbody>
</table>

Table 5.3 Contacts between C3bot and RalA

<table>
<thead>
<tr>
<th>C3bot residue</th>
<th>RalA residue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gln 92</td>
<td>Glu 99, Ala 102</td>
</tr>
<tr>
<td>Asn 93</td>
<td>Ser 22 (2), Gly 23, Phe 107</td>
</tr>
<tr>
<td>Lys 94</td>
<td>Gly 23</td>
</tr>
<tr>
<td>Gly 95</td>
<td>Gln 72 (3)</td>
</tr>
<tr>
<td>Val 96</td>
<td>Ser 22, Gly 23 (2), Gly 71 (2), Gln 72, Phe 107</td>
</tr>
<tr>
<td>Ile 97</td>
<td>Gly 71 (2), Asp 74, Tyr 75 (2)</td>
</tr>
<tr>
<td>Asn 98</td>
<td>Arg 79 (3), Tyr 82 (4), Phe 107 (2)</td>
</tr>
<tr>
<td>Gly 99</td>
<td>Asp 106, Phe 107 (4), Gln 110 (4)</td>
</tr>
<tr>
<td>Phe 100</td>
<td>Arg 79 (2), Gln 110 (4)</td>
</tr>
<tr>
<td>Pro 101</td>
<td>Gln 110</td>
</tr>
<tr>
<td>Ile 105</td>
<td>Asp 74, Tyr 75 (2), Ala 76 (3), Arg 79 (2)</td>
</tr>
<tr>
<td>Glu 109</td>
<td>Tyr 75 (18)</td>
</tr>
<tr>
<td>Asp 112</td>
<td>Tyr 75 (3)</td>
</tr>
<tr>
<td>Arg 219</td>
<td>Gln 72 (3)</td>
</tr>
<tr>
<td>His 220</td>
<td>Gln 72</td>
</tr>
</tbody>
</table>

Numbers in superscript are the number of contacts between the listed residues

The area of C3bot involved in the formation of this interface is distant from the active site of C3bot. Although part of helix-3 is involved in the binding to RalA none of the residues that form part of the α-3 helix (Tyr 79, Asn87, Arg91) are involved (Holbourn et al, 2006) and when superposed onto the structure of wild type C3bot (PDB code 1G24) (Han et al, 2001) none of them have suffered a significant alteration in position or side chain conformation.
Unlike the original complex of C3bot-RalA (Holbourn et al, 2005) the binding region of RalA has changed dramatically and occludes the nucleotide binding site and has several bonds with residues on the important switch II loop. Gly71, Gln71 and Tyr75 are all part of the switch II loop and form the central segment of the interface, Tyr75 in addition to forming a hydrogen bond with Asp112 of C3bot is also involved in over 20 van der Waal interactions.

The involvement of switch II in this interface also results in a change in the nucleotide dependence of the C3bot-RalA interaction. In wild type RalA upon GTP binding switch II undergoes a large conformational shift and interacts heavily with the third phosphate group of the GTP molecule. However, when C3bot is bound as shown here it prevents switch II undergoing any movement and results in GDP dependant binding as reported by Pautsch and co-workers (Pautsch et al, 2005).
This interface is highly similar to an already determined structure (PDB code 2A78) (Pautsch et al, 2005) though there are some subtle differences in the hydrogen bonding patterns. In the reported structure there are two additional hydrogen bonds involving the RalA residues Glu97 and Asp106 that are missing in this structure. Additionally in this structure there is an additional hydrogen bond between Gln110 of RalA and Gly99 of C3bot. The largest change involves the C-terminal of C3bot that in the reported structure (Pautsch et al, 2005) is involved in van der Waal interactions with a segment of switch II of RalA. In our structure the C-terminal is involved in forming the crystal interface described below.

5.4.5 The crystal interface

The second interface that forms between C3bot 1-RalA 2 and RalA 1-C3bot 2 is much smaller and involves helix α4 of RalA packing against the C-terminus of the C3 molecules (figure 5.5). The residues involved in this interface are listed below in tables 5.4 and 5.5. This interface has a buried surface area of 710.4Å² and involves some of the important residues seen in the original C3bot-RalA complex (Holbourn et al, 2005). This interface is made up of the C-terminus of C3bot and part of β-strand 2 that stacks against the face of helix 4 of RalA as shown in figure 5.4.

This interface is small and contains only 3 hydrogen bonds and 11 residues involved in van der Waal interactions (5 from C3bot and 6 from RalA). This interaction does however involve helix α4 of RalA. This is the helix that forms the centre of the interface in the original C3bot-RalA structure and involved residues from this helix interacting directly with the ARTT loop of C3bot. In this complex two of the residues involved in forming the interface with the ARTT loop are involved here, Glu140 and Glu147.
5.4.6 The ARTT loop

The original complex between C3bot and RalA (Holbourn et al, 2005) demonstrated a novel conformation of the ARTT loop that involved a kink being formed and the side chain of becoming twisted to place it in the centre of the ARTT loop in the area normally occupied by the Cβ atom of Ala210 for the wild type (Han et al, 2001) and the initial reported structure around the switch II interface (Pautsch et al, 2005). In this structure the ARTT loop of both of the C3bot molecules resembles that of the wild type and does not possess the conformation seen in the previous C3bot-RalA structure. This would suggest that the ARTT loop only folds into that unique orientation upon the binding of RalA directly at the ARTT loop.
Table 5.4 Potential hydrogen bonds between C3bot and RalA

<table>
<thead>
<tr>
<th>C3bot residue</th>
<th>RalA residue</th>
<th>Distance (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asp 166 OD2</td>
<td>Glu 140 OE1</td>
<td>2.92</td>
</tr>
<tr>
<td>Met 244 O</td>
<td>Asn 144 ND2</td>
<td>3.03</td>
</tr>
<tr>
<td>Thr 246 O</td>
<td>Asn 144 OD1</td>
<td>3.02</td>
</tr>
</tbody>
</table>

Table 5.5 Contacts between C3bot and RalA

<table>
<thead>
<tr>
<th>C3bot residue</th>
<th>RalA residue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asp 166</td>
<td>Glu 140</td>
</tr>
<tr>
<td>Met 244</td>
<td>Ser 138, Glu 140, Glu 141</td>
</tr>
<tr>
<td>Gly 245</td>
<td>Asn 144</td>
</tr>
<tr>
<td>Thr 246</td>
<td>Asn 144</td>
</tr>
<tr>
<td>Asn 249</td>
<td>Glu 147, Gln 148</td>
</tr>
</tbody>
</table>

5.4.7 Implications for other proteins

The unusual interaction between C3bot and RalA was also seen for the related C3 exoenzymes C3lim and C3cer but not for the closely related C3stau (Wilde et al, 2002a). A sequence alignment of the C3 exoenzymes is shown below in figure 5.6 highlighting the residues involved in the interactions around the switch II interface. A similar alignment of RalA and related GTPases is presented in figure 5.7.

C3bot, C3lim and C3cer are all seen to interact with RalA whilst C3stau does not. In the original complex (Holbourn et al, 2005) there were clear protein sequence differences that could explain the results from amongst the C3 family. In this structure the residues around the switch II interface vary dramatically between all 4 C3 exoenzymes. The lack of binding observed by C3stau can be explained by serious differences at many of the binding residues. In the case of the charged residues C3botGln92 and C3botAsn93, C3stau instead has a serine and alanine respectively. Similarly at C3botGly99 C3stau has a lysine. The importance of the loss of this glycine has been observed for a Gly99Asp mutant that demonstrated drastically reduced binding (Pautsch et al, 2005).

For C3lim and C3cer that also show binding to RalA there are still differences between some of the residues at the interface. In the case of C3lim these changes are
fairly minor save for the exchange of asparagine 92 to alanine that would result in the loss of the hydrogen bonds and van der Waal interactions with Glu99 and Ala102 of RalA. Importantly C3lim maintains Gly99 that is important for formation of this interface. C3cer has several substitutions including Gln92 becoming methionine, Asn93 to histidine, Asn98 to aspartate and Gly99 to serine. These may result in the low level of binding seen between C3cer and RalA, the lowest of the three C3 proteins that do interact with RalA. Unlike the larger changes for C3stau none of these alterations is substantial enough to completely disrupt the interface.

The specificity seen amongst the GTPases is less pronounced. Although the sequence differences between RalA and the related Rho, Rac and Cdc42 GTPases are obvious and include the universal substitutions of Gly23 to alanine, Tyr82 to serine and Gln110 to glutamate the differences between RalA and Ras are limited. RalA and Ras share all of the residues around the switch II interface save for Glu99 which is a lysine in Ras. However the interaction between C3bot and RalA takes place involves the main chain carbonyl of Glu99 and so an exchange to lysine should not make such a large difference. Mutagenesis work however has suggested that the important difference between RalA and Ras is actually at Ala103 (aspartate in Ras) (Pautsch et al, 2005). Presumably the extended length of aspartate cause steric clashes with the nearby residues of C3bot (Gln92 and Asn93) and disrupt their important interactions around the interface region.
Figure 6.6 Sequence alignment of the C3 exoenzymes constructed using the web-based program T-Coffee (Notredame et al, 2000). The residues involved in hydrogen bonds at the switch II interface are highlighted in turquoise. The three residues involved in hydrogen bonds at the crystal interface are highlighted in grey.
Figure 6.7. Sequence alignment of the GTPases constructed with the web-based served T-Coffee (Motredame et al., 2000) highlighting in turquoise the residues involved in forming hydrogen bonds at the switch II interface. Residues involved in interactions at the crystal interface are highlighted in red.
5.4.8 Unanswered questions

While there is much mutagenic data (Pautsch et al, 2005) to support the formation of a complex centred around the switch II interface as the biologically significant one there are also several problems with this model.

The interactions around the switch II interface do block access to RalA\textsuperscript{Thr\textregistered46} on the switch I region of RalA that is glucosylated by \textit{C. sordelli} lethal toxin and so in agreement with that observation made by Wilde and co-workers (Wilde et al, 2002a). However the regions of C3bot involved in binding to RalA are distant (>20Å) from the active site and so do not offer a mechanism to inhibit Rho binding and ribosylation. Additionally Pautsch and co-workers (Pautsch et al, 2005) found that their "switch II" complex illustrated no changes compared with wild type C3bot upon binding of NAD (PDB code 2A9K) and so cannot explain the 5-fold increase in NAD glycohydrolysis observed for the C3bot-RalA interaction. Equally the switch II interface may only form when RalA is in its GDP bound form rather than in either the GDP or GTP bound forms as initially reported (Wilde et al, 2002a).

5.5 Conclusions

We have determined an alternative arrangement of a complex between C3bot-RalA that is centred on switch II of RalA. This complex is highly similar to an already published structure with a different space group with 2 RalA molecules and 2 C3bot molecules in the asymmetric unit. Whilst the interface around the switch II region is larger than the original complex centred on the ARTT loop it does not offer a means by which the ribosylation of RhoA can be inhibited, nor how the glycohydrolytic of NAD can undergo such a dramatic increase in rate. However the recent finding that this alternative arrangement can be found in the original complex due to crystallographic symmetry lends more support to this structure being the true biologically relevant complex and the implications of this will be discussed further in chapter 8.
6.0 The (almost) structure of the C3bot-RhoA complex

6.1 Aims of obtaining the structure of the C3bot-RhoA complex

The interaction between C3bot and RhoA is highly specific. This specificity is shared with the other members of the C3 family, save for C3stau that extends the specificity to RhoE (Wilde et al., 2001). In spite of mutagenesis data the exact nature of the interaction between C3bot and RhoA is still unknown. The aim of this part of the project was to determine the structure of the C3bot-RhoA complex and gain a complete understanding of the mechanism and specificity of the interaction.

6.2 Interactions between C3bot and RhoA

The RhoA-C GTPases have long been identified as the sole substrates for the C3 family of ADPRTs (Aktories et al., 1989; Sekine et al., 1989). The Rho GTPase family is a large one that acts as a key regulator of a wide range of cellular functions. The number of proteins regulated by RhoA amounts to approximately 1% of the human genome (Jaffe & Hall, 2005) making the Rho GTPases essential cellular switches. Their functions include, but are not limited to: regulation of actin cytoskeleton and stress fibres, regulation of NADPH oxidase, cell cycle control, secretion, apoptosis, transcriptional activation, cell adhesion, cell transformation and transcriptional activation (Wilde & Aktories, 2001; Etienne-Manneville & Hall, 2002; Aktories et al., 2004; Jaffe & Hall, 2005).

The Rho GTPases have a similar structure to other members of the GTPase superfamily with a central 6 stranded β-sheet surrounded by a variable number of α-helices, seven in the case of RhoA. RhoA is modulated by nucleotide exchange between an inactive GDP bound form and an active GTP bound form (Wilde & Aktories, 2001). The two switch regions which undergo large conformational changes upon GDP/GTP exchange are central to this process and switch I includes Asn41, the exposed asparagines that is the target for ribosylation by C3bot and its
relatives (Sekine et al, 1989). The location of the switch regions and Asn41 adjacent to the nucleotide binding site can be seen in figure 6.1 (Wei et al, 1997).

**Figure 6.1 The structure of RhoA**

![Image of RhoA structure with highlighted switch regions and Asn41]

**Figure 6.1.** The crystal structure of RhoA (PDB code 1FTN) (Wei et al, 1997) in its inactive GDP bound form. The switch regions are highlighted in gold, GDP is shown bound in the nucleotide binding site and the target residue of the C3 family of exoenzymes (Asn41) is shown in ball-and-stick form.

The areas of Rho that are necessary for ribosylation by C3bot have been studied extensively. Mutagenic and chimeric studies using Rho-Ras chimeras has confirmed that the N-terminal half of RhoA contains all of the residues necessary for C3bot recognition. As can be seen in figure 6.1 Asn41 is solvent exposed. Several other residues around the switch I region are also solvent exposed and mutagenesis of Val38, Phe39 and Val43 all resulted in inhibition of ribosylation by C3bot. Similarly a Ras mutant with the equivalent Ras residues mutated to those found in RhoA became a viable target for C3bot ribosylation (Wilde et al, 2000). Another residue
that may be important for recognition and ribosylation is Glu40 that may play a role in the transferase reaction by interacting directly with the 2'OH of the ribose ring in a similar manner to that of Asp696 in the recently determined PAETA-eEF2 structure (Jorgensen et al, 2005).
6.3 Methods

The expression and purification was based on methods previously described by Evans et al (Evans et al, 2004) and myself (Holbourn et al, 2005). Initial attempts to obtain a complex of C3bot and RhoA followed a similar route to that used to generate the C3bot-RalA complex. A second attempt was made using conditions that C3bot and RhoA were known to form a complex in, with the absence of NAD.

6.3.1 Expression and purification of C3bot

A C3bot-MBP fusion construct in BL21(DE3) cells was kindly provided by Dr. J.M. Sutton (Health Protection Agency UK). C3bot was purified exactly as described earlier.

6.3.2 Expression of RhoA

The human RhoA gene clone (as a pGEX-2T fusion) was a kind gift of Dr Hiroshi Koide, Department of Stem Cell Biology, Graduate School of Medical Science, Kanazawa University, Japan. The plasmid was then transferred to the host strain JM109 for expression. The clone was inoculated into 100ml of Terrific Broth supplemented with 0.5% glucose and 100μg/ml ampicillin and grown overnight at 37°C. The culture was diluted 1:50 into 11 of fresh medium and grown for 4 hours at 37°C until an OD600 of ~0.8 was reached. Expression was induced by addition of IPTG to a final concentration of 1mM and grown for a further 4 hours at 25°C. Cells were harvested by centrifugation (Beckman Avanti J-25).

6.3.3 Purification of RhoA

Cell pellets were resuspended in 50mM BisTris, 0.5M NaCl pH7.0 1mg/ml lysozyme and left on ice for 45 minutes. The cells were lysed by sonication
(Soniprep 150) and the cell debris removed by centrifugation (Beckman Avanti J-25).

The clarified cell lysate was applied to a GST-trap column (GE Healthcare) equilibrated with 50mM BisTris pH 7.0, 0.5M NaCl, and was eluted with 20mM Tris pH 8.0, 0.2M NaCl, 10mM reduced glutathione as a single peak. The peak was dialysed into 20mM Hapes pH 7.3, 150mM NaCl, 5mM MgCl₂, 2.5mM CaCl₂ overnight at 4°C. It was then treated with 1U thrombin (New England Biolabs) /100μg protein overnight at 4°C. RhoA was separated from the GST fusion tag by reapplication to the GST-trap column. The RhoA protein was concentrated in a 10kDa centrifugal concentrator (Amicon) and applied to a Superdex 200 gel filtration column (GE Healthcare) equilibrated with 20mM Hapes pH 7.3, 20mM NaCl, 5mM MgCl₂, 1mM EDTA. The purified sample was further concentrated to 10mg/ml in a 10kDa centrifugal concentrator (Amicon) and stored for crystallisation.

6.3.4 Crystallisation of the C3bot-RhoA complex

Attempts were made through two methods to obtain a stable C3bot-RhoA complex. The first method was simply mixing the proteins together in a 1:1 ration and incubating at 4°C. The second method was similar to the one used by Christian Wilde and co-workers and involves using an excess of RhoA and a specific binding buffer (Wilde et al, 2000).

6.3.4.1 Incubation at 4°C

For crystallisation RhoA and C3bot were incubated together overnight in a ratio of 1:1 giving a working concentration of 5 mg/ml. Crystallisation trials were performed as hanging drops using 2μl of protein with 2μl of reservoir solution in the drop. Structure screens I and II (Hamptons Research) were used to generate initial
crystallisation conditions. Small crystals were observed in condition 43 of structure screen I (0.2M (NH₄)₂SO₄, 30% PEG 4000).

6.3.4.2 Incubation with C3bot-RhoA binding buffer

Pure C3bot and RhoA were independently buffer exchanged into binding buffer, 50mM Hepes pH 7.3, 2mM MgCl₂, 1mM EDTA, 1mM DTT, 0.3mM PMSF, using a 10kDa spin column (Amicon). RhoA and C3bot were then mixed together in a ratio of 5:2 and incubated at 4°C for >6 hours. Crystallisation trials were performed at 4°C as hanging drops using 1.75µl of protein with 1.75µl of reservoir solution in the drop, giving final protein concentrations of 3.5mg/ml RhoA and 1.5mg/ml C3bot in the crystallisation drops. Structure screens I and II (Hampton Research) were used to generate initial crystallisation conditions.

6.3.5 Data collection and refinement

Diffraction data were collected from a single cryo-cooled (100K) crystal with 25% ethylene glycol on station PX14.2 of the synchrotron radiation source at Daresbury (UK) using an ADSC Quantum 4 CCD detector. The data were processed and scaled using the HKL2000 package (Otwinowski, 1993a; Otwinowski, 1993b). The symmetry and systematic absences were consistent with the C2 space group (unit cell dimensions a=104.8Å, b=75.7Å and c=102.5Å, β=102.5°) with two copies of a C3bot1-RhoA complex per asymmetric unit and the crystals containing ~53% solvent. Data reduction was carried out using the program TRUNCATE (CCP4, 1994).

The molecular replacement program PHASER (McCoy et al, 2005) was unable to find any copies of RhoA in the asymmetric unit however and subsequent molecular replacement with 4 copies of C3bot obtained a viable molecular replacement solution (Rₑₚₑₚ 28.3, Rₑᵣₑ 34.6 and a figure of merit of 70%). The space group was also consistent with one of the published structures of C3bot (Menetrey et al, 2002).
6.4 Results and discussion

6.4.1 Experimental results

Expression and purification of RhoA followed the same protocol as RalA and yielded ~6mg/l of culture. The crystals grown in 0.2M (NH₄)₂SO₄, 30% PEG 4000 at 4°C grew over the course of ~48 hours and are shown below in figure 6.2. Data collection at station PX14.2 of the SRS Daresbury was straightforward and the data was processed in space group C2 with the following unit cell dimensions: a=104.8Å, b=75.7Å, c=120.6Å, β=102.5°.

Figure 6.2 Potential crystals of the C3bot-RhoA complex

Figure 6.2. Crystals of the potential C3bot-RhoA crystals grown in 0.2M (NH₄)₂SO₄, 30% PEG 4000. Crystals were grown at 4°C and grew over ~48 hours.

Molecular replacement using AMoRe (Navaza, 1994), MolRep (Vagin & Teplyakov, 1997) and PHASER (McCoy et al, 2005) failed to find any copies of RhoA in the unit cell. However a molecular replacement solution with 4 C3bot molecules as search model was produced with an R_crys 28.3% and R_free 34.6%. Comparison with the published structures of C3bot found that the unit cell
dimensions were very similar to those of the C3bot structure published by Menetrey and co-workers \( (a=109\text{Å}, b=75.6\text{Å}, c=123.5\text{Å}, \beta=102.4) \) (Menetrey \textit{et al}, 2002) that crystallised in 80mM citric acid pH 3.0, 22.5% PEG3350, 100mM \( \text{Li}_2\text{SO}_4 \).

The crystallisation trials performed using the C3bot-RhoA binding buffer has failed to yield crystals to date.

6.5 Conclusions

We have managed to purify pure C3bot and RhoA. Attempts to determine the structure of a C3bot-RhoA complex have so far been unsuccessful although we have discovered a previously unreported crystallisation condition for C3bot.
7. *Streptococcus pyogenes* ADP-ribosyltransferase SpyA

7.1 Aims in purifying SpyA

SpyA from *S. pyogenes* has been identified as an ADPRT from sequence analysis and conserved motifs and more recently has been biochemically shown to ADP-ribosylate actin and other host cell proteins. We hope to purify, crystallise and determine the structure of SpyA and compare it with the structures of other members of the ADPRT family.

7.2 SpyA from *Streptococcus pyogenes*

*Streptococcus pyogenes* is a Gram positive pathogenic bacteria that is responsible for a host of pathogenic effects ranging from the potentially fatal toxic-shock syndrome and necrotising facitis to mild pharyngitis and tonsillitis. *S. pyogenes* is carried as a commensal bacterium in 5-15% of healthy individuals and only when the immune system is compromised or when it gains entry to the blood stream (e.g. from an infected wound) does it show a pathogenic effect.

In order for *S. pyogenes* to cause such a wide range of diseases it produces a variety of toxins and effectors to facilitate it pathogenic effects. These include: highly potent streptococcal superantigenic toxins such as SpeA and C (Barsumian *et al.*, 1978; Schlievert, 1993); a cysteine protease, SpeB, that cleaves a range of cellular targets (Ohara-Nemoto *et al.*, 1994); 2 potent cytolysins, Streptolysin S and O (Billington *et al.*, 2000; Palmer, 2001); and the clot dissolving streptokinase (Lottenberg *et al.*, 1987).

The SpyA ORF in the *S. pyogenes* genome was identified by two groups as an ADPRT from its sequence. In the first case from the genomic analysis that defined ORF Spy0428 as an ADPRT (Ferretti *et al.*, 2001); in the second ART_STREPY was identified from a PSI-BLAST search (Pallen *et al.*, 2001). In both cases the 250
amino-acid gene product was identified from the 3 conserved motifs common to all ADPRTs: (i) A potentially catalytic glutamate; (ii) a β/α region with a serine/threonine motif; and (iii) a conserved arginine (Domenighini & Rappuoli, 1996). This protein was termed SpyA (Coye & Collins, 2004) and is 250 amino acids in length with a molecular mass of 24.9kDa and a putative 30 residue signal peptide at its N-terminus that may be involved in pathogenesis (Pallen et al, 2001). The SpyA gene has a high degree of identity with both C3stau (25% identity and an additional 22% similarity) and C3bot (23% identity and an additional 23% similarity). SpyA has now been characterised as an ADPRT and is able to hydrolyse β-NAD and ribosylate poly-L-arginine (Coye & Collins, 2004).

Eukaryotic expression of SpyA has been linked with a breakdown in cell cytoskeleton (Coye & Collins, 2004), a tactic used by many pathogens used to resist phagocytosis (Gruenheid & Finlay, 2003). SpyA may thus play a role in preventing the phagocytosis of S. pyogenes (Coye & Collins, 2004). Although, like the related C3 exoenzymes, SpyA lacks an obvious eukaryotic binding domain it has been shown that Streptococcal NAD-glycohydrolase gains access to host cells via a pore formed by the cholesterol dependant streptolysin O (Madden et al, 2001) and SpyA may follow a similar route into eukaryotic cells (Coye & Collins, 2004).

The intracellular targets of SpyA are still being examined although radio-labelling experiments have highlighted actin, tropomysin and vimentin as possible targets of SpyA. All three of those proteins are heavily involved in the regulation of, or formation of, the cell cytoskeleton. ADP-ribosylation of actin prevents polymerisation of actin and results in the eventual destruction of the actin cytoskeleton (Wegner & Aktories, 1988). Tropomysin is another structural protein that binds to actin filaments in cells providing them with structural stability and vimentin forms intermediate filaments in eukaryotic cells (S. V. Perry, 2001).

Amongst the ADPRT family SpyA is unusual falling into two different categories. Based on its small size (24.9kDa), high pl (pl 9.1), sequence similarity and lack of a
SpyA is a novel ADPRT from *S. pyogenes* that ADP-ribosylates actin and possesses all of the motifs common to ADPRTs. Due its chemical characteristics, small size and sequence similarity it is probably a unique member of the C3bot family.
7.3 Methods

The SpyA gene in the pET15b vector (Novagen) was a kind gift from Dr. Carleen Collins, Department of Microbiology, University of Washington.

7.3.1 Expression of SpyA

The SpyA plasmid was transformed through heat shocking to BL21(DE3)codon+ expression cells (Novagen) and grown overnight on LB-gar plates supplemented with 100µg/ml ampicillin. Single colonies were used to inoculate 10ml of ZYM-505 auto-induction media overnight at 37°C (Studier, 2005) (1% N-Z amine, 0.5% yeast extract, 5mM Na₂SO₄, 2mM MgSO₄, 25mM (NH₄)₂Cl₂, 25mM KH₂PO₄, 20mM Na₂PO₄, 0.5% glycerol, 0.05% glucose, 50nM FeCl₃, 20nM CaCl₂, 10nM MnCl₂, 10nM ZnSO₄, 2nM CoCl₂, 2nM CuCl₂, 2nM NiCl₂, 2nM Na₂MoO₄, 2nM Na₂SeO₄, 2nM H₃BO₃) supplemented with 100µg/ml ampicillin. The overnight cultures were then diluted 1:1000 into 11 fresh ZYM-5052 media (1% N-Z amine, 0.5% yeast extract, 5mM Na₂SO₄, 2mM MgSO₄, 25mM (NH₄)₂Cl₂, 25mM KH₂PO₄, 20mM Na₂PO₄, 0.5% glycerol, 0.05% glucose, 0.2% lactose, 50nM FeCl₃, 20nM CaCl₂, 10nM MnCl₂, 10nM ZnSO₄, 2nM CoCl₂, 2nM CuCl₂, 2nM NiCl₂, 2nM Na₂MoO₄, 2nM Na₂SeO₄, 2nM H₃BO₃) supplemented with 100mg/l ampicillin and left to grow for overnight at 37°C. Partially soluble protein expression was observed on an SDS-page gel. Cells were harvested by centrifugation (Beckman Avanti J-25).

7.3.2 Isolation of inclusion bodies

Cell pellet was resuspended in ~40ml of lysis buffer (20mM Tris pH 8.0, 0.5M NaCl, 0.33mg/ml lyosyme) and stirred for ~30 minutes at room temperature. The cells were then lysed through two passes through the French press and inclusion bodies isolated through centrifugation (Beckman Avanti J-25). Both passes through the French press and centrifugation were performed at room temperature. Inclusion bodies were then solubilised in ~20ml of solubilisation buffer (50mM Tris pH 8.0,
0.5M NaCl, 6M Guanidine HCl) and stirred at room temperature for ~2 hours. Any remaining inclusion bodies were removed through centrifugation (Beckman Avanti J-25) and by passage through a 0.45μm syringe filter (Millipore).

7.3.3 Purification via Ni²⁺ affinity chromatography

The clarified inclusion bodies were loaded onto ~10ml of fast flow H6 affinity resin (GE Healthcare) equilibrated with solubilisation buffer and charged with 0.4M NiSO₄. The column was then washed with ~3 column volumes of wash buffer (50mM Tris pH 8.0, 0.5M NaCl, 7M urea) and the SpyA eluted by an isocratic step into elution buffer (50mM Tris pH 8.0, 0.5M NaCl, 7M urea, 0.5M imidazole). Fractions containing pure SpyA were pooled prior to refolding.

7.3.4 Refolding of SpyA

A 24 condition refolding screen based on an amalgam of commercially available was performed to obtain refolding conditions suitable for SpyA. 50μl of purified SpyA was added slowly to 1ml of each refolding condition and then left overnight at 4°C. Precipitate was removed using centrifugation and the supernatant containing refolded protein was visualised on an SDS-page gel.

The remaining pure SpyA was then dripped slowly over ~2 hours into 500ml of refolding buffer containing 50mM Tris pH 8.5, 9.6mM NaCl, 0.4mM KCl, 0.5% triton X-100, 0.05% Peg3350, 0.4M sucrose, 2mM MgCl₂, 2mM CaCl₂ and 5mM β-Mercaptoethanol and left for 24-48 hours without stirring at 4°C. This was then centrifuged at high speed for 1 hour to remove precipitate and other contaminants (Beckman Avanti J-25).
7.3.5 Purification of refolded SpyA

The refolded SpyA was applied by very slow overnight loading onto a 5ml Hitrap column (GE Healthcare) charged with 0.4M NiSO₄ and equilibrated with refolding buffer. SpyA was eluted as a single peak with 20mM Tris pH8.5, 120mM NaCl and 0.5M imidazole and stored at 4°C.
7.4 Results and discussion

7.4.1 Experimental methods

SpyA has proved a challenging and difficult protein to purify. Initial purification strategies focused on a soluble purification protocol. When expressed in autoinduction media SpyA overexpresses well and in large amounts. The expressed SpyA was found to be partially soluble. Soluble purification strategies using Ni\textsuperscript{2+} affinity chromatography followed by size exclusion chromatography failed to yield pure, homogenous SpyA suitable for crystallisation. Efforts to improve the purity of the soluble SpyA involved trials with SP-sepharose ion exchange chromatography to take advantage of the relatively high pI (pI 9.1) of SpyA, and trials with a Cibracon blue (GE Healthcare) dye column to take advantage of the NAD binding characteristic of SpyA. Reverse phase chromatography was also attempted but even this failed to yield pure SpyA, in addition solubility problems were encountered after lyophilisation of the sample after reverse phase chromatography.

It was decided that after the success of the refolding screen for several other proteins that inclusion body purification may yield better results. To increase the yield of insoluble SpyA the cell lysis steps: treatment with lysosyme, mechanical lysis in the French press and centrifugation were all performed at room temperature. This resulted in the vast majority of expressed SpyA being present in the inclusion bodies. Initially attempts were made to refold the SpyA before any purification steps were taken but this led to very poor refolding results and many problems with precipitation. The lack of refolding may have been due to other contaminating protein species present in the inclusion bodies and the refolding sample so it was though that purification of the solubilised proteins before refolding may remove many contaminants and increase the efficiency of refolding.

Initial attempts at this using pre-packed Hitrap (GE Healthcare) and His-select (Promega) columns resulted in the columns malfunctioning and losing the protein.
sample. Changing the protocol to non-prepacked FF H6 affinity media resulted in effective purification of the unfolded SpyA as see in the figure 7.1.

**Figure 7.1 Insoluble purification of SpyA**

<table>
<thead>
<tr>
<th>Hitrap Fractions</th>
</tr>
</thead>
<tbody>
<tr>
<td>66</td>
</tr>
<tr>
<td>45</td>
</tr>
<tr>
<td>35</td>
</tr>
<tr>
<td>29</td>
</tr>
<tr>
<td>24</td>
</tr>
<tr>
<td>20</td>
</tr>
</tbody>
</table>

**Figure 7.1**. SDS-page gel illustrating the partial purification of SpyA after histidine affinity chromatography performed in the presence of 7M urea.

The refolding screen is based on an amalgam of several commercially available refolding screens. Initial trials with the refolding screen resulted in 4 of the conditions giving positive results: (1) Condition A5 containing 50mM Tris pH 8.5, 9.6mM NaCl, 0.4mM KCl, 0.05% PEG3350, 0.5% triton X-100 2mM MgCl₂, 2mM CaCl₂, 400mM sucrose and 5mM β-Mercaptoethanol; (2) Condition A6 containing; 50mM Tris pH 8.5, 9.6mM NaCl, 0.4mM KCl, 0.5% triton X-100 and 5mM β-Mercaptoethanol; (3) Condition B1 containing 100mM Tris pH 8.5, 240mM NaCl, 10mM KCl, 0.05% PEG3350, 5mM β-Mercaptoethanol; and (4) condition B2 containing 50mM Tris pH 8.5, 240mM NaCl, 10mM KCl, 2mM CaCl₂, 0.4M sucrose and 1mM DTT. When done on larger scale than the 50μl trial scale buffer A5 gave better results than either A6 or B1 and buffer B2 was discarded as DTT can lead to problems with Ni²⁺ affinity columns. After refolding the sample was purified again using Ni²⁺ affinity chromatography. This second step served two purposes: (1) to remove any further contaminants; and (2) to concentrate the 500ml volume of
refolded SpyA to a more manageable volume. After refolding and the second stage of purification SpyA was also seen to dimerise (figure 7.2).

**Figure 7.2 Pure SpyA and dimerised SpyA**

Figure 7.2. An SDS-page gel illustrating the purity of SpyA after a refolding and a second round of histidine affinity chromatography. The peak from the Hitrap column contained a mixture of monomer and dimerised SpyA. After storage for ~24 hours at 4°C the SpyA had either formed a dimer or precipitated out of solution.

Whether this dimerisation is biological or merely due to the buffer conditions is unknown at this time, though no members of the C3 family of ADPRTs are known to exit as dimers.

Due to ongoing solubility problems no crystallisation has been possible as yet. The solubility problems may be due to the buffer exchange from refolding buffer into 20mM Tris pH8.5, 120mM NaCl and 0.5M imidazole following the second purification step, and eluting the SpyA in refolding buffer supplemented with imidazole may result in stable, soluble SpyA.
7.4.2 Homology modeling with the C3 exoenzymes

SpyA shows a high degree of sequence identity with C3bot (23%) and C3stau (25%) and a sequence alignment of the members of C3 family with SpyA show large regions of conserved sequence.

**Figure 7.3 Sequence alignment of SpyA with related ADPRTs**

The sequence alignment in figure 7.3 was constructed with T-coffee (Notredame et al., 2000) and highlights the key regions of the ADPRT family. A homology model of SpyA was constructed using C3bot [PDB code 1G24] (Han et al., 2001) as a
template and the SWISS-MODEL web server (Schwede et al, 2003). The SpyA model is shown below in figure 7.4.

**Figure 7.4** The homology model of SpyA

Like C3bot the model of SpyA consists of a mixed $\alpha/\beta$ core. A five stranded mixed $\beta$-sheet ($\beta_1$, $\beta_2$, $\beta_3$, $\beta_4$, $\beta_6$) stacks perpendicularly against a two stranded anti-parallel $\beta$-sheet ($\beta_5$, $\beta_7$) with four helices ($\alpha_1$, $\alpha_2$, $\alpha_3$, $\alpha_4$) arrayed around the three stranded sheet plus a fifth long helix ($\alpha_5$) flanking the five stranded sheet. The two smaller helices at the top of the model ($\alpha_6$, $\alpha_7$) are likely to only be loop regions in the actual structure. The structural motifs that are common to all of the C3bot-like
and Iota-like toxins are also conserved in SpyA, both in the sequence and in the structural model.

Like the ADPRT family SpyA has many of the conserved features common to ADP-ribosyltransferases. It has a catalytic glutamate residue (Glu 187) on the ARTT loop and in common with all of the actin ribosylating ADPRTs a second glutamate two residues upstream (Glu 185) as can be seen in figure 7.5.

**Figure 7.5 A close-up view of the active site of the SpyA model**

*Figure 7.5. A close-up view of the active site of SpyA showing the 5 motifs common to the α-3 helix ADPRTs. This includes the ARTT loop (in blue) containing the catalytic glutamate and a second substrate determining Glu/Gln two residues upstream. Although in the case of SpyA there is no conserved aromatic residue. The STT motif (purple), the Arg/His motif (yellow) and the PN loop (green) are also shown. The α-3 helix (red) contains a conserved tyrosine and asparagine but lacks the arginine seen in C3bot.*
It is likely that the change in this second glutamate from glutamine to glutamate is plays a part in the altered substrate specificity. All of the actin modifying toxins possess a Glu-x-Glu motif at the ARTT loop and alterations to this Gln/Glu have been seen to cause dramatic changes in substrate specificity, either from asparagine to arginine for C3lim (Vogelgesang & Aktories, 2006) or from cell surface to cytoplasmic targets in EctoART (Ritter et al, 2003). An E187A mutant has been shown to abolish both the hydrolysis and transferase reactions confirming Glu187 is the catalytic glutamate (Coye & Collins, 2004). Unlike the C3 or Iota like toxins however it appears that SpyA lacks a conserved aromatic residue equivalent to Phe209 of C3bot instead possessing a valine residue. This is thought to be critical for the action of both the actin modifying Iota toxin (Tsuge et al, 2003) and the Rho-modifying C3bot (Wilde & Aktories, 2001) it is possible that SpyA may have a slightly different way of recognising its substrate. A hypothesis partially born out by the observation that SpyA appears to target actin, tropomysin and vimentin (Coye & Collins, 2004). This broad substrate specificity has not been observed for the other actin modifying toxins. The other structural motifs are also intact in SpyA. It possesses an STS motif, although, in the case of SpyA this changed to an STT motif. The conserved Arg/His motif maintains the hydrophobic-aromatic-arginine pattern seen in the other C3-like amino acids and it also possesses the α-3 motif with a Y-X7-I-N sequence, though in common with the Iota-like toxins it lacks the arginine immediately after this motif.

7.5 Conclusions

We have managed to optimise a refolding based purification protocol for SpyA, a novel ADPRT from S. pyogenes. SpyA has been shown to be an ADPRT with broad substrate specificity and a high sequence similarity to the C3 family of ADPRTs. Small changes in the sequence, particularly in the normally conserved ARTT loop may account for this broad substrate specificity. Physical similarity to the C3 family whilst possessing enzymatic action more closely related to the Iota-toxin family has led to it being classified as a unique member of the C3bot family.
8. General discussion

8.1 Interactions between C3bot and RalA

We have managed to purify and determine the structure of two alternative complexes of C3bot and RalA alongside a structure of the first complex with NAD bound into the active site of C3bot. However in the initial complex determined, a symmetry related arrangement also gives rise to an interface that is the same as the "alternative complex" detailed in chapter 5. This gives reason to suspect that initial complex centred on the ARTT loop may in fact be a crystallographic artefact and not a true biological interface necessitating a close examination of the two complexes to investigate which is more likely to be the true biological interface.

The two complexes are highly different binding to vastly different areas of each molecule. The ARTT loop complex involves C3bot being bound directly at the active site completely blocking the 'face' of the molecule and obscuring the active site (Holbourn et al, 2005). By contrast the areas involved in RalA binding in the 'switch II' complex are distant from the active site (>25Å) (Pautsch et al, 2005). The different areas are highlighted below in figure 8.1A. Similarly the areas of RalA involved in the different modes of binding are distant from each other. In the original complex (Holbourn et al, 2005) the area is on the far side of the RalA molecule with respect to the nucleotide binding site. In the second complex the interface is formed in a groove formed by helix α3 and the switch II loop of the nucleotide binding site. In the switch II complex the position of the binding interface also results in the complex only forming being with GDP bound RalA. The areas that form the interfaces are highlighted below in figure 8.1B. Both of these different complexes, the original one based around the ARTT loop and the second one similar to that reported by Pautsch and co-workers (Pautsch et al, 2005) centred around the switch II region of RalA do not fully answer the questions posed by the initial biochemical experiments (Wilde et al, 2002a).
8.2 Biological basis of the competing interfaces

Each of the proposed models fits some of the criteria that the original biological work ascribed to the interaction between C3bot and RalA. The initial complex (Holbourn et al, 2005) described in chapter 2 offers a clear mechanism to bring about the inhibition of Rho ribosylation. The interaction of RalA about the ARTT loop obscures the front of C3bot and prevents access of other proteins, such as RhoA, to the active site, thereby inhibiting the ribosyltransferase reaction. The second structure however involves binding to the rear of the C3bot distant from the active site and does not involve any residues crucial to NAD binding. Equally the position of the RalA binding site towards the rear of the C3bot molecule leaves the active site open to solvent and would allow RhoA easy access to the binding site.
The position of the interface region in the second, switch II, model of the complex also offers no explanation for the increased rate of glycohydrolysis observed upon C3bot interacting with RalA (Wilde et al, 2002a). The original complex by interacting directly with the ARTT loop may be able to offer a plausible explanation for this. The unique conformation of the ARTT loop involves the side chain of Ser207 flipping into the centre of the ARTT loop and forming ionic interactions with additional residues of the ARTT loop possibly granting the normally flexible a level of rigidity. Also the PN loop in the structure detailed in chapter 4 possesses some subtle changes including a shift in the position of Gln182 that prevents it interacting with the Gln212 of the ARTT loop. However as both the ARTT loop interface and switch II interface are present in the asymmetric unit it is quite difficult to draw a meaningful conclusion from the small re-arrangements in the PN loop. A mutant replacing Gln182 with alanine and removing this same interaction has been seen to almost double the rate of glycohydrolysis without having a significant effect on the NAD binding strength (Menetrey et al, 2002).

However as in the complex presented in chapter 2 an electrostatic surface representation of the complex centred around the Switch II interface shows a huge change in the electrostatic surface around the active site of C3bot. This is shown below in figure 8.2. Similar to the ARTT loop complex there is a large change in surface electrostatic potential upon complexation with RalA that may account for the increase in NAD hydrolysis without any structural change in C3bot. Similar to the changes seen in the ARTT loop interface the potential of the NAD site becomes increasingly negative. As this interaction also occurs on the 'back' of C3bot relative to the active site it may be possible that this large shift in electrostatic charge may the be the reason that Rho ribosylation is inhibited as unlike in the ARTT interface where RalA blocks access to the active site there is no obvious method of inhibition in the complex based around the Switch II interface.
Figure 8.2 Electrostatic surface potential of the C3bot-RalA complex arranged around the switch II interface. The C3bot-NAD is on the left and the RalA is on the right.

The nucleotide dependant nature of the interaction also varies between the differing complexes. In the original complex the interaction takes place around the α4-helix of RalA distant from the nucleotide binding site and the important switch regions. This allows the RalA to be present in both GDP and GTP bound form in accordance with the initial biochemical findings (Wilde et al, 2002a). The work by Pautsch and co-workers reporting the structure of the switch II complex found that this complex between C3bot and RalA is only permissible when RalA is bound to GDP (Pautsch et al, 2005). The restriction of this complex to RalA-GDP is due to the involvement of switch II in the interface preventing it from moving into its GTP bound conformation due to steric clashes with C3bot. The mutagenesis work performed by Paustch and co-workers suggesting that loss of several of the residues of RalA involved at the ARTT loop interface had no effect on binding and their complex only forming with RalA-GDP may have been caused by their preparations only
forming the a single species of the complex, namely the complex centred around switch II.

### 8.3 Biochemical comparison of the interfaces

In addition to biological evidence there are numerous biochemical tools that can be used to analyse each interface and give an indication as to which interface is more likely to represent the true biological interface. One set of criteria has been developed that analyses the buried surface area between the sub-units and includes statistics on the shape and bonds at the interface (Jones & Thronton, 1996). This has been compiled into a web-based resource; the protein-protein interaction server, and the results for the ARTT loop interface and the Switch II interface are displayed below in table 8.1.

<table>
<thead>
<tr>
<th>Table 8.1 Protein-protein interaction server results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Protein interface parameters</strong></td>
</tr>
<tr>
<td>Accessible surface area (Å²)</td>
</tr>
<tr>
<td>% total surface area</td>
</tr>
<tr>
<td>Length &amp; Breadth (Å)</td>
</tr>
<tr>
<td>% Polar atoms</td>
</tr>
<tr>
<td>% Non-polar atoms</td>
</tr>
<tr>
<td>Hydrogen bonds</td>
</tr>
<tr>
<td>Gap volume</td>
</tr>
<tr>
<td>Gap volume index</td>
</tr>
</tbody>
</table>

The data listed in table 8.1 suggest several pertinent facts when comparing the two interfaces. The accessible surface area is the most obvious difference between the two interfaces. When compared to a structural comparison of the interfaces in the PDB the average accessible surface area (ASA) for proteins of 24kDa was found to be ~1000Å² (Jones & Thornton, 1996); close to the size for the switch II interface but much larger than the ARTT loop interface. The strength of the switch II interface compared to the ARTT loop interface is also apparent from its dimensions that are roughly similar to the ARTT loop interface but contains almost double the number
of hydrogen bonds (6 compared with 11). A second method of assessing the 'quality' of an interface is the "gap volume index" that gives a score to the shape complementarity of the two surfaces. This is defined as a ratio between the gap volume between the two molecules and the interface accessible surface area. For the hetero-complexes examined the gap index varied between 0.35Å to 5.17Å with a mean of 2.5Å. An assessment of the two interfaces using gap index as a measure of the quality would also suggest that the switch II interface is a chemically 'better' interface.

Analysis of the individual residue B-factors can also give an indication of which residues are involved in formation of the interface as when compared to the free C3bot or RalA there should be a corresponding drop in B-factors of the residues that are involved in forming an interface. Graphs of the B-factors comparing the wild-type C3bot and RalA with the structures presented in chapter 2 and chapter 5 are shown below in figure 8.3A and B. The B-factors for C3bot comparing the structures presented in chapters 2 and 5 and the original wild type structure (Han et al, 2001) do not show a great deal of difference when the important binding residues are compared. A similar trend is seen for RalA when the structures from chapter 2 and 5 are compared with the wild type GDP bound RalA (Nicely et al, 2004) when there is no stabilisation of the binding residues observed. Whilst this lack of change doesn't give an indication in favour of either of the possible interfaces the similarity between the B-factor plots of the two complexes could suggest that there is little or no difference between them as they both contain the Switch II interface and the addition of the ARTT loop interface does not make a substantive alteration to the stability of the complex.
Figure 8.3 B-factor plots of C3bot and RalA

A. B-factor plot of C3bot with the ARTT interface structure (blue), Switch II interface structure (green) and the wild type (PDB code 1G24) (Han et al., 2001). B. The B-factor plots of RalA with the ARTT interface structure (blue), Switch II interface structure (green) and the wild type (PDB code 1U8Z) (Nicely et al., 2004). The blue and green bands mark the residues at the ARTT loop interface and Switch II interface respectively.
8.4 Crystal polymorphism

The growth of both forms of the C3bot-RalA complex in the same crystallisation conditions is an unusual example of crystal polymorphism as polymorphism normally involves changing of some conditions in the crystallisation conditions. This change can be in pH, temperature (Jolles & Berthou, 1972), salts (Vaney et al, 2001) or changes in other crystallisation components (Berthou & Jolles, 1973). The nature of how this polymorphism comes about may suggest an equilibrium related formation in which the initial conditions favour the ARTT loop complex then as time progresses the equilibrium shifts to favour the larger and more stable complex presented in chapter 5.

8.5 Conclusions

Initially it was thought that the complexes presented in chapters 2 and 5 were competing complexes that answered different biological questions. It was only recently that a symmetry related interface identical to the Switch II interface was found in the initial published complex (Holbourn et al, 2005). The presence of the Switch II interface in all of the structures found to date and the strong mutagenic data suggests that it may be the biologically relevant interface.

The biophysical characteristics of the two interfaces also favour the Switch II interface as a true interface as the ARTT loop interface has a very small accessible surface area with respect to all other published complexes and a very large gap volume index that lies far outside the expected range. By comparison the Switch II interface falls into the expected range of surface area and gap volume index for a complex between two proteins of approximately 24kDa.

The comparison of the two possible interfaces has shown that mutagenic and crystallographic factors strongly favour the interface centred at “switch II” over the ARTT loop interface. However this is still troubling as despite this wealth of
evidence the Switch II interface does not show an obvious method through which the inhibition of Rho ribosylation comes about suggesting that there may still be elements of the interaction by C3bot and RalA that are poorly understood.

8.6 Future work

With the multiple forms of the interface between RalA and C3bot there is still some confusion as to which is the biologically active one, or, if the biological observations come about from a an equilibrium between the two forms. We are in the process of attempting to produce crystals of C3bot-RalA with GTP rather than GDP soaked into the binding site of RalA to confirm that the interaction between C3bot and RalA is nucleotide independent.

Several other biophysical methods of determining the biological complex also exist. Both nuclear magnetic resonance (NMR) and mass spectroscopy tools could be used to determine which parts of the C3bot and RalA molecules are involved in the interface. Mass spectroscopy has been successfully used to investigate the accessibility of surface residues in other protein complexes. The most common methods for using mass spectroscopy in this manner involve either using hydrogen – deuterium exchange (HDx) or labelling the reactive primary amine side chains of surface lysine residues with small groups such as acetate or biotin (Borch et al, 2005)(Suckau et al, 1992)(Scholten et al, 2006). In HDx the the protein is suspended in D₂O and mass spectroscopy can give information as to the parts of the molecule that are surface assessable due to hydrogen- deuterium exchange on the backbone amides. With labelling strategies exposed lysines become labelled by a small molecule, such as using acetic anhydride to acetylate exposed lysine residues (Scholten et al, 2006)(Suckau et al, 1992), in order that after tryptic digest fragments of the molecule were surface exposed show a change in their mass. NMR can also be used in a similar manner to probe intermolecular contacts. The most common method of NMR used to study protein interfaces and one that would be applicable to the complex between C3bot and RalA is “chemical shift perturbation
mapping". In this method one of the proteins is labelled with $^{15}$N and its $^{15}$N-$^1$H NMR profile is monitored as the second protein is added. The molecules that are involved closely at the interface will then show a degree of chemical shift allowing you to observe which part of the molecule is forming the biologically relevant interface (Zuiderweg, 2002). Ideally either of these methods may resolve the question and definitively illustrate which complex is formed in solution.

The fine details of the recognition and binding between C3bot and RhoA is still unknown. To achieve a stable complex between C3bot and RhoA suitable for crystallisation we are awaiting Carba-NAD (Slama & Simmons, 1988) (Slama & Simmons, 1989) from collaborators. This non-hydrolysable NAD analogue would bind to the C3bot active site and allow the C3bot to move into its NAD-bound form and successfully bind RhoA.

The unique C3-related ADPRT SpyA has now had a purification protocol optimised for it. The next stage will be to begin crystallisation studies to obtain diffraction quality crystals of SpyA. This may also lead to attempts to obtain crystals of a complex between SpyA and its substrates.
Appendix A

Basic introduction to crystallography

X-ray crystallography is the most commonly used tool for determining the three dimensional structure of biological macromolecules at, or near, atomic resolution. This has allowed understanding of drug binding, enzymatic mechanism and the underlying factors that are involved in protein-protein recognition. It is a complex process involving biochemistry, physics and mathematics. The science of crystallography improved over the years as more powerful sources of X-rays, advances in computing and automation of many of the steps of the process have allowed more structures to be determined and more information to be gleaned from the experimental data obtained. These steps are shown below in figure A1 and will be discussed throughout this chapter. The basic principles of X-ray diffraction and the steps involved in the determination of a structure will also be discussed.

Figure A1 The process of structure determination

![Figure A1: Flow diagram of the steps involved in macromolecular structural determination.](image-url)
Protein purification

Prior to crystallisation a large amount of homogenous pure protein must be acquired. With modern techniques this has been simplified by the availability of recombinant vectors that are designed with ready to use cloning sites and a wide selection of affinity tags to aid in the purification process. The selection of vector and subsequent method or purification can have a profound effect on the quality of protein obtained and the conditions of any resulting crystals. There are several factors to consider when selecting a purification method:

- Source of the protein
- Expression system
- Selection of fusion tag ability to cleave off the tag
- Method of purification

The choice of vector can affect the final product considerably. The most commonly used expression system is *E. coli* as it is cheap and straightforward. As it is a prokaryotic system it does not allow for any post-translational modification of the recombinant protein. Yeast and insect cell expression are more complex eukaryotic systems that allow post-translational modification of recombinant proteins and other difficult to generate eukaryotic proteins.

Purification is most commonly performed with affinity chromatography using a variety of fusion protein tags. The most common of these are hexa-histidine (His\(_6\)) tags, glutathione S-transferase (GST) fusion proteins and maltose binding protein (MBP) fusion proteins. The use of affinity tags allows a straightforward initial purification before more traditional chromatography methods such as ion-exchange, gel filtration or reverse phase chromatography are employed to obtain the \(~99\%\) purity required for crystallisation.

The purity of the protein produced is of paramount importance as contaminants or heterogeneity of the sample may produce a sample that will fail completely to
crystallise or produce crystals that do not diffract. This may come about from contaminating species that disrupt the orderly structure of the crystal and interfere in the orderly nature of the crystalline lattice (Rhodes, 2000; Blow, 2002).

**Crystallisation**

Crystals are ordered arrays of molecules held together by non-covalent interactions. The growth of ordered protein crystals of suitable size that diffract X-rays to a usable resolution is the limiting step of structure determination. The complexity involved in producing suitable protein crystals is understandable when one considers the variables that may affect crystal growth. These vary from more controllable variables such as the homogeneity of the protein, concentration of the protein, pH, temperature, salt and precipitant concentrations to less obvious variables such as cleanliness of the crystal trays, vibrations, gravity, convection and the age and source of the protein. The process of crystallisation involves the controlled precipitation of a protein from solution to form ordered crystals. It involves aggregation of protein molecules that form interactions that resemble those found in the eventual protein. This aggregate exists in an equilibrium that supports both growth and dissolution of the crystal nuclei until it achieves a critical size where only growth is supported and addition protein molecules occurs in an orderly manner expanding the crystalline lattice. These processes occur in supersaturated solution where the concentration of protein exceeds its solubility value. The supersaturation phase can be further divided into the labile zone where both nucleation and growth can occur and the metastable zone where only crystal growth occurs and these phases can be seen in figure A2A.
The most commonly used technique for growing protein crystal is through vapour diffusion techniques. In vapour diffusion experiments the protein concentration is increased through diffusion of water from a drop containing protein and mother liquor to a reservoir of more concentrated mother liquor. In hanging drop vapour diffusion shown above in figure A2B the drop containing protein is suspended from a siliconised coverslip above the reservoir and once supersaturation occurs both nucleation and crystal growth can occur simultaneously. Other methods of vapour diffusion techniques include the sitting drop and sandwich drop techniques. Due to the wide range of precipitants, salts, ions, additive and buffer conditions that can be used to induce crystal growth a common starting point is to use pre-made crystal screens such as the Molecular Dimensions structure screens (Jancarik et al, 1991) or Hamptons strategy screens (Jancarik et al, 1991) to try and find an initial condition that yields small crystals or crystalline precipitate. These pre-made screens of different conditions contain ready made solutions that span a wide range of common
buffers, salts and precipitants. Such a beginning can be used to obtain an initial condition to that can then be optimised to yield crystals that are suitable for X-ray diffraction experiments. Optimisation of crystal conditions commonly includes screening around the initial condition varying all of the conditions and components by small increments in an effort to improve the size or quality of the crystals (Blow, 2002; Rhodes, 2000).

The unit cell

Crystals are ordered three-dimensional arrays of molecules held together by non-covalent interactions. The smallest repeating unit that contains the protein is defined as the ‘unit cell’. This unit cell is representative of the crystal as a whole, the crystal being built of many repeated unit cells arranged together like bricks in a wall. Within the unit cell all atoms are assigned Cartesian co-ordinates (x, y and z) to define their position with one of the vertices selected as the origin and assigned the co-ordinates x=0.0, y=0.0 and z=0.0 (0,0,0). The edges of the unit cell are defined by 3 vectors a, b and c with lengths a, b and c. The angles separating these vectors are assigned as α, β and γ, a generalised unit cell is shown figure A3.

Figure A3: The unit cell

Figure A3 The generalised Unit Cell with edges a, b and c and angles α, β, and γ separating them.
The unit cell may also possess its own internal symmetry, known as crystallographic symmetry, and the smallest unit within the unit cell that can be rotated and translated to form the unit cell is called the asymmetric unit. The combination of the shape of the unit cell and the possible internal symmetries within the unit cell give rise to the crystallographic 'space group'. There are seven classified crystal systems that derive from the basic six dimensions of the unit cell as shown in table A1.

Table A1 The fourteen Bravais Lattice

<table>
<thead>
<tr>
<th>Name</th>
<th>Bravais lattice types</th>
<th>Lattice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Triclinic</td>
<td>P</td>
<td>$a \neq b \neq c, \alpha \neq \beta \neq \gamma$</td>
</tr>
<tr>
<td>Monoclinic</td>
<td>P, C</td>
<td>$a \neq b \neq c, \alpha = \gamma = 90^\circ \neq \beta$</td>
</tr>
<tr>
<td>Orthorhombic</td>
<td>P, C, I, F</td>
<td>$a \neq b \neq c, \alpha = \beta = \gamma = 90^\circ$</td>
</tr>
<tr>
<td>Tetragonal</td>
<td>P, I</td>
<td>$a = b \neq c, \alpha = \beta = \gamma = 90^\circ$</td>
</tr>
<tr>
<td>Trigonal</td>
<td>P</td>
<td>$a = b = c, \alpha = \beta = 90^\circ, \gamma = 120^\circ$</td>
</tr>
<tr>
<td></td>
<td>or R</td>
<td>$a = b = c, \alpha = \beta = \gamma &lt; 120^\circ \neq 90^\circ$</td>
</tr>
<tr>
<td>Hexagonal</td>
<td>P</td>
<td>$a = b \neq c, \alpha = \beta = 90^\circ, \gamma = 120^\circ$</td>
</tr>
<tr>
<td>Cubic</td>
<td>P, I, F</td>
<td>$a = b = c, \alpha = \beta = \gamma = 90^\circ$</td>
</tr>
</tbody>
</table>

These seven crystal systems give rise to 14 Bravais lattices depending on the arrangement of lattice points within the unit cell. When combined with internal symmetry operators such as rotation, translation, inversion and screw axes gives rise to 230 space groups though only 65 are allowable for proteins due to the chiral nature of amino acids. (Blow, 2002)
Bragg's Law

W.L. Bragg was the first to discover in 1913 that diffraction of X-rays by crystals may be compared with reflection of light by two parallel planes (Bragg, 1913). This has given rise to Bragg's Law which predicts the angle of reflection for any diffracted beam of X-rays by sets of parallel planes. For this diffraction to result in a 'diffraction spot' constructive interference must take place and the path difference in the waves must be an integral number of wavelengths.

To demonstrate this rule consider the construct given in figure A1.4 where two parallel planes (P1 and P2) are separated by an interplanar space (d) and illuminated with X-rays at an angle of incidence and reflection (θ) giving rise to the incident rays (1 and 2) and reflected waves (1' and 2').

For diffraction to occur the path difference between two waves must be an integral number of wavelengths apart to create constructive interference. In the construction above the path difference is equal to 2L.

$$2L = n\lambda \quad [1]$$

The length of L may also be expressed trigonometrically as;

$$L = d \sin \theta \quad [2]$$


$$n\lambda = 2d \sin \theta$$

As in X-ray crystallography the wavelength of the X-ray source (λ) is known, and the angle, θ, can be measured directly. The interplanar spacing can be calculated for each measured reflection.
Figure A4. Bragg's Law occurs when two parallel planes (P1 and P2) are illuminated with X-rays at an angle of incidence and reflection ($\theta$) and separated by an interplanar space (d) that is equal to an integral number of wavelengths.

Resolution

Resolution in crystallography is a measure of the minimum interplanar distances in the measured dataset. Due to the atomic scale of crystallography the angstrom ($\AA = 10^{-10}\,\text{m}$) is the unit of measurement used. The resolution also determines the amount of information that can be drawn from a crystallographic structure as summarised on table A2.
Table A2 Limits of resolution

<table>
<thead>
<tr>
<th>Resolution (Å)</th>
<th>Details Visible</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.0 Å</td>
<td>The outline of the molecule and positions of secondary structural features such as helices and sheets. The polypeptide backbone can be traces and with the amino acid sequence the peptide side chains can be fitted</td>
</tr>
<tr>
<td>3.0 Å</td>
<td>Bound water molecules, co-factors, ions and ligands can be observed</td>
</tr>
<tr>
<td>2.0 Å</td>
<td>Individual atoms are almost visible and details of solvent structure are visible</td>
</tr>
<tr>
<td>1.5 Å</td>
<td>Multiple confirmations of side chains are visible as are individual hydrogen atoms</td>
</tr>
<tr>
<td>1.0 Å</td>
<td></td>
</tr>
</tbody>
</table>

From reflections to structure

In X-ray diffraction the X-rays striking the crystal are diffracted by the electron clouds around each atom and arrive at the detector in an ordered array of spots known as the diffraction pattern as shown below in figure A5. These spots are termed 'reflections' as they are diffracted from the crystal as if reflected by planes of atoms, each spot representing a discrete plane of atoms.

The position of the spot can be measured accurately and assigned three indices in reciprocal space of the diffraction pattern. These indices, termed Miller indices, are designated by $h$, $k$ and $l$ with the central beam position taken as the origin with co-ordinates $(h,k,l) = (0,0,0)$ or $hkl =000$ and other reflections numbered as whole numbered co-ordinates counting out from this origin.
Figure A5: A typical diffraction pattern

Figure A5: An example of a diffraction pattern obtained at the SRS Daresbury. The black spots each represent one reflection and the large white circle in the centre is the beam stop that blocks the direct beam. The circles illustrate the resolution ranges as you move away from the direct beam. This data were collected with crystals of a complex between C3bot-RalA at station A1.6 at SRS Daresbury.

The greater the intensity of the X-rays reaching the detector the darker the spot and this may be measured by the optical absorbance. Thus from each reflection there are two piece of information available (1) the position, $hkl$, of the reflection and (2) the intensity, $I_{hkl}$, of the reflection.

The Ewald sphere is a geometric construction that can aid in visualising the reciprocal lattice and is constructed from a sphere of radius $1/\lambda$, drawn around the crystal position with the direct beam terminating at the origin of reciprocal space.
Figure A6: The Ewald sphere. A sphere of radius $1/\lambda$ is drawn around the crystal with the direct beam passing through the origin of the reciprocal lattice. A diffraction spot appears whenever the reciprocal lattice point contacts the Ewald sphere. Adapted from (Blow, 2002).

A diffraction ‘spot’ will be produced whenever the circle of radius $1/\lambda$ touches a reciprocal lattice point, an example of this is shown in figure A6. When a reciprocal lattice point comes into contact with circle Bragg’s law is satisfied as the angle between the incoming and outgoing beams is $2\theta$ and the length of the vector, $V$, is $1/d_{hkl}$. Extrapolated into 3 dimensions as a crystal rotates in an X-ray beam different reciprocal lattice points come into the contact with the Ewald sphere and as they do so give rise to a diffraction spot conforming to a particular set of equivalent parallel real-space $hkl$ planes (Rhodes, 2000; Blow, 2002).

The vector, amplitude and phase from a particular set of Bragg planes are termed the structure factor ($F$). The structure factors for the points of the crystal lattice correspond to a Fourier transform of the electron density at each point within the
unit cell. The structure factor equation relies on treating the reflections as waves, and is permitted due to the electromagnetic nature of X-rays. The equation for a simple wave form can be defined as follows:

\[ f(x) = F \cos 2\pi (hx + \alpha) \]

Where \( f(x) \) specifies the height of the wave at any particular horizontal position measured in wavelengths. \( F \) specifies the amplitude; \( h \) specifies the frequency and \( \alpha \) specifies the phase. In a structure factor calculation the contribution from every atom in the unit cell must be considered, however any waveform no matter the complexity can be described as the sum of many simple waveforms, where \( F_n \) is the contribution from atom n.; this is termed a Fourier series.

\[ f(x) = F_0 \cos 2\pi (0x + \alpha_0) + F_1 \cos 2\pi (1x + \alpha_1) + \cdots + F_n \cos 2\pi (nx + \alpha_n) \]

In Fourier theory any periodic wave function may be used to describe the function and another basic wave description may be used in place of \( f(x) = F \cos 2\pi (hx + \alpha) \);

\[ f(x) = \cos 2\pi (hx) + i \sin 2\pi (hx) \]

In this equation the waveform is split into a complex number in the form of \( a + ib \) and the term \( \cos 2\pi (hx) + i \sin 2\pi (hx) \) can be compressed into \( e^{2\pi i hx} \). However the equations for electron density are 3 dimensional wave forms so the \( (hx) \) term must be expanded to \( (hx + ky +zx) \).

In terms of electron density the Fourier series must reflect the contribution of all the atoms in the unit cell and must include an additional term, \( f_j \), that represents the scattering factor of the atom based upon a mathematical function treating the atom as a simple sphere of electron density and varies form element to element dependant on differing \( Z \) values (number of electrons) of the elements. This results in the structure factor equation having the following form:
\[ F_{hkl} = \sum_{j=1}^{n} f_j e^{2\pi i (hx_j + ky_j + zx_j)} \]

This describes the atomic structure factor of reflection \( hkl \) as a Fourier series with the amplitude contribution from its atomic type \( (f_j) \) and its phase contribution from its location in the unit cell \((x_j, y_j, z_j)\). The same equation may also be written as the sum of contributions of electron density;

\[ F_{hkl} = \sum \rho(x,y,z) e^{2\pi i (hx+ky+zx)} dV \]

Where \( \rho(x,y,z) \) is the average value of a volume element centred at \((x,y,z)\), and \( dV \) represents the total values of all the possible integral values of \( x, y \) and \( z \) in the unit cell. The volume elements taken are small enough so that the average density at each point becomes precisely equal to the value at that point.

Due to the reversible nature of a Fourier series this equation may be rewritten in terms of electron density as the transform of structure factors, where \( V \) is the volume of the unit cell;

\[ \rho(x,y,z) = \frac{1}{V} \sum_n \sum_k \sum_l F_{hkl} e^{2\pi i (hx+ky+zx)} \]

**X-ray sources and data collection**

The goal of data collection in an X-ray crystallography experiment is to collect a full data set accurately, rapidly and as efficiently as possible. To enable this there have been advances in technology at every stage of this processes including X-ray sources, detector design and data collection strategies and experimental design.

The most basic requirement of crystallography is an X-ray source that can produce intense, tightly focused monochromatic X-rays with wavelengths ranging from 0.5-1.6\( \text{Å} \). X-rays sources in protein crystallography are of one of two types; laboratory sources or Synchrotron radiation sources. Laboratory sources such as sealed tube or rotating anode devices accelerate electrons to approximately 45-50kV towards metal...
targets such as copper or molybdenum that then emit X-rays at a characteristic wavelength (CuKα = 1.54Å, MoKα = 0.71Å). These X-rays are then filtered through a crystal monochromator before being focused by curved mirrors. Laboratory sources however are of quite low intensity and low flux and the inability to tune the wavelength results in some techniques such as MAD or MIR not being easy to exploit. Synchrotron radiation sources result in X-rays that are very intense with a very high flux, high brilliance and tuneable wavelengths. This allows very fast, high resolution data collection from crystals and the ability to select suitable wavelengths for Isomorphous replacement or anomalous scattering experiments. Synchrotron radiation is produced from particle storage rings where very high energy electrons (2GeV) are accelerate to 99% of the speed of light and as they are curved around the ring emit high energy X-rays. These X-rays are further tuned by insertion devices such as undulators and wigglers that cause additional bending of the light to further increase the radiation. Through a series of monochromators and focusing mirrors X-rays of a selected wavelength can then be focused on the crystal.

Detection of X-rays is an important part of the crystallographic process and there are several types of X-ray detector available to collect the diffraction pattern. In modern X-ray experimental set-ups there are 3 types of detectors used: (1) Image plates are plastic sheets coated with phosphorescing crystals such as BaF:Eu²⁺ that absorb the X-rays and can be read by laser scanners. (2) Area detectors combine a wide dynamic scanning range with the ability to read directly without a scanning step. They are constructed from sets of perpendicular detector wires, one set is the anode, the perpendicular set the cathode and the X-rays hitting these wires are recorded. The most recent detector systems are (3) charge coupled devices (CCDs). These are solid state devices that are coated with phosphors that emit light when excited by X-rays. These are formed into arrays of μm sized pixels connected to optical fibres that directly count the photons.

Data collection can take place at either room temperature or at cryogenic temperatures. At room temperature crystals are mounted in thin walled glass
capillary tubes partially filled with mother liquor. Glass capillaries are used as glass has minimal X-ray absorption and can be corrected for. Cryogenic data collection is the normal mode of data collection now for most biological macromolecules. In cryogenic data collection the crystals are flash frozen in a stream of liquid nitrogen (80-100K) into an amorphous glass. A cryoprotectant is added to prevent the formation of ice crystal during data collection. The cryogenic temperatures also prevent the diffusion of free radicals from the sites of radiation damage. Commonly used cryoprotectants include glycerol, various weight polyethylene glycols and xylitol. Cryogenic collection has the advantage of limiting the amount of radiation damage to the crystal by the X-ray beam giving the crystal a longer lifetime in the beam and allowing a complete dataset to be collected from a single crystal. It also reduces the mechanical stress on crystal allowing the use of smaller and more delicate crystals and allows collection at higher resolutions due to reductions in atomic thermal factors. Cryogenic collection unfortunately does have several drawbacks; it can lead to an increase in crystal mosaicity due to small unfavourable changes in the crystal and lead to small changes in the unit cell volume which can lead to non-isomorphism between native and heavy metal data sets. Finally if the cryoprotectant is not optimal it may lead to the formation of ice rings that interfere with the diffraction information from the protein.

Data collection is a complex process and in order to collect a high quality data set it is important to formulate the correct strategy based upon the quality of the crystal and the X-ray experimental set-up. A preliminary examination of a single diffraction image can supply information on the unit cell, internal symmetry and the resolution limit of the crystal. This information can then be used to formulate a data collection strategy that will maximise the amount of reflections recorded. There are many factors that must be considered and these include: crystal to detector distance, exposure time, mosaicity and oscillation of the crystal. The crystal to detector distance determines the resolution the data can be collected to and has an effect on the spot size on the diffraction pattern. Longer exposure times can increase the intensity each spot allowing better recording of weak high resolution reflections.
Long exposure times can also lead to overloads as spots become saturated and longer exposure times lead to an increase in radiation damage to crystals. The molecules in a protein crystal do not form a perfect lattice but a mosaic of roughly aligned structures. The amount of disorder in a crystal is termed the 'mosaicity' and causes the tight collimated X-ray beam to become spread into a narrow cone. This results in some reflections being spread over two or more images. These reflections are termed partial reflection while reflections recorded as a single image are full reflections. As the mosaicity increases the amount of partial reflections increase and the spots are broadened, high levels of mosaicity result in spots overlapping and data being lost. The angle of oscillation required to collect a complete data set is related to the symmetry of the unit cell. As Friedel's law states that reflection $hkl$ is equivalent to $-h-k-l$ rotation of 180° will result in a complete dataset being collected. In higher symmetry unit cells the angle of rotation required is less as there are more symmetrically related reflections. In practice data collection is carried out over a larger range of oscillation than required to increase the redundancy of the data to improve the signal-to-noise ratio (Blundell & Johnson, 1976; Rhodes, 2000; Blow, 2002).

Data reduction and scaling

Data reduction and scaling involves processing the raw data collected to index each reflection in a diffraction pattern and then scaling all of the patterns into a complete dataset of indices and measured intensities. This process can be automated using one of several computer programs, the HKL package (Otwinowski, 1993a; Otwinowski, 1993b) and MOSFILM (CCP4, 1994) being the most commonly used. The HKL package consists of 3 separate programs: (1) DENZO- which carries out autoindexing, refinement and integration of each image. (2) XdisplayF- for visualising each mage and intensity measurements. (3) SCALEPACK- for scaling of the indexed dataset. With increased speed of detector readout and data processing the indexing and data collection can be performed in synchrony giving a real time assessment of the quality of the data being collected (Otwinowski, 1993a;
Otwinowski, 1993b). In autoindexing Fast Fourier Transforms are used to determine the values for indices in each direction, equivalent to finding the independent real space axes, and then once indices are determined in all three directions a set of possible lattices are selected based on the rules of symmetry laid down in the International Tables for Crystallography (Hahn, 1987). The unit cell lattice with the highest symmetry and least amount of distortion is used as an initial reference for the rest of the data processing. A complete dataset may include distinct blocks of data from different crystals or different areas of the same crystal, the intensity of the X-ray beam may vary over the time course of the experiment and radiation induced damage may adversely affect diffraction from the crystal. These factors result in the absolute intensities not being consistent from one block of data to the next. For this reason the data requires scaling. Scaling requires comparing reflections of the same index but in different blocks of data and rescaling the intensities so that they become identical relative to each other. SCALEPACK in the HKL suite performs this process and the quality of the data may be monitored through three statistical measures. The completeness is measure of how inclusive the data is and is calculated by dividing the number of unique reflections measured by the total number of unique reflections possible. A completeness of less than 95% may indicate a problem in the data collection strategy. The signal to noise ratio, or $I/\sigma I$, is the ratio of intensity divided by the error in intensity. The $I/\sigma I$ effectively defines the resolution limit as data with an $I/\sigma I$ of less than two cannot be reliably distinguished from the background and cannot be used. Finally the $R_{sym}$ gives an indication of the reliability of the data by comparing the mean difference between the symmetrical related reflections to the mean magnitude of the measured values and is calculated from the following equation:

$$R_{sym} = \frac{\sum_h \sum_i |I_h - I_{hi}|}{\sum_h \sum_i I_{hi}} \times 100$$

Where $I_h$ is the weighted mean measured intensity of the observations $I_{hi}$. Essentially the $R_{sym}$ is a measure of the disagreement between the symmetrical related intensities.
that should be identical. For a typical dataset of 2Å an acceptable $R_{sym}$ ranges from 10-12%. (Blow, 2002)

**The phase problem**

When X-rays are diffracted by a crystal they give rise to reflections that correspond to points in the reciprocal lattice. Each reflection therefore represents a wave with both amplitude and a phase. However when the diffracted ray hits the detector all of the phase information is lost and only the amplitude, intensity, of the wave can be recorded. This loss of phase information is known as the “phase problem”. There are several methods that can be utilised to solve this problem: (1) Molecular replacement (MR) that involves using a previously solved similar structure as a search model to obtain initial phase information (Rossmann & Blow, 1962). (2) Isomorphous replacement (IR) that involves soaking several heavy metals into the structure and estimating the phases from the changes in the vector components of the $hkl$ co-ordinates (Green et al, 1954). (3) Anomalous scattering involves the use of an atom that has its absorption frequency close to the wavelength of the X-ray source and phase information is gleaned from disruptions to Friedel’s law (Hendrickson, 1991).

**Molecular replacement**

Molecular replacement relies on starting with a search model that has a good degree of similarity to the target molecule. One of the most important factors that determine the success of molecular replacement is the quality of the search model chosen. A sequence identity of at least 30% is generally needed to find a solution. A well refined structure search model of at least 2.0-2.5Å will also result in calculated phases that are closer to the true (observed) phases.

The next step is to obtain a rotation matrix, $[R]$, and translation vector, $t$, to apply to the search model, $M$, to solve the solution of the target structure, $X$:
This search is a six dimensional search involving 3 rotational and 3 translational parameters. This can be solved by separating the search into 2 3-dimensional searches: one for the rotation matrix, the second for the translation vector as shown in figure A7.

Figure A7: Molecular replacement

![Figure A7: A schematic of the steps involved in molecular replacement. This involves rotating a search model by a rotation matrix \([R]\) and applying a translation vector \(t\) to find the target molecule.](image)

These searches are performed by calculating the structure factors for the search model and using the Patterson function. The Patterson function is used as it relies only on the squared intensities of the data and requires no phase information. An intramolecular Patterson map is independent of the positions of the atoms within the unit cell and comparing the intramolecular maps of the targets and search model can supply the correct rotation matrix to apply to the search model. Once the rotation matrix has been applied to the search model and intermolecular Patterson function is performed to supply the translation vector to place the search model in the correct position within the unit cell (Rossmann & Blow, 1962). There are several programs available that automate this process including AMoRe (Automated Molecular
Replacement) (Navaza, 1994), MolRep (Vagin & Teplyakov, 1997) and most recently PHASER (McCoy et al, 2005) all of which are part of the CCP4 program suite (CCP4, 1994).

**Isomorphous replacement**

Isomorphous replacement relies upon adding heavy atoms to the molecule in an effort to perturb the structure factors and from the manner in which they are perturbed make an informed deduction of the likely phase (Green et al, 1954). In practice two data sets are collected, one with the native protein and the second with the heavy metal derivative soaked protein. The atoms in this case are 'heavy' in the sense that they possess many more electrons and as atoms scatter proportionally to the square of the number of electrons a few heavy atom may have an obvious effect. For example a uranium atom will have a contribution equivalent to 225 carbon atoms as it possesses 15 times as many electrons. It is important when using heavy metal soaked crystals that the heavy metals replace are integrated into the crystal without significant disruption to the crystal lattice and that the only differences in the dataset are due to the heavy metals.

By treating structure factors as vectors it is possible to construct the following equation in which the structure factors for the heavy metal derivative, $F_{ph}$, represents the structure factor of the heavy metals, $F_h$, plus the structure factors of the protein, $F_P$. This vector addition results in two possible phases for $F_P$ if the vectors are added together to form a pair of triangles as shown below in figure A8.

To resolve the ambiguity between the two solutions a second derivative may be prepared that binds to the protein at different sites to the first one, in this case the technique is labelled multiple isomorphous replacement (MIR). The information derived from the second heavy metal will prove consistent with only of the two possible solutions for the phase. As shown in figure A8B adding a second
derivative, $F_{H2}$, gives two possible phases for the protein, $F_{p1}$ and $F_{p3}$, however only one of these will be in agreement with the phases obtained from the first soak.

**Figure A8: Vector summation for isomorphous replacement and anomalous scattering**

In practice errors in the model for the heavy atoms in the derivative crystal, errors in measuring the amplitudes of the structure factors and tiny changes to the crystal lattice mean that the vectors do not exactly meet up and the points where they meet become smeared. The ambiguity is related to lack of closure between the vectors and is monitored by the value $R_{Cullis}$. A high value of $R_{Cullis}$ may indicate that the derivative data set is unusable (Rhodes, 2000; Blow, 2002).

**Anomalous scattering**

Anomalous scattering/ dispersion also relies on certain properties of heavy metals. In anomalous scattering it is the small changes in phase and amplitude that occur when the wavelength of the X-rays lie near the absorption edge of the heavy metal. It is this small change in both amplitude and phase that gives the technique its name. The lighter atoms of the protein do not exhibit this behaviour as their absorption edges lie far from the range of wavelength used in crystallography. The most
common use of this technique now relies on introducing selenomethionine into the protein structure in place of methionine as the selenium will give strong anomalous dispersion within the range of wavelengths commonly used in synchrotron sources (Hendrickson, 1991).

The effect of anomalous scattering is to break Friedel’s law as the amplitude of the mates is no longer equal. That is the amplitude of \((h \, k \, l) \neq (-h -k -l)\). This alteration in the amplitudes in the structure factors of Friedel pairs can be illustrated by vectors on a Harker diagram in a similar manner to that of Isomorphous replacement. By performing the data collection at several wavelengths around the absorption edge we can build up a Harker diagram with multiple solutions analogous to that built in MIR. Using multiple wavelengths to obtain the phases is called multiple-wavelength anomalous dispersion (MAD) (Blow, 2002) (Rhodes, 2000).

**Model building and refinement**

The phases obtained from molecular replacement, isomorphous replacement or anomalous scattering can then be used to construct a model of the structure in a molecular graphics programs such as O (Jones et al, 1991) or COOT (Emsley & Cowtan, 2004). This allows an electron density map of the phases to be displayed and a model of the protein structure to be fitted into it. In the case of molecular replacement this can be done fairly quickly if the search model used is sufficiently similar to the target protein. In the case of novel proteins or proteins that have no suitable relatives then the entire protein must be fitted carefully into the electron density map. This can however now be automated by several programs such as ARP/wARP (Lamzin & Wilson, 1993) or SOLVE/RESOLVE (Terwilliger & Berendzen, 1999; Terwilliger, 2002). This initial model is then ready for refinement.

To improve the quality of the electron density and obtain a high quality interpretable structure the model obtained from molecular replacement, MIR or MAD must be refined. Refinement is a cyclical process in which the aim is to improve the
agreement between the observed data and the atomic model. Iterative cycles continue to improve the fit between the observed and calculated data allowing for improvement in phases and the observation of finer details of the structure. Every atom has at least 4 parameters (an x, y and z co-ordinate, B-factor and sometimes the occupancy) which leads to a poor observation: parameter ratio and overfitting of the data is possible so validation is an important aspect of refinement. The traditional method of measuring validation is by monitoring the R factor, the fractional agreement between the observed and calculated structure factors.

\[ R_{\text{cryst}} = \frac{\sum |F_{\text{obs}}| - |F_{\text{calc}}|}{\sum |F_{\text{obs}}|} \]

Where \( F_{\text{obs}} \) represents the structure factors derived from the recorded intensities and \( F_{\text{calc}} \) represents the calculated structure factors of the current model. The standard technique to avoid overfitting the data is to exclude a test-set of data (5-10%) from refinement. This test set is then used to calculate \( R_{\text{free}} \) which uses only this sub-set of data and avoids the problem of overfitting leading to a more valid model. This process, known as cross-validation, was first introduced by Alex Brünger (Brünger, 1992).

The target for crystallographic refinement is the search for a global energy minimum and can be expressed by the following equation:

\[ E = E_{\text{chem}} + w_{\text{xray}}E_{\text{xray}} \]

Where \( E_{\text{chem}} \) is the empirical potential energy for all the atoms describing covalent and non-covalent interactions, \( w_{\text{xray}} \) is the weighting given to each term and can be critical in achieving a valid model and \( E_{\text{xray}} \) represents the difference between the observed and calculated diffraction data.

\[ E_{\text{chem}} = \sum k_b(r-r_0)^2 + \sum k_d(\theta-\theta_0)^2 + \sum k_p \cos(m\varphi - d) + \sum k_{\omega}(\omega-\omega_0)^2 + \sum (ar^{-12} + br^{-6} + cr^{-1}) \]

bonds angles dihedrals chiral, planar atom pairs
The parameters of $E_{chem}$ stem from experimental and theoretical observations and by introducing these restraints it is possible to improve the ratio of refined parameters: observations as the restraints introduced force a parameter to have certain values or set up relationships between different parameters.

In order to improve the model and find the global energy minima of the target function complicated mathematical optimisation programs are used. The two main methods of optimisation employed in crystallography software use either Least Square (LSQ) method or a maximum likelihood model for optimising the target function. LSQ moves the target function to the weighted sum of squares of the differences between the observed and calculated quantities. Though LSQ methods may have problems with the many local minima of the target function and rely on the assumption that initial phase is correct. Maximum likelihood selects the model that is most consistent with the observed data, the consistency being a ratio between the probable and all the possible models. In order to utilise a maximum likelihood optimisation method it is important to account for all sources of error, including coordinate errors, missing atoms and errors in the B-factors. The B-factors represent the thermal motion of each atom and large B-factors can indicate errors in the model.

Practically this optimisation is achieved through simulated annealing. Annealing is a physical process in which a solid is heated to a sufficient temperature to form a liquid of randomly arranged particles and then slowly cooled in order for the particles to order themselves in the lowest energy state. In simulated annealing the model is ‘heated’ to a temperature of 5000K and then slowly cooled in small steps to fall into the lowest energy state to improve the quality of the model. The most common refinement programs used are REFMAC (Murshudov et al, 1997), as part of the CCP4 suite, (CCP4, 1994) and CNS (Brünger et al, 1998). With each cycle of refinement the parameters defining each atom are brought closer to the observed values. Each atom may be defined by at least 6 parameters: three spatial co-ordinates
(x, y, z), the scattering factor, the occupancy and the B-factor. Neither of these optimisation protocols can correct large errors in the model however. In order to correct gross errors manual rebuilding of the model in molecular graphics such as O (Jones et al, 1991) or COOT (Emsley & Cowtan, 2004) must be performed with aid of the electron density maps calculated from a given set of phases. The electron density maps normally used are the $2F_{\text{obs}}-F_{\text{calc}}$ and the $F_{\text{obs}}-F_{\text{calc}}$ maps. The $F_{\text{obs}}-F_{\text{calc}}$ highlight areas in the electron density that show the greatest disagreement between the observed and calculated phases and can also be used to identify the location of solvent atoms, ions or ligands that are present in the crystalline lattice. Solvent atoms can be added automatically to a model during refinement by the WATERPICK module of CNS (Brünger et al, 1998). Ligand molecules and ions generally have to be added manually and the atomic models and parameters of common ligands and ions can be accessed from web based databases such as the Hicup web-server (Kleywegt & Jones, 1998). Manual rebuilding to correct gross errors leads to an improved model that can then be further refined to give a better agreement between the observed and calculated data and lead to an improvement in the R factor. This iterative process of refinement followed by manual rebuilding continues until no further rebuilding can be performed and the structure has a satisfactory R-factor and satisfies a range of validation criteria. Satisfactory values for both $R_{\text{cryst}}$ and $R_{\text{free}}$ are approximately 0.2 for $R_{\text{cryst}}$ and 0.25-0.28 for $R_{\text{free}}$ (Blow, 2002).

Structural analysis, validation and deposition

Once a model structure has been refined to a suitable $R_{\text{cryst}}$ there are several programs and methods available to ascertain the validity of the structure. The most commonly used method to evaluate of the geometry is the Ramachandran plot. This is a method to assess the $\phi$ and $\psi$ angles of the polypeptide backbone to ensure that they lie within the allowed ranges to form secondary structure elements. An indicator of the accuracy of the structure may also be taken from the root mean square deviation (RMSD) of bond angles and bond lengths from the ideal. In the
case of a well refined structure the RMSD of the bond lengths should be 0.005-0.010Å. The program PROCHECK (Laskowski et al, 1993) assesses various factors of the overall geometry of the structure including bond lengths, bond angles, torsion angles of side chains and planarity of aromatic rings. The role of PROCHECK in this is to assess how closely the values of these parameters in the model match those of the ‘ideals’ and highlight residues or parts of the model that have chemically unfavourable parameters and may indicate a poor model. The web based servers “What if” (Vriend, 1990) and “MolProbity” (Davis et al, 2004) are alternatives to PROCHECK and validate the structure based on the same chemical criteria.

If these all point to a valid and structurally sound model then the functional analysis of the structure can be performed. Examination of the intra- and intermolecular interactions, ligand binding sites and superposition of the native structure over mutant or wildtype structures yield a large of amount important information on protein structure and function. Additionally structures involving large multi-domain proteins can illustrate important information about domain folding and domain architecture while structures of protein complexes can give important insights into the nature of protein-protein interactions and the protein-protein recognition. To assist in this process there are many programs that are designed to assist in this process such as parts of the CNS program (Brünger et al, 1998) or programs such as CONTACT and AREAIMOL in the CCP4 program suite (CCP4, 1994).

Once a structure has been completed it can then be submitted to the Protein Data Bank (PDB) (Bernstein et al, 1977), which is a global repository of all X-ray, NMR and theoretical models. Deposition normally involves depositing the atomic coordinates, structure factors and experimental details with the PDB through an online deposition program such as AutoDep.
Appendix B

Bacterial pathogens that utilise Type III secretion systems

Enteric pathogens lead to very large amounts of morbidity and mortality worldwide and are a major focus for microbial, genetic and medical research. These gram negative gastrointestinal pathogens have evolved a large number of strategies designed to withstand the diverse responses of the immune system, rapidly colonise a host and cause disease. Genetic analysis of pathogenic bacteria has shown that many of the proteins and virulence factors responsible for harmful effects of these enteropathogens are organised into genetic clusters termed pathogenicity islands and it is the presence of these pathogenicity islands that are linked to the disease causing effects of the bacteria. These pathogenicity islands may be shared through horizontal gene transfer allowing many species of otherwise unrelated bacteria to harbour closely related virulence genes. The type III secretion system (TTSS) is an example of this possessing an elaborate and complex secretion system that is found in organisms as distantly related as the *Yersinia* and *Erwinia* species (Hueck, 1998). The type III secretion system is an elaborate organisation of genes that results in a contact dependant 'syringe' that injects virulence factors into host cells directly crossing the bacterial inner and outer membrane and the host outer membrane in one step. This enables pathogens to directly inject effector protein into the cytosol of the host cell where they may interact with their cellular targets resulting in the symptoms of pathogenesis. This specialised system is large with over 20 proteins forming the secretion apparatus and are conserved across the whole range of bacteria possessing this system suggesting a common ancestor (Hueck, 1998) (Galan & Bliska, 1996). The range of bacteria that posses this TTSS is very large and includes many human pathogens that are responsible for highly dangerous diseases such as the *Yersinia* spp. (bubonic plague and tuberculosis), *Salmonella* spp (typhoid fever and gastroenteritis), *Shigella flexneri* (dysentery), enteropathogenic (EPEC) and enterohemorrhagic (EHEC) *Escherichia coli*, *Burkholderia pseudomallei* (mellidosis), *Pseudomonas aeruginosa*, *Chlamidia pisitachii* (chlamydia), the
animal pathogen *Heamophilus somnus* and several plant pathogens including *Pseudomonas syringae*, *Ralstonia* spp, *Erwinia* spp and *Xanthomonas* spp. Type III secretion (TTSS) is the most complex of the bacterial secretion systems. It allows bacteria to in one step translocate toxins directly into the cytosol of a eukaryotic cell bypassing the periplasm and the eukaryotic membrane entirely. The TTSS was first discovered in *Yersinia* (Rosqvist *et al*, 1994) and has been extensively studied in the three *Yersinia* species, *Y. pestis*, *Y. pseudotuberculosis* and *Y. enterocolitica*. The TTSS apparatus is highly complex and features cytoplasmic, transmembrane and extracellular domains (Blocker *et al*, 2001). The cytoplasmic and transmembrane domains combine to form the basal body that is responsible for protein export and provide the energy for the transport of the effectors into the eukaryotic cells. The basal body is composed of a large number of different proteins, 24 in *Yersinia* (Cornelis *et al*, 1998) of which 11 are common to nearly every gram negative bacteria possessing a TTSS (Hueck, 1998). Many of the proteins common to the basal body also exhibit similarity to proteins found in the bacterial flagellar system (Galan & Bliska, 1996; Hueck, 1998). The basal body also possesses an ATPase that is required to drive the proteins through the needle complex (Cornelis, 2000b; Hueck, 1998). Is common to all of the TTSS found to date and is one of the most highly conserved constituents of the TTSS apparatus. In *Yersinia*, the ATPase, YscN shows a high degree of similarity to the α and β sub-units of the F₁ component of the bacterial F₀F₁ proton translocating ATPase. Although its precise location is still unknown its similarity to the soluble F₁ component of the ATPase suggests that it is cytosolic and associates with the membrane bound proteins that form the secretion apparatus (Hueck, 1998). The basal body itself is comprised of two upper and two lower rings that cement a long needle to the bacterial membrane and an example of this shown in figure B1.1. The basal body structure has a pore through the centre of it, approximately 25Å in diameter that runs the full length of the needle through the pore at the other end gaining direct access to the eukaryotic cytosol. All effector molecules, either fully or partially unfolded must pass through this pore. Although the basal body is composed of many proteins there are believed to be three main proteins that form the bulk of the assembly. These were originally identified from
the *Salmonella* TTSS apparatus (Kimbroough & Miller, 2000; Sukhan *et al.*, 2001; Yip *et al.*, 2005) and include InvG, PrgK and PrgH. InvG which has orthologues in *Yersinia* (YscD), *Shigella* (MxiD) and EPEC (EscC) is a member of the secretin superfamily and forms a large multimeric annular complex in the outer membrane that can mediate protein passage (Crago & Koronakis, 1998). PrgK similarly has orthologues present in other TTSS bacteria; MxiJ (*Shigella*), YscJ (*Yersinia*) and EscJ (EPEC). PrgH however has one orthologue found so far, MxiG in *Shigella*. PrgH and PrgK are both believed to be associated with the inner membrane of the bacteria and form the base of the basal body though little is known about PrgH. The structure of EPEC orthologue of PrgK, EscJ, has been determined (Yip *et al.*, 2005) and has been modelled as a 24 monomer ring structure with extensive regularly spaced charged regions that may act as anchoring points for the other transmembrane components of the basal body in EPEC cells such as EscR, EscS, EscT, EscU and EscV. A deep negatively charged trench on the periplasmic side of the ring may also act as an anchor for the needle-like superhelix of EscF that forms the syringe like needle that protrudes from the membrane into the extracellular environment (Yip *et al.*, 2005). The dimensions of the needle itself vary from bacteria to bacteria and in EPEC the length can vary from 75-260nm with a diameter of 10-12nm (Daniell *et al.*, 2001) though the needle itself is formed from a superhelix of a single protein. In *Yersinia* the length of the needle formed by monomers of YscF is controlled by YscP that acts as a molecular ruler to halt the extension of the needle when YscP is fully extended from tip of the nascent needle to the base of YscP that is anchored in the membrane (Joumet *et al.*, 2003). The NMR structure of the needle protein, BsaL, from *Burkholderia pseudomallei* has been determined (Zhang *et al.*, 2006) and is seen to be a small 2-helix bundle stabilised by inter-helix hydrophobic contacts that can be modelled to form a large superhelix needle structure. The needle may also carry an extension at the tip that is involved in the penetration of and pore forming at the eukaryotic membrane. The structures of two extension proteins, EspA from EPEC (Yip *et al.*, 2005) and LcrV from *Yersinia* (Derewenda *et al.*, 2004) has been determined. EspA is part of the needle extension proteins in the EPEC needle whilst LcrV acts to insert the tip of the
Yersinia translocon into the eukaryotic membrane and form part of the pore through the eukaryotic membrane (Derewenda et al, 2004).

Figure B1 Schematic diagram of a typical TTSS apparatus

---

Though many of the components of the TTSS apparatus are known there is little structural information on most of them as the majority exist in multimeric states or are membrane associated. A generic TTSS needle is shown in figure B1 alongside the determined structures of the known TTSS components. While the components of the TTSS apparatus have shared sequence and structural homology the range of virulence factors and effectors and chaperones vary widely between bacterial species. In recent years the structures of many chaperones and TTSS effectors have
been determined showing the diversity of the TTSS. A selection of effectors and chaperones are shown below in figures B2 and B3 respectively. Bacterial specific effectors and chaperones are discussed in more detail below.

**Figure B2 A selection of TTSS effector molecule structures**

![Figure B2](image)

Figure B3 A selection of TTSS effector chaperone structures

*Figure B3.* A selection of TTSS chaperone molecules. These include the *Yersinia* chaperone SycT [PDB code 2BHO] (Locher et al, 2005), the *P. syringae* chaperone AvrpPhF [PDB code 1S21] (Singer et al, 2004) and the generic chaperone Spa15 [PDB code 1RY9] (van Eerde et al, 2004) from *Shigella flexneri.*

**TTSS in *Yersinia spp.***

There are 3 *Yersinia* species that have been found to be pathogenic to human and rodents. They are: *Y. pestis* the causative agent of bubonic plague; *Y. enterocolita* that causes a broad range of gastrointestinal problems (Hueck, 1998) and *Y. pseudotuberculosis* that also causes gastrointestinal problems. *Yersinia enterocolita* and *Yersinia pseudotuberculosis* are transmitted via contaminated foods and water (Hueck, 1998) where they localise to the small and large intestine and Peyer’s
patches. The bacteria colonise the Peyer’s patches and are taken up by M-cells (Zaharik et al, 2002). The bacteria may then destroy the M-cells thus gaining access to the underlying tissue and bypassing the epithelial barrier. Invasin mediated uptake via β1-integrin promotes efficient bacterial attachment and entry into the M-cells (Bergman et al, 1994). Once through the epithelial barriers and through the M-cells the Yersinia can then enter the lymphatic system (Hanski et al, 1989; Siebers & Finlay, 1996). Y. pestis infections initiated by fleabite result in the development of bubonic or septicemic plague (Perry & Fetherston, 1997). Symptoms of plague include fever, headaches, local edema and swelling regional lymph nodes that results in black swellings called buboes. Septicemic plague occurs when the bacteria bypass the lymphatic system and enter the bloodstream and disseminate around the body. In all 3 species once the bacteria enter the lymph system they are disseminated around the body and gather near the spleen and liver where the replicate rapidly in the extracellular space leading to necrotic lesions and eventual death (Holmstrom et al, 1995; Straley & Cibull, 1989). The ability of *yersinia* to disseminate via the lymphatic system and replicate extracellularly stems from its unusual ability to evade the immune system in which the TTSS plays a crucial role. It does this through three general methods: (Aepfelbacher, 2004) Antagonising uptake by phagocytes and other immunological components allowing it replicate extracellularly (Hacker et al, 1997; Hanski et al, 1991; Simonet et al, 1990; Aepfelbacher, 2004). Counteracting pro-inflammatory cytokine release; and (Aepfelbacher, 2004) Apoptosis of macrophages (Ruckdeschel, 2002; Zhang & Bliska, 2005). These three effects are brought about wholly by the TTSS of the *Yersinia* bacteria. The TTSS apparatus is found on a single 70kB plasmid in *Yersinia* called the pYV plasmid that contains the genes encoding the TTSS needle, the secreted effectors, regulators of Yop secretion/translocation and the adhesion YadA that can mediate attachment and uptake of *Yersinia* by M-cells (Cornelis, 1998; Hoiczyk et al, 2000). The *Yersinia* TTSS apparatus was discovered in the 1900’s through investigation into the calcium and temperature dependence of *Yersinia* (Michiels et al, 1990) where it was noticed that *Yersinia*, in the absence of Ca²⁺ ions and at 37°C, do not grow, but instead, secrete large numbers of proteins.
into media. This led to genetic analysis that identified the 29 Yop secretion genes (Bergman et al., 1994) involved in the TTSS needle, the eight effector Yersinia outer proteins (Aepfelbacher, 2004) and their specific Yop chaperones (Buttner et al., 2005). Of the 29 Ysc proteins involved 10 have orthologs found in all other TTSS identified. YscC is a secretin protein and forms a 200Å ring in the outer membrane with 50Å pore in the centre of it and is linked with another lipoprotein YscW (Koster et al., 1997) YscD, -R, -U, -V, -S and -T have been shown to span the inner membrane (Cornelis, 2000a) and interacts with the cytoplasmic ATPase YscN (Woestyn et al., 1994). The function of YscJ, the counterpart to EscJ in E.coli, is not known yet though it is probable that it spans the inner and outer membranes in a similar manner to EscJ (Yip et al., 2005). While the Ysc genes that form components of the injectisome are conserved in many other TTSS there is a much larger variation amongst the effectors produced.

There are three broad categories of effector produced by Yersinia spp: (Aepfelbacher, 2004) Mediators that regulate secretion and expression of other Yops; (2) translocators that play an important role in transporting the cytotoxic Yops into the host cells; and (3) the Yops with direct antihost functions (Hueck, 1998). YopN is the most well studied of the regulatory effectors and is directly responsible for the Ca^{2+} dependant nature of Yop secretion by acting as calcium sensor (Boland et al., 1996; Forsberg et al., 1991; Torruellas et al., 2005). YopN in complex with YscB (Day & Plano, 1998) and another protein termed TyeA (Iriarte et al., 1998) blocks the secretion channel from the bacterial cytosol (Cheng et al., 2001) until contact with a eukaryotic cell occurs accompanied by an increase in temperature to 37°C and a drop in Ca^{2+} levels. When a drop in Ca^{2+} ions is registered by the needle protein, YscF, it undergoes a conformational change that results in the YscD-YopN-TyeA complex moving and subsequently leaving the channel unobstructed for the secretion of the cytotoxic Yops (Day et al., 2003; Hoiczyk & Blobel, 2001; Torruellas et al., 2005). The recent structure of the YscD-YopN-TyeA offered structural details on how this pore blocking and unblocking can occur (Schubot et al., 2005). There are three translocation associated effectors in Yersinia: YopB,
YopD and LcrV. All of these are necessary to form a pore in the eukaryotic membrane to allow the efficient translocation of the cytotoxic yops directly into the host cytoplasm (Boland et al, 1996; Cornelis, 2002c; Hakansson et al, 1996; Ramamurthi & Schneewind, 2002). Both YopB and YopD show hydrophobic domains in their sequence (Hakansson et al, 1993). YopB contains a domain reminiscent of the RTX toxins (Hueck, 1998) and both have been show to be able to form pores in liposomes (Cornelis, 2002b). In addition to the pore forming translocases there is also YopK a modulator of YopB/D that is essential for translocation of effectors across the eukaryotic membrane (Holmstrom et al, 1997).

The Yop effectors

There are 6 well characterised effectors that have been found to contribute to the anti-phagocytotic and pathogenic effects of Yersinia spp: YopH, a phosphotyrosine phosphatase, YopE, a Rho-GAP, YopT, a cysteine protease, YopJ, a cysteine protease, YopO, a serine/threonine kinase and YopM, a leucine rich protein.

YopH is a 468 residue protein tyrosine phosphatase. It has distinct N and C terminal domains linked by a proline rich linker region and structures of the individual domains have been determined. The N-terminal domain contains a TTSS secretion signal, the chaperone binding site and a cleft that functions as a substrate targeting domain, binding directly to tyrosine phosphorylated proteins (Khandelwal et al, 2002). The C-terminal domain contains the active site phosphate binding loop with the characteristic C(X)3R(S/T) motif and a second substrate targeting site (Ivanov et al, 2005; Zhang & Bliska, 2003). The catalytic activity of YopH inhibits several phagocytotic pathways (Fallman et al, 1997) including the β1-integrin pathway required for phagocytosis and in doing so inhibits oxidative bursts by preventing degranulation in neutrophils (Persson et al, 1999) along with suppressing the release of macrophage chemoattractant protein 1 (Sauvonnet et al, 2002).
The targets for YopH include focal adhesion kinase (FAK), Fyn-binding protein (Fyb) and Crk-associated substrate (Cas) all of which are involved in activation of Rac-1 GTPase leading to actin polymerisation and internalisation of bacteria (Aepfelbacher, 2004; Cornelis, 2002a; Isberg & Barnes, 2001; Weidow et al, 2000). The importance of YopH has been shown in mice assays were a YopH null mutant was attenuated over 3.4million fold (Kerschen et al, 2004).

YopJ (also called YopP in Yersinia enterocolitica) acts as a potent inhibitor of the MAPK and NF-κB signalling pathways and results in blocking of cytokine production and induction of apoptosis in macrophages (Aepfelbacher, 2004; Aepfelbacher et al, 1999; Ruckdeschel, 2002; Zhang & Bliska, 2005). YopJ is a 288 residue cysteine protease of the CE clan (Viboud & Bliska, 2005) that is thought to cleave ubiquitin or a ubiquitin-like modification on its targets marking them for destruction. As the NF-κB pathway is controlled at several stages by ubiquination the exact target of YopJ being unknown at present (Orth, 2002). The importance of YopJ in infection is still unclear as with mouse studies YopJ null mutants showed a 1.5 fold or 64 fold change in virulence in Y. pestis (Perry & Fetherston, 1997) and Y. pseudotuberculosis (Monack et al, 1998) respectively.

The YopM protein is the only effector not to have a clear enzymatic activity (Viboud & Bliska, 2005) and also varies in size dramatically between the Yersinia species. The YopM effector contains a variable number of a ~20 amino acid leucine rich repeating unit. The number of repeats can vary from 12-20 and in the structure of Y. pestis YopM (Evdokimov et al, 2001) shown in figure 1.3 there are 15 repeats organised into a horseshoe shape and four of these monomers may be arranged as a ring with an inner pore of 35Å (Evdokimov et al, 2001). Though it is known that YopM is transported to the nucleus via a vesicle associated pathway (Skrzypek et al, 2003) its exact role in virulence and its biological function in the nucleus is unknown. It has been shown that YopM is associated with a depletion in several cytokines and YopM knockout mutants have their virulence attenuated by 100,000 fold (Kerschen et al, 2004; Viboud & Bliska, 2005).
While the three Yops mentioned above have diverse targets, the remaining three Yop effectors, Yop E, -O and -T, all target the Rho-GTPases, a common target for many bacterial toxins (Barbieri et al, 2002). The Rho-GTPases are good targets for effectors exerting anti-phagocytotic effects as they manage aspects of the cell cytoskeleton, especially the formation of actin filaments, as well as mediating many other downstream pathways.

YopE is a GTPase activating protein (GAP) on RhoA, Rac-1 and Cdc42. The GAP activity encourages the hydrolysis of GTP to GDP thus switching the GTPase from an active to inactive form. YopE is a 219 residue protein and the GAP domain (residues 96-219) is homologous to the GAP domains in other TTSS effectors such as SptP from *Salomonella* and exoS from *Pseudomonas aeruginosa* (Aepfelbacher, 2004; Andor et al, 2001; Black & Bliska, 2000; Fu & Galan, 1999; Goehring et al, 1999). The active site of the YopE contains an arginine finger motif that intrudes into the GTP binding site of the GTPase and promotes efficient hydrolysis of GTP to GDP inactivating the GTPase.

Through deactivation of RhoA the actin cytoskeleton is heavily disrupted and phagocytosis of the *Yersinia* is prevented. YopE mediated antiphagocytotic behaviour has been observed in macrophages, neutrophils and epithelial cells (Fallman et al, 1995; Grosdent et al, 2002; Ruckdeschel et al, 1996). In addition to the antiphagocytic activity YopE has also been implicated in reducing cytokine production through blocking of the Rac-1 - Caspase-1 interactions in macrophages (Schotte et al, 2004). The antiphagocytic effects of YopE are important to the overall virulence of *Yersinia* spp. and YopE null mutants have been shown be to 10,000 fold less virulent than wild type strains (Perry & Fetherston, 1997).

YopO (called YpkA in *Y. pseudotuberculosis*) is a multi-domain protein with multiple properties of ~730 amino acids (729 in *Y. enterocolita*, 732 in *Y. pestis* and *Y. pseudotuberculosis*). It contains an N-terminal serine/threonine kinase domain
and actin binding and Rho-GTPase binding domains. The actin binding domains and Rho-GTPase binding domains are contained within the C-terminal domain of YopO (Barz et al, 2000; Cornelis, 1998; Dukuzumuremyi et al, 2000). YopO kinase activity is activated by actin binding results in autophosphorylation and phosphorylation of basic substrates, including actin (Juris et al, 2000).

Figure B4. The Yop effector targets in the cell

[Diagram showing the targets of Yersinia effectors. During infection GEF’s are activated via β1-integrin signalling. GTPase activation then triggers phagocytosis and also, via MAP kinase (MKK) and NH-κB (IKK) pathways, gene transcription. The Yop effectors interfere with these pathways at nearly every step. YopH inhibits the initial steps of the β1-integrin pathways halting phagocytosis. YopT inhibits phagocytosis by cleaving RhoA off the membrane. YopE inhibits phagocytosis and cytokine release via its GAP activity and YopO also targets GTPases, though via an unknown mechanism. YopJ prevents the release of cytokines by blocking activation of MKK and IKK and YopM is trafficked to the nucleus where it has its cytoxic effect via an unknown mechanism. This figure was adapted from (Viboud & Bliska, 2005).]
The Rho/Rac binding of YopO is independent of both the nucleotide bound state of the Rho GTPase and the kinase activity of YopO (Juris et al, 2002). Although YopO has been shown to inhibit phagocytosis and disrupt the actin cytoskeleton (Fallman et al, 1997) the cellular function of YopO is still unclear. There is no clear substrate or mechanism for its action though it has been shown to synergise with other Yops to allow the bacteria to colonise and persist in different tissues (Logsdon & Mecsas, 2003).

YopT is a 322 residue cysteine protease of the CA (papain) family that proteolytically cleaves RhoA, Rac and Cdc42 just before a C-terminal cysteine linked geranylygeranyl group that anchors the GTPase to the cellular membrane (Shao & Dixon, 2003). This results in the GTPase becoming released from the cellular membrane thereby inactivating it. This inactivation of RhoA disrupts the formation of stress fibres and leads to cell rounding and inhibition of phagocytosis, similar to the effects of YopE. A more detailed discussion of YopT and the work performed on this protein follows.
**YopT of *Yersinia Pestis***

**Aims of cloning and purifying YopT**

*Yersinia* spp. produce a range of effector molecules that modulate their pathogenicity. Amongst them is YopT, a cysteine protease, that catalyses the proteolytic cleavage of RhoA away from the plasma membrane leading to a breakdown of the actin cytoskeleton. YopT has also been identified as the defining member of a large family of cysteine proteases from many different organisms. The aim of this project was to clone, purify and subsequently determine the structure of YopT to understand its catalytic mechanism and substrate selectivity.

**The history of YopT**

*Yersinia* spp. are able to resist the immune system of their host and replicate extracellularly due to the presence of a complicated TTSS apparatus. In *Yersinia* spp. this is TTSS is called the Yop virulon and is encoded on a single 70kb plasmid, the pYV plasmid (Cornelis & Wolf-Watz, 1997; Iriarte & Cornelis, 1998). This plasmid carries a series of 29 Ysc proteins that are required to construct the TTSS and a series of effector Yops that give rise to the characteristic pathogenic behaviour of the *Yersinia* (Iriarte & Cornelis, 1998).

When the sequence of the pYV plasmid was analysed (Cornelis & Wolf-Watz, 1997) two previously unrecognised ORFs (ORF1 and ORF2) were identified located between YopQ and YopM. These two ORFs were adjacent, the start codon of ORF2 actually overlapping with the stop codon of ORF1 by one nucleotide. ORF1 coded for a 322 residue protein and ORF2 coded for a smaller acidic (pI 4.4) 130 residue protein. ORF2 was shown to be 69.7% and 23.0% similar, to SycE and SycH respectively (Iriarte & Cornelis, 1998; Wattiau et al, 1994; Wattiau & Cornelis, 1993) and hydrophobic moment plots of ORF2 strongly resembled those of SycE suggesting that ORF2 may also have been a chaperone protein. By analogy with the
Using a polymutant *Yersinia* strain (YopH, YopO, YopP, YopE and YopM null) called ΔHOPEM (Boland & Cornelis, 1998) expression of a 35.5kDa protein band representing ORF1 was easily observed. This was only expressed alongside ORF2 and at 37°C in the absence of Ca²⁺ confirming its identity as a new Yop protein subsequently titled YopT (Iriarte & Cornelis, 1998). Initial work performed also showed that the cytotoxic effect of YopT led to disruption of the actin cytoskeleton and cell rounding (Iriarte & Cornelis, 1998). As YopT was seen to interfere with the actin cytoskeleton of cells initial identification of the YopT target focused on RhoA and the related Cdc42 and Rac-1. RhoA, Rac-1 and Cdc42 are known regulators of the cell cytoskeleton. RhoA is crucial to the regulation of stress fibres, Cdc42 and Rac-1 mediate filopodia and membrane ruffles respectively (Hall, 1998). YopT was shown to affect the electrophoretic mobility of [³²P]-labelled RhoA and induce an acidic shift in the pI of RhoA (from pI 6.3 to pI 5.9) through covalent modification (Zumbihl et al, 1999). This covalent modification of RhoA also led to its localisation in the cytosol (Sorg et al, 2001; Zumbihl et al, 1999) away from the membrane where it is active (Hall, 1998).

YopT’s identity as a cysteine protease came to light when Shao and co-workers (Shao et al, 2002) identified 19 sequences from a wide array of bacterial pathogens infecting both animals and plants that showed a degree of sequence similarity and contained highly conserved cysteine, histidine and aspartate (C139, H258 and D274 in YopT) residues (Shao et al, 2002). Subsequent mutational analysis of the C139, H258 and D274 confirmed that all three residues are required for YopT cytotoxicity. Yeast Cdc42 was found to be a target of YopT from a multi-copy suppressor screen and subsequent GST pull-down assays demonstrated that RhoA and Rac-1 were also substrates for YopT. However, these interactions were only observed when the GTPases were post-translationally prenylated. RhoA, Rac-1 and Cdc42 all undergo a similar post-translational modification at their C-terminal CAAX box (C, cysteine;
A, aliphatic; X, any residue) (Zhang & Casey, 1996). The CAAX box is a recognition motif for prenylation, either geranylgeranylation or farnesylation, of the cysteine residue prior to the proteolytic removal of the –AAX tripeptide and methyl esterification of the cysteine (Zhang & Casey, 1996). This modification allows the GTPase to be anchored into the membrane. YopT was unable to bind GTPase mutants missing the CAAX box and a GFP-CAAX box mutant could be co-precipitated with YopT, indicating that the CAAX box is vital for the interaction between YopT and the Rho-GTPases (Shao et al., 2002). Cleavage at the C-terminal of the Rho-GTPases leads to the loss of the prenyl group and displacement of RhoA from the membrane into the cytosol (Shao et al., 2002). The target for YopT recognition has been identified as a polybasic sequence of amino acids upstream of the prenylated cysteine (Shao et al., 2003). The exact point of cleavage, directly N-terminal to the prenylated cysteine, has been confirmed with biochemical and mass-spectroscopic methods. As illustrated in figure B5, a polybasic sequence is present in RhoA, Rac-1 and Cdc42 adjacent to the CAAX box and the prenylated cysteine. As the interaction between YopT and Rho-GTPase occurs at the C-terminus the cleavage is independent of the nucleotide bound state of RhoA (Shao & Dixon, 2003) and is independent of the prenyl group attached to the cysteine (Shao & Dixon, 2003).

Figure B5. C-termini cleavage sites of the Rho-GTPases

<table>
<thead>
<tr>
<th>Protein</th>
<th>Cleavage Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>RhoA</td>
<td>ARRGGKKSGCLVL193</td>
</tr>
<tr>
<td>Rac-1</td>
<td>LCPPPVKKRKCLLLL192</td>
</tr>
<tr>
<td>Cdc42</td>
<td>EPPETQPKRCCIF191</td>
</tr>
</tbody>
</table>

Figure B5. The C-terminus sections of RhoA, Rac-1 and Cdc42. The polybasic sequences rich in arginines and lysines are shown and the CAAX box is underlined. The arrow denotes where the YopT cleavage takes place. This figure is adapted from (Shao et al., 2003).

To summarise, YopT is a cysteine protease that assists Yersinia in overcoming phagocytosis by targeting RhoA, Rac-1 and Cdc42. In the host cell it binds to a C-
terminal polybasic sequence and the CAAX box, and cleaves the GTPase N-terminal to the prenylated cysteine that anchors the GTPase to the membrane. The GTPase is then left in the cytosol, where it is inactive leading to a disruption of actin cytoskeleton and inhibition of *Yersinia* phagocytosis. In *Yersinia* pathogenesis the role of YopT and YopE overlap somewhat as both target the RhoGTPases leading to disruption of the actin cytoskeleton. Both are necessary for resistance to phagocytosis and YopT mutants have been shown to be less resistant to macrophages (Grosdent *et al*, 2002). However, it may be that YopE and YopT act synergistically as YopE disrupts the formation of actin-rich phagocytic cups almost immediately, whilst the YopT effect is slower requiring 15-30 minutes. Also the YopE effect is reversible and transient, whilst the YopT effect is a permanent effect. This co-operation and redundancy between YopE and YopT generates phagocytic resistance (Aepfelbacher *et al*, 2003).

The initial studies into YopT by Shao and co-workers uncovered a wide variety of ORF's that showed similar invariant C/H/D residue suggesting that YopT may define a new family of cysteine proteases common to plant and animal pathogens (Shao *et al*, 2002). The sequence alignment is shown below in figure B6 and includes ORFs from all three *Yersinia* species, *Haemophilus somnus*, the plant pathogen *Pseudomonas syringae* and the enteropathogenic *Escherichia coli* O157. Cysteine proteases are grouped together into ~40 different families related by sequence identity (Barrett, 1994; Barrett *et al*, 2001). YopT however shows very low sequence identity to any of the current families, but the secondary structure predictions for YopT, shown in figure B6, and many of the related ORFs show some similarity to the papain family (CA clan) of cysteine proteases. This hypothesis was confirmed by the recent report of the structure of AvrpPhb a *P. syringae* type III effector that is part of the YopT family of cysteine proteases (Zhu *et al*, 2004) and strongly resembles the papain family cysteine proteases, part of the CA clan. Biochemical inhibition studies using the aspartyl protease inhibitor pepstatin, the serine protease inhibitor PMSF, EDTA, thiol protease inhibitor N-ethylmaleimide and the CA-clan specific inhibitor E64 (L-trans-epoxysuccinyl-leucylamide-(4-
guanido)-butane) (Barrett et al, 2001) are also consistent with YopT being a cysteine protease in the CA clan (Shao et al, 2002).

Analysis of the related ORFs also highlighted interesting features including three broad groups of related proteins. (1) Small type III effectors proteins ranging from 30-40kDa in size such as the Psuedomonas effectors. (2) Large >300kDa multi-domain proteins that possessed a YopT-like domain such as LifA from E. coli or toxinB from the enteropathogenic E. coli O157 and (3) the Haemophilus P76 immunoglobulin binding proteins. YopT along with the Haemophilus proteins from both H. somnus and H. ducreyi may also form a sub-family as they have a much higher sequence similarity than the other proteins found to date (Shao & Dixon, 2003).
Figure B6. Sequence alignment of YopT and three other related cysteine proteases; P76 from *H. somnus*, HopPtoC from *P. syringae* and; toxB from *E. coli* 0157:H7. The secondary structure assignment was determined by the YopT model build by the 3d-Jigsaw server (Bates et al, 2001). The catalytic triad (C139, H258 and D274 in YopT) are shaded in yellow. The 3 conserved aromatic residues that form stacking Van der Waals interactions are shaded in red, as is the conserved glutamate that interacts with the conserved tryptophan. The amide contributing residue that forms the oxyanion hole is shaded in turquiose. Other areas of conserved sequence are shaded in grey.
Methods

YopT expression and purification was attempted through two different strategies. The first was cloning into pET22b (Novagen Inc., UK) and following a similar protocol to that employed in the purification and structural determination of the related cysteine protease AvrpPhb from *P. syringae* (Zhu *et al.*, 2004). The second strategy was to clone YopT into the solubility enhancing pMalp2x vector (New England Biolabs) as an MBP fusion protein that ultimately failed.

**Expression and purification of the YopT-pET22b construct**

The YopT gene was received as freeze-dried plasmid in the pDR4TOPO plasmid provided by Etelechon GmbH. Primers were designed to introduce a 5' NdeI site and a 3' SalI site prior to cloning YopT into pET22b (Novagen Inc., UK).

Forward primer (NdeI site underlined):

5' GGG CAT ATG ATG AAC AGC ATT CAT GGC 3'

Reverse primer (SalI site underlined):

5' CGC GTC GAC GAA TTC CAC TTC TTT TGA 3'

The PCR was performed using the Boehringer Expand™ PCR system and the following protocol: 95°C 30 seconds, 30 cycles of 95°C 30 seconds, 45°C 1 minute, 72°C 1 minute and 30 seconds, followed by a 7 minute hold at 72°C and storage at 4°C. The resulting DNA visualised on a 1% agarose gel and purified using the Qiagen DNA purification kit (Qiagen Ltd. UK).

The purified YopT DNA and the pET22b plasmid (Novagen Inc., UK) underwent restriction digestion using the NdeI and SalI (Promega) restriction enzymes with a 2 hour incubation at 37°C. Restriction digested plasmid and YopT DNA were ligated overnight at 4°C with T4 DNA ligase (Promega). Ligated plasmids were then transformed by heat-shock into BL21 cloning cells and the resulting transformants
were plated out onto LB-agar plates supplemented with 100μg/ml ampicillin and incubated overnight at 37°C.

Successful transformants were screened for by PCR using Taq polymerase (Promega) and the following protocol: 95°C 1 minute, 25 cycles of 95°C 30 seconds, 45°C 1 minute, 72°C 1 minute and 30 seconds, followed by a 7 minute hold at 72°C and storage at 4°C. The resulting DNA visualised on a 1% agarose gel and successful colonies were used to inoculate 10ml of luria broth supplemented with 100μg/ml ampicillin overnight at 37°C. The plasmid DNA was recovered using the Promega wizard miniprep kit and 1.25μl of the purified YopT-pET22b was then used to transform BL21(DE3) expression cells (Novagen Inc., UK). Successful insertion of the YopT gene was confirmed with T7 forward and reverse sequencing (MWG Biotech).

Initial expression of soluble YopT proved difficult and optimisation of the growth conditions resulted in a low temperature expression scheme strategy. With the advent of the work by Studier and colleagues (Studier, 2005) autoinduction media was used instead of the low temperature expression in terrific broth and yielded significantly higher yields of soluble YopT.

Single colonies of pET22b-YopT were grown overnight at 37°C in 10ml ZYM-505 media (1% N-Z amine, 0.5% yeast extract, 5mM Na2SO4, 2mM MgSO4, 25mM (NH4)2Cl2, 25mM KH2PO4, 20mM Na2PO4, 0.5% glycerol, 0.05% glucose, 50nM FeCl3, 20nM CaCl2, 10nM MnCl2, 10nM ZnSO4, 2nM CoCl2, 2nM CuCl2, 2nM NiCl2, 2nM Na2MoO4, 2nM Na2SeO4, 2nM H3BO3) supplemented with 100μg/ml ampicillin. The overnight cultures were then diluted 1:1000 into 1l fresh ZYM-5052 media (1% N-Z amine, 0.5% yeast extract, 5mM Na2SO4, 2mM MgSO4, 25mM (NH4)2Cl2, 25mM KH2PO4, 20mM Na2PO4, 0.5% glycerol, 0.05% glucose, 0.2% lactose, 50nM FeCl3, 20nM CaCl2, 10nM MnCl2, 10nM ZnSO4, 2nM CoCl2, 2nM CuCl2, 2nM NiCl2, 2nM Na2MoO4, 2nM Na2SeO4, 2nM H3BO3) supplemented with 100mg/l ampicillin and left to grow for overnight at 37°C. Soluble protein
expression was observed on an SDS-page gel. Cells were harvested by centrifugation (Beckman Avanti J-25).

Harvested cell pellets was re-suspended in lysis buffer (20mM Hepes pH 7.3, 500mM NaCl, 0.3mg/ml lysozyme) and stirred on ice for 30 minutes. Cells were lysed via sonication (Soniprep 150) and the inclusion bodies removed by high speed centrifugation (Beckman Avanti J-25).

Clarified cell lysate was loaded onto a 5ml Hitrap affinity column (GE Healthcare) charged with NiSO$_4$ and equilibrated with 50mM Hepes pH 7.3, 500mM NaCl and 10mM imidazole. YopT was eluted as a single peak with 50mM Hepes pH 7.3, 500mM NaCl and 500mM imidazole. The protein was then applied to a Superdex 200 gel filtration column (GE Healthcare) equilibrated with 20mM Tris pH 8.5, 500mM NaCl, 10mM β-mercaptoethanol with a slow flow speed of 0.1ml/min to separate full length YopT from its 25kDa fragment. The purified YopT and the 25kDa YopT fragment were concentrated separately in a 10kDa centrifugal concentrators (Amicon) to 10mg/ml and stored at 4°C for crystallisation.

**Crystallisation of YopT**

Crystallisation trials were performed as hanging drops using 1.75µl of protein with 1.75µl of reservoir solution in the drop, giving a final protein concentration of 5mg/ml in the crystallisation drops. Structure screens I and II (Hampton Research) were used to generate initial crystallisation conditions. Optimisation trials centering around 100mM sodium cacodylate or Hepes pH 6.5, 0.2M magnesium acetate, 15% PEG8000 were performed. Plates were incubated at either 16°C or 4°C.

Micro-seeding experiments were performed as follows: A 24 well plate was set-up with 100mM Hepes pH 6.5, 0.2M magnesium acetate, 10-15% PEG8000 with 10mg/ml YopT and was left to equilibrate overnight at 16°C. A drop containing crystals was resuspended in 800ul of 100mM Hepes pH 6.5, 0.2M magnesium acetate...
acetate, 12.5% PEG8000 and mixed vigorously for ~10 minutes. Then 0.5μl of the solution was added to each crystal drop on the equilibrated plate. The plate was then incubated for at 16°C until crystals appeared.

Streak-seeding experiments were performed as follows: A 24 well plate set-up with 100mM Hepes pH 6.5, 0.2M magnesium acetate, 10-15% PEG8000 and 10mg/ml YopT was left to equilibrate overnight at 16°C. A cat’s whisker was then brushed through a drop containing micro-crystals and then brushed across the surface of each drop on the equilibrated plate. The plate was then incubated for at 16°C until crystals appeared. Additive screens were set-up by the addition of small amounts of the additive to the crystallisation reservoir. Additives used were 1-10% MPD, 1-20% ethanol, 5-20% isopropanol and 1-10% 1,6-hexandiol.

**Crystallisation of the YopT fragment**

Crystallisation trials were performed as hanging drops using 1.75μl of protein with 1.75μl of reservoir solution in the drop, giving a final protein concentration of 5mg/ml in the crystallisation drops. Structure screens I and II (Hampton Research) were used to generate initial crystallisation conditions. The initial condition from the structure screen consisted of 100mM Tris pH 8.5, 50% MPD and 0.1M (NH₄)₂SO₄. Optimisation screens were performed varying: pH of the Tris buffer, percentage of MPD and concentration of (NH₄)₂SO₄.
Results and discussion

Experimental results

The expression and purification of YopT posed several problems during optimisation. The initial choice to use the pET22b vector (Novagen) was based on the work by Zhu and colleagues (Zhu et al, 2004) in purifying the AvrpPhB cysteine protease from P. syringae. AvrpPhb was identified as a member of the same family of 30-40kDa type III effectors as YopT (Shao et al, 2002). Initial cloning into pET22b was straightforward and was confirmed with sequencing, double digestion using NdeI and Sall restriction enzymes (Promega) and colony PCR. Expression was more difficult however. Standard trials using 1mM IPTG to induce expression at 37°C failed to produce soluble YopT. The protocol used by Zhu and colleagues using M9 minimal media but with expression overnight at 20°C induced with 0.4mM IPTG was also unsuccessful at producing soluble YopT (Zhu et al, 2004). However, this did lead a successful protocol using low temperature, 20°C, and a low IPTG concentration, 0.4mM, in terrific broth. This protocol led to high levels of expression as shown below in figure B7A. Later trials with autoinduction media at 37°C resulted in greatly increased yields from 8mg/ml obtained at 20°C in terrific broth to >20mg/ml.

The initial purification of YopT was straightforward. Once soluble YopT had been obtained histidine affinity chromatography was an effective and efficient method of purifying YopT away from the majority of cell contaminants, although in all cases, a second band of ~25kDa was co-purified with full length YopT as can be seen in figure B2.3B. This second band was thought be a breakdown product of YopT as without addition of 10mM β-mercaptoethanol the YopT degraded rapidly increasing the ratio of YopT fragment: full length YopT in the solution. The presence of this 25kDa fragment made gel filtration chromatography more difficult as both the fragment and the full length YopT tended to elute as a series of peaks. Reverse phase chromatography was attempted but freeze-dried YopT was unable to return to
solution. Both histidine affinity chromatography and anion-exchange chromatography failed to separate the fragment from the full length protein. The presence of this 25kDa fragment was unexpected, as, unlike AvrpPhB YopT is not known to be proteolytically cleaved before activation (Zhu et al, 2004) and other groups expressing YopT have not mentioned the appearance of this 25kDa fragment. The fragment co-purified alongside YopT on a nickel affinity column so is likely formed from the C-terminus of YopT as the pET22b vector possesses a C-terminal hexahistidine tag.

Figure B7 Expression & purification of YopT

Figure B7: In 3A is the expression of YopT in terrific broth before induction (+0), 3 hours after induction with 0.4mM IPTG at 20°C (+3) and expression after overnight incubation at 20°C directly before the cells were harvested (O/N). The component of the cell pellet is divided into the soluble and insoluble fractions after lysis by sonication. 3B shows the purity after initial metal affinity chromatography (Hitrap) and the component of the 3 peaks witnessed upon gel filtration. The full length YopT (~40kDa) and the YopT fragment (~25kDa) can be observed.
Crystallisation of YopT and the fragment of YopT have led to encouraging initial results. Micro-crystals and thin needles like those shown below in figure B8A have been grown for the full length YopT. Although none of the needles grown to date have been large enough for diffraction studies. Crystals of the YopT fragment, shown below in figure B1.8, grew in 100mM Tris pH 8.5, 50% MPD and 0.1M (NH₄)₂SO₄ but were too small for diffraction study. Further optimisation of the crystallisation conditions is in progress.

**Figure B8 Crystals of full length YopT**

A ![Image](image1)

B ![Image](image2)

*Figure B8: Crystals of (a) the full length YopT protein and (b) the 25kDa breakdown product of YopT.*

**Homology modelling with AvrpPhB**

YopT and AvrpPhB were the cysteine proteases that defined this family (Shao *et al*, 2002). Both possessed a conserved catalytic triad (C139/H258/D274 in YopT) along with similar predicted secondary structure. In common with most other members of this family of cysteine proteases they are also thought to share a conserved core domain.
Using the AvrpPhB structure (PDB code 1UKF) (Zhu et al., 2004) the automated modelling program 3d-Jigsaw Server (Bates et al., 2001) was used to build a model of the full length YopT protein. The central 50 residues around the active site features superimposed well on the AvrpPhB structure with an RMSD of 1.67Å and is shown below in figure B9. The model has an α/β fold with two lobes formed primarily of α-helices stacked above and below a five stranded anti-parallel β-sheet. Two long loop regions (residues 230-244 and 284-301) protrude away from the core of the molecule as long coils but are more likely areas where the automated modelling program was unable to predict accurate secondary structural elements.

Figure B9 Homology model of YopT effector from *Yersinia pestis*

**Figure B9:** Homology model of YopT prepared by 3d-Jigsaw server (Bates et al., 2001) using the related cysteine protease AvrPphB from *Pseudomonas syringae* as the initial model (PDB code 1UKF) (Zhu et al., 2004).
The catalytic triad form the active site just under the β-sheet with C139 lying at the very beginning of α2 helix and H258 and D274 lying β-strands 3 and 4 respectively. A detailed look at active site of reveals many similarities to both papain and AvrpPhB structures and is shown in figure B10. The catalytic triad are conserved and in the same conformation as is found in AvrpPhB and the rest of the papain family (Jia et al, 1995) suggesting that YopT may share a similar catalytic mechanism. This involves the cysteine and histidine forming a thiolate-imadozolium pair with the catalytic aspartate stabilising the ion pair and orienting the active site correctly. Mutagenesis studies confirm that loss of any of those 3 residues results in the loss of YopT activity (Shao et al, 2002).

**Figure B10 Important residues around the active site of the predicted YopT structure**

![Figure B10](attachment:image.jpg)

**Figure B10.** A close up of the active site of YopT showing the catalytic triad and the conserved hydrophobic residues that stack above each other.
There are also 3 hydrophobic residues that are conserved throughout the YopT family of proteases (Shao et al, 2002) (Zhu et al, 2004). In YopT these are: Phe273, Pro275 and Trp146. Phe273 and Pro275 flank the catalytic aspartate on the β4 strand and Trp248 lies at the tail end of the α2 helix that the catalytic Cys139 lies on. Phe273, Pro275 and Trp146 stack above each other with strong van der Waal interactions. These interactions hold the superstructure of the active site together and hold helix α2 and strand β4 in close proximity so the catalytic aspartate is correctly positioned to stabilise the thiolate-imadozolium.

In the AvrpPhB structure the equivalent tryptophan also forms a hydrogen bond with a glutamate residue (E279 in YopT) that is 5 residues after the catalytic aspartate and on the strand β5 separated from strand β4 by a short turn formed by the conserved proline (Zhu et al, 2004). The importance of Trp148 in YopT was highlighted by mutagenic studies that demonstrated a W148A mutant lost all activity (Shao et al, 2002). The other important features of the cysteine proteases are the oxyanion hole that stabilises the P1 residue, the residue directly N-term to the cleavage site, and the S1, S2 and S3 sites that interact with the side chains of the P1, P2 and P3 residues of the substrate.

The P1, P2 and P3 residues are the three residues directly upstream of the point of proteolytic attack. In the case of YopT’s cleavage of RhoA the P1, P2 and P3 residues are a glycine, serine and lysine respectively as can be seen in figure B2.1. In papain and AvrpPhB the oxyanion hole is formed from the amide group of an asparagine residue (Zhu et al, 2004) (Schroder et al) (Menard et al, 1991). Although YopT lacks an asparagine at this position the loop upon which the equivalent reside, Gly 137, lies is shorter than in papain or AvrpPhB resulting in the amide group of Gly137 in YopT lying in a similar position to the amide group of the asparagines in papain or AvrpPhB. The S1, S2 and S3 sites in papain and AvrpPhB are involved in the substrate specificity, however in YopT the specificity is governed by the prenyl group attached to the cysteine and a short polybasic sequence lying further upstream of the cleavage site.
**Figure B11.** The S1, S2 and S3 sites of YopT (Blue) compared with AvrPphB (red). The catalytic triad of YopT is shown. In all cases the AvrPphB residue is numbered in italics. The difference in the loop regions around the oxyanion hole is highlighted.

In AvrPphB the S1 site is centred around Y175 whilst in the equivalent position in YopT there is a lysine (K178). S2 in AvrPphB is formed by arginine 205 whilst in YopT the equivalent residue is Leu251. The S3 site is formed by a pocket of several residues at the tip of helix α4 and is comprised of leucine, phenylalanine and tyrosine in AvrPphB but lysine, isoleucine and aspartate (K178, I182 and D190) in YopT. These are shown in figure B11. In the related AvrPphB the nature of S1-S3 provide complimentary binding sites for the P1, P2 and P3 residues (Zhu et al, 2004). In YopT the same strong link between residue and binding site is not so obvious, especially as can be seen from figure B11 there are differences in the P1, P2 and P3 residues between the three known substrates of YopT. From our structural model it
is not immediately obvious how substrate specificity is controlled and where the prenyl group or polybasic sequence interact with the YopT molecule.

**Conclusions**

We have cloned and optimised a purification strategy to isolate pure YopT cysteine protease. We have managed to grow small needle-like crystals, though at present these are too small to be of use in diffraction experiments. Homology modelling with the structure of the related cysteine protease AvrPPhB from *Pseudomonas syringae* shows conserved features around the active site including the catalytic triad, oxyanion hole and a trio of hydrophobic residues. As AvrPPhB and YopT share different substrates and YopT has 3 substrates the manner in which the polybasic sequence and prenyl group determine specificity in YopT is still unknown.
P76 High weight immunoglobulin binding protein from 
*Haemophilus somnus*

Aims of cloning and purification of the YopT-like domain of P76

*H. somnus* possesses P76, a large immunoglobulin binding protein on its surface. The C-terminal half of this 76kDa protein has been identified as having strong sequence identity to the YopT effector from *Yersinia pestis*. We hope to clone, purify and subsequently determine the structure of the YopT-like domain to gain insight into its substrate and to investigate if it shares similar structural features to the rest of the YopT-like cysteine protease family.

An introduction to Immunoglobulin binding proteins in *H. somnus*

*Haemophilus somnus* is a gram negative pleomorphic bacterium that causes many debilitating and serious conditions in cattle and sheep. Symptoms caused by *H. somnus* are serious and include thrombotic meningoencephalitis, septicemia, pneumonia, infertility, abortion, myocarditis and arthritis (Harris & Janzen, 1989; Humphrey & Stephens, 1983; Kwiecien & Little, 1991; Miller et al, 1983; Stephens et al, 1981). However many cattle carry *H. somnus* as a commensal on the genital mucosa (Corbeil et al, 1997). One of the differences that have been noted between the virulent strains and the asymptomatic strains is their resistance to being killed by the complement system in bovine serum. This gives rise to invasive, virulent serum resistant strains and asymptomatic commensal serum sensitive strains (Corbeil et al, 1985). In *H. somnus* resistance to complement mediated killing comes about from the lipooligosaccharide composition and the presence of immunoglobulin binding proteins on the cell surface (Inzana et al, 1988; Widders et al, 1989). These immunoglobulin binding proteins (IgBPs) consist primarily of either high molecular weight IgBPs (HMW-IgBPs) that bind to bovine immunoglobulin 2 (IgG2) or a 41kDa outer membrane protein with a lower affinity for several immunoglobulins including IgG1 and IgG2 (Yamall et al, 1988). The HMW-IgBPs are a series of proteins with molecular weights ranging from 120 to 350kDa and more recently,
P76 was also shown to be a member of the HMW-IgBPs as well (Corbeil et al., 1997). All of the IgBPs, including P76, are coded for by a single, long 12kb insert (Corbeil et al., 1988). More recent work has shown that the entire 12kb insert codes for a single ORF, IbpA, which encodes all of the HMW-IgBPs and P76 but has many translation initiation sites (Tagawa et al., 2005). IbaA is coupled with another ORF, IbaB, that is consistent with a 66kDa outer membrane protein and meets the requirements for two-partner secretion pathway for transporting large virulence exoproteins (Jacob-Dubuisson et al., 2001). This two protein secretion couple is also seen for LspA1 and A2 from *H. ducreyi* coupled with LspB (Tagawa et al., 2005).

Within the IbpA sequence there are two regions; one in the centre of the insert, the other in the P76 region, that are homologous to Fc binding domains of *S. pyogenes* and play a similar role in inhibition of complement mediated killing. Binding to Fc domain of IgG2 in this manner may prevent the immunoglobulin achieving the correct conformation to activate the complement cascade (Cole et al., 1993; Johnsson et al., 1994; Pack et al., 1996). The IpbA ORF also encodes several adhesion motifs that may aid in adhesion to epithelial cells (Tagawa et al., 2005). The role of the HMW-IgBPs in *H. somnus* involves binding to the IgG2 to prevent complement mediated killing and also to assist in adhesion to epithelial cells. To this end the HMW-IgBPs, including P76, are presented to the extracellular matrix as a series of surface fibrils coating the *H. somnus* bacterium (Corbeil et al., 1997). These surface fibrils bear a strong resemblance to the surface virulence factors from other pathogens including YadA form *Yersinia* spp. (Kapperud et al., 1987) filamentous hemagglutinin of *B. pertussis* (Reiman et al., 1989) and a family of high molecular weight proteins (HMW1 and HMW2) from *H. influenzae* (St Geme et al., 1993). All of these surface fibrils are involved in adhesion to epithelial cells, and in the case of YadA, resistance to complement mediated killing (Pilz et al., 1992).

P76 along with LspA1 and LspA2 of *H. ducreyi* possess the most closely related cysteine protease domains to the *Yersinia* spp. (Shao et al., 2002) though in the case of P76 the function is unknown.
Methods

Expression of the YopT-like domain of *H. somnus* P76 gene was attempted through three different strategies. These were: cloning into pET22b (Novagen) with a hexahistidine tag, cloning into pMalp2x (New England Biolabs) as an MBP fusion protein and cloning into the ligation independent vector ysbLIC based on pET28 (Novagen). Ultimately only the LIC clone was successful.

Expression and purification of the P76-ysbLIC construct

The ysbLIC vector was a gift from Dr. Philip Leonard and was based on pET28 (Novagen) and possesses an N-terminal hexahistidine tag.

The P76 gene fragment was received as freeze-dried plasmid in the pDR4TOPO plasmid provided by Etelechon GmbH. Primers were designed with the appropriate ends for insertion in to the ysbLIC vector.

Forward primer (LIC specific ends underlined):

5' CAC CAC CAC CAC ATG GAA TTT CTC AAA GAA CTT GCT AAA AAA GGT AG 3'

Reverse primer (LIC specific ends underlined):

5' GAG GAG AAG GCG CGT TAG TCG ACT TAG AAT TCG TTA TTT TTT TTG TAG TTA A 3'

The PCR was performed using KOD polymerase (Novagen) and the following protocol: 94°C 2 minutes, 35 cycles of 94°C 30 seconds, 45°C 30 seconds, 72°C 25 seconds, followed by a 7 minute hold at 72°C and storage at 4°C. The resulting DNA was visualized on a 1% agarose gel and purified using a DNA purification kit (Promega).
For the insertion into the ysbLIC vector 0.2pmol of the purified P76 DNA was incubated at 22°C for 30 minutes with T4 DNA polymerase (Novagen) 2.5mM dATP and 5mM DTT, the reaction was stopped by heating to 75°C for 20 minutes. After treatment with T4 DNA polymerase P76 DNA was incubated with 50ng of the ysbLIC vector for 10 minutes at room temperature before addition of EDTA and incubation at room temperature for a further 10 minutes. 1.25μl of the annealing reaction was transformed into Nova-Blue (Novagen) and subsequently plated out onto LB-agar plates supplemented with 30μg/ml kanamycin.

Successful transformants were screened for by PCR using KOD polymerase (Novagen) and the following protocol: 95°C 1 minute, 35 cycles of 95°C 30 seconds, 45°C 30 seconds, 72°C 25 seconds, followed by a 7 minute hold at 72°C and storage at 4°C. The resulting DNA visualised on a 1% agarose gel and successful colonies were used to inoculate 10ml of luria broth supplemented with 300μg/ml kanamycin overnight at 37°C. The plasmid DNA was recovered using the Promega wizard miniprep kit and 1.25μl of the purified P76-ysbLIC was then used to transform BL21(DE3) expression cells (Novagen Inc., UK). Successful insertion of the P76 gene was confirmed with T7 forward and reverse sequencing (MWG Biotech).

Single colonies of ysbLIC-P76 were grown overnight at 37°C in 10ml luria broth supplemented with 30μg/ml kanamycin. The overnight cultures were then diluted 1:100 into fresh terrific broth supplemented with 0.5% w/v glycerol and 30mg/l kanamycin and left to grow for ~3 hours at 37°C until an OD$_{600}$ ~0.6 was reached. Protein expression was induced with 1mM IPTG for 5 hours at 25°C and high levels of soluble expression were observed as a 36kDa band on an SDS-page gel. Cells were harvested by centrifugation (Beckman Avanti J-25).

Harvested cell pellets were re-suspended in lysis buffer (20mM BisTris pH 7.0, 500mM NaCl, 0.3mg/ml lysozyme) and stirred on ice for 30 minutes. Cells were
lysed via sonication (Soniprep 150) and the inclusion bodies removed by centrifugation (Beckman Avanti J-25).

Clarified cell lysate was loaded onto a Hitrap column (GE Healthcare) charged with NiSO$_4$ and equilibrated with 50mM Hepes pH 7.3, 500mM NaCl. P76 was eluted as a single peak with the addition of 50mM Hepes pH 7.3, 500mM NaCl, 500mM imidazole. Fractions containing P76 were further purified using reverse phase chromatography (Jupiter C-25 column) and an acetonitrile gradient. P76 was eluted as a single band at ~47.5% acetonitrile. The fractions containing P76 were lyophilised overnight and re-suspended in distilled water and concentrated to 10mg/ml with a 10kDa cut-off concentrator (Amicon) for crystallisation trials.

**Crystallisation**

Crystallisation trials were performed as hanging drops using 1.75µl of protein with 1.75µl of reservoir solution in the drop, giving a final protein concentration of 5mg/ml in the drops. Structure screen I and II (Hampton Research) were used to generate initial conditions with two conditions offering possible crystals. The conditions were: (1) structure screen II condition 37, 100mM Tris pH 8.5, 1.0M Li$_2$SO$_4$, 10mM NiCl$_2$ and (2) structure screen II condition 6, 100mM sodium acetate pH 4.5, 1.0M 1,6-Hexanediol, 10mM CoCl$_2$. Grid screening around these two possible conditions were set-up. No crystals have been observed as yet.
Results and discussion

Experimental results

Initial cloning into the pET22b construct was straightforward and gave positive results in many of the colonies obtained after transformation. However, correct insertion of the P76 gene no expression was observed in either luria broth, terrific broth or auto-induction media.

Cloning into the pMal-p2x vector was more successful and it expressed well. However efforts to purify the P76-MBP fusion protein using ion-exchange and amylose affinity chromatography both failed to purify P76.

Figure B12 Expression and purification of the YopT-like P76 domain

Figure B12. A: The expression of the P76-LIC construct with time points before induction (+0), 2 hours after induction with 1mM IPTG (+2) and 5 hours after induction with 1mM IPTG (+5). The soluble and insoluble contents of the cell after lysis are also shown and it P76 can be seen clearly in the soluble fraction. B: the purity of P76 after nickel affinity (Hitrap) and reverse phase chromatography (HPLC).
Cloning into the ysbLIC vector was markedly and a large soluble expression band was observed on an SDS-page gel shown above in figure B12A. Purification was straightforward with no optimisation of either the metal chelating step or reverse phase step and pure P76 was still soluble after freeze-drying. This resulted in a yield of 9mg of P76 per litre of culture that was pure, homogeneous, soluble at high concentration and suitable for crystallisation. The purity of P76 can at each stage of the process is shown above in figure B12B.

Crystallisation yielded little initial success. However after ~2 months small crystallioids were observed in the two conditions mentioned previously. Optimisation screens designed around condition 6 and 37 from structure screen II (Hampton) have as yet yielded no crystals. For the condition based on structure screen II number 6 grid screening was performed around the concentrations of CoCl₂ and 1,6-hexandiol. For the other condition grid screens were performed with varying concentrations of NaCl and Li₂SO₄.

**Homolgy modelling of P76 using AvrpPhB**

P76 was the YopT homolog identified with the highest sequence similarity to YopT and using the plant effector molecule AvrpPhB [PDB code 1UKF] (Zhu et al., 2004), a model of the structure was built by the web-based server CPH model (Lund et al., 2002). The CPH model server identified the sequence of the P76 domain as having 16.3% identity with AvrpPhB and a homology model was built with an RMSD of 1.68Å around 50 residues of the active site. The model, however, does omit the first 135 residues of the P76 fragment indicating substantial differences at the N-terminus. The P76 model was comprised of the same secondary structure arrangement as YopT and AvrpPhB (Zhu et al., 2004) with a 4 stranded anti-parallel β-sheet sandwiched between two lobes comprised of loop regions and α-helices. The catalytic of triad of P76 were conserved with C139, H262 and D277 superimposing exactly on all of the homology models examined. The model of P76 also possessed the stacked hydrophobic rings F276, P278 and W146 that interact strongly via van
der Waals interactions and hold helix α-1 and strands β-2 and β-3 together in the same manner as seen in the YopT model and the structure of AvrpPhB and are shown below in figure B3.2. The P76 also possesses a glutamate residue (Glu282) on strand β-4 that is oriented in a manner to form a hydrogen bond to the conserved tryptophan residue (Trp146).

Figure B13 A close-up view of the P76 hypothetical active site

Other areas of importance around the S1, S2 and S3 sites also strongly resemble the features seen in AvrpPhB and the model of YopT. The oxyanion hole amide is available for the amino group of Gly137, similar to the YopT model. However, the
S1 site and S2 sites show marked differences to YopT with the S1 site substituting a isoleucine (Ile176) for the tyrosine and at the S2 site the change is from a leucine to a serine (Ser 254). The large pocket involving the cleft between helices α2 and α3 forming the S3 site is also different to that seen in YopT as seen below in figure B13.

In P76 the cleft is lined with hydrophobic isoleucine and leucine residues (Ile 176 and Leu182), except at the top adjacent to the loop connecting helix α2 and α3, where there is a glutamine (Gln 195). These differences in the S1, S2 and S3 binding sites may suggest that whilst P76 shows a strong similarity to YopT in overall structure it is likely to possess a different substrate.

**Conclusions**

We have successfully cloned and optimised a purification strategy to isolate the YopT-like domain of P76 from *H. somnus*. Crystallisation experiments are still ongoing. Homology modelling with the structure of the related cysteine protease AvrpPhB from *Pseudomonas syringae* shows conserved features around the active site including the catalytic triad, oxyanion hole and a trio of hydrophobic residues that stack via van der Waals interactions and maintain the structure of the active site. A comparison of the substrate binding sites in the P76 model with the YopT model highlighted differences that may be related to the as yet unknown target of the cysteine protease domain of P76.
HopPtoC (AvrPpci2) from *Pseudomonas syringae* pv tomato *DC3000*

**Aims of cloning and purifying HopPtoC**

Plant pathogens have a broad range of TTSS effectors. Several of these have been identified as being cysteine proteases that are related to YopT from *Yersinia* spp. The structure of one of these, AvrpPhB, has been determined but the others have not. We hope to clone, purify and subsequently determine the structure of HopPtoC to understand its catalytic mechanism and possibly shed some light on its as yet unknown role in plant pathogenicity.

**An introduction to type III effectors in *Pseudomonas syringae***

*Pseudomonas syringae* is a gram negative bacteria composed of many strains that infect different plant species causing a broad range of diseases (Jin *et al.*, 2003). However, individual strains of *P. syringae* tend to be specific to only a few species of plants or even a select range within a plant species and this give rise to different pathovars (pv.). In *P. syringae* infection the bacteria enter the apoplast of the plant and multiply extracellularly resulting in tissue necrosis. This ability to multiply extracellularly and cause disease is mediated by a TTSS and the effector proteins that are secreted by it (Alfano & Collmer, 1997).

In *P. syringae* the TTSS assembly is encoded for by the hypersensitive reaction and pathogenicity (hrp) genes and secretes avirulence proteins (Avr) and Hrp outer proteins (Hops). These names have come about from experiments that describe a hypersensitive reaction that occurs when strains of *P. syringae* are exposed to non-host plant species or resistant plant species. This hypersensitive response results in a rapid physiological response in infected plant tissue that culminates in programmed cell death and a cessation of bacterial growth. This cessation of bacterial growth led to avirulence and the effectors that elicited this response being named avirulence.
proteins (Avr) (Shao et al, 2002) (Collmer et al, 2002). In susceptible plants however these effectors (both Hops and Avr proteins) have been shown to contribute to the pathogenicity of P. syringae (Innes, 2001) and the consensus currently is that they contribute to pathogenicity in one of two ways: (1) by evading and/or suppressing the plant immune response; and (2) by initiating the release of water and nutrients from the host cells into the apoplastic space. This allows the extracellular proliferation of bacteria that leads to necrosis and other symptoms of disease.

The complete genomic analysis of the tomato and arabadopsis pathogen Pseudomonas syringae pv. tomato DC3000 (Petnicki-Ocwieja et al, 2002) has allowed many elements of the TTSS to be identified. The secretion apparatus is comprised of 27 hrp genes of which 8 are conserved amongst all bacteria that possess a TTSS and the bacterial flagellum assembly genes (Alfano & Collmer, 1997; He, 1997). This TTSS leads to the secretion of ~36 effector genes (Collmer et al, 2002; Guttman et al, 2002; Petnicki-Ocwieja et al, 2002). The vast number of which have no known targets and only putative functions.

The HopPtoC gene (AvrPpiC2pto) was one of the effectors identified by Hrp dependant secretion as a type III effector. This led to it being classified as a Hop protein and led to it being renamed. It was originally identified as a P. syringae tomato homologue of an avirulence protein originally found in P. syringae pisi (Vivian & Mansfield, 1993). The HopPtoC gene possesses a putative N-terminal sequence export signal pattern that is common to many of the identified Hop effectors (Petnicki-Ocwieja et al, 2002). Supporting evidence from micro array analysis and RNA blotting confirm that it is a secreted by the Hrp assembly (Collmer et al, 2002; Fouts et al, 2002; Zwiesler-Vollick et al, 2002) as a type III effector though its substrate and role in pathogenesis is unknown. HopPtoC was identified as a cysteine protease through the conserved catalytic triad and other conserved hydrophobic residues (Shao et al, 2002) suggesting that it may play a similar role to YopT or one of the YopT family of cysteine proteases.
Methods

HopPtoC expression and purification was attempted through two different strategies. The first was cloning into pET22b (Novagen Inc., UK) and following a similar protocol to that employed in the purification and structural determination of the related cysteine protease AvrpPhb from *P. syringae* (Zhu et al., 2004). The second strategy was to clone HopPtoC into the solubility enhancing pMalp2x vector (New England Biolabs) as an MBP fusion protein thought proved unsuccessful...

Expression and purification of the HopPtoC-pET22b construct

The 817bp HopPtoC gene was received as freeze-dried plasmid in the pDR4TOPO plasmid provided by Etelechon GmbH. Primers were designed to introduce a 5' NdeI site and a 3' SalI site prior to cloning HopPtoC into pET22b (Novagen Inc., UK).

Forward primer (NdeI site underlined):  
5' GGG CAT ATG ATG ACG ATT GTT AGT GGT C 3'  
Reverse primer (SalI site underlined):  
5' CGC GTC GAC GAA TTC ATC AAT TTT AGA GG 3'

The PCR reaction and subsequent restriction digestion, ligation and transformations were performed following the protocol used in creating the YopT-pET22b construct.

HopPtoC-pET22b was expressed in auto-induction media in the same manner as YopT.

Purification of the insoluble pET22b-HopPtoC protein was attempted through 3 different methods; on the column refolding, batch refolding using the protocol developed in our laboratory (Holloway et al, 2001) and using an optimised refolding...
condition determined from a home-made refolding screen. The optimised refolding was the only method that proved successful.

Inclusion bodies were isolated and solubilised as above. A 24 condition refolding screen based on an amalgam of commercially available was performed to obtain refolding conditions suitable for HopPtoC. 50μl of solubilised protein was added slowly to 1ml of each refolding condition and then left overnight at 4°C. Precipitate was removed using centrifugation and the supernatant containing refolded protein was visualised on an SDS-page gel.

The remaining solubilised protein was then dripped slowly over ~2 hours into 500ml of refolding buffer containing 10mM Tris pH 8.5, 9.6mM NaCl, 0.4mM KCl, 0.5% triton X-100 and 5mM β-mercaptoethanol and left without stirring for >24 hours at 4°C. This was then centrifuged at high speed for 1 hour to remove precipitate and other contaminants (Beckman Avanti J-25) and applied to a 5ml Hitrap column (GE Healthcare) charged with NiSO₄ and equilibrated with 20mM Tris pH8.0, 0.5M NaCl. HopPtoC was eluted as a single peak with 20mM Tris pH8.0, 0.5M NaCl and 0.5M imidazole. The protein was then applied to a Superdex 200 gel filtration column (GE Healthcare) equilibrated with 20mM Tris pH 8.0, 500mM NaCl, 5mM β-mercaptoethanol. The purified sample was concentrated in a 10kDa centrifugal concentrator (Amicon) to a maximum concentration of 2mg/ml.

**Crystallisation of HopPtoC**

Crystallisation trials were performed as hanging drops using 1.75μl of protein with 1.75μl of reservoir solution in the drop, giving a final protein concentration of 1mg/ml in the crystallisation drops. Structure screens I and II (Hampton Research) were used to generate initial crystallisation conditions. At present no optimisation trials have been performed.
Results and discussion

Experimental methods

The initial cloning of HopPtoC into pET22b proved straightforward and high levels of HopPtoC could be observed in the inclusion bodies as seen in figure B14A. Attempts were made to obtain soluble HopPtoC using low temperature growth, minimal media and heat shocking the cultures directly prior to induction but soluble HopPtoC expression was never achieved. Initial purification attempts of HopPtoC centred on the GE Healthcare on column refolding protocol (GE Healthcare).

Figure B14 Expression and purification of HopPtoC

Figure B14. A: The expression of HopPtoC in terrific broth with time points before induction of expression (+0), 3 hours after induction with 0.4mM IPTG at 20°C (+3), and after overnight expression at 20°C (O/N). The insoluble (Insol) and soluble (sol') fractions of the lysed cell pellet are also shown and an arrow marking the 30kDa band corresponding to HopPtoC. B: An SDS gel illustrating the purity of HopPtoC at various stages of purification. These stages include refolding in buffer A6 (10mM Tris pH 8.5, 9.6mM NaCl, 0.4mM KCl, 0.5% triton X-100 and 5mM β-mercaptoethanol), directly after the Ni⁺ affinity Hitrap column and after the final step of size exclusion chromatography.
It involved a slow gradient of urea from 6M to 0M whilst the protein was physically bound to the column. Although this was successful once and led to very pure protein the protein subsequently precipitated out of solution before further purification steps could be taken.

The batch refolding protocol involving arginine has worked well for proteins in our laboratory in the past (Holloway et al, 2001). The use of arginine in the refolding buffer resulted in ion-exchange chromatography being used as the primary purification step as arginine can interfere with histidine based tags binding to affinity columns. The calculated pI of HopPtoC is 6.63 (Gasteiger et al, 2005) and both Q-sepharose chromatography at pH8.0 and SP-sepharose chromatography at pH 5.5 failed to bind significant amounts of HopPtoC. Hence purification via arginine based refolding was abandoned and attempts to optimise an improved refolding condition were begun. The refolding screen was designed based on an amalgam of several commercially available refolding screens but was focused on avoiding refolding conditions containing arginine. The insistence on arginine free conditions was to allow the easy use of the hexahistidine tag on the pET22b vector. Initial screening of HopPtoC in the refolding screen gave positive results in four different conditions: (1) Condition A5 containing 50mM Tris pH 8.5, 9.6mM NaCl, 0.4mM KCl, 0.05% PEG3350, 0.5% triton X-100 2mM MgCl₂, 2mM CaCl₂, 400mM sucrose and 5mM β-mercaptoethanol; (2) Condition A6 containing: 50mM Tris pH 8.5, 9.6mM NaCl, 0.4mM KCl, 0.5% triton X-100 and 5mM β-mercaptoethanol; (3) Condition B1 containing 100mM Tris pH8.5, 240mM NaCl, 10mM KCl, 0.05% PEG3350, 5mM β-mercaptoethanol; and (4) condition D6 containing 50mM Borate pH 9.5, 264mM NaCl, 11mM KCl, 0.05% PEG3350, 0.5% Tween 80, 550mM Guanidine HCl and 5mM β-Mercaptoethanol. The choice was made to focus on condition A6 as it contained the least components and avoided the use of 550mM Guanidine HCl and 440mM sucrose that may have led to downstream problems with the selection of compatible chromatography techniques. After successful refolding conditions were obtained purification through Ni²⁺ affinity and size exclusion chromatography in a similar manner to the protocol for
YopT gave pure homogenous HopPtoC as shown in figure B14B. Concentration of HopPtoC was difficult however as at concentrations >2mg/ml the protein began to precipitate. The very low concentration of HopPtoC in the crystallisation drops may be partly responsible for the lack of any crystals or conditions that may be suitable for optimisation.

Homology modelling with AvrpPhB

HopPtoC is one of 19 ORF’s identified as possessing strong sequence similarity to YopT (Shao et al, 2002). The plant effector molecule AvrpPhB [PDB code 1UKF] (Zhu et al, 2004), was used to construct a homology model of HopPtoC the web-based geno3d server (Combet et al, 2002). The geno3d model server identified the sequence of the HopPtoC as having 20% identity with AvrpPhB and a homology model was able to be built with an RMSD of 2.93Å when superimposed onto the central 50 residues.

The model is however missing the first 57 N-terminal residues. While the core of the molecule superimposes over the core structure of AvrpPhB well many of secondary structural elements have been lost. Whilst AvrpPhB has a seven stranded anti-parallel and the models of both YopT and P76 have 5 stranded anti-parallel β-sheets the model of HopPtoC has only a 3 stranded anti-parallel β-sheet. Although sections of HopPtoC modelled as loop regions do lie where the missing strands are found. As in the other related cysteine proteases there are two lobes composed primarily of α-helices above and plane of the β-sheet, though in the HopPtoC model many of these helices are drastically shortened.
Figure B15 The active site of the homology model of HopPtoC

Figure 15. The active site of the homology model of HopPtoC illustrating the catalytic triad (Cys 72, His 213 and Asp 230), the conserved hydrophobic trio (Pro 231, Phe 229 and Trp 77), the amide of Ser 67 that forms the oxyanion hole and the conserved Glu 235 that forms a hydrogen bond with Trp 72 to stabilise the secondary structure elements around the active site.

There is also a large 25 residue loop between residues 82-107 that is modelled as an unstructured loop but is likely to adopt an additional secondary structural element not present in the AvpPhB structure. The catalytic triad is still arranged in the same spatial organisation as seen in AvrpPhB and the other models although the β-strand containing the catalytic histidine (His 213) has been modelled here as unstructured coil and is shown in figure B15. The other conserved features around the active site including the 3 stacking hydrophobic rings (Pro 231, Phe 229 and Trp 77) are also present and can be seen in figure B4.2 to still directly stack above another one via Van der Waals interactions and the bond between the Trp 77 and Glu 235 that lies
on strand β-3 but is the conserved glutamate residue that is seen in all of the other structures.

**Conclusion**

We have cloned, and using a novel refolding screen, purified the *Pseudomonas syringae* type III effector HopPtoC. Although efforts to crystallise HopPtoC have so far been unsuccessful a homology model based on the related *P. syringae* type III effector showed a conserved secondary structure and a strongly conserved core active site. This core active site was centred on the catalytic C/H/D triad and a series of 3 conserved hydrophobic residues that stack above one another and play a role in maintaining the conformation of the active site.
Toxin B from *E. coli* O157:H7

Aims of cloning & purifying the YopT-like domain of toxB

Enteropathogenic *E. coli* is a major cause of mortality and morbidity worldwide. Its pathogenicity is closely related to its TTSS and a family of putative adhesion factors. ToxB has been identified as an adhesion factor that also plays a role in regulating the TTSS. ToxB has also been identified as possessing a YopT-like cysteine protease domain though its exact role in *E. coli* pathogenesis is unknown. We hope to clone, purify and subsequently determine the structure of the YopT-like domain to gain some insight as to what is substrate specificity might be and to investigate if it shares similar structural features to the rest of the YopT-like cysteine protease family.

An introduction to Enteropathogenic *E. coli*

Enteropathogenic *E. coli* (EPEC) and its close relative enterohemorrhagic *E. coli* (EHEC) are major causes of mortality and morbidity worldwide. EPEC is the leading cause of infant diarrhoea in the developing world (Donnenberg et al, 1997; Hueck, 1998). EHEC strains are more dangerous as they produce shiga-like toxins that lead to hemorrhagic colitis and the potentially fatal haemolytic uremic syndrome (HUS). *E. coli* O157:H7 is the most prevalent strain associated with outbreaks of HUS (Griffin & Tauxe, 1991) and is responsible for over 73,000 cases and 60 deaths in the US annually, with non-O157 EHEC strains responsible for an additional 37,000 cases (Gansheroff & O'Brien, 2000; Mead et al, 1999). The pathogenicity of O157:H7 strains are due to 3 main factors; the ability to produce shiga-like toxins, the ability to induce enterohemolytic activity and the ability to form attaching and effacing (A/E) lesions (Tatsuno et al, 2001; Tozzoli et al, 2005).

In the course of EHEC infection the bacteria are introduced, most commonly by ingestion of contaminated food and water, to the lower intestine where it can adhere
to epithelial cells. Intimate attachment involves the formation of actin-rich pedestals and the effacement of microvilli beneath the bacteria. These pedestals topped by the attached bacteria form the A/E lesions. The formation of the A/E lesions is reliant on a TTSS that secrets several adherence factors and effectors molecules including EspA, EspB, EspD, Tir and intimin. These secreted proteins are all encoded on a single pathogenicity island known as the locus for enterocyte effacement (LEE) (Jia et al, 1995; McDaniel & Kaper, 1997; Elliott et al, 1998). Of these effectors only one, Tir (translocated intimin receptor), has had its structure determined (Luo et al, 2000). During the formation of the A/E lesions the TTSS tranlocates the effectors, apart from intimin, into the host cell cytoplasm. In the cytoplasm instigate a range of host cell signalling instigating amongst other things actin nucleation and the insertion of Tir into the host membrane where it becomes available to bind to intimin and form the A/E lesions (DeVinney et al, 1999; Ebel et al, 1998; Kenny et al, 1997).

In O157 strains of EHEC the presence of an additional 93-104kb plasmid may contain additional virulence factors that increase the pathogenicity seen in the O157 strains (Karch et al, 1987; Schmidt et al, 1996). This plasmid, dubbed pO157, contain several other virulence factors including a catalase peroxidase (katP) (Brunder et al, 1996), a serine protease (espP) (Brunder et al, 1997), enterohemolysin (ehxA) (Schmidt et al, 1994) and toxin B (toxB) (Burland et al, 1998; Makino et al, 1998; Tozzoli et al, 2005).

ToxB, a large putative virulence gene, was initially discovered in 4 different strains under 4 different names before being defined in the 0157:Sakai strain (Tatsuno et al, 2001). The full length ToxB is a very large protein, 3169 residues (361.9kDa) in size, and has been found in a large number of EHEC strains (Stevens et al, 2004; Toma et al, 2004; Tozzoli et al, 2005). The sequence indicates a glycosyltransferase domain in the N-terminal that shows ~23% identity with the large Clostridium difficile toxins A and B (Tatsuno et al, 2001) and the C-terminal has a cysteine protease domain that shows similarity to YopT of Yersinia pestis (Shao et al,
ToxB has also been linked with two other large virulence factors in *E. coli* that it shares 28% identity with EPEC lymphostatin (lifA) and EHEC factor for adherence (Efa1) (Tatsuno *et al*, 2001).

The exact function of ToxB is still unclear but is has now been linked to 3 areas of activity involved in the adhesion to epithelial tissues and the formation of A/E lesions. The main role of ToxB seems to be as an adhesion factor that assists in bacterial adherence to epithelial cells. Several experiments have now shown the strains lacking ToxB, or possessing ToxB mutants, lose much of their ability to bind to host cells. Its function as an adhesion factor is similar to that of Efa-1 and the presences of both in some strains of O157 lead to added redundancy and an improved ability to adhere to host cells (Stevens *et al*, 2004). Its second function seems to involve inhibition of lymphocyte activity. In this regard it functions in a manner similar to that seen for lifA (Klapproth *et al*, 2000; Tatsuno *et al*, 2001; Stevens *et al*, 2004). This anti-lymphocyte ability results in inhibition of interleukin -2 and -4 production along with other pro-inflammatory cytokines and inhibition of lymphocyte proliferation (Klapproth *et al*, 2000). The third and possibly most important role of ToxB is in the regulation of the TTSS effectors encoded by the LEE. ToxB has now been shown to affect the production and secretion of EspA, EspB, EspD and Tir (Tatsuno *et al*, 2001; Stevens *et al*, 2004; Toma *et al*, 2004), though this regulation of LEE effector secretion is also seen for the closely related Efa1 (Stevens *et al*, 2004). The mechanism by which this secretion comes about is still unclear though it is thought that ToxB may be membrane associated and interferes with targeting of the LEE effectors or is involved in sensing extracellular signals required for full production and secretion of the LEE effectors (Tatsuno *et al*, 2001).

The exact mechanism of ToxB action in EHEC virulence is still unclear though it is one of number of large virulence genes that are grouped together as putative adhesions and possible targets for therapeutic vaccines preventing intestinal colonisation by EHEC (Toma *et al*, 2004). It has multiple roles in EHEC
pathogenesis including bacterial adhesion, resistance to lymphocytic activity and regulation of the TTSS effectors involved in the formation of A/E lesions. These multiple roles are also confined to well defined domains on the gene including the glycosyltransferase domain and the YopT-like cysteine protease domain. The cysteine protease domain is shared with Efa-1 (Shao et al, 2002; Stevens et al, 2004) and may play a role in either the adhesion to host cells or the regulation of the LEE effector secretion or an as yet unknown function.
Methods

Expression of the YopT-like domain of EHEC toxB gene was attempted through two different strategies. These were: cloning into pET22b (Novagen) with a hexahistidine tag and cloning into the ligation independent vector ysbLIC based on pET28 (Novagen). The LIC clone proved to be the successful strategy.

Expression and purification of the toxB-ysbLIC construct

The ysbLIC vector was a gift from Dr. Phillip Leonard and was based on pET28 (Novagen) and possesses an N-terminal hexahistidine tag.

The toxB gene fragment was received as freeze-dried plasmid in the pDR4TOPO plasmid provided by Etelechon GmbH. Primers were designed with the appropriate ends for insertion into the ysbLIC vector.

Forward primer (LIC specific ends underlined):

5' CAC CAC CAC CAC ATG TTA AAC CTC GAA GGT CTT ATC AAA AAG 3'

Reverse primer (LIC specific ends underlines):

5' GAG GAG AAG GCG CGT TAG TCG ACT TAG AAT TCG TTA TTT AAC GAT TCA CT 3'

The PCR and subsequent annealing and transformations were performed following the protocol used to generate the P76-ysbLIC construct.

The toxB-ysbLIC construct was expressed in auto-induction media following the protocol used for YopT.

Harvested cell were re-suspended in lysis buffer (20mM Tris pH 8.0, 500mM NaCl, 0.3mg/ml lysozyme) and stirred on ice for 30 minutes. Cells were lysed via
sonication (Soniprep 150) and the inclusion bodies isolated by centrifugation (Beckman Avanti J-25). After cell lysis the inclusion bodies containing toxB were resuspended in 20mM Tris pH 8.0, 0.5M NaCl, 2% triton x-100 and 2M urea and centrifuged at high speed (Beckman Avanti J-25) to remove lipids and other soluble contaminants. This wash step was repeated twice. The washed inclusion bodies were solubilised in 20ml 8M urea, 20mM Tris pH 8.0, 0.5M NaCl, 5mM β-mercaptoethanol and stirred at room temperature for 2 hours.

The solubilised protein was then dripped slowly over ~2 hours into 500ml of refolding buffer containing 10mM Tris pH 8.5, 240mM NaCl, 10mM KCl, 0.05% PEG 3350 and 5mM β-mercaptoethanol and left for >24 hours at 4°C. This was then centrifuged at high speed for 1 hour to remove precipitate and other contaminants (Beckman Avanti J-25).

Refolded toxB was applied to a 5ml Hitrap column (GE Healthcare) charged with NiSO₄ and equilibrated with 50mM tris pH 8.0, 0.5M NaCl. ToxB was eluted as a single peak with 20mM tris pH 8.0, 0.5M NaCl and 0.5M imidazole. The protein was then applied to a Superdex 200 gel filtration column (GE Healthcare) equilibrated with 50mM tris pH 8.0, 500mM NaCl. The purified sample was concentrated in a 10kDa centrifugal concentrator (Amicon) to 10mg/ml and stored at 4°C for crystallisation.

**Crystallisation of toxB**

Crystallisation trials were performed as hanging drops using 1.75μl of protein with 1.75μl of reservoir solution in the drop, giving a final protein concentration of 5mg/ml in the drops. Structure screen I and II (Hampton Research) were used to generate initial conditions with one condition suitable for optimisation; 0.4M sodium potassium tartarate. Optimisation screens have as yet failed to yield crystals suitable for diffraction experiments.
Results and discussion

Experimental methods

The initial cloning of toxB into pET22b proved difficult. Repeated attempts at transformation of the toxB-pET22b plasmid into either cloning or expression strains of *E. coli* failed and several strains and techniques were attempted. This included both heat-shock and electroportation to transform cells and a wide range of cell lines including DH5α, BL21(DE3) (Promega), BL21(DE3)codon+ (Promega), Rosetta (Novagen), JM109 (Promega) and NovaBlue (Novagen). After no expression was observed in the pET22b construct the decision was made not to clone toxB into pMalp2x as it was thought that that might suffer from equally large problems.

The cloning into ysbLIC was straightforward and required no optimisation of the LIC cloning protocol. As with HopPtoC toxB was expressed in large amounts but was only found in the insoluble fraction of the lysed cell pellets as can be seen in figure B16A. The refolding screen was designed based on an amalgam of several commercially available refolding screens but was focused on avoiding refolding conditions containing arginine. The insistence on arginine free conditions was to allow the easy use of the hexahistidine tag on the ysbLIC vector as arginine can interfere with binding to a metal chelation column and result in a lack of protein binding. Initial screening of toxB in the refolding screen gave positive results in two different conditions: (1) Condition B1 containing 100mM Tris pH8.5, 240mM NaCl, 10mM KCl, 0.05% PEG3350, 5mM β-Merceptoethanol and (2) condition C1 containing 10mM Mes pH6.5, 264mM NaCl, 11mM KCl, 0.05% PEG3350, 550mM Guanidine HCl, 2.2mM MgCl₂, 2.2mM CaCl₂, 440mM sucrose. The choice was made to focus on condition B1 as it contained less solutions and avoided the use of 550mM Guanidine HCl and 440mM sucrose as these can lead to downstream problems with the selection of compatible chromatography techniques. Initial purification steps using a Ni²⁺ affinity column yielded little or no success and only μg amounts of pure toxB were obtained. To yield large amounts of pure toxB it was
found that altering the solubilisation buffer from 6M guanidine HCl to 8M urea resulted in a significantly higher amount of toxB being successfully refolded and binding to the Ni$^{2+}$ affinity column as shown in figure B16B. Subsequent purification using gel filtration and concentration for crystallisation was straightforward. The cause for the dramatic improvement in yield, from µg to ~12mg/l, due to changing from guanidine HCl to urea is unknown though it may be involve differences in the stabilisation of the flexible regions of toxB between guanidine HCl and urea.

**Figure B16 Expression and purification of the YopT-like domain of toxB**

*Figure B16. A: Expression of toxB in the initial expression trial in terrific broth with time points before induction (+0), 3 hours after induction with 0.4mM IPTG (+3) and 5 hours after induction with 1mM IPTG (+5). The contents of the soluble and insoluble fractions of the lysed cell samples are also shown. B: The purity of toxB after a Ni$^{2+}$ affinity Hitrap column followed by gel filtration.*

Crystallisation has so far proved unsuccessful. Structure screen I, condition 40 (Hampton) gave a network of large branchlike structures in the drop that grew rapidly. Attempts to optimise this condition have centred on altering the
concentration of sodium potassium tartarate as well as varying the concentration of
the protein drop to slow down growth and produce crystals suitable for structural
studies.

**Sequence similarity with the YopT cysteine protease family**

The 239 residue fragment of toxB shows some sequence similarity to YopT and a
family of cysteine proteases. However unlike the other homolgs of YopT no model
of the YopT-like toxB fragment has been made. All of the web-based modelling
programs, including Geno3D (Combet et al, 2002), 3D jigsaw server (Bates et al,
2001) and CPH models (Lund et al, 2002), failed to build a model of toxB though
both Geno3d and CPH models were able to identify the AvrpPhB structure as a
similar domain. The inability to build a model of toxB may stem from some subtle
differences in the identity of the conserved residues found in the YopT family.
Although a conserved catalytic triad of C/H/D was observed there are some other
important differences. ToxB, and its closely related *E. coli* adhesions LifA and Efa1,
lack the conserved tryptophan, phenylalanine and glutamate and so may have a
slightly different arrangement around their active site. This difference may also
explain the inability of the web-based modelling programs to build a plausible
model of the YopT-like domain of toxB. As this fragment represents only a single
domain of a 320kDa protein it is possible that the rest of the protein may play a part
in binding to its still unknown target.

**Conclusions**

We have successfully cloned and optimised a purification strategy to isolate the 329
residue YopT-like domain of enterohemorrhagic *Escherichia coli* putative adhesion
toxB. In doing so the refolding screen was optimised and alterations to the
solubilisation buffer had a critical impact on protein refolding. Attempts to
crystallise this protein have so far proved unsuccessful. Though similar to YopT the
target, function and nature of toxB is at present still unknown.
General discussion

The YopT family of cysteine proteases

We have cloned, expressed and purified YopT from *Yersinia pestis* and 3 other cysteine proteases that have been shown to be members of a large family of bacterial cysteine proteases (Shao *et al*, 2002). These proteases all share a conserved catalytic triad of C/H/D and several other conserved features around the active site. Homology modelling using the structure of AvrpPhB (Zhu *et al*, 2004) has been used to build models of three of these cysteine proteases: YopT, HopPtoC and the YopT-like domain from P76 while all of the available web-based servers failed to be able to construct a model of the YopT-like domain of toxB.

Superimposition of the 3 models with the structure of AvrpPhB shows a general conserved structure centred around the anti-parallel β-sheet with two lobes comprised of α-helices and coil regions. While there may be differences in the two lobe regions, the active sites are highly conserved and all of the catalytic residues can be seen to overlay atop one another in figure B17. Whilst the conserved histidine of HopPtoC may appear to be adrift from the others it may be due to the poorer quality of the HopPtoC model that failed to identify the β-strand containing His-213. All of the models built also possessed a tryptophan, phenylalanine and proline that stack above one another and form a stack of strong van der Waal bonds that may play a strong role in supporting the active site by keeping the α-helix containing the catalytic cysteine and the tryptophan in close proximity to the β-sheet and the catalytic histidine and aspartate. The importance of these interactions is also partially supported by data from a YopT W146A mutant that lost all its activity (Shao *et al*, 2002).

The biological activity and importance of many of these YopT-family cysteine proteases is still unknown, only two of the 19 ORF’s identified (Shao *et al*, 2002) have had their substrate identified. The GTPase RhoA for YopT (Iriarte & Cornelis,
1998) and RPS5 for AvrpPhB (Simonich & Innes, 1995) though in both cases the catalytic triad and proteolytic activity are required for cytotoxicity (Shao et al., 2002). The importance of the cysteine protease activity in the other ORFs is unknown. Particularly for those that possess the cysteine protease domain as a small domain of a much larger protein such as for P76 and toxB. In both these cases the need or importance of the protease domain is unknown.

**Figure B17 Overlay of the active site of AvrpPhB and the 3 homology models**

---

**Figure B17.** A close-up view of the active site of the YopT family of cysteine proteases. The three catalytic residues are shown coloured according to their protein with YopT (blue), HopPtoC (green), YopT-like domain of P76 (red) and AvrpPhB (yellow). The conserved anti-parallel β-strand is also shown with the 3 hydrophobic residues and the conserved glutamate that act as a scaffold to hold the active site together.
Future work

We have managed to purify all 4 cysteine proteases and crystallisation trials have begun with all 4 of these. In the case of YopT and P76 further optimisation trials are planned in an attempt to improve the crystal quality and yield crystals suitable for diffraction experiments. In the case of toxB and HopPtoC more crystallisation trials are planned in an attempt to obtain initial crystallisation conditions.
11 Appendix C

List of publications


References


computer-based archival file for macromolecular structures. *Eur J Biochem* 80: 319-324


porcine brain cytosol and its characterization as a GTP-binding protein highly homologous to the rho gene product. *FEBS Lett* **243**: 70-76


Carroll SF, Collier RJ (1987) Active site of Pseudomonas aeruginosa exotoxin A. Glutamic acid 553 is photolabeled by NAD and shows functional homology with glutamic acid 148 of diphtheria toxin. *J Biol Chem* **262:** 8707-8711


Cornelis GR (1998) The *Yersinia* Yop virulon, a bacterial system to subvert cells of the primary host defense. *Folia Microbiol (Praha)* **43**:253-261

Cornelis GR (2000a) Molecular and cell biology aspects of plague. *Proc Natl Acad Sci USA* **97**:8778-8783


229


Kapperud G, Namork E, Skurnik M, Nesbakken T (1987) Plasmid-mediated surface fibrillae of Yersinia pseudotuberculosis and Yersinia enterocolitica:
relationship to the outer membrane protein YOP1 and possible importance for pathogenesis. *Infect Immun* 55: 2247-2254


Li M, Dyda F, Benhar I, Pastan I, Davies DR (1996) Crystal structure of the catalytic domain of Pseudomonas exotoxin A complexed with a
nicotinamide adenine dinucleotide analog: implications for the activation process and for ADP ribosylation. *Proc Natl Acad Sci U S A* **93**: 6902-6906


Mead PS, Slutsker L, Griffin PM, Tauxe RV (1999) Food-related illness and death in the united states reply to dr. hedberg. *Emerg Infect Dis* **5**: 841-842


Otwinowski Z. (1993a) *DENZO*: an oscillation data processing programme for macromolecular crystallography. Yale university, New Haven, CT.

Otwinowski Z. (1993b) *SCALEPACK*: Software for the scaling together of integrated intensities measured on a number of separate diffraction images. Yale university, New Haven, CT.


Phan J, Lee K, Cherry S, Tropea JE, Burke TR, Jr., Waugh DS (2003) High-resolution structure of the Yersinia pestis protein tyrosine phosphatase YopH
in complex with a phosphotyrosyl mimetic-containing hexapeptide.

*Biochemistry* **42**: 13113-13121


Rossmann M, Blow D (1962) The detection of sub-units within the crystallographic asymmetric unit. Acta Crystallogr 15: 24-31


Sandkvist M (2001b) Type II secretion and pathogenesis. Infect Immun 69: 3523-3535


molecular weight-immunoglobulin binding proteins. *Microb Pathog* **39**: 159-170


Vriend G (1990) WHAT IF: a molecular modeling and drug design program. *J Mol Graph* **8**: 52-56


sequence of a Staphylococcus aureus exfoliative toxin B plasmid and identification of a novel ADP-ribosyltransferase, EDIN-C. *Infect Immun* **69**: 7760-7771


